A Case-based Reasoning System for Radiotherapy Treatment Planning for Brain Cancer

Rupa Jagannathan

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy
July 2013

Abstract

In this thesis, a novel case-based reasoning (CBR) approach to radiotherapy treatment planning for brain cancer patients is presented. In radiotherapy, tumour cells are destroyed using ionizing radiation. For each patient, a treatment plan is generated that describes how the radiation should be applied in order to deliver a tumouricidal radiation dose while avoiding irradiation of healthy tissue and organs at risk in the vicinity of the tumour. The traditional, manual trial and error approach is a time-consuming process that depends on the experience and intuitive knowledge of medical physicists. CBR is an artificial intelligence methodology, which attempts to solve new problems based on the solutions of previously solved similar problems. In this research work, CBR is used to generate the parameters of a treatment plan by capturing the subjective and intuitive knowledge of expert medical physicists stored intrinsically in the treatment plans of similar patients treated in the past.

This work focusses on the retrieval stage of the CBR system, in which given a new patient case, the most similar case in the archived case base is retrieved along with its treatment plan. A number of research issues that arise from using CBR for radiotherapy treatment planning for brain cancer are addressed. Different approaches to similarity calculation between cases are investigated and compared, in particular, the weighted nearest neighbour similarity measure and a novel non-linear, fuzzy similarity measure designed for our CBR system. A local case attribute weighting scheme has been developed that uses rules to assign attribute weights based on the values of the attributes in the new case and is compared to global attribute weighting, where the attribute weights remain constant for all target cases. A multi-phase case retrieval approach is

introduced in which each phase considers one part of the solution. In addition, a framework developed for the imputation of missing values in the case base is described.

The research was carried out in collaboration with medical physicists at the Nottingham University Hospitals NHS Trust, City Hospital Campus, UK. The performance of the developed methodologies was tested using brain cancer patient cases obtained from the City Hospital. The results obtained show that the success rate of the retrieval mechanism provides a good starting point for adaptation, the next phase in development for the CBR system. The developed automated CBR system will assist medical physicists in quickly generating treatment plans and can also serve as a teaching and training aid for junior, inexperienced medical physicists. In addition, the developed methods are generic in nature and can be adapted to be used in other CBR or intelligent decision support systems for other complex, real world, problem domains that highly depend on subjective and intuitive knowledge.

Acknowledgments

I would like to express my profound gratitude to my supervisor, Professor Sanja Petrovic for giving me the opportunity of working with her on this research project and also for her constant support and guidance on all matters throughout my doctoral studies.

This research would not have been possible without the advice and expertise of medical physicists Angela McKenna and Louise Newton at the City Hospital, Nottingham University Hospitals Trust, NHS, UK, who took time out of their busy schedules to discuss with us the radiotherapy treatment planning process in detail and provide us with brain cancer patient data. Thank you!

Thanks to the Engineering and Physics Science Research Council (EPSRC), UK, for sponsoring this project (Ref No. EP/C549511/1).

I would like to thank the staff and students of the ASAP research group in the School of Computer Science at the University of Nottingham for making my time with them enjoyable and productive, in particular, Nishikant Mishra, for the interesting discussions about our work, Victor Huddlesfield for his help with IT and Christine Fletcher and Deborah Pitchfork for their help with administrative issues.

I would also like to thank Dr Andrew Parkes and Dr Joana Matos Dias for their invaluable suggestions and feedback, which significantly improved the thesis.

I am grateful to my parents C.S. Jagannathan and Sharada Jagannathan, my brother Arun Jagannathan and my parents-in-law Nirmala Rao and Raghavendra Rao for their never ending support, encouragement, good wishes and help.

A very big thank you to my friends Jessica Nicholson, Louise Stephen and Amy Kritz for always being there for me.

Last but by no means least; I would like to thank my husband, Abhijit Rao, not only for his assistance with figures and formatting in this thesis but for all his help, support and encouragement in my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis presents a novel case-based reasoning approach to radiotherapy treatment planning. The radiotherapy treatment planning problem for brain cancer at the Nottingham University Hospitals Trust, City Hospital Campus, Nottingham, UK is investigated. All concepts are tested and validated using real world patient data from the City Hospital. The methodologies designed for the case-based reasoning system can be adapted to be used in decision support systems for similar problems. This chapter introduces the radiotherapy treatment problem in section 1.1. Design considerations of the case-based reasoning system are discussed in section 1.2. Section 1.3 details the motivation driving our research work on radiotherapy treatment planning. The research objectives are outlined in section 1.4. The scope, contribution and layout of the remainder of the thesis are described in section 1.5 and dissemination, including publications of the work done so far is listed in section 1.6.

1.1 Radiotherapy Treatment Planning

Radiotherapy is a type of cancer treatment. In the UK, 4 out of 10 patients receive radiotherapy either alone or in conjunction with other forms of treatment such as surgery or chemotherapy (Cancer_Research_UK, 2010). Radiotherapy is based on the concept that fast proliferating cells are more sensitive to ionising radiation than healthy cells. Therefore, cancerous or tumour cells can be destroyed by subjecting them to high energy x-rays or gamma rays. However, excessive radiation adversely affects all cells, including healthy tissue and critical organs. The aim of radiotherapy is to deliver a tumouricidal dose over the tumour region while minimizing the radiation received by healthy tissue and critical organs (also called organs-at-risk or OAR) in the vicinity of the tumour. Therefore, in radiotherapy treatment planning (RTP), a detailed treatment plan is created for each patient that describes exactly how a patient should be irradiated.

The protocol and guidelines for treatment planning differ widely based on the type and location of the cancer. The research presented in this thesis focuses on brain cancer cases. The close proximity of the tumours to vital organs such as the spinal cord, brain and sensory organs, such as the eyes, makes radiotherapy treatment planning especially challenging since overdosing the organs at risk can fundamentally impair the patient's quality of life. On the other hand, if the tumour is not successfully treated it can prove fatal.

Challenges of Treatment Planning

Currently in many hospitals, including our project collaborator, the Nottingham University Hospitals Trust, NHS, Nottingham City Hospital Campus, treatment planning is done manually using a trial and error approach called forward planning. The planning parameters are adjusted iteratively to achieve an acceptable dose distribution of irradiation. Generating a good treatment plan can take from a few hours to a few days in complicated cases and requires the expertise of one or more experienced medical physicists. In addition, with advances in technology, the complexity of treatment planning has increased. The literature on automated treatment planning systems mainly focuses on complex mathematical models or rule-based inference engines to generate an optimal treatment plan. Some of the problems, however, that these systems face are as follows:

- Radiotherapy treatment planning is a complex, computationally expensive problem, in particular, when treatment plan generation involves calculating the dose and its distribution of a potential treatment plan (Petrovic et al., 2011). According to Meyer et al. (2005) treatment planning can take from a few hours to several days. Schreibmann et al. (2003) state that the the optimisation of six beam angles with a resolution of 5° requires about $3*10^{20}$ computations.

- The treatment plan usually offers a compromise or trade-off between tumour control and side effects due to radiation of healthy tissue. This trade-off is not trivial to model using mathematical formulations or rules. (Vineberg et al., 2002).

At the City Hospital, a treatment planning system called Oncentra (Nucletron, 2011) is used. Oncentra has the functionality to generate treatment plans based on the patient images; however, this is not used by the treatment planners. Currently, Oncentra is used to view the the dose distribution resulting from treatment plan parameter configurations. The plan parameters are iteratively evaluated by visualising the resulting dose distributions in a manual trial and error fashion. The treatment planning procedure is described in detail in section 5.1.2. Discussion with staff at the City Hospital revealed further challenges of treatment planning and why the medical physicists are reluctant to use Oncentra to generate treatment plans:

- The treatment plan has to take into account hospital policies, capabilities and guidelines.
- At times, the plans generated by the existing treatment planning system are not clinically acceptable, which is part of the reason why medical physicists prefer to manually generate treatment plans. There are also physical constraints about beam placement, for instance, the radiation beams should not be directed at the patient from directly underneath the treatment bed. These kind of practical constraints are often not taken into account by existing treatment planning systems.
- According to the medical physicists at the City Hospital, it is difficult to see how a treatment plan has been derived using current automated treatment planning systems, which often work like a "black box", where the planner is only aware of the inputs and outputs. This can be an issue, in general, with clinical automated decision support systems, which work like a black box (Anooj, 2011, Kawamoto et al., 2010). This reduces the confidence of the user in the system.

However, the largest drawback of these systems is that they completely disregard the experience gained by expert physicians. After years of practice, many senior oncologists and medical physicists have gained a lot of empirical knowledge about which treatment plan configurations are suitable for particular cancer cases (Kalet and Paluszynski, 1990, Shepard et al., 1999). They are often aware of what works well even when the exact underlying causes are not entirely understood. The aim of our research is to create a decision support system that incorporates the wealth of experience possessed by experts, by applying case-based reasoning to generate a good treatment plan for a cancer patient.

1.2 Case-Based Reasoning

In case-based reasoning (CBR), problems are solved based on the solutions of similar past problems (Kolodner, 1993). The case base consists of past cases, which are stored along with their solution. Given a new case, the CBR system calculates the similarity between the new case and each case in the case base and then retrieves the most similar case. The solution of the retrieved case is used in the solution of the new case. Usually, the solution of the retrieved case is adapted to fit the specific requirements of the new case. The new case along with is adapted solution can be stored in the case base for future retrieval.

Advantages of CBR

The advantages of applying CBR to radiotherapy treatment planning are as follows:

- CBR captures subjective knowledge & experience, gained by medical physicists over many years. In the City Hospital, only senior medical physicists are allowed to perform treatment planning and usually, two or more medical physicists are involved in generating treatment plans for each patient. Further, each generated treatment plan is carefully examined by the consultant oncologist before it is approved for treatment. These steps not only help in the validation of the generated treatment plans but also improve the understanding and skill of planners to create good plans.
- Since the treatment plans are not generated from scratch, computation time is low.
- CBR can consider previous successes, errors and failures.
- The institution's capabilities and preferred protocols are inherently present as part of the cases in the case base. Further, if the protocols or the guidelines of an institution change, this can be easily incorporated due to the modular and flexible nature of

cases, for instance by representing them as an additional attributes or incorporating rules specifying the changed parameters.

- Knowledge from several experts can be pooled. This is useful, as in reality, there exists a possibility that individual experts might generate slightly different treatment plans for the same patient. By including knowledge from a number of different experts or institutions, the CBR system could retrieve several treatment plans for the same target patient case to allow the user to compare or fuse knowledge from different plans.
- CBR can also be used in unusual or complex situations.
- The cases are naturally available and are stored as patient data.
- Since CBR models human reasoning, it is very easy to provide an explanation of how a solution has been derived. This increases the confidence of the user in the system.

Disadvantages of CBR

CBR is not suited to all domains or problems and when approaching a problem, the motivation of using CBR and its suitability in a domain has to be carefully considered. The points below detail the common disadvantages of CBR:

- Limited applicability: CBR is highly suited to problems, where the underlying theory is very complex or not entirely understood and where reasoning depends on subjective knowledge or experience of experts. However, when a problem can be described and solved using precise mathematical formulations or rules, CBR might lead to less accurate or inferior results in comparison.
- CBR works best when knowledge can be encoded in the form of cases. The case attributes have to be carefully chosen based on their relevance to the solution of a problem. In many situations, the cases can be generated automatically (though this requires an initial effort in automating case generation or knowledge encoding) though in some situations, they have to be created by hand or with the help of a domain expert. In our CBR system, the only inputs required to the system are the patient image files, which are used by medical physicists in treatment planning; hence no additional information is required from experts though the input data has to be pre-processed to formulate cases.

- If a problem domain requires a large number of cases to cover all solutions, CBR systems can have large storage requirements. Also, if the case base is large, retrieval time can be time consuming. However, a proper design of the system or maintenance of the case base, such as removing redundant cases or combining information in case clusters can reduce storage requirements and retrieval time (Haouchine et al., 2007, Lawanna and Daengdej, 2010). Currently, in the developed CBR system for radiotherapy treatment planning the size of the case base does not require large storage space or long retrieval times. In the future, case base maintenance techniques will be employed to deal with storage requirements and retrieval time.
- The retrieval stage of a CBR system depends highly on the cases available in the case base. If the problem and solution coverage of the case base are limited, or if the available case solutions are of sub-optimal quality, the performance of the system is negatively affected as well. The CBR system for radiotherapy treatment planning depends highly on the quality of the treatment plans of the cases generated by medical physicists. Currently, we are assuming that the treatment plans of cases in the case base generated by medical physicists represent a good solution with respect to a particular patient case. However, we are aware of the fact that this assumption might not necessarily always be valid. The next stage of CBR design, i.e. adaptation of retrieved treatment plans, will work not only towards customising treatment plans to fit the specifics of the new patient but also towards improving the treatment plans based on factors other than the existing treatment plans in the case base.

1.3 Motivation

Currently, radiotherapy treatment planning at the Nottingham City Hospital is done manually using a trial and error approach. First, the medical physicist generates a treatment plan based on their experience that is deemed potentially suitable for the current patient. Then the dose distribution of the treatment plan is evaluated. If any dose violations are found a new treatment plan is generated. This process is repeated till a satisfactory dose distribution is obtained. This is a time consuming process that can take from a few hours to a few days and requires the expertise of one or more experienced medical physicists. Though automated treatment planning systems for radiotherapy treatment planning have been widely discussed in the literature they are less commonly applied in practice for the reasons outlined in section 1.1.

The motivation driving this work is to design a CBR system, which overcomes the problems of existing automated treatment planning systems and aids medical physicists with treatment plan generation.

Since CBR systems are based on the concept that similar cases have similar solutions, the quality of a CBR system depends heavily on the design of the similarity measure in the retrieval stage. The commonly used weighted nearest neighbour similarity measure (wNN) (Cover and Hart, 1967) matches each attribute in the target case to its corresponding attribute in the archive case. The aggregate similarity is given by the weighted sum of the individual attribute similarities. Identifying the attribute weights, which assign the relative importance of each attribute with respect to the solution, is imperative and much of the work in this research focuses on determining global and local attribute weights, which vary depending on the attribute values of the target case. The local weights are assigned using rules learnt from the system during the training phase.

The development of decision support systems using real world data often suffers from the problem of having initially insufficient data available making parameter training difficult or unreliable. In this work, strategies have been designed that take into account the size and content of the case base available for training.

A drawback of the wNN similarity measure is that it does not take into account the distribution of individual case attribute similarities. If the attribute similarity distributions show wide variation, there is a possibility that the numerical attribute similarity values are not comparable and cannot be directly combined into an aggregate similarity value for the case. Further, the aggregate similarity is always a linear function of the individual case attribute similarity values. In other words, even a low attribute similarity value contributes to the aggregate similarity. However, at times it is beneficial that a low similarity value is heavily penalized to ensure that case solutions unsuitable to the target case are not retrieved. To introduce this kind of non-linearity in the similarity computation and to take into account the distribution of attribute similarities, a similarity measure based on fuzzy set theory has been designed.

Most CBR systems aim to find a single optimal combination of attribute weights that ensures that the case with the most suitable solution to the target case is retrieved. In some cases, however, some attributes are more relevant with respect to particular parts of the solution rather than the entire outcome. Our experiments have shown that if we split the solution into separate parts we can optimise the retrieval mechanism with respect to each solution part. We have therefore designed a two-phase retrieval system that uses different weight settings to retrieve two different parts of the solution.

A common problem with case-based reasoning systems is that the collected cases can be incomplete. Incomplete cases are often discarded but this not only reduces the size of the case base but also wastes potentially useful information present in the existing data of these cases. We have studied the use and the design of an imputation method framework to estimate missing values in a case-based reasoning system.

1.4 Research Objectives

The author's research and this thesis focus on the following salient issues:

- Analysis of the radiotherapy treatment planning problem, in particular, as applied to brain cancer.
- Investigation of radiotherapy treatment planning at the City Hospital with a focus on their guidelines, policies and requirements of an automated treatment planning system.
- Review of existing methods in radiotherapy treatment planning and the advantages and disadvantages of case-based reasoning systems in healthcare problems.
- Analysis of brain cancer patient data obtained from the City Hospital.
- Identifying and extracting relevant data from DICOM image patient files.
- Design and development of a CBR system based on the radiotherapy treatment planning at the City Hospital using CBR. Our work focusses on designing the retrieval mechanism of the CBR system including
    1) Attribute selection and weighting.

2) Design of a similarity measure suitable to the problem domain and its data and comparison with the commonly used wNN similarity measure.
3) Validation of the developed CBR system.

- Design of a framework for imputation of missing values in a CBR system.

1.5 Scope, Contribution and Layout of the Thesis

This section outlines the scope of our work, the contributions made in the fields of radiotherapy and CBR and the layout of the thesis.

1.5.1 Scope

The aim of this research is to design and develop a CBR system for radiotherapy treatment planning for brain cancer patients. In particular, the work concentrates on the retrieval stage of the CBR system. The goal is to develop a prototype CBR system that, given a new patient case, is capable of retrieving a similar case from the case base whose solution is relevant to the new case. This would provide a good starting point for adaptation, which is the next stage of a CBR system. Often the retrieved case is not exactly the same as the target case with respect to the attribute values. The aim of adaptation is to adjust the retrieved treatment plan to fit the specific requirements of the target case. Methods to carry out adaptation include adaptation by a domain expert, by using a case base of adapted cases or by using rules. However, while adaptation is an important component of a CBR system, its implementation is not within the scope of this thesis.

All experimental results have been validated using real world brain cancer patient data from the City Hospital. However, it has to be noted that the validation of the experimental results is based on the existing treatment plans generated by medical physicists with the assumption that these treatment plans represent good treatment plans for a case. It is possible that for every target case superior or equally good treatment plans with different parameters could be generated. However, in practice, this possibility is difficult to account for as it requires evaluating each solution of the CBR by a domain expert. Expert validation is planned for the future, but does not fall within the scope of the research work presented in this thesis. The imputation of missing values for a CBR system has been validated using prostate cancer patient data from the City Hospital.

1.5.2 Contribution

The contribution of this work can be summarised as follows:

- CBR has not been applied previously to radiotherapy treatment planning for brain cancer. In addition, in general, CBR has hardly been applied to radiotherapy treatment planning. A decision support system based on CBR would overcome the problems of existing approaches and have many advantages as outlined in sections 1.1 and 1.2 and in Chapter 2:.
- Investigation and comparison of similarity measures, namely:
    1) Weighted nearest neighbour (wNN) similarity measure
    2) Novel non-linear, fuzzy similarity measure that takes into account the distribution of attribute similarity values across the case base and allows weighting of similarity and dissimilarity between cases. The fuzzy membership functions are generated based on attribute similarity values found across the case base. We have also investigated the novel use of local fuzzy membership functions defined for each target case.
- Case attribute weighting:
    1) Determination and analysis of global attribute weights
    2) Design of a novel local attribute weighting mechanism that assigns attribute weights based on the attribute values of the target case. The local weights are assigned using rules that are learnt by the CBR system during the training stage and selected using rule evaluation measures. A novel rule evaluation measure is introduced that gives an assessment on the reliability of feedback obtained during the training phase based on the content of the case base.
    3) In addition, a novel method of generating feedback about the retrieval performance during the training phase for continuous solution parameters is introduced.

- Development of a multi-phase retrieval system in which each phase uses attribute weights in the similarity measure that are customised with respect to parts of the solution.
- Utilization of textual information in the patient DICOM images: This method eliminates the need for complex image processing tools. The information extracted from the DICOM images is used to compute the case attributes.
- The validation of the retrieval mechanism and the developed concepts shows that CBR is a feasible methodology for radiotherapy treatment planning decision support systems. The success rate shows that the treatment plans of the cases retrieved by the CBR system for new brain cancer patients provide a reasonable starting point for adaptation.

1.5.3 Thesis Layout

The remainder of the thesis is organized as follows:

Chapter 2:

The radiotherapy treatment problem is described in detail in Chapter 2. This chapter also includes a comprehensive literature review of approaches used in radiotherapy treatment planning.

Chapter 3

CBR and related concepts are described in this chapter. Key concepts, advantages and challenges of CBR systems are outlined. Common applications of CBR are listed, in particular, in health care. The similarity measure, global and local attribute weighting, fuzzy set theory and validation concerns are discussed. Finally, the problem of missing values in CBR systems is described.

Chapter 4

This chapter deals with the development of a framework for the imputation of missing values in a case-based reasoning system. A simple filtering imputation technique is introduced and compared to commonly used imputation methods.

Chapter 5

This chapter gives an overview of the architecture of the CBR system. First, the treatment planning process at the City Hospital, on which the CBR system is based, is discussed. The patient DICOM image data is described in detail, which supplies the input data to the system. Case representation and case attributes are explained. Finally, an overview of the retrieval mechanism is supplied. The filtering mechanism which selects a subset of cases from the case base given a new case is explained.

The following chapters discuss the retrieval mechanism and related concepts in detail. All methods are validated and compared using real world test cases. The results are analysed and discussed.

Chapter 6

This chapter discusses validation concerns and outlines different validation methods including their applications, advantages and disadvantages. The random base line error of the system is presented, which serves as the basic starting point for improving the performance of the retrieval mechanism.

Chapter 7

The weighted nearest neighbour (wNN) similarity measure is introduced in this chapter. The determination of global attribute weights is explained. The local attribute weighting scheme using rules to assign attribute weights is described in detail, including the rule evaluation measures used and designed. The reliability and accuracy of the feedback about the performance of the retrieval mechanism during the training phase is discussed and an alternative method of obtaining performance feedback is introduced.

Chapter 8

Chapter 8 discusses the fuzzy, non-linear similarity measure that takes into account the distribution of attribute similarity values across the case base. The use of both global and local fuzzy membership functions is discussed.

Chapter 9

The two-phase retrieval mechanism is explained in this chapter, which allows customisation of the retrieval mechanism with respect to the part of the solution determined in that phase.

Chapter 10

This chapter summarises the major findings and conclusions of this work. The test results about the performance of the retrieval mechanism and related concepts are presented and compared. The contribution to the fields of radiotherapy treatment planning and CBR is discussed in detail. Interesting avenues of future research are proposed.

Appendix A: DICOM RT Image Files

In this appendix, snapshots of diagrams of DICOM files, from the Radiotherapy RT DICOM supplement to the DICOM standard, are presented.

Appendix B: Medical Dictionary

This appendix contains the descriptions of relevant medical terms, which are used throughout the thesis.

1.6 Dissemination of Results

This section lists the journal and conference papers, where our work has been published. The research work has also been presented at a number of workshops and university seminars.

Journal Papers

- Jagannathan, R., Petrovic, S., McKenna, A., Newton, L. (2012) *A Novel Two Phase Retrieval Mechanism for a Clinical Case-Based Reasoning System for Radiotherapy Treatment Planning*. International Journal on Artificial Intelligence Tools. 21(4) pp. 1240017.

Reviewed Conference Papers

- Jagannathan, R. and Petrovic, S. (2012). *A Local Rule-based Attribute Weighting Scheme for a Case-based Reasoning System for Radiotherapy Treatment Planning*. Computer Science Case-Based Reasoning Research and Development. Lecture Notes in Computer Science, (7466), pp. 167-181, International Conference on Case-Based Reasoning, September 2012, Lyon, France.
- Jagannathan, R., Petrovic, S., McKenna, A., Newton, L. (2010). *A Fuzzy Non-linear Similarity Measure for Case-Based Reasoning Systems for Radiotherapy Treatment Planning*. Proceedings of the 6th IFIP Conference on Artificial Intelligence, Applications and Innovations, October 2010, Larnaca, Cyprus.
- Jagannathan, R. and Petrovic, S. (2009). *Dealing with Missing Values in a Clinical Case-Based Reasoning System*. Proceedings of the IEEE International Conference on Computer Science and Information Technology, ICCSIT 2009, pp. 120-124.

Short Paper

- Jagannathan, R., Petrovic, S., McKenna, A., Newton, L. (2011). *A CBR System for Radiotherapy Treatment Planning*. Operational Research Annual Conference (OR53).

Abstract

- Jagannathan, R., Petrovic, S., McKenna, A., Newton, L. (2010) *A Case-based Reasoning System for Radiotherapy Treatment Planning for Brain Cancer*. 24th European Conference on Operational Research (EURO-24).

Related Work

- Jagannathan, R., Petrovic, S. *Validation Concerns of Case-based Reasoning Systems For Real World Health Care Applications"*. Expert Systems With Applications. (*In preparation*).

Seminars and Workshops

- "A Case-based Reasoning System for Radiotherapy Treatment Planning for Brain Cancer – Research" presented at the Doctoral Consortium, International Conference on Case-based Reasoning (ICCBR), Lyon, France, 3rd September 2012.

- "A Two-phased Case-based Reasoning System for Radiotherapy Treatment Planning" presented to the Automated Scheduling, Optimization and Planning Research Group, School of Computer Science, University of Nottingham, 1st March, 2011.

- "A Case-based Reasoning System for Radiotherapy Treatment Planning – Brain Cancer" presented to the Automated Scheduling, Optimization and Planning Research Group, School of Computer Science, University of Nottingham, 2nd March, 2010.

- "A Framework for the Imputation of Missing Values in a Clinical Case-based Reasoning System" presented to the Automated Scheduling, Optimization and Planning Research Group, School of Computer Science, University of Nottingham, 22nd January 2009

- "Missing Values in a Clinical Case-based Reasoning System" presented at the workshop on LANCS Healthcare Modelling PhD Symposium, Cardiff University, 18th- 20th January, 2009.

- "Methods in Radiotherapy Treatment Planning" presented at the workshop on Radiotherapy Planning and Scheduling, Coventry University, 27th February, 2008.

Chapter 2

Approaches to Radiotherapy Treatment Planning

Cancer covers a group of diseases, which are characterised by the abnormal growth of the cells of bodily tissue. Cancerous cells divide and grow uncontrollably thereby forming tumours. Benign tumours are localised whereas malignant tumours spread to other parts of the body. The most common treatment methods for cancer include surgery, chemotherapy and radiotherapy.

In radiotherapy, the areas containing the tumour are subjected to ionising radiation, either in the form of high energy x-rays or gamma-rays. Though fast proliferating cancer cells are more sensitive to radiation than healthy tissue, all cells are damaged by prolonged exposure to radiation. Radiotherapy, therefore, requires a trade-off between achieving tumour control and avoiding normal or healthy tissue complications. Healthy tissue complications are site dependent. For instance, for head and neck and brain cancer common complications following radiotherapy include salivary hypo function, problems involving dry mouth, the sensory organs and the teeth (Parliament et al., 2004). A major aim of radiotherapy therefore lies in focussing the radiation on the tumour while minimising the radiation received by healthy tissue and critical organs (also called organs-at-risk or OAR) in the vicinity of the tumour.

The radiation is produced by a machine called a linear accelerator (linac) (Levitt, 2008) shown in Figure 2. 1. The linac is mounted on a gantry that rotates around the patient lying on the patient bed. The radiation is focussed into a narrow beam using a collimator. Multi-leaf collimators have the capability of further shaping the radiation beam to conform to the tumour.

Figure 2. 1 Linear Accelerator (Trilogy, 2008)

Since the radiation beam has to traverse through healthy tissue before it reaches the tumour site, the radiation is usually applied sequentially using several beams from different angles, which intersect at the tumour. This reduces the radiation received by the healthy tissue in the path of the beams, but the total radiation to the tumour remains constant. Figure 2.2 shows a brain tumour with the radiation dose received described with the help of isodose lines. The orange dot, for instance, at the isocentre of the tumour, where the beams intersect, receives 110% of the total dose. The green line represents the area that receives 95% of the total dose. The OAR under consideration is the eye. The blue isodose line around the eye denotes that the eye would receive 20% of the radiation.

The medical dictionary found in the appendix of this thesis provides a list of medical terms used in this thesis and their definitions.

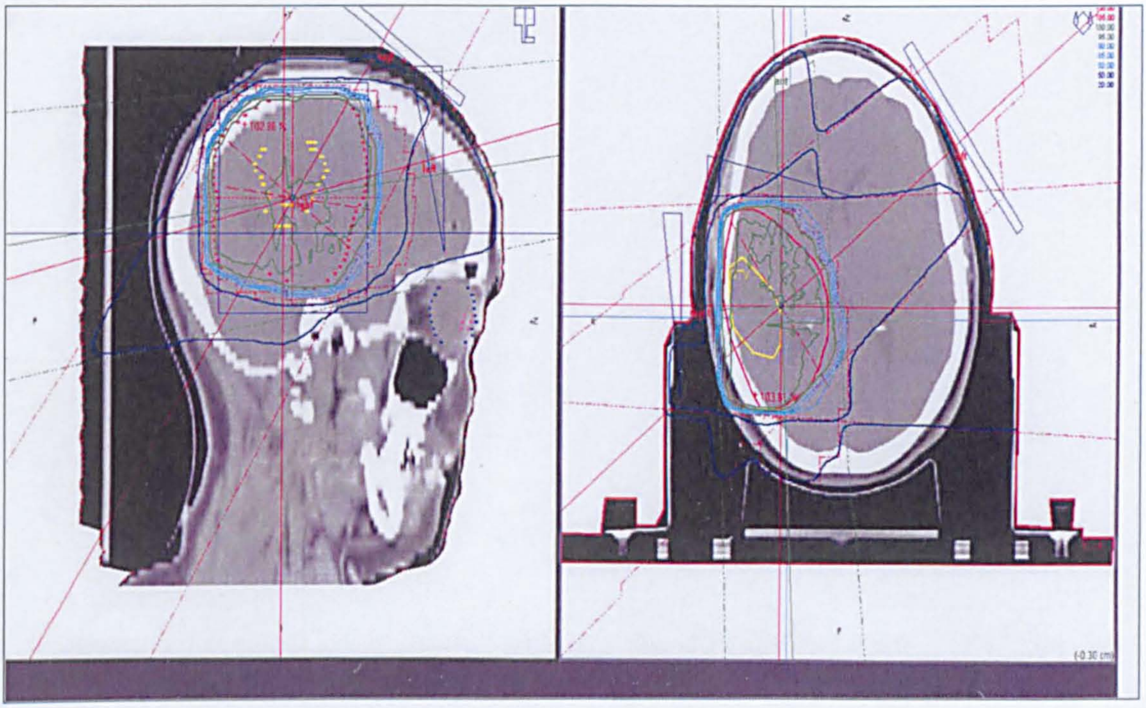


Figure 2.2. CT image showing isodose lines and the eye as OAR

2.1 Treatment Planning Aims and Challenges

The aim of radiotherapy treatment planning is to generate a treatment plan, which describes how the radiation is applied in order to achieve maximum tumour control while minimizing the radiation received by healthy tissue and OARs. It is important to note that maximum tumour control does not necessarily mean applying the maximum dose possible. Hamacher et al (2002) observe that an ideal dose distribution, in which no healthy tissue and OAR are overdosed and the tumour tissue is not underdosed, is usually not feasible. Therefore the goal becomes to minimise the impact of dose violations to the most sensitive OAR. Further, the dose that can be applied to the tumour is limited since usually tumour cells and healthy cells are interspersed in the tissue so the radiation should be appropriate to kill the tumour cells without killing the healthy cells (Holder, 2004). After discussion with staff at the City Hospital, it transpired that treatment goals also depend on other factors:

Age of Patient/Stage of Cancer

Younger patients are often treated more aggressively whereas in older or terminally ill patients more importance is given to the quality of the remaining life rather than completely removing the tumour. For example, in the case of head, neck and brain cancers, in a younger patient, the eyes might be sacrificed in order to maximise tumour control (in the hope of possibly repairing them later) whereas in older patients, the treatment would be tailored to try to save the eyes even at the expense of tumour control.

Organs-at-Risk

Different OARs respond differently to radiation (Holder, 2004). For instance, the liver will fail if particular sensitive regions of the liver receive an excessive dose. It can, however, function reasonably well in the event of the entire liver receiving a low radiation dose. These kinds of organs are known as *chain organs* as opposed to *rope organ*, which will fail when the entire organ receives even just a low dose.

The general treatment goals can be described as follows:

- Uniform, homogenous coverage of tumour or planning treatment volume (PTV)
- Avoiding hot spots (regions of overdosing in the OAR) and cold spots (regions of underdosing in the PTV)
- Conformance of beams to PTV (as close to the edge of the PTV as possible)
- Avoidance of radiation to healthy tissue and organs at risk

The decision variables or the treatment plan parameters mainly concern the dose and the beam configuration and include:

Total radiation dose:

The radiation dose depends on several factors based on the location and type of the tumour. The total radiation is applied in fractions as a course over a specific duration of time. Since healthy tissue recovers from radiation faster than tumour cells, applying the radiation in sessions allows the healthy tissue to recover while still maintaining tumour control.

Beam Configuration

The beam configuration is influenced by the specific patient anatomy, which includes the location of the tumour, the shape of the tumour and the OARs in the vicinity of the tumour.

- Number of beams: In 3D conformal radiotherapy, the radiation is applied using a number of beams, which are applied sequentially from different directions in order to reduce the total radiation dose to the healthy tissue in the path of the beams. The number of beams can vary from 2-9 beams but is often limited (to a constant number of three or four) according to hospital policy for the sake of ease and effectiveness of implementation (Schreibmann et al., 2003). A larger number of beams allows more flexibility in planning and closer conformance of the radiation to the tumour. However, the disadvantage of using a larger number of beams is that with the beam number, the treatment time increases and also the process of planning becomes more complicated. Further, treatment using fewer beams reduces the probability of the patient moving during treatment, thereby improving its precision.
- Beam weights: The beam weights denote the intensity of each beam. Given a total prescribed radiation dose that the tumour has to receive, the individual beams can be weighted differently to make up the total dose.
- Angle of beams: The beams are applied at an angle to ensure that they conform to the tumour volume while avoiding as much as possible the organs at risk. The beams can be applied coplanar (all lying in one plane) or non-coplanar. Non-coplanar beams can produce superior treatment plans; however, they substantially increase the computational effort and increase the complexity of planning. At the City Hospital, 3D conformal radiotherapy is carried out using non-coplanar beams for brain cancer.
- Wedges: Wedges are metallic wedge-shaped blocks, which are placed in the path of the beam to attenuate the radiation.
- Multileaf collimator settings: The leaves of the collimator shape the radiation beam. When multi-leaf collimators are used, the process of treatment planning is often referred to as *Conformal Radiotherapy*.

Figure 2.3 shows a schematic diagram of the region to be treated including the radiation beams and wedges. The GTV denotes the Gross Target Volume, which is the volume of tissue that contains the cancerous cells (Burnet et al., 2004). The CTV or Clinical Target Volume includes the GTV and a margin, which accounts for a possible spread of the tumour cells. The PTV or Planning Target Volume includes the CTV and a margin, which allows for uncertainties in planning and delivery of the radiation. Usually the radiation is made to conform to the area of the PTV. For planning purposes, usually it is sufficient to consider only the PTV (since it contains the GTV, anyway). When PTV coverage is not possible, then the GTV is considered.

The radiation beams are applied sequentially, with or without wedges and intersect to focus the radiation on the PTV.

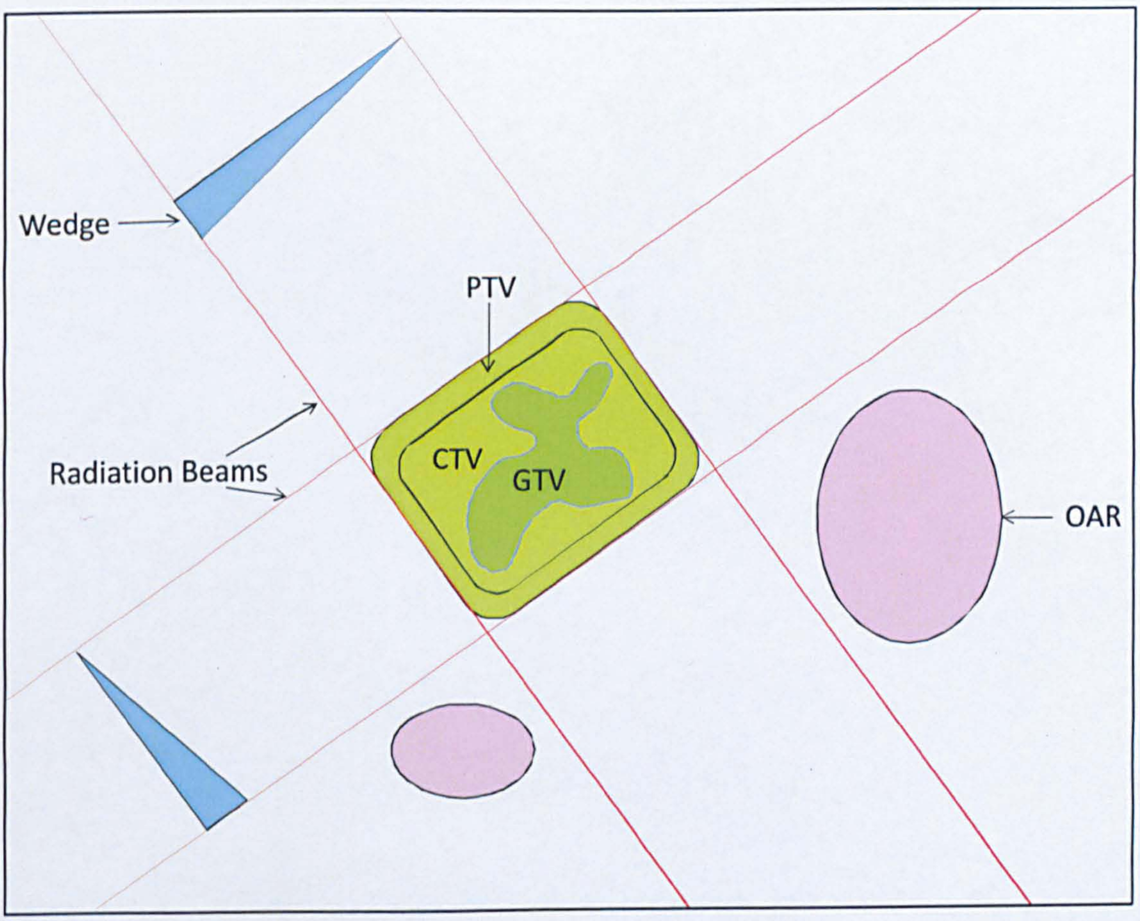


Figure 2.3. Schematic showing tumour volumes (GTV, CTV and PTV), radiation beams, wedges and OAR

2.2 Types of Treatment Planning

Radiotherapy treatment planning is usually carried out by medical physicists and oncologists by using an iterative process of trial and error called *forward planning*, in which the generated treatment plan is modified till an acceptable dose distribution is obtained. The general process of forward planning is outlined as follows:

1) The oncologist views the computed tomography (CT) patient images and outlines the GTV, CTV, PTV and OAR on the images using the treatment planning system. The treatment planning system used by the hospital for viewing patient images, placing beams and wedges and evaluating the resulting dose distribution and should not be confused with the CBR system that we are developing, which generates treatment plan parameters based on the outlines of the PTV and OAR drawn by oncologists on the CT images. In the City Hospital, the treatment planning system Oncentra (Nucletron, 2011) is used for the above purposes but not to compute the treatment plan parameters, which are manually determined by the medical physicists using a trial and error method. The fitness of the solution, i.e. whether a treatment plan is acceptable or not, is determined by the medical physicists after viewing the dose distribution displayed by Oncentra.
2) Then the beams are placed so that they intersect at the isocentre (usually, the centre of the tumour). The other parameters, such as beam weights, wedges and multileaf settings, are determined.
3) The dose distribution is calculated by the treatment planning system and evaluated with respect to the planning goals.
4) The beam configuration and the other parameters are modified.
5) Steps 2-4 are repeated till the dose distribution is satisfactory.

The outlined procedure is a time consuming and inefficient process. For this reason, *inverse planning* is becoming more popular. In inverse planning, the consultant oncologist specifies the dose distribution objectives to the tumour and the organs at risk (also called the prescribed dose) based on the available patient information and the patient's image files. For certain types of cancer, such as head and neck or brain cancer, many institutions use fixed dose limits to the tumour and organs at risk. The treatment planning parameters are then determined to achieve a dose distribution that is as close as possible to the prescribed dose objectives. Bär et al. (2003) did a comparative study of forward and inverse treatment planning for head and neck cancer and found that better tumour coverage and OAR sparing could be achieved using inverse planning. Oldham et al. (1995) did a comparative study of forward planning carried out by a human planner and inverse planning using an automated treatment planning system. They found that inverse planning produced superior treatment plans exhibiting higher tumour control and lower normal tissue complications probability. The biggest advantage they found was that the automated treatment system was able to generate the treatment plan about 20 times faster than the human planner. Hamacher et al. (2002) stated that inverse planning could produce superior plans to forward planning but was mathematically more challenging.

3D Conformal Radiotherapy, which allows the radiation to conform to the PTV more closely, provides the capability to treat more complicated cases, however at a computational expense. An advanced treatment modality known as Intensity Modulated Radiation Therapy (IMRT) is quickly becoming very popular. In IMRT the beam is divided into a large number of beamlets. The intensity of each beamlet can be modified allowing greater control over the shape of the beam, which leads to better tumour conformity and organ sparing. Another advantage of IMRT is that optimised treatment plans can be generated that use fewer number of beams, thereby reducing treatment time (Cozzi et al., 2004) (Nutting et al., 2001). IMRT is becoming more popular and our project collaborators are planning to acquire an IMRT system in the near future as well. However, according to Webb (2001) only about 30% of all radiotherapy cases require IMRT and advocates 3D Conformal Radiotherapy with the correct selection of non-coplanar beams.

With advances in state-of-the-art radiotherapy technology and treatment modalities such as IMRT or 3D Conformal Radiotherapy, thousands of clinically feasible plans are possible (Jain and Kahn, 1993), which can result in long computation times of treatment plans (Good, 2012), in particular when non-coplanar beams are used or many OARs have to be considered (Meyer et al., 2005). Planning becomes even more complicated when the OARs are of different sizes and radio sensitivities as is the case with head, neck and brain cancer. To assist manual planning and to be able to exploit the advances in technology to produce superior treatment plans, the use of decision support systems and automated treatment planning systems has been widely researched. The following chapters give an overview of the commonly used approaches.

2.3 Approaches to Radiotherapy Treatment Planning

The approaches to radiotherapy treatment planning (RTP) found in the literature can be classified broadly into numerical optimisation methods and knowledge based methods. Optimisation methods include deterministic linear and non-linear mathematical programming models, which can be solved using exact or heuristic methods. Examples of knowledge based methods include rule based systems and case-based reasoning systems. Table 2. 1 lists some of the methods widely used in radiotherapy treatment planning.

2.3.1 Numerical Optimisation Methods

Numerical optimisation techniques use an objective or a cost function to evaluate the quality of a given solution and are used to drive the optimisation procedure. The objective function is optimised by varying the decision variables subject to predefined constraints to achieve a desired outcome. For instance, the objective function commonly refers to the minimum tumour dose with dose limits to healthy tissue as the constraint. The treatment plan parameters, such as beam or wedge configurations, are adjusted to achieve the highest minimum tumour dose possible without violating the dose constraints imposed to healthy tissue. With the advance in computing capabilities a variety of optimisation methods have been adopted in radiotherapy treatment planning. Often, an attempt is made to optimise a single parameter or a combination of a few parameters, while keeping the other parameters constant. For instance, many approaches focus on optimising the weight or the orientation of beams, while keeping the number of beams and wedges fixed. Oldham et al (1998) compared the relative benefit of optimising beam weights, beam orientation and wedge angles in the treatment plan of a brain tumour patient. The aim was to improve the dose distribution of the treatment plan through a) the optimisation of beam weights and wedges angles and b) the optimisation of beam orientations, beam weights and wedge angles. A downhill-simplex optimisation algorithm was employed to minimize the radiation received by healthy tissues and OAR and to minimise the non-uniformity of the radiation over the tumour region. They found that the optimisation of both combinations of parameters achieved a better dose distribution than the standard plan but the improvement was more pronounced when the beam orientations were considered in addition to the beam weights and wedge angles. This shows that restricting optimisation to one of two parameters while keeping the others constant may lead to sub-optimal plans.

Ehrgott (2010) provides a survey of mathematical methods and models used in parameter optimisation in IMRT. A comprehensive survey of continuous mathematical optimisation methods applied to beam and beamlet intensities in IMRT can be found in (Reemtsen and Alber, 2009)

Linear Programming Methods

In a linear optimisation problem, both the objective as well as the solution constraints can be formulated as linear expressions (Dantzig and Thapa, 1997, Dantzig and Thapa, 2003). The general form of a linear program can be written as :

$$\begin{array}{ll} \text{Maximise or minimise} & c^T x \\ \text{Subject to} & Ax \leq b \\ \text{and} & x \geq 0 \end{array} \tag{2.1.}$$

where x is a decision variable vector, c and b are known coefficients and A is a matrix of coefficients.

Rosen et al (1995) recommend using linear programming for radiotherapy treatment planning when the objective function and constraints can be expressed in linear form. Since the dose during irradiation is deposited in a linear fashion, the use of linear programming in radiotherapy planning has been widely researched in the literature (Holder, 2004). The main advantage of linear programming is the speed and ease of formulation (Shepard et al., 1999). As early as 1968 (Bahr et al., 1968), linear programming was applied to radiotherapy planning as an improvement on the trial and error method of manual forward planning. Shepard et al (1999) used linear programming to optimise the weights of beamlets in IMRT. They tested a variety of objective functions and constraints. One formulation aims at minimising the total radiation dose applied to the region but with a lower bound on the tumour dose to ensure sufficient radiation to the tumour. Another formulation that focuses on dose uniformity minimises the maximum deviation of the applied dose from the prescribed dose subject to an upper and a lower bound on the tumour dose and an upper bound on the dose to healthy tissue and OAR. Zhang et al. use linear programming in a two stage approach to dose optimisation in IMRT (2010). In the first stage an approximate dose is calculated and optimised while in the second stage a Monte-Carlo Kernel algorithm is used with linear programming. They found that their approach achieved good organ sparing. Hamacher et al. (2002) employ a multi-criteria linear programming approach to build a data base of Pareto optimal solutions. Each solution represents a treatment plan that is optimised with respect to dose constraints of a particular OAR. The physician can then choose the most appropriate plan from the database. Many approaches concentrate on reducing overdosing to OARs and healthy tissue. An advanced approach by Romeijn et al (Romeijn et al., 2006) penalises not only overdosing to OAR but also underdosing to the PTV. The disadvantage of linear programming (Holder, 2004) is that often it is impossible to find a solution that satisfies all of the constraints imposed by physicians. This problem can be partially overcome by changing constraints to penalties, however, certain constraints are hard constraints (for instance, feasible beam angles). Also, methods like the simplex algorithms usually terminate at the boundary of the solution, which means that some regions will attain either the upper or lower dose limits placed on them. Since setting the dose limits is not an exact science, this is not desirable.

Non-Linear Approaches

To extend the range of possible objective functions and constraints, non-linear programming can be used. The general form of a non-linear program is (Shepard et al., 1999):

$$\begin{aligned} &\text{Minimise} && f(x) \\ &\text{subject to} && g(x) \leq 0 \\ &\text{and} && l \leq x \leq u \end{aligned} \qquad 2.2.$$

where $f(x)$ is the non-linear objective function and $g(x)$ represents the set of constraints, l and u are the lower and upper bounds placed on variables x.

Rosen et al. (1995) recommend using quadratic programming to match the prescribed dose distribution to the actual one when the number of variables is small. As there are no restrictions on linearity a variety of formulations are allowed. For instance, Redpath et al. (1976) used quadratic programming in order to improve the range of possible formulations. Their objective function aims to minimize the variance of the dose at selected points of the tumour in order to preserve dose uniformity. Another possible approach is to minimise the weighted squared differences between the applied dose and the prescribed dose over the irradiated region (Shepard et al., 1999). Gibbons et al. (2000) have outlined a method, in which they start with a large number of uniformly placed beams. The beam weights are optimized using a non-linear least squares algorithm and beams with negligible weights are removed. For each remaining beam, four replicas are created. In each replica a 60° wedge is added in four different orientations. Then another optimisation algorithm is run to find the best beam and wedge arrangement. The objective function used in this study represents the difference between the tumour dose and the dose to the healthy tissue. Sonderman and Abrahamson (1985) noted a number of disadvantages that arise when using a quadratic formulation for the radiotherapy treatment planning problem: First, a minimum dose over the tumour region subject to the dose constraints on healthy tissue cannot be guaranteed. Second, multiphase treatments cannot be modelled using the objective function. It is difficult to limit the number of beams that appear in the solution. They propose using an approach, which combines linear and non-linear programming. They use homogeneity and an integral dose model with objective functions that minimize the maximum dose over the region. Wilkens et al. (2007) use linear and non-linear objective functions in a goal programming approach. The planning parameters to achieve the highest priority goals are first calculated and then turned into hard constraints while calculating planning parameters for lower priority goals. The largest drawback of non-linear programming is that the optimisation process often becomes very time-consuming with an increase in the complexity of the problem. As mentioned in section 1.1, numerical optimisation methods usually require the dose to be computed in order to evaluate the generated treatment plans. Including dose calculations treatment planning can take from a few hours to several days. (Petrovic et al., 2011, Schreibmann et al., 2003, Meyer et al., 2005). However, with constant advances in computing, methods such as parallel computing or using cloud computing could be employed in the future to considerably reduce the computation time. Also, there exists a possibility that the computed solution is only locally optimal (Wilkens et al., 2007).

To overcome the problem of local minima, the use of heuristic optimisation algorithms has been widely studied in the radiotherapy planning literature.

2.3.2 Heuristic Approaches

Heuristic approaches can be used when an exhaustive search is not feasible or is too time-consuming. A heuristic is a search methodology, which seeks a good solution in a reasonable amount of time. Heuristics cannot guarantee optimal solutions. Common heuristics are simulated annealing, evolutionary algorithms and tabu search. Heuristics have been widely applied to the radiotherapy treatment planning problem due to the large computational requirements of treatment plan parameter optimisation. The following sections provide an overview of the most commonly used heuristics.

Simulated Annealing

Simulated annealing (Kirkpatrick et al., 1983) uses an iterative random search, in which changes are accepted not only if they improve the current solution but at times also if they are worse. In the latter case, they are accepted with a probability$P$, given by the Metropolis Criterion:

$$P = e^{\frac{f(x)-f'(x)}{T}} \qquad 2.3.$$

where $f(x)$ is the cost of the current solution, $f'(x)$ is the cost of the solution under consideration and T is a control parameter known as the temperature. The advantages of simulated annealing are that it is able to escape local minima (by accepting worse solutions with probability P), and is also easy to implement and generally applicable (Burke and Kendall, 2005). For these reasons, simulated annealing has been widely studied for use in treatment planning. Rosen et al. (1995) recommend using simulated annealing to create conformal treatment plans when the objective function and constraints are in the form of complicated non-analytic functions, such as maximising tumour control probability or minimising normal tissue complication probability. They compared the performance of four different simulated annealing algorithms when computing beam weights that maximised the minimum tumour dose subject to dose volume constraints. They found that the VSGA (variable step size generalized simulated annealing) algorithm generated plans with the highest minimum tumour dose. Webb (1989) investigated the use of simulated annealing for determining optimum beam weights given the dose prescription. Each beam is divided into a number of beam elements. The weights are iteratively increased to optimise the consistency between the resulting dose distribution and the prescribed dose distribution. After each incremental weight increase, the difference in the actual and prescribed dose distribution is calculated. If the difference reduces, the increment is accepted. If the difference increases, the increment is still accepted with a probability of KT (taken from the thermal annealing analogy). This reduces the possibility of becoming trapped in a local minimum. The simulated annealing schedule initially starts with a high value of KT and is then gradually lowered. Pugachev et al. (2001) investigated the optimisation of beam orientation in IMRT. They used a simulated annealing algorithm to optimise the beam orientation at various cancer sites and compared the use of coplanar and non-coplanar beams. They found that optimizing the beam orientations improved the treatment plans markedly but also noted that algorithms, which rely on predefined objective functions, were not able to consider all factors required to generate the beam treatment plan for a patient. Morrill et al. (1991) used simulated annealing to initially identify the best objective functions and constraints but suggested that once these were identified that other methods could be used for treatment parameter computation.

The main difficulty that arises with simulated annealing is that though simulated annealing is capable of finding a global minimum, there is no guarantee that it will do so in a finite amount of time (Rosen et al., 1995). Another disadvantage is that simulated annealing is very suited for use with discrete variables and adapting it for use with continuous variables is not trivial. Finally, the efficiency of the simulated annealing algorithm depends on the selection of the generating function and the annealing schedule and choosing these a priori is not easy.

Evolutionary algorithms

Evolutionary algorithms are meta-heuristic approaches, which are characterised by candidate solutions, maintained in a population of solutions, competing for *survival.*

Commonly used evolutionary algorithms are genetic algorithms, genetic programming and ant algorithms. Similar to simulated annealing, the advantage of evolutionary algorithms lies in their ability to escape local minima.

Genetic algorithms (GA) represent the candidate solutions as alphabetical strings of fixed length. The solutions are evolved by employing operators such as crossover, mutation or replacement. The quality of a solution is evaluated using a fitness measure. Langer et al. (1996) compared genetic algorithms with simulated annealing to optimize beam weights in radiotherapy. They found that using the same set of constraints, the GA allowed a higher tumour dose and improved the overall dose distribution. Wu et al. (2000) used a genetic algorithm to optimize the selection of beam weights to achieve dose uniformity subject to an upper bound to the maximum allowable doses applied to OAR. The GA operators used include uniform crossover, arithmetical crossover, geometrical crossover, Gaussian mutation and uniform mutation. The algorithm was tested on three different tumour sites, i.e. a brain tumour, abdominal tumour and a chest tumour and was found to perform well in all three. Lei and Li (2008) employ a DNA genetic algorithm to find beam angles that result in an optimised dose distribution. Due to the large solution space of beams, treatment plan generation can be prohibitively slow. Nazareth et al. (2009) use a distributed computing platform and GAs to simultaneously optimise the beams angles and the dose distribution. A drawback of GAs is that constructing an efficient algorithm requires the selection of many parameters. The long computation time also prove problematic in computationally intensive problems such as IMRT.

Another evolutionary heuristic commonly used in radiotherapy treatment planning are based on swarm intelligence, which include the ant colony optimization algorithm and swarm optimisation. Ant colony optimisation is an iterative algorithm that is inspired by the foraging behaviour of ants. At each iteration, a number of artificial ants create solutions by visiting previously unvisited vertices. The vertex to be visited is chosen using a stochastic function (Merkle and Middendorf, 2005). Li at al. (2005) prefer ant colony optimisation to GA since their search is more efficient owing to the use of distributed computing, constructive feedback and greedy search methodology. They employ an ant colony optimisation algorithm to study beam angle optimisation for IMRT for prostate cancer. Pei et al. (2011) also used an ant colony algorithm to determine IMRT beam angles in a feasible amount of time.

Particle swarm optimisation is an iterative evolutionary search methodology that improves each candidate solution or particle based on the best known position of that particle and the overall best known position in that iteration. BASPSO (Y. Li et al., 2005) uses particle swarm optimisation to select beam angles. They found that their algorithm was more efficient than using a genetic algorithm and the resulting plan was deemed superior than the plan obtained using manual planning. However, though the method is promising BASPSO has only been evaluated using two patient cases so far.

2.3.3 Other Hybrid Approaches

Many approaches in the literature use hybrid methods to improve the performance of their algorithms. Bertsimas et al. (in press (doi:10.1016/j.cor.2012.06.009)) propose a hybrid approach that uses simulated annealing and linear programming to select optimal beam angles and to calculate their intensity. According to Rocha et al. (2012) most methods, including heuristics, evaluate too many cost functions, which make their use prohibitive in radiotherapy treatment planning, in particular in beam angle optimisation. They propose the use of radial basis functions to optimise beam angles in IMRT for head and neck cancer. Knowles and Corne (2000) used an artificial neural network in order to find suitable beam weights and wedge positions. The artificial neural network was trained using existing plans generated by human planners. They were able to generate successful treatment plans for prostate and breast cancer. However, they assumed that the beam number was fixed and that the beams were first manually placed by the planner.

Gilio (1998) employs a meta heuristic process based on tabu search that aims to find different beam configurations for brain, lung, prostate and pancreas cancer. A number of approaches using the Boltzmann transport equation have been investigated for radiotherapy planning.

In order to utilise the advantages of varied techniques, hybrid techniques are employed. Haas et al. (1998) employed a multi-objective genetic algorithm to optimise treatment plans. The generated plans are then ranked using a Pareto algorithm. The final selection is left to the clinician. Though their approach was successful, they noted that due to the large number of candidate solutions, treatment plan generation was too slow for practical clinical use. They suggested using the similarity between cancer types in an artificial neural network to improve the speed of planning. Lim et al. (2000) first apply a fast and efficient least square algorithm to obtain a good beam configuration. They then use a genetic algorithm to avoid being trapped in a local optimum.

2.3.4 Knowledge Based Methods

Knowledge based methods do not use mathematical algorithms as the reasoning mechanism but attempt to extract unknown information by manipulating existing knowledge. In spite of knowledge based methods being over-taken in many areas by optimisation techniques, they are still widely used in various clinical decision support systems. Common knowledge-based mechanisms include rule-based reasoning, case-based reasoning or hierarchical organisation of knowledge. In radiotherapy planning, they find use in several aspects such as treatment planning, assigning specific protocols to patients, developing new protocols and as training tools (Kalet and Paluszynski, 1990). Many knowledge based methods employ rules to generate treatment plans, since it is easy to transform basic clinical knowledge into a set of rules. Also, as knowledge increases, rules can easily be modified or added to the system. Kalet et al. have developed a rule-based system that uses a prototype treatment plan from the database and then applies a set of heuristic rules to refine the prototype plan:

"If the dose level is low within the target and cannot be raised with existing beams because of normal tissue tolerance limits, add another beam to the combination"

or

"If there is a cold spot within a beam's path, increase this beam's contribution in the beam combination".

A rule based system for lung cancer, CARTES (Computer Aided Radiotherapy Expert System) (Nariainen et al., 1987) uses social and clinical patient information in the treatment decision making. It can also be used as critiquing system that cross-checks the diagnosis of a physician with the stored patient data and treatment objectives.

RADONCOL (Ionescu-Farca and Willi, 1991), a rule based system for head and neck cancer uses clinical information about the tumour to first determine the treatment modality and in the case of radiotherapy, recommend the dose, fractionation and beam configuration based on a database of prototypical plans. CAVCAV (Haton, 1992) is another rule-based expert system that uses clinical patient information to specify the beam configuration and protective devices for sensitive organs first at a moderate dosage and then for a higher dosage in the boost phase of the treatment. Finally it schedules any further irradiation that is required taking into account the radiation already applied in the first two phases and the availability of the patient for treatment. A drawback of rule based systems is that a large number of rules are required to cover complex treatment problems. Also, some type of knowledge, in particular intuitive knowledge, cannot easily be encoded in rules. Langlotz et al. (1985) recognized the need for a system that could handle non-standard, complicated cases, which could not be solved using algorithmic knowledge. They designed a knowledge based treatment planning system that first generated a number of treatment plans using current and past patient information, then evaluated these plans and finally ranked them according to individual patient goals. Prentzas and Hatzilygeroudis argue (2007) that rules are brittle in the sense that they can't deal with non-standard problems or when information required to fire a rule is missing. They also state that a major drawback of rule-based systems is that they don't take into account experience of decision making.

The problems encountered in rule based reasoning, can be avoided by using case based reasoning (CBR). However, CBR has not been widely applied in radiotherapy treatment planning systems yet. A knowledge-based method using mutual information to generate treatment plans for prostate cancer uses the treatment plan of the best match found in the case base for the treatment plan of the new patient (Chanyavanich et al., 2011). Mishra (2008), Mishra et al. (2009, 2008) and Petrovic et al. (2011) designed a CBR system for treatment planning for prostate cancer. Their system suggests a treatment dose in phase I and phase II of treatment based on clinical tumour information and the dose volume histogram values, which give the permissible radiation to OAR. Schlaefer and Dieterich (2011) propose a CBR system to guide the robotic arm that applies radiation beams in radiosurgery. They show that their case-based approach reduces treatment time while maintaining high plan quality. Berger (1994) designed a CBR system, called Roentgen that aids radiotherapy planning for thorax cancer. Based on the geometry of the new patient, Roentgen retrieves a case from the case base that best matches the new case. The solution of the retrieved case is then tailored to match the specific details of the new patient. The resulting treatment plan is evaluated for dose violations and repaired if any faults are found. However, no implementation details or experimental evaluation of Roentgen are furnished.

2.4 Conclusion

The approaches used in automated treatment planning systems, their applications, advantages and disadvantages are summarized in Table 2. 1. As seen in the previous section, radiotherapy treatment planning is a complex and time-consuming process. The advent of new technologies such as 3D Conformal Radiotherapy and IMRT have made manual, iterative treatment planning prohibitively time-consuming. Both numerical optimisation based methods and rule-based methods suffer from the drawback that their efficient working depends on the knowledge or algorithm encoded in the system and is therefore limited in nature. They usually concentrate on optimising a few parameters, while keeping the other constants. Exploiting the flexibility offered by varying all or most parameters is computationally expensive and very complex to design (Dieren et al., 2000). However, in clinical practice, in particular in brain cancer, planning involves the consideration and tweaking of a large number of parameters to obtain a good treatment plan that is customised to each patient. Most of the approaches outlined above generate standard plans. However, in complex or unusual cases, standard plans are often insufficient (Kalet and Paluszynski, 1990). Instead, clinicians employ the knowledge gained by years of experience to intuitively design a treatment plan. Another major drawback is that the algorithms in many treatment planning systems work like a black box and it is difficult for the user to see how a treatment plan has been derived, which is one of the reasons why in practice treatment planning systems are not that widely employed in spite of the large body of research that has gone into creating them.

The motivation for the research done in this work lies in overcoming these problems. The advantages of applying CBR to the radiotherapy treatment problem have been listed in section 1.2.

The next chapter explains the concepts of CBR and discusses common applications of CBR, in particular in clinical decision support system.

Table 2. 1: Summary of most common approaches in radiotherapy treatment planning

Method	Advantages	Drawbacks	References in RTP
Optimisation Methods			
Linear Programming	Easy formulation, quick implementation	Can have difficulties finding a feasible solution	(Bahr et al., 1968, Rosen et al., 1995, Shepard et al., 1999, Holder, 2004, Hamacher and Kuefer, 2002, Romeijn et al., 2006, Zhang et al., 2010)
Non-Linear Programming	Allows more complex formulations	Computationally expensive, might get stuck in local optima	(Redpath et al., 1976, Sonderman and Abrahamson, 1985, Rosen et al., 1995, Shepard et al., 1999, Gibbons et al., 2000)
Simulated Annealing	Avoids becoming trapped in local optima	Requires a carefully devised cooling schedule	(Webb, 1989, Morrill et al., 1991, Rosen et al., 1995, Pugachev et al., 2001, Webb, 2005, Aleman et al., 2008)
Evolutionary Algorithms	Problem independence	Inefficient if number of variables is large	(Langer et al., 1996, Knowles and Corne, 2000, Li et al., 2003, Li et al., 2005, Merkle and Middendorf, 2005, Lei and Li, 2008, Nazareth et al., 2009, Ahmad and Bergen, 2010, Yongjie and Jie, 2010, Pei et al., 2011)
Knowledge Based Methods			
Rule-based Reasoning	Does not require elaborate calculations	Difficult to encode complex problems as rules	(Langlotz et al., 1985, Nariainen et al., 1987, Kalet and Paluszynski, 1990, Ionescu-Farca and Willi, 1991, Haton, 1992)
Case-Based Reasoning	Intuitive, quick, uses experience	Complex problems require large number of cases in case base	(Berger, 1994, Mishra, 2008, Mishra et al., 2008, Mishra et al., 2009, Chanyavanich et al., 2011, Petrovic et al., 2011, Schlaefer and Dieterich, 2011)

Chapter 3

Case-Based Reasoning

Case-based reasoning (CBR) is a knowledge-based artificial intelligence (AI) methodology that models human reasoning. In CBR, the solution to new problems is based on the solutions of past similar problems (Kolodner, 1993). The case archive or case base of a case-based reasoning system contains a database of problems and the solutions that were found for them. Each problem is represented in the form of a case. Given a new problem or target case, the case base is scanned and the case most similar to the new case is retrieved. The solution of the retrieved case is adapted to work with the new case.

It has to be noted that by modelling human reasoning it is meant that CBR models the process of human reasoning in general, i.e. solutions are inferred based on situations encountered in the past. CBR does not attempt to infer a solution by exactly following the reasoning process or steps that a human reasoner would use to solve the problem at hand. In fact, CBR is useful in situations, where the exact reasoning process employed by a human is not known entirely or is difficult to follow.

Watson provides a brief history of CBR (1994). He refers to the work of Schank and Abelson (1977), which studies the nature of knowledge and human reasoning, as the precursor of CBR. Much of the pioneering work in CBR was carried out by Janet Kolodner (1993, 1983, 1992.) (to mention just a few publications of her research work). A lot of the early work was done in the legal domain, since the practice of law often depends on the notion of precedence and previous cases (Aamodt and Plaza, 1994). Since then, CBR, however, has been applied in many domains from medical decision support systems to solving engineering problems.

This chapter provides an overview of the key ideas of CBR, in particular, focussing on the concepts, which are relevant to the work presented in this thesis. Common applications and related work, in particular focussing on medical CBR systems is presented.

3.1 Advantages of CBR

CBR can be applied to any problem, where knowledge can be stored in the form of cases. It is especially applicable to problems, where the underlying theory is difficult to model or not fully known and problems, which depend on the experience and knowledge of human experts. Traditional artificial intelligence methodologies, such as rule based systems, have received much criticism for excluding the unique human element of reasoning. Chalmers et al. (1992) argue that artificial intelligence focuses on concepts while ignoring high level perceptions fundamental to human reasoning, which leads to distorted models. CBR attempts to overcome this difficulty by reasoning based on the experience of human experts. It therefore includes contextual, perceptive and intuitive knowledge, which marks human reasoning. Other advantages include:

- CBR avoids rules (Leake, 1996): Generating rules is a time consuming and laborious process. Also the number of rules required to cover a problem domain might be prohibitively large.
- Knowledge Base Improvement: Most CBR systems constantly update their knowledge base with new, relevant cases that often have just been solved by the CBR system. Also knowledge can be pooled from several experts, which reduces human error and bias and provides a larger knowledge base.
- CBR considers previous successes, errors and failures: Knowledge of errors and failures is important since it can warn the user of common mistakes and potential problems. Currently, the developed CBR system only considers successful treatment plans. However, if cases with unsuccessful, infeasible or erroneous treatment plans were available, those could be added to the case base (flagged as unsuccessful) in

order to alert or warn the user when the treatment plan generated after adaptation has similarities with a plan flagged as unsuccessful.

- The institution's capabilities and preferred protocols are inherently present as part of the cases in the case base: In other problem solving systems, these need to be encoded either as rules or constraints, which might not be trivial. However, in CBR systems this information is already present in the cases that have been obtained from the institution.
- CBR can also be used for solving unusual, complex or incomplete problems (Bonzano et al., 1997b): In other systems, unusual and complex problems require specific rules or algorithms. Further, many systems cannot deal with problems, in which information required for problem solving is missing. Chapter 4 describes a framework that the author developed of how missing information can be dealt with in CBR systems.
- Provision of explanation: Since CBR models human reasoning, it is very easy to provide an explanation of how a solution has been derived whereas numerical optimisation or rule based systems often work like a "black box", where the user is only aware of the inputs and the outputs of the system. Providing an explanation based on the experience of experts in the field and successful previous cases increases the confidence of the user in the system. Explanations in CBR systems can be provided in different forms, depending on the domain and the goals of the user. Simple types of explanation involve displaying the retrieved case rather than merely its solution or the case attributes of the retrieved and target case, which were found to be similar and those, which were found to be non-similar in order to increase the transparency of the reasoning system (Sørmo et al., 2005)
- Modularity: Each case is a self-contained unit of information that can be easily added or removed to the system (Prentzas and Hatzilygeroudis, 2007).

3.2 Disadvantages of CBR Systems

Case-based reasoning is usually used in domains where no clear and exhaustive rules exist to solve problems or where the underlying theory is not fully understood or cannot easily be explained using mathematical formulations. However, in problems, which can be solved easily and quickly using mathematical formulations, CBR might lead to less accurate results. Some disadvantages of CBR have been listed in section 1.2. Other disadvantages include:

- Inability to express general knowledge: Cases usually contain quite specialised knowledge and it not always straight forward to design or modify an existing CBR system to solve more general problems. However, some work has been done on building general CBR development platforms or CBR problems solvers such as Colobri Studio (Recio-García, 2008).

- Small case base: Many real world applications suffer from a problem of acquiring enough case knowledge to cover a sufficiently large problem space in a domain. Case collection is often a difficult and time consuming, though vital, process. The lack of a sufficient number of cases hinders the construction and inference process of a case-based system.

- Inference efficiency problems (Prentzas and Hatzilygeroudis, 2007): The inference efficiency can be reduced when the case base is very large and therefore, retrieval time is prohibitive. A larger case base can considerably improve the retrieval accuracy. Usually, a more accurate retrieval mechanism reduces the need for a very complex adaptation mechanism and vice versa. This means that a trade-off has to be done when designing CBR systems or the saturation point has to be found at which increases in the case base size and therefore retrieval time are not offset by savings in adaptation time. Case base maintenance attempts to deal with this issue by closely monitoring the cases that are added to the case base or by clustering similar cases in the case base.

3.3 Key Concepts of a CBR System

The main parts of a CBR system are the case base and the inference engine. The case base contains a database of past cases. Each case is made up of case attributes or features, often in the form of key value pairs. Selecting the correct attributes to represent cases is crucial. The attributes not only have to describe each case accurately but more importantly, they have to describe the case relevant to finding the solution. Figure 3.1 shows the architecture of a basic CBR System. CBR systems are popularly described using the *"4 Re"* (Aamodt and Plaza, 1994):

RETRIEVE:

This stage retrieves the case from the case base that is most similar or relevant to the target case

REUSE:

The solution is adapted to fit the specific details and requirements of the target case.

REVISE:

The adapted solution is evaluated and if necessary modified.

RETAIN:

The target case and its generated solution are stored in the case base for future retrieval.

Figure 3.1. Components of a case based reasoning system

3.3.1 Case Representation and the Case Base

A case can be defined as a "contextualized piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of the reasoner" (Leake and Kolodner, 1996). The effectiveness and efficiency of a CBR system depend heavily on the representation and structure of the cases (Aamodt and Plaza, 1994). Each case represents a particular problem situation or scenario in the problem domain. The case representation depends on the requirements of the problem domain and the format in which information is available. The main components of a case are the problem description, the solution used to solve the problem and the outcome, once the solution was applied. Common approaches to represent cases include feature vectors (propositional cases), structured (relational) cases and textual (semi-structured) cases (Bergmann et al., 2005). The case representation depends on the problem and the retrieval mechanism or the similarity measure. For instance, distance based similarity measures, such as the nearest neighbour similarity measure, often represent cases as feature vectors, which represent the case as a vector of key-value pairs. Selecting the relevant features or case attributes, which describe the case relevant to the solution, is crucial. Attribute selection is discussed in detail in Chapter 5 in the context of radiotherapy treatment planning. The cases are contained in the case base. The organisation of cases in the case base depends on the retrieval mechanism. Distance based similarity measures often use a flat structure whereas inductive systems use a hierarchical organisation of cases.

3.3.2 Retrieval

This stage retrieves the case from the case base that is most similar or relevant to the target case. The two main concerns of the retrieval mechanism are correct retrieval, i.e. retrieving the case whose solution is most suitable for the target case, and efficiency of the retrieval mechanism. According to Park and Han (2002), correct and efficient retrieval is ensured by good case representation, indexing and the similarity measure. According to Stahl (2005) a good retrieval mechanism should be able to determine the most useful case from the case base, distinguish between useless and useful cases, rank the most useful cases and estimate their utility.

Retrieval mechanisms can be broadly divided into two categories, inductive and computational retrieval.

In the inductive approach, cases are organised hierarchically. They are not stored separately but form interconnected parts of the case base structure. The case base is searched for a case similar to the target case by traversing down a hierarchical indexing structure such as a decision tree. Inductive retrieval is ideally used when the solution or the goal outcome are very well defined (Park and Han, 2002). Many inductive retrieval algorithms are based on ID3, which is an iterative algorithm to construct decision trees based on case histories (Quinlan, 1986). ID3 first constructs a decision tree from a randomly chosen subset of the training data set called the window. The tree is then evaluated by classifying the objects of the remaining training set. Incorrectly classified objects are added to the window and the tree is constructed again from this window. ID3 can be used to construct a decision tree in a reasonable amount of time even when many variables are present. Heider (1996) designed an inductive CBR system based on ID3 for fault finding of aircraft engines. Jarmulak et al. (2000) use C4.5, which is an extended version of ID3, to construct a decision tree for their case-based reasoning system for tablet formulation. Inductive retrieval provides a reduced search space, which improves the speed of retrieval (Main et al., 2000). Li et al. (2012) use an inductive CBR system in supply chain trust diagnosis and found that it improved the predictive capability of their system. The disadvantage is that building the indexing structure is a complex task and cases cannot be easily added to the case base. This method is also intolerant to cases with missing or incomplete information.

The computational approach uses an explicit similarity function, which computes the similarity between the target case and the cases in the case base with respect to the case attributes. The cases are then ranked according to their similarity and the most similar case (or a specific number of cases) is retrieved along with its solution. The case base is flat in structure, i.e. each case is stored separately. The advantage of this kind of retrieval mechanisms is that they are easy to implement and maintain. Since the cases are modular they can be easily added to the case base. However, the similarity measure has to be carefully designed. Also, retrieval time can be very large for large case bases reducing the retrieval efficiency. To improve the efficiency of the retrieval stage, cases can be clustered in the case base and the clusters can be indexed so that the retrieval mechanism only has to identify the relevant cluster and then retrieve the most similar case in that cluster rather than searching through the entire case base. The retrieval efficiency can also be improved by searching the case base in parallel (Leake and Kolodner, 1996). Wess et al. (1994) use a k-d tree, or a multidimensional binary search tree to increase the speed of retrieval. The main component of computational retrieval is the similarity measure, which is discussed in detail in section 3.5 and in Chapter 7, the latter in the context of radiotherapy treatment planning.

Another consideration in retrieval is that the best retrieval mechanism does not necessarily have to retrieve the most similar case. Retrieval can be guided by other factors. For instance, adaptation guided retrieval attempts to retrieve a case that is suitable for adaptation even if it might not be the most similar case (Smyth and Keane, 1998). In diversity conscious retrieval, when a number of cases are suitable to be retrieved, the system attempts to offer maximum diversity between the retrieved cases. Compromise-driven retrieval also aims to increase the diversity of the retrieved cases. It allows the retrieval of cases, which don't exactly match the user requirements but offer a compromise that is possibly acceptable to the user. Retrieval can also be explanation guided (McSherry, 2005) to exploit the ability of CBR systems to explain how a solution has been derived and why it is applicable.

3.3.3 Adaptation

In adaptation, the retrieved solution is modified to fit the specific characteristics of the target case. Adaptation is carried out to account for differences between the target case and the retrieved case. However, many commercial CBR systems skip this step or leave adaptation to the user. This is commonly called *Null Adaptation* (Wilke and Bergmann, 1998). According to Ji et al. (2012), the method of adaptation can either involve reducing the need for adaptation or improving the adaptation algorithm. They propose a methodology that fulfils both these points in a CBR system for construction cost estimation in Korea. Adaptation is often carried out using IF-THEN rules. The rules are frequently domain specific and hand coded by experts (Berger, 1994). Though this results in very specific and accurate rules, the process of generating rules is tedious and time-consuming (Li et al., 2009a). Another problem is that since CBR is often used in domains where the underlying theory is difficult to understand, generating rules can be difficult. Another possibility is to use case-based reasoning for the adaptation stage as well (Leake et al., 1996). This requires the use of a second case base that contains instances of case adaptation and can be memory intensive. Hanney and Keane (1996) learn adaptation knowledge from the case-base. They examine the differences between all cases in the case base and infer rules according to the corresponding differences in the solution.

3.3.4 Retaining New Cases & Case Base Maintenance

Case base maintenance refers to optimising the efficiency and the performance of the case base reasoning system with respect to several factors including the retrieval efficiency, the problem coverage and the quality of the solutions (Mantaras et al., 2005). CBR systems learn by adding new cases to their case base. When a solution has been retrieved and adapted for a target case, it can be added to the case base for future use, in particular if it represents a situation that is not covered with the existing cases. However, storing every new case does not necessarily improve the quality of the CBR system. When the case base becomes too large, the efficiency of the retrieval mechanism reduces. On the other hand, a large case base covering more problem scenarios might reduce the amount of adaptation required. In the early stages of a CBR system, every new case added to the case base possibly has a large impact on problem coverage, however, this impact reduces as the case base grows since the new cases might overlap with existing cases and not offer any new knowledge. A trade-off between retrieval and adaptation cost is often required (Mantaras et al., 2005).

When the size of the case base becomes too large, deleting cases might become necessary. The simplest method is random deletion (Lawanna and Daengdej, 2010). This, however, can substantially reduce the competence of the CBR system. A strategy to select which cases to delete or retain from the case base is vital. Smyth and Keane (1995) delete or retain cases based on the coverage, the number of target cases that a case can cover and reachability, the number of cases that can be used as solution for a target case.

Another question when retaining cases is what information should be retained. Instead of just retaining the new solution, more information can be recorded as well, for instance, information that explains how the solution has been derived or how successful the outcome was (Mantaras et al., 2005).

3.4 Applications of CBR

CBR can be applied to any problem domain, in which information can be suitably encoded as cases. The underlying requirement is that similar cases should have similar solutions. It is particularly useful, when problem solving depends on expert experience or intuitive reasoning or when the underlying theory is difficult to encode or not very well understood. One of the main domains of CBR is in decision support systems. Historically, one of the earliest fields of interest in CBR was the field of law, which naturally deals with information encoded in the form of cases or precedents. HYPO (Ashley, 1991) is a CBR system that models how attorneys review past cases or precedents for trade secret disputes and infer legal arguments. Rissland et al. (2005) provide a survey of legal systems using CBR. A review of legal CBR systems with respect to argument schemes can be found in Wyner and Bench-Capon (2007).

An overview of commercial CBR systems, ranging from customer-support help desk applications to engineering problem solving can be found in Allen (1994). A variety of CBR applications has been reported in the literature. The use of CBR in electronic commerce has also started emerging recently. Lenz (1999) describes his experience of using CBR for a Virtual Travel Agency. CBR also finds use in diagnostic systems, both in health care and in the industries. The oil and gas industry has used CBR system to reduce drilling costs and to increase safety. An interesting overview of CBR in drilling operations can be found in Shokouhi et al. (2011). Watson (1994) also gives a good overview of CBR systems, in general.

3.4.1 Applications in Healthcare

CBR has been widely applied in clinical applications. In many ways, CBR is ideally suited for the medical domain. Successful decision making in medicine depends on the patient's clinical information, the facilities available, and the physician's knowledge and clinical experience. After years of experience, often the combination of these factors becomes intuitive for many physicians (Holt et al., 2006). Also, the underlying theory behind medical knowledge is often not clearly understood or defined, but knowledge can be easily described through cases. Another important factor is that it is very easy to provide an explanation of how a solution has been derived using expert clinical knowledge, which increases the confidence of the user in the system. Other reasons for using CBR in healthcare are (Bichindaritz and Marling, 2006):

1) Guidelines, which are frequently used in medicine, can be easily incorporated into a CBR system.
2) The medical literature often uses and quotes anecdotal patient cases.
3) Medical professionals naturally reason with examples.
4) Medicine uses a large body of data, making it ideal for knowledge based decision support.
5) Cases are easily available as patient information is naturally stored by hospitals.

Medical CBR systems find applications in diagnostics, classification, treatment planning and tutoring or training. Related, though they are not clinical applications, are patient record organisation and scheduling tasks, such as treatment scheduling or nurse rostering. CASEY (Koton, 1988), one of the first clinical CBR systems, was designed to diagnose heart failure in patients based on the patient data of previously treated patients. Phuong et al. (2000) introduce a CBR system for medical diagnosis of tuberculosis and other lung diseases. Care Partner (Bichindaritz et al., 1998) is a web based system, combining case-, rule-based reasoning and information retrieval, that assists clinicians with the follow up care of stem cell transplant patients. KASIMIR (D'Aquin et al., 2006) is a CBR system that aids decision support in breast cancer treatment. Frize and Walker (2000) developed a CBR system to aid assessment of patient status and facilitate diagnosis and treatment decisions. More recently, Aamodt et al. (2010) describe a CBR that aids the assessment and diagnosis of depression in palliative care. Bruland et al. (2010) use a hybrid approach to deal with uncertainty in medical decision making. They propose a system that uses a Bayesian network to model medical knowledge that is well understood and can be easily encoded, but uses CBR when such models are not available. Another hybrid approach integrates CBR with rule-based reasoning in a clinical decision support system in order to deal with high complexity problems, low experienced new staff and changing medical conditions in the intensive care unit of a hospital (Kumar et al., 2009). A summary of CBR used in medical applications can be found in (Holt et al., 2006). Begum et al. (2011) provide a more recent survey of trends and developments of CBR systems in healthcare.

3.5 Similarity measures

The main component of the retrieval mechanism is the similarity measure. Since CBR is based on the premise that similar cases have similar solutions, the similarity between two cases is an indication of how applicable the solution of a case is to the target case. Therefore, in order to retrieve cases with suitable solutions, the choice and design of the similarity measure is an important consideration when creating a CBR system.

According to Tversky (1977), the concept of similarity is fundamental in knowledge and behaviour theory and is used by humans to classify objects, form concepts and make generalisations. The notion of similarity plays a big role, not only in CBR systems, but also in classification and pattern recognition systems, and many methods and algorithms used in CBR are borrowed from classification theory. Reviews of similarity measures in CBR can be found in (Cunningham, 2009, Richter, 1992, Liao and Zhang, 1998). This chapter deals with similarity measures that can be found in computational retrieval systems.

A very popular algorithm is the k - Nearest Neighbour (kNN) method (Cover and Hart, 1967), in which the similarity between cases is a function of the distance between cases in the attribute space. The kNN algorithm is a lazy learning algorithm, which means that no computation (or generalisation) is performed till the point of retrieval of a similar case. This reduces the need for training though it can increase the actual retrieval time. Also, as will be seen later, if attribute weights are used, these are often learnt using training. The kNN algorithm is parametric and therefore makes no assumptions about the underlying distribution of the data used (Park and Han, 2002). In CBR, *k* denotes the number of cases retrieved. The similarity between attribute values is usually computed as the inverse of the Euclidian distance between two attributes. The effectiveness of the kNN algorithm reduces when there are many irrelevant case attributes present (Jiang et al., 2007). This problem can be overcome by selecting a subset of attributes, which are most relevant with respect to the solution in a problem domain. This is known as feature selection. Feature selection is a special case of feature (or attribute) weighting, where each attribute is assigned a weight according to its significance with respect to the solution. When feature weights are used, the algorithm is known as the weighted k nearest neighbour (wkNN) algorithm. Feature selection and weighting are discussed in detail in section 3.6. The weighted kNN finds application in both CBR and classification tasks due to its simplicity, robustness and effectiveness (Jiang et al., 2007). As explained previously, since the similarity computation is delayed till the retrieval stage, as in all lazy learning algorithms, the retrieval time for large case bases can be considerable. Another disadvantage of kNN is that the algorithm does not generalize well and the performance suffers when noisy data is present (Ricci and Avesani, 1995). Some of these problems can be solved using the weighted kNN algorithm. Also, the value of *k* has to be carefully determined. The method of calculating the similarity value between cases depends on the type of data available and the problem domain.

The similarity measure is often an inverse function of the distance between the individual case attributes. Let C_T be the target case and C_c be a case from the case-base. $v_{T,l}$ and $v_{C,l}$ denote the values of case attribute *l*. A basic distance measure (Cunningham, 2009) is the Minkowski Distance D_M:

$$D_M(C_T, C_C) = \left(\sum_l \left(v_{T,l} - v_{C,l} \right)^p \right)^{1/p} \qquad 3.1.$$

When $p = 1$, then the Minkowski Distance is known as Manhattan Distance and when $p = 2$, it is known as Euclidian Distance. Varying the value of p, changes the weight of the most dissimilar attributes. Both the Manhattan and the Euclidian Distance are commonly used in CBR systems, especially in the nearest neighbour algorithm. However, these distance computations assume that the attribute values are numerical in nature. Many problem domains use nominal or categorical data or often use different types of data. Ordinal values can be converted into numerical ranks. However, even when all attributes are numerical in nature or have been transformed into numerical form, they still need to follow the same interval scale so that they are comparable.

When the attributes are nominal or categorical in nature, the Tversky similarity measure can be used (Liao and Zhang, 1998):

$$Sim_{Tversky} = \frac{\alpha * common}{\alpha * common + \beta * different} \quad 3.2.$$

where $common$ or $different$ denotes the number of attributes whose values are similar or different. The constants α and β represent the weights of $common$ and $different$, respectively. Perner (1999), for instance, used the Tversky similarity measure in their CBR system for image segmentation and Champin and Solnon (2003) used it for measuring the similarity between labelled graphs.

3.5.1 Fuzzy Similarity Measures

Fuzzy set theory allows partial membership of a variable value to a set (Zadeh, 1965). In classical set theory, an object either fully belongs or does not belong to a set, whereas in fuzzy set theory an object can partially belong to a set with a membership degree of μ, which normally lies in the interval [0,1] (Kilr and Yuan, 1998). A membership function is used to compute the membership degree of a variable to a fuzzy set. Figure 3. 2 shows an example of a crisp set of all the real numbers between 1 and 3. The membership degree μ of variable x to the set *Medium* is defined by expression 3.3.

$$\mu = \begin{cases} \mathbf{1}\, \boldsymbol{for}\ \mathbf{1} \le \boldsymbol{x} \le 3 \\ \mathbf{0}\, \boldsymbol{for\ all\ other\ x} \end{cases} \quad 3.3.$$

Figure 3. 3 shows the membership functions of fuzzy sets *low, medium* and *high.* For instance, the membership degree of variable x to the fuzzy set *low* is given by expression 3.4.

$$\mu = \begin{cases} \mathbf{1}\, \boldsymbol{for\ x} \le \mathbf{1} \\ \mathbf{0}\, \boldsymbol{for\ x} \ge \mathbf{2} \\ \dfrac{\mathbf{2} - \boldsymbol{x}}{\mathbf{1}}\ \boldsymbol{for}\ \mathbf{1} \le \boldsymbol{x} \le \mathbf{2} \end{cases} \quad 3.4.$$

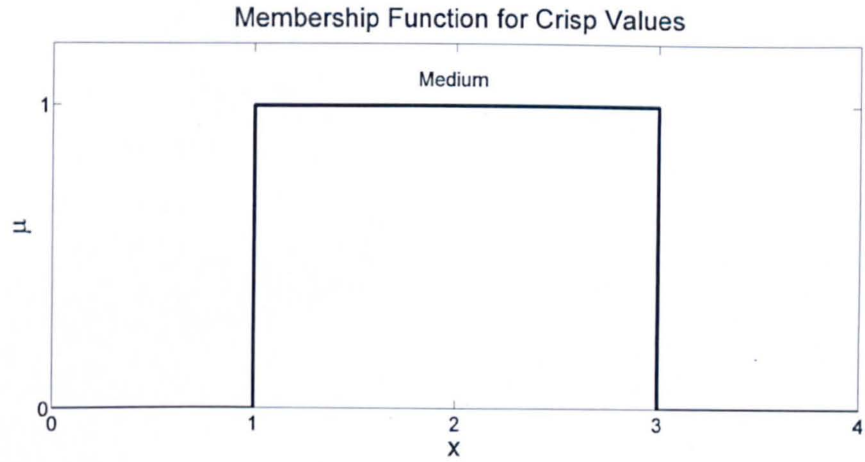


Figure 3. 2: Example of a membership function for a crisp set

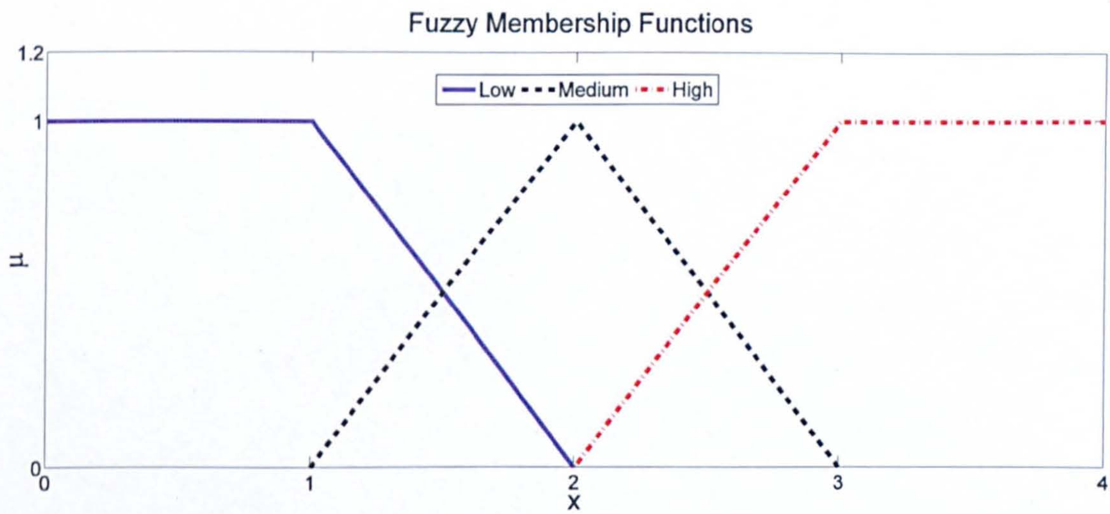


Figure 3. 3: Example of fuzzy membership functions

Fuzzy set theory or fuzzy logic is capable of capturing the meaning of vague human expressions (for example, *very small, small, medium, large*). It is also very useful to model uncertainty or imprecision (Kilr and Yuan, 1998).

Fuzzy logic has been widely used in the similarity measures of CBR systems. Hüllermeier et al. (1999) state that the concept of similarity and fuzzy set theory are very closely related since membership values of a variable can be thought of as degrees of similarity. They also suggest that fuzzy logic provides a useful tool to model and process uncertainty, which is often present in problems, which are ideally solved using CBR.

Bonissone and Mantaras (1998) suggested that fuzzy logic is very applicable to CBR since the cases stored in a CBR system are inherently fuzzy in nature as the usefulness of the case solution is normally a matter of degree as evaluated by the similarity measure. They use a fuzzy CBR system to estimate the values of residential properties. Fuzzy logic also deals efficiently with information that is imprecise or linguistic in nature. According to Weber-Lee et al. (2006) a fuzzy similarity measure is superior to a mere weighted sum of attribute values in modelling human reasoning. They employed a fuzzy similarity measure in a CBR system that forecasts cash flow accounts to assist financial management decisions. Aggour et al. (2003) propose the use of fuzzy logic in the case representation, case retrieval and similarity calculation of a CBR system. Fazel Zarandi et al. (2011) use a fuzzy clustering model in the retrieval mechanism of their CBR system for value engineering. Fuzzy CBR systems are also popular in forecasting domains. San Pedro et al. (2005) used a fuzzy CBR system to forecast tropical cyclones. CAREFUL (Case Retrieval Based on Fuzzy Logic) is a CBR system that uses fuzzy logic to represent cases and also in the retrieval step (Jaczynski and Trousse, 1994). Fuzzy indexing and retrieval was used in a CBR system that aids the design of rubber compounds of tire threads (Bandini and Manzoni, 2001). Fuzzy logic is also very useful in clinical CBR systems as it can define inexact medical terms. A fuzzy case-based reasoning system has been designed by Begum et al. to classify and diagnose stress in individuals (2007). Another advantage of using fuzzy sets is that when the attribute values are expressed in terms of membership functions, it eliminates the need for normalisation of the attribute values (Song et al., 2007). In our research, we have used a fuzzy similarity measure to accurately describe the non-linearity of similarity of case attributes between cases.

3.6 Attribute selection and attribute weights

The case attributes describe the problem in terms that are relevant to the problem solution. Case attribute selection is an important step when designing a CBR system not only to improve the accuracy of the retrieval mechanism (Li et al., 2009b) but also to further domain understanding, reduce data storage and possibly attribute measurement requirements (John et al., 1994, Guyon and Elisseeff, 2003). Wettschereck et al. (1997)

argue that the kNN algorithm is inherently biased since it permits the use of redundant, irrelevant, and noisy attributes. Once a subset of relevant case attributes has been identified, their relative significance has to be determined. Not all case attributes are equally important in determining the similarity between two cases. Most similarity measures therefore employ weights that signify the relative importance of each attribute.

Guyon and Elisseef (2003) outlined a number of questions that guide the choice of attribute selection algorithms. They suggest first using a linear predictor in a forward feature selection method or if feasible to try several feature selection methods, including both linear and non-linear and choose the one, which results in the lowest error. Contrary to common practice, they recommend using redundant features and show that this can lead to noise reduction and better class separation, in particular if the features are not highly correlated. In order to avoid overfitting of weights, they suggest looking at seemingly useless attributes, in conjunction with the other attributes available. They show that often an attribute that is useless on its own can prove to be useful in the presence of another attribute.

A large body of research exists on both attribute selection and attribute weighting. Wettschereck et al. (1997) provide a comprehensive review of feature weighting methods for lazy learning algorithms. A thorough discussion of the feature subset selection problem using the wrapper method (and the most important filter methods) can be found in Kohavi and John (1997).

3.6.1 Attribute Selection or Attribute Weighting

Attribute selection can be viewed as a subset of attribute weights where the permissible weights values are binary, that is '0' or 1. John et al. (1994) define three categories of attribute relevance, *strong relevance,* which refers to features that are indispensable in the inference mechanism, *weak relevance,* which are features that at times improve prediction accuracy and *irrelevant features,* which never contribute to prediction accuracy. Attribute selection distinguishes between relevant and irrelevant features, whereas attribute weighting also distinguishes between strong and weak relevance. Wettschereck et al. (1997) suggest using feature selection when the features are either highly correlated (and therefore could be redundant) or when features are completely irrelevant as opposed to domains, where the features vary in relevance, in which case feature weighting is more relevant.

Filter or Wrapper Methods

Attributes and their weights are often selected after consultation with domain experts. However, some studies have considered the use of automated algorithms to find an optimal set of case attributes. Attribute selection algorithms can be classified into *wrapper* and *filter* methods, also known as *performance* or *preset* biases, respectively (Wettschereck et al., 1997).

Filter methods are not based on any knowledge of the algorithm used and generally constitute a pre-processing step (Das, 2001). In other words, they do not use any performance feedback from the inference engine to select attributes or learn weight settings but use general characteristics of the training data. In classification tasks, a common method is to use class separability to select or weight attributes. An attribute that can distinguish between classes takes on the same value or the same range of values for all examples of the class and it has different values of all examples of other classes (Dash et al., 2000). Yu and Liu (2003) use a correlation-based filter method for attribute selection in classification tasks. An attribute is selected if it is highly correlated to the class and shares a low correlation with other attributes (thereby avoiding redundancy of attributes.) Many feature selection algorithms are based on the mutual information between attributes (Hanchuan et al., 2005, Torkkola, 2003, Fleuret, 2004, Dogan et al., 2008). RELIEF (Kira and Rendell, 1992) is another well-known filter algorithm. Filter methods are less computationally expensive and can be used when a large number of attributes are present. However, they are, in general, less accurate than wrapper methods (Yu and Liu, 2003, Das, 2001).

Wrapper methods use feedback about the system's performance to select attributes. The performance of the system is usually assessed using cross validation methods (Guyon and Elisseeff, 2003). The attribute subset or vector of attribute weights that offers the best performance is chosen. If the number of attributes under consideration is small an exhaustive search for an optimal set can be performed. Otherwise, more efficient search strategies are required. Kirsopp et al. (2002) have explored the use of heuristics to determine the optimum case attributes in a CBR system. They successfully use a combination of random search, hill climbing and forward sequential selection. A genetic algorithm has been employed to determine the set of attributes and their weights in a CBR system for personnel rostering (Beddoe and Petrovic, 2006). Maldonado and Weber (2009) use support vector machines for a backward feature selection algorithm. In *backward feature elimination,* the algorithm commences with the full set of attributes and iteratively removes attributes, which are deemed redundant. This is in contrast to *forward feature selection,* which starts with an empty set and then adds attributes, which are required. A clustered feature weighting approach has been adopted for a CBR system for yield management in the semiconductor industry (Ha et al., 2008). A wrapper method was used based on the sensitivity, activity, saliency and relevance of attributes with respect to the error obtained in the yield prediction. Munoz-Avila and Huellen (1996) iteratively update feature weights in the CBR system depending on the number of times a case was adequately retrieved with that feature weight. Every time a case is correctly retrieved, the amount by which its attribute weights are adjusted reduces, till the best weight is found. Apart from increasing the performance accuracy of CBR systems, Wettschereck (1995) found that the wrapper method works better than other feature weighting methods since it requires less pre-processing of the data, can be used with case attributes that are correlated to each other and it increases the rate of learning. However, wrapper methods can be computationally expensive if a large amount of evaluations is required (Bermejo et al., 2011). Bermejo et al. propose a hybrid approach that uses both wrapper and filter methods for feature selection in high dimensional datasets. The hybrid approach uses the GRASP metaheuristic to reduce the number of evaluations required by the wrapper method. They found that their algorithm considerably speedens up the feature selection process.

3.6.2 Local Attribute Weights

Attribute weights can be global or local. Traditionally, CBR systems use global weights, which remain constant over the domain, i.e. an attribute is always assigned the same weight irrespective of the attribute value or the values of other attributes. Local attribute weights can vary from case to case or with every run of the algorithm. Most CBR systems use global weights, however in certain situations this can be overly constraining or inappropriate (Wettschereck et al., 1997). It has been shown that in human reasoning the importance of an attribute changes depending on the context or the values of other attributes. Aha and Goldstone (1992) have explained that the importance of the attribute "date of deadline" might vary with respect to the value of the attribute "upcoming computer downtime before deadline". This kind of human reasoning can be transferred to artificial intelligence and seems applicable especially in medicine. For example, when determining the risk of developing diabetes, the importance of the attribute "obesity" might vary with respect to the value of "family history of diabetes".

Ricci and Avesani (1995) discuss the use of a local similarity metric in a nearest neighbour algorithm. They define the local metric as a "metric that depends on the point in the input space from which the distance is taken". For example, if the value of feature A is greater than a specific threshold, only features B & C will be used in the similarity computation. They assert that the use of a problem specific local metric improves the accuracy of the similarity computation. They show that the metric they designed, called Asymmetric Anisotropic Similarity Metric (AASM), improves the accuracy and for the same accuracy requires fewer cases in the case base. The term *anisotropic* refers to the fact that the metric is local and the term *asymmetric* denotes that the weight of an attribute in a case changes based on the value of the corresponding attribute in the target case, i.e. is it smaller or larger. Bonzano et al. (1997a) compare the effectiveness of local and global weights applied to their CBR system for Air Traffic Control and find that local weights achieve a lower error rate. The local attributes weights are updated based on the weighting scheme described above also by Munoz-Avila and Huellen (1996). Park et al. (2004) trained a neural network to learn the pattern of local attribute weights in a CBR system by reducing the distance between training cases of the same class and increasing the distance between cases of different classes. Once the neural network is trained, it assigns attribute weights during retrieval based on the attribute values of the target case. Nunez et al. (2003) use an entropy based local weighting scheme (EBL), which is based on the concept that a range of values of a feature might be more significant than another range. They assign a high weight to the significant attribute range and a low weight to attribute values, which fall outside this range. A drawback of this method is that it does not work well if the case attributes are correlated. The value difference metric (VDS) similarity measure assigns attribute weights based on the value of the attribute in that case (Stanfill and Waltz, 1986). Howe and Cardie (1997) argue that using a different attribute weight for each case instance might not always be applicable. They implement weighting on a coarser scale, where the attribute weights of a nearest neighbour algorithm are local with respect to a class. They use a class distribution weighting scheme (CDW), in which the measure of the degree to which a feature takes on a unique set of values for each class is converted into a corresponding weight value. Mesghouni et al.(2011) propose both a local and global weighting scheme using self-organising maps for feature selection. Both algorithms are tested using a wide variety of datasets with good results. However, no comparison of the local and global feature weighting algorithm is provided.

In our work, we have designed a local rule-based weighting scheme that assigns weights to the attributes of the similarity measure based on the attribute values of the target case as described in Chapter 7. The local weights and the rules to assign them are non-linear in nature and are inferred based on the performance of the retrieval mechanism rather than regression methods. We compared the performances of the similarity measure using local weights and global weights.

3.7 Imputation of Missing Values

Most real-life knowledge-based applications encounter missing values in their database. Values can be missing for several reasons including incorrect data entry, erroneous or skipped measurements or equipment faults. Missing values cause problems such as loss of effectiveness, inability of the system to process data with missing values and biasing of the data compared to the original dataset (Farhangfar et al., 2007). In CBR systems, the retrieval mechanism depends on the case description and might become less effective if case attributes are missing. McSherry (2001) shows that the precision (the number of relevant cases retrieved for a target case) and the recall (the number of cases that are relevant among the retrieved cases) deteriorate as the number of missing values in the case base increases.

The approaches to deal with missing values in CBR systems generally fall in the following three categories:

1) Use of a retrieval mechanism that tolerates missing values.
2) Case-wise deletion, where cases with missing values are discarded.
3) Imputation or replacing the missing value by making an informed guess.

In CBR, mainly the first two approaches have been applied (McSherry, 2001, Ricci and Avesani, 1995, Song and Shepperd, 2007). Many CBR systems deal with missing values at the time of retrieval by using a standard distance value for incomplete cases. In McSherry's (2001) nearest neighbour based retrieval mechanism, the similarity between attributes is calculated by awarding a point for each pair of similar attributes, giving no points for each missing attribute and subtracting a point for non-similar attribute pairs. The concept used by Ricci and Avesani (1995) is similar. A distance of '1' is assigned to two equal attributes, a distance of '0' if they are unequal, and a distance of '0.5' if one of them is missing. These methods work well if the data is nominal or if the similarity is based on exact matches.

Case-wise deletion is a very simple and very common approach to dealing with missing values. It is also referred to as list-wise deletion, case deletion or complete case analysis (Song and Shepperd, 2007). In this method, only complete cases are considered at the time of retrieval and cases containing missing values are discarded. The main drawback of this method is the loss of information. Cases having one or more missing values could be deleted in spite of being very similar to the target case with respect to the existing attributes. If there are many cases with missing values, it could lead to an unacceptable reduction in the size of the case base. Also, this approach is only valid if case attributes are missing completely at random, since otherwise it might introduce a strong bias in the case-base (Song and Shepperd, 2007).

Imputation, replacing the missing value with an estimate, is very popular in other database applications, especially for applications with clinical data (Abdala and Saeed, 2004, Vorobieva et al., 2007, Gilchrist et al., 2008, Barnard and Meng, 1999). The advantage of imputation is that it preserves valuable knowledge by not only considering the existing information in cases with missing values but also by making an informed guess about the value of the missing attribute.

Imputation methods range from simple substitutions of the missing attribute value with the mean of the entire database to complex statistical or hybrid methods. The choice of method depends mostly on (a) the type of data (b) the reason why the data is missing i.e. the mechanism of missingness (Rubin, 1976) and (c) the source of data (Gilchrist et al., 2008).

The type of data can be binary (for example yes/no, high/low), nominal (no numerical order), ordinal (data can be put in numerical order) or continuous. Some imputation algorithms are more suitable for specific types of data. For example, substituting missing data with the attribute mean (arithmetic average) or median (the value below and above which half of the observations fall) of the dataset cannot be applied to nominal data; instead the mode (most frequently occurring value) is more suitable. For ordinal or binary data we can use the median or the mode but further processing is required if the resulting value is not in the required form, i.e. an integer (Diamantaopoulis and Schlegelmilch, 2002).

Regression based imputation methods (Qin et al., 2009) are used normally for continuous data but can be applied to ordinal and binary data as well (Gilchrist et al., 2008). To find more similar cases the kwNN can be used (Song and Shepperd, 2007, Abdala and Saeed, 2004, Wasito and Mirkin, 2006). When k=1, the method is called *Hot-Deck* (Gilchrist et al., 2008). Methods based on the k-nearest neighbour algorithm substitute the missing value with a value taken from k cases that are most similar to the one with the missing value.

The mechanism of missingness also influences the choice of imputation method. There are three different mechanisms of missingness (Scheffer, 2002):

1) When data is "Missing Completely at Random" the missingness does not depend on the missing attribute or any other attribute in the case base.
2) When data is "Missing at Random" the missingness does not depend on the missing attribute, but might depend on another attribute in the case base.
3) When data is "Not Missing at Random" the missingness depends on the actual attribute that is missing.

Most methods work well if the data is 'Missing Completely at Random' or 'Missing at Random'. If the data is 'Not Missing at Random' an imputation algorithm that is based on Rubin's multiple imputations can be used to avoid biasing the data (1996). Donders et al. (2006) give a comprehensive introduction to the imputation of missing values and demonstrate why multiple imputation is superior to case-wise deletion or substituting the missing value with the mean of all values of that attribute.

The source of data should also be considered in the choice of the imputation method. Gilchrist et al. (2008) propose using all the data from a single patient for the imputation rather than multiple patients, when the data has been collected over time.

According to Barnard and Meng (1999), a sensible imputation model should include as much as possible the information available in the existing data set. Also, a balance has to be found between using a method that is too simplistic (for example, using the mean of the entire data set) and therefore inadequate and an overly complex model, which might not be practical for an application and could increase the possibility of implementation errors.

An evaluation of some popular imputation algorithms can be found, along with their applications, advantages and disadvantages in Hu et al. (2000). Acock (2005) offers a comprehensive summary of dealing with missing values.

In our research work we designed an imputation method using filtering for highly correlated attributes. Further, a frame work has been developed that considers the fact that a case has imputed values at retrieval time.

3.8 Conclusion

This chapter has demonstrated that the inference method of CBR has been applied in a large variety of applications, in particular in health care. The design of the retrieval mechanism and the similarity measure are very important. When the wNN similarity measure or any of its variations is used the attributes weights have to be carefully chosen. So far, most CBR systems use global attribute weights; however, local attribute weighting schemes can often result in a more accurate similarity calculation between two cases. In Chapter 7 a novel rule based local weighting scheme is introduced that assigns weights to an attribute based on the values of all attributes in the target case. The performance of this method is contrasted with the performance of global attribute weights in the similarity measure. Fuzzy set theory and its application in the similarity measure of CBR systems has been explained in detail. Chapter 8 describes a novel use of fuzzy set theory that takes into account the distribution of attribute similarities in a case base. In Chapter 9, a multi-phase retrieval mechanism is presented in which the parameters of the similarity measure in each retrieval phase are optimised with respect to a single solution parameter.

Chapter 4

Missing Values in a Clinical Case-based Reasoning System

This chapter deals with the imputation of missing values in a CBR system. As seen in section 3.7, imputation of missing values is important in order to preserve the information present in cases where some attributes are missing. Case wise deletion is not appropriate as shown in section 4.2. Assigning the similarity that involves missing values a constant and standard value is not appropriate either since in our CBR system, the similarity is based on the difference in attribute values rather than on exact matches, which makes this approach unsuitable. For these reasons, it was decided to develop an algorithm that could impute missing values. The problem of missing values was encountered in a CBR system (called RTP-CBR) for radiotherapy treatment planning for prostate cancer developed previously by members of the research group (Song et al., 2007, Petrovic et al., 2011) for the City Hospital. The reason why values are missing is not always known. Some attributes may be omitted during data transfer. Since many of the patient records are hand written, illegibility too gives rise to missing values. Previously, in the developed RTP-CBR system for prostate cancer, incomplete cases were discarded. Due to the scarcity of cases available, however, this was deemed to be a waste of patient information, which motivated us to design an imputation algorithm. Since the case attributes are correlated (however, not linearly), they are very suitable for imputation.

When dealing with missing values, another important question that arises is that once a missing value has been imputed how is it used in an application? Multiple imputations (Rubin, 1976) take into account the inherent uncertainty associated with imputed values but more work needs to be done to reflect this uncertainty when using the imputed value in an application. In CBR, one has to keep in mind that the similarity of a case in the case base to the target case could be erroneously high or low if the imputed value of a case attribute happens to be very different from the original value.

This chapter describes a methodology of dealing with missing values in the case base of a CBR system. A novel imputation method, called filter imputation, for ordinal, correlated data is described. Further, a framework is presented for dealing with missing values in a case-based reasoning system that uses a nearest neighbour based retrieval mechanism. The framework, which consists of a series of steps that can be used with any imputation method, considers the quality of the imputation method and the inherent uncertainty in the similarity calculation during case retrieval. The framework is evaluated using the CBR system for prostate cancer (described in section 4.1). Real-world data on prostate cancer patients are used in the experiments. It has to be noted that this is a different CBR to the one for radiotherapy treatment planning for brain cancer, referred to in the remainder of the thesis. Section 4.2 demonstrates the short comings of using case-wise deletion to deal with missing values in a CBR system. An imputation approach using the common weighted nearest neighbour (wNN) algorithm is described in section 4.3.1. Section 4.3.2 presents the novel filter imputation method that was developed for the CBR system for prostate cancer. The filter approach was compared with the wNN method and a simple substitution of the missing attribute by the mode of the case-base as presented in section 4.4. Section 5 discusses the imputation framework and demonstrates its application in the RTP-CBR system for prostate cancer with the help of two real prostate cancer cases obtained from the City Hospital.

4.1 A CBR System for Radiotherapy Planning in Prostate Cancer

To facilitate understanding of the main issues of this chapter, the previously developed case-based reasoning system named RTP-CBR for radiotherapy treatment planning for prostate cancer is described in this section. Further details can be found in

(Song et al., 2007). This CBR system proposes a radiation dose for a new prostate cancer patient based on previously treated "similar" prostate cancer patients. The case base contains 47 complete cases. Each case represents a patient and is described by two groups of attributes shown in Table 4.1.

Group I provides clinical information about the progress of the tumour. It consists of 3 discrete, ordinal attributes, namely the clinical stage, the MRI value and the Gleason score and one continuous attribute, the Prostate Specific Antigen (PSA) value. Group II consists of 8 continuous attributes, related to as DVH (dose-volume histogram) whose values give the percentage dose received by the different percentages of volume of irradiated organs. The solution part of each case is the amount of radiation dose prescribed for the cancer patient. Using a nearest neighbour similarity measure, based on fuzzy set memberships of attributes, the four most similar cases to the target case are retrieved and passed on to the adaptation stage to calculate the dose prescription for the new patient. For the sake of numerical calculations, the clinical stage and the MRI stage values have been converted to corresponding rank values, as shown in Table 4.1.

Table 4. 1: Prostate cancer patient attributes used to describe cases.

Attribute	Description	Data type	Values present in our case base	Rank values
Group I Attributes:				
Clinical Stage	Describes cancer extent	Ordinal	T1a,T1b,T1c,T2a, T2b, T3a	T1a=1, T1b=2, T1c=3,...T3a=6
MRI Stage	Describes cancer extent	Ordinal	T1, T2, T3	T1=1, T2=2, T3=3
Gleason Score	Describes cancer grade	Ordinal	6,7,8,9	-
PSA Value	Prostate Specific Antigen: Protein secreted by prostate gland cells indicating presence of cancer cells	Continuous	5 - 35	-
Group II Attributes:				
DVH Values	Distribution Volume Histogram giving distribution of given radiation dose	Continuous Percentages	1-100%	-

In this research, we only consider missing values of the Gleason score and the clinical stage. Continuous attributes, such as the PSA and the group II DVH values will be handled in our future research.

Table 4. 2 Spearman's rank correlation coefficients between attributes of group I

Attribute	Clinical stage	MRI value	Gleason score	PSA value
Clinical Stage	1	0.456	0.464	0.350
MRI value	0.456	1	0.278	0.274
Gleason Score	0.464	0.278	1	0.133
PSA value	0.350	0.274	0.133	1

Table 4. 2 shows the Spearman's rank correlation coefficient of group I attributes. The Spearman rank correlation coefficient is a statistical, non-parametric measure of correlation for ordinal data, which does not make any assumptions about the distribution of the data (Diamantaopoulis and Schlegelmilch, 2002). It is calculated using the complete case base prior to any imputation.

4.2 Case-wise Deletion

Despite serious drawbacks, this method is attractive due to its simplicity of use. We tested the effect of case-wise deletion on the RTP-CBR system by deleting 10% (5 cases out of 47), 20% (10 cases) and finally 50% (24 cases) of the cases in our case base and comparing the resulting similarity of retrieved cases with that of cases retrieved from the original complete case base. The graph in Figure 4.1a) shows the average similarity of the 4 retrieved cases for 10 different target cases randomly chosen from our case base. We can see that for many target cases the average similarity of the retrieved cases reduces as more cases are removed from the case base, which means that cases with less suitable solutions are retrieved. To reconstruct a worst-case scenario we deleted the four most similar cases for each of the target cases as shown in Figure 4.1 b) and calculated the average similarity of the next four cases. The graph shows a significant reduction in the average similarity for all target cases between the first 4 most similar cases and the next 4 most similar cases. It can be concluded that case-wise deletion would adversely affect the performance of the CBR system.

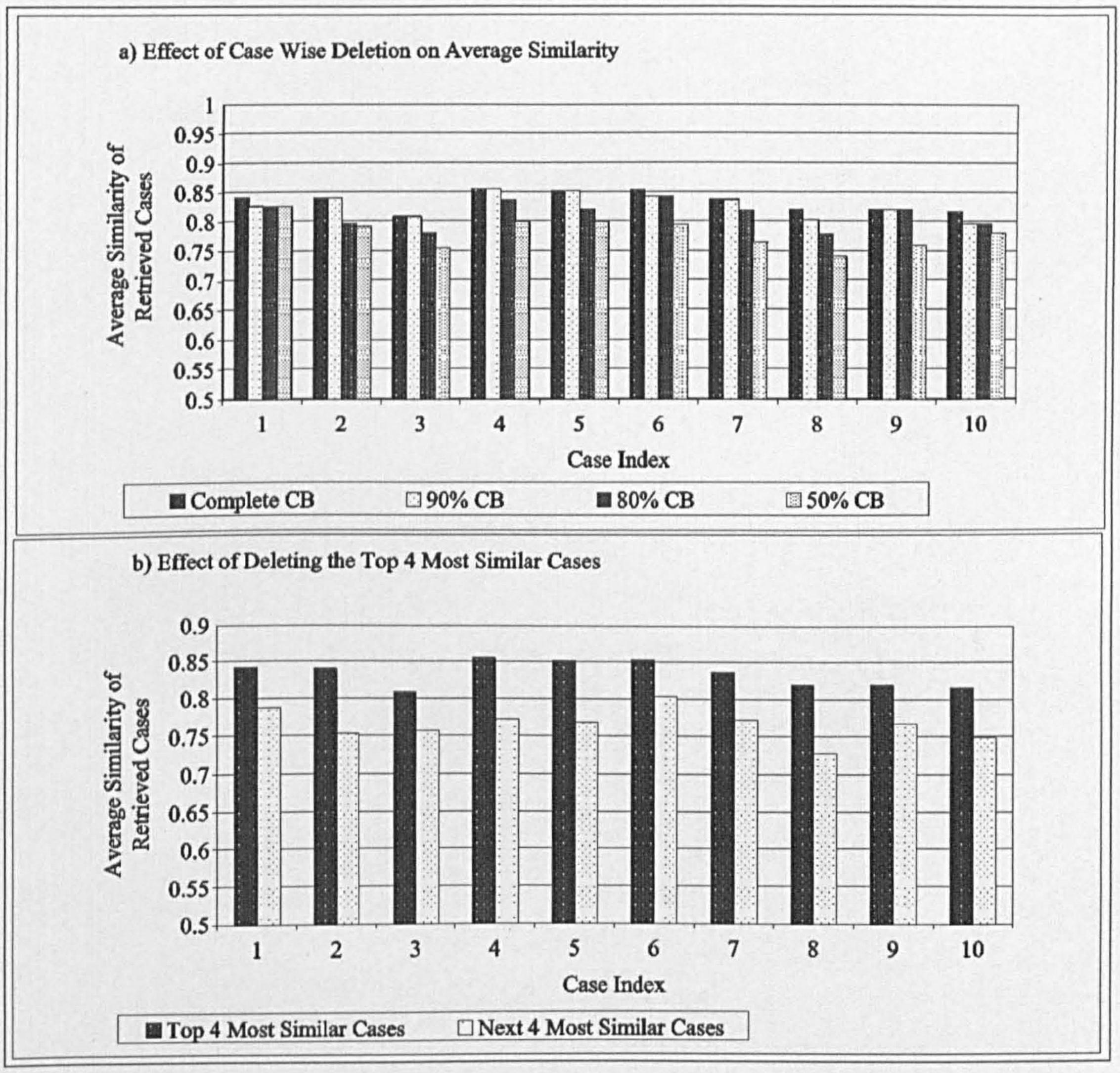


Figure 4.1: Graph (a) Average similarity of 4 most similar retrieved cases when using all (complete CB), 90%, 80% and 50% of cases in the case base. Graph (b): Average similarity of top 4 most similar retrieved cases and the next 4 most similar cases if top 4 retrieved cases were deleted.

4.3 Imputation of Missing Values

This section explains two approaches to imputing missing values. The first approach is based on the commonly used weighted k nearest neighbour method (wkNN). Then the developed filter method is described.

4.3.1 Weighted k Nearest Neighbour Imputation Approach

For our data, which is discrete, ordinal and "Missing at Random", the weighted k-nearest neighbour (wkNN) algorithm, fulfils the prescribed requirements. It utilizes the information contained in the entire case base and the correlation between attributes. It is applicable to most types of data and, if designed appropriately, works well with all mechanisms of missingness. Significantly, the implementation is quick and easy.

The wkNN imputation uses the four group I attributes, the clinical stage, the MRI value, the Gleason score and the PSA value. The value of the missing attribute is imputed by the mode, the most frequently occurring value of the attribute, of the k nearest neighbours. For ordinal data, both the mode and the median of a dataset can be used, but lower imputation errors were obtained with the mode. If there is more than one mode among the k nearest neighbours, then by convention the smaller mode value is used. Consider a case c_m having case attributes l_n, n = 1,2,3,...N. The missing attribute is denoted by l_m. As an example, let us assume that the clinical stage is missing. The *k* nearest neighbours are found based on the other three attributes, i.e. the Gleason score, the MRI stage and the PSA value. The nearest neighbours are identified by computing the distance between case c_m and each of the cases in the case base. The distance D between the case c_m and a case c_i from the case base is given by

$$D(C_m, C_i) = \left(\sum_{l=1}^{3} w_{l_n} \left(v_{m,l_n} - v_{i,l_n} \right)^2 \right)^{1/2} \qquad 4.1.$$

where v_{m,l_n} and v_{i,l_n} are the normalized values of attribute l_n of cases c_m and c_i respectively; w_{l_n} is the weight of attribute l_n, given by Spearman's rank correlation coefficient between attribute l_n and the missing attribute l_m as shown in Table 4. 2. Using trial and error method, we found that we obtained the best results with k=8.

The disadvantage of the wkNN imputation method is that finding appropriate weights can be a difficult and time-consuming task, especially if the correlation between attributes is not clearly defined (Gilchrist et al., 2008). The value of 'k' has to be carefully chosen. If it is too large it might include cases that are quite dissimilar and therefore irrelevant for the imputation while if it is too small the imputed value will be vulnerable to outliers or extreme values and biasing. For these reasons, we developed an imputation method, which retains the good characteristics of the wkNN algorithm but overcomes its shortcomings.

4.3.2 A Filter Approach to Imputation

We propose a new filter imputation method that is based on the correlation between attribute values. However, it is not required to know the exact values of the correlation coefficients but only which attributes are more correlated than others.

Consider a case c_m with missing attribute l_m. Let l_n, n=1, 2, 3,...N, denote the other attributes in the case. From the case base, the attribute that has the highest correlation with the missing attribute is identified. If attribute l_n has a high correlation with l_m, all cases in the case base that have the same value of l_n as c_m are extracted. The other cases are discarded. The value that occurs most frequently, i.e. the mode of the attribute of the extracted cases gives the imputed value. The specificity between the case with missing values and the filtered cases can be increased by filtering using other attributes as well. So, if for example, attributes $l_1, l_2,...l_N$ of a case are correlated to attribute l_m, then all cases that have the same values of $l_1, l_2,...l_N$ as the case with the missing value, are extracted. The remaining cases are discarded. The mode of l_m in the extracted cases provides the imputed value. The attributes used to filter are chosen depending on how strong the correlation between them and the missing attribute is. The system is programmed to only filter by using an attribute l_n of case C if there is a predefined minimum number of cases (in our system, 2 cases) that have the same value of I_n. Alternatively, we can use the nearest value to I_n that does have a match in the case base.

The filter imputation method works well for imputation in our case base, since all case attributes are correlated to each other as shown in Table 4. 2. According to Table 4. 2, the correlation between the clinical stage and the other attributes is in the descending order of Gleason score, MRI stage and PSA value.

The filter method requires discrete attributes. Therefore the PSA values, which are continuous in nature, are divided into three groups 1, 2 and 3, which represent the clinical categories "low", "medium" and "high probability" of prostate cancer, respectively. The groups are defined as follows

$$PSAGroup = \begin{cases} 1 \; if \; PSA < 10.5 \\ 2 \; if \; 10.5 \leq PSA \leq 15.5 \\ 3 \; if \; PSA > 15.5 \end{cases} \qquad 4.2.$$

As an example let us consider the case given in Table 4.3, which misses the value for the clinical stage. The PSA value of 7.1 falls in $PSAGroup$ 1. The case base is filtered by extracting cases that contain a Gleason score of '7', an MRI value of T2 and a PSA Group of 1. The remaining cases are discarded. The mode of the clinical stage values of the filtered set of cases gives the imputed value.

Apart from the case in Table 4.3, there are 21 cases in the case base with a Gleason score of 7, 25 cases with an MRI stage of T2 and 16 cases that fall in PSAGroup 1. If we filter using all three attributes we obtain 3 cases. The mode of the clinical stage in these 3 cases is T1c. So the missing clinical stage value is replaced with T1c.

Table 4.3: Example of a case with missing clinical stage.

Case	Clinical stage	MRI value	Gleason score	PSA value
Cm	?	T2	7	7.1
C_1	T1c	T2	7	4.6
C_2	T1c	T2	7	8.7
C_3	T1b	T2	7	11.2

4.4 Test Results

We tested the wkNN and the filter imputation method on our case base of 47 cases by using a leave-one-out strategy. Each of the 47 cases was consecutively made the target case by removing the clinical stage attribute and imputing its value using the other 46 cases. The difference between the imputed value of the attribute and the original value denotes the error and gives an indication of the success of the imputation. The error is averaged over all 46 retrieval runs. The procedure is repeated by consecutively removing the Gleason Score from the target cases. The results were also compared to the most basic imputation method i.e. substituting the missing value by the mode of the entire case base.

The normalised root mean square error *NRMSE*, $NRMSE \in [0,1]$, is used to measure the quality of the imputation method. It is based on the *RMSE*, the root mean square value of the error, between the imputed value and the actual value and the number of possible values that an attribute can take. A lower *NRMSE* value indicates higher quality of the imputation method.

$$NMRSE = \frac{RMSE}{(x_{max} - x_{min})} \qquad 4.3.$$

where x_{max} and x_{min} are the maximum and minimum value of the attribute found in the case base.

Figure 4. 2 shows the *NRMSE* achieved by the filter imputation method, when filtering by using 1, 2 or 3 attributes, the wkNN method and using the mode of the entire case base. Let us denote by Filter-I, II and III a filter using one, two or three attributes respectively. If the clinical stage is missing, Filter-I extracts cases from the case base with respect to the Gleason score, which has the highest correlation with the clinical stage, Filter-II with respect to both the Gleason score and the MRI stage, which has the second highest correlation with the clinical stage and Filter-III with respect to the Gleason score, the MRI stage and the PSAGroup value. If the Gleason score is missing, Filter-I extracts cases from the case base with respect to the Clinical stage, Filter-II with respect to both the Clinical stage and the MRI stage and Filter-III with respect to the clinical stage, the MRI

stage and the PSAGroup value. In the graph, "CB mode" denotes the method of using the mode of the entire case base to provide the imputed value.

The clinical stage has a higher correlation with the other attributes than the Gleason score, i.e. for the clinical stage the lowest Spearman's rank correlation coefficient has a value of 0.350, whereas for the Gleason score the lowest Spearman's rank correlation coefficient is 0.133 as given in Table 4. 2. Therefore, the imputation of the clinical stage shows an overall lower *NRMSE* value than the Gleason score for all filters, indicating better imputation success.

The performance of the investigated imputation methods is presented in Figure 4. 2a. When the clinical stage is missing, using the mode of the entire case base gives the highest error. The filter method fares better than the wkNN. As the specificity of the filter i.e. the number of attributes that are used for filtering, is increased, the performance improves.

For the Gleason score imputation, we can see in Figure 4. 2a that the wkNN has the lowest *NRMSE*. Since the correlation between the Gleason score and the other attributes is smaller, the filter method, which heavily relies on the correlation between attributes, fares worse. This is confirmed by the fact that the filter method does well with Filter-I, when filtering only using the clinical stage, since the correlation between the clinical stage and Gleason score is still good. It performs poorly when filtering using the MRI and PSA attributes as in Filter-II and Filter-III.

The interesting question that arises is how much are the imputation methods affected by the size of the case base. We carried out an experiment to test the performance of Filter-I, II, III and the wkNN imputation method with different sizes of the case base, i.e. with 10%, 20% and 50% of cases removed from the case base. As shown in Figure 4. 2 b) the *NRMSE*s of neither the filter nor the wkNN imputation methods show a strong or regular dependence on the number of cases in the case base.

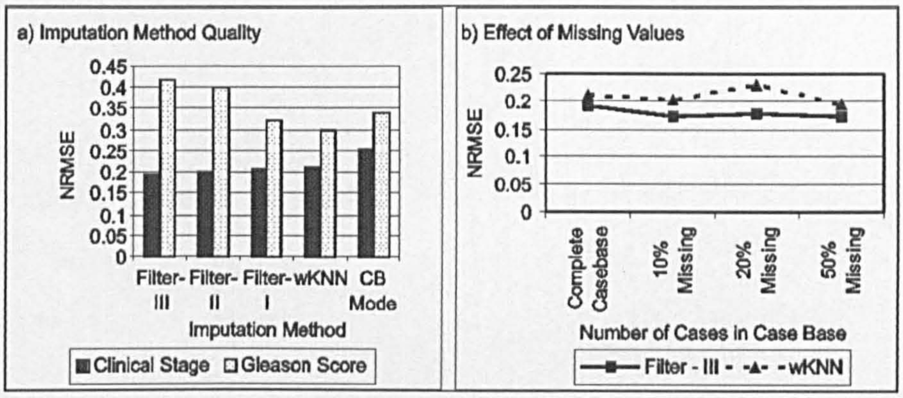


Figure 4. 2: *NRMSE* of the different imputation methods. The graphs show the *NRMSE* for (a) the clinical stage and Gleason score imputation (b) the clinical stage imputation if 10%, 20% or 50% of cases in the case base are missing.

4.5 Imputation Framework

Once missing values in the case base have been imputed, the most important question is how to use these values during case retrieval. Using a high quality imputation method increases the likelihood of finding a close match to the original value of an attribute. However, no imputation method can provide a guarantee that the imputed value is the same as or even close to the original value. Therefore, if there are two similar cases, a complete case and a case with imputed values, then the complete case should preferentially be retrieved. Two situations can occur that the original value of a case attribute is closer to the target case value than the imputed value or that the imputed value is closer. In the first situation, the case with the imputed value might not be retrieved in spite of having a suitable solution. In this situation, the next most similar case would be retrieved, which might slightly reduce the effectiveness of the CBR system but would still be acceptable. However, if the imputed value is closer to the target case value than the actual value, a case with a very unsuitable solution could be retrieved. This situation could hazardously affect the final solution suggested by the CBR system and has to be avoided. Since the original value of the missing attribute is not known, it is impossible to estimate, which situation might occur. Therefore, retrieving a case with an imputed value carries a potential risk, which has to be reflected in the retrieval process.

The success of an imputation depends not only on the imputation method but also on the type of data in the case base. A particular imputation method might not perform equally well for all attributes in the case base. However, it might not be feasible or practical to use several imputation methods in one system. Also, it might not be necessary to use the most successful or reliable imputation method. Depending on the resources and time available, in some instances, a method that is fast and easy to implement might be preferable to a more successful but complex method. However, if a less reliable imputation method is used, this needs to be considered in the retrieval process.

For these reasons we have devised a framework in which the missing values in a CBR system are imputed and information about the quality of the imputation is fed back into the CBR system through the similarity calculation. The similarity between the target case and a case with imputed values is lowered based on the quality of the imputation method. The steps of the framework are given below:

Imputation Method Selection for Each Attribute

The most appropriate imputation method is chosen for each attribute based on the type of attribute data, the mechanism of missingness and the data source. Depending on the resources available a single imputation method can be used as well for imputation of all missing attribute values.

Calculation of the Quality Measure of the Imputation Method.

First, the attributes that contain missing values in the case base are identified. Then all complete cases from the case base are extracted to form an evaluation set of cases. For each identified attribute, values are removed from the evaluation set to simulate missing values. A leave-one-out strategy is employed to make each case consecutively the target case and impute its "missing value". The difference between the imputed value and the actual value is calculated. The quality measure Q of the chosen imputation methods is calculated in terms of the *NRMSE* as given in expression 4.4. If there is more than one attribute missing in a case, the quality measure Q is given by the product of the individual quality measures for each missing attribute.

$$Q = \frac{1}{1 + NRMSE} \qquad 4.4.$$

Imputation of Missing Values

The missing values in the case base are imputed using the imputation method selected in the first step.

Marking Cases

Cases with imputed attributes are marked. Each imputed attribute is marked with a score given by the quality measure Q for that imputation method. Attributes with original values are given a score of 1. The total quality measure of a case is given by the product of the individual attribute scores.

Modified Similarity Calculation

To reflect the uncertainty of imputation, the similarity between the target case and a marked case in the case base is modified based on the quality measure as shown in expression 4.5.

$$\boldsymbol{Sim_{imp} = Q \times Sim} \qquad 4.5.$$

where Sim is the similarity between the target case and the case in the case base (without considering whether the case attribute values have been imputed); Sim_{imp} is the modified similarity that reflects the quality Q of the imputation method if an attribute value has been imputed. Since, $\mathrm{NRMSE} \in [0,1]$, $Q \in [0.5,1]$. Therefore Q lowers the value of the computed similarity Sim. This reflects the fact that a case with imputed values might not have as good a solution as a similar but complete case without missing values, and should therefore be given lower priority during the retrieval process.

The lower limit of the similarity of a case with imputed values is given by Sim_{min}, which is obtained by substituting the missing attribute value by a value that has the maximum distance from the value of the attribute in the target case. Therefore, we scale Sim_{imp} to the range $[Sim_{min}, Sim]$ using the following expression

$$Sim_{scaled} = Sim_{min} + [(Sim - Sim_{min}) \times Sim_{imp}] \qquad 4.6.$$

Retrieval

The cases most similar to the target case based on the modified similarity measure Sim_{scaled} are retrieved and passed on to the adaptation stage.

4.6 Evaluation of the Proposed Framework

We provide two examples to illustrate the application and performance of the imputation framework.

Table 4. 4 Example containing target case A and selected cases from the RTP-CBR case base

Row reference	Case reference	Similarity to A	Similarity rank	Clinical stage	MRI value	Gleason score	PSA value
1	Target	-	-	T1c	T2	7	7.1
2	1	0.892932	1	T1c	T1	7	6.8
3	2	0.878205	2	T1c	T2	7	8.7
4	3	0.856848	3	T1c	T2	7	12
5	...	...	...	...	...	...	...
6	11	0.769769	11	**T2b / ?**	T2	7	9.7
7	...	...	...	...	...	...	...
8	39	0.649	39	**T2b / ?**	T2	9	9
9	...	...	...	...	...	...	...

<u>Example 1</u>

Let us suppose that a case base contains the cases shown in Table 4. 4. Case *A* is the target case. Cases are listed in descending order of their similarity to target case *A*. Consider case *11*, which is the 11th most similar case to the target case. The similarity between case *11* and target case *A* is $Sim_{A\text{-}11}$ = 0.7698 as given by our CBR system. The clinical stage value of the target case *A* is T1c and that of case *11* is T2b. To demonstrate the significance of our framework let us assume that the clinical stage of case *11* is missing. If the clinical stage is imputed using the filter imputation method, Filter-III, a value of T1c is obtained for the clinical stage. The imputed value of the clinical stage increases the similarity between the target case and case *11* to 0.8569, which would make case *11* the 3rd most similar case to the target case. Since the CBR system retrieves the 4 most similar cases, case *11* would be retrieved and used in the solution for target case *A*.

The aim of the framework is to lower the similarity to reflect that the retrieved case contains imputed attribute values. This is done as follows:

First the quality measure of the filter imputation method is calculated for the clinical stage. Since, *NRMSE* = 0.1919, *Q=1/(1+NRMSE)* = 1/(1 + 0.1919) = 0.8389

Following expression 4.5,

$$Sim_{imp} = Q * Sim = 0.8389 * 0.8569 = 0.7188$$

Since the clinical stage of target case *A* is T1c, the clinical stage value in our case base that has the furthest distance from T1c is T3a. Therefore, substituting the clinical stage value in case *11* with T3a, Sim_{min} = 0.7582. Following expression 4.6,4.6

$$Sim_{scaled} = Sim_{min} + \left[\left(Sim - Sim_{min}\right) * Sim_{imp}\right]$$

$$Sim_{scaled} = 0.7582 + [(0.8569 - 0.7582) * 0.7188] = 0.8290$$

With a similarity of 0.8290, case *11* is only the 5th most similar case to the target case and would therefore not be retrieved by the CBR system. This example shows how the erroneously high similarity of a case with imputed values is brought down reflecting the uncertainty in its imputation.

Example 2

The following example shows the use of the framework when two attributes in a case are missing and are imputed using different imputation methods. The clinical stage and the Gleason score of target case *A* are T1c and '7' respectively. Consider case *39* in, Table 4. 4, which is the 39th most similar case to case *A* with a similarity of 0.6469. Let the clinical stage and the Gleason score be removed from case *39*. Using Filter – III to impute the clinical stage and the wkNN imputation method to impute the Gleason score we obtained values of T1c and 7 for the clinical stage and the Gleason score respectively. Without the framework the similarity $Sim_{A\text{-}39}$ = 0.8449, which would make case *39* the 4th most similar case to the target case. However, the actual similarity of case *39* to target case *A* is very low and hence its solution is not likely to be suitable for case *A*. Let us now apply the framework. The respective *NRMSE* for the clinical stage and the Gleason score for the filter and wkNN methods are: 0.1919 and 0.2997. Therefore,

$$Q = Q_{ClinicalStage} * Q_{Gleason} = [1/(1+0.1919)] * [1/(1+0.2997)] = 0.6455$$

Following expression 4.5,

$$Sim_{imp} = Q * Sim = 0.6454 * 0.8449 = 0.5452$$

Using 'T3a' for the clinical stage and '9' for the Gleason score in case *39*, we obtain the minimum similarity value, Sim_{min}= 0.6105. Following expression 4.6,,

$$Sim_{scaled} = Sim_{min} + \left[\left(Sim - Sim_{min}\right) * Sim_{imp}\right].$$

$$Sim_{scaled} = 0.6105 + [(0.8449 - 0.6105) * 0.5452] = 0.7383.$$

With this similarity, this case is the 19th most similar case to the target case and rightly would not be retrieved or used in the solution for target case.

From the two examples, we can see that even with a generally successful imputation method, there is a possibility of the imputation wrongly driving up the similarity of a case with imputed values. However, the proposed framework avoids this situation by modifying the similarity value based on the quality of the imputation method.

4.7 Conclusion

This chapter presents the filter imputation method that requires correlated data to impute missing values in a case-based reasoning system. We proposed a framework that goes beyond the imputation of missing values by considering the inherent uncertainty of imputation and feeding back the quality of the imputation method to the similarity calculation. We discussed the suitability of case-wise deletion and other common mechanisms that deal with missing values in CBR systems and found them to be unsuitable for use with our data in the medical CBR system for radiotherapy planning. Case-wise deletion in particular, disregards a large amount of existing useful information and deteriorates the retrieval mechanism of CBR.

Among imputation methods, wkNN method works well for ordinal data but requires care in choosing the exact weights and the value of '*K*', the number of retrieved cases. These drawbacks are overcome with the filter method, which works exceedingly well for highly correlated data but does not require knowledge about the exact correlation values. With smaller correlations (in the developed CBR system, correlations with a Spearman's rank correlation coefficient of less than 0.350), the case base should not be filtered using all attributes but only using the ones that show a larger correlation to the missing attribute. In case bases, where many different attributes are missing, the filter method can be adjusted to filter by those attributes that have fewer missing values.

The framework presented allows us to use several imputation methods or less efficient but quick imputation methods in a CBR system by giving feedback about the quality of imputation used for each attribute. Cases with imputed values are still considered in the case base but are given lower priority than similar but complete cases to reflect the possibility of imputation error.

In the future, we will study imputation of continuous attributes and how the percentage of missing values influences the decision about the number of filtering stages or which attributes to use for filtering. We will also further look into ways that improve the filter method when the number of cases that provide exact matches for the filter attributes is small.

The filter imputation method and the imputation framework have been applied to prostate cancer cases. In the future, we will apply and test these techniques to the developed radiotherapy treatment planning CBR system for brain cancer cases.

Chapter 5

Architecture of the CBR System

This chapter provides an overview of the developed CBR system for radiotherapy treatment planning. Extended and comprehensive discussions have been held with oncologists and medical physicists at the City Hospital to discuss the treatment planning procedure. It has to be noted that our CBR system does not aim to faithfully replicate the steps of manual treatment planning but instead aims to provide the functionality of generating an acceptable treatment plan. However, an understanding of the manual process is vital in order to thoroughly understand and fulfil the aims, requirements and considerations of a decision support system developed for treatment planning. Section 5.1 reflects on the discussions held with staff regarding treatment planning at the City Hospital. The input data, i.e. the patient DICOM image files, used in both manual treatment planning and the CBR system is described in section 5.2 in order to explain how the case attributes are derived. Section 5.3 gives an overview of the developed CBR system with a focus on the case attributes. Prior to retrieval of the most similar case, the case base is filtered to identify cases that are compatible with the target case. This step is explained in section 5.4.

5.1 Radiotherapy Treatment Planning for Brain Cancer at the City Hospital

Currently, there are about 5000 new brain cancer cases every year in the UK. The Nottingham University Hospitals Trust, City Hospital Campus in Nottingham treats about 150 cases each year. The City Hospital has a modern, dedicated radiotherapy facility, which treats different types of cancer. Each type of cancer follows a different protocol and guidelines. In order to demonstrate the differences between the planning procedures between different types of cancer, Table 5.1 outlines the major differences between prostate and brain cancer treatment planning at the City Hospital.

Table 5.1. Differences between prostate and brain cancer treatment planning

	Prostate Cancer	**Brain Cancer**
Treatment aims	Maximise radiation dose, while minimising dose to OARs	Uniform dose distribution with no over - or underdosing
Dose	Dose is determined during planning	Dose is constant
Number of beams	4	2-6
Angle of beams	90degrees, co-planar	Any angle, co- or non-coplanar
Treatment planning inputs	Uses clinical patient and dose volume histogram information	Uses patient DICOM images showing tumour and OAR outlines

Another type of cancer is head and neck cancer. Since the treatment planning steps and the data considered for planning are exactly the same for both head and neck cancer cases and brain cancer cases, both types could potentially be handled by the final CBR system. However, there are vital differences that would have to be considered additionally. In head and neck cancer, often in addition to target coverage, the radiation fields need to be matched to ensure coverage of cancer nodules, especially if the tumour is located in the neck region. Some cases of head and neck cancer patients are treated in two phases. Head and neck tumours are grouped according to their location whereas brain tumours are grouped according to their pathological tissue type; however, the tumour type affects only the tumour or planning target volume (PTV) outline and not the medical physicist's process of planning. Presently, the case base includes a couple of head and neck cancer cases. However, they are very straight forward cases that use only one treatment phase and do not have cancer nodules and are therefore comparable to brain cancer cases.

5.1.1 Treatment Planning Aims

In brain cancer radiotherapy at the City Hospital, the radiation dose to the tumour volume is constant and is set by the oncologist. This dose is called the *prescribed dose*. The main goals of treatment planning include:

1) Uniform PTV coverage: The dose to the PTV should be between 95 – 105% of the prescribed dose.
2) Avoiding hot spots: Hot spots are areas that receive a dose of more than 110% of the prescribed dose. Hotspots often occur when two beams are placed close to each other and the radiation overlaps. Hotspots can be reduced by increasing the weight of the wedges of the overlapping beams, increasing the weight of the beams or balancing, i.e. creating a second hotspot opposite to the original one, which results in both hotspots being of a lower dosage. The closer hot spots are to the organs at risk the more undesirable they are.
3) Conformance of radiation to the planning target volume: The PTV contains the tumour. The treatment planning parameters have to be adjusted so that the radiation follows the edge of the PTV as closely as possible. This is done by first adjusting the number and angle of beams and then fine tuning the arrangement by changing the multileaf collimator settings.

The organs at risk (OAR) that have to be avoided by the radiation beams include: the left and right eye, the lens of the eyes, the optic chiasm, the left and right optic nerve, the brainstem and the spinal cord. The treatment aims are customised to the patient. In palliative care, which focuses on pain management, the dose at times is compromised to spare OAR. For example, the dose could be reduced to avoid damage to the eyes. In radical treatment, tumour control is more important and the eyes might not be spared if they can be treated later on using surgery. Sometimes a compromise between conforming the radiation to the tumour and avoiding critical organs is made. For example, in order to avoid irradiating the eyes, the PTV is shaped to include a little of the brain.

5.1.2 Treatment Planning Procedure at the City Hospital

This section outlines the manual treatment planning procedure at the City Hospital. Planning is done based only on the images, i.e. using only the contours of the planning target volume (PTV) and organs at risk (OAR) as drawn by oncologists on the computed tomography (CT) slices. At the City Hospital, the radiation dose is prescribed according to a protocol for brain cancer. The beam configuration is then adjusted iteratively to achieve the prescribed radiation dose. For primary tumours, 70 Gy is administered in 7 weeks. For postoperative tumours, 66 Gy is administered if there are residual cancer cells remaining even after surgery and 60Gy if no residual cancer cells are found. Treatment planning is done manually using a trial and error approach, which can be viewed as a greedy constructive search method. The steps involved in manual treatment planning at the City Hospital are described below.

1) The oncologist views the CT and magnetic resonance imagining (MRI) images and outlines all structures of interest (planning target volume (PTV), clinical target volume (CTV), gross target volume (GTV) and the organs at risk (OAR) on the CT slices of a patient using a software called PROSOMA, which is a 3D simulation and visualisation software tool for radiotherapy (MEDCOM, 2012). On PROSOMA, the CT images can be fused with the corresponding MRI images to view soft tissue, such as the tumour, with reference to bony structures, such as the skeleton.
2) The CT slices are transferred to the treatment planning system ONCENTRA (Nucletron, 2011) for treatment planning.
3) First an initial beam is placed. There are no restrictions or guidelines for beam placement except that it should encompass the tumour area and avoid the OAR.
4) A *normalisation point* is set within the PTV. This is the point that receives 100% of the dose. All doses and isodose lines are relative to this point. The *isocentre* is the point at which all beams meet. Usually the normalisation point and the isocentre are identical. However, at times they need to be offset to ensure that the normalisation point lies within the tumour tissue (for example, if there are metal structures in the body).
5) The first beam is normally placed to enter the head at a location that is as close to the tumour as possible. For example, if the tumour is on the left superior side of the brain,

then the first beam would be placed so that it enters the head from the left superior side and passes through the isocentre.

6) Gradually more beams are placed such that the radiation conforms to the PTV outline.
7) Once the beams have been placed provisionally, the resulting dose distribution is viewed using the treatment planning system. If there are any dose violations, or if there is room for improvement of the dose distribution, then wedges are added, beam and wedge weights and angles are changed and the collimator leaves are modified.

Step 6) and 7) are repeated till a satisfactory dose distribution is obtained. This means that the entire PTV receives 95 - 105% of the prescribed dose and no healthy tissue or OAR is overdosed.

The medical physicists usually generate between three to four treatment plans and show them to the consultants along with disadvantages and advantages of each plan. Based on the oncologist's inputs the physicists may modify the treatment plan.

5.2 Input Data

Each patient study consists of a set of images taken during one or several scans such as MRI or CT scans. Each scan or image acquisition consists of a series of images, also called slices, which show the patient anatomy at different cross sections or orientations. The MRI scans show soft tissue such as the tumour or OAR and the CT scans show bony or skeletal structures and provide an anatomical reference. As seen in the previous section, treatment planning mainly depends on the geometrical location of the PTV and the OAR. Manual planning is performed based on a visual examination of the PTV and OAR outlines on the CT and MRI patient images. Since planning is done based exclusively on the patient CT and MRI images, these form the sole inputs to the CBR system.

5.2.1 The DICOM Standard

The patient images are in DICOM format. DICOM or *Digital Imaging and Communications in Medicine* (National Electrical Manufacturers Association) is a standard for storing, printing and transmitting clinical images. DICOM was developed by the

American College of Radiology (ACR) and National Electrical Manufacturers Association (NEMA) in order to ensure interoperability of medical imaging equipment and to facilitate communication with other networked devices such as servers, workstations and printers, used in the hospital. Equipment from multiple vendors can be interfaced as long as they are DICOM compatible. The standard consists of 20 separate parts, which can be found on the website of NEMA: http://medical.nema.org/. An overview of the salient points of DICOM along with practical advice on its usage and applications can be found in (Bidgood et al., 1997). More details on DICOM images used in this research work and excerpts from DICOM CT, RS and RP files can be found in Appendix A.

General DICOM Data File

A DICOM data file consists of both image information and header or textual information such as the patient ID or the hospital name. The header of a DICOM file is optional and included in the data file. It consists of a 128 bytes preamble, followed by 4 bytes called the DICOM prefix. The information is encoded as data elements. The structure and encoding of information is described in part 3. 5 of the DICOM standard (National Electrical Manufacturers Association, 2011a).

DICOM RT

Over the years, supplements and extensions have been added to the DICOM standard to deal with the particulars of medical specialities. Between 1997 and 1999, seven radiotherapy (RT) objects were created, described in supplement 11 (National Electrical Manufacturers Association, 1997) and 29 (National Electrical Manufacturers Association, 1999). The seven RT objects consist of the RT Structure Set (RS), RT Plan (RP), RT Dose (RD), RT Image (RI), and RT Treatment Record. RT Treatment Record is further divided into RT Beams Treatment Record, RT Brachy Treatment Record and RT Treatment Summary Record (Law and Liu, 2009).

In the City Hospital, a radiotherapy patient directory usually contains the CT DICOM image slices, RS, RP, RD and RI DICOM files. In order to extract data, we converted the DICOM files into text files using the open source DICOM toolkit *DCMTK* from the Offis Computer Science Institute (Offis, 2009). *DCMTK* is written in ANSI C and C++ and comprises a collection of libraries and applications that deal with DICOM files.

RT Structure Set (RS)

The RS DICOM images contain information regarding the structure outlines as drawn by the oncologist on the patient image. Examples of structures, also called regions of interest (ROI) include the GTV, CTV, PTV, OAR, body contour and reference points. Each ROI is numbered and described using tag descriptions such as *ROIName*, *ROINumber*, and *ROIDisplayColour* among others. The structure outlines are recorded in the form of their [x\y\z] coordinate triplets with the data element tag *(3006,0050)* called *ContourData*. Each ROI can have several *ContourData* fields, often one for each image slice. These files do not contain any pixel data but only textual information.

RTPlan (RP)

The RP images do not contain any pixel data either but only textual information regarding the treatment plan parameters. Among others, they record the isocentre coordinates, the beam number and for each beam provide the gantry, collimator and patient couch angle, beam weight, use of wedges and their weights, and multi leaf collimator settings.

RT Dose (RD)

RD images contain information about the prescribed dose and the dose distribution such as the isodose lines, dose volume histogram (DVH) values, etc.

RT Image (RI)

RI images are acquired or calculated using conical geometry such as digitally reconstructed radiographs (Law and Liu, 2009).

RT Treatment Record

The three RT treatment record objects added in 1999 include the RT Beams Treatment Record, RT Brachy Treatment Record, and RT Treatment Summary Record (Law and Liu, 2009):

1) RT Beams Treatment Record: This RT image contains textual information generated by the treatment planning system regarding the beam configuration, equipment details and dose information.

1) RT Brachy Treatment Record: In brachytherapy, the radiation does not originate from an external source such as a linear accelerator but from a radioactive material that is placed inside the body close to the tumour. If brachytherapy is used, treatment information is contained in the RT Brachy Treatment Record.
2) RT Treatment Summary Record: This record summarises information regarding both external beam radiation and brachytherapy.

5.3 The CBR System

Figure 5.1 shows the main components of our CBR system. The case base contains cases of previously treated patients. The cases in the case base consist of the case description in the form of case attributes and the treatment plan parameters. The case attributes are grouped into two groups, explained in section 5.3.1. This section also describes in detail the case representation and the treatment plan parameters. Given a target case, the retrieval mechanism first filters out from the case base a group of cases, which are comparable to the target case with respect to group I attributes (as explained in section 5.4). The filtered cases from the case base are then made available for similarity calculation. The similarity with respect to group II attributes between the target case and every case of the filtered case base is computed and the most similar case and its treatment plan are retrieved. The solution of the retrieved case is used in the solution of the target case. The retrieved treatment plan can be presented to the medical physicists as a starting point for planning or can be passed on to the adaptation stage. The scope of this thesis only includes retrieval, however. Adaptation will be carried out in future research work (as described in Chapter 10). The remainder of the thesis presents work that focusses on two parameters of a treatment plan; the number of beams and the beam angles. It has to be noted that the entire treatment plan of a case is retrieved, which includes the number of beams, their angles, wedges, the leaf settings of the collimator, etc. However, currently the CBR system is designed to retrieve treatment plans with suitable beam numbers and beam angles for the target case as these are determined in the first step of manual treatment planning by medical physicists in the City Hospital. Once the number of beams and their beam angles are determined, the other parameters are investigated. A possibility would be to determine only the beam angles as these implicitly define the number of beams. However, the medical physicists at the City Hospital attach more importance to the beam number as it is more straightforward to adjust the angles once a good number of beams has been found.

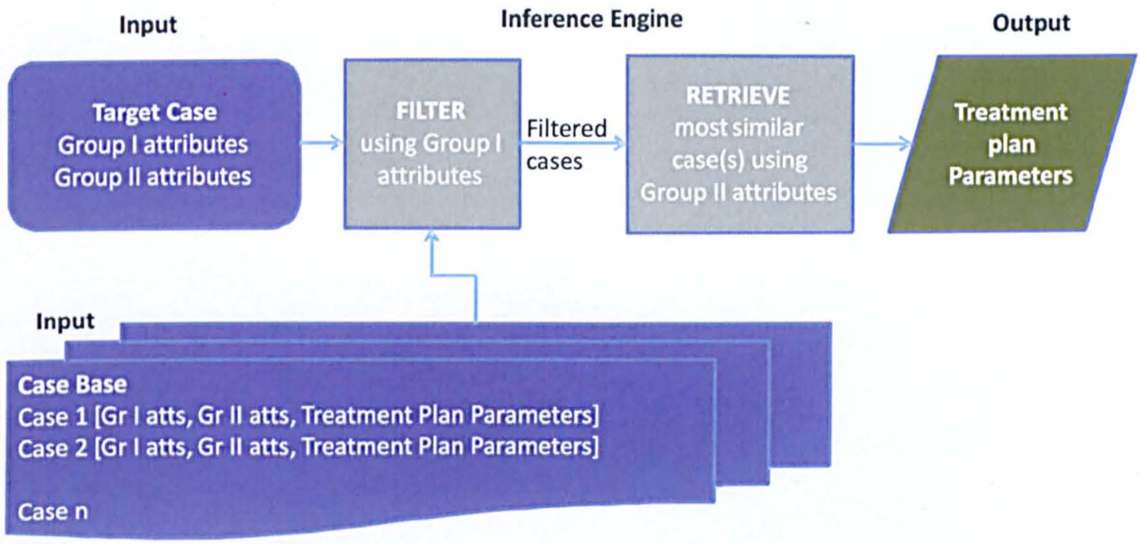


Figure 5.1: CBR system overview

5.3.1 Case Representation

Cases are stored in the case base in the form of key value pairs called case attributes or case features. The case attributes have to be carefully selected so that the similarity calculation between two cases accurately represents the relative applicability of the solution of a case to the target case. In the proposed CBR system, the case attributes have been selected after consultation with hospital staff at the City Hospital. Since treatment planning aims to focus the prescribed radiation on the PTV while avoiding the OAR, the treatment plan parameters are largely determined based on the location and dimensions of the PTV and the spatial relationship between the PTV and the OAR. These attributes determine the geometry of a patient and are computed using the 3D coordinates representing the outlines of the PTV and OAR structures, which are extracted from the DICOM RS files. We have identified eight case attributes that describe the patient in terms of attributes relevant to the treatment plan. The case attributes shown in Table 5.2. can be divided into two groups: Patient information and geometrical descriptors.

Table 5.2: Case attributes containing patient information and geometrical descriptors.

Attribute	Attribute label	Weight label	Attribute values	Data type	Similarity calculation
Group I: Patient Information					
OAR	-	-	String [left lens, right lens, chiasm, left optic nerve, right optic nerve, brainstem	nominal	Exact match
Patient Position	-	-	String [HFS, HFP]	nominal	Exact match
Group II: Geometrical Descriptors					
Angle	A	w_A	0 – 360 Degrees	continuous	Partial match using similarity measure
Distance	E	w_E	mm	continuous	Partial match using similarity measure
Volume	V	w_V	mm^2	continuous	Partial match using similarity measure
Body PTV Ratio	R	w_R	Ratio	continuous	Partial match using similarity measure
PTV Body Distance	Dt	w_{Dt}	mm	continuous	Partial match using similarity measure
PTV OAR Position	P	w_P	[0,1]	logical	Partial match using similarity measure

Group I Attributes:

The attributes in group I are used to filter the case base according to the requirements of the target case. The data type of these attributes is nominal and they require an exact match during retrieval. These attributes have been selected mainly for practical purposes to extract from the case base cases that are comparable to the target case. The procedure of filtering the case base in described in section 5.4. The group I attributes include:

- Organs at Risk (OAR): Depending on the location of the tumour, the oncologist decides which OAR are to be considered when generating the treatment plan. These OAR are outlined on the patient images and their names are recorded by the treatment planning system in the DICOM RS Structure file with the tag *(3006,0026)* described as *ROIName*. Currently, the CBR system only retrieves cases that consider the same OAR as the target case. The reason for this is practical in nature for ease of implementation. The group II attributes are computed with respect to the PTV and each OAR. The cases in the case base do not all include the same OAR. Therefore calculating the similarity, between cases that have different OAR and therefore consider a different number of attributes (with respect to each OAR) is not straightforward, in particular, as it is not currently known why not all cases contain information about a standard set of OAR. The reason for this has to be investigated if the entire case base irrespective of OAR is considered in every retrieval. A concern is that filtering the case base with respect to OAR reduces the cases available for retrieval for a target case and in the future we will investigate the reason for missing OAR information and how to implement a retrieval mechanism that can consider all cases in the case base.
- Patient Position: The patient position denotes the position in which the patient is lying on the patient couch at the time of imaging and during radiotherapy treatment. At the City Hospital, this can be either head first prone (HFP) or head first supine (HFS). Head first means that the head of the patient is positioned towards the front of the equipment. Supine or prone mean that the patient lies face up or face down, respectively, on the patient couch. This attribute is important since the 3D coordinate information in the DICOM image is given relative to the patient position. The value of this attribute is found in the CT DICOM image file under tag (0018,5100) described as 'PatientPosition'. Currently, the CBR system only retrieves cases with the same patient position as the target case

Group II Attributes - Geometrical Descriptors

The geometrical descriptors shown in Figure 5.2 describe the PTV and the spatial relationship between the PTV and the OAR. These attributes are used in the similarity measure of the retrieval mechanism. The geometrical descriptors are mainly computed using the 3D coordinates of the structure outlines obtained from the DICOM RS image files.

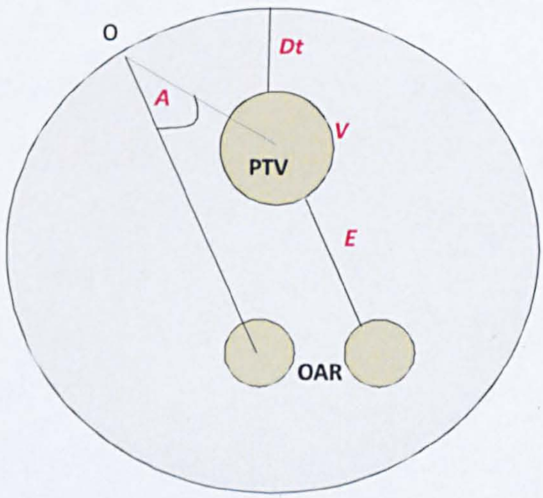


Figure 5.2: Group II attributes, A, E, V and Dt

Angle *A*

The angle *A* (in degrees) gives information about where the tumour is located in relation to the OAR and is calculated as the dot product angle between the line connecting the centre of the PTV and the origin of the DICOM image coordinate system and the line connecting the centre of the OAR and the origin as shown in Figure 5.2. The centre coordinates of the PTV and the OAR are computed using the structures coordinates as described in section 5.3.3. This value is calculated for each OAR separately.

Distance *E*

The distance is the minimum edge to edge distance connecting the outline of the PTV and the OAR given in mm as shown in Figure 5.2. This value is calculated for each OAR. The distance is significant since a large distance allows for more flexibility when placing the beams whereas a smaller distance between PTV and OAR requires a more complicated plan with possibly a large number of beams.

Volume V

The volume is defined as the volume of the PTV given in mm^2 as shown in Figure 5.2. It is either directly extracted from the DICOM structures file or it is estimated from the structures coordinates as described in section 5.3.3.

Body – PTV volume ratio, *R*

This is the ratio of the tumour volume to the volume of the entire patient body. Similar treatment plans frequently have similar PTV – body ratios, even if the actual PTV or body volumes differ. The volume of the body is estimated as described in section 5.3.3.

Body – PTV distance, *Dt*

This attribute denotes the minimum edge to edge distance in mm between the outline of the PTV and the outline of the body as shown in Figure 5.2. This value provides positional information about the tumour and describes how close the tumour is to the surface of the body.

PTV – OAR Spatial Relationship *P*

This is defined by the relative position of the PTV with respect to the OAR. This attribute is divided into six positional labels that take values [0,1] depending on if they are true or false as shown in Table 5. 3

The data type of geometrical descriptors is continuous and therefore, they are matched partially during similarity calculation, i.e. the difference in the value of the attribute is an indication of how similar the target case is to a case in the case base. In order to determine the total similarity between the target case and the cases in the case base, the partially matched attributes have to be combined in a similarity measure.

Table 5. 3. Values taken by case attribute *P*.

Positional label	**Value**	**If positional label value = 1, then**
Left	0 or 1	The PTV is placed left of the OAR
Right	0 or 1	The PTV is placed right of the OAR
Inferior	0 or 1	The PTV is placed inferior to the OAR
Superior	0 or 1	The PTV is placed superior to the OAR
Posterior	0 or 1	The PTV is placed posterior to the OAR
Anterior	0 or 1	The PTV is placed anterior to the OAR

5.3.2 Case Solution – Treatment Plan - Decision Parameters

The solution of a case consists of the treatment plan. For a case stored in the case base, this is the plan that was used to treat the patient. For the target case, the treatment plan has to be determined and constitutes the output of the retrieval mechanism of the CBR system. As described in section 2.1, the treatment plan parameters include the beam number, beam angles, beam weights, wedges, multileaf collimators settings, etc. So far, we have considered the two main parameters, i.e. the number of beams and their angles in the developed prototype CBR system. These parameters are found in the DICOM RP image file of a previously treated patient.

Number of beams:

In order to reduce the localised radiation received by healthy tissue, the radiation is applied using several beams of lower intensity that enter the body at different angles. This means that the total dose to the PTV remains the same but the dose received by healthy tissue in the path of each beam is reduced.

In the City Hospital, for brain cancer patients, the number of beams ranges normally from 2-6. A large number of beams reduces irradiation of healthy tissue in the path of the individual beams. Also, better conformance to the tumour is possible with a large beam number. However, if too many beams are used, the patient treatment time becomes too long, which increases the strain on the patient and also the hospital load. For palliative patients (whose treatment concentrates on improving the quality of life rather than tumour control), in general only 2-3 beams are used. This is to reduce treatment time and therefore patient discomfort. For radical patients (where treatment focuses on tumour control), at least 3 beams are used. In general, once the beam number is fixed, it is easier to tweak angles rather than change the beam number of a retrieved plan. The number of beams of a treated patient is given in the DICOM RP file using tag (300a,0080) and tag description *NumberOfBeams* as shown in Figure 5.3. The value representation *IS* stands for *integer string*.

```
(300a,0080) IS [3]                # 2, 1 NumberOfBeams
```

Figure 5.3: Number of beams as shown in DICOM RP file.

Angle of Beams

The angle of a beam determines the point at which the radiation beam enters the patient's head to pass through the isocentre of the tumour. The angle of a beam is adjusted to avoid irradiating the OAR. The beams can be coplanar (lying in one plane) or non-coplanar (lying in multiple planes), which increases the possible number of angles. The total angle consists of the gantry angle of the linear accelerator and the angle of the patient couch. The gantry angle is the angle by which the gantry has been rotated. The patient couch can also be rotated. This angle is called the patient couch or the patient support angle. Together they determine the angle at which the radiation enters the body of the patient. These two angles have to be specified for each beam applied. The DICOM representation of the gantry angle and the patient couch angle are shown in Figure 5.4. The value representation *DS* stands for *Decimal String*.

```
(300a,011e) DS [270.00]           # 6, 1 GantryAngle
(300a,0122) DS [0.00]             # 4, 1 PatientSupportAngle
```

Figure 5.4: Gantry angle and patient support (patient couch) angle as given in DICOM RP file.

5.3.3 Attribute Extraction and Data Pre-processing

This section describes how information is extracted from the DICOM image files and converted into the case attributes or plan parameters. The case attributes are computed for the cases in the case base as well as the target case. The treatment plan parameters, which constitute the archived case solutions, are computed for the cases in the case base only. The data is pre-processed using scripts written in MATLAB and C++. The pathway of this process is shown in Figure 5.5.

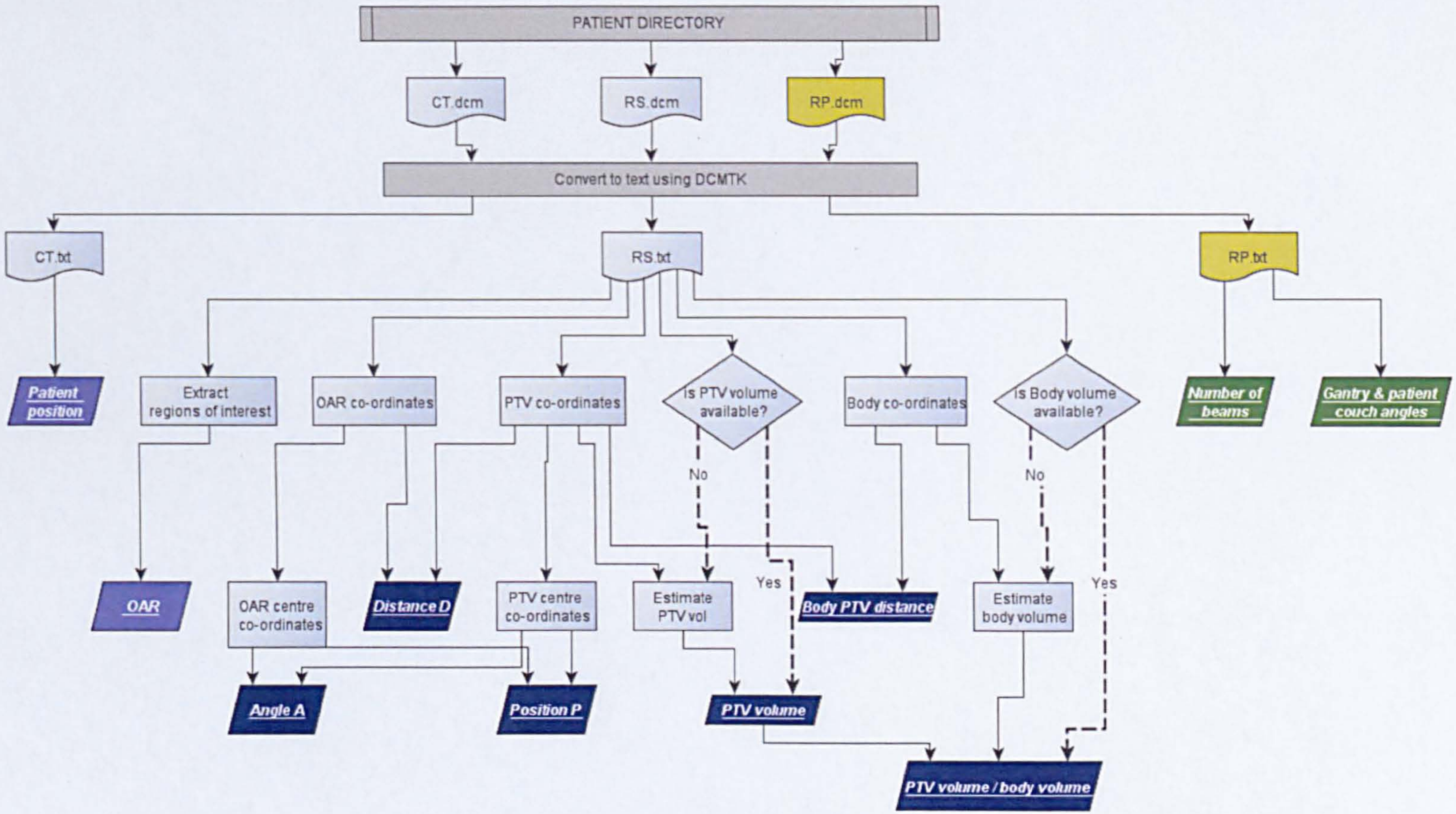


Figure 5.5: Data pre-processing pathway and extraction of case attributes

The radiotherapy structures (RS), plan (RP) and the CT DICOM image files are converted into text files using the *DCMTK* library utilities.

Group I attributes

The group I attributes are determined as follows:

1) The OAR are extracted directly from the converted RS text file.
2) The patient position is extracted directly from the converted CT text file.

Group II Attributes

The regions of interest (ROI), recorded in the RS file, and the 3D coordinates of their outlines are extracted for the *body*, *PTV* and *OAR*. The centre point (x_{Centre}, y_{Centre}, z_{Centre}) of the PTV and OAR structure outlines is computed using the 3D coordinates. The group II case attributes are computed as follows:

1) The angle *A* is calculated using the centre points of the PTV and each OAR of the patient. The angle *A* is recorded separately for each OAR.
2) The case attribute distance *D* is given by the minimum edge-to-edge distance between the PTV and each OAR. The distance *D* is recorded separately for each OAR.
3) In some RS files, the volume of the regions-of-interest has been recorded by the clinician at the time of outlining the structures on the files. In those cases, the PTV volume *V* (or body volume) can be extracted directly. The corresponding tag is (3006,002c) described as ROIVolume. However, in most patient RS files, this data element is missing and has to be estimated. This is done as follows: Using the 3D structures coordinates of the PTV, a closed surface is created using a utility MATLAB script called MyRobustCrust.m from the MathWorks File exchange (Luigi, 2009a). MyRobustCrust.m is a simple surface reconstruction program that uses a crust algorithm to return a tight triangulated surface from a set of 3D points. The volume enclosed by the surface is then computed using the MATLAB script SurfaceVolume.m (Luigi, 2009b). The accuracy of volume estimation was determined by running the volume estimation MATLAB script on the patient files, for which

the volume was recorded by the City Hospital and looking at the difference between the recorded and the estimated volume shown in Table 5.4. The average error of 0.87 mm^3 was deemed acceptable.

Table 5.4. : Error between estimated PTV volume and true PTV volume

Case	Estimated PTV volume	True PTV volume	Absolute error
2	432.3096	432.989	0.679391
5	34.89097	34.708	0.182968
8	79.50744	77.113	2.394438
9	211.2388	211.149	0.089837
10	261.9985	262.816	0.817459
11	229.9911	229.45	0.541146
12	446.3048	444.887	1.417784
Average error			0.87471757

4) The volume of the body is either extracted from the RS file if available or estimated as described in the previous point. The ratio of the PTV and body volume is calculated to give attribute *R*.
5) The minimum edge-to-edge distance *Dt* between the body structure outlines and the PTV structure outline is computed using the 3D structure coordinates.
6) The position of the PTV relative to an OAR (attribute *P*) is computed by determining if the centre point of the PTV structure is left/ right, superior/inferior or anterior/posterior to the centre point of the OAR structure. This attributes is also recorded separately for each OAR.

The first five case attributes of group II, i.e. *A, E, V, R* and *Dt* have to be normalised so that they can be used in the similarity calculation. An attribute value v_l is normalized using expression 5.1:

$$v_{N,l} = \frac{v_l - v_{min}}{v_{max} - v_{min}} \qquad 5.1.$$

Treatment Plan parameters

The treatment plan parameters are determined for the cases in the case base. The number of beams, gantry angle and patient support angle of each beam are extracted directly from the RP text file. The gantry angle and the patient support angle together constitute the plan parameter beam angle.

5.4 Filtering using Group I attributes

Prior to retrieval, the CBR system identifies cases in the case base, which share the same group I attributes with the target case. This is done partly to capture similarities arising due to group I attributes but mainly in order to keep the treatment plans compatible and comparable. For instance, the beam angles in a treatment plan for a patient lying on the patient couch with patient position *HFS* (head first supine) have to be interpreted differently for a patient who is *HFP* (head first prone). In addition, presenting only cases with the same OAR as the target case to the retrieval mechanism, simplifies the similarity calculation. In future research work, the CBR system can be modified to allow retrieval of cases with different patient position or OAR to the target case. For example, if a *HFP* case is retrieved for a *HFS* target case, the beam angles of the treatment plan could be rotated to be *HFS* compatible. The cases filtered from the case base with respect to group I attributes are available for the next stage. The retrieval mechanism then computes the similarity between the target case and each of the filtered cases in order to determine the most similar case to the target case.

5.5 Conclusion

This chapter presented an overview of the CBR system. The architecture has been designed with a focus on radiotherapy treatment planning, however, the general design concepts such as filtering the case base initially for specific attributes can be used in other CBR systems. The DICOM patient image data, which is used by the medical physicists for treatment planning is also the sole input data to the CBR system. The data obtained from the DICOM files is processed to extract the case attributes and also the treatment plan parameters. The DICOM standard is widely used in hospitals and clinics in the UK and extracting the data, whenever possible, from the textual key-value pairs in the DICOM file is a simple and quick alternative to using image processing algorithms. The inference engine consists effectively of two stages: filtering of the case base and retrieval of the most similar case to the target case. In the following sections, the term case attributes refers to the six geometrical descriptors, i.e. the group II attributes unless otherwise specified. The design of the similarity measure is described in detail in the following chapters. Validation concerns are discussed in Chapter 6. The weighted nearest neighbour (wNN) method and attribute weighting is discussed in Chapter 7. Chapter 8 outlines a variation of the similarity measure using fuzzy sets. Retrieval is carried out in two stages, where each stage focuses on retrieving a treatment plan with respect to one of the treatment plan parameters. This process is described in Chapter 9.

Chapter 6

Validation

Validation refers to the process of ensuring that the proposed system fulfils its intended purpose. The extent to which the system fulfils this purpose is known as its performance or success rate. A validation fault or error in a CBR system usually manifests itself as the suggestion of an ineffective solution to the current problem. In our case, this means that the treatment plan suggested by the CBR system is not suitable to provide a starting point to treat a new patient (or is suitable for adaptation).

Though CBR systems have been widely studied in the literature there is no standard method of validating them. A thorough evaluation of a completed and ready to use clinical decision support system should include validation at different levels, for instance:

- System level (for example, retrieval accuracy or correctness)
- User level (user friendliness or user satisfaction)
- Process level (time or cost gained through use of automated system)
- Clinical outcome level (health improvements as measured by clinical markers or improvements in quality of life, number of complications, increase in expected life span)

In this work, validation and evaluation mainly refer to quantifying the performance of the retrieval mechanism at system level.

Althoff (1997) describes a comprehensive evaluation process of the INRECA CBR system that includes qualitative descriptions with respect to predefined decision criteria and quantitative analysis of the prediction accuracy on training cases, retrieval and building speed. McSherry (2001) uses the concepts of retrieval *precision* and *recall* to evaluate interactive CBR systems. *Precision* refers to the number of cases among the retrieved cases whose solution is relevant to the target case whereas *recall* refers to the number of retrieved cases among all cases in the case base whose solution is relevant to the target case (Brüninghaus and Ashley, 1998). The most commonly applied validation measure, however, is retrieval accuracy, that is the number of times that the solution of the retrieved case is relevant to the target case (Bonzano et al., 1997b, Liao et al., 2000, Bellazzi et al., 1998, Petrovic et al., 2011). This is often expressed as the error between the retrieved solution and the known (or expected solution) of the target case. In this work, validation mainly refers to retrieval accuracy (also called retrieval error).

Validation can fulfil two purposes, performance estimation or model selection. In performance estimation, the accuracy or success rate of the chosen model is computed. In model selection, a technique of cross validation is used to determine the parameters of the inference engine by using feedback from the performance error. Cross validation for model selection is discussed further in section 6.2. In this work, validation has two distinct purposes:

Estimation of the performance of the retrieval mechanism in the CBR system: A number of different methods have been designed to retrieve the most similar case from the case base with respect to the target case. The different methods, in particular, variations on the similarity measure and attribute weights determination, are compared based on their retrieval performance or the retrieval error. Computation of the retrieval error is described in section 6.1.1.

To guide the determination of free parameters in the similarity measure (using a wrapper approach): The free parameters refer to variables such as attribute weights, the value of k in the k-nearest neighbour method or the weights of the fuzzy sets in the fuzzy, non-linear similarity measure. The nearest neighbour similarity measure and weights determination is discussed in Chapter 7 whereas the fuzzy similarity measure is introduced in Chapter 8.

The performance of many decision support systems is evaluated by human experts. The solution generated by the decision support system is compared to the solution suggested by an expert with thorough knowledge of the application domain (for instance, a medical physicist in the case of radiotherapy treatment planning). Human validation carries the risk of introducing subjectivity or bias into the validation. A more significant and prohibitive issue is the high cost of using human experts who often do not have the time to carry out a sufficiently extensive evaluation of the solutions suggested by the decision support system.

Another possibility is to use the dose distribution of the treatment plan of the retrieved case as applied to the target case as validation. This requires the computation of the radiation dose deposited in the patient's OAR and PTV. A treatment plan is considered as valid if the applied dose complies with the medical dose description. However, computing the deposited radiation dose is not trivial and currently, this functionality does not exist in the developed CBR system. Further, we do not have free access to Oncentra, the treatment planning system used by the medical physicists in the City Hospital to view the computed dose distribution of treatment plans. CERR (Computational Environment for Radiotherapy Research) (Deasy et al., 2003) is a software platform for developing and sharing research results and functions and radiotherapy treatment planning. It is also capable of displaying the dose distribution of treatment plans. Unfortunately, it cannot be used currently with non-coplanar beams.

The problems outlined above can be avoided by validating the system using test cases whose solution is already known. In CBR systems, in particular, the human expertise is intrinsically available in the cases in the case base (Gonzalez et al., 1998). The solutions of the cases in the case base are known. If a case with a known solution (called a test case) is made the target case, then the difference between the retrieved solution and the known solution of that case is an indication of the performance of the retrieval mechanism.

According to Gonzalez et al. (1998) the salient points that have to be considered in validation with respect to a CBR system include the selection of validation criteria, the test case set design and the development of the test drivers. These can be applied to both model selection and performance estimation and are described below in more detail.

This chapter discusses the design considerations and techniques used to both validate the CBR system and to determine the free parameters of the retrieval mechanism. Section 6.1 discusses the salient points of validation mentioned by Gonzalez et al. with respect to the developed CBR system. Section 6.2 introduces cross validation techniques, with a focus on the techniques that were used in this work to determine the free parameters of the retrieval mechanism. The challenge of learning free parameters from small sets of training cases is very briefly addressed. The baseline random retrieval error of the CBR system is presented in section 6.3. This error gives an indication of the random retrieval performance which measures the CBR system as compared to a system randomly retrieving a case from the case base.

6.1 Validation Considerations

This section describes the factors to be taken into account during validation according to Gonzalez et al. (1998).

6.1.1 Validation Criteria

The performance of the system is judged based on the validation criteria. The validation criteria provide the standard against which the output of the CBR system is judged. The validation criteria should be relevant to the solution of the CBR system. In many cases, the output parameters of the solution can be used as validation criteria themselves. Another aspect of the validation criteria is how close the output of the system has to be to the desired standard (Gonzalez et al., 1998). The standard can be established either by human experts or by known results of test cases.

Gonzalez et al. (1998) suggested that the test cases used to evaluate a CBR system can be obtained directly from the case base. According to medical physicists at the City Hospital, for each patient case there exists a unique configuration of treatment plan parameters that constitute the best treatment plan for that patient. We are very aware that this assumption might not always be valid in reality. Research done on generating radiotherapy treatment plans using optimisation methods, have shown that in some instances, the automated optimisation techniques were able to generate treatment plans that were superior with respect to the dose distribution than the plans generated manually by medical physicists (Y. Li et al., 2005). However, as computing the dose distribution is currently not feasible, the evaluation in this research work was based on comparing the treatment plans of retrieved test cases with the known treatment plans of the test target case. The medical physicists believe that the treatment plans of the cases in the case base constitute good, successful treatment plans and the evaluation of the system is based on this assumption. A point to support this assumption is that all the treatment plans of the cases in the case base have had a positive treatment outcome for the patient. The limitation of this approach is that the generated treatment plans currently, can only be as good in quality as the existing treatment plans in the case base. Another concern is that if there exists the possibility that different possible treatment plans are valid to treat one patient case, the CBR system might validly retrieve a treatment plan but it would be considered as an unsuccessful retrieval if the retrieved treatment plan though equally valid is different to the existing treatment plan of the test case that was generated by the planner. Keeping these limitations in mind, the treatment plans of the cases in the case base are viewed as the standard that the CBR retrieval mechanism should aim to predict.

Therefore, the validation criteria to evaluate the performance of the retrieval mechanism of the developed CBR system are provided directly by the solution parameters or the treatment plan parameters i.e. the number of beams and their angles. More specifically, it is the error between the beam number and beam angles of the retrieved treatment plan and the known treatment plan of a test case that constitutes the validation criteria.

Beam Number Error E_{BN}

E_{BN} denotes the error or difference in the beam number *BN* between the retrieved treatment plan and the expected value (obtained from the known treatment plan of the target case). If the number of beams is exactly the same in both the retrieved treatment plan and the known treatment plan of the test case, the retrieval is deemed successful. The strict limit on E_{BN} has been imposed to reflect the fact that the beam number is difficult to adapt from the retrieved plan since changing the beam number of a treatment plan requires changing all the angles as well. Therefore, the standard of the validation criteria beam number error is given by $E_{BN} = 0$.

Beam Angles Error E_{BA}

E_{BA} denotes the error in the beam angles between the retrieved treatment plan and the expected value (obtained from the original treatment plan of the target case). Expressing the error between the beam angles of two treatment plans is not as straightforward as the beam number error. E_{BA} denotes the difference between the angle of a beam in one treatment plan and the angle that is numerically closest to it in the other treatment plan. In other words, to obtain a fair estimate of the error, the beams are paired up so that the distance between the two beams of a pair is minimised. If the number of beams in the treatment plans is not the same, then the angles of the extra beams are not considered in the error calculation. Also, the angles are circular about 360°.

For instance, consider *treatment plan 1* with gantry angles at 70°, 110° and 350° and *treatment plan 2* with gantry angles of 10° and 60°. If we pair up the closest angles and discard the extra angle in *treatment plan 1*, we obtain a gantry angle error of $E_{BA\_Gantry} = |70° - 60°| + |350° - 10°| = 10° + 20° = 30°$.

The error is calculated separately for the gantry angle error E_{BA_Gantry} and patient support angle error E_{BA_PS}. The total error is given by $E_{BA} = E_{BA\_Gantry} + E_{BA\_PS}$.

As the value of E_{BA} is continuous and lies in the interval of [0,360], it cannot be expected that the gantry and patient support angle of the known treatment plan of the target case and the treatment of the retrieved case have exactly the same values. In other words, it is highly unlikely that E_{BA}=0°, even when treatment plans are similar. However, for the sake of computing the success rate of the retrieval with respect to beam angles, an upper limit of E_{BA} has to be set, above which the retrieval is deemed unsuccessful and below which the retrieval is deemed as successful. In discussions with medical physicists at the City Hospital it transpired that an average beam angle error of less than approximately 30° per treatment plan was deemed as an acceptable error margin, since an average angle difference of less than 30° does not substantially influence the dose distribution. Also, below 30°, it would be easy for the medical physicists to adjust the beam angles by small amounts to achieve the desired dose distribution. Therefore the upper limit of the beam angles error was set as E_{BA} <30°. However, we are aware that this upper limit has been set empirically. It would be interesting to view the dose distributions of treatment plans with differences in their beam angles error and track the corresponding changes in dose distribution.

The pseudo code of the procedure to calculate E_{BN} and E_{BA} is shown in Figure 6.1.

```
// Find the difference in beam number E_BN and corresponding beam angles
E_BA between case 1 and case 2


Read beam number (BN1) from RP DICOM file from treatment plan 1
Read beam number (BN2) from RP DICOM file for treatment plan 2
Read gantry and patient couch angle list (GA1 & PA1) from treatment
plan 1
Read gantry and patient couch angle list (GA2 and PA2) from treatment
plan 2


for (m=1 to BN1)
    for (n=1 to BN2)
         for each PAIR(m,n)
          Calculate GA_DIFF(m,n)= GA1(m)- GA2(n)
          Calculate PA_DIFF(m,n)=PA1(m) - PA2(n)
         end
    end
end
// Total number of pairs = BN1 * BN2
//Number of possible combinations of GA or PA pairs = BN1*BN2 /
min(BN1, BN2)
if BN1 < BN2
    possible combination set of GA/PA pairs S = BN2
    size of S = BN1
else
    possible combination set of GA/PA pairs S = BN1
    size of S = BN2
end
p=1
for (m=1 to BN1)
    for (n=1 to BN2)
         Calculate GA_SUM(p) = sum(GA_DIFF(m,n))
         Calculate PA_SUM(p) = sum(PA_DIFF(m,n))
         p = p+1
    end
end
Find MIN_GA_SUM = min(GA_SUM(p))
Find MIN_PA_SUM = min(PA_SUM(p))
GA_ERROR = MIN_GA_SUM
PA_ERROR = MIN_PA_SUM
Calculate TOTAL_BA_ ERROR E_BA = GA_ERROR + PA_ERROR
Calculate TOTAL_BN_ERROR E_BN = |BN1 - BN2|
```

Figure 6.1: Pseudo code for calculating the beam number error E_{BN} and the beam angle error E_{BA} between the retrieved treatment plan and the actual treatment plan

6.1.2 Test Cases

The data used to validate a system consists of two distinct sets: A training data set and a test data set. The training data set is used to determine the free parameters of the system, such as the weights of the similarity measure. The test set is used to estimate the performance of the system. In order to obtain an unbiased evaluation, it is important that the test cases have not been seen by the system. In other words, the test cases have not been used to design the system (as opposed to training cases, which are used to make design decisions about the system). The test case set should be generic enough to test the system but not so extensive that testing becomes too expensive and impractical (Gonzalez et al., 1998).

Currently the test set consists of 22 brain cancer cases randomly selected from the case base.

6.1.3 Test Drivers Development

According to Gonzalez et al. (1998), test drivers development refers to the automated process of evaluating the system using test cases. The test cases are sequentially made the target case. For each test case, the most similar treatment plan is retrieved. The retrieval error E_{BN} and E_{BA} between the beam number and angles of the retrieved treatment plan and the beam number and angles of the known treatment plan are calculated. The average retrieval error in terms of E_{BN} and E_{BA} of all test cases gives an indication of the performance of the system.

6.2 Cross Validation

Cross-validation methods can be used to both estimate the accuracy of a system and to train the system. In training, the accuracy error of the retrieval mechanism is used to guide the selection of free parameter values. For example, we have utilised cross validation techniques to find the attribute weights in the similarity measure as described in Chapter 7 using training cases.

Frequently, the training cases constitute 2/3 of the entire data set and the test cases 1/3 (Kohavi, 1995). A larger test set at the expense of the training set increases the error of the system, whereas a smaller test set might not be able to precisely predict the error of the system. In this work, the training set consists of 64 cases and the test set consists of 22 cases. All cases are real brain cancer patient cases obtained from the City Hospital.

Cross validation techniques are employed in order to obtain the maximum information from the training data available. The simplest method to evaluate the performance of a system (be it for model selection or performance estimation) would be to use the entire data set (that is, all available cases with known solutions) and calculate the average error of all cases. If the available data is extremely limited, the validation process runs the risk of the results over fitting the training data. This means that the validation results using the training data cannot be generalised to another data set, since the parameters of the system were chosen to only fit the one set of training data. This is especially problematic if the available training data set is small and the number of parameters that have to be determined is large. The error obtained tends to be overly optimistic as it represents the *best case* scenario, i.e. the parameters have been fine tuned to give the lowest error on this data set. The aim of cross validation techniques is to maximise the information by splitting the available training data into exclusive subsets.

Common cross-validation techniques include leave-one-out cross-validation, *k*-fold cross validation and boot strapping among others.

The following sections describe the computation of the retrieval error used to give an indication of the performance of the retrieval mechanism and the cross validation methods used to determine the free parameters of the system.

6.2.1 Retrieval Error Computation

In this work, 22 brain cancer patient cases obtained from the City Hospital are used exclusively as the test set in order to estimate the performance of the retrieval mechanism. It is important to note that this test set is different to the validation sets discussed in the cross validation techniques below. The case base consists of 86 cases. Each test case *(i)* is consecutively made the target case, the most similar case is retrieved and the retrieval error with respect to the beam number E_{BNi} and the beam angles E_{BAi} is calculated. The error obtained over 22 cases is averaged to obtain E_{BN} and E_{BA}.

$$E_{BN} = \frac{1}{n}\sum E_{BNi} \qquad 6.1.$$

$$E_{BA} = \frac{1}{n}\sum E_{BAi} \qquad 6.2.$$

where i, $i = 1,2,3....22$, is the index of the current target case and n, $n = 22$, is the number of test cases.

6.2.2 k-Fold Cross Validation

In *k*-fold cross validation the data set is randomly divided into '*k*', usually equally sized, subsets. In each run or fold, one of the subsets is made the test or validation set and the other *(k-1)* subsets constitute the training data. The process of cross validation is then carried out *k* times so that each subset forms the validation set exactly once. The results from each fold are then averaged over *k* times. The subsets are randomly selected. However, to avoid introducing a bias into the results more than one data splits can be used. In this case, the entire *k*-fold cross validation procedure is repeated a fixed number of times. Each repetition randomly partitions the data differently so as to introduce variation in the way the cases are grouped into validation and training sets. The results of each repetition are averaged again. A low error is an indication of high performance accuracy on the training cases. However, if the variation in error is high, then this means that the results might not be generalizable to other data. For this reason, it is important that not only the error but also the variance in error between folds is taken into account.

In the context of the developed CBR system, one subset is treated as the set of target cases and the other *k-1* subsets make up the case base. In each fold, for each target case of the current subset, the most similar case is retrieved and the beam number error, E_{BN}, and the beam angle error, E_{BA}, between the retrieved treatment plan and the known treatment plan of the target case are calculated. The error values obtained for each target case in the *kth* set are averaged to give E_{BN_k} and E_{BA_k}. This process is repeated *k* times so that each case has been made the target case exactly once and has acted as a case in the case base *k-1* times. This results in *k* values of E_{BN_k} and E_{BA_k}, which are averaged again *k* times to give E_{BN_AVG} and E_{BA_AVG}. Then the entire dataset is partitioned randomly again into *k* folds and the process is repeated and the error results of each repetition are averaged again to obtain E_{BN_AVG3} and E_{BA_AVG3}. The index '3' in E_{BN_AVG3} and E_{BA_AVG3} is due to three repetitions, which seem to be a good trade-off between accuracy and run-time in the weights determination algorithm. The variance between the average errors in *k* folds is computed and averaged over all repetitions. The variance is given by expression 6.3.

$$v = \frac{1}{n-1}\sum_{i=1}^{n}\left(E_i - E_{avg}\right)^2 \qquad 6.3.$$

Where *i* is the index of the current target case, *n* is the number of cases in each subset and E_i is the retrieval error for the target case (E_{BNi} in beam number retrieval or E_{BAi} in beam angles retrieval).

Selecting the number of folds *k*: The number of folds, in general, is selected based on the number of training cases available. A value of *k*=10 is common. Increasing the number of folds, reduces the performance error of the system on the training data, however, it increases the variation. Another disadvantage is that the computation time increases with the number of folds. A lower value of *k* reduces the computation time and also the variation. The error tends to be higher and more conservative.

The pseudo code to calculate E_{BN_AVG3} and E_{BA_AVG3} is given in Figure 6.2. The average variance is calculated similarly. Henceforth in this thesis, for the sake of clarity, E_{BN} and E_{BA} denote both E_{BN_AVG3} and E_{BA_AVG3} of cross validation and also the average beam number and beam angles retrieval error obtained when testing the performance of the retrieval mechanism using test cases.

The advantage of using *k*-fold cross validation is that the available data set is used efficiently. Each case is used both as target case and in the case base. Overfitting is avoided by using both the average error as well as the average variance in error between folds.

```
// Computation of cross validation error: k fold cross validation

Set of cases available for cross validation: S_All
Total number of cases available for cross validation: T
Number of folds: k
Number of cases in each subset: n
T = k*n
Number of repetitions = 3

for (z=1 to 3)
    for (x=1 to k)
          Set of current target cases = S_x
          Set of cases in case base = S_All - S_x
          for (y=1 to n)
               Target case = C_T,x,y
               Compute similarity between C_T,x,y and each case in case
               base
               Retrieve most similar case
               Calculate E_BN_ky and E_BA_ky
          end
```

$$E_{BN\_kx} = \frac{1}{n}\sum_y^n E_{BN\_ky}$$

$$E_{BA\_kx} = \frac{1}{n}\sum_y^n E_{BA\_ky}$$

```
    end
```

$$E_{BN\_Avg} = \frac{1}{k}\sum_x^k E_{BN\_kx}$$

$$E_{BA\_Avg} = \frac{1}{n}\sum_x^k E_{BA\_kx}$$

$$z = z + 1$$

```
end
```

$$E_{BN\_Avg3} = \frac{1}{3}\sum_x^k E_{BN\_Avg}$$

$$E_{BA\_Avg3} = \frac{1}{3}\sum_x^k E_{BA\_Avg}$$

Figure 6.2: Calculation of *k*-fold cross validation error

6.2.3 Leave-One-Out Cross Validation

Leave-one-out cross validation is a special case of *k*-fold cross validation, where *k* is equal to the number of all training cases available. In the context of a CBR system, each case in the training data is consecutively made the target case and the remaining cases, constitute the case base. If the number of training cases is n then n retrieval runs are performed. In each run, the number of cases in the case base is n-1, with the one remaining case constituting the target case. In general, leave-one-out cross validation can be used if the data available for training is very small since it improves the accuracy of the system on the training cases. However, the variance in error between the folds increases substantially and over fitting is a common problem. This means that the results are not very generalizable. However, leave-one-out cross validation is a popular technique in CBR systems to estimate the performance of a CBR system (Burke et al., 2006, Cheetham and Price, 2004, Mishra et al., 2009). Another disadvantage is that if the data set is very large, this method can be very time-consuming.

6.2.4 Considerations Arising from Small Training Sets

The design of decision support systems for real world applications often suffers from the problem of having only a small set of training data available, in particular, during the initial design stages. In this work, additional methodologies have been designed that take into account the small size of the training data set to ensure that the feedback obtained from the retrieval error calculation during cross validation is both reliable and generalizable. Besides considering both the variance in error between folds and the average error when using *k*-fold cross validation, an attempt has been made to assess the quality of the feedback obtained from the retrieval error based on the number of cases available for retrieval having suitable treatment plans for the target case (i.e. E_{BN} =0 or $E_{BA} \leq 30°$). The developed methodology is applied to the local case attributes weighting scheme described in section 7.2. Further, an alternative feedback parameter based on the contents of the case base instead of the absolute value of the retrieval error is introduced in section 7.3.

6.3 Baseline Random Retrieval Accuracy

The baseline random retrieval accuracy refers to the retrieval accuracy achieved if a random case was retrieved from the case base rather than the case most similar to the target case as determined. The retrieval accuracy when using an intelligent retrieval mechanism such as a similarity measure is expected to be much higher than the random retrieval accuracy. This test ensures that cases are not just selected at random by the retrieval mechanism (Beddoe and Petrovic, 2006).

To calculate the average random retrieval accuracy, the 22 test cases are consecutively made the target case and a random case is retrieved from the case base. The beam number and beam angle error between the retrieved treatment plan and the known treatment plan of each case is computed. The average beam number and beam angle errors E_{BN} and E_{BA} of all 22 test cases are calculated as described in section 6.2.1. This process is repeated 10 times and the results are averaged to give an estimate of the random retrieval accuracy. The random retrieval accuracy in terms of the beam number error RE_{BN} and the beam angles error *REBA in* our case base is shown in Table 6. 1

The success rate is the percentage of the number of target cases out of all target test cases with E_{BN}= 0 with respect to beam number retrieval and $E_{BA} \leq 30°$ with respect to beam angles retrieval. The retrieval error, the number of correct retrievals and the success rate are used as indicators in this work to measure the performance of the retrieval mechanism with respect to the beam number error. RE_{BN} and RE_{BA} constitute the base line random retrieval error against which the performance of the developed concepts in the retrieval mechanism are measured, as explained in the chapters to follow.

Table 6. 1: Random beam number and beam angles accuracy

	Average error	Success rate (%)
Random beam number error RE_{BN}	0.77	27
Random beam angles error RE_{BA}	40.48°	36

6.4 Conclusion

This chapter presented the techniques used to validate the various retrieval mechanisms and variations on the similarity measure. They are put in the context of the developed CBR system prototype for radiotherapy treatment planning. The standards against which the quality of the retrieval is measured is currently provided by the treatment plans that were manually generated by medical physicists for the existing cases in the case base while making a working assumption that these treatment plans represent good and feasible treatment plans. The method of computing the retrieval error between the beam number and beam angles of the retrieved plan and the known plan of a training or test case was described. It also introduced cross validation, which will be used to determine the free parameters of the retrieval mechanism, such as the case attribute weights, the value of *k* in the k-nearest neighbour method or the weights of the fuzzy sets in the fuzzy, non-linear similarity measure. It has to be noted that the results of cross validation to determine the parameters of the system using the case base are necessarily heavily dependent on the contents and the coverage of the case base. This means that a low retrieval success rate is not necessarily a reflection on the performance of the chosen design parameters in the retrieval mechanism but could be due to the fact that a suitable case is not present in the case base. This issue is discussed in detail in section 7.3 and a methodology that takes this point into account when determining the design parameters is proposed and implement with promising results. The baseline random retrieval accuracy of the system on the test cases was calculated in order to obtain a baseline for comparison with intelligent retrieval mechanisms. According to Smyth and Keane (1998), retrieval should be adaptation guided, which means that during retrieval preference should be given to retrieve solutions that are easy to adapt rather than necessarily represent the largest similarity between cases. Currently, adaptation guided retrieval is not considered in this work. However, in the future, it would be interesting to keep this possibility in mind when defining the retrieval error. The remainder of the thesis focuses on the weights determination in the wkNN similarity measure (Chapter 7), the fuzzy non-linear similarity measure (Chapter 8) and the two-phase retrieval mechanism (Chapter 9). These methods are compared by applying the method described in 6.2.1 on the 22 test cases obtained from the City Hospital.

Chapter 7

The wNN Similarity Measure

The most important stage of the retrieval mechanism is the similarity computation. Following the filtering of compatible cases for the target case, the similarity between the target case and each case in the case base (or a selected subset of cases in the case base) is calculated with respect to the case attributes. The case (or a number of cases) with the largest similarity is retrieved. The treatment plan of the retrieved case (or cases) is used to form the solution of the target case. Since CBR systems are based on the notion of similar cases having similar solutions, the definition of similarity is crucial as the CBR system depends on a good retrieval engine that is capable of retrieving cases whose solution is relevant to the target case.

In the retrieval mechanism of the developed CBR system the similarity is computed using the k - weighted nearest neighbour algorithm (wkNN), described in section 3.5. The nearest neighbour (NN) algorithm, traditionally used in classification and pattern recognition problems to assign objects to classes, has been widely used in CBR (Chang et al., 2012, Kwong et al., 1997, Ahmed et al., 2011) owing to its ease of implementation and the fact that it does not make any assumptions about the distribution of the underlying data. In NN classification, an object is classified by assigning it to the known class of its nearest examples (or neighbours) in the solution space. The solution space in CBR can be viewed as a collection of clusters, where each cluster contains similar solutions (Blanzieri and Ricci, 1999). The nearest neighbour is found using a similarity measure based on the input space.

When the attribute values are numeric and continuous in nature, then a commonly used distance metric is the Euclidian distance (Cost and Salzberg, 1993), which is used in the wNN similarity measure in the developed CBR system.

Let C_T be the target case and C_C be a case from the case-base with attributes A = angle between planning target volume (PTV) and organ at risk (OAR), E = distance between PTV and OAR, V = PTV volume, R = ratio between PTV and body volume, Dt = distance between nearest edge of body to PTV and P = position of PTV with respect to OAR. The Euclidean distance, D_{wNN}, between cases is calculated as follows:

$$D_{wNN} = \sqrt{\sum_{l=A,E,V,R,Dt,P} w_l\left(v_{T,l} - v_{C,l}\right)^2} \qquad 7.1.$$

where, $v_{T,l}$ and $v_{C,l}$ denote the attribute values of attribute l, where l = A, E, V, R, Dt, P of target case C_T and case C_C respectively. The similarity S_{wNN} between C_T and C_C is given by:

$$S_{wNN} = 1 - D_{wNN} = 1 - \sqrt{\sum_{l=A,E,V,R,Dt,P} w_l\left(v_{T,l} - v_{C,l}\right)^2} \qquad 7.2.$$

Not all attributes contribute equally to the similarity calculation. The weight of attribute l, denoted by w_l, indicates the relative importance of an attribute. Careful selection of the attribute weights is crucial in order to ensure that the solution of the most similar case really is suitable to apply to the target case. A special case of attribute weighting is attribute selection, in which the weights of an attribute can take values [0, 1]. Often attribute weights are set with the help of domain experts. However, a lot of work has been carried out in the literature on both attribute selection and weighting in order to develop automated methods of weight determination as discussed in section 3.6.

In section 5.3.1, we described geometrical descriptors, which attempt to capture factors that are considered important by treatment planners at the City Hospital during manual planning. However, due to the differences between manual planning and an automated decision support system and owing to the complex and subjective nature of treatment planning, these attributes are essentially empirically determined and their actual usefulness as predictors of the solution in the similarity measure needs to be confirmed. Irrelevant or wrongly weighted attributes can add noise and reduce the accuracy of the similarity measure.

In order to determine the set of relevant attributes and their weights we used a wrapper method. The application of wrapper methods in attribute weighting and selection has been discussed in detail in section 3.6. In wrapper methods, the search for attributes or attribute weights is guided by feedback from the system on its performance on training cases. The attribute weights used in the similarity measure are initialised to an arbitrary value. The training cases consist of brain cancer patients treated in the past. Therefore, the solution of these cases, that is the treatment plan parameters, is known. The aim is to find attribute weights, which result in an average low retrieval error on the training cases. One of the training cases is made the target case and the most similar case from the case base is retrieved. The difference in the treatment plan parameters of the retrieved case and the known treatment plan parameters of the target case, gives an indication of the performance of the retrieval mechanism and hence, the weights used in the similarity measure. This difference in the treatment plan parameters constitutes the retrieval error as described in section 6.1.1. The weights set, resulting in the minimum error, are selected to be used in the similarity measure. Usually a number of training cases is used and the average retrieval error is fed back to the inference engine to alter the weights accordingly.

Wrapper methods are popular in attribute weighting algorithms since they incorporate the actual inference or classification engine (Saeys et al., 2007). They inherently take into account the correlation between features and are also simple to design. A major disadvantage is the risk of over fitting, in particular, when the number of training cases available is small or biased. Wrapper methods can also be computationally expensive depending on the size of the training data set, the number of attributes, the range of permissible values that each attribute weight can take and if an exhaustive or heuristic search is carried out.

The remainder of this chapter describes the weights analysis carried out to determine the importance of the attributes independently and in relation to the attribute values in the target case. Section 7.1 describes the determination of global attribute weights, which take the same value every time the retrieval mechanism is run unlike local context sensitive attribute weights, whose values change based on the attribute values of the target case as described in section 7.2. In section 7.3 an alternative method of obtaining feedback about the retrieval performance is proposed that takes into account the contents of the case base, which is an important consideration in small case bases.

7.1 Determination of Global Attribute Weights

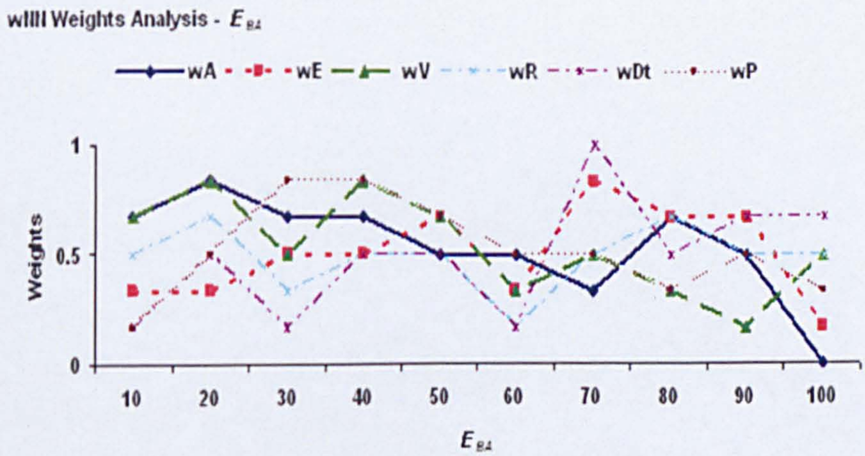
Global attribute weights are constant, i.e. the value of the attributes weights are the same every time the similarity calculation algorithm is called during the retrieval stage of the CBR system. In contrast, local weights (discussed in section 7.2.), can take different values every time the similarity calculation algorithm is called.

The weights are usually expressed in values taken from the range [0, 1]. The binary situation results in feature selection, where '0' signifies that the attribute is not used in the similarity calculation and '1' signifies that the attribute is one of the main contributors to the similarity calculation. The interval or step between the values that a weight can take is selected based on the problem domain and if the increase in computational complexity is acceptable or not. Kohavi et al. (1997) demonstrated that using a vector of weights with more than one or two non-zero values, made their algorithm more unstable, since the large number of weights tried to fit the training data rather than modelling the true weights of the attributes. In our experiments, the attribute weights were allowed to take values from the set [0, 0.5, 1], which provided a good trade-off between accuracy and computational complexity and had a reduced risk of over fitting. The aim of the attribute weighting algorithm is to find a vector of attribute weights used in the similarity measure, which accurately describes the importance of each attribute with respect to the treatment plan parameters and therefore results in a low retrieval error. The vector of attribute weights is denoted by W_n, where $W_n = [w_A, w_E, w_V, w_R, w_{Dt}, w_P]$, $n = 1 \ldots 729$, resulting in $3^6 = 729$ combinations of W_n. The following sections describe the experiments performed to accurately determine the attribute weights to be used in the similarity measure of the developed CBR system.

7.1.1 Preliminary Results

The similarity measure was initially trained on a reduced data set containing 41 brain cancer cases obtained from the City Hospital along with the treatment plans that were used. In order to find a weight vector, W_n, that resulted in a low retrieval error on the training cases, the leave-one-out cross validation algorithm described in section 6.2.3 was used. Each case in the case base was consecutively made the target case and the remaining 40 cases constituted the case base. For each target case, the most similar case in the case base was retrieved using the similarity measure shown in expressions 7.1 and 7.2. The system was set to retrieve only the most similar case, i.e. k=1. The weights in the similarity measure took values from the set [0, 0.5, 1.0] which resulted in 729 weight combinations, for weight vector $W_n = [w_A, w_E, w_V, w_R, w_{Dt}, w_P]$, n=1...729. For every weight vector, the treatment plan of the retrieved case was compared to the known treatment plan of the target case and the beam number error E_{BN} and the beam angles error E_{BA} were computed as described in section 6.1.1. The error values E_{BN} and E_{BA} were averaged over all 41 cases. To prevent over fitting the weights to the data, we avoided selecting the weight vector that resulted in the lowest error. Instead, the trend in the variation in error with respect to W_n was studied. To obtain a visual representation of any correlation between the attribute weights W_n and the error, the weight vectors were ranked and plotted in ascending order of beam number error E_{BN} and the beam angles error E_{BA}. Then a moving average of the weights was obtained with a period *PE*, where *PE* is equal to 10% of the number of experimental runs. A moving average was used to smooth the data and remove short term fluctuations. An experimental run refers to a retrieval process for a single target case with a given attribute weight vector. Figure 7.1 shows the results obtained for the beam angle error E_{BA} and beam number error E_{BN} versus the averaged attribute weights. Along the x-axis of the graph, the cases are ranked in ascending order of retrieval error. The y-axis represents the weights vector averaged over 10% of the ranked cases resulting in the retrieval error. From the graph, it appears that as the values of angle weight w_A and the volume weight w_V reduce, E_{BA} increases. With respect to the beam number error E_{BN}, the error appears to increase as w_E reduces. However, from the graph it is difficult to draw a conclusion about the attributes *R*, *Dt* and *P*.

The conclusion, we can draw from the graph is that there is a probable relationship between the attribute weights of attributes *A*, *E* and *V* and both the beam angle error E_{BA} and beam number error E_{BN}. We observed trends in the attribute values versus treatment plan parameters to avoid the risk of over fitting the weights due to the small case base. The advantage of the leave-one-out cross validation technique is the simple implementation, short running time and the fact that it utilises all the information available in the training cases. However it also increases the risk of over fitting. We carried out a full weights analysis using a larger case base and the more robust *k*-fold cross validation technique described in the following section.



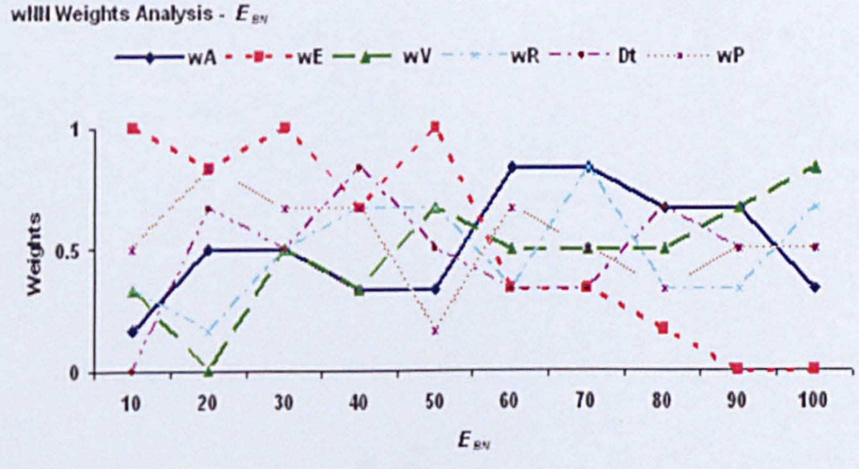


Figure 7.1: Graph of the attribute weights versus beam angle error E_{BA} and beam number error E_{BN} with weights, $l = A, E, V, R, Dt, P$

7.1.2 Full Case Base - Weights Determination

The preliminary results were obtained on our initial (small) case base of 41 cases using leave-one out cross validation. During our research work more real world brain cancer cases were collected from the City Hospital, which enabled us to accurately determine the attribute weights and an appropriate similarity measure. The experiments were repeated on the larger case base of 86 cases along with the treatment plans used.

To determine the attribute weights, a *k*-fold cross validation technique was used (described in section 6.2.2), where *k* is the number of folds. For the sake of clarity, we will henceforth denote the number of folds as *n* so as to avoid confusion with parameter *k*, which denotes the number of cases retrieved by the retrieval mechanism of the CBR system. We used 64 cases (out of 86) in the weights determination; the remaining 22 cases from the case base form the test set. These 64 cases were divided into *n* sets of training cases, where $n = 4, 8, 16$. We consecutively made each set of training cases the target set and used the other $n - 1$ sets as the case base. The similarity was calculated between the target cases and each case in the case base using expression 7.1 and 7.2. The difference in the beam parameters (i.e. beam number and beam angles) was calculated between the treatment plan of the retrieved case and the original treatment plan of the target case. E_{BN} and E_{BA} form the average errors of the beam number and beam angles of the treatment plans, respectively. The results were averaged over *n* sets. To allow for variation in the distribution of training cases in the sets, this experiment was repeated three times; each time the *n* sets contained different cases. Again, the results were averaged. The entire experiment was repeated for each combination or vectors of weight values, where weights took values from the set [0, 0.5, 1]. Then, the weight vector that offered the lowest average error values was determined. To avoid over fitting the results to the data, we noted both the error values, *E*, for each weight set and the average variance in the error between folds, *V*, so as to ensure robustness of our findings.

To find the best trade-off between *E* and *V*, *E* and *V* were normalised to take values between [0,1] and the average of their sum, A_{EV}, was used to determine the best weight vector. Normalisation was done by considering the values of *E* and *V* as cost so that larger values of A_{EV} indicate a superior weight vector resulting in lower values of *E* and *V*. We also studied, which value of *n*, i.e. the number of folds, used in the cross validation, would give us a larger value of A_{EV} and therefore identify a more robust method of cross validation. As expected, for both the beam number and the beam angles, V increases considerably as the value of n increases but with little improvement in E. A good trade-off was obtained using n=4, which was therefore used during cross validation. Table 7. 1 shows the weight vectors that resulted in the largest value of A_{EV} on the training cases both with respect to the beam number and the beam angles and with respect to k, the number of cases retrieved, where k takes values from [1,3,5]. In the case of beam number retrieval, we can see that the smallest value of A_{EV} is obtained with weight vector W = [0.5, 1, 0.5, 1, 1, 1] when three cases are retrieved, i.e. k =3. The error and variance is substantially larger in the un-weighted case, i.e. W = [1, 1, 1, 1, 1, 1], which demonstrated that weights do play an important role in this similarity measure. Though the difference in E and V between k=1, k=3 and k=5 is not large, we use k=3 in the developed CBR system for beam number retrieval. In the case of beam angles retrieval, we can again see that the un-weighted retrieval results in much larger values of E and V. A small value of A_{EV} is obtained with k=1 or k=3. However, since the error is considerably lower and the variation only slightly larger for k=1, we use in the developed CBR system, k=1 for beam angles retrieval.

An interesting finding was that better results were obtained when separate weight vectors were used to retrieve the treatment plan to suggest the number of beams and the treatment plan to suggest the beam angles. In other words, the importance of attributes depends on the treatment plan parameter. Henceforth, in our experiments, the retrieval mechanism was run twice for each target case, first using the weight vector that resulted in the lowest beam number error (and variance during k-fold cross validation) and then using the weight vector that resulted in the lowest beam angles error (and variance during k-fold cross validation). A framework to use this kind of multi-phase retrieval is presented in Chapter 9, which discusses the two phase retrieval mechanism in detail.

Table 7. 1: Results of cross-validation error and variance in error between folds on training cases

Number of cases retrieved, k	w_A	w_E	w_V	w_R	w_{Dt}	w_P	Training cases error, E	Training cases variance, V	Summed average of normalised E and V, A_{EV}
Beam Number									
k=1	0	0	0	0	0	0.5	0.38	0.007	0.92
k=3	0.5	1	0.5	1	1	1	0.36	0.006	**0.93**
k=5	0.5	1	1	0.5	0	0	0.36	0.007	0.92
k=3	1	1	1	1	1	1	0.40	0.013	0.64
Beam Angles									
k=1	1	1	1	1	0.5	0.5	23.46	16.447	**0.93**
k=3	0.5	1	0.5	1	1	1	27.88	16.160	**0.93**
k=5	0.5	1	1	0.5	0	0	28.22	19.618	0.78
k=1	1	1	1	1	1	1	25.57	22.949	0.76

7.1.3 Evaluation of Global Weights Using Test Cases

We evaluated the performance of the weighted nearest neighbour similarity measure using 22 test cases, consisting of real brain cancer patient cases obtained from the City Hospital, with the global weights determined in the previous. Validation was done using the method described in section 6.2.1. The 22 test cases are consecutively made the target case. For each target case, the most similar case(s) is retrieved and the error between the treatment parameters of the known solution of the target cases and the treatment plan of the retrieved case(s) is computed. The retrieval mechanism is run twice first with the weight vector W_{BN} = [0.5, 1, 0.5, 1, 1, 1], optimised to retrieve a treatment plan suggesting beam numbers and then with weight vector W_{BA} = [1, 1, 1, 1, 0.5, 0.5] optimised to retrieve a treatment plan suggesting the beam angles. The error in both situations is averaged over all test cases.

Table 7.2 shows the results obtained for both beam number retrieval (using weights vector W_{BN} = [0.5, 1, 0.5, 1, 1, 1]) and beam angles retrieval (using weights vector W_{BA} = [1, 1, 1, 1, 0.5, 0.5]). The error value denotes the average error obtained over 22 test cases (either E_{BN} or E_{BA}). The success rate refers to the number of successful retrievals, that is in the case of beam number retrieval, this value denotes the percentage of the number of cases out of all 22 cases where E_{BN} =0 and in beam angles retrieval, it denotes the percentage of the number of cases out of all 22 cases, where $E_{BA} \leq 30$ degrees.

We can see from the results that an intelligent retrieval mechanism can definitely retrieve relevant cases. However, the success rate is still rather low and the remainder of the thesis describes more sophisticated weighting methods and similarity measures that attempt to improve the success rate of the retrieval phase of the CBR system.

Table 7.2: Beam number and beam angles error obtained on test cases using wNN – global weights

k	w_A	w_E	w_V	w_R	w_{Dt}	w_P	**Error**	**Success rate (%)**
Beam Number Retrieval								
3	0.5	1	0.5	1	1	1	0.36	68.18
Beam Angles Retrieval								
1	1	1	1	1	0.5	0.5	32.27°	59.0909

7.2 Local Context Sensitive Weights

In the previous section, we used global attribute weights in the similarity measure. Context sensitive attributes are weighted based on the values of the attributes in the target case. Plotting single attribute weights versus the resultant beam number and beam angles error, E_{BN} and E_{BA}, does not show a clear or direct relationship. However, as described by staff at the City Hospital, it is conceivable that the importance of attributes changes with respect to their own value or the value of other attributes. For instance, medical physicists pay special attention if the tumour volume (PTV) is small, and therefore the number of possible beam directions that both avoid the organs-at-risk (OAR) and irradiate the PTV is limited. So if the target case has a small PTV, the importance of attribute *V* increases. Similarly, according to medical physicists, if the distance between the PTV and OAR is small, then the angles and position between the PTV and OAR become more important and consequently in a target case with a small distance value we could weight the angle and/or position between the PTV and OAR highly. In this manner, rules could be formulated such as "IF *V* is small THEN $w_V = 1$" or "IF *E* is small THEN $w_A = 1$ AND $w_P = 1$".

Previously, we plotted attribute weight values against the resultant average error both with respect to beam number and beam angles using a limited set of cases (Jagannathan et al., 2012). When visually examining the graphs, the error showed a variation with respect to the attribute weights. Rules were formulated that reflected attribute weights showing a small error on the training data. Preliminary experiments carried out on the reduced case base of 41 cases showed promising results.

This section presents a more accurate and objective method that we have designed for our current work to learn the weight assignment rules based on specific evaluation criteria of the retrieval error obtained from training data. In addition, this method also considers the effect of the correlation between attribute values on the importance of the attributes. The attribute values of the training cases are first assigned to two groups or clusters, *Large* or *Small* using the k-means clustering algorithm (as described in section 7.2.1). Attribute weights are assigned based on which clusters the attributes in the target case belong to. Then the rules are generated based on feedback about the retrieval performance on the training cases using the leave-one-out cross validation measure, as described in section 6.2.3. The rules are pre-screened using two rule evaluation measures known as support and confidence (described in section 7.2.3). Finally, for each combination of attribute values clusters, a rule is selected based on a novel concept called the random retrieval probability (*RRP*), which takes into account the likelihood of a successful retrieval being due to correct attribute weighting rather than a bias in treatment plan parameters in the case base (described in section 7.2.4). A flowchart delineating clustering, rule generation, pre-screening and selection is provided in Figure 7.2.

7.2.1 Clustering

Clustering can be defined as the unsupervised classification of objects, where unlabelled data is separated into discrete clusters (Rui and Wunsch, 2005). Objects in one cluster are similar, while objects in different clusters are dissimilar. A widely used clustering technique is the k-means algorithm, which is popular due to its ease of implementation, simplicity and efficiency. The k-means algorithm groups objects by minimizing the squared Euclidean distance between the mean of each cluster and the objects in the cluster (Jain, 2010).

Since Howe and Cardie (1997) suggested using different local weights for each attribute value is impractical and can lead to over fitting, we categorize the values of each attribute found in our case base into two groups, *Large* and *Small*. The local weights are then assigned to the clusters of an attribute rather than attribute values. The first step in determining the impact of attribute values on the six case attributes *A, E, V, R, Dt* and *P* is to define what attribute values constitute a large or small value for an attribute. The range of clusters *Large* and *Small* for each attribute is obtained from the training cases in the case base. Each case from the case base is assigned to one of the two clusters (*Large* or *Small*) based on the attribute value of the PTV-OAR distance *E* (*large E* or *small E*), PTV volume *V* (*large V* or *small V*), PTV-OAR volume ratio *R* (*large R* or *small R*) and body-PTV distance *Dt* (*large Dt* or small *Dt*). The PTV-OAR angle *A* and the position *P* are not strictly monotonically increasing, which makes it difficult to assign them to distinct groups such as *Large* and *Small*. For this reason, the effect of *A* and *P* on the importance of case attributes is not considered in this study. The effect of the other four attributes values on *A* and *P*, however, is taken into account. This means that each case is assigned weights for its six attributes and is a member of four clusters based on the values of its attributes *E, V, R,* and *Dt*.

The clusters are determined for each attribute separately using the k-means function from the MATLAB statistical toolbox. This function iteratively partitions the input data with the aim of minimising the total sum over all clusters obtained by summing the distance of each object within a cluster to the cluster centroid. The input to the k-means clustering function consists of the values of an attribute of all training cases in the case base. The distance between objects in a cluster and the cluster centroid is set to be the Euclidian distance. The k-means function is run with five replicates. In each replicate a different centroid point, randomly selected from the training data, is used as starting point and then the partitions with the lowest total sum of the distances of each object in the cluster to the cluster centroid, are determined. We also use an online update phase, which ensures that the solution is a local minimum. In other words, moving any single object to a different cluster would increase the total sum of distances.

The clusters that are created are not equal in size. This, however, is deemed acceptable since equal sized clusters result in less distinction between the clusters. However, we have restricted our work to only two clusters (*Large* and *Small*) per attribute since using more clusters results in the size of clusters being exceedingly small due to the limited available data for training. The ranges of attribute values found among the training cases and the centroid of the clusters *Large* and *Small* as determined by the k-means algorithm for each attribute are shown in Table 7.3. Each case is now represented by a vector of attribute clusters [E_{Cl}, V_{Cl}, R_{Cl}, Dt_{Cl}], where the subscript CL = [*Large*, *Small*] denotes the cluster that the attribute value belongs to.

Table 7.3: Range of attribute values and the centroids of their respective clusters *Large* and *Small*

Attribute	Attribute range	CentroidSmall (mm)	CentroidLarge(mm)
E	2.0mm - 99.8mm	18.95	53.02
V	24.9mm3 -729.8mm3	194.35	417.70
R	6.4 - 255.9	19.68	158.93
Dt	0.06mm - 53.7mm	9.38	44.70

7.2.2 Rules Generation

Once the case attributes *E*, *V*, *R*, and *Dt* of all cases are assigned to clusters *Large* and *Small*, the next step is to determine what effect a *Large* or *Small* attribute value has on the significance of the case attributes. The effect is formulated in the form of IF THEN rules. First, the candidate IF THEN rules are generated and then a number of rules are selected based on the rule evaluation measures *confidence* and *support*. A rule Rq, where q=1, 2,...n_A, n_A = number of rules or antecedents, can be expressed in the following form:

IF $[E_T, V_T, R_T, Dt_T] = Aq$ THEN $[w_{A,q}, w_{E,q}, w_{V,q}, w_{R,q}, w_{Dt,q}, w_{P,q}] = C_q$

where, $[E_T, V_T, R_T, Dt_T]$ denotes the attribute cluster vector of the target case, $Aq = [Eq, Vq, Rq, Dtq]$ is the antecedent of rule R_q, and weight vector, $C_q = [w_{A,q}, w_{E,q}, w_{V,q}, w_{R,q}, w_{Dt,q}, w_{P,q}]$ is the consequent of rule R_q. The weights, $w_{l,q}$, can take values from [0, 0.5, 1], where l = A, E, V, R, Dt, P.

Let n_A be the number of possible antecedents and n_C be the number of possible consequents that can be formulated. Since each antecedent vector consists of four attributes (*E*, *V*, *R* and *Dt*), which can take one of two possible values (*Large* or *Small*), $n_A = 2^4 = 16$. Therefore, we require 16 rules for weights assignment. As the weights can take values from the set [0, 0.5, 1], the number of consequents available for six attributes (*A*, *E*, *V*, *R*, *Dt* and *P*) is $n_C = 3^6 = 729$. This means that the number of rules, n_R, that are generated by a combination of antecedents and consequents is: $n_R = n_A * n_C = 11664$.

Given an antecedent A_q, q=1, 2,..n_A, all training cases that are compatible with antecedent A_q are identified and then used as the set of target cases $S_{A,q}$ during cross validation of rule R_q. In order to determine the retrieval error obtained with each rule, i.e. each antecedent-consequent combination, we use the leave-one-out strategy described in section 6.2.3. Each of the training cases in the identified set $S_{A,q}$ is consecutively made the target case and the remaining cases constitute the case base. For each target case, the most similar case in the case base is retrieved using expressions 7.1 and 7.2, where the consequent C_r, r = 1,2,.. n_C, supplies the attribute weights vector $[w_{A,r}, w_{E,r}, w_{V,r}, w_{R,r}, w_{Dt,r}, w_{P,r}]$. The retrieval error with respect to E_{BN} and E_{BA} is computed for each rule. A rule R_q is deemed feasible with respect to a target case, if the antecedent A_q matches the attribute values of the case and if the weights used in the similarity measure, supplied by consequent C_r result in a successful retrieval. During beam number retrieval, a successful retrieval occurs if the beam number in the retrieved plan is the same as the beam number in the known treatment plan of the target case, i.e. if E_{BN} =0. During beam angles retrieval, a successful retrieval occurs if the average difference in beam angles in the retrieved plan and the known treatment plan of the target case is smaller than an empirically pre-set threshold, i.e. $E_{BA} <= 30°$. The rules, which result in successful retrieval on the training cases, constitute the set of feasible rules.

7.2.3 Pre-screening Using Rule Evaluation Measures

It may happen that more than one consequent results in successful retrieval and is therefore associated with the same antecedent. From the set of feasible rules we need to find a limited set of 16 rules, which will uniquely assign a consequent or weight vector to each antecedent or attribute values vector. In order to determine the most appropriate and relevant rules, rule evaluation measures are used as constraints. Two rule evaluation measures commonly used in data mining are the confidence and support of a rule (Ishibuchi and Yamamoto, 2004). A higher confidence or support value indicates a more appropriate rule.

If D is a set containing m training cases then $D(A_q)$ is the number of cases, which are compatible with antecedent A_q and $[D(A_q) \cap D(C_r)]$ is the number of cases that are compatible with both antecedent A_q and consequent C_r. In other words, $[D(A_q) \cap D(C_r)]$ represents the number of cases with attribute values E_q, V_q, R_q, Dt_q in which the retrieval was successful when weights $w_{A,r}$, $w_{E,r}$, $w_{V,r}$, $w_{R,r}$, $w_{Dt,r}$, $w_{P,r}$ were used in the wNN similarity measures described by expressions 7.1 and 7.2.

The confidence, *con*, measures the validity of rule R_q. It is the percentage of all cases compatible with antecedent A_q that are also compatible with consequent C_r.

$$con\left(A_q \Rightarrow C_r\right) = \frac{|D(A_q) \cap D(C_r)|}{|D(A_q)|} \qquad 7.3.$$

The support, *sup*, measures the coverage of rule R_q. It is the percentage of all training cases, which are compatible with both antecedent A_q and consequent C_r.

$$sup\left(A_q \Rightarrow C_r\right) = \frac{|D(A_q) \cap D(WC_t)|}{\mathrm{m}} \qquad 7.4.$$

Though the confidence and support can directly be used as evaluation measures, according to Ishibuchi and Yamamoto (2004), the confidence criterion selects rules, which cover only a small number of compatible training cases but have a low retrieval error. The support criterion selects rules based on many compatible training cases but could result in a high retrieval error. They found that they obtained a good trade-off between generalisability and retrieval error on various different data sets when using the product *CSP* of the confidence and support, i.e.

$$CSP = con\left(A_q \Rightarrow C_r\right) * sup\left(A_q \Rightarrow C_r\right) \qquad 7.5.$$

For each antecedent, the rules with the largest value of *CSP* are selected. Then, among the pre-screened rules, one rule is selected for each antecedent using the random retrieval probability (*RRP*) rule evaluation measure described in section 7.2.4.

7.2.4 Rule Selection Based on Instance Weighting Using Random Retrieval Probability

From the pre-screened rules, a single rule per antecedent has to be selected. This is done by using a novel instance weighting algorithm that gives an indication of the quality of information gained from a retrieval instance. Not every successful retrieval indicates that a rule accurately describes the relationship between attribute significance and weights and will obtain good results outside the training phase when using unseen cases. If the case base is small or biased the average retrieval error based on the training cases can be skewed if the solution parameter values are not equally distributed. For example, let us assume that a large number of treatment plans in the case base use four beams and let us further assume that the target case happens to have four beams as well. This will result in a low retrieval error even though the number of beams in practice might not usually be 4 as is indicated by the training cases. In that situation a low retrieval error might not be indicative of the performance of the retrieval mechanism but might just mean that the probability of retrieving a case with the correct solution is very large since the solution parameter is uncharacteristically over represented in the case base. Therefore, we require a way to quantify the validity of a successful retrieval as opposed to a random retrieval. In a random retrieval, a case is retrieved at random from the case base without calculating the similarity of cases. The random retrieval probability (*RRP*) of an instance refers to the probability of a random retrieval being successful. In other words, what is the likelihood of successful retrieval if a random case is retrieved from the case base (instead of the most similar case) given a particular target case? If *RRP* is small the information we infer from the instance when using the weighted similarity measure is valid. If *RRP* is large, then we do not know if the retrieval is successful due to correct weights used in the similarity measure or due to a solution parameter bias in the case base, which results in a large likelihood of a randomly successful retrieval.

In our CBR system, we define the random retrieval probability *RRP*, which considers cases available after filtering based on OAR of the target case. For a given target case, let C_{Right} denote the number of cases in the filtered case base, where E_{BN} =0 in the beam number retrieval or $E_{BA} \leq$ 30deg in the beam angles retrieval. Let C_{Wrong} denote the number of cases, where $E_{BN} \neq 0$ or $E_{BA} > 30$ degrees. Then for a given target case, the random retrieval probability of a retrieval instance is given by:

$$RRP = \frac{C_{Right}}{C_{Right} + C_{Wrong}} \tag{7.6.}$$

where the number of cases in the filtered case base is given by $(C_{Right} + C_{Wrong})$. At special conditions, the following expression applies:

$$\begin{gathered} IF\ C_{Right} = 0\ THEN\ RRP = n/a \\ IF\ C_{Wrong} = 0\ THEN\ RRP = 1 \\ IF\ C_{Right} = C_{Wrong}\ THEN\ RRP = 0.5 \end{gathered} \tag{7.7.}$$

If there are no cases available for a given target case that would result in a successful retrieval, i.e. C_{Right} =0, then *RRP* becomes irrelevant or not applicable since then for the retrieved case, necessarily, $EBN \neq 0$ or $E_{BA} > 30$ degrees. On the other hand, if all cases available for retrieval would result in a successful retrieval, then *RRP* =1. If the number of cases that result in a successful retrieval and the number of cases that do not result in a successful retrieval are equal, then *RRP* is 50%. In other words, if *RRP* < 50%, then the retrieval is considered to be not entirely random.

For each rule (i.e. combination of antecedent and consequent) the retrieval mechanism is run and the number of successful retrieval instances over the training cases is noted. The *RRP* values of all successful retrieval instances are averaged to represent the average *RRP* of a rule. The average *RRP* of a rule constitutes the final rule evaluation measure to select a unique consequent for each antecedent. For a given antecedent, the consequent (from the set of pre-screened rules) with the lowest average *RRP* is chosen. If more than one consequent results in the same lowest *RRP* value, then one of those consequents is chosen arbitrarily.

//Rule generation mechanism for local attribute weighting

//m = number of training cases

//Clustering

FOR each training case

Assign attribute values v_E, v_V, v_R, v_{Dt} to clusters *Large* or *Small* using k-means algorithm

//Rule generation

FOR each antecedent A_q = [E_q, V_q, R_q, Dt_q] FROM 1 TO n_A

FOR each consequent C_r = [$w_{A,r}$, $w_{E,r}$, $w_{V,r}$, $w_{R,r}$, $w_{Dt,r}$, $w_{P,r}$] FROM 1 TO n_W

FOR each target case C_T FROM 1 TO n

IF [E_T, V_T, R_T, D_T] = A_q //antecedent matches target case attribute clusters

Antecedent counter $cnt_{A,q}$ = $cnt_{A,q}$ +1

//Rule pre-screening

Retrieve most similar case to target case using wNN similarity measure with

[$w_{A,r}$, $w_{E,r}$, $w_{V,r}$, $w_{R,r}$, $w_{Dt,r}$, $w_{P,r}$] = C_r

IF retrieved case has E_{BN} = 0 // during beam number retrieval

[OR]

IF retrieved case has E_{BA} $\leq$ 30degrees // during beam angles retrieval

THEN retrieval = successful // rule R_q: $A_q \rightarrow C_r$ is selected

Rule counter $cnt_{Aq,Cr}$= $cnt_{Aq,Cr}$ +1

Random retrieval probability of instance: RRP_T = $C_{Right}/(C_{Right} + C_{Wrong})$, $RRP \neq 0$

Confidence con = $cnt_{Aq,Cr}$ / $cnt_{A,q}$

Support sup = $cnt_{Aq,Cr}$ /n

Confidence support product *CSP* = *con* * *sup*

//Rule Selection

Average RRP of rule R_q = $\sum_{T=1,2,..n} RRP_T/n$

FOR each antecedent A_q FROM 1 TO n_A

select consequent C_q with highest *CSP*. Among selected consequents, select consequent C_q with lowest *RRP*.

Rule R_q : A_q => C_q

Figure 7.2: Pseudo code showing clustering, rule generation, pre-screening and selection of rules for local attribute weighting

7.2.5 Evaluation of Local Weights Using Test Cases

We tested the rule generation and evaluation algorithm for local weights assignment using real brain cancer patient data from the City Hospital. The set of training cases was of size 64, while 22 cases constitute the test cases. From the set of 11 664 possible rules that can be formed using all combinations of antecedents and consequents, 4340 rules resulted in successful beam number retrieval while 5096 rules resulted in successful beam angles retrieval. The two rule evaluation measures, i.e. the support-confidence product *CSP* and the average random retrieval probability *RRP*, were used to select 16 weights assignment rules for beam number retrieval and 16 rules for beam angles retrieval. The 22 test cases (with known treatment plans) were used as target cases and the most similar case was retrieved for each using the wNN similarity measure defined in expressions 7.1 and 7.2 with local weights assigned to the attributes of the target cases using the appropriate rules. The retrieval errors E_{BN} and E_{BA} for each target case were computed.

Table 7. 4 shows the average beam number error, E_{BN}, and the average beam angles error, E_{BA}, obtained using local weights over the test cases. The success rate refers to the percentage of test cases in which the retrieval was successful, that is $E_{BN} = 0$ or $E_{BA} \leq 30^{\circ}$. We can see that the success rate using local weights is much better than the success rate obtained using global weights (Table 7.2) for both beam number and beam angle retrieval. We conclude that the importance of attributes does vary with respect to the attribute values in the target case and generating rules learnt from the training data, the weights can accurately be assigned to the attributes of the target case.

Table 7. 4: Beam number error E_{BN} and beam angles error E_{BA} using local weights

	Average error	**Success rate (%)**
Beam number retrieval	0.27	77.3
Beam angles retrieval	25.04 degrees	72.7

7.3 Determination of Retrieval Accuracy Based on Contents of the Case Base

Developing decision support systems using real world data often suffers from the problem of obtaining sufficient data, especially in the development stage. Any CBR system is only as good as its case base since it largely depends on the availability of cases in the case base that are sufficiently similar to the target case. Also, in the training phase when making design decisions about system parameters such as attribute weights, training using only a small number of cases can lead to over fitting. Ideally, sufficient data should be available even at the design stage. However, in practice, this is often not possible since acquiring real world date can be a long process, in particular since it depends on many external factors that are not under the control of the developers. In the development of the prototype CBR system for radiotherapy treatment planning, we have exclusively used real brain cancer patient cases from the City Hospital but data acquisition is a gradual, slow and continuously on-going process. Using real world data in clinical systems is preferable to generating artificially cases as it allows the developers to make design decisions that are accurately based on practical considerations that are likely to be encountered when the system is used by the intended end users.

In this work we have attempted to overcome the problem of a small case base (and the number of training cases available) using several methods:

- The choice and design of the similarity measure is guided by results obtained by observing error trends rather than absolute values.
- Cross validation methods are used in the training phase that efficiently make use of all the information available in the existing cases. Also, we use both the average error in treatment plan parameters and the variance in error between folds to ensure generalizability of the obtained results.
- The Random Retrieval Probability (*RRP*) gives an indication of the quality of the feedback obtained from a successful retrieval during cross validation. It measures the ability of the similarity measure to intelligently select a suitable case, by rating the probability of the successful retrieval being due to the limited, biased contents of the cases rather than the quality of the similarity measure. This is important, in particular, when the number of cases available for retrieval after the filtering stage is small and

therefore, the ratio of the number of cases with the correct solution to the number of cases with a wrong solution becomes more important.

In the following sections, we introduce an alternative error calculation during cross validation that takes into account the availability of a case in the case base with a suitable treatment plan for the target case rather than the absolute retrieval error value.

7.3.1 Generating Local Weights Assignment Rules Using an Alternative Retrieval Error Calculation

Currently, the retrieval success during cross validation is based on the absolute error in the treatment plans. So far, the beam number error E_{BN} and the beam angles error E_{BA} have been calculated based on the difference in treatment plan parameters between the retrieved case and the original known treatment plan of the target case. This is a commonly used method of error calculation in CBR systems (Aha and Bankert, 1994) and classification systems. When the error of the solution parameters is binary as is in the case of beam number error (that is E_{BN} is either '0' or '1'), this method is acceptable. However, in the case of continuous error values in solution parameters such as the beam angles error E_{BA}, the error occupies a range of values. This means that the error between the treatment plans of cases in the case base and the target case can be ranked.

Consider target case C_T and cases C_1 and C_2 from the case base. Then,

IF

$E_{BA}(C_T, C_1) > E_{BA}(C_T, C_2)$

THEN

similarity $S(C_T, C_1) < S(C_T, C_2)$

Similarly, the error E_{BA} and similarity values $S(C_T, C_n)$, where $n = 1,2, \dots N$ and N is the number of cases in the case base, between target case C_T and all other cases in the case base can be ranked.

However, usually error calculations in cross validation methods only consider the absolute error between the solution of the retrieved case and the known solution of the target case. If the case base is sufficiently large and complete, in the sense that it covers all possible cases, then this is not a problem, since it is assumed that a case exists in the case base that is very similar to the target case and therefore, has an appropriate solution. The error between the solution of the retrieved case and the known solution of the target case therefore accurately represents the quality of the similarity measure, the retrieval mechanism or the attribute weights used in the similarity measure.

In the current wrapper method, the parameters of the CBR system, such as the attribute weights are trained only based on the absolute error between the treatment plans of the target case and the most similar case. This is sufficient if the case base is large enough and covers all possible target cases. However, in smaller case bases, with insufficient coverage, the possibility exists that the retrieved case in spite of having a large beam angle error has, in fact, the best treatment plan for the target case compared to the solutions of all other cases in the case base. In other words, the absolute error of the retrieved case in small case bases is not an accurate indication of the quality of the similarity measure but is biased by the contents of the case base. The similarity measure should be capable of retrieving the case with the most suitable treatment plan for the target case available for retrieval in the case base. Therefore during the rule learning/weights assignment training phase, we need to take into account not only the absolute error between treatment plan parameters but also if the retrieved case has the most similar treatment plan compared to all other cases available for retrieval.

To illustrate this issue, consider target case C_T from our case base. There are four cases $C_{C,1}$, $C_{C,2}$, $C_{C,3}$ and $C_{C,4}$ in the case base that consider the same OAR as target case C_T. In other words, $C_{C,1}$, $C_{C,2}$, $C_{C,3}$ and $C_{C,4}$ are the cases available for retrieval for target case C_T. Table 7.5 shows the beam angle error E_{BA} between the treatment plan of C_T and the treatment plans of $C_{C,1}$, $C_{C,2}$, $C_{C,3}$ and $C_{C,4}$. Using the wNN similarity measure with the local weights assigned as described in section 7.2, the most similar case to target case C_T is case $C_{C,1}$. The treatment plans of C_T and $C_{C,1}$ are shown in Figure 7.3. We can see in the figure that their beam angle error $E_{BA}(Plan_{T,x}, Plan_{C,1})$ is 35.7°. The condition, for the retrieval to be deemed successful is $E_{BA} \leq 30$. It can be seen that with the absolute error value of 35.7°, the retrieval is not deemed successful. However, from Table 7.5, it is clear that error E_{BA} between target case C_T and case $C_{C,1}$ is smaller than E_{BA} between target case C_T and any of the other three cases. There exists no case in the case base with the same OAR whose treatment plan is more similar to C_T with respect to the beam angles than case $C_{C,1}$. In this case, the large value of E_{BA} between the treatment plans of C_T and case $C_{C,1}$ is not necessarily due to inappropriate weights being used in the similarity measure but due to the contents of the case base. This example demonstrates that the absolute value of E_{BA} is not always a good indicator of the performance of the retrieval mechanism and on its own is not a reliable parameter to guide learning of the weights or rules during the training phase.

Treatment plan of case from case base C_T		
Beam	Gantry Angle (degrees)	Patient Couch Angle (degrees)
1	30	300
2	308	0
3	270	0
4	166	0
$E_{BA}(Plan_T,Plan_{C,1})$		35.7

Treatment plan of case from case base $C_{C,1}$		
Beam	Gantry Angle (degrees)	Patient Couch Angle (degrees)
295	295	90
2	272	0
3	80	320
4	315	0

Figure 7.3: Treatment plans of target case C_T and a case $C_{C,1}$ from the case base. The treatment plans show the gantry and patient couch angle (in degrees) of each beam. The error $E_{BA}(Plan_T,Plan_{C1})$ is also shown.

Table 7.5: Beam angle error E_{BA} between target case C_T and four cases from the case base

C_T	$C_{C,1}$	$C_{C,2}$	$C_{C,3}$	$C_{C,4}$
E_{BA}	35°	65°	41°	63°

The random retrieval probability, introduced in the previous sections considers the contents of the case base by taking into account the probability of a successful retrieval being due to a bias in the solution parameter values in the case base rather than an accurate similarity measure. However, it does not take into account the possibility of the treatment plan of the retrieved case being the most suitable to the target case in spite of a large error.

In the following section, we introduce an alternative training technique for generating the rules that assign local attribute weights during beam angles retrieval that takes into account the contents of the case base by not only using the absolute error in treatment plans and the random prediction probability but also examines if the retrieved case is the most similar case to the target case of all cases available for retrieval in the case base with respect to their treatment plans.

7.3.2 Extended Local Attribute Weights Rule Generation Method

This alternative method of rules generation for local attribute weights assignment is based on the steps outlined in section 7.2. The attributes of the target case are assigned to the clusters *Large* or *Small*. A rule R_q consist of antecedent A_q = [Eq, Vq, Rq, Dtq] and consequent $C_q = [w_{A,q}, w_{E,q}, w_{V,q}, w_{R,q}, w_{Dt,q}, w_{P,q}]$, where the weights, $w_{l,q}$, can take values from [0, 0.5, 1], and $l = A, E, V, R, Dt, P$. The rules are generated as described in section 7.2 and evaluated using the leave-one-out strategy. In the previous method, a rule was selected based on the number of training cases that resulted in successful retrievals (i.e. $E_{BA} \le 30\,^\circ$) and the average random retrieval probability (*RRP)* of the rule over all training cases.

In order to take into account the content of the case base, a condition is introduced that limits what is a successful retrieval during the training stage. Consider target case C_T and case C_C. Let $Plan_{C,T}$ and $Plan_{C,C}$ be the known treatment plans of target case C_T and case C_C, respectively. Let $Plan_{C,C}$ be the treatment plan in the case base that has been found to be the most similar plan to the treatment plan of target case C_T. Since the treatment plan parameters of the training cases are known, the case in the case base (with same OAR as C_T) with the most similar treatment plan $Plan_{CC,\ MostSim}$ to the treatment plan $Plan_{C,T}$ of target case C_T can be determined. A retrieval of case C_C is successful if the condition in expression 7.8 is satisfied. This ensures that either the treatment plan $Plan_{CC}$ of case C_C is the most similar treatment plan to the target case in the case base or that $E_{BA} \le 30$ degrees. The aim of this condition is that even if the retrieval error E_{BA} is large, the retrieval is still deemed successful if the retrieved case has the most similar treatment plan of the available cases in the case base to the target case. In addition, we specify the condition that a successful retrieval must not be random, i.e. *RRP* < 0.5, where *RRP* is given by expression 7.6.

$$\text{IF} \quad \begin{array}{c} Plan_{C,C} = Plan_{C,MostSim} \\ OR\ E_{BA} \leq 30 \\ AND\ RRP < 0.5 \end{array} \quad \text{THEN} \quad \begin{array}{c} \text{Retrieval} \\ \rightarrow \\ \text{Successful} \end{array} \qquad 7.8.$$

In local weights assignment as previously mentioned, the final 16 rules for local attribute weights are generated using the rule evaluation measures of the support and confidence, given in expressions 7.3 and 7.4, respectively and their product *CSP*.

7.3.3 Evaluation Using Test Cases

The alternative local attribute weights assignment method that considers the contents of the case base is evaluated as described in section 7.3. In addition, condition 7.8 is applied when deciding if a retrieval has been successful or not. The average beam angles error and the success rate over the 22 test cases is shown in Table 7. 6. We can see that there is an improvement in E_{BA} and the success rate of retrieval when considering both the absolute error and the contents of the case base when learning the local weights rules during the training stage.

Table 7. 6: Beam angles error when using local attribute weights assigned using rules that satisfy condition 7.8.

	Average error, E_{BA}	Success rate (%)
Beam angles retrieval	22.99°	81.8

7.4 Conclusion

This chapter introduced the weighted nearest neighbour similarity measure (wNN), on which the retrieval mechanism in this work is based. The wNN similarity measure is easy to implement and it is effective, however, the attribute weights have to be carefully chosen. The weights analysis showed that different weights lead to optimal results during beam number and beam angles retrieval. For both the beam number retrieval and the beam angles retrieval, the global weights were computed using the k-fold cross validation technique.

The disadvantage of using global weights is that they ignore the impact of the attribute values on their importance. For this reason, we designed a novel local weighting mechanism, where the weights are assigned using rules based on the attribute values of the target case. The rules are generated using a supervised learning approach in which feedback about the retrieval success of the wNN similarity measure on training cases is used to guide the weights determination. The rules are pre-screened using the rule evaluation measure of the product of the confidence and support, often used in data mining. A novel concept introduced in this chapter, called the random retrieval probability, takes into account how reliable the feedback obtained about retrieval success is. Another advantage of the local weighting algorithm is that since clustering and rule generation are done offline using the archived cases in the case base, they do not affect the retrieval time. Therefore, when presented with a target case, the retrieval mechanism runs quickly using the pre generated weight assignment rules. The clusters and rules, however, can be updated when a large number of cases has been added to the case base. The success rate obtained using local weights shows a marked improvement over the results obtained using global weights. A comparison of all methods used in this work will be given in section 10.1.

This chapter also presents a variation of the local attribute weights rule generation algorithm that is very effective with small training case bases that potentially do not have sufficient coverage, i.e. not all target cases have similar cases in the case base with a suitable treatment plan for the target case. In essence, rather than using only the absolute error of continuous solution parameters between the target case and cases from the case base, the question of whether the case with the most similar treatment plan to the treatment plan of the target case has been retrieved, guides the learning of rules for local attribute weights. As shown, experimental results prove that this method is able to obtain better results even with a small case base. It has to be noted that this method is only applied during determination of the parameters of the retrieval mechanism and not to compute the success rate of the retrieval mechanism. The success rate gives an indication of how the CBR would perform when in clinical (or commercial) use. In that situation, retrieving the case that is more similar to the target case compared to the treatment plans of the other cases in the case base available for retrieval is not an option if the absolute retrieval error is high. If the similarity between the target case and none of the cases in the case base is sufficiently high (which would mean that none of the treatment plans are suitable to be used in the solution), a more reasonable alternative would be to not retrieve a case for that target case and possibly display a suitable error message. For this reason, when validating the performance of the retrieval mechanism, in this work, only the absolute error between the treatment plans of target and retrieved case is used. However, an interesting question is what similarity value can be considered as "sufficiently high" and what would constitute a suitable threshold below which a case is not retrieved. The investigation of this question is planned in future research work.

An interesting avenue of future research, we are planning to explore is the possibility of the local weights based not only the target case attribute values but also on the attribute values of the cases in the case base. One method to do this is to weight attribute values depending on the similarity between an attribute in the target case and the corresponding attribute in the case from the case base.

Currently, the attribute values are assigned to the crisp classes, *Large* or *Small*, obtained using the *k*-means clustering algorithm. However, since the attribute values are continuous the boundaries of each class are artificially generated. In the future, we plan to use fuzzy sets to obtain a more accurate representation of how large or small an attribute value is and how its value affects the significance of attribute weights.

Further, instead of clustering attributes another possibility would be to cluster sets of patients. Each patient case would be assigned to a cluster containing similar patients. The attributes weights would then be different for different clusters of similar patients. Currently, this approach is impractical due to the small case base but it would be interesting to compare the current approach of clustering attributes with clustering patient cases.

Chapter 8

The Fuzzy, Non-linear Similarity Measure

The weighted nearest neighbour (wNN) similarity measure computes the aggregate similarity between two cases as the sum of the individual attribute values. However, in order to compute the sum, it assumes that the similarity values computed between attributes are linear and comparable, that is a similarity value S_A with respect to attribute A denotes the same extent of similarity as similarity value S_B with respect to attribute B, if $S_A = S_B$. This, however, is not always necessarily accurate. To give a very simple example, in a randomly drawn sample of humans, the probability that the majority of this sample has a similar number of fingers on their hands is rather high. However, the probability that the majority of the sample has a similar hair colour is much lower. The numerical similarity values of attribute *number of fingers* and *hair* colour between humans are not necessarily comparable since *number of fingers* has a larger probability of being similar.

In the developed CBR system, the values of the attributes are first normalised, but this only accounts for the scale or range of attribute values and not for the variability of the similarity in between the extreme values of small and large attribute similarity between two cases. In order to obtain an idea of the distribution of attribute similarities, we calculated the similarity between each case and every other case (with the same OAR) in our case base considering one attribute a time. Figure 8.1 shows the similarity (as a function of the difference in attribute values) calculated using a leave-one-out strategy between each case in the case base consecutively used as a target case and the other cases in the case base. The similarity values are arranged in ascending order. It can be seen that in spite of the similarity values for each attribute being similar at the extreme points of small and large similarity, the similarity curves differ for each attribute in between. Figure 8.2 shows the frequency distribution of the similarities calculated similarly as above for each attribute. Again, it can be seen that though the frequency of larger similarity values increases in general for all attributes, the actual similarity values differ substantially between attributes. It can be seen from the graph that the similarity values of all attributes, except *E* and *V*, concentrate towards the higher end of the similarity values spectrum, in particular for attributes *A*, *R* and *Dt*. An example of the uneven distribution of attribute similarity values, consider the 15th percentile of the similarity values of attribute *R*, which is 0.89. This means that only 15% of similarity values between cases have a similarity of less than 0.89. In contrast, about 93% of the case similarities calculated with respect to attribute *E* have a value below 0.89. The graphs confirm that the numerical attribute similarity values are not necessarily comparable.

Based on the contents of the case base, the question therefore arises if the numerical similarity values of normalised attributes are actually comparable and can they be summed to provide the total similarity between two cases. In order to sensibly compare attribute similarities and generate an aggregate similarity measure, the CBR system has to understand how to interpret the numerical value of the attribute similarity with respect to the numerical values of the other attributes. In this work, we investigate how to facilitate the interpretation of numerical similarity values with the help of the similarity values between the target case and the other cases in the case base. In other words, the similarity between two cases is interpreted as *Large*, *Medium* or *Small* depending on how it compares numerically to the similarity between the target case and all cases in the case base with respect to an attribute. We propose the use of fuzzy sets to describe for each attribute what similarity value constitutes a large, medium or small similarity based on the similarity values found in the case base.

Section 8.1 gives an overview of the steps involved in creating the fuzzy system. Section 8.2 describes the fuzzy membership functions that assign fuzzy membership grades to attribute similarity values, followed by section 8.3, which introduces the fuzzy, non-linear similarity measure. A variation of this method using local membership functions for each target case is presented in section 8.4. The fuzzy similarity measure with globally and locally defined fuzzy membership functions is validated in section 8.5 before the chapter concludes with section 8.6.

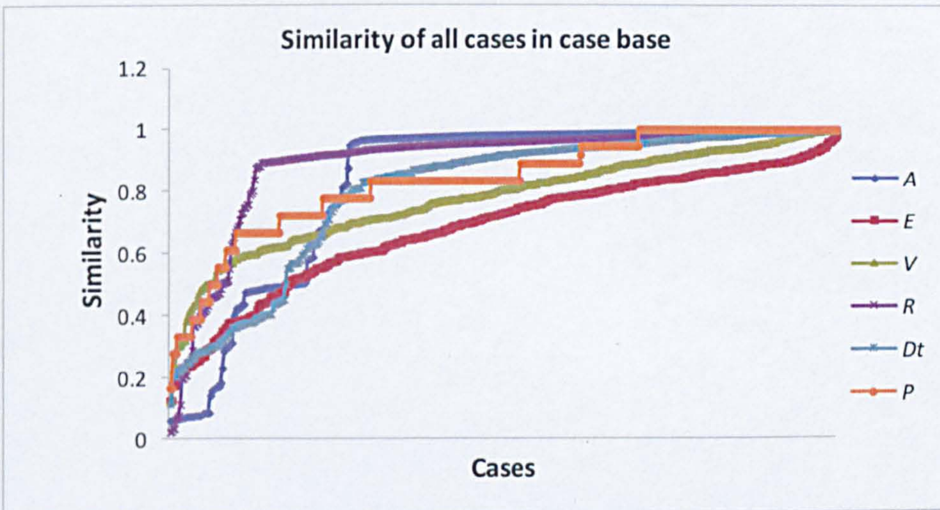


Figure 8.1: Similarity between each case and every other case in the case for attributes *A, E, V, R, Dt* and *P*

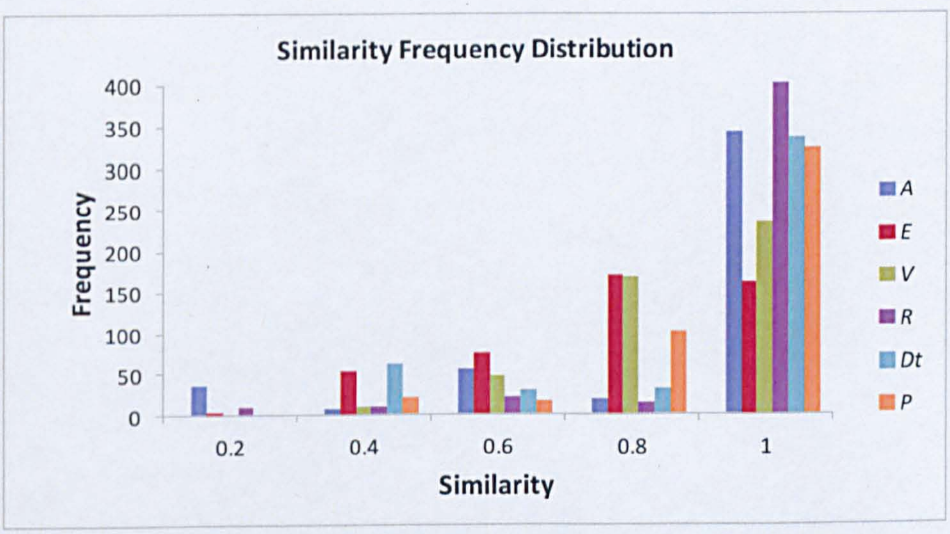


Figure 8.2: Frequency distribution of similarity values between each case and every other case in the case for attributes *A, E, V, R, Dt* and *P*.

8.1 The Fuzzy Inference System

The fundamentals of fuzzy set theory were introduced in section 3.5.1. In general, a fuzzy inference system (FIS), a term borrowed from fuzzy controllers, consists of three steps:

1) Fuzzification: The crisp input values are fuzzified with the help of fuzzy membership functions.
2) Aggregation: The fuzzified output values are combined into a single aggregate value.
3) Defuzzification: In control systems, the aggregate fuzzy output value is defuzzified using membership functions to obtain a crisp value. However, since we are only interested in comparing the relative similarity values between the target case and cases in the case base, given by the aggregate fuzzy value, defuzzification is not necessary.

Case retrieval using the fuzzy similarity measure proceed as follows: Consider two cases, target case C_T and case C_c from the case base. The attribute similarity between the cases with respect to attributes *A, E, V, R, Dt* and *P* is given by similarity S_l, where l = *A, E, V, R, Dt,* and *P*. Each attribute similarity S_l between C_T and C_c is assigned a membership grade to fuzzy sets *Large, Small* and *Med* using membership functions as described in section 8.2. In aggregation, the fuzzy membership grades of sets *Large* and *Med* and *Small* of all attributes are summed to give a *Large, Med* and *Small* fuzzy component, which are then combined to arrive at the total fuzzy similarity value *S* as described in section 8.3. The case (or cases) with the largest fuzzy similarity value *S* are retrieved to be used in the solution of the target case.

8.2 Fuzzy Membership Functions

The membership function of a fuzzy set assigns to each object of the set a grade of membership (Zadeh, 1965). The membership grades usually range from '0' to '1'. Fuzzy membership functions can be learnt from existing data (Hong and Lee, 1996, Nauck and Kruse, 1993, Aha and Bankert, 1994) or defined a priori, usually with the help of domain experts. Often the membership functions take on particular geometric forms such as triangular or trapezoidal. The main advantage of using triangular membership functions lies in its simplicity (Pedrycz, 1994) and therefore it provides a good starting point, in particular in the absence of detailed domain information.

In this work, the fuzzy sets take the form of triangular membership functions as well. A triangular membership function is defined by the support and the model point of the triangle. In order to approximately model the distribution of the similarity values among the cases in the case base of an attribute, the left and right support and model point of the triangular membership function are given by the minimum, maximum and average values, respectively, of the similarities found in the case base. Expression 8.1 represents the rules used to assign membership grades μ_{Small}, μ_{Med}, μ_{Large} of attribute similarity S to fuzzy sets *Small, Med* and *Large,* respectively, where S_{min}, S_{avg} and S_{max} are the minimum, average and maximum values of the similarities found in the training case base between each case and every other case with respect to an attribute. The shapes of the fuzzy membership functions are shown in Figure 8.3.

$$\mu_{Small} = \begin{cases} 1 & for\, S_l < S_{min} \\ 0 & for\, S_l > S_{avg} \\ \dfrac{S_{avg} - S_l}{S_{avg} - S_{min}} & for\, S_{min} \le S_l \le S_{avg} \end{cases}$$

$$\mu_{Med} = \begin{cases} \dfrac{S_l - S_{min}}{S_{avg} - S_{min}} & for\, S_{min} < S_l < S_{avg} \\ \dfrac{S_{max} - S_l}{S_{max} - S_{avg}} & for\, S_{avg} < S_l < S_{max} \\ 0 & for\, S_{min} \le S_l \ge S_{max} \end{cases} \qquad 8.1.$$

$$\mu_{Large} = \begin{cases} 1 & for\, S_l > S_{max} \\ 0 & for\, S_l < S_{avg} \\ \dfrac{S_l - S_{avg}}{S_{max} - S_{avg}} & for\, S_{avg} \le S_l \le S_{max} \end{cases}$$

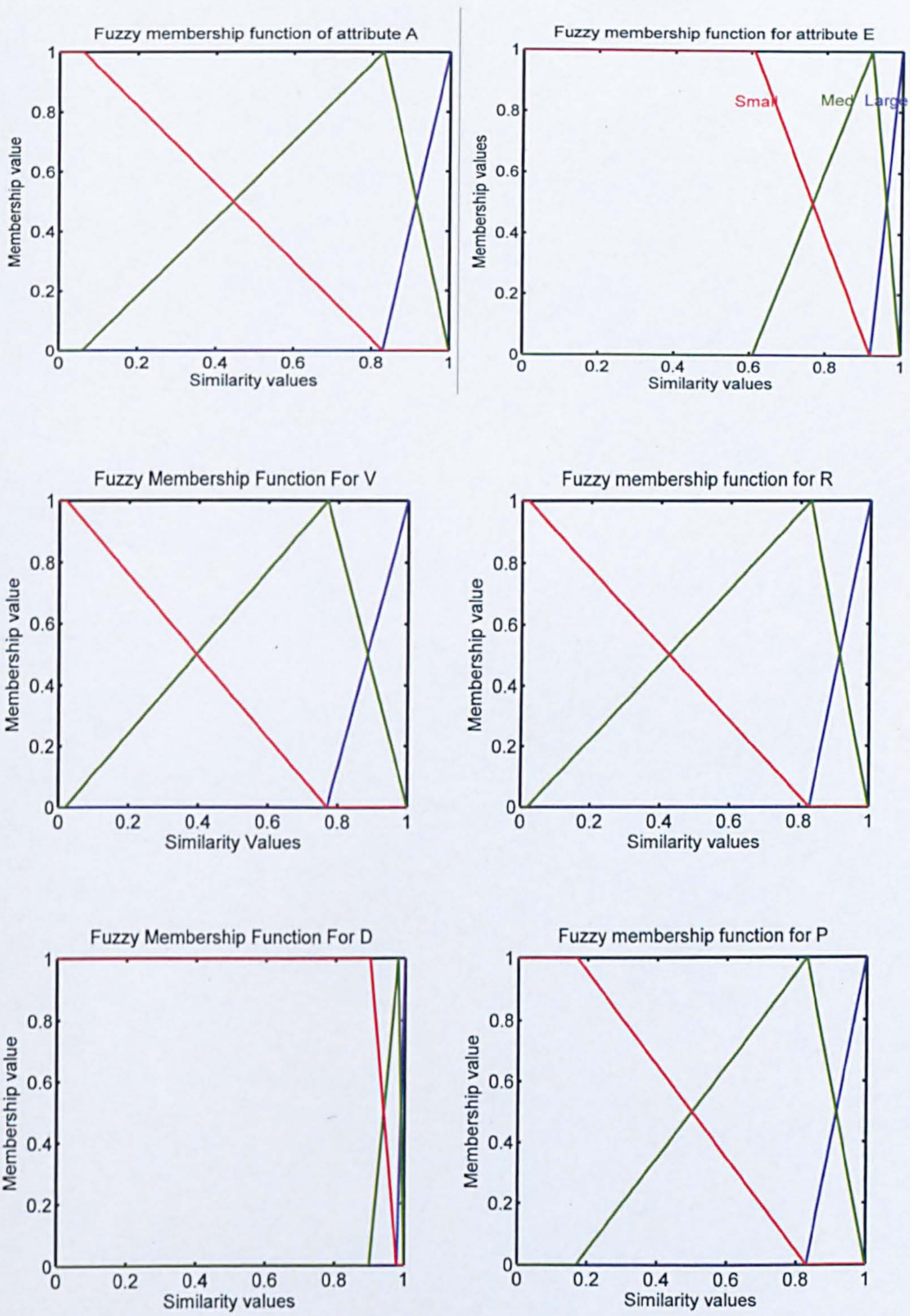


Figure 8. 3: Fuzzy membership functions of attributes *A, E, V, R, DT* and *P* for fuzzy sets *Large* (blue), *Med* (green) and *Small* (red).

8.3 The Fuzzy, Non-Linear Similarity Measure

This section describes the aggregation of the fuzzy membership grades of the six attributes to fuzzy sets into the aggregate similarity value used in the retrieval process. Aggregation is carried out using the fuzzy, non-linear similarity measure, which is loosely based on the wNN similarity measure. However, instead of calculating the weighted sum of the attribute differences, we define the fuzzy sets *Large, Med* and *Small,* which denote large similarity, medium similarity and small similarity, respectively, for each attribute. The fuzzy membership functions of the three sets are defined for each attribute *A, E, V, Dt, R* and *P* based on the minimum, maximum and average of the corresponding similarity values found across the case base. They, therefore, give a realistic indication of what constitutes a relatively large similarity, medium similarity or small similarity for an attribute. Given a target case C_T and a case from the case base C_C, the membership degree of the attribute similarity between these two cases to fuzzy sets *Large, Med* and *Small* is computed for each attribute l, $l = A, E, V, R, Dt, P$. The aggregate similarity consists of the *Large, Med* and *Small* component M_s, defined as the sum of the membership degrees of the attribute similarities to their corresponding fuzzy sets *Large, Med* and *Small,* shown in expression 8.2.

$$M_s = \sum_{l=A,E,V,R,Dt,P} w_l \mu_{l,s} \qquad 8.2.$$

$$S_F = w_{Large} \sum_{l=A,E,V,R,Dt,P} w_l(\mu_{l,Large}) + w_{Med} \sum_{l=A,E,V,R,Dt,P} w_l(\mu_{l,Med}) - w_{Small} \sum_{l=A,E,V,R,Dt,P} w_l(\mu_{l,Small}) \qquad 8.3.$$

where s = *Large, Med, Small,* w_l denotes the weight of *attribute* l, $l = A, E, V, R, Dt$ and P, and $\mu_{l,s}$ is the membership degree of the attribute similarity to the fuzzy sets *Large, Med* and *Small.* The terms w_{Large}, w_{Med} and w_{Small} denote the weights of the large, medium and small fuzzy components, which will be explained in section 8.3. A large value of component M_{Large}, M_{Med} and M_{Small} indicates a large, medium and small aggregate similarity between two cases, respectively. That is, M_{Large} has a positive effect on the aggregate similarity between two cases, while M_{Small} has a negative or penalizing effect. M_{Med} either adds to or penalises the aggregate similarity based on the sign of w_{Med}. The aggregate similarity S_T between two cases is defined as the net contribution of M_{Large}, M_{Med} and M_{Small} as shown in expression 8.3.

An added advantage of using expression 8.3 to compute the similarity measure between two cases is that the similarity and dissimilarity can be expressed and therefore weighted separately. The wNN sums the weighted attribute similarity values; hence the aggregate similarity is always a function of the attribute similarity values. In other words, large attribute similarity values act to increase the aggregate similarity by a large amount while small attribute similarity values also increase the aggregate similarity, but by a smaller amount. That is, no matter how similar or dissimilar two cases are to each other with respect to an attribute, the attribute similarity always contributes positively to the aggregate similarity. An alternative method of case retrieval would be to consider the dissimilarity between cases or to measure the extent by which two cases are different from each other. In this work we have designed a similarity measure that combines similarity and dissimilarity between cases. Large similarity values contribute positively to the aggregate similarity. Small similarity values (which indicate dissimilarity) are subtracted from the aggregate similarity. In other words, small similarity values act to penalize the aggregate similarity.

8.3.1 Determination of Weights of Similarity and Dissimilarity

The weights w_{Large}, w_{Med} and w_{Small} in expression 8.3 determine the importance of the fuzzy sets *Large, Med* and *Small* respectively. By varying the values of w_{Large}, w_{Med} and w_{Small} we can control the contribution that the large, medium and small similarity component have on the aggregate similarity measure, namely w_{Small} emphasizes the dissimilarity between cases, while w_{Large} emphasizes the similarity between cases. By increasing the values of w_{Large} and w_{Small}, the non-linearity of the similarity measure is increased. Therefore, the solution of a case with large attribute similarities is very suitable for the target case and conversely, the solution of a case with very large attribute dissimilarity is very unsuitable. By changing the values of w_{Large}, w_{Med} and w_{Small}, we can control the extent of non-linearity and therefore indicate the contribution of large and small similarity between attribute values to the aggregate similarity.

In order to determine the values of w_{Large}, w_{Med} and w_{Small}, we ran the retrieval mechanism using a leave-out technique to find the fuzzy component weights that would yield the smallest beam number error, E_{BN} and beam angles error, E_{BA} for the training cases. The weights w_{Large} and w_{Small} can take values from the set [0, 0.5, 1] and w_{Med} can take values from the set [-1, -0.5, 0, 0.5, 1]. The local attribute weights are provided using the rules determined in section 7.2 for beam number retrieval and the rules learnt based on both the absolute retrieval error and the contents of the case base in section 7.3 for beam angles retrieval. As described previously with the wNN similarity measure, the three most similar cases are retrieved for beam number retrieval (k=3). The mode of the beam numbers in the three retrieved cases is chosen for the solution of the target case. With respect to the beam angles retrieval, the most similar case is retrieved (k=1).

Table 8. 1 shows the error results obtained when running the retrieval mechanism using the fuzzy similarity measure on the training cases with different values for w_{Large}, w_{Med} and w_{Small}. In the case of the beam number error E_{BN}, it can be seen that when only the large similarity component or only the small similarity component are used, the error is quite large, indicating that both components provide important information about the similarity between two cases. The lowest value of E_{BN} was obtained when w_{Large}=1, w_{Med} =0 and w_{Small}=1. In the case of the beam angles error E_{BA}, using exclusively the high similarity component gives better results than using unity weights or exclusively the dissimilarity component. However, again the best result is obtained when using both the similarity and the dissimilarity component of the fuzzy similarity measure.

Table 8. 1: Beam Number Error E_{BN} and Beam Angles Error E_{BA} obtained using the fuzzy similarity measure on the training cases

W_{Large}	W_{Med}	W_{Small}	**Error**
Beam Number Error E_{BN}, k=3			
1	1	1	0.38
1	0	0	0.44
0	0	1	0.48
1	0	1	**0.36**
Beam Angles Error E_{BA}, k=1			
1	1	1	25.12
1	0	0	21.44
0	0	1	32.25
1	0	1	**20.23**

8.4 Local Membership Functions

In the previous sections, the membership functions that were used to assign fuzzy membership grades to the crisp attribute similarity values were global, that is the same membership functions were used for all target cases. In this section, we examine the use of defining the membership functions online (i.e. while the retrieval takes place) with respect to the attribute similarities of a target case.

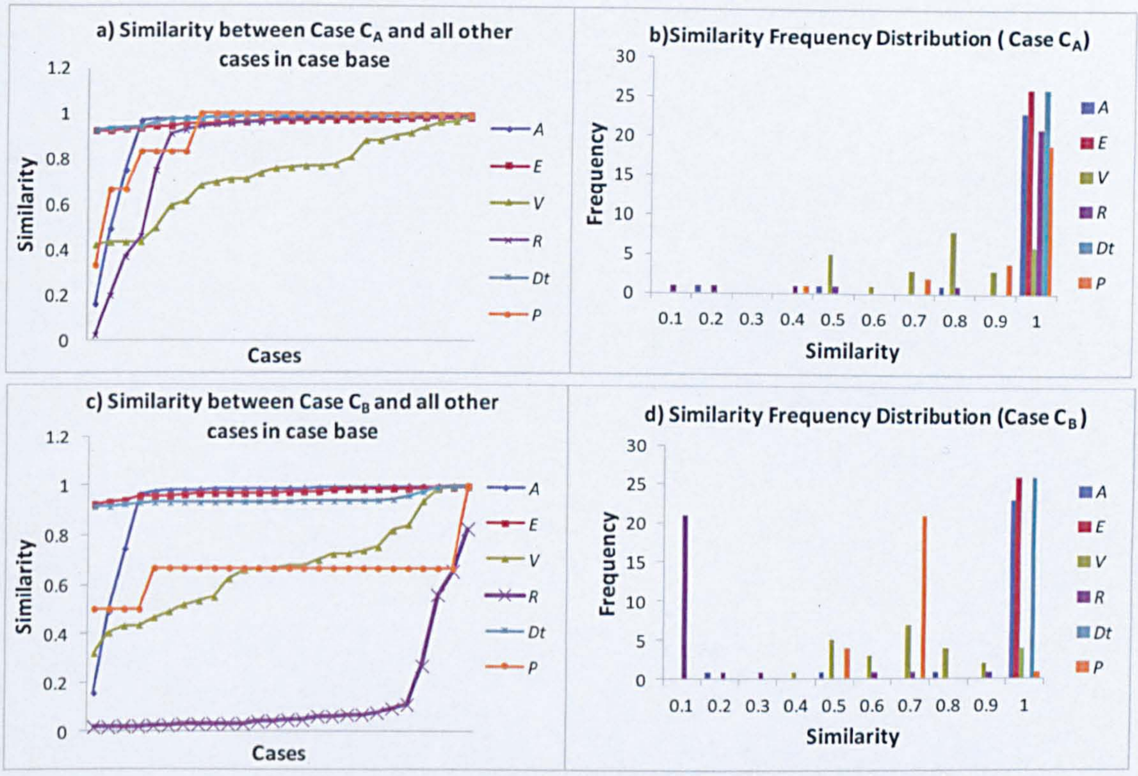


Figure 8.4: Graph a) and c) show the similarity between two cases C_A and C_B and the other cases in the case base. Graph b) and d) show their frequency distribution

To illustrate the rationale behind introduction of a new approach to defining membership functions, let us consider two random cases from the case base, C_A and C_B. Figure 8.4 a) and c) show the attribute similarities between C_A and C_B, and all other cases in the case base. For both cases, it can be seen that the similarity values concentrate between 0.9 and 1 with respect to attributes *A, E* and *Dt*. However, the attribute similarities with respect to attributes *R* and *P* are very different for cases C_A and C_B. Similarly, graphs c) and d), which present the frequency distribution of the attribute similarity values for cases C_A and C_B show considerable differences. Comparing Figure 8.4 with Figure 8.1 and Figure 8.2, which show the distribution of attribute similarities over all cases, it can be seen that the similarity distribution for a single case is very different. It is likely that the fuzzy membership functions for fuzzy set *Large, Med* and *Small* are not appropriate for every target case, since the maximum, minimum and in particular, the average similarity values of the target case and the cases in the case base show wide variation.

For this reason, we define the membership functions individually for each target case. Given a target case, the CBR system computes the shape of the membership function using expression 8.1. However, the values of S_{min}, S_{max} and S_{avg} are computed online for each target case based on the similarity values between the target case and the cases in the case base available for retrieval. The membership grades and fuzzy components of all attributes are then combined as described in section 8.2. The case with the highest aggregate fuzzy similarity value is then retrieved to be used in the solution of the target case.

8.5 Evaluation of Fuzzy Similarity Measure Using Test Cases

In order to test the performance of the fuzzy similarity measure both with global and with local fuzzy membership functions, the retrieval mechanism was evaluated using the 22 test cases. Each test case was consecutively made the target case. The attribute similarities between the test target cases and the cases in the case base were fuzzified first using the globally defined membership functions as described in section 8.2. The fuzzified attribute similarity values were combined into an aggregate similarity value using expression 4.4 with fuzzy weights $w_{Large} = 1$, $w_{Med} = 0$, $w_{Small} = 1$ as determined during the training phase. During beam number retrieval, the three cases with the highest fuzzy similarity were retrieved. E_{BN}, as done previously, was given by the difference in the computed mode of the treatment plans of the three retrieved cases and the beam number in the known treatment plan of the target case. During beam angles retrieval the case with the highest fuzzy similarity was retrieved as done previously. The beam angles error E_{BA} is given by the difference in beam angles between the treatment plan of the retrieved case and the beam angles of the known treatment plan of the test target case. The values of E_{BN} and E_{BA} are averaged over all 22 test cases. The averaged error values are shown in Table 8.3. The procedure was then repeated using the locally defined fuzzy membership functions for each test target case. The averaged values of beam number error E_{BN} and the beam angles error E_{BA} using locally defined fuzzy membership functions are shown in Table 8. 2.

It can be seen that the beam angles error only shows a marginal improvement as compared to using the wNN similarity measure with local weights, which indicates that the beam angles retrieval is more stable with respect to the distribution of the attribute similarity. This is further confirmed by the fact that using locally defined fuzzy membership functions that model the distribution more accurately does not improve the success rate any further either. The beam number error increases when using the fuzzy similarity measure as opposed to the wNN similarity measure with local weights. However, there is a substantial improvement when using local fuzzy membership functions. We conclude that the distribution of the attribute similarities is significant, however it has to be taken into account accurately for each target case by using locally defined membership functions.

Table 8. 2: Beam number error E_{BN} and beam angles error E_{BA} obtained using the fuzzy similarity measure on the test cases when using fuzzy global weights

	Error	**Success Rate (%)**
Beam Number Error, E_{BN}	0.36	68
Beam Angles Error, E_{BA}	22.18°	82

Table 8.3: Beam number error E_{BN} and beam angles error E_{BA} obtained using the fuzzy similarity measure with local fuzzy membership functions.

	Error	**Success Rate (%)**
Beam Number Error, E_{BN}	0.23	82
Beam Angles Error, E_{BA}	22.76°	82

8.6 Conclusion

This chapter introduced the non-linear fuzzy similarity measure. The advantages of this similarity measure over the wNN similarity measure are twofold:

1) It takes into account that the distribution of attribute similarities varies for each attribute and therefore the numerical attribute similarity values cannot be merely summed to give an accurate representation of the true similarity between two cases.
2) By grouping the similarity into large, medium and small similarity components, both similarity and dissimilarity can be expressed and separately weighted. More experiments on a larger case base or benchmark data are required to gauge the benefit in this work. However, the fuzzy similarity measure looks promising to be applied to other CBR systems where the attribute similarity values have been found to be not numerically comparable.

Using locally defined fuzzy membership functions for each target case, substantially improves the success rate for beam number retrieval, indicating that the distribution of attribute similarity values indeed is an important factor in beam number retrieval. However, beam angles retrieval appears to be more stable to the distribution of attribute similarities. In future work, we are planning to investigate further the reason behind why the beam angles retrieval is not considerably improved using the fuzzy similarity measure and if there is a link between the attribute values and the beam angles retrieval performance when using the fuzzy similarity measure.

In our experiments, we did not notice an increase in computation time when using locally defined fuzzy membership functions that are generated online for every target case. However, with an increase in the size of the case base, the computation time might become more significant. A possible solution would be to generate a library of fuzzy membership functions for the existing cases in the case base using a leave-one-out approach. With a large case base, we could also define membership functions for clusters of similar cases. A new target case would then be assigned to a cluster of similar cases and the membership functions defined for that cluster applied.

Chapter 9

The Two-Phase Retrieval System

As seen in the previous sections, the weight vectors that result in the smallest beam number error E_{BN} are different to the weight vectors that result in the smallest beam angles error E_{BA} during the learning phase using training cases. In other words, the importance of attributes varies depending on which treatment plan parameter of the solution we are aiming to retrieve. When retrieving a treatment plan that suggests an appropriate beam number to use for the target case, the optimum weights are different to the optimum weights to be used when retrieving a treatment plan that suggests the beam angles.

Retrieving a single treatment plan would result in the practical problem of combining the beam number error E_{BN} and the beam angles error E_{BA}. A simple method of combining E_{BA} and E_{BN} is to normalise them to values between [0,1] and then take the average of the normalised values of E_{BA} and E_{BN}. When learning the weights of the wNN similarity measure during the training phase using k-fold cross validation, the attribute weights vector that results in the lowest error value, E_{BT}, with respect to the combined beam number and beam angles error is shown in Table 9.1.

Table 9.1: Attribute weights in wNN similarity measure that result in lowest average error E_{BT} during weights training phase.

Weights						E_{BN}	E_{BA}	E_{BT}
w_A	w_E	w_V	w_R	w_{Dt}	w_P			
0	1	1	0	1	1	**0.38**	**28.05**	**0.47**

Table 9.2: Attribute weights in wNN similarity measure showing weights vector for single beam number and beam angles retrieval.

Plan parameter	Weights						*Error*
	w_A	w_E	w_V	w_R	w_{Dt}	w_P	
Beam number	0.5	1	0.5	1	1	1	**E_{BE} =0.36**
Beam angles	1	1	1	1	0.5	0.5	**E_{BA}=23.46**

However, when using the weights vectors, which in section 7.1 were shown to give the best results for beam number (W_{BN} = [0.5, 1, 0.5, 1, 1, 1]) and beam angles retrieval (W_{BA} = [1, 1, 1, 1, 0.5, 0.5]) in the wNN similarity measure with respect to the training cases we get beam number error, E_{BN} = 0.36 and beam angles error, E_{BA} = 23.46° as shown in Table 9.2. E_{BN} is slightly smaller when using a weights vector optimised for beam number retrieval but E_{BA} is considerably smaller when using a weights vector optimised for beam angles retrieval. So, using the weights that result in the smallest combined error of beam angles and beam number is sub-optimal. In other words, it is more advantageous to retrieve two cases along with their treatment plans using optimized weights for each treatment plan parameter.

So far, in the work presented in this thesis, the retrieval mechanism was called twice for every target case, once with attribute weights (global or local) optimised to retrieve a treatment plan to suggest the beam number and then once again with attribute weights (global or local) optimised to retrieve a treatment plan to suggest the beam angles. In essence, two treatment plans were retrieved in parallel for every target case. However, if two treatment plans are retrieved in parallel, the practical question arises of how the two treatment plans are used further, in particular if the parameters of the treatment plans are incompatible with each other (for instance, if the number of beams is different in the treatment plans). For this reason, a simple methodology, called the two phase retrieval method, of retrieving two treatment plans in sequence has been designed and is described in this chapter.

When using a multi-phase retrieval mechanism, the order in which the treatment plans are retrieved is rather important as explained in section 9.1. The two phase retrieval procedure is outlined in section 9.2. Section 9.3 presents the error and success rate obtained on the test cases using the two-phase retrieval mechanism and section 9.4 concludes this chapter.

9.1 Order of Retrieval

In radiotherapy treatment planning, merely retrieving two treatment plans in parallel leads to practical implementation problems since the beam number of both treatment plans from the two retrieved cases have to be the same. For example, consider treatment plans P_{BN} and P_{BA}, which are retrieved using the optimal weights W_{BN} (for beam number retrieval) and W_{BA} (for beam angles retrieval), respectively. Let us suppose that in treatment plan P_{BN}, the number of beams used is 4. If plan P_{BA} has 5 angles, then one of the angles has to be discarded since we only want to use 4 beams. However, choosing which angle to discard is not straightforward. Similarly, if plan P_{BA} has 3 angles, then the angle for the required fourth beam would be missing. To avoid conflicts such as these the treatment plans in the two phase retrieval method are retrieved in sequence. However, the order of retrieval is vital. A few of the concerns that have to be taken into account when deciding on the retrieval order are:

- Ease of implementation: If we determine the beam angles in the first phase and then the number of beams the system would have no way of determining which beam angles to choose if the number of beams in P_{BA} is larger or smaller than the number of beams in P_{BN} retrieved in the second stage.
- Adaptation: In CBR, adaptation is a very important module. After the most similar case has been retrieved, the solution of the retrieved case is normally adapted to fit the particular problem details of the target case. It is recommended to design the retrieval process in such a way as to facilitate the adaptation process (Smyth and Keane, 1998). Solution parameters that are more difficult to adapt are given higher preference in the

retrieval process. During consultation with staff at the City Hospital, it transpired that given the number of beams, the beam angles can easily be tweaked to achieve the desired dose distribution. However, adding another beam or removing a beam from the suggested beam configuration is considerably more complicated since the angles would have to be readjusted for the new beam number.

Figure 9.1 shows the two-phase retrieval mechanism.

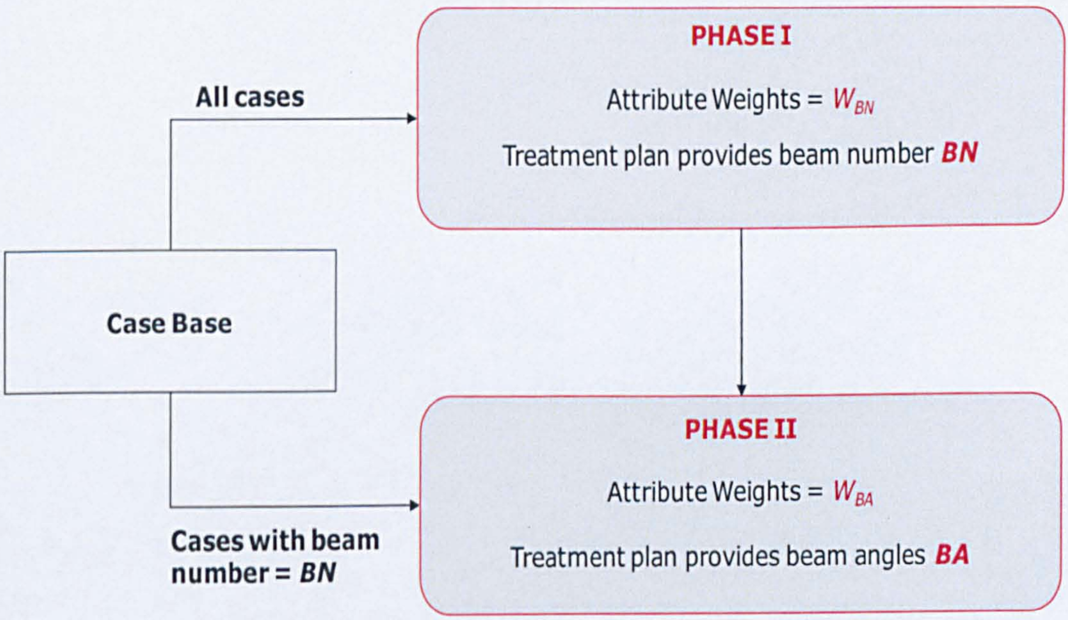


Figure 9.1: Schematic showing two-phase retrieval mechanism

9.2 Two-Phase Retrieval Procedure

This section outlines the procedure of the two phase retrieval mechanism.

Phase I

The aim of phase I in the CBR system for radiotherapy treatment planning is to retrieve a case with a treatment plan that will suggest the number of beams for the target case. During retrieval any similarity measure can be used. In the experiments to follow, the fuzzy similarity measure with locally defined membership functions for fuzzy sets *Large, Small* and *Med* was used to calculate the similarity between two cases.

Given a target case, the similarity between the target case and all cases with the same organs at risk in the case base is computed with respect to each attribute. The fuzzy membership grades of the attribute similarities are calculated using expression 8.1 and the aggregate similarity between the target cases and the cases in the case base is determined using expression 8.3 with weights $w_{Large} = 1$, $w_{Med} = 0$ and $w_{Small}=1$ (as determined in section 8.3.1). The local attribute weights $W_{BN}= w_A, w_E, w_V, w_R, w_{Dt}$ and w_P are assigned to each target case using the generated rules for beam number retrieval as described in section 7.2. The case with the largest fuzzy similarity is retrieved and the number of beams *BN* used in the retrieved treatment plan is stored.

Phase II

The aim of phase II is to retrieve a case with a treatment plan that suggests the beam angles to be used in the solution of the target case. The case base used in phase I is filtered and all cases with beam number *BN* are extracted. These cases form the case base for phase II. The fuzzy similarity is calculated as described above with weights $w_{Large} = 1$, $w_{Med} = 0$ and $w_{Small}=1$ (as determined in section 8.3.1). Local attribute weights $W_{BA}= w_A, w_E, w_V, w_R, w_{Dt}$ and w_P are assigned using the generated rules for beam angle retrieval. The most similar case is retrieved from the filtered case base and its treatment plan is used in the solution of the target case to suggest the number of beams and their angles.

9.3 Evaluation of Two-Phase Retrieval Using Test Cases

In order to evaluate the performance of the two phase retrieval mechanism, we ran the retrieval mechanism using the test cases.

Each test case was consecutively made the target case. The attribute similarities between the test target cases and the cases in the case base were fuzzified using local fuzzy membership functions. The aggregate similarity was then computed using expression 8.3 and weights $w_{Large} = 1$, $w_{Med} =0$, $w_{Small} =1$, which were identified to result in the lowest retrieval error values as described in section 8.3.1. In phase I, the three cases with the highest fuzzy similarity were retrieved and the mode of the beam numbers in the three treatment plans was stored. The procedure is repeated in phase II using the reduced case base filtered with respect to *BN*, the number of beams suggested in phase I. The most similar case was retrieved and its treatment plan used in the solution of the target cases. The averaged error values of E_{BN} and E_{BA} are shown in Table 9.3.

Table 9.3: Beam number error E_{BN} and beam angles error E_{BA} obtained during two phase retrieval with test cases

	Error	Success Rate (%)
Beam Number Error, E_{BN}	0.23	81
Beam Angles Error, E_{BA}	28.83°	68

From the results, it can be seen that the beam angles error E_{BA} has considerably increased. This is very likely to be due to the reduced number of cases that are available for retrieval in phase II for each target case. Table 9. 4 shows the number of cases available in phase I and phase II of the two phase retrieval mechanism for the 22 test cases. In phase I for each target case the cases in the case base that consider the same organs at risk as the target case are available for retrieval. However, among the cases available for retrieval in phase I, in phase II only the cases whose treatment plans contain the number of beams suggested in phase I are now available for retrieval. It can be seen in Table 9. 4 that the number of cases available for retrieval in phase II is considerably reduced compared to phase I, which explains why the beam angles retrieval error E_{BA} was considerably lower in the previous experiments, when all cases with the same OAR were available for retrieval during beam angles retrieval.

Table 9. 4: Number of cases available in phase I and phase I for retrieval for each test case

Test case index	Number of cases available for retrieval in phase I	Number of cases available for retrieval in phase II
1	4	4
2	27	17
3	27	17
4	27	17
5	27	17
6	26	19
7	27	17
8	26	19
9	27	17
10	26	19
11	27	17
12	26	19
13	26	19
14	26	19
15	26	19
16	26	19
17	26	19
18	4	2
19	2	1
20	27	17
21	2	1
22	2	1

9.4 Conclusions

This chapter introduced the two phase retrieval mechanism. Using a multi-phase sequential retrieval mechanism is appropriate when the importance of attributes varies with respect to the parameters of the solution. Retrieving cases sequentially helps to avoid contradictions between solutions. However, care has to be taken when deciding on the sequence of retrieval.

In our work, two-phase retrieval is expected to work well with a larger case base but due to the small case base, the results of beam angles retrieval in phase II are inadequate. However, it has to be noted that this means that the moderate performance of the retrieval mechanism in phase II is due to the contents of the case base rather than the quality of the retrieval mechanism, the similarity measure or its parameters. The two phase retrieval mechanism is implemented mainly for practical reasons to avoid conflicts in beam number when two treatment plans are retrieved to suggest beam number and beam angles. Retrieving just a single treatment plan with attribute weights optimised to reduce the combined beam number and beam angles error, shows larger retrieval error values. It is expected that the error E_{BA} will reduce and the success rate will considerably increase as more cases become available. E_{BN} remains unchanged since essentially there is no difference between phase I and the beam number retrieval described in the previous chapters. When adjusting the attribute weights based on each part of the solution, care has to be taken to not overfit the weights. Currently, when the solution only focusses on two parameters, namely the beam number and beam angles, this is not a concern. However, this issue needs to be considered in the future if the retrieval mechanism is divided into more parts in order to customise the weights for all treatment plan parameters.

In future work, the effect of the size of the case base on the retrieval error in later phases will be studied in detail. We are also interested in investigating what size can be considered to be large enough to ensure that a suitable treatment plan for any target case is available and the trade-off between case base size and retrieval speed.

Chapter 10

Conclusion

This thesis presents the development stages of a prototype decision support system that uses case-based reasoning techniques to facilitate radiotherapy treatment planning for brain cancer by suggesting treatment plan parameters of previously treated patient cases that are similar to the new patient. The work, presented in this thesis, focuses on the retrieval stage of the CBR system, in particular the similarity measure. Several methodologies based have been suggested to use in a CBR system for radiotherapy treatment planning. However, I would like to stress that the work presented is still in progress and on-going and there is plenty of scope for improvements.

This chapter provides a summary and discussion of the work carried out and the contents of the thesis. Section 10.1 summarises the experimental results obtained when testing the performance of the CBR systems on test cases. The performance of the novel concepts introduced in this work are compared and analysed. The contribution of this work, both with respect to the application domain, radiotherapy treatment planning and case-based reasoning (CBR), is outlined in section 10.2. Though the CBR system has been designed with keeping the application of radiotherapy treatment planning in mind, the concepts are applicable in a wide range of domains and can be adopted or adapted for use in other decision support systems as discussed in section 10.3. The chapter concludes with future work directions in section 10.5.

10.1 Summary of Performance Test Results

Validation of the CBR approaches presented in this thesis and also the cross validation methods used to determine the design parameters used the known treatment plans of the test cases obtained from the City Hospital as validation standards. The performance of the retrieval mechanism and the similarity measure is assessed using 22 brain cancer patient cases from the City Hospital as test cases. The same set of test cases has been used for all performance measurements so that the results can be compared and the improvement when using the developed concepts can be quantified.

Table 10. 1 and Table 10. 2 given below show the results obtained when running the CBR system using the various concepts outlined in the previous chapters in terms of beam number error E_{BN} and beam angles error E_{BA}, respectively. The final success rate of the retrieval mechanism with respect to beam number retrieval (phase I of the two phase retrieval mechanism) is 82% when using the fuzzy similarity measure with locally defined fuzzy membership functions and local attribute weights. With respect to beam angles retrieval, the success rate is 82%, when the entire case base is available for retrieval and when using the fuzzy similarity measure with locally defined fuzzy membership functions and local attribute weights. The success rate deteriorates to 68% in the two phase retrieval mechanism as the cases available for retrieval are limited as explained in section 9.3. From the results obtained, it can be seen that CBR is a viable technique in the generation of treatment plans for brain cancer radiotherapy. The success rate obtained provides a good starting point for the next stage of development in the CBR system, the adaptation phase.

Table 10. 1: A summary of all test results comparing the performance of the various retrieval and similarity measure mechanisms in terms of the treatment plan beam number error E_{BN} and the success rate.

Retrieval method	Chapter	Average error, E_{BN}	Success rate (%)
Random retrieval	Chapter 6	0.77	27
Weighted nearest neighbour	Chapter 7	0.36	68
Weighted nearest neighbour with local attribute weights	Chapter 7	0.27	77
Non-linear, fuzzy similarity	Chapter 8	0.36	68
Fuzzy similarity using local membership functions	Chapter 8	0.227	82
Two phase retrieval	Chapter 9	0.227	82

Table 10. 2: A summary of all test results comparing the performance of the various retrieval and similarity measure mechanisms in terms of the treatment plan beam angles error E_{BA} and the success rate

Retrieval method	Chapter	Average error, E_{BA}	Success rate (%)
Random retrieval	Chapter 6	40.48°	36
Weighted nearest neighbour	Chapter 7	32.27°	60
Weighted nearest neighbour with local attribute weights	Chapter 7	25.04°	73
Learning weights w.r.t. contents of case base	Chapter 7	22.99°	82
Non-linear, fuzzy similarity	Chapter 7	22.18°	82
Fuzzy similarity using local membership functions	Chapter 8	22.76°	82
Two phase retrieval	Chapter 9	28.83°	68

10.2 Contribution

The contribution of the research work can be viewed both in terms of improvements to treatment plan generation in brain cancer radiotherapy and the contribution made to the field of CBR.

10.2.1 Contribution in Radiotherapy Treatment Planning

The main contributions made to the field of radiotherapy treatment planning for brain cancer are as follows:

Problem Analysis

The radiotherapy treatment planning problem has been studied and analysed based on the literature and discussions with medical physicists at the Nottingham University Hospitals NHS Trust, City Hospital Campus. This thesis has presented the aims, guidelines and the challenges faced in radiotherapy treatment planning. The key parameters of a treatment plan have been identified and the manual trial and error based planning process used at many hospitals, including the City Hospital, has been outlined.

The need for an automated treatment planning systems has been identified. Manual treatment planning is a time consuming process requiring expertise and experience. After carefully reviewing the existing literature, current methods, which mainly focus on numerical optimisation or rule based algorithms, have been described, including their applications, advantages and disadvantages.

Advantages of CBR in Radiotherapy Treatment Planning for Brain Cancer

It has been found that many of the issues with manual treatment planning and automated treatment planning systems that were identified from the literature and after discussions with medical physicists at the City Hospital can be solved by CBR. Some of the defining characteristics of manual radiotherapy treatment planning make CBR a natural choice as the inference engine of an automated treatment planning system. Currently, manual treatment planning is a somewhat intuitive process that requires the expertise and experience of senior medical physicists, which makes CBR highly appropriate and applicable not just in radiotherapy treatment planning, but also in the wider domain of decision making in health care, in particular, in the absence of a well-defined mathematical model as is the case with radiotherapy treatment planning.

CBR can generate treatment plans quickly, it can take into account successful and failed plans used in the past and can work with incomplete or unusual cases. CBR also enables following the guidelines and protocols of an institution, which are inherently present in historical cases. Another important advantage is that with CBR, unlike many other automated treatment planning system, it is easy to see how a treatment plan has been derived. In other words, it does not work like a black box, which considerably increases the confidence of the user in the system.

Quick and Easy Input Data Extraction and Pre-processing

The CBR system is designed such that it only uses the information contained in the patient DICOM image directories as input. This means that no additional information has to be acquired by the user of the CBR system, reducing overheads and saving time and money. C++ and MATLAB code was developed to automatically extract and pre-process the data.

Identification of Relevant Case Attributes

Following detailed discussions with medical physicists at the City Hospital about manual treatment planning key geometrical parameters were identified that would model the manual planning process and could be used in an automated decision support system. The attributes model geometrical information about the tumour and describe the spatial relationship between the tumour (PTV) and organs-at-risk (OAR). They include the angle between the lines connecting the centroids of the PTV and the OAR to the origin of the patient coordinate system, the minimum edge to edge distance between the outlines of the PTV and OAR structures, the volume of the PTV, the ratio between the PTV and the patient body volume, the minimum edge to edge distance between the PTV and the body outlines and the location of the PTV with respect to the OARs. The attributes were calculated using the PTV and OAR 3D outline coordinates extracted from the patient DICOM files.

Treatment Plan Parameters (Decision Variables)

Currently, the CBR system focusses on recommending two treatment plan parameters, the number of beams used and their angles. However, since the entire treatment plan is retrieved it also contains information about the other parameters such as wedges or collimator angles. (The performance of the retrieval mechanism though, currently, is assessed only based on the beam configuration parameters, i.e. the beam number and the beam angles.)

Determination of Design Parameters and Performance Validation Using Real World Domain Cases

Design decisions about the retrieval mechanism and the similarity measure used in the CBR system were made using feedback about the retrieval error between the treatment plan parameters of the retrieved case and the known treatment plan of the target case. The parameters of the developed CBR system were trained using real brain cancer patient cases as training cases to accurately simulate real world situations. Further, the performance of the developed concepts for the retrieval mechanism and the similarity measure was assessed using real brain cancer patient cases as test cases. This gives an accurate indication of how well the system would perform in a real world application.

Success Rate of Retrieval Mechanism

The final success rate of the two phase retrieval mechanism is 88% for beam number retrieval and 68% for beam angles retrieval. It is expected that the beam angles success rate will improve considerably with a larger case base. However, this success rate is deemed as a promising starting point for adaptation. The prototype CBR system developed so far was briefly demonstrated to the medical physicists at the City Hospital. A few randomly chosen brain cancer patient cases were selected as target cases and the treatment plan of the retrieved case for each target case was evaluated by the medical physicists and deemed to be acceptable to be used for the target patients. An extended, more comprehensive and structured evaluation by the medical physicists of the performance of the retrieval mechanism is planned.

Improvements Expected in Medical Case of Brain Cancer Radiotherapy Patients

The work described in this thesis and the results obtained from the performance assessment provide evidence that CBR can be used as valid and successful inference method in a decision support system for radiotherapy treatment planning. The quantitative results (i.e., the success rate of the retrieval mechanism) and the qualitative advantages of using CBR (such as the ability to provide an explanation of how a treatment plan has been derived, the expected increased user confidence, the ability to work with incomplete and unusual cases, etc.) suggest that CBR in general, and the developed methodologies in particular, can considerably improve medical care in the following ways:

- Currently, manual treatment planning takes from a few hours to a few days. The time saved by medical physicists by using the CBR system means that they are freed up to concentrate on their other responsibilities. By speeding up the generation of treatment plans, the treatment of patients can start sooner, a factor, which is very important in the treatment of cancer patients. This also means that more patients can be treated increasing throughput of the radiotherapy department, which apart from the financial benefit in terms of resources, will improve patient care in general.
- The treatment plan generated by the CBR system can serve well as a starting point in training of junior and inexperienced medical physicists.

10.2.2 Contributions to the Field of Case-based Reasoning

The following contributions were made to the field of Case-based Reasoning:

Literature Review into CBR and Related Applications and Techniques

The fundamentals of CBR and the components of a CBR system have been explained. Relevant CBR systems, in particular in the field of health care, have been critically reviewed. The design considerations of CBR systems in general and specific to the problem domain of radiotherapy treatment planning have been listed. The background, requirements and motivation of the designed novel concepts in this work have been discussed by drawing examples from the literature of existing approaches, including their limitations, applications, advantages and disadvantages.

Imputation of Missing Values in the Case Base

A common problem in CBR systems, in particular in health care, is incomplete data. Missing values in the cases in the case base pose a big challenge not only during functioning of the system but also during the design stage. An easy to implement and quick imputation method that preserves the existing information in incomplete cases and imputes the missing values by using the correlation between attributes has been introduced. Further, a novel framework consisting of a step by step procedure has been outlined that adjusts the calculated similarity value between two cases if one or both of the cases have one or more imputed attribute values based on the confidence and success rate associated with an imputation method.

Validation in CBR Systems

Validation is important in the development of a CBR system both in terms of determining the design parameters of the system, for example of attribute weights and in evaluating the performance of the retrieval mechanism. The advantages of validation using feedback about the performance of the CBR system at least at the design stage are that it is objective and accurate. Further, it also reduces validation constraints in terms of time and cost associated with human evaluators. As validation is done using real world brain cancer cases, the human expertise is inherently available in the case base.

However, when real world data is used the size of the case base is often small during the design stage, which makes the choice of cross validation technique particularly important when design decisions have to be made in the absence of a large case base. The advantages, disadvantages and applications of a number of cross validation techniques have been discussed and compared. In this research work, mainly *k*-fold cross validation and leave-one-out cross validation were used. An attempt was made to avoid overfitting, which is a serious problem when using a small set of training cases, by not only using the error but also the variance in error between folds during *k*-fold cross validation during the training stage.

wNN Similarity Measure and Determination of Global and Local Attribute Weights

The most important component of the retrieval mechanism is the similarity measure, which has to be designed such that the computed similarity between two cases is relevant to the solution of the target case. The choice of similarity measure also depends on the case representation and the case attributes. In this work, the weighted *k*-nearest neighbour (kwNN) similarity measure was investigated and global weights of case attributes were determined. The retrieval mechanism using the kwNN similarity measure with global weights achieved a moderate success rate but resulted in the interesting finding that the lowest beam number retrieval error and beam angles retrieval error were obtained with different weight vectors. For this reason, in the work following, two different treatment plans were retrieved, one to suggest the number of beams and one to suggest the beam angles. Each plan was retrieved using optimised attribute weights based on the treatment plan parameter in question.

A local case attribute weighting scheme was designed, where weights are assigned to clusters of attribute values rather than individual attribute values to avoid over fitting. A rule-based methodology was developed to weight attributes in the target case. The rules were learnt based on which local weights resulted in successful retrievals on training cases. The rules were then pre-selected using the rule evaluation measures *support* and *confidence* that indicate the coverage and validity of a rule respectively. A third novel rule evaluation measure was introduced called the random retrieval probability (*RRP*) that gives an indication of the probability of obtaining an acceptable retrieval error based on the contents of the case base. By using this method, a very good success rate was obtained for both beam number and beam angles retrieval, indicating that local weights assigned using rules selected with the three described rule evaluation measures more accurately describe the importance of an attribute with respect to the treatment plan parameters than global rules.

Finally, an alternative method to learn rules for local attribute weights assignment was presented. The method based on the contents of the case base is useful with continuous plan parameters such as the beam angles. Its advantage is that it not only uses the absolute error but also considers the differences between the treatment plans of two cases when providing feedback about the retrieval success during the training phase. As expected, this resulted in an improved success rate during beam angles retrieval.

Non-linear, Fuzzy Similarity Measure using global and locally defined membership functions

A similarity measure using fuzzy sets was designed to take into account differences in the distribution of attribute similarity values, which might result in them not being directly comparable. The use of fuzzy sets also allows separately weighting large and small similarity components, which effectively represent similarity and dissimilarity. In this application, the results indicate that exclusively using the similarity or dissimilarity does not result in a good retrieval performance. However, by using both similarity and dissimilarity but ignoring average similarity values, which possibly "dilute" the computed similarity between the attributes of two cases good results are obtained with the fuzzy similarity measure, in particular, for beam number retrieval. A method for using local fuzzy membership functions for each target case has been presented, which substantially improves the beam number retrieval performance. With respect to the beam angles using globally or locally defined membership functions does not improve the results considerably indicating that the beam angles of the treatment plans are less sensitive to the distribution of attribute similarities than the beam number.

Two Phase Retrieval Mechanism

As mentioned previously, better retrieval performance is obtained when different weight vectors are used for beam number and beam angles retrieval. This has led to the implementation of a two phase retrieval system in which the attribute weights used in each phase are optimised with respect to part of the solution, i.e. individual parameters of the treatment plan. The advantage of this method lies mainly in its practicality. Retrieving a single treatment plan that suggests both beam numbers and beam angles is sub-optimal. However, retrieving two treatment plans in parallel and independently can lead to conflicts in the parameters of the two treatment plans, which is avoided using the two phase retrieval mechanism. A drawback of this method is that the case base available in phase II is substantially reduced. In small case bases such as ours, this leads to deterioration in the success rate of beam angles retrieval in phase II. However, it is expected that the success rate will improve considerably as the case base grows. Performance testing the system is a regular process during development and will take part as and when new cases become available.

Determination of system parameters on a small set of real world training cases

Developing decision support systems using real world data is often preferred as it is able to model more accurately the requirements and challenges of the application domain and take into account practical considerations. However, a common problem, especially in healthcare applications, is the availability of real world data, in particular, during the design and development stage. Since collecting real world data can be a very time consuming process, often development has to be done on an initially small set of data. During training, when using a feedback loop about the performance of the system to guide the determination of the system parameters, common problems encountered when using a small set of training cases include over fitting and the reliability of the feedback obtained from the system's performance, which can bias the results.

- To avoid overfitting, in this work, not only the retrieval error between treatment plan parameters was used but also the variance in error between folds when using *k*-fold cross validation to determine global attribute weights.
- The *random retrieval probability* assesses the quality of the feedback obtained about the retrieval error by taking into account the contents of the case base. It measures, for a given target case, the ratio among the cases available for retrieval between the cases having suitable treatment plans and unsuitable treatment plans for the target case. In situations, where a successful retrieval is not necessarily due to the performance of the similarity measure or its parameters but the favourable existence of a large number of suitable treatment plans, the feedback obtained is discounted.
- Another strategy to take into account the contents of the case base during the training phase is outlined in section 7.3. In this method, feedback about the retrieval performance that guides determination of system parameters is obtained not only from the absolute error between solution parameters but also by the similarity between the treatment plans. The purpose of this method is to take into account that in small case bases a retrieval is successful if it retrieves the case with the best treatment plan in spite of having a large retrieval error as in this situation, the large retrieval error is due to the contents of the case base rather than the performance of the retrieval mechanism.

10.3 Applicability to Other Domains

Owing to its ability to capture subjective experience and intuitive knowledge CBR is applicable to many applications. The CBR system introduced in this thesis has been developed with a focus on radiotherapy treatment planning. However, the concepts introduced are applicable to many other domains.

Healthcare applications

The prototype CBR system has been developed with a focus on radiotherapy treatment planning for brain cancer at the City Hospital. However, the system can directly be used in any other hospital to assist brain cancer treatment planning, as long as similar planning protocols are followed. Also, the CBR system can be used for treatment planning in other types of cancer that follow a similar procedure such as lung cancer or head and neck cancer. Straight forward head and neck cancer patients can be directly added to the case base or more complex head and neck cancer cases, with additional attributes that consider specific head and neck issues such as cancerous nodules and two-stage treatment planning.

Case Attributes

Exploiting the spatial relationship between objects in an image as introduced by Berger (1994) and further explored in our work is highly applicable in any domain that works with images, in particular other health care applications where diagnosis or treatment is based on the image. In domains using the DICOM image standard, the standard used in most clinical applications but also other domains, highly sophisticated image processing might not be necessary if as in our work, the required data about the image and image objects can be extracted from the key value pairs found in the DICOM image header.

Attribute Weighting

Attribute Weighting is a significant aspect of most decision support systems, not just CBR systems. For instance, it is also widely used in classification systems. In many applications the importance of attributes is not entirely independent of each other and the importance of parameters often depends on their values or the values of other parameters. The method of generating and selecting rules using clustering of attribute values and rule evaluation measures such as the confidence and support of a rule and the *random retrieval probability* of a retrieval instance is highly applicable to any weighting scheme, in which the importance of attributes depend on its own value and the values of other attributes.

The Fuzzy Similarity Measure

The fuzzy similarity measure is recommended for use in any similarity calculation where the distributions of attribute similarities are not comparable. This is easily measured by plotting and comparing the distribution of the attributes used. Defining fuzzy membership functions is not an easy task and in the absence of domain expert advice, they are often defined rather arbitrarily. Using the minimum, average and maximum values of attribute similarities (or data values in other applications) to specify the support and centre of the membership function represents a slightly more accurate method of defining the fuzzy membership functions based on the data itself. The accuracy can be increased by using locally defined fuzzy membership functions. Many applications would benefit from being able to separate the similarity and dissimilarity component of the similarity between two cases and weighting them independently. In some applications, using the similarity is more important whereas in others using the dissimilarity is more important, while in others again using a combination of both with different weights is the most appropriate.

Multi-Phase Retrieval

Multi-phase retrieval is particularly applicable in multi criteria decision making, as is often the case in health care domains. It can be applied whenever the solution of a CBR system consists of more than one parameter though the maximum benefit is obtained when the importance of attributes or the value of other parameters of the inference engine differs with respect to each solution parameter. When this is the case, an obvious challenge will be to make the various retrieved case solutions for a target case compatible and solve contradictions between treatment plans. In particular, in health care, the solution parameters, such as medicines, often have to be compatible with each other. By using multiple phases in sequence rather than parallel, where each phase uses the case base filtered according to the solution of the previous phase, this problem can be avoided. Care has to be taken, however, when small case bases are used as the size of the case base effectively reduces in each phase.

Applications of Imputation Frame Work

Missing values are a common problem in databases. The filter imputation method can be modified to impute missing values in any data mining or data processing application as it is not limited to CBR. Similarly, the imputation framework could be useful in other applications as well as the validity of imputation in any domain is a concern that needs to be considered.

10.4 Scope and Applicability of CBR to RTP

The research presented in this thesis demonstrates that CBR is a promising technique in the radiotherapy treatment planning problem. However, its disadvantages and limitations need to be carefully considered as well. For instance, a serious drawback of CBR systems is that the quality of the generated solutions is limited by the quality of the cases in the case base. As applied to the radiotherapy treatment planning problem this means that if the treatment plans stored in the case base are sub-optimal, the solutions generated by the CBR system are necessarily sub-optimal as well. For this reason it is important to periodically review the case base and to replace cases for which better treatment plans have been found. A good adaptation module can help with this issue as well. In adaptation, the retrieved treatment plan is evaluated for its suitability to the target patient case and adapted to correct dose violations. At this point, the treatment plan could be improved by a domain expert or possibly by optimisation techniques. In this type of hybrid system, the role of the CBR system would be to retrieve a treatment plan that provide a good starting point for optimisation, as suggested in section 10.5

10.5 Future Work

This thesis has described the retrieval mechanism of a CBR system for radiotherapy treatment planning in brain cancer. Test results reveal that the developed concepts work well in the retrieval mechanism of a CBR system for radiotherapy treatment planning for brain cancer. However, more work can be done to improve the performance of the retrieval mechanism, in particular, and the CBR system, in general.

Adaptation

Adaptation is a stage of CBR where the solution of the retrieved plan is modified with respect to the specifics of the target case. The performance of the case retrieval mechanism and its similarity measure is crucial to the working of a CBR system. However, frequently, differences do exist between the target case and the retrieved case. Once the most similar case has been retrieved, its solution generally has to be adapted to the specific needs of the target case. Adaptation can be done by adjusting the beam configuration according to the geometric displacement in the location of the tumour and OAR structures of the target case compared to the retrieved case. Another method would be to evaluate the dose distribution of the treatment plan to identify dose violations of OAR and confirm tumour coverage. The adaptation module of the CBR system then has to adjust the plan parameters of the retrieved case to resolve the violations. This could be done using if-then rules, for instance "If there is a hot spot (area of high dose) in the OAR, add another beam to reduce the dose intensity of all beams". Adaptation can also be carried by using another case-based reasoning inference mechanism. This requires the use of another case base that contains examples of adapted treatment plans. This kind of case base contains pairs of cases with their treatment plans including the process that was used to adapt the first treatment plan to make it suitable for the second case. Adaptation would consist of retrieving a case pair from the adaptation case base that is similar to the retrieved case-target case pair and applying the same adaptation process. Currently, work is done by the research group to apply a knowledge light method of adaptation (Mishra et al., 2009) that identifies the differences in solution parameters with respect to corresponding differences in case attributes and adjust the solution parameters accordingly.

Application to IMRT

Currently, there is work in progress to apply the developed CBR approaches to IMRT head and neck cancer cases. The aim is initially to determine the number of beams and their angles and it is expected that the developed CBR system can be applied mainly directly to the IMRT patient cases to determine these two parameters. The main difference between IMRT and 3D conformal radiotherapy, however, is that in IMRT, each beam is divided into a large number of beamlets, each of which can have a different intensity or weight in order to allow closer shaping of the radiation beam to the tumour. We are going to investigate how to extend the functionality of the CBR system to include beamlet intensities along with the other treatment plan parameters.

Extension of Case Attributes

The chosen case attributes clearly are related to the treatment plan parameters as demonstrated by the improved success rate when using an intelligent retrieval system instead of retrieving a random case for the target case. However, it would be interesting to investigate the use of other attributes to describe cases. At the moment we cannot be certain that the chosen attributes entirely or accurately describe the patient with respect to the treatment plan parameters. Additional attributes could be other geometrical or spatial attributes such as the shape of the PTV/OARs, their orientation, or their position with respect to the patient's anatomy (Berger, 1994). Currently, we only use information that can be extracted from the patient DICOM image files to simplify and reduce information collection for the user of the CBR system. However, a possibility would be to incorporate additional clinical or patient information such as the age of the patient. The age and therefore the expected life expectancy of the patient influences the treatment aims. Young patients with a potentially long life expectancy are treated aggressively with a focus on cure (radical treatment) whereas elderly patients are treated with a focus on pain management and preserving the quality of life of the patient's remaining life (palliative treatment) rather than cure. For instance, in younger patients, organs such as the eyes might be sacrificed in order to completely destroy all tumour cells whereas with elderly patients current practice is to try to preserve the eyes, even at the cost of not completely irradiating the tumour. This clearly has implications on treatment planning and the determination of the planning parameters. Another attribute that also affects treatment planning is termed as the fitness of a patient and takes into account the patient's mobility and general health. Information such as the age or the fitness of a patient, which is commonly present in patient records, could easily be incorporated in the CBR system.

Treatment plan parameters

Currently, the CBR system retrieves two cases and their treatment plans, one to suggest the number of beams and one to suggest their angles. In both cases, the entire treatment plan is retrieved, so technically other planning parameters such as wedges, collimator angles or beam weight, are available to be used for the target case, however, the retrieval mechanism is not optimised for these additional parameters and their suitability has not been evaluated. Treating a patient requires all treatment plan parameters, even though the number of beams and their angles are the most important ones and influence the choice and values of the others. The additional parameters could be retrieved in subsequent phases of a multi-phase retrieval mechanism with the appropriate weights or they could be grouped and taken from the treatment plan in phase I or phase II.

Hybrid methods

A lot of interest is shown in the literature in hybrid methods that solve a problem using two or more methodologies. In radiotherapy treatment planning an excellent approach would be to use CBR to short list treatment plans for a patient. These treatment plans can be input into a second stage that uses mathematical modelling or optimisation methods to accurately calculate the parameter values. The advantage of such a system would be that the optimisation module does not have to start generating a treatment plan from scratch but starts with a feasible one. Using optimisation methods it might be possible to obtain more accurate results for the treatment plan parameters.

Another interesting possibility is to use CBR to determine the beam numbers and then use optimisation techniques to determine the angles. The beam angles of the treatment plan of the retrieved case (for instance in phase I of the two-phase retrieval mechanism) with the beam number fixed by this treatment plan could be used to provide a starting point for a local search algorithm to find optimal beam angles.

Validation of Retrieval Performance

A limitation of the evaluation method used in this research is that for practical considerations (discussed in section 6.1.1), the system is designed and evaluated by comparing the retrieved treatment plans of test target cases with the known or existing treatment plans of these cases using the assumption made by medical physicists at the City Hospital that the treatment plans of all cases in the case base constitute good and valid treatment plans. However, in reality this assumption might not always be valid and it also limits the evaluation of the system to exclusively the cases contained in the case base. As mentioned previously, the quality of a CBR system is only as good as its case base and ideally the aim of the CBR system is to generate treatment plans that are as good as manually generated plans or better.

Therefore, an important method of evaluation (both in evaluating the performance of the CBR system and also in cross validation to determine design parameters) that is planned in the future is to let the retrieved plan for a target case be evaluated by the medical physicists themselves or an oncologist. This offers logistical problems as the time of medical physicists is highly limited, however, it is expected that it would not only provide a second method of evaluation but would ultimately increase the confidence of users in the system if they have been clinically tested. A compromise would be to use a full treatment planning system, which is capable of calculating the dose distribution based on the DICOM image files and the treatment plan parameters. The dose distribution resulting from a treatment plan would show if a plan is feasible for a patient and areas of under dosing and over dosing. However, medical physicists would be able to evaluate all aspects of the CBR system, including factors such as user friendliness of the system.

Missing Values

The algorithm for missing values has been evaluated using prostate cancer patient cases. It would be very interesting to see how the algorithms perform in the developed CBR system for brain cancer radiotherapy. Also, in the future, we will study imputation of continuous attributes and how the percentage of missing values in the case base affects the performance of the imputation method or influences the choice or parameters of imputation methods.

Local Attribute Weighting

Based on our experiments with a reduced case base of 41 cases and the full case base of 86 cases, the attribute weights change slightly as the contents of the case base changes. This means that the rules for assigning the local weights should be updated whenever a large number of cases are added to the case base. Also, during clustering, the attribute values are assigned to crisp clusters, *Large* or *Small,* which create artificial boundaries between attributes that are continuous in nature. Using fuzzy sets as attribute value clusters might enable us to obtain a more accurate representation of how large or small an attribute value is and how its fuzzy membership grade affects the significance of attribute weights.

Fuzzy Similarity Measure

Currently, the aggregate similarity value between two cases is computed as the sum of the large and medium fuzzy components (representing similarity) from which the small fuzzy component (representing dissimilarity) is subtracted. However, there are many other types of aggregation methods described in the literature. In this work, we have tested several others such as using the fuzzy component with the maximum value or the product/ratio of the fuzzy components. Preliminary results obtained were inferior compared to the current method but it cannot be ruled out that other methods described in the literature such as using polynomials or splines would not give more accurate results.

Currently, weights in the fuzzy similarity measure are crisp. It would be interesting to use fuzzy weights such as *small, slightly small, medium, slightly large, large.* An interesting approach would also be to use fuzzy rules in the computation of similarity between cases. The rules could assign fuzzy similarity values depending on the magnitude of similarity between attributes of two cases. The fuzzy attribute similarity value could then be aggregated to form the total similarity. This approach would eliminate the need to define sets of attribute weights.

Testing concepts using a larger case base and benchmark data

A challenge of this work has been to design and test the retrieval mechanism using a small case base. Problems arising from the small case base include mainly the coverage of problem scenarios, the possibility of overfitting when determining design parameters using cross validation techniques and also bias in the case base. The treatment plans of the cases in the case base were deemed as successful by medical physicists; however their quality has not been independently verified, for instance using alternative techniques such as optimisation algorithms. As the quality of a CBR system is highly dependent on the quality of the case base, the case base is a vital factor and deficiencies in the case base deteriorate the performance of the system. For instance, currently, there exists a possibility that for a given target case, no similar or suitable case exists in the case base. A larger case base would help to alleviate some of these problems. Collection of brain cancer patient cases is an on-going process and it is important to evaluate how the size or added content of the case base affects the retrieval performance. It would also be interesting to test the concepts, such as the local attribute weighting scheme using rules, the fuzzy similarity measure and the concept of *random prediction probability* with data sets of other domains or standardised data sets.

A few examples of websites that provide data repositories for a variety of domains are mentioned below:

1) http://www.statsci.org/datasets.html

This website contains links to datasets such as raw data, pre- and fully processed data and statistical data that can be used for training and teaching

2) http://www.stat.ucla.edu/data/

Links to a large variety of datasets can be found on the webpages of the University of California at Los Angeles.

3) http://data.worldbank.org/

The online world bank of data contains a wealth of information in datasets on a large number of topics from health to financial data.

Appendix A: DICOM RT Image Files

In this appendix, snapshots of diagrams of DICOM files, from the Radiotherapy RT DICOM supplement to the DICOM standard, are presented. All files shown in the tables are part of a DICOM patient directory containing DICOM image files obtained from the City Hospital. In the City Hospital, a radiotherapy patient directory usually contains the CT DICOM image slices, RS (structure file), RP (treatment plan file), RD (radiation dose file) and RI (image file).

General DICOM Data File

A DICOM data file consists of both image information and header or textual information such as the patient ID or the hospital name. The header of a DICOM file is optional and included in the data file. It consists of a 128 bytes preamble, followed by 4 bytes called the DICOM prefix. The information is encoded as data elements. The structure and encoding of information is described in part 3. 5 of the DICOM standard (National Electrical Manufacturers Association, 2011a). Each data element of a DICOM file consists of the following parts as shown in Table 1 :

Table 1: DICOM Data Element

Data Element Tag	Value Representation	Value Field	Value Length	Value Description
(0010,0010)	PN	[Anonymous1]	10, 1	Patient Name

1) DICOM Tag: The tag uniquely identifies each data element. It consists of a *group number* and an *element number,* usually in hexadecimal format. In the example shown

in Table 5.4, the group number *0010* relates to personal patient information and along with the element number *0010* describes the patient's name.

2) Value Representation: This field relates to the data type and format. For instance, *PN* is a character string standing for *Person Name.*
3) Value Field: The value field denotes the actual value of the element, in the example this is the patient's name.
4) Value Length: This denotes the length of the value field.
5) Value Description: A textual description of the data element.

A list of data elements including their tag, value representation and description can be found in part 3.6 of the DICOM standard (National Electrical Manufacturers Association, 2011b).

The diagrams below contain parts of the following DICOM files:

<u>DICOM CT patient file</u>

The DICOM CT patient files contain pixel information (*PixelData*) to display the CT images of the relevant region in the patient's body. In addition, they contain textual information about the patient, such as the patient name (*PatientsName*), birth of date (*PatientsBirthDate*), details about the study, such as the date when the study was started (*StudyDate*) or the image acquisition date (*AcquisitionDate*) and details about the institution, such as the institution's name (*InstitutionName*) and address (*InstitutionAddress*). An excerpt from a DICOM CT patient file is shown in Figure 1. The pixel data is contained in the data element tag (7fe0,0010) as shown in the example in Figure 2.

```
# Dicom-File-Format

# Dicom-Meta-Information-Header
# Used TransferSyntax: LittleEndianExplicit
(0002,0000) UL 216                         # 4, 1 MetaElementGroupLength
(0002,0001) OB 00\01                       # 2, 1 FileMetaInformationVersion
(0002,0002) UI =CTImageStorage             # 26, 1 MediaStorageSOPClassUID
(0002,0003) UI [1.3.6.1.4.1.2452.6.2046433570.1277320819.1401479102.978761052] # 62, 1
MediaStorageSOPInstanceUID
(0002,0010) UI =LittleEndianExplicit       # 20, 1 TransferSyntaxUID
(0002,0012) UI [1.2.250.1.59.3.0.3.5.3]    # 22, 1 ImplementationClassUID
(0002,0013) SH [ETIAM_DCMTK_353]           # 16, 1 ImplementationVersionName
(0002,0016) AE [STORESCU]                  # 8, 1 SourceApplicationEntityTitle

# Dicom-Data-Set
# Used TransferSyntax: LittleEndianExplicit
(0008,0005) CS [ISO_IR 100]                # 10, 1 SpecificCharacterSet
(0008,0008) CS [ORIGINAL\PRIMARY\AXIAL\CT_SOM5 SEQ]  # 34, 4 ImageType
(0008,0016) UI =CTImageStorage             # 26, 1 SOPClassUID
(0008,0018) UI [1.3.6.1.4.1.2452.6.2046433570.1277320819.1401479102.978761052] # 62, 1
SOPInstanceUID
(0008,0020) DA [20100216]                  # 8, 1 StudyDate
(0008,0021) DA [20100216]                  # 8, 1 SeriesDate
(0008,0022) DA [20100216]                  # 8, 1 AcquisitionDate
(0008,0023) DA [20100216]                  # 8, 1 ContentDate
(0008,0030) TM [112650.265000]             # 14, 1 StudyTime
(0008,0031) TM [113926.328000]             # 14, 1 SeriesTime
(0008,0032) TM [114205.916347]             # 14, 1 AcquisitionTime
(0008,0033) TM [114205.916347]             # 14, 1 ContentTime
(0008,0050) SH [469199]                    # 6, 1 AccessionNumber
(0008,0060) CS [CT]                        # 2, 1 Modality
(0008,0070) LO [SIEMENS]                   # 8, 1 Manufacturer
(0008,0080) LO [Nottingham City  Hospital] # 26, 1 InstitutionName
(0008,0081) ST [Hucknall Road Nottingham District GB]  # 36, 1 InstitutionAddress
(0008,0090) PN (no value available)        # 0, 0 ReferringPhysiciansName
(0008,1010) SH [C36617]                    # 6, 1 StationName
(0008,1030) LO [Head^01_HeadNeckSeq]       # 20, 1 StudyDescription
(0008,103e) LO [Head Seq 3.0 H30s]         # 18, 1 SeriesDescription
(0010,0010) PN [Patient39]                 # 10, 1 PatientsName
(0010,0020) LO [University]                # 10, 1 PatientID
(0010,0030) DA (no value available)        # 0, 0 PatientsBirthDate
(0010,0040) CS (no value available)        # 0, 0 PatientsSex
(0018,0015) CS [HEAD]                      # 4, 1 BodyPartExamined
(0018,0050) DS [3]                         # 2, 1 SliceThickness
(0018,0060) DS [130]                       # 4, 1 KVP
(0018,1020) LO [VA47C]                     # 6, 1 SoftwareVersions
(0018,1030) LO [01_HeadNeckSeq]            # 14, 1 ProtocolName
(0020,1041) DS [-660]                      # 4, 1 SliceLocation
[...]
```

Figure 1: Example of a DICOM CT Image File

(7fe0,0010) OW 0013\0018\001c\ [....\....\....\....] \0015\0012\0012\0016\001a\0018\
\0043\0047\004b\004a\0047\0047\004a\004b\0048\0045\0043\0047 # 524288, 1 PixelData

Figure 2: DICOM pixel data in DICOM CT image file

RT Structure Set (RS)

The RS DICOM images contain information regarding the structure outlines as drawn by the oncologist on the patient image. Examples of structures, also called regions of interest (ROI) include the GTV, CTV, PTV, OAR, body contour and reference points. Each ROI is numbered and described using tag descriptions such as *ROIName*, *ROINumber*, and *ROIDisplayColour* among others. The structure outlines are recorded in the form of their [x\y\z] coordinate triplets with the data element tag *(3006,0050)* called *ContourData* as seen in Figure 3. Each ROI can have several *ContourData* fields, often one for each image slice. These files do not contain any pixel data but only textual information.

(3006,0050) DS [11.870000\-331.810000\-615.000000\2.000000\-332.450000\-615.000000\-
9.460000\-331.810000\-615.000000\-17.110000\-329.900000\-615.000000\-24.750000\-327.040000\-
615.000000\-34.940000\-323.530000\-615.000000\-42.580000\-320.350000\-615.000000\-
49.270000\-315.250000\-615.000000\-50.230000\-307.930000\-615.000000\-43.860000\-
301.880000\-615.000000\-35.260000\-303.150000\-615.000000\-30.480000\-309.520000\-
615.000000\-26.340000\-310.160000\-615.000000\-21.090000\-312.550000\-615.000000\-
11.530000\-313.180000\-615.000000\-4.690000\-306.340000\-615.000000\3.910000\-292.010000\-
615.000000\12.190000\-301.560000\-615.000000\12.830000\-307.930000\-615.000000\13.780000\-
311.110000\-615.000000\20.790000\-310.160000\-615.000000\29.700000\-305.380000\-
615.000000\42.120000\-305.700000\-615.000000\50.880000\-309.040000\-615.000000\53.910000\-
315.570000\-615.000000\47.220000\-320.990000\-615.000000\34.800000\-326.080000\-
615.000000\19.510000\-330.540000\-615.000000] **# 962,84 ContourData**

Figure 3: Contour Data in DICOM File

RTPlan (RP)

The RP images do not contain any pixel data either but only textual information regarding the treatment plan parameters. Among others, they record the isocentre coordinates, the beam number and for each beam provide the gantry, collimator and patient couch angle, beam weight, use of wedges and their weights, and multi leaf collimator settings. Figure 4 shows an excerpt from a DICOM RP file.

```
(300a,011e) DS [270.00]                 # 6, 1 GantryAngle
    (300a,011f) CS [NONE]                  # 4, 1 GantryRotationDirection
    (300a,0120) DS [90.00]                # 6, 1 BeamLimitingDeviceAngle
    (300a,0121) CS [NONE]                  # 4, 1 BeamLimitingDeviceRotationDirection
    (300a,0122) DS [0.00]                 # 4, 1 PatientSupportAngle
    (300a,0123) CS [NONE]                  # 4, 1 PatientSupportRotationDirection
```

Figure 4: RP DICOM image file

Appendix B: Medical Dictionary

3D Conformal Radiotherapy	Three-dimensional (3D) conformal radiation therapy is a technique where the beams of radiation used in treatment are shaped to match the tumor.
beamlets	In intensity modulated radiation therapy (IMRT), the beam is divided into a large number of beamlets that can be modulated individually to achieve better conformation of the radiation to the tumour volume
benign tumour	Tumour that does not metastatize (spread to other parts of the body)
Brachytherapy	Form of radiotherapy where radiation source is placed next to the tumour inside the body
Chemotherapy	Form of cancer treatment using drugs
Clinical stage	Label indicating extent of cancer
Clinical Target Volume (CTV)	Treatment volume including tumour and a margin around the tumour to allow for sub-clinical spread of tumour
Cold spots	Regions of underdosing in planning target volume
Collimator	Device used to shape the radiation into a narrow beam
Computed tomography (CT)	X-ray imaging modality that computes the image from slices taken of the body
Coplanar	Lying on the same plane
Dose violations	Areas that don't conform to the treatment plan. Violations could be areas of over or underdosing or non-uniform or incomplete tumour coverage
Dose volume histogram (DVH)	A plot of a cumulative dose-volume frequency distribution, which graphically summarizes the simulated radiation distribution within a volume of interest of a patient which

	would result from a proposed radiation treatment plan.
Forward planning	Treatment planning method, which involves a trial and error method of modelling plan parameters such as the beam configuration, evaluating the resulting dose distribution and modifying the plan parameters again based on dose violations. This procedure is repeated till an acceptable dose distribution is obtained.
Gantry	A device for rotating the radiation delivery apparatus around the patient during radiation therapy.
Gleason score	System of grading prostate cancer.
Gray	Unit of Radiation
Gross Target Volume (GTV)	Area containing tumourous cells visible in image.
Hot spots	Areas of overdosing
Intensity Modulated Radiation Therapy (IMRT)	Advanced mode of high-precision radiotherapy that uses computer-controlled linear accelerators to deliver precise radiation doses to tumourous cells.
Inverse planning	Method of treatment planning, where first the desired dose distribution is determined and then the required treatment plan parameters calculated.
Linear accelerator (Linac)	Device used to apply radiation to cancer patients.
Magnetic resonance imaging (MRI)	Imaging modality based on nuclear magnetic resonance of hydrogen nucleii. Usually used to image soft tissue.
Malignant tumour	Tumour that metastasizes (spreads to other part of body)
Multileaf collimator	Collimator that uses leaves to shape the radiation beam.
Non-coplanar	Beams can be applied from any angle and do not have to lie on the same plane.
oncologist	Physician specialising in cancer treatment
Organs-at-risk (OAR)	Organs in the vicinity of the tumour that are at risk from radiation
Patient support angle	Angle of patient bed.
Planning Target Volume (PTV)	Volume incorporating CTV and GTV and an additional margin to allow of uncertainties in planning and radiation delivery.
Prescribed dose	Dose to be applied to the PTV.

Prostate specific antigen (PSA)	Substance produced by prostate gland. Elevated amounts in the blood can be an indication of prostate cancer.
Radiotherapy	Form of cancer treatment that uses ionising radiation to kill tumour cells.
Tumour	Tissue with abnormal growth.
Wedges	Wedge shaped metallic blocks used to attenuate radiation.

Bibliography

AAMODT, A., GUNDERSEN, O. E., LOGE, J. H., WASTESON, E. & SZCZEPANSKI, T. Case-based reasoning for assessment and diagnosis of depression in palliative care. IEEE 23rd international symposium on computer-based medical systems (CBMS), 12-15 Oct. 2010. 480-485.

AAMODT, A. & PLAZA, E. 1994. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications,* 7, 39-59.

ABDALA, O. T. & SAEED, M. Estimation of missing values in clinical laboratory measurements of ICU patients using a weighted k-nearest neighbors algorithm. Computers in Cardiology, Sept. 2004. 693-696.

ACOCK, A. C. 2005. Working with missing values. *Journal of Marriage and Family,* 67, 1012-1028.

AGGOUR, K. S., PAVESE, M., BONISSONE, P. P. & CHEETHAM, W. E. Soft-CBR: A self-optimizing fuzzy tool for case-based reasoning. Lecture Notes in Computer Science, 2003.

AHA, D. W. & BANKERT, R. L. 1994. Feature selection for case-based classification of cloud types: An empirical comparison. *Association for the Advancement of Artificial Intelligence.*

AHA, D. W. & GOLDSTONE, R. L. 1992. Concept learning and flexible weighting. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society.* Erlbaum.

AHMAD, S. U. & BERGEN, S. W. A. 2010. A genetic algorithm approach to the inverse problem of treatment planning for intensity-modulated radiotherapy. *Biomedical Signal Processing and Control,* 5, 189-195.

AHMED, M. U., BEGUM, S., FUNK, P., XIONG, N. & VON SCHEELE, B. 2011. A multi-module case-based biofeedback system for stress treatment. *Artificial Intelligence in Medicine,* 51, 107-115.

ALEMAN, D., KUMAR, A., AHUJA, R., ROMEIJN, H. & DEMPSEY, J. 2008. Neighborhood search approaches to beam orientation optimization in intensity modulated radiation therapy treatment planning. *Journal of Global Optimization,* 42, 587-607.

ALLEN, B. P. 1994. Case-based reasoning: Business applications. *Communications of the ACM,* 37, 40-42.

ALTHOFF, K.-D. Validation of case-based reasoning systems. Proceedings, 5. Leipziger Informatik-Tage, 1997.

ANOOJ, P. N. 2011. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules and decision tree rules. *Central European Journal of Computer Science,* 1, 482-498.

ASHLEY, K. D. 1991. *Modeling legal arguments: Reasoning with cases and hypotheticals,* Bradford Books.

BAHR, G. K., KEREIAKES, J. G., HORWITZ, H., FINNEY, R., GALVIN, J. & GOODE, K. 1968. The method of linear programming applied to radiation treatment planning. *Radiology,* 91, 686-693.

BANDINI, S. & MANZONI, S. Application of fuzzy indexing and retrieval in case based reasoning for design. Proceedings of the 2001 ACM symposium on Applied computing, 2001 Las Vegas. 462 - 466.

BÄR, W., SCHWARZ, M., ALBER, M., BOS, L. J., MIJNHEER, B. J., RASCH, C., SCHNEIDER, C., NÜSSLIN, F. & DAMEN, E. M. F. 2003. A comparison of forward and inverse treatment planning for intensity-modulated radiotherapy of head and neck cancer. *Radiotherapy and Oncology,* 69, 251-258.

BARNARD, J. & MENG, X. 1999. Applications of multiple imputation in medical studies: From AIDS to NHANES. *Statistical Methods in Medical Research,* 8, 17--36.

BEDDOE, G. R. & PETROVIC, S. 2006. Selecting and weighting features using a genetic algorithm in a case-based reasoning approach to personnel rostering. *European Journal of Operational Research,* 175, 649--671.

BEGUM, S., AHMED, M. U., FUNK, P., NING, X. & FOLKE, M. 2011. Case-based reasoning systems in the health sciences: A survey of recent trends and developments. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,* 41, 421-434.

BEGUM, S., AHMED, M. U., FUNK, P., XIONG, N. & V. SCHÉELE, B. Classify and diagnose individual stress using calibration and fuzzy base-based reasoning. International Conference on Case-Based Reasoning 2007, LNAI, 2007.

BELLAZZI, R., MONTANI, S. & PORTINALE, L. 1998. Retrieval in a prototype-based case library: A case study in diabetes therapy revision. *In:* SMYTH, B. & CUNNINGHAM, P. (eds.) *Advances in case-based reasoning.* Springer Berlin Heidelberg.

BERGER, J. 1994. Roentgen: Radiation therapy and case-based reasoning. *Proceedings of the Tenth Conference on Artificial Intelligence for Applications.*

BERGMANN, R., KOLODNER, J. & PLAZA, E. 2005. Representation in case-based reasoning. *The Knowledge Engineering Review,* 00, 1-4.

BERMEJO, P., GÁMEZ, J. A. & PUERTA, J. M. 2011. A grasp algorithm for fast hybrid (filter-wrapper) feature subset selection in high-dimensional datasets. *Pattern Recognition Letters,* 32, 701-711.

BERTSIMAS, D., CACCHIANI, V., CRAFT, D. & NOHADANI, O. in press (doi:10.1016/j.cor.2012.06.009) A hybrid approach to beam angle optimization in intensity-modulated radiation therapy. *Computers and Operations Research.*

BICHINDARITZ, I., KANSU, E. & SULLIVAN, K. M. Case-based reasoning in care-partner: Gathering evidence for evidence-based medical practice. Proceedings of the 4th European Workshop on Advances in Case-Based Reasoning, Lecture Notes In Computer Science, 1998. 334 - 345.

BICHINDARITZ, I. & MARLING, C. 2006. Case-based reasoning in the health sciences: What's next? *Artificial Intelligence in Medicine,* 36, 127-135.

BIDGOOD, W. D., HORLL, S. C., PRIOR, F. W. & SYCKLE, D. E. V. 1997. Understanding and using DICOM, the data interchange standard for biomedical imaging. *Journal of the American Medical Informatics Association,* 4, 199-212.

BLANZIERI, E. & RICCI, F. Probability based metrics for nearest neighbor classification and case-based reasoning. *In:* ALTHOFF, K.-D., BERGMANN, R. & BRANTING, L. K., eds. Proceedings of the 3rd International Conference on Case-Based Reasoning Research and Development (ICCBR-99), volume 1650 of LNAI, 1999. Springer Berlin / Heidelberg, 73-73.

BONISSONE, P. P. & MANTARAS, R. L. D. 1998. *Fuzzy case-based reasoning systems, handbook on fuzzy computing (f 4.3),* Oxford University Press.

BONZANO, A., CUNNINGHAM, P. & SMYTH, B. Learning feature weights for CBR: Global versus local. Proceedings of the 5th Congress of the Italian Association for Artificial Intelligence 1997, London, UK. Springer-Verlag, 417--426.

BONZANO, A., CUNNINGHAM, P. & SMYTH, B. 1997b. Using introspective learning to improve retrieval in CBR: A case study in air traffic control. *In:* LEAKE, D. &

PLAZA, E. (eds.) *Case-Based Reasoning Research and Development.* Springer Berlin / Heidelberg.

BRULAND, T., AAMODT, A. & LANGSETH, H. Architectures integrating case-based reasoning and bayesian networks for clinical decision support. *In:* SHI, Z., VADERA, S., AAMODT, A. & LEAKE, D., eds. 6th IFIP TC 12 International Conference, IIP 2010, Manchester, UK, October 13-16, 2010, 2010. Springer Boston, 82-91.

BRÜNINGHAUS, S. & ASHLEY, K. D. Evaluation of textual CBR approaches AAAI (Association for the Advancement of Artificial Intelligence) Workshop on Textual Case-based Reasoning, 1998. AAAI Press, 30-34.

BURKE, E., PETROVIC, S. & QU, R. 2006. Case-based heuristic selection for timetabling problems. *Journal of Scheduling,* 9, 115-132.

BURKE, E. K. & KENDALL, G. (eds.) 2005. *Search methodologies: Introductory tutorials in optimization and decision support techniques*: Springer.

BURNET, N. G., THOMAS, S. J., BURTON, K. E. & JEFFERIES, S. J. 2004. Defining the tumour and target volumes for radiotherapy. *Cancer Imaging,* 4, 153-161.

CANCER_RESEARCH_UK. 2010. *Radiotherapy* [Online]. Cancer Research UK. Available: http://cancerhelp.cancerresearchuk.org.

CHALMERS, D. J., FRENCH, R. M. & HOFSTADTER, D. R. 1992. High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental and Theoretical Artificial Intelligence,* 4, 185--211.

CHAMPIN, P.-A. & SOLNON, C. 2003. Measuring the similarity of labeled graphs. *Proceedings of the 5th international conference on Case-based reasoning: Research and Development.* Trondheim, Norway: Springer-Verlag.

CHANG, P. C., LIN, J. J. & DZAN, W. Y. 2012. Forecasting of manufacturing cost in mobile phone products by case-based reasoning and artificial neural network models. *Journal of Intelligent Manufacturing,* 23, 517-531.

CHANYAVANICH, V., DAS, S. K., LEE, W. R. & LO, J. Y. 2011. Knowledge-based IMRT treatment planning for prostate cancer. *Medical physics,* 38, 2515.

CHEETHAM, W. & PRICE, J. 2004. Measures of solution accuracy in case-based reasoning systems. *In:* FUNK, P. & GONZÁLEZ CALERO, P. A. (eds.) *7th european conference, eccbr 2004, madrid, spain, august 30 - september 2, 2004.*: Springer Berlin / Heidelberg.

COST, S. & SALZBERG, S. A weighted nearest neighbor algorithm for learning with symbolic features. Machine Learning, 1993. 57--78.

COVER, T. M. & HART, P. E. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory,* 13, 21--27.

COZZI, L., FOGLIATA, A., BOLSI, A., NICOLINI, G. & BERNIER, J. 2004. Three-dimensional conformal vs. Intensity-modulated radiotherapy in head-and-neck cancer patients: Comparative analysis of dosimetric and technical parameters. *International Journal of Radiation Oncology*Biology*Physics,* 58, 617-624.

CUNNINGHAM, P. 2009. A taxonomy of similarity mechanisms for case-based reasoning. *IEEE Transactions on Knowledge and Data Engineering,* 21, 1532--1543.

D'AQUIN, M., LIEBER, J. & NAPOLI, A. 2006. Adaptation knowledge acquisition: A case study for case-based decision support in oncology. *Computational Intelligence,* 22, 161-176.

DANTZIG, G. B. & THAPA, M. N. 1997. *Linear programming: 1: Introduction,* Springer Verlag.

DANTZIG, G. B. & THAPA, M. N. 2003. *Linear programming: 2: Theory and extensions: ,* Springer Verlag.

DAS, S. 2001. Filters, wrappers and a boosting-based hybrid for feature selection. *Proceedings of the Eighteenth International Conference on Machine Learning.* Morgan Kaufmann Publishers Inc.

DASH, M., LIU, H. & MOTODA, H. Consistency based feature selection. Proceedings of the Fourth Pacific Asia Conference on Knowledge Discovery and Data Mining, 2000. Springer-Verlag, 98-109.

DEASY, J. O., BLANCO, A. I. & CLARK, V. H. 2003. CERR: A computational environment for radiotherapy research. *Medical physics,* 30.

DIAMANTAOPOULIS, A. & SCHLEGELMILCH, B. 2002. *Taking the fear out of data analysis,* London, Thomson.

DIEREN, E., NOWAK, P., WIJERS, O., KOSTE, J., EST, H., BINNEKAMP, D., HEIJMEN, B. & LEVENDAG, P. 2000. Beam intensity modulation using tissue compensators or dynamic multileaf collimation in the three dimensional conformal radiotherapy of primary cancers of the oropharnx and larynx, including the elective neck. *International Journal of Oncology Biology and Physics,* 47, 1299-1309.

DOGAN, S. Z., ARDITI, D. & GÜNAYDIN, H. M. 2008. Using decision trees for determining attribute weights in a case-based model of early cost prediction. *Journal of Construction Engineering and Management,* 134, 146--152.

DONDERS, A. R. T., VAN DER HEIJDEN, G. J. M. G., STIJNEN, T. & MOONS, K. G. M. 2006. Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology,* 59, 1087-1091.

EHRGOTT, M., GÜLER, Ç., HAMACHER, H. & SHAO, L. 2010. Mathematical optimization in intensity modulated radiation therapy. *Annals of Operations Research,* 175, 309-365.

FARHANGFAR, A., KURGAN, L. A. & PEDRYCZ, W. 2007. A novel framework for imputation of missing values in databases. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans,* 37, 692-709.

FAZEL ZARANDI, M. H., RAZAEE, Z. S. & KARBASIAN, M. 2011. A fuzzy case based reasoning approach to value engineering. *Expert Systems with Applications,* 38, 9334-9339.

FLEURET, F. 2004. Fast binary feature selection with conditional mutual information. *Journal of Machine Learning Research,* 5, 1531-1555.

FRIZE, M. & WALKER, R. 2000. Clinical decision-support systems for intensive care units using case-based reasoning. *Medical Engineering and Physics,* 22, 671-677.

GIBBONS, J. P., MIHAILIDIS, D. N. & ALKHATIB, H. A. A novel method for treatment plan optimization. Proceedings of the 22nd annual EMBS international conference, July 2000 Chicago, IL.

GILCHRIST, J., TOWNSEND, D., ENNET, C., FRIZE, M. & BARICIAK, E. Discrimination of inconsistencies in medical data. MemeA - IEEE International Workshop on Medical Measurements and Applications, May 9-10 2008 Ottawa, Ontario, Canada. 87--92.

GILIO, J. P. 1998. Tabu search: An automatic treatment planning algorithm for conformal radiotherapy. *Medical physics,* 25, 796.

GONZALEZ, A. J., XU, L. & GUPTA, U. M. 1998. Validation techniques for case-based reasoning systems. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on,* 28, 465-477.

GOOD, D. 2012. *Inter-institution application of knowledge-based imrt treatment planning.* MSc, Duke University.

GUYON, I. & ELISSEEFF, A. E. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research,* 3, 1157--1182.

HA, S. H., JIN, J. S. & YANG, J. W. 2008. Predictive performance of clustered feature-weighting case-based reasoning. *Proceedings of the 4th international conference on Advanced Data Mining and Applications.* Chengdu, China: Springer-Verlag.

HAAS, O. C. L., BUMHAM, K. J. & MILLS, J. A. Hybrid optimisation technique for radiotherapy treatment planning. Proceedings of the 1998 IEEE International Conference on Control Application, September 1998 Trieste, Italy.

HAMACHER, H. W. & KUEFER, K. H. 2002. Inverse radiation therapy planning - a multiple objective optimization approach. *Discrete Applied Mathematics,* 118, 145-- 161.

HANCHUAN, P., FUHUI, L. & DING, C. 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 27, 1226-1238.

HANNEY, K. & KEANE, M. T. 1996. Learning adaptation rules from a case-base. *Proceedings of the Third European Workshop on Advances in Case-Based Reasoning.* Springer-Verlag.

HAOUCHINE, M.-K., CHEBEL-MORELLO, B. & ZERHOU, N. 2007. Case base maintenance approach. *International Conference on Industrial Engineering and Systems Management.* Beijing, China.

HATON, M. C. Knowledge-based decision making for radiotherapy planning. *In:* MORUCCI, J. P., PLONSEY, R., COATRIEUX, J. L. & LAXMINARAYAN, S., eds. Proceedings of the Fourteenth Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1992.

HEIDER, R. Troubleshooting CFM 56-3 engines for the boeing 737 using CBR and data-mining. *In:* SMITH, I. & FALTINGS, B., eds. Advances in Case-based Reasoning, 1996. Springer, 513.

HOLDER, A. 2004. Operations research and health care: A handbook of methods and applications. *In:* BRANDEAU, M. L., SAINFORT, F. & PIERSKALLA, W. P. (eds.). Kluwer Academic Publishers.

HOLT, A., BICHINDARITZ, I., SCHMIDT, R. & PERNER, P. 2006. Medical applications in case-based reasoning. *The Knowledge Engineering Review,* 20, 289—292.

HONG, T.-P. & LEE, C.-Y. 1996. Induction of fuzzy rules and membership functions from training examples. *Fuzzy Sets and Systems,* 84, 33-47.

HOWE, N. & CARDIE, C. 1997. Examining locally varying weights for nearest neighbour algorithms. *In:* LEAKE, D. & PLAZA, E. (eds.) *Case-Based Reasoning Research and Development: Second International Conference on Case-Based Reasoning, Lecture Notes in Artificial Intelligence.*

HU, M., SALVUCCI, S. M. & COHEN, M. P. Evaluation of some popular imputation algorithms. American Statistical Association, 2000.

HÜLLERMEIER, E., DUBOIS, D. & PRADE, H. 1999. Fuzzy rules in case-based reasoning. *Conf. AFIA99 Raisonnement à Partir de Cas.*

IONESCU-FARCA, F. & WILLI, A. Radoncol: An expert system for treatment advice in radiation oncology. Annual Meeting of the Swiss Society of Radiation Biology and Medical Physics, 1991 Kerzers, Switzerland. Max huber Verlag, 29-36.

ISHIBUCHI, H. & YAMAMOTO, T. 2004. Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets and Systems,* 141, 59-88.

JACZYNSKI, M. & TROUSSE, B. Fuzzy logic for the retrieval step of a case-based reasoner. Proceedings of the European Workshop on Case-Based Reasoning'94, 1994. 313–321.

JAGANNATHAN, R., PETROVIC, S., MCKENNA, A. & NEWTON, L. 2012. A novel two phase retrieval mechanism for a clinical case-based reasoning systems for radiotherapy treatment planning *International Journal on Artificial Intelligence Tools,* 21, 1240017.

JAIN, A. K. 2010. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters,* 31, 651-666.

JAIN, N. L. & KAHN, M. G. Clinical decision-support systems in radiation therapy. First international Symposium on 3D Radiation Treatment Planning and Conformal Therapy, April 1993 St Louis, USA.

JARMULAK, J., CRAW, S. & ROWE, R. Self-optimising CBR retrieval. Proceedings 12th IEEE International Conference on Tools with Artificial Intelligence, 2000. 376-383.

JI, S., PARK, M. & LEE, H. 2012. Case adaptation method of case-based reasoning for construction cost estimation in korea. *Journal of Construction Engineering and Management,* 138, 43-52.

JIANG, L., CAI, Z., WANG, D. & JIANG, S. 2007. Survey of improving k-nearest-neighbor for classification. *Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery - Volume 01.* IEEE Computer Society.

JOHN, G. H., KOHAVI, R. & PFLEGER, K. Irrelevant features and the subset selection problem. Machine Learning: Proceedings of the Elevent International Conference, 1994.

KALET, I. & PALUSZYNSKI, W. 1990. Knowledge-based computer systems for radiotherapy planning. *Americal Journal of Clinical Oncology,* 13(4), 344-351.

KAWAMOTO, K., FIOL, G. D., ORTON, C. & LOBACH, D. F. 2010. System-agnostic clinical decision support services: Benefits and challenges for scalable decision support. *The Open Medical Informatics Journal.*

KILR, G. & YUAN, B. 1998. Basic concepts and history of fuzzy set theory and fuzzy logic. *In:* RUSPINI, E. H., BONISSONE, P. P. & PEDRYCZ, W. (eds.) *Handbook of fuzzy computation.* Institute of Physics.

KIRA, K. & RENDELL, L. A. 1992. A practical approach to feature selection. *Proceedings of the ninth international workshop on Machine learning.* Aberdeen, Scotland, United Kingdom: Morgan Kaufmann Publishers Inc.

KIRKPATRICK, S., GELATT, C. D. & VECCHI, M. P. 1983. Optimization by simulated annealing. *Science,* 220, 671-680.

KIRSOPP, C., SHEPPERD, M. & HART, J. Search heuristics, case-based reasoning and software project effort prediction. Proceedings of genetic and evolutionary computation conference, GECCO 2002, 2002.

KNOWLES, J. D. & CORNE, D. W. 2000. Evolving neural networks for cancer radiotherapy. *In:* CHAMBERS, L. (ed.) *The practical handbook of genetic algorithms applications.* Chapman & Hall/CRC.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection". Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1995. Morgan Kaufmann.

KOHAVI, R. & JOHN, G. H. 1997. Wrappers for feature subset selection. *Artificial Intelligence,* 97, 273-324.

KOHAVI, R., LANGLRY, R. & YUN, Y. The utility of feature weighting in nearest neighbor algorithms. European Conference on Machine Learning, ECML'97, 1997.

KOLODNER, J. 1983. Reconstructive memory:A computer model. *COGNITIVE SCIENCE,* 7, 281-328.

KOLODNER, J. 1993. *Case-based reasoning,* San Francisco, Morgan Kaufmann.

KOLODNER, J. L. 1992. An introduction to case-based reasoning. *Artificial Intelligence Review,* 6, 3--34.

KOTON, P. Reasoning about evidence in causal explanations. Proceedings of AAAI (Association for the Advancement of Artificial Intelligence)-1988, 1988. Morgan Kaufmann.

KUMAR, K. A., SINGH, Y. & SANYAL, S. 2009. Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in icu. *Expert Systems with Applications,* 36, 65-71.

KWONG, C. K., SMITH, G. F. & LAU, W. S. 1997. Application of case based reasoning in injection moulding. *Journal of Materials Processing Technology*, 63, 463--467.

LANGER, M., BROWN, R., MORRILL, S., LANE, R. & LEE, O. 1996. A generic genetic algorithm for generating beam weights. *Medical physics*, 23, 965-971.

LANGLOTZ, C., FAGAN, L. M., TU, S. W., WILLIAMS, J. & SIKIC, B. Onyx: An architecture for planning in uncertain environments. International Joint Conference on Artificial Intelligence, 1985. 447-449.

LAW, M. & LIU, B. 2009. Informatics in radiology: Dicom-RT and its utilization in radiation therapy. *Radiographics*, 3, 655--67.

LAWANNA, A. & DAENGDEJ, J. 2010. Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems. *International Journal of Computer and Information Engineering*, 4, 10-18.

LEAKE, D. 1996. CBR in context: The present and future. *In:* LEAKE, D. (ed.) *Case-based reasoning: Experiences, lessons, and future directions.* Menlo Park: AAAI Press/MIT Press.

LEAKE, D. B., KINLEY, A. & WILSON, D. Acquiring case adaptation knowledge: A hybrid approach. Proceedings of the Thirteenth National Conference on Artificial Intelligence, 1996. AAAI Press, 684--689.

LEAKE, D. B. & KOLODNER, J. 1996. A tutorial introduction to case-based reasoning. *In:* LEAKE, D. B. (ed.) *Case-based reasoning: Experiences, lessons and future directions.* AAAI Press / MIT Press.

LEI, J. & LI, Y. A DNA genetic algorithm for beam angle selection in radiotherapy planning. IEEE International Conference on Cybernetics and Intelligent Systems, 2008.

LENZ, M. 1999. Experiences from deploying cbr applications in electronic commerce. *Proceedings of the German Workshop on CBR.*

LEVITT, S. H. 2008. *Technical basis of radiation therapy: Practical clinical applications,* Springer.

LI, H., LI, X., HU, D., HAO, T., WENYIN, L. & CHEN, X. 2009a. Adaptation rule learning for case-based reasoning. *Concurrency and Computation: Practice and Experience,* 21, 673-689.

LI, H., SUN, J., WU, J. & WU, X. J. 2012. Supply chain trust diagnosis (SCTD) using inductive case-based reasoning ensemble (icbre): The case of general competence trust diagnosis. *Applied Soft Computing,* 12, 2312 - 2321.

LI, Y., YAO, D., CHEN, W., ZHENG, J. & YAO, J. Ant colony system for the beam angle optimization problem in radiotherapy planning: A preliminary study. The 2005 IEEE Congress on Evolutionary Computation, 2005.

LI, Y., YAO, J. & YAO, D. 2003. Genetic algorithm based deliverable segments optimization for static intensity-modulated radiotherapy. *Phys. Med. Biol,* 48, 3353–3374.

LI, Y. F., XIE, M. & GOH, T. N. 2009b. A study of mutual information based feature selection for case based reasoning in software cost estimation. *Expert Systems with Applications,* 36, 5921—5931.

LIAO, T. W. & ZHANG, Z. 1998. Similarity measures for retrieval in case-based reasoning systems. *Applied Artifical Intelligence,* 12, 267-288.

LIAO, T. W., ZHANG, Z. M. & MOUNT, C. R. 2000. A case-based reasoning system for identifying failure mechanisms. *Engineering Applications of Artificial Intelligence,* 13, 199--233.

LIM, C. J., AHMAD, S. U., LEYMAN, R. & ER, M. H. A multi-modal approach to radiotherapy treatment planning optimization. Proceedings of the 22nd Annual EMBS International Conference, July 2000 Chicago IL.

LUIGI, G. 2009a. Myrobustcrust.M. MathWorks.

LUIGI, G. 2009b. Surfacevolume.M. MathWorks.

MAIN, J., DILLON, T. S. & SHUI, S. C. K. 2000. *Soft computing in case based reasoning,* Springer Verlag, London.

MALDONADO, S. & WEBER, R. 2009. A wrapper method for feature selection using support vector machines. *Information Sciences,* 179, 2208-2217.

MANTARAS, R. L. D., MCSHERRY, D., BRIDGE, D., LEAKE, D., SMYTH, B., CRAW, S., FALTINGS, B., MAHER, M. L., COX, M. T., FORBUS, K., KEANE, M., AAMODT, A. & WATSON, I. 2005. Retrieval, reuse, revision and retention in case-based reasoning. *Knowledge Engineering Review,* 20, 215-240.

MCSHERRY, D. Precision and recall in interactive case-based reasoning. *In:* SPRINGER, V., ed. ICCBR: Case-Based Reasoning Research and Development, 2001. 392--406.

MCSHERRY, D. 2005. Explanation in recommender systems. *Artificial Intelligence Review,* 24, 179-197.

MEDCOM. 2012. *Medcom - prosoma* [Online]. Available: http://www.medcom-online.de/applications/cancer/prosoma/prosoma.htm.

MERKLE, D. & MIDDENDORF, M. 2005. Swarm intelligence. *In:* BURKE, E. K. & KENDALL, G. (eds.) *Search methodologies: Introductory tutorials in optimization and decision support techniques.* Springer.

MESGHOUNI, N., GHEDIRA, K. & TEMANI, M. 2011. Unsupervised local and global weighting for feature selection. *Advances in Swarm Intelligence,* 283-290.

MEYER, J., HUMMEL, S. M., CHO, P. S., AUSTIN-SEYMOUR, M. M. & PHILLIPS, M. H. 2005. Automatic selection of non-coplanar beam directions for three-dimensional conformal radiotherapy. *The British Journal of Radiology,* 78, 316--327.

MISHRA, N. A case-based reasoning approach to radiotherapy treatment planning. 18th Triennial Conference of the International Federation of Operational Research Societies (IFORS), July 2008 Sandton, South Sandton, South Africa.

MISHRA, N., PETROVIC, S. & SUNDAR, S. A novel case based reasoning approach to radiotherapy planning. Proceedings of the 18th Triennial Conference of the International Federation of Operational Research Societies (IFORS), 13-18 July 2008 2008 Sandton, South Africa. 66.

MISHRA, N., PETROVIC, S. & SUNDAR, S. A knowledge-light nonlinear case-based reasoning approach to radiotherapy planning. Proceedings of the 2009 21st IEEE International Conference on Tools with Artificial Intelligence, 2009.

MORRILL, S. M., LANE, R. G., JACOBSON, G. & ROSEN, I. 1991. Treatment planning optimization using constrained simulated annealing. *Physics in Medicine and Biology,* 36, 1341-1361.

MUNOZ-AVILA, H. & HUELLEN, J. 1996. Feature weighting by explaining case-based planning episodes. *Proceedings of the Third European Workshop on Advances in Case-Based Reasoning.* Springer-Verlag.

NARIAINEN, K., OJALA, A., RANTANEN, J. & SARANUMMI, N. 1987. Cartes—a knowledge system for decision support. In therapy design and quality assurance in cancer therapy. Technical Research Center of Finland.

NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION DICOM - digital imaging and communications in medicine.

NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION 1997. Digital imaging and communications in medicine (DICOM). *Supplement 11: Radiotherapy Objects.*

NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION 1999. Digital imaging and communications in medicine. *Supplement 29: Radiotherapy Treatment Records and Radiotherapy Media Extensions.*

NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION 2011a. Digital imaging and communications in medicine (DICOM). *Part 5: Data Structures and Encoding.* Virginia, USA: National Electrical Manufacturers Association,.

NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION 2011b. Digital imaging and communications in medicine (DICOM). *Part 6: Data Dictionary.* Virginia, USA: National Electrical Manufacturers Association.

NAUCK, D. & KRUSE, R. A fuzzy neural network learning fuzzy control rules and membership functions by fuzzy error backpropagation. IEEE International Conference on Neural Networks, 1993. 1022-1027.

NAZARETH, D., MALHOTRA, H., JONES, M., BAKHTIARI, M. & BRUNNER, S. 2009. *Optimization of beam angles for intensity modulated radiation therapy treatment planning using genetic algorithm on a distributed computing platform.*

NUCLETRON. 2011. *Nucletron - oncentra* [Online]. Available: http://www.nucletron.com/Oncentra/Pages/default.aspx.

NUNEZ, H., SANCHEZ-MARRE, M. & CORTES, U. 2003. Improving similarity assessment with entropy-based local weighting. *Proceedings of the 5th international conference on Case-based reasoning: Research and Development.* Trondheim, Norway: Springer-Verlag.

NUTTING, C., ROWBOTTOM, C., COSGROVE, V., HENK, J., DEARNALEY, D., ROBINSON, M., CONWAY, J. & WEBB, S. 2001. Optimisation of radiotherapy for carcinoma of the parotid gland: A comparison of conventional, three-dimensional conformal, and intensity-modulated techniques. *Radiotherapy and Oncology,* 60, 163-172.

OFFIS 2009. DICOM tool kit DCMTK. 3.5.4 ed. Oldenburg, Germany: Offis Computer Science Institute.

OLDHAM, M., KHOO, V. S., ROWBOTTOM, C. G., BEDFORD, J. L. & WEBB, S. 1998. A case study comparing the relative benefit of optimizing beam weights, wedge angles, beam orientations and tomotherapy in stereotactic radiotherapy of the brain. *Physics in Medicine and Biology,* 43, 2123--2146.

OLDHAM, M., NEAL, A. & WEBB, S. 1995. A comparison of conventional 'forward planning' with inverse planning for 3d conformal radiotherapy of the prostate. *Radiotherapy and Oncology,* 35, 248-262.

PARK, C.-S. & HAN, I. 2002. A case-based reasoning system with the feature weights derived by analytic hierarchy process for bankruptcy prediction. *Expert Systems with Applications,* 23(1), 255-264.

PARK, J. H., IM, K. H., SHIN, C.-K. & PARK, S. C. 2004. MBNR: Case-based reasoning with local feature weighting by neural network. *Applied Intelligence,* 21, 265-276.

PARLIAMENT, M. B., SCRIMGER, R. A., ANDERSON, S. G., KURIEN, E. C., THOMPSON, H. K., FIELD, G. C. & HANSON, J. 2004. Preservation of oral health–related quality of life and salivary flow rates after inverse-planned intensity-modulated radiotherapy (IMRT) for head-and-neck cancer. *International Journal of Radiation Oncology, Biology, Physics,* 58, 663-673.

PEDRYCZ, W. 1994. Why triangular membership functions? *Fuzzy Sets and Systems,* 64, 21-30.

PEI, X., CAO, R., LIU, H., JIN, C., CHENG, M., ZHENG, H. & WU, Y. 2011. Beam orientation optimization using ant colony optimization in intensity modulated radiation therapy. *World Academy of Science, Engineering and Technology.*

PERNER, P. 1999. An architecture for a CBR image segmentation system. *Engineering Applications of Artificial Intelligence* 12, 749-759.

PETROVIC, S., MISHRA, N. & SUNDAR, S. 2011. A novel case based reasoning approach to radiotherapy planning. *Expert Systems with Applications,* 38, 10759-10769.

PHUONG, N. H., THANG, V. V. & HIROTA, K. 2000. Case based reasoning for medical diagnosis using fuzzy set theory. *Biomedical Fuzzy and Human Sciences,* 5, 1--8.

PRENTZAS, J. & HATZILYGEROUDIS, I. 2007. Categorizing approaches combining rule-based and case-based reasoning. *Expert Systems,* 24, 97-122.

PUGACHEV, A., LI, J. G., BOYER, A. L., HANCOCK, S. L., LE, Q.-T., DONALDSON, S. S. & XING, L. 2001. Role of beam orientation optimization in intensity modulated radiation therapy. *International Journal Radiation Oncology Biology Physics,* 50, 551-560.

QIN, Y., S, Z., ZHU, X., ZHANG, J. & SHANG, C. 2009. Pop algorithm: A kernel based imputation to treat missing values in knowledge discovery from databases. *Expert Systems with Applications,* 36, 2794--2804.

QUINLAN, J. R. 1986. Induction of decision trees. *Machine Learning,* 1, 81-106.

RECIO-GARCÍA, J. A. 2008. *Jcolibri: A multi-level platform for building and generating case-based reasoning systems.*

REDPATH, A. T., VICKERY, B. L. & WRIGHT, D. H. 1976. A new technique for radiotherapy planning using quadratic programming. *Physics in Medicine and Biology,* 21, 781-791.

REEMTSEN, R. & ALBER, M. 2009. Continuous optimization of beamlet intensities for intensity modulated photon and proton radiotherapy. *Handbook of Optimization in Medicine.* Springer US.

RICCI, F. & AVESANI, P. 1995. Learning a local similarity metric for case-based reasoning. *Proceedings of the First International Conference on Case-based Reasoning, ICCBR-95.* Sesimbra, Portugal.

RICHTER, M. M. Classification and learning of similarity measures. Jahrestagung der Gesellschaft fur Klassifikation, 1992. Springer Verlag, 1--8.

RISSLAND, E. L., ASHLEY, K. D. & BRANTING, L. K. 2005. Case-based reasoning and law. *The Knowledge Engineering Review,* 20, 293-298.

ROCHA, H., DIAS, J., FERREIRA, B. & DO CARMO LOPES, M. 2012. Incorporating radial basis functions in pattern search methods: Application to beam angle optimization in radiotherapy treatment planning. *Computational Science and Its Applications–ICCSA 2012,* 1-16.

ROMEIJN, H. E., AHUJA, R. K., DEMPSEY, J. F. & KUMAR, A. 2006. A new linear programming approach to radiation therapy treatment planning problems. *Operations Research,* 54, 201–216.

ROSEN, I. I., LAM, K. S., LANE, R. G., LANGER, M. & MORRIL, S. M. 1995. Comparison of simulated annealing algorithms for conformal therapy treatment planning. *International Journal of Oncology Biology and Physics,* 33, 1091-1099.

RUBIN, D. 1976. Inference and missing data. *Biometrika,* 63, 581--592.

RUBIN, D. B. 1996. Multiple imputation after 18+ years. *Journal of the American Statistical Association,* 91, 473-489.

RUI, X. & WUNSCH, D., II 2005. Survey of clustering algorithms. *Neural Networks, IEEE Transactions on,* 16, 645-678.

SAEYS, Y., INZA, I. & LARRAÑAGA, P. 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics,* 23, 2507-2517.

SAN PEDRO, J., BURSTEIN, F. & SHARP, A. 2005. A case-based fuzzy multicriteria decision support model for tropical cyclone forecasting. *European Journal of Operational Research,* 160, 308-324.

SCHANK, R. C. & ABELSON, R. P. 1977. *Scripts, plans, goals and understanding: Inquiry into human knowledge structures (artificial intelligence)* Hillsdale, New Jersey LAWRENCE ERLBAUM ASSOCIATES.

SCHEFFER, J. 2002. Dealing with missing data. *Research Letters in the Information and Mathematical Sciences,* 3, 153--160.

SCHLAEFER, A. & DIETERICH, S. 2011. Feasibility of case-based beam generation for robotic radiosurgery. *Artificial Intelligence in Medicine,* 52, 67-75.

SCHREIBMANN, E., LAHANAS, M., URICCHIO, R., THEODOROU, K., KAPPAS, C. & BALTAS, D. 2003. A geometry based optimization algorithm for conformal exterman beam radiotherapy. *Physics in Medicine and Biology,* 48, 1825--1841.

SHEPARD, D. M., FERRIS, M. C., OLIVERA, G. H. & MACKIE, T. R. 1999. Optimizing the delivery of radiation therapy to cancer patients. *SIAM Review,* 41, 721-744.

SHOKOUHI, S. V., SKALLE, P. & AAMODT, A. 2011. An overview of case-based reasoning applications in drilling engineering. *Artificial Intelligence Review,* 1-13.

SMYTH, B. & KEANE, M. Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems. Proceedings of the 13th International Joint Conference on Artificial Intelligence, 1995. 377-382.

SMYTH, B. & KEANE, M. T. 1998. Adaptation-guided retrieval: Questioning the similarity assumption in reasoning. *Artificial Intelligence,* 102, 249--293.

SONDERMAN, D. & ABRAHAMSON, P. G. 1985. Radiotherapy treatment design using mathematical programming models. *Operations Research,* 33, 705-725.

SONG, Q. & SHEPPERD, M. 2007. A new imputation method for small software project data sets. *The Journal of Systems and Software,* 80, 51--62.

SONG, X., PETROVIC, S. & SUNDAR, S. A case-based reasoning approach to dose planning in radiotherapy. The seventh international conference on case-based reasoning, ICCBR 07, Workshop Proceedings, August 2007 Belfast, Ireland. 348-357.

SØRMO, F., CASSENS, J. & AAMODT, A. 2005. Explanation in case-based reasoning–perspectives and goals. *Artificial Intelligence Review,* 24, 109-143.

STAHL, A. Learning similarity measures: A formal view based on a generalized CBR model. Proceedings of the 6th International Conference on Case-Based Reasoning (ICCBR 2005), 2005 Chicago, USA. Springer.

STANFILL, C. & WALTZ, D. 1986. Toward memory-based reasoning. *Communications of the ACM,* 29, 1213--1228.

TORKKOLA, K. 2003. Feature extraction by non-parametric mutual information maximization. *Journal of Machine Learning Research,* 3, 1415-1438.

TRILOGY. 2008. *Http://www.Peninsula.Org/images/cancer center* [Online].

TVERSKY, A. 1977. Features of similarity. *Psychological Review*, 84, 327-352.

VINEBERG, K. A., EISBRUCH, A., COSELMON, M. M., MCSHAN, D. L., KESSLER, M. L. & FRAASS, B. A. 2002. Is uniform target dose possible in imrt plans in the head and neck? *International Journal of Radiation Oncology Biology and Physics*, 52, 1159-1172.

VOROBIEVA, O., RUMYANTSEV, A. & SCHMIDT, R. A CBR solution for missing medical data. 5th Workshop on CBR in the Health Sciences, 2007.

WASITO, I. & MIRKIN, B. 2006. Nearest neighbours in least-squares data imputation algorithms with different missing patterns. *Computation Statistics and Data Analysis*, 50, 926--949.

WATSON, I. 1994. Case-based reasoning: A review. *Knowledge Engineering Review*, 9.

WEBB, S. 1989. Optimisation of conformal radiotherapy dose distributions by simulated annealing. *Physics in medicine and biology*, 34, 1349--1370.

WEBB, S. 2001. *Intensity-modulated radiation therapy* C. G. Orton, J. A. E. Spaan, J. G Webster, Taylor and Francis.

WEBB, S. 2005. Optimizing radiation therapy inverse treatment planning using the simulated annealing technique. *International Journal of Imaging Systems and Technology*, 6, 71--79.

WEBER-LEE, R., BARCIA, R. M. & KHATOR, S. K. Case-based reasoning for cash flow forecasting using fuzzy retrieval. Case-Based Reasoning Research and Development: Proceedings of the 1st International Conference on Case-Based Reasoning, 2006. 510--519.

WESS, S., ALTHOFF, K.-D. & DERWAND, G. 1994. Using k-d trees to improve the retrieval step in case-based reasoning. *Selected papers from the First European Workshop on Topics in Case-Based Reasoning*. Springer-Verlag.

WETTSCHERECK, D. & AHA, D. 1995. Weighting features. *In:* VELOSO, M. & AAMODT, A. (eds.) *Case-based reasoning research and development.* Springer Berlin / Heidelberg.

WETTSCHERECK, D., AHA, D. W. & MOHRI, T. 1997. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review,* 11, 273-314.

WILKE, W. & BERGMANN, R. 1998. Techniques and knowledge used for adaptation during case-based problem solving. *Proceedings of the 11th International Conference on Industrial and Engineering Applications of Artificial In telligence and Expert Systems: Tasks and Methods in Applied Artificial Intelligence.* Springer-Verlag.

WILKENS, J., ALALY, J., ZAKARIAN, K., THORSRAD, W. L. & O'DEASY, J. 2007. IMRT treatment planning based on prioritizing prescription goals. *Physics in medicine and biology,* 52, 1675-1692.

WU, X., ZHU, Y., DAI, J. & WANG, Z. 2000. Selection and determination of beam weights based on genetic algorithms for conformal radiotherapy treatment planning. *Physics in Medicine and Biology,* 45, 2547--2558.

WYNER, A. & BENCH-CAPON, T. 2007. Argument schemes for legal case-based reasoning. *Proceedings of the 2007 conference on Legal Knowledge and Information Systems: JURIX 2007: The Twentieth Annual Conference.* IOS Press.

Y. LI, YAO, D., YAO, J. & CHEN, W. 2005. A particle swarm optimization algorithm for beam angle selection in intensity-modulated radiotherapy planning. *Physics in Medicine and Biology,* 50, 3491–3514.

YONGJIE, L. & JIE, L. 2010. A feasible solution to the beam-angle-optimization problem in radiotherapy planning with a DNA-based genetic algorithm. *IEEE Transactions on Biomedical Engineering,* 57, 499-508.

YU, L. & LIU, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), 2003 Washington, DC, US.

ZADEH, L. 1965. Fuzzy sets. *Information and Control,* 8, 338–353.

ZHANG, H. H., MEYER, R. R., WU, J., NAQVI, S. A., SHI, L. & D D'SOUZA, W. 2010. A two-stage sequential linear programming approach to IMRT dose optimization. *Physics in medicine and biology,* 55, 883.