Approaches to Knowledge-Light Adaptation in Case-Based Reasoning for Radiotherapy Treatment Planning

Gulmira Khussainova, MSc

Thesis submitted to The University of Nottingham
for the degree of Doctor of Philosophy

January 2016

Abstract

In radiotherapy, ionised radiation beams are used to destroy cancerous cells. A radiotherapy treatment plan needs to be created to deliver a sufficient radiation dose to cancerous cells while sparing nearby organs at risk and healthy tissue. The development of such a treatment plan is a time consuming trial and error process which can take from a few hours up to a few days.

This thesis builds on the previously developed Case-Based Reasoning (CBR) system for radiotherapy treatment planning for brain cancer that was developed in collaboration with Nottingham University Hospitals NHS Trust, City Hospital Campus, UK. The original CBR system focused on the retrieval stage of CBR, where the most similar case was retrieved for the new patient case. The results obtained were promising but adaptation needed to be performed for them to be suitable for the new patient.

Testing of the CBR system by medical physicists has revealed that some of the retrieved radiation beams were not suitable for the tumour position of the new cases and thus could not be used. To avoid this the clustering of cases by their tumour positions was implemented to only retrieve cases with similar tumour positions. The revised CBR system should now retrieve treatment plans with better suited beams.

Adaptation requires a lot of domain knowledge which is often difficult to acquire. In this research we present adaptation approaches which are knowledge-light, i.e. they utilise knowledge available in the case base without requiring interaction with medical experts. Adaptation methods based on machine learning algorithms, in particular neural networks, the naive Bayes classifier, and support vector machines, were developed. Also, an adaptation-guided retrieval approach is presented, in which the case is retrieved only if it can be adapted. In addition, a pair of similar cases are retrieved with it, which guide the adaptation process. The developed knowledge-light adaptation methods have improved the results of the original CBR system. In addition, the proposed adaptation methods are general and could be used in domains where the available amount of knowledge is limited.

Acknowledgements

My sincere thanks go to my two supervisors, Professor Sanja Petrovic, and Dr. Rupa Jagannathan, for the support and guidance they have given to me over the last four years. I have learned a lot from both of you.

My thanks also go to Angela McKenna, Louise Newton and other staff at Nottingham University Hospitals NHS, City Hospital Campus, UK, for providing us with data, their expertise and their valuable time.

I would like also to thank the Business School, School of Computer Science at the University of Nottingham, and EPSRC (LANCS Science and Innovation Grant (EP/F033214/1)) for their support, as their funding made this project possible.

In addition, many thanks go to my fellow research students and people I met during my PhD who supported me in this path and from whom I learnt something new. Lastly I would like to thank my friends and family for their constant support and encouragement.

To my dearest mother and sister

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis, adaptation approaches in Case-Based Reasoning (CBR) for radiotherapy treatment planning (RTP) are presented. The research presented here builds on the decision-support system that used CBR for radiotherapy treatment planning of brain cancer which was developed in collaboration with the Nottingham University Hospitals Trust, NHS, Nottingham City Hospital Campus and is described in (Jagannathan and Petrovic, 2012) and (Jagannathan et al., 2012).

1.1 Background and General Motivation

Radiotherapy is a type of cancer treatment that uses ionising radiation beams to control cancer growth. A detailed treatment plan has to be created to determine exactly how the patient will be irradiated. A treatment plan includes the number of radiation beams, beam angles, dose fraction in each treatment session, beam wedges, and so on. RTP is usually a time-consuming process. Different software systems are used to facilitate the planning process, e.g. Helax and Oncentra. However, while following treatment protocol different hospitals have different approaches in creating a treatment plan. The development of the treatment plan is subject to the knowledge and experience of the medical experts. Experienced planners learn carefully from their past patient cases and in general, these treatment planning systems use complex mathematical models that neglect to capture the expertise and experience gained from the physicists.

The treatment plan at Nottingham City Hospital is produced in a trial and error process that requires the expertise of a group of medical experts including oncologists and medical physicists and can take from 2-3 hours to a few days. Currently, Oncentra planning software is used to facilitate the treatment plan development process. This system can generate the treatment plan parameters such as number of beams, beam angles and wedges for the new patient. However, the medical physicists use the software only for the visualisation of patient images (the tumour and surrounding organs at risk) and the radiation dose distribution. The reason for not using the treatment plan suggested by the system is that the medical physicists do not know the reasoning of the mechanism behind the created treatment plan that the software uses. Therefore, during meetings with the medical physicists they have commented that a decision support system where a reasoning mechanism could be followed up and understood easy would be especially useful.

CBR has been widely used for developing knowledge-based decision support systems in the medical domain. CBR is based on using experience gained when solving similar problems in the past to understand and solve new problems (Kolodner, 1992). The CBR process comprises of the steps of reasoning, which take into consideration all previous similar events and features that could help in solving a new problem. When creating a treatment plan for a new patient, medical staff try to remember/retrieve similar cases that they have encountered in the past along with their treatment plans and try to map them to the new case. Therefore, using CBR seems to be a natural decision in designing an application that will help medical staff make a treatment plan.

The general aim of the project is to develop a CBR system that uses knowledge of medical expertise and can be used in a clinical setting to assist medical physicists. Such decision support system would automate the radiotherapy treatment plan development process and save the valuable time of the medical physicists.

The CBR system was developed by Jagannathan et al. (2012) where given the new case, the system retrieved the most similar case from the previously developed stored treatment plans. Experiments showed promising results, but suggested that an adaptation of the retrieved solution needs to be performed to better address the requirements of the new case. The research presented in this thesis builds upon this work and investigates adaptation mechanisms to improve the results of the current retrieval system. Adaptation is a crucial phase of the CBR that fits the retrieved solution to the requirements of the new problem and determines the applicability of a CBR system in the real world.

1.2 Motivation and Aims of the Thesis

The aim of the research presented in this thesis is to investigate adaptation mechanisms and explore to what extent they can enhance the usefulness of the retrieved case for solving a new problem. The radiotherapy treatment planning is used as a domain and the aim is to perform adaptation of the retrieved solution of the previously developed CBR retrieval system to have a better suited treatment plan for the new case.

Literature review of the state-of-the-art of CBR shows that although CBR has been widely used in the last few decades in a variety of domains, the research into adaptation phase of CBR is limited. Early CBR applications mainly consisted of CBR retrieval only and adaptation, if needed, was left to the system users or proposed for the future work. The motivation for the work presented in this thesis is to add to the knowledge of the CBR adaptation.

1.3 Research Questions and Objectives of the Thesis

Adaptation methods that are present in the literature were reviewed. Adaptation approaches were considered with respect to their applicability to the research problem presented in this thesis.

The adaptation phase is knowledge-intensive, i.e. it requires a lot of domain knowledge. Most of the adaptation methods reported in the literature are knowledge-intensive. The knowledge required for adaptation is often obtained from the system users or experts in the relevant field. However, they are not always available, especially with the busy schedule of the medical experts which is the case in our research. Therefore, we look at the sources of knowledge needed for adaptation that are available to us in the domain and the case base. Hence, the adaptation methods presented in this thesis are knowledge-light, i.e. they utilise the domain knowledge available in the case base. We investigate to what extent the knowledge that is already present in the case base is sufficient for adaptation phase and how it can be utilised for adaptation purposes.

The main research questions addressed in this thesis can be briefly summarised as follows:

- Chapter 4

  - Which additional attributes can improve the retrieval system and how to include them in the retrieval phase of the CBR system?
  - To what extent can the clustering of cases which considers a new carefully selected attribute of cases improve the success rate of the CBR retrieval?

- Chapter 5

  - How useful could be the application of several machine learning-based adaptation methods, including neural networks (NNs), the naive Bayes classifier and support vector machines (SVM) to CBR adaptation?

- Chapter 6

  - How does the performance of adaptation-guided retrieval approaches compare to the performance of the CBR retrieval system with no adaptation?

All developed adaptation approaches are evaluated against the results obtained by the previously developed CBR system which contains only retrieval but no adaptation. If CBR adaptation phase improves on the result of the CBR retrieval phase the adaptation is deemed successful.

1.4 Publications and Presentations

The following peer reviewed academic papers have been produced as a result of this research:

- Sanja Petrovic, Gulmira Khussainova and Rupa Jagannathan. ”Knowledge-light adaptation approaches in case-based reasoning for radiotherapy treatment planning”, *Artificial intelligence in medicine*, Volume 68, pp.17-28, March 2016.

- Gulmira Khussainova, Sanja Petrovic and Rupa Jagannathan. ”Retrieval with Clustering in a Case-Based Reasoning System for Radiotherapy Treatment Planning”, Journal of Physics: Conference Series 616(1), 2015, *Improving Healthcare: new challenges, new approaches.* Coimbra, Portugal. May, 2015, paper presented

In addition, the work was presented at several workshops and conferences:

- Gulmira Khussainova, Sanja Petrovic and Rupa Jagannathan. ”Adaptation in Case-Based Reasoning for Radiotherapy Planning”, OR 55 Annual Conference, Exeter, UK. 05 September, 2013, presentation

- Gulmira Khussainova, Sanja Petrovic and Rupa Jagannathan. ”Adaptation in a case-based reasoning system for radiotherapy treatment planning for brain cancer”, 4th Student Conference on Operational Research (SCOR 2014). Nottingham, UK. 3 May 2014, presentation

- Gulmira Khussainova, Sanja Petrovic and Rupa Jagannathan. ”Knowledge-Light and Machine Learning-Based Adaptation for Case-Based Reasoning in Radiotherapy Treatment Planning”. ’Life Beyond PhD’ conference, Cumberland Lodge, Egham, UK. 27 August 2014, presentation and poster

- Gulmira Khussainova, Sanja Petrovic and Rupa Jagannathan. ”Knowledge-Light Adaptation Approaches in Case-Based Reasoning for Radiotherapy Treatment Planning”, 22nd International Conference on Case-based Reasoning (ICCBR) Doctoral Consortium, Cork, Ireland. 3 September 2014, presentation

1.5 Structure of the Thesis

The thesis contains seven chapters in total. The relevant literature review is presented in Chapter 2. The detailed description of the existing CBR system upon which the research here builds up is in Chapter 3. Chapter 4 describes the revised CBR retrieval system with the cases clustered according to their tumour position. Knowledge-light machine learning-based adaptation methods using the knowledge available in the case base are presented in Chapter 5. A knowledge-light adaptation-guided retrieval method is introduced in Chapter 6. The final conclusion and future work directions are provided in Chapter 7.

Chapter 2

Literature Review

2.1 Introduction

This chapter will provide a review of the relevant literature for the areas of work studied in this thesis. There are two main areas: case-based reasoning and radiotherapy treatment planning.

The first to be covered is an overview of the radiotherapy treatment planning (RTP) problem. Depending on the type of cancer tumour there are different radiotherapy treatment methods that can be used. This section will cover some of the widely used existing RTP delivery methods. The radiation treatment is delivered by radiation beams, therefore a short overview of beam configuration parameters is presented afterwards. RTP is a time consuming procedure; the approaches which were developed to facilitate this problem will be discussed.

Case-Based Reasoning (CBR) is presented next. CBR has been used in many domains and in this research we look into CBR applied particularly to the radiotherapy treatment planning problem. The structure of the CBR system will be given in detail, including the retrieval, adaptation and maintenance phases. The adaptation phase of CBR is an important part of our research, therefore a more comprehensive overview on it will be given. CBR applications in healthcare will be reviewed further. The relevance of the CBR approach to the RTP problem and the motivation behind this combination will be explained, followed by examples of the existing CBR systems in RTP.

2.2 Radiotherapy Treatment Planning

Radiotherapy is a type of treatment that uses ionising radiation beams to control cancer growth. Radiotherapy treatment planning aims at delivering a sufficient radiation dose to cancerous tumour cells while sparing or minimising damage to healthy organs in the tumour surrounding area. This type of treatment is based on the different abilities of healthy and cancerous cells to recover from radiation damage. The total radiation dose is delivered in a large number of small fractions referred to as fractionation of radiation, which allows healthy cells to recover between treatments. In ideal dose distribution the radiation dose to healthy tissue should be as low as possible, and the shape of the dose should be conformal to the tumour and avoid healthy surrounding tissue, in order to reduce radiation side effects. As there are many types of cancer, there are many RTP modalities, for example, three-dimensional (3D)-conformal radiotherapy, intensity-modulated radiation therapy, volumetric arc therapy, etc. In addition, different hospitals might have different treatment protocols which means that the type of RTP for the very same type of cancer can vary from one hospital to the next, even within one country.

During the RTP process medical images such as computed tomography (CT) and magnetic resonance imaging (MRI) are taken to detect the location of the tumour. Oncologists then determine the required radiation dose to treat the patient and pass the information to the medical physicists for a detailed treatment plan design, which includes determination of radiation beam number, beam angles, dose fraction in each treatment session, beam wedges, etc. A detailed treatment plan has to be created to determine how exactly the patient will be irradiated to achieve the required dose distribution.

RTP is usually a time consuming trial and error process and needs the expertise of one or more experienced medical physicists and oncologists. Treatment planning is done with the help of a planning software. The software helps medical physicists to visualise the tumour area and see the radiation distribution for a particular beam configuration. Even with the help of such a software the RTP process still requires a lot of 'manual' labour. Planning can take from a few hours up to 3-4 days, depending on the complexity of the plan. For example, Esch et al. (2002) review five radiotherapy units and report that the mean physicist's time typically needed for verification of the treatment plan per patient is 6 hours in a radiotherapy unit in Leuven, 6 hours in Berlin, 3 hours in Helsinki, 10 hours in Copenhagen and 40 hours in Reggio Emilia.

In Nottingham City Hospital a detailed RTP is created in the same trial and error fashion described above for each patient, to determine how exactly a tumour will be irradiated to achieve the required dose distribution. Developing a treatment plan is highly dependent on the expertise and experience of the medical staff. It requires the expertise of a group of medical physicists, which often involves back and forth communication with oncologists; the designed plan has to be approved by the oncologist, and if it is not, the plan is redesigned.

2.2.1 Linear Accelerator

Linear Accelerator is a device used to apply radiation beams to the tumour region by means of a gantry, which can be rotated around the patient. Radiation can be delivered to the tumour from any angle by rotating the gantry and moving the treatment couch. Figure 2.1 illustrates a linear accelerator moving around the couch. Radiation beams are shaped by a multileaf collimator incorporated into the head of the machine (see Figure 2.2). Figure 2.3 illustrates the multileaf collimator with leaves positioned to shape the beam. The patient lies on a moveable treatment couch and lasers are used to ensure patient is in the proper position. The treatment couch can move in different directions including up, down, right, left, in and out.

After the gantry of the linear accelerator is rotated and positioned at a selected angle, the multileaf collimator leaves move in and out to shape the beam in order to conform the radiation to the outlines of the tumour. Beam arrangement in radiotherapy treatment planning is crucial. Its configuration is dependant on the tumour size, shape and its location with respect to the organs at risk (OARs). The beam configuration should be such that most of the radiation dose is applied to the centre of the tumour with the OARs receiving as little of the dose as possible.

There are different attributes that characterise the beam:

Beam number - In 3D conformal radiotherapy treatment the number of beams used varies between 2-9 (Schreibmann et al., 2004). Hospital policy usually limits the number of beams to ease the planning procedure and to reduce the treatment time in each session and therefore, patient discomfort. The more beams are applied, the more time it takes to compute and deliver the treatment plan. In Intensity Modulated Radiotherapy Treatment (IMRT), each beam consists of many beamlets for each of which the intensity

FIGURE 2.1: Linear accelerator and a patient couch (St Rita's Medical Centre, 2015)

FIGURE 2.2: Beam emission from linear accelerator illustration (RAD planning, 2015)

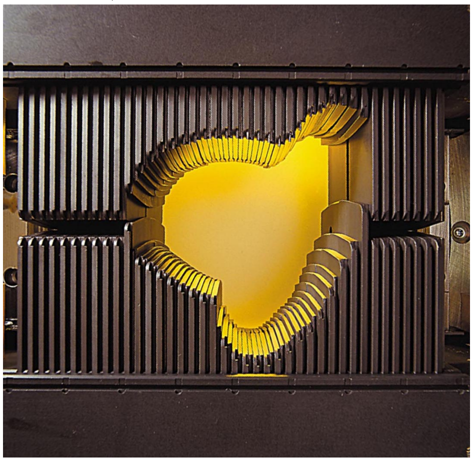

FIGURE 2.3: Multileaf collimator (Varian Medical Systems, 2016)

needs to be defined and that requires longer computational time.

Beam angle - The angle at which the radiation beam enters the head. Beam angles can be coplanar (lying in one plane) or non-coplanar.

Beam weight - Radiation intensity of the beam. Individual beams can have different weights. Contribution of each beam to the total radiation dose is determined by the beam weight.

Beam wedge - Wedges modify the shape of the beam. The radiation beam intensity decreases when a wedge filter is placed in the path of it. Metallic blocks have been the primary means of producing the wedged fields.

There are different types of linear accelerators and new more advanced ones can deliver different forms of RTP. The next section will give an overview of the most common presently used types of RTP delivery methods.

2.2.2 Types of Radiotherapy Application

Radiotherapy planning can use either a forward or an inverse planning approach. In inverse planning, the required dose distribution is specified and then a treatment plan is created, so that it fits the prescribed dose distribution, while in forward treatment planning the planning parameters are adjusted iteratively to achieve an acceptable radiation dose distribution. The literature on radiotherapy planning covers both aspects (Rocha et al., 2012b; Oldham et al., 1998; Liber and Bresson, 1999).

The choice of which radiotherapy treatment planning method to use depends on the cancer type and the facilities available at the treatment centre. At the Nottingham City Hospital, 3D conformal radiation therapy is used for brain cancer. 3D conformal radiation therapy is a technique where the radiation beams used in treatment are shaped to match the tumour. 3D conformal radiation therapy focuses the radiation precisely on the tumour, while avoiding the healthy surrounding tissue. It makes it possible to use higher levels of radiation in treatment, which are more effective in shrinking and killing tumours (University of Pittsburg Cancer Institute, 2015). It requires the use of 3D imaging methods and is typically accomplished with a set of pre-defined radiation beams shaped according to the target volume projection on the image. The radiation beams typically have uniform intensity across the field, or the intensity is modified by devices such as wedges or compensating filters (Schreibmann et al., 2004).

In IMRT, each radiation beam consists of many smaller beamlets, usually about 150, each of different intensity, formed by a multileaf collimator. Thus, it allows a range of different radiation intensities to be applied across the tumour. IMRT planning is usually divided into three smaller problems: the geometry problem, intensity problem, and realisation problem. The geometry problem consists of finding the minimum number of beams and corresponding angles that satisfy the treatment goals using optimisation algorithms. In practice, the number of beams and the beam angles are manually selected by the treatment planner based on his or her experience. After the number of beams is defined the intensity problem is solved, in which the optimal beamlet weights are identified. The outcome of the intensity problem is a set of optimised radiation intensity maps (one for each beam angle). After an acceptable set of intensity maps is produced, one must find a suitable method for the delivery (realisation problem). Sometimes the produced intensity maps cannot be implemented, as the multileaf collimator cannot be positioned to deliver it, thus they need to be revised (Rocha et al., 2012b). IMRT allows a larger volume of normal tissue to receive lower radiation doses (Elith et al., 2011), but this comes at a cost of increased treatment time.

Both 3D conformal and IMRT radiation therapies are widely used and reported in literature due to their improved conformity to the tumour and avoidance of nearby healthy OARs. However, they take a longer computational time, in particular when non-coplanar beams are used. Therefore, the use of the decision support systems for the RTP has been widely researched. There are some new radiotherapy treatment planning delivery methods which are gaining attention:

- Tomotherapy: is a treatment system that is a combination of a computed tomography scanner and a linear accelerator. The treatment dose is delivered when a patient is translated through the bore of the machine as its gantry continuously rotates. Tomotherapy is capable of producing high-quality plans that increasingly spare dose to the surrounding OARs (Elith et al., 2011).

- Volume Modulated Arc Therapy (VMAT): delivers the dose to the entire tumour in a single arc rotation. The treatment is delivered on a linear accelerator using a cone beam that continuously rotates around the patient. The cone beam is modulated by the dynamic multileaf collimator, variable dose rate and variable gantry speed to generate IMRT-quality dose distributions in a single optimised arc around the patient

(Elith et al., 2011).

- Image guided radiotherapy (IGRT): is a conformal radiation treatment guided by specialised imaging tests, such as CT scans, ultrasound or X-rays. Imaging of the patient before or during the treatment session improves the ability to concentrate the dose on the target volume. IGRT can lower complication rates and potentially escalate the tumour dose. The patient is imaged just before being treated, so that the process of applying radiation beams is made more precise. It is especially popular for moving organs, e.g. lung cancers, to take into account the path of the organ movement while breathing (Dawson and Sharpe, 2006). IGRT is often used in combination with other RTP types, as additional imaging helps to generate a more successful treatment plan delivery.

Tomotherapy and VMAT are rotating gantry IMRT techniques and their application can be found in the literature to a lesser extent compared to IMRT and 3D conformal approaches. These approaches are rapidly gaining favour in the medical community owing to decreased treatment time and thus, improvement in the patient's quality of life.

2.2.3 Existing Approaches to RTP

When determining suitable values for planning parameters such as the number of beams, beam angles, dose in different phases of treatment, number of wedges and their configuration, various methods can be used. This section will introduce in detail methods used to solve the RTP problem.

Optimisation methods

The literature on RTP offers a variety of methods, which can be classified as optimisation and knowledge-based methods. Optimisation methods developed for RTP include linear and non-linear programming, quadratic programming, and heuristic optimisation methods, such as simulated annealing, tabu search, genetic algorithm and ant colony optimisation.

There are various treatment planning methods available to facilitate RTP process, most of them being optimisation-based methods (Holder and Salter, 2004; Ferris et al., 2003; Morrill et al., 1991; Rocha et al., 2012b). They are highly specialised and designed to specifically solve a particular problem, i.e. typically only some of the treatment plan parameters are addressed in an optimisation function and the rest are added by the end user or targeted in a separate optimisation function (Oldham et al., 1998; Rocha et al., 2012b).

A number of papers provide a survey of optimisation methods for radiotherapy treatment (Oldham et al., 1998; Holder and Salter, 2004; Ehrgott et al., 2008). A survey of the use of mathematical optimisation in IMRT is given in (Ehrgott et al., 2008). Optimisation methods represent the problem in a mathematical formulation with a multiple-criteria objective function. In (Ehrgott and Johnston, 2003; Ehrgott and Winz, 2008) a multi-criteria integer programming model was demonstrated to solve some of the problems encountered in IMRT treatment planning such as the beam direction optimisation and dose trade-off optimisation. However, the set up of the multileaf collimator to deliver the optimised plan has not been implemented and was suggested for future work. Holder and Salter (2004) examined the shortcomings of linear, non-linear and quadratic models in optimisation for radiation oncology, including 3D conformal and IMRT. Linear programming optimisation models were reviewed to be generally infeasible in RTP, as the region defined by linear models was too small and did not provide significant improvements in the treatment quality. In addition, they could not find a solution that would satisfy all of the constraints imposed by medical physicists. After linear models the most prevalent seen in the RTP domain were quadratic optimisation models (Redpath et al., 1976; Hoegele et al., 2012). However, the models described suffer from the same infeasibility problem as linear models. Rocha et al. (2012a) reported that IMRT was commonly solved by using a range of optimisation techniques including mixed integer linear programming.

Heuristic approaches

Heuristic approaches produce a satisfactory solution in a reasonable amount of time. However, the solution might not be an optimal solution. Heuristic approaches are widely used where application problems are too complex or computationally expensive to find an optimal solution. Application of heuristic approaches to radiotherapy planning can be found in (Rocha et al., 2012b; Teodorovic et al., 2013). A binary probabilistic tabu search was developed for inverse treatment planning in IMRT (Rocha et al., 2012b). Ehrgott and Johnston (2003) used heuristics in optimisation of beam directions in IMRT planning. De Neve et al. (1996) made use of heuristic methods in optimisation of radiation dose for head and neck cancer radiotherapy treatment. The technique efficiently spared the spinal cord and allowed the delivery of higher target doses than with conventional techniques, but at the expense of extended execution times. Purdie et al. (2011) reported a successful application of heuristic methods in automated IMRT planning for breast cancer. In total, 87% of generated treatment plans were deemed to be equal to corresponding clinical plans.

Other drawbacks of optimisation approaches are the need for large computational efforts and the possibility of a poorly defined problem. The review of models (Legras et al., 1982; Morrill et al., 1991; Holder and Salter, 2004; Ferris et al., 2003) addressing different aspects of radiotherapy treatment design showed that in general optimisation methods do not sufficiently address the clinical issues that arise in treatment planning, as they fall short when compared to the design process conducted by a physician. These treatment planning systems use complex mathematical models that neglect to capture the expertise and experience gained from physicists. Experienced planners often know what configurations are suitable for a cancer patient even when the underlying theory is not thoroughly understood (Jagannathan et al., 2012; Reyes et al., 2015).

Knowledge-based methods

Knowledge-based methods aim at utilising knowledge available in the domain and used by the domain end user to simulate the natural problem solving process as much as possible. These methods do not use mathematical algorithms but make use of domain knowledge or past experience in determining the parameters of the treatment plan. Artificial intelligence methods such as rule-based reasoning, case-based reasoning and hierarchical organisation of knowledge have been employed to automate the radiotherapy planning process (Berger, 1994; Liber and Bresson, 1999; Song et al., 2007; Mishra et al., 2009; Petrovic et al., 2011).

One of the most common approaches is rule-based reasoning. Liber and Bresson (1999) presented CASIMIR RBR, which used rule-based reasoning in breast cancer diagnosis and treatment, but the obtained solution needed to be revised by the committee. Therefore, the system was extended into CASIMIR RBR/CBR with the addition of CBR that simulated the reasoning of the medical committee. The drawback of rules-based reasoning is that a lot of rules are needed to generate a problem solver. In addition, in some complex domains, including RTP, the established protocols to form the rules are not available, therefore this approach cannot be used. CBR overcomes these problems. However, CBR has not been widely used in radiotherapy planning. A detailed overview of the state-of-the-art of CBR applications in the RTP problem will be presented in Section 2.3.

Hybrid approaches

There are also systems where optimisation methods and knowledge-based methods are combined to enhance each other's performance; such a system was developed by Teodorovic et al. (2013), in which CBR was enhanced with bee colony optimisation to determine the thyroid cancer radioiodine dose. Bee colony optimisation was used to determine attribute weights in the similarity measure of the CBR system. There are also examples of the use of CBR and machine-learning tools in the hybrid systems for other domains in healthcare. In particular, Biswas et al. (2014) used CBR and neural networks for the feature weighting mechanism in the retrieval phase for classification purposes. The CBR system was validated on various clinical and non-clinical datasets and it was claimed that the two approaches complement each other, resulting in superior accuracy. In this research we also make use of machine learning tools in conjunction with CBR for classification purposes in adaptation; the details are presented in Chapter 5.

One of the differences between developing a knowledge-based method and an optimisation method is the additional amount of knowledge required to build the decision-support system. The additional knowledge can create an advantage for the system's decision making mechanism, but a disadvantage in terms of the amount of time required for obtaining the relevant knowledge. When creating a treatment plan clinicians often use their intuition and experience. The underlying mechanism behind the treatment plan created using optimisation models is often not known to the end user, which is why their use is not widely employed in practice, e.g. the medical physicists at Nottingham City Hospital mentioned it. With knowledge-based systems it is usually easy to offer an explanation of how a solution was derived, as it can be seen in the rules or cases that were applied. The type of knowledge-based reasoning that is relevant to this thesis is case-based reasoning, and its overview and application in healthcare and RTP domains will be presented in the next section.

2.3 Case-Based Reasoning

Often the way people solve their problems is by remembering similar experiences they have had in the past and forming decisions based on that past experience. CBR is the reasoning process we often subconsciously use in our daily activities; when we come across a problem the solution to it will often be based on our experience of dealing with similar events or problems. CBR is based on using old experience to understand and solve new problems. CBR was first introduced by Kolodner (1992), who demonstrated its application in everyday activities of human life. The CBR process is comprised of the steps of reasoning that take into consideration all previous similar events and features that could help in solving a new problem. The essential idea behind CBR is that similar problems have similar solutions.

CBR solves new problems by retrieving similar previously solved problems and their solutions. The cases and their solutions are stored in an archive also known as the case base. Given a new case, for which a solution is to be found, the most similar case in the case base to the new case is retrieved along with its solution. The solution of the retrieved case is used to form the solution of the new case. As CBR reasons using past experience it is easy to provide an explanation to the end user.

The retrieved solution can be applied directly to the new problem, for example in tasks such as classification and evaluation. However, tasks that require design or planning usually need revision or adaptation of the retrieved solution in order to meet the exact needs and specific requirements of the new problem.

CBR was studied by researchers in many different forms and subjects. Aha (1998) argued that contributions from reasoning from cases are present in many disciplines. It was suggested that CBR has existed and has been researched under different names since long before it was known as CBR. He defined CBR as a type of *lazy problem solving* approach, which is characterised by the storage of cases and their demand-driven recalling as a response to queries. It is a form of problem solving in which computation is not performed until there is a request for it. Lazy approaches rely on the availability of cases rather than the difficult task of extracting rules. Pattern recognition, feature selection and weighting are considered to be part of lazy problem solving approaches too. The most imperative reason for using a lazy problem solving approach is that it is highly intuitive; experts solve problems based on past experience of a similar case, which suggests a form of CBR.

CBR gained recognition especially in the last few decades and has been used in many research domains, including engineering, pharmacy, chemistry and medical domains (Vong and Wong, 2010; Guessoum et al., 2014; Bichindaritz and Marling, 2010; Schmidt et al., 1999). CBR was noted to be especially useful for complex domains or domains in which the underlying problem is not well understood.

There are many advantages of using CBR such as:

- CBR can be used to find solutions to problems when the complete understanding of the domain is not captured or is very complex.
- It allows the reusing of past solutions, which helps to avoid solving problems from scratch and thus saves time.
- It allows learning from mistakes by retrieving incorrect cases (Hammond, 1986).
- Cases can be used for a number of different problem solving purposes as they contain more information, while rules can generally be reused only for the purpose for which they were defined (Aha, 1998).
- CBR can be easily explained, understood and can, therefore, be straightforwardly used and applied.

The main disadvantages of CBR are:

- It is more difficult to capture domain knowledge in a case base with a limited number of cases.
- CBR requires larger storage requirements and demands careful attention for the definition of case retrieval and adaptation (Aha, 1998).
- CBR might lead to more inaccurate results than when mathematical formulations are used to solve a problem.

One of the characteristics of CBR is that in order to utilise this technique in a particular domain, cases with successful, correct or desirable outcomes or design should be available. That is, the case base should contain successful cases from which the system can learn.

A CBR system usually comprises retrieval, reuse, revision and retainment phases (Aamodt and Plaza, 1994). Figure 2.4 illustrates the standard CBR cycle. In the retrieval

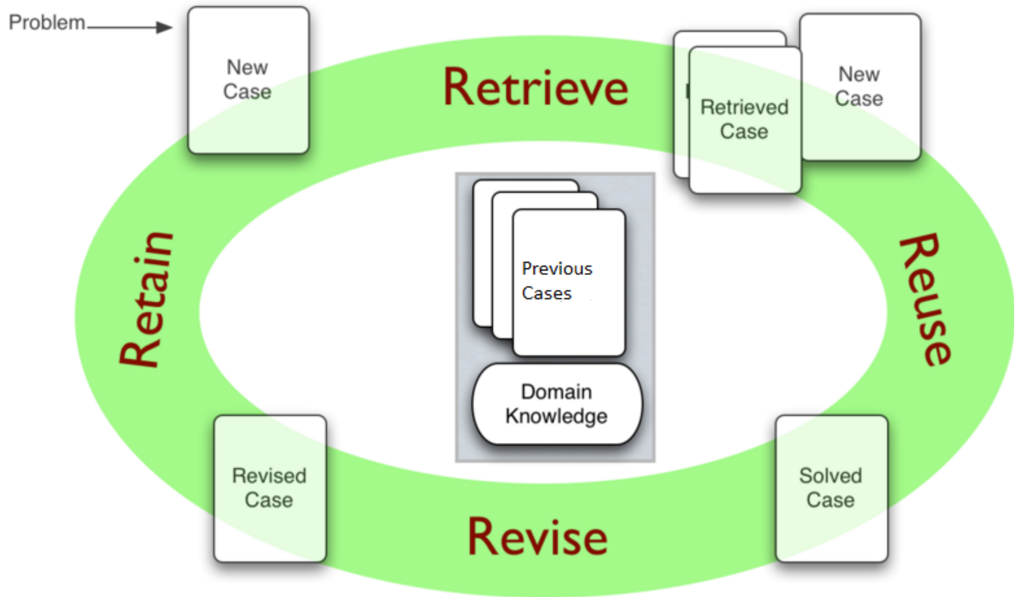

FIGURE 2.4: CBR '*4 R*' cycle. (Cbrwiki, 2012)

phase, a problem similar to the new one is retrieved from the case base. The solution of the retrieved case is then reused for the new case. It can be reused directly without any changes to the new case, or it might need to be revised to meet all of the requirements of the new case. If the case is assessed to be useful for future problem solving, it is stored with its solution in the case base for future retrieval during the retainment phase. In the literature, these stages of CBR are often referred to as '*4 R*' - cycle of retrieving, reusing, revising and retaining.

They will be described in the remaining part of the chapter with the focus on applications in healthcare and, in particular, cancer and RTP domains.

2.3.1 Retrieval

The retrieval stage is the first stage in the CBR system. It is considered to be a very important phase of the CBR process, as the method chosen for this stage will determine the success of selecting the most suitable case. In the early CBR systems the main focus was on the retrieval part of CBR, as the successful implementation of it will affect the efficiency of retrieving cases which are similar to the new (target) problem case (Aamodt and Plaza, 1994). Generally, all retrieval mechanisms rely on finding a case to be retrieved that will be the most similar to the new case, on the premise that similar cases have similar solutions and hence the solution of the retrieved case will be a good match for the new problem case.

Retrieval can be considered as a pattern recognition process (Leake et al., 1995). Every case is represented by the attribute values that characterise the problem. The new case attribute values are compared to the attribute values of the cases in the case base. In the literature, the nearest neighbour method or k-nearest neighbour method (k-NN) (Cover and Hart, 1967) are the most widely used. The nearest neighbour similarity measure is distance-based. In k-NN method, k denotes the number of cases retrieved. It computes the similarity between the attribute values using Euclidean distance between two attributes. The efficiency of k-NN is reduced if there are too many irrelevant attribute values present; therefore it is important to select attributes that are most relevant to the solution. In addition, the value of k should be chosen carefully; if it is too small then the chosen set might be susceptible to outliers and if it is too large some cases included might be dissimilar and thus impair the similarity mechanism. The attributes are selected based on their importance with respect to the solution case. The importance of an attribute is defined using attribute weights, which determine the relative contribution of each attribute to the similarity between two cases. When weights are used, the algorithm is known as weighted k-NN. A careful selection of the attribute weights is imperative to ensure that the solution of the retrieved case is applicable to the new problem.

The choice of the algorithm to be used in the retrieval mechanism is often made through a trial and error process by observing which method gives good results on known test data. In the CBR system by Jagannathan et al. (2012) the k-NN-based similarity measure is used for the retrieval process.

In many real-world problems attributes describing a case have different measurement units. Therefore, often normalisation of the attributes values is performed. Alternatively, they are represented by fuzzy sets, which allows representation of the attributes with different values or scaling systems. In the CBR system for radiotherapy planning (Jagannathan et al., 2012) attributes are represented using fuzzy sets and then the k-NN method is applied to retrieve similar cases.

The solution of the retrieved case can be applied directly to the new case as it is or adapted to suit the conditions of the new case. In classification systems, where the solution for the retrieved case can be applied directly to the new problem, a CBR system consisting of the retrieval phase only is often sufficient. For example, in De Paz et al. (2009) the CBR system designed for gene classification for leukaemia consists only of the retrieval phase. The revision step is limited to assessment and validation of the retrieval phase by the system users, i.e. if the reasoning process is clear and satisfactory then the solution is accepted. Raman et al. (2011) used a CBR system for classification of malignant masses in mammograms, in which the system consisted of the retrieval phase only. However, apart from classification systems, most other domains, where design and planning is involved, require the revision of the retrieved solution to be used for the new problem case (Hennessy and Hinkle, 1992; Vong and Wong, 2010; Sanchez-Ruiz et al., 2008; Bichindaritz and Marling, 2010). Vong et al. (2002) implemented a CBR system in the design of a hydraulic production machine in which the retrieval phase was implemented using the production rules available in the domain. However, the retrieved solution had to be adapted by the domain experts. Hennessy and Hinkle (1992) developed a CBR system for autoclave loading in which appropriate shapes of the moulds to be used are suggested by the system. The CBR system retrieved the cases, which then needed to be adapted by the system users. In the CBR system for the diagnosis of chronic obstructive pulmonary disease the retrieval phase retrieved solutions that needed revision in order for them to be applicable to the new cases (Guessoum et al., 2014).

2.3.2 Adaptation

In the adaptation phase, the solution of the retrieved case is adapted to match the new problem case. Craw et al. (2006) highlighted the incremental nature of adaptation as a further reasoning after the retrieval phase. Smyth and Keane (1998) pointed out that adaptation adds intelligence to what would be simple pattern recognition. Adaptation is a crucial phase that makes it possible for the CBR system to be used in the real world. Bichindaritz and Montani (2011) claim that lack of the adaptation phase is responsible for most of the developed CBR applications not being used in practice in medical domains. Often in practical applications the adaptation step is left out and the retrieved case solution is applied to the new problem without any modification, or adaptation is left to the end user (Bichindaritz and Marling, 2010). In the literature this approach is often referred to as null adaptation (Wilke and Bergmann, 1998; McSherry and Stretch, 2011).

CBR reasons using knowledge contained in the case base. However, once the similar case is retrieved more knowledge is needed, often outside of cases available, to adjust it for the new problem. Thus, to perform CBR adaptation, additional knowledge is required. Adaptation is considered as the knowledge-intensive phase of the CBR system. It is largely accepted that one of the major difficulties of CBR is knowledge elicitation and case adaptation. Chebel-Morello et al. (2013) reviewed that the time system users and developers dedicated to these phases were between 37.7% and 45.9%, of the total time they respectively spent in the development of the system. Craw et al. (2006) reviewed that the dominance of classification problems for which the early CBR systems were designed has resulted in a limited need of adaptation. However, the recent increase of CBR systems application in non-classification domains has brought attention to the need for adaptation and subsequently accentuated the knowledge engineering demand.

Knowledge acquisition is a key element prior to starting adaptation phase implementation. Literature suggests two ways of obtaining knowledge for adaptation: first, from the available domain and case base, and second from the domain users. However, with the former approach often more substantial knowledge is required than one can get from the case base. Thus, the majority of adaptation approaches currently reported in the literature are based on acquiring adaptation knowledge from domain experts or involving them at least in some stage of the adaptation phase implementation (Hennessy and Hinkle, 1992; Vong et al., 2002; Reyes et al., 2015). Reyes et al. (2015) implemented an adaptation knowledge acquisition process for an industrial machine design problem by involving domain experts. They suggested that knowledge acquisition for adaptation is a dynamic process, as knowledge constantly changes, hence, the dynamic nature of the adaptation phase is key to the successful reuse of knowledge. Therefore, they suggest that for up-to-date adaptation knowledge, the adaptation learning should be guided by experts, through their interactions with the system.

Fuchs et al. (2014) reviewed research studies of adaptation in CBR and concluded that a general model of adaptation has not been developed yet. This is due to the large amount of domain knowledge usually required by adaptation, which is not always readily available or easy to acquire. In addition, as the knowledge is highly domain dependant, it is difficult to design a generic model for adaptation.

Craw et al. (2006) proposed that for domains represented by numerical attributes the adaptation part is more straightforward compared to domains represented by symbolic ones, as there exists a natural means of dissimilarity calculation between the new and retrieved solutions. Symbolic attributes are represented either by text, shapes or signs.

However, while this is true for linear domains, most of the domains for which CBR is developed are complex or non-linear, in which such calculations are not simple. The works of Sanchez-Ruiz et al. (2008) and Chebel-Morello et al. (2013) have shown that, with the help of experts, symbolic domains are easier to adapt even though symbolic dissimilarity cannot be calculated. This can be explained by obvious domain knowledge that can be seen from symbols, i.e. the incorrectly retrieved solution is visible. For example, in the CLAVIER CBR system, designed for autoclave loading by Hennessy and Hinkle (1992), if the retrieved case had a wrong shape layout (e.g. a triangle instead of a circle) the experts corrected the retrieved case and the adapted case was used for future reference. In (Blansche et al., 2010; Cojan et al., 2011; Mota and Agudo, 2012) CBR systems were developed for the cooking domain. Owing to the symbolic nature of the domain, adaptation rules could be created with the help of the domain experts. In adaptation rules ingredients were divided into being either acceptable or unacceptable for the certain recipe, e.g. beef was an unacceptable ingredient for an apple pie.

Adaptation methods that are found in the literature are variations of the three types of adaptation presented by Wilke and Bergmann (1998): generative, transformational and compositional adaptation. *Generative adaptation* implies an insertion of a problem-solver into the system. Such a problem-solver needs knowledge that helps to construct a solution from scratch, rather than an explanation of differences between problem descriptions that lead to differences in the solution. In *Transformational adaptation*, the retrieved case solution is transformed into a new solution for the new problem. This type of adaptation is done by modifying, adding, deleting or substituting some elements of the solution. For example, in the process of the adaptation stage for personal computer (PC) configuration, an adaptation based on substitution of some elements of the solution is used to get a better result (Wilke and Bergmann, 1998). The adaptation process was captured in rules that could be stored and applied for future use. Substitution-based adaptation was useful for domains in which the inference mechanism is clearly structured (Gonzalez-Calero et al., 1999; Abdul et al., 2014). For example, in the work of Gonzalez-Calero et al. (1999), substitution-based adaptation was guided by the explicitly represented dependency relations within a case, in order to guide adaptation for a solution. The reasoning between the relation and concept description was carried out using description logics. The adaptation was performed by the domain users. Gonzalez-Calero et al. (1999) compares substitution adaptation to the simple rule-based adaptation process described in the works of Leake et al.

(1995). Both case studies included a case base of substitution cases or rule-based adaptation cases, from which a search engine looked for cases. If the adaptation case was retrieved it was applied directly for adaptation purposes. Knowledge acquisition of the adaptation cases was an interactive process between the system and domain user in all of the above discussed systems. *Compositional adaptation* is the combination of both generative and transformational adaptation.

Methodologies that deal with adaptation can be divided into four groups: rule-based adaptation, case-based adaptation, adaptation-guided retrieval and machine learning-based adaptation. While the former two are knowledge-intensive approaches, which require the involvement of domain users, the adaptation-guided retrieval and machine learning-based adaptation methods that are found in the literature can be classed as either knowledge-light or knowledge-intensive approaches depending on the amount of knowledge invested to implement them.

Rule-Based Adaptation

In rule-based adaptation, rules can be derived after reviewing and revising the retrieved cases to specify, under a certain condition, how to modify the value of a solution feature, or how to insert or delete certain features in the case representation in order to adapt a solution for the new problem. Such an approach to rule-based adaptation was employed in the CBR system for radiotherapy treatment planning for prostate cancer by Mishra et al. (2011) in which experts' personal opinions were modelled by rules. Koton (1988) used rule-based adaptation in a CBR system for heart failure. The treatment procedure followed a rule-based protocol in the form of the *if-then* rules, which were used in adaptation. For example, in the case of the absence of the specific symptom in the new patient that was present in the retrieved patient, the verdict associated with this symptom that leads to diagnosis is dropped during adaptation. D'Aquin et al. (2006) went a step further and introduced an automatic adaptation knowledge acquisition process for learning adaptation rules from experts. However, such a learning process required constant communication with the experts at the development stage.

Adaptation Case Base

When using an adaptation case base, during the development phase domain experts adapt cases suggested by the CBR system and the adapted solutions along with the adaptation procedures are stored for future reference. They can then be used for adaptation when the difference between the new problem and the retrieved case is similar to the difference between the case which generated the proposed adaptation and its corresponding retrieved case.

In some domains a separate adaptation case base is not created and instead the adapted solution is added to the existing case base, and thus, naturally enhances its knowledge base. Such an approach is referred to as case-based adaptation. Vong et al. (2002) claimed that case-based adaptation is another level of CBR outside of the *'4 R'* cycle, which helps CBR system to adapt itself, as incorrect cases are also saved to learn from in the adaptation process. However, they point out that it is domain dependent and, in order to enhance adaptation capability, the features which can be modified, added or deleted should be known prior to constructing the case-based adaptation level. If enough adaptation case examples are gathered, such an adaptation history could help in creating adaptation rules. In the CBR system developed by Hennessy and Hinkle (1992) for autoclave loading case-based adaptation was employed. The solution of the retrieved case adapted by the expert was added to the case base. Hence, the adaptation case base expanded gradually with the use of the system. Vong et al. (2002) proposed that integrating experts into the adaptation process was the most natural way of acquiring domain knowledge and developed a CBR system in the design of a hydraulic production machine in which adaptation was performed with the help of domain experts.

Knowledge-Light Approach

Generally, an adaptation phase requires a large amount of domain specific knowledge. However, CBR is often used in poorly understood and complex domains, thus, the knowledge needed for adaptation is not always available or can be difficult to obtain. Domain experts or system end users who could provide such knowledge are not always available in some domains, e.g. medical domain. Contrary to the knowledge-intensive approaches, knowledge-light approaches do not rely on the use of domain-specific knowledge. Wilke et al. (1997) emphasised that knowledge-light refers to methods which do not presume a lot of knowledge acquisition work before learning, but instead use already acquired knowledge inside the system, which is inherently part of the cases in the case base. They presented a framework that focuses on the knowledge-light approaches. In knowledge-light approaches the knowledge available in the case base is used 'as it is' without any, or very little, additional knowledge. The knowledge-light approaches can be useful when the sources of additional knowledge for adaptation are not available. Hanney and Keane (1996) did some early work in this area. They presented an inductive learning algorithm to extract adaptation knowledge from the cases in the case base. The algorithm built pairs of cases and used the feature differences of these case pairs to build adaptation rules. This approach is based on the assumption that the differences that occur between cases in the case base are representative of the differences that will occur between future problems and the case base. The experiments were reported only in the context of linear domains.

Adaptation-Guided Retrieval

Despite the practice in CBR of performing adaptation upon the retrieved case, one of the often encountered problems is that the adaptation phase cannot be performed due to the retrieved solution not being a good starting point for adaptation.

Smyth and Keane (1998) suggested that it was unwarranted to assume that the most similar case was also the most appropriate for adaptation. Moreover, it is argued that retrieval based solely on similarity can generate an incapacity for CBR to solve a problem, by proposing an inadequate solution. They suggest that the similarity measure should be expanded with the inclusion of adaptation knowledge about whether the case can be easily modified to fit the new problem case. Consequently, they present a new criterion, case adaptability, to evaluate the adaptation effort in order to improve the retrieval performance. They propose adaptation-guided retrieval, in which the retrieval mechanism retrieves a case that is easy to adapt to suit the requirements of the new case. This is further stressed in the work of D'Aquin et al. (2006) and Fuchs et al. (2014). In adaptation-guided retrieval a similar case is one that can be easily adapted to the new case. In the adaptation-guided retrieval, the standard CBR system that follows the procedure of retrieving the most similar case is reconsidered when the retrieved solution cannot be adapted. While this approach is essentially aimed at the retrieval phase of the CBR system, the changes to the retrieval are done with consideration of case adaptability, thus this approach is often tied to the adaptation phase.

The knowledge-light adaptation-guided retrieval approach to adaptation was introduced by McDonnell and Cunningham (2006). This approach was developed for non-linear systems where the influence of each attribute on the output value is not linear throughout the range of values for that attribute. Adaptation was performed using the knowledge contained in the case base itself. The retrieval phase searched for both a case similar to the new case and a pair of cases that were similar to the pair (i.e. new case and retrieved case) and which were therefore appropriate for adaptation. These cases were all used in solving the problem given in the new case. Their knowledge-light adaptation approach was used to predict the values of house prices. The price of a house suggested by the retrieved case was adapted based on the difference in values of attributes such as the number of bedrooms and locations of neighbouring houses. The knowledge-light adaptation approach used differences in attribute values between cases as a tool for adaptation purposes. In RTP, a knowledge-light approach to adaptation was applied in the CBR system for dose planning in prostate cancer (Mishra et al., 2009). In particular, the calculation of how much to violate the constraint on the radiation dose limit imposed on OAR was performed by using the retrieved case and a pair of cases with similar differences in the input attribute values as the differences between the new and the retrieved case.

2.3.3 Maintenance

After adaptation is performed, the new problem case, or sometimes the adapted case, are added to the case base. Adding or retaining cases in the case base is part of the case base maintenance. Leake and Wilson (1998) define case base maintenance as revision of organisation and contents of the case base in order to facilitate future reasoning. Case base maintenance helps to keep it in the most competent and useful form, by deleting cases that are redundant or noisy (Craw et al., 2007). There are fewer reports in the literature on case base maintenance in CBR compared to the retrieval and adaptation phases. Maintenance in CBR systems is often a natural result of the adaptation phase. Knowledge required for adaptation learning adds to the issue of general case base maintenance. For example, adding adaptation cases into the case base during the adaptation process as described earlier would ensure natural up-to-date maintenance of the whole CBR system. Sometimes the maintenance of the case base can require additional knowledge, which may lead to a high involvement of experts at this stage (Hammond, 1986).

Hammond (1986) suggested not only to store successfully adapted cases in the case base but also failure cases, claiming that by doing so the planner would be able to anticipate and avoid those failures in the future. This is similar to a 'failure-driven memory' approach which was introduced by Schank (1982). No case or rule is deleted, but instead stored for capture of the experience. In this approach, when the CBR system fails to find a case suitable for the new case and retrieves the incorrect case, the experts identify and index the important features that differentiate between the incorrectly retrieved case and the new case. Memories of past failures are used to warn the planner of impending problems and memories of past repairs are called upon to tell the planner how to deal with them. By storing failures as well as successes, the planner is able to anticipate and avoid future plan failures. This simple approach allows large systems to be easily built and maintained by unaided experts. A ripple-down rule maintenance approach was presented by Kang and Compton (1995) for the classification domain. The ripple-down rule methodology is essentially a way of storing and retrieving cases. The cases and associated justifications are added incrementally to the case base when a case is misclassified. These cases are then used to create rules that avoid retrieving a misclassified case in the future.

2.3.4 Application of the CBR in Healthcare

CBR has been used in medical applications in such diverse areas as clinical diagnosis, treatment and epidemiology. Holt et al. (2005) discussed the medical applications that have been using CBR in the last two decades and grouped them into the following categories: systems for classification, diagnosis, planning and tutoring. It was indicated that the use of CBR was successful in many applications. However, they pointed out that adaptation in CBR was a problem in many domains, including the medical domain, due to the adaptation knowledge being expensive to gain. Consequently, only a few applications were used in clinical settings.

The first applications of CBR in medicine appeared in the early 1980s (Kolodner and Kolodner, 1987; Koton, 1988). One of the earliest CBR systems in the medical domain is CASEY, which was developed for diagnosing heart failure by (Koton, 1988). CASEY was built on the top of a complete model-based diagnostic system which used a case based library. The CASEY system took a description of a patient's symptoms and produced a causal model of possible states that could lead to the solution. Given a new case, CASEY tried to find a case with similar symptoms. Then the retrieved case was adapted to meet the differences between the new and old case. The system proved to be more efficient in adapting old cases than the previously used model-based system. The SHRINK CBR system developed in (Kolodner and Kolodner, 1987) was used to aid psychiatric diagnosis and treatment. While SHRINK was never fully implemented, it was used later as a basis for MNAOMIA (Bichindaritz, 1996), a CBR system designed to diagnose and treat psychiatric eating disorders. Adaptation was left out, as knowledge present in MNAOMIA CBR system was deemed to be insufficient to perform this phase. Bulu et al. (2012) used CBR in the detection and retrieval of breast masses from a digital mammography archive. The retrieval part of the CBR system was implemented with no report on adaptation. Diaz et al. (2006) demonstrated a CBR application (Gene-CBR) that was used for cancer classification. The system classified a new case and the result was examined by an expert. Adaptation was performed by system users who accepted a solution and reconsidered previous classifications in light of the new evidence, or rejected it and looked for another solution. Huang et al. (2012) used CBR in their decision support system for breast cancer classification, which consisted only of the retrieval phase.

Bichindaritz and Montani (2011) reported that early CBR applications in health sciences focused on diagnosis but it was pointed out that there have been no systems implemented in the clinical setting yet. Schmidt et al. (2003) reviewed adaptation techniques used in medical CBR systems. The only technique that they found to be general enough to solve many medical adaptation problems is the application of adaptation rules. Although the technique is a general one, the content of such rules had to be domain specific. However, for some complex medical tasks the generation of adaptation rules was often too time consuming and sometimes impossible due to the domain complexity. Adaptation rules that were created based on the existing standards or protocol used in medical practice were reported to be especially useful (Mishra et al., 2011; Guessoum et al., 2014).

Liber and Bresson (1999) made use of both rule-based reasoning and case-based reasoning in their CASIMIR/CBR system for breast cancer treatment planning. The knowledge for breast cancer treatment is presented in the form of protocol and used as a rule system for creating a treatment plan. However, it does not produce successful enough solutions; therefore they are revised by a medical committee. CASIMIR consisted of a rule-based reasoning part of the system based on the protocol, and a CBR part. CBR employed the experience and knowledge of a therapy committee in adjusting a treatment plan suggested by the rule-based reasoning part. Although in CASIMIR different approaches to adaptation were attempted, e.g. using similarity paths, a lack of adaptation knowledge was reported.

Schmidt and Gierl (2001) developed the ICONS CBR system for antibiotics therapy advice. While the system needed to be developed further to be used in a clinical setting it found its use for tutoring purposes of the junior staff at the initial stage. The adaptation approach implemented in the ICONS system was rule-based adaptation, where, if the retrieved case and new problem case had different contraindications, a set of therapies associated with them were added to the retrieved solution or removed accordingly. In the second adaptation approach the end user provided extra information about the case which led to adaptation being performed not on a single case, but on a statistical evaluation of that specific information from the number of cases. The adaptation based on this statistical evaluation was conducted manually by the system user and details of it were not reported. Guessoum et al. (2014) applied CBR for the diagnosis of a chronic obstructive pulmonary disease caused by tobacco. The adaptation rule set was created based on known clinical treatment protocol, where rules were applied according to the weights of the symbolic attributes of the case. Evaluation of the CBR system demonstrated that with adaptation it produced better results compared to the null adaptation approach.

CBR in Radiotherapy Treatment Planning

The RTP problem, introduced in Section 2.2, is a trial and error approach, i.e. when creating a treatment plan for a new patient, medical staff try to retrieve/remember similar cases that they have encountered in the past, find their treatment plans and try to map them to the new case. This process is similar to the CBR process, therefore using it seems to be a natural decision in designing an application that will help medical staff to make a treatment plan.

CBR systems for radiotherapy treatment planing have been reported in the literature, although to a limited extent. A list of CBR systems developed for radiotherapy treatment planing problem are presented in Table 2.1.

The first CBR system for radiotherapy treatment planning for thorax cancer, Roentgen, was introduced by Berger (1994). Roentgen retrieved cases based on the location

	System implemented	Cancer site	References
Radiotherapy Planning Systems	ROENTGEN	Lung cancer	Berger (1994)
	CBR in Robotic Radiosurgery	Prostate cancer	Schlaefer and Dieterich (2011)
	Dempster-Shafer rule based CBR system	Prostate cancer	Song et al. (2007), Petrovic et al. (2011)
	Trade-off planning	Prostate cancer	Mishra et al. (2011)
	2-phase retrieval CBR system	Brain cancer	Jagannathan et al. (2012)

Table 2.1: CBR systems in radiotherapy treatment planning

and geometrical measures of relevant organs and the tumour, and treatment constraints. The dose distribution in treatment planning was determined according to a patient's internal geometry, hence a good match had to be similar with respect to the geometry of the new patient. In addition, the dose delivered to the tumour and healthy tissues of the patient in the retrieved plan had to meet the constraints of the new patient. However, there are no published results of this study.

CBR was used in robotic radiosurgery for beam generation planning by Schlaefer and Dieterich (2011). The purpose of using CBR was to reduce the time it took to search for treatment beams. The results have shown a considerable reduction of runtime of algorithm for beam generation, namely almost up to 50% compared to the conventionally used heuristic-based beam selection. They reported that the CBR system was especially efficient in domains where treatment protocols are established, e.g. prostate cancer.

Song et al. (2007) and Petrovic et al. (2011) used CBR to determine the dose plan for prostate cancer patients in radiotherapy treatment planning. The dose plan was determined based on the trade-off between the expected benefit of cancer control by delivering a high radiation dose and the risk of side effects caused by radiation of the surrounding OARs. A CBR approach using fuzzy set theory in the similarity measure was proposed (Song et al., 2007). No adaptation was reported in Song et al. (2007); instead multiple cases were retrieved and the fusion of their solutions was suggested as the dose for the new patient. The Dempster-Shafer theory was applied in the retrieval phase to fuse plans from multiple retrieved similar cases to strengthen agreements and decrease conflicts in situations where oncologists might have different opinions of the treatment dose for the same patient (Petrovic et al., 2011). Mishra et al. (2011) developed a CBR system for dose planning in prostate cancer to determine the extent by which dose limits set for OAR can be exceeded to better fight cancer cells. In (Mishra et al., 2011; Petrovic et al., 2011) adaptation rules were created based on a treatment protocol.

A CBR system for brain cancer radiotherapy treatment planning was developed by Jagannathan et al. (2012). The main focus of their research was on the retrieval phase of the CBR system. The similarity measure used local weights assigned to case attributes that depended on the values of other case attributes (Jagannathan and Petrovic, 2012). The treatment plan was retrieved with the beam number and beam angles suggested for the new problem case. This system does not include an adaptation phase. The research presented in this thesis builds on this work and investigates to what extent adaptation can improve the plan proposed by the retrieval phase of the CBR system.

2.4 Conclusions

This chapter first provides the background on the radiotherapy treatment planning problem. Different approaches were developed to solve the radiotherapy treatment planning problem including optimisation and knowledge-based methods. Optimisation methods fall short in sufficiently addressing the clinical issues that arise in the design process, when compared to the process conducted by the medical experts. Knowledge-based methods are better in this regard as they try to simulate the natural problem solving used by the domain experts. Rule-based reasoning is a popular approach, however in some cancer domains rules for radiotherapy treatment planning are difficult to formulate.

Although many CBR applications have been developed for the medical domain, very few of them are used in clinical practice due to the lack of an adaptation phase. The reported studies that included the CBR adaptation phase were mostly knowledge-intensive, i.e. domain experts were involved at some stage of the adaptation development.

The application of CBR in RTP is very limited and all of the reported studies have demonstrated a need for adaptation.

The literature on CBR adaptation has demonstrated the need for additional knowledge acquisition to perform adaptation. Adaptation was identified as a knowledge-intensive phase of the CBR. Typically additional knowledge is obtained through the domain users. However, the time of medical domain experts is limited and often not available. In this thesis, we present adaptation approaches that do not require the presence of the medical experts and rely solely on the knowledge available in the case base.

Chapter 3

Existing CBR system for Radiotherapy Treatment Planning for Brain Cancer

3.1 Introduction

This chapter describes the CBR system for radiotherapy treatment planning developed in (Jagannathan et al., 2012; Jagannathan and Petrovic, 2012), upon which our research builds. The CBR system was created for 3D conformal radiotherapy treatment planning of brain cancer which was developed in collaboration with the Nottingham University Hospitals Trust, NHS, Nottingham City Hospital Campus. The purpose of creating such a system was to facilitate the planning process and save medical physicists' time.

Section 3.2 describes the architecture of the CBR system including the representation of cases, treatment plan parameters that are used in the system, the similarity measure and the description of the CBR retrieval system. Validation results will be presented in Section 3.3, followed by the summary of the chapter in Section 3.4.

3.2 Data Used in the System

The CBR system contains a case base of 86 brain cancer patient records received from Nottingham City Hospital containing descriptions of patient cases and their treatment plans.

Patient data were received in the DICOM (Digital Imaging and Communications in Medicine) file format (NEMA, 1998). DICOM is a standard for handling images in clinical practice. Each DICOM patient record contains a computer tomography (CT) image data file, a radiotherapy structure file containing the outlines of the structures, such as the planning treatment volume (PTV) and OARs, and a radiotherapy plan containing the treatment plan parameters, such as the number of beams and the beam angles. The DICOM files were converted to text files for further processing.

3.2.1 Case Representation

Each case is described by attributes, which were defined in consultation with medical physicists. In the Nottingham City Hospital, after examining the image files of the patients, the PTV containing the tumour and OARs are outlined in the treatment planning system by the oncologist. The coordinates of the outlines are saved in the patient DICOM files. The following OARs are considered in the treatment plan: the right and left lens of the eyes, right and left optical nerves, the spinal cord, optic chiasm and brainstem.

In the Nottingham City Hospital, after the PTV and OARs are outlined, oncologists decide on the radiation dose and pass these details to the medical physicists. The treatment plan is then generated by the medical physicists who place beams on the patient image and determine the acceptable dose distribution in a trial and error fashion using the treatment planning system.

The case attributes have to be carefully selected so that the similarity calculation between the two cases accurately represents the relative applicability of the solution of a case to the new case. Since the treatment planning aims to focus the prescribed radiation on the PTV while avoiding the OARs, 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 OARs. These attributes give information about the geometry of a patient. The following case attributes that describe the patient in terms of attributes relevant to the treatment plan were identified and are graphically presented in Figure 3.1.

Vol - volume of PTV. Value in the range [24.87-729.82 mm^3]

A - angle between the line connecting the origin of the DICOM patient-based coordinate system in the image (O) with the centre of the PTV and with the centre of the OAR. This attribute gives information about the location of the tumour in relation to the OAR. Value in the range [0°-360°]

E - minimum distance between the edge of the PTV outline and OAR outline. This attribute is important since a large distance allows for more flexibility when placing the beams, whereas a smaller distance between PTV and OAR might require a more complicated plan and possibly a larger number of beams. Value in the range [2.04-674.87 mm].

R - the ratio of the PTV volume to the volume of the patient's body. This attribute is relevant because medical physicists have noted that similar treatment plans frequently have similar PTV and body ratios, even if the actual PTV or body volumes differ. Value in the range [6.40-255.92]

Dt - minimum distance between the body edge and PTV outline edge. This attribute provides positional information about the tumour and describes how close the tumour is to the surface of the body. Value in the range [0-53.66 mm].

P - the relative position of the PTV with respect to the OAR. This attribute is relevant because medical physicists use different beam angles based on the location of the tumour in the head. This attribute contains six labels: left, right, posterior, anterior, superior and inferior.

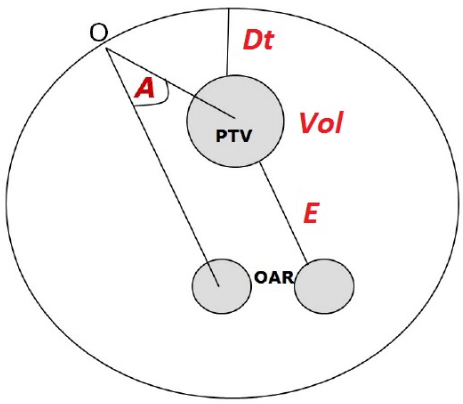

FIGURE 3.1: A graphical representation of the attributes of a case (Jagannathan et al., 2012)

The values range for the attributes are presented in the brackets next to the attribute description. In order to perform the similarity calculation all attributes were normalised to values between 0 and 1 (Jagannathan et al., 2012).

Considering all OARs, a total of 17 attributes listed below are used in the experiments presented further in this research:

- 1. PTV Volume

- 2. PTV-Body Ratio
- 3. PTV-Body Distance
- Distance PTV-OAR:
  - 4. Distance PTV-left lens
  - 5. Distance PTV-right lens
  - 6. Distance PTV-left optical nerve
  - 7. Distance PTV-right optical nerve
  - 8. Distance PTV-brainstem
  - 9. Distance PTV-chiasm
  - 10. Distance PTV-spinal cord
- Angle PTV-OAR:
  - 11. Angle PTV-left lens
  - 12. Angle PTV-right lens
  - 13. Angle PTV-left optical nerve
  - 14. Angle PTV-right optical nerve
  - 15. Angle PTV-brainstem
  - 16. Angle PTV-chiasm
  - 17. Angle PTV-spinal cord

3.2.2 Treatment Plan Parameters in the CBR System

At this stage of the development the solution of the CBR system includes two treatment plan parameters: the number of beams to be applied to the PTV and the angles of those beams. The beam angle determines the point at which radiation beams enter the head. The total angle consists of the gantry angle of the linear accelerator and the patient couch angle. For each beam both of these parameters have to be specified. In Figure 3.2, the beams are shown as red lines entering the head at different angles and passing through the tumour.

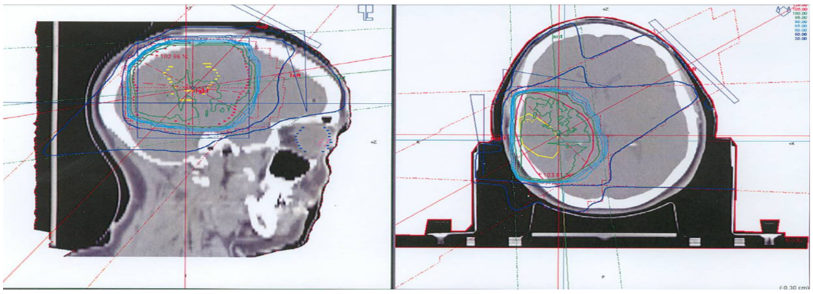

FIGURE 3.2: Beam angles projected on PTV (Jagannathan et al., 2012)

3.2.3 Case Retrieval and the Similarity Measure

Not all cases contain attributes with respect to all the OARs, as depending on the tumour location some OARs are not affected by the radiation beams. For example, if the tumour is located in the front part of the head then the brainstem might not be considered as an OAR, as no risk is posed to it due to its large distance from the tumour. Therefore, prior to CBR retrieval, the case base is filtered to contain only cases that contain the same OARs as the new case. Then the similarity between the new case and the filtered cases in the case base is computed.

The similarity measure employed in the CBR system uses local weights assigned to the case attributes. The values of local attribute weights change in each instance based on the attribute values of the new case. Experiments showed that different weights of the attributes lead to optimal results during beam number and beam angles retrieval. For this reason a rule-based algorithm was developed, which assigns local weights to the attributes based on the values of the new case attributes. The rules, which assign the local attribute weights are produced using feedback about the retrieval success of the similarity measure on training cases (Jagannathan and Petrovic, 2012).

The similarity measure which is used in the retrieval is as follows. Let us denote by C_n and C_{cb} a new case and a case from the case base, respectively. The distance, d, between the cases is calculated using the Euclidean distance:

$$d(c_n, c_{cb}) = \sqrt{\sum_{l=1,\ldots,6} w_l (V_{n,l} - V_{cb,l})^2} \tag{3.1}$$

where, $V_{n,l}$ and $V_{cb,l}$ denote the attribute values of attribute l, where $l = 1,\ldots,6$ for the attributes A, E, Vol, R, Dt, P of the new case and a case from the case base, respectively, while w_l is the weight of attribute l. The similarity, s, between C_n and C_{cb} is calculated using:

$$s(c_n, c_{cb}) = 1 - d(c_n, c_{cb}) = 1 - \sqrt{\sum_{l=1,\ldots,6} w_l (V_{n,l} - V_{cb,l})^2} \tag{3.2}$$

3.2.4 CBR Two Phase Retrieval

The retrieval was carried out in two phases, where in the first phase a treatment plan that suggests the number of beams is retrieved and in the second phase a treatment plan that suggests the beam angles is retrieved. The reason for having two phases in the retrieval is that different attribute weights are used for beam number retrieval and beam angle retrieval. The first phase of the retrieval determines the number of beams to be used in the treatment plan. The proposed number of beams is stored and used in the second phase in which only cases in which treatment plans have the number of beams proposed in the first phase are available for the retrieval. Among these cases the case most similar to the new case is retrieved in the second phase and it suggests the beam angles. The reason for finding the correct beam number first is that medical physicists consider this parameter to be one of the most important elements of the treatment plan. It is easier to adjust beam angles during the treatment planning than the number of beams, therefore it is more important that the CBR system is able to correctly determine the beam number. Furthermore, if the retrieved case has an unsuitable beam number then, in the adaptation of the beam angles, missing or extra beams need to be considered, which is not trivial. Therefore, in this research, we perform first beam number adaptation, followed by beam angle adaptation.

3.3 Validation

The performance of the system is evaluated in the validation stage. Gonzalez et al. (1998) suggested that test cases to evaluate the system can be obtained directly from the case base. The treatment plans in the case base were generated by the medical physicists and validated by the oncologists and are thus considered as the gold standard against which we evaluate the treatment plans suggested by the CBR system. The advantage of this validation approach is that it does not require any input from domain experts whose time is often limited and expensive. However, in the future we plan to get the system fully evaluated by the medical physicists.

The leave-one-out cross validation method was used to evaluate the retrieval mechanism where each case is repeatedly made a new case and the remaining cases are used as a case base from which to retrieve a similar case. This is a popular validation method used when the available dataset is small, as it improves the accuracy of the training sets. However, the disadvantage of this method is that if the error varies a lot when different samples of data are used, it might mean that the results are not very generalisable. Nevertheless, this is a popular method that is commonly used in CBR systems (Smyth and McKenna, 2001; Craw et al., 2006). The leave-one-out cross validation method will be further used for validation purposes in this thesis.

In the experiments conducted, for each new case the treatment plan retrieved by the CBR system is compared with the actual treatment plan suggested by the medical physicists. The beam number error, E_{BN}, is defined as the difference between the beam number of the retrieved case and the actual beam number used in the treatment plan. The retrieval is deemed to be successful if $E_{BN} = 0$.

The beam angles error, E_{BA}, is calculated as the average of the gantry and patient couch angles errors between the retrieved and new cases. The order in which the beams are applied is not fixed and therefore in the comparison of two cases the best correspondence between beam angles has to be established. The beam angles are matched between two cases so that the minimum beam angle error is achieved. The retrieval is successful if E_{BA} $<30\,^{\circ}$. This threshold was considered by the medical physicists to be acceptable for the new case. If the number of beams in the new case is different from that of the retrieved case, then the beam number is chosen from the case with the lowest number. Then the beam angles among two cases are matched in such a way that E_{BA} is minimum.

For example, let us suppose that the *treatment plan New* has patient couch angles $0\,^{\circ}$, $108\,^{\circ}$, $306\,^{\circ}$ and the *treatment plan Retrieved* has patient couch angles $10\,^{\circ}$ and $90\,^{\circ}$. The closest angles are paired and the extra beam is discarded from the *treatment plan New*. The patient couch angle error between the *treatment plan New* and the *treatment plan Retrieved* is then given by $E_{BApc} = |0\,^{\circ}\text{-}10\,^{\circ}| + |108\,^{\circ}\text{-}90\,^{\circ}| = 10\,^{\circ} + 18\,^{\circ} = 28\,^{\circ}$.

Similarly the error is calculated for gantry angles, $E_{BAgantry}$. E_{BA} is calculated as follows $E_{BA} = E_{BApc} + E_{BAgantry}$.

Out of 86 cases in the case base, 80 cases were used in the leave-one-out cross validation. Some of the cases could not be used due to missing OARs details or having the same OAR multiple times, e.g. two right lenses specified. The success rate was 61% for the beam number retrieval, i.e. in 49 out of 80 cases the system retrieved a case with the same beam number as suggested by the medical physicists. The success rate of the beam angle retrieval was 39%, i.e. out of 80 cases in 31 cases E_{BA} was less than the defined threshold.

3.4 Summary

This chapter has introduced the existing CBR system that was developed by Jagannathan et al. (2012). The case base, architecture of the CBR system, similarity measure used and validation of the system were presented. The results of the CBR retrieval are a good starting point, however they suggest the need for further adaptation.

The aim of the research presented in this thesis is to build on the existing CBR system. The developed CBR retrieval system is considered as it is with its attributes, similarity measure and identified treatment plan parameters, i.e. beam number and beam angles. The remainder of this thesis will focus on adaptation of the treatment plans retrieved by the CBR retrieval system presented here. We will look at the differences in the attributes between the new and retrieved case and use this information to adapt the retrieved beam number and beam angles to fit the new case.

Chapter 4

Clustering in CBR Retrieval

4.1 Introduction

The CBR system developed by Jagannathan et al. (2012), which was described in Chapter 3, was tested by medical physicists using the new patient cases. It was discovered that some of the retrieved cases were not suitable to be adapted for the new cases. In particular, the treatment plans of some of the retrieved cases could not be adapted to the new case because their tumour positions were very different to the tumour position of the new case. This resulted in highly unsuitable beam angles. Treatment planning aims to deliver the prescribed radiation to the centre of the tumour. Therefore, the placement of beams is largely determined by the position of the tumour.

This motivated us to revise the retrieval mechanism of the existing CBR system by adding a clustering stage that clusters cases based on their tumour positions. A number of well-known clustering methods were investigated and employed in the retrieval mechanism. Results using the available case base have shown that the success rate of the new CBR retrieval is higher than that of the original system.

The continuation of this chapter is as follows. Section 4.2 provides the insights learned from the CBR retrieval system being tested by the medical physicists. Methods implemented for determining the optimal number of clusters for better identification of the tumour position and clustering itself are described in Section 4.3.1. Computational results of the clustering of cases in the current case base and the performance comparison of the original and revised CBR systems are presented next in Sections 4.5 and 4.6, respectively. Finally, conclusions are drawn in Section 4.7.

4.2 CBR Retrieval System Validation by the Medical Physicists

The developed CBR system was presented to the medical physicists who used it for the treatment planning of the six new patients. This section describes the insights learned from this validation.

4.2.1 Isocentre and the Tumour Position

The PTV or tumour position in the head/brain is described by the three coordinates X, Y and Z, as illustrated in Figure 4.1. In the anatomical coordinate system that is used in radiotherapy treatment planning the X coordinate increases from the left side of the body towards the right, and thus, it defines the left and right position of the tumour, the Y coordinate increases towards the anterior part of the body and defines the anterior and posterior position of the tumour, and the Z coordinate increases towards the superior part of the body and therefore determines the superior and inferior position of the tumour.

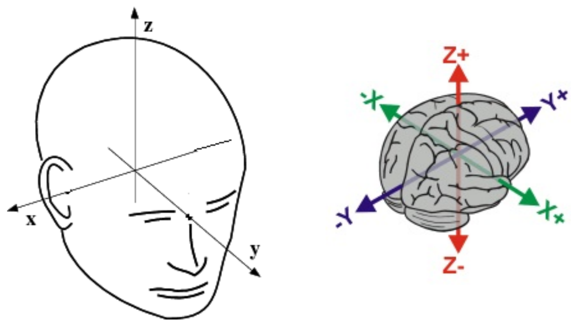

FIGURE 4.1: Anatomical coordinate system of the head and brain (Athinoula A. Martinos Centre for Biomedical Imaging, 2014; Rorden, 2002)

The beams are placed such that the radiation is focused at the point, known as the isocentre, where all beams meet and ideally this point should be placed at the centre of the tumour. Since beam angles are focused at the centre of the tumour position, from now on when we refer to the tumour position we assume the centre of the tumour.

The isocentre coordinates were extracted from the treatment plans recorded in the cases. In all cases in our case base the tumour position in the patient case and the isocentre positions in the treatment plan have coordinates that are close to each other. Therefore, it can be concluded that beams in the treatment plan are positioned well and form good solutions for the cases. Table 4.1 presents examples of two randomly selected cases from the case base, *a* and *b*, with their PTV centres (i.e. tumour position centres) and isocentre positions in the coordinate system (X, Y, Z).

	Case a	*Case b*
PTV (tumour) centre	-25.91 \-273.76 \-645.00	-16.92 \-176.43 \-599.49
Isocenter position	-29.91 \-275.01 \-645.00	-33.08 \-223.48 \-627.00

TABLE 4.1: An example of two cases with their tumour positions and isocentres in their treatment plan

Visually it can be observed from the table that the PTV centre and isocentre are very close in *Case a*. However, the isocentre is not that close in *Case b* to the PTV centre, which can be explained by an attempt to avoid the irradiation of the nearby OARs in the corresponding treatment plan.

4.2.2 CBR Adaptation: Medical Physicists Overview

In all six cases used in the validation, the system retrieved treatment plans whose beam number was the same as that generated by the medical physicists in the past.

For two out of the six new patient cases the retrieved cases provided treatment plans that were a good starting point for adaptation; for two cases the retrieved cases suggested treatment plans with some acceptable beam angles and some beam angles that medical physicists would not use; finally, for the two new cases the suggested beam angles could neither be used nor adapted for the new case.

When the medical physicists validated the CBR system they noted that adapting the retrieved cases was difficult. Particularly, they observed that the isocentre of the retrieved treatment plan was often misplaced with respect to the tumour centre position of the new patient. The tumour position relative to the OAR is captured in the similarity measure by attribute P, described earlier. However, this attribute was not sufficient to define the tumour position, as the isocentre of the suggested beams in the retrieved treatment plan did not match the tumour centre of the new case in most of the cases.

The tumour position is an important factor in treatment planning as it defines how and from which side of the head beams will enter. For example, if the tumour is located in the left side of the head then the beams should be placed to enter the head from the left side, because having beams entering from the right side would mean radiation passing through a lot of healthy tissue. In the cases where the retrieved beams are misplaced in such a manner it is easier to design a new treatment plan from scratch than to adapt the retrieved treatment plan. Therefore, in order to avoid retrieving such non-adaptable treatment plans, it was decided to take into account the tumour position of cases during the CBR retrieval system.

We analysed the isocentres of the retrieved cases and tumour centres of the six new cases (*Case 1*,...,*Case 6*) that were adapted (see Tables 4.2 - 4.4). The analysis has revealed that the proximity of the isocentre of the retrieved case to the tumour position of the new case corresponded with the adaptability of the cases. Table 4.2 provides an overview of isocentres of the retrieved cases that were a good match for the new cases and Table 4.3 provides information on the cases retrieved that were not adaptable, i.e. a bad match.

	Coordinates (Left/Right(X), Anterior/Posterior(Y), Inferior/Superior(Z))	Beam number
Case 1 (tumour position)	[-25.16 \-11.66 \22.00]	4
Retrieved case *R1* (isocentre)	[-26.56 \-14.56 \22.50]	4
Case 2 (tumour position)	[18.34 \-33.06 \21.30]	4
Retrieved case *R2* (isocentre)	[32.95 \-217.95 \-597.00]	4

TABLE 4.2: New cases for which the CBR system retrieved an adaptable case

	Coordinates (Left/Right(X), Anterior/Posterior(Y), Inferior/Superior(Z))	Beam number
Case 3 (tumour position)	[-22.26 \4.54 \6.00]	4
Retrieved case *R3* (isocentre)	[-6.46 \-221.06 \-564.00]	4
Case 4 (tumour position)	[16.64 \34.44 \27.00]	4
Retrieved case *R4* (isocentre)	[-18.49 \-225.89 \-571.50]	4

TABLE 4.3: New cases for which the CBR system did not retrieve an adaptable case

The two retrieved cases, case *R1* and case *R2*, which were good starting points for beam angle adaptation, were compared with respect to their isocentres' proximity to the tumour position of the new case with the two cases, case *R3* and case *R4*, which could not be adapted. It can be seen from Table 4.2 and Table 4.3 that the isocentres of the retrieved cases that were adaptable are closer to the tumour position of the new cases compared to the isocentres of the non-adaptable cases. However, in the retrieved case *R2*, Z coordinate, defining inferior/superior position of the isocentre, is not close to the Z coordinate of the

Case 2 but could still be adapted. Retrieved cases *R3*, and especially *R4*, which had isocentres with Z coordinates located far from the Z coordinate of the tumour position of the new cases, similar to the case *R2*, could not be adapted. From these examples we can assume that the left/right (X) and anterior/posterior (Y) coordinates are more important in defining suitability of beams to the tumour position than inferior/superior (Z) coordinates. Both *Case 1* and *Case 2* had 4 beams. The angles of the retrieved case *R1* were better suited for *Case 1* adaptation: out of four beams, three beams were a very good match while one beam was misplaced and required a lot of adaptation. It can be seen that the retrieved case *R1* isocentre matches the tumour centre of *Case 1* well. The angles of the retrieved case *R2* required more adaptation, that is more time was taken for adaptation and more changes were conducted to arrive at a suitable plan for *Case 2*, namely two beams were a good match with no change and two beams needed adaptation.

Cases retrieved for *Cases 3* and *Case 4* in the Table 4.3 were deemed to be non-adaptable by the medical physicists. *Cases 3* and *4* describe patients with posterior tumours, i.e. tumours located at the back of the brain for which cases *R3* and *R4* with anterior tumours were retrieved which had isocentres located in the anterior part of the brain. This resulted in beams entering the head from the wrong side and radiation passing through a lot of healthy tissue. The beams could not be adapted and new beam angles needed to be generated for the new cases.

	Coordinates (Left/Right(X), Anterior/Posterior(Y), Inferior/Superior(Z))	Beam number
Case 5 (tumour position)	[39.54 \16.44 \-22.00]	4
Retrieved case 5r (isocentre)	[31.69 \-229.51 \-601.50]	4
Case 6 (tumour position)	[6.74 \11.74 \16.50]	3
Retrieved case 6r (isocentre)	[-30.00 \27.30 \4.50]	3

TABLE 4.4: New cases for which the CBR system retrieved cases that took a long time to adapt

Case 5 was treated with 4 beams and *Case 6* with 3 beams. Retrieved $R5$ and $R6$ cases presented in Table 4.4 suggested beam angles which could be adapted but required a lot of time for adaptation. In addition, both retrieved cases $R5$ and $R6$ had one beam each that could not be adapted and were discarded. In both these cases the medical physicists added a newly generated beam to achieve the required dose distribution.

The evaluation of the retrieved treatment plans for the six new cases has shown that metrics used to compute the retrieval error of BA and BN do not give an indication of how suitable the retrieved plan is for adaptation. It is possible that the retrieved treatment plan is difficult to adapt or is unsuitable for adaptation in spite of having a small retrieval error.

4.3 Clustering in CBR Retrieval

To address the issues discussed in the previous section the retrieval in the current CBR system is revised to include the matching of the isocentre in the treatment plan and the tumour position of the new case.

If the tumour positions of the retrieved and new cases match, then the treatment plan of the retrieved case should be suitable for adaptation of the new case. After the OARs filtering, we filter the case base based on the tumour positions of the cases. However, in different cases tumour positions usually do not match exactly. It is not expected that a case will be found in the case base with the tumour positioned in exactly the same location as the new case. The aim is to consider only the cases in which the tumour position is very close to the new case, so that beam angles in the retrieved treatment plan can be adapted. Clustering can be defined as grouping of items that are similar to each other. For this purpose cases are clustered by their tumour positions so that during retrieval the system considers only the cases that are in the same cluster as the new case. The system then computes the similarity between the new case and the cases in the same cluster as the new case, and retrieves the most similar case.

The tumour position similarity information was incorporated in the separate clustering of cases, rather than being included in the similarity measure presented in the original CBR retrieval system (Chapter 3). The reasoning behind this approach is that if, in the CBR similarity measure, the tumour position is not similar but other attributes have strong similarity a case can still be retrieved. However, medical physicists stated that beams should not enter from the wrong side of the head, therefore filtering cases by clusters rules out the possibility of retrieving cases with different tumour positions.

Initial clustering experiments were conducted to check the plausibility and validity of the clustering idea based on the observation of the aforementioned adaptation of six cases, in order to generalise the observation over the whole case base. Clusters were selected with manually defined thresholds for X, Y and Z coordinates in a trial and error fashion. The case could be retrieved if the coordinates were in the defined threshold range: for example, if the X coordinate of the case A is +/-20 ° within the range of the X coordinate of the new case N. The results of the experiments proved that setting up thresholds for X, Y and Z led to the isocentre of the retrieved case being closer to the tumour position of the new case. This was motivation for investigation of the available automatic clustering algorithms in the retrieval mechanism.

4.3.1 Clustering Methods

This section will introduce clustering algorithms. The k-means clustering algorithm (MacQueen, 1967) is used to cluster cases based on their tumour positions. Other clustering algorithms introduced in the next section are used to determine an optimal number of clusters k to perform k-means clustering.

4.3.2 K-means Clustering

K-means is a classical robust clustering technique used when the number of clusters in the dataset is fixed a priori. The k-means clustering algorithm was first introduced by MacQueen (1967) and an efficient version of the algorithm was proposed by Hartigan (1975). This clustering algorithm is fast and simple to use which makes it a popular choice for cluster analysis.

The k-means clustering algorithm divides M data samples in N dimension into k clusters so that the average squared distance from the data points to the cluster centres, referred to as the cluster mean, is minimal. The k-means clustering is represented as follows:

$$J = \sum_{i=1}^{c} \sum_{x \in c_i} (||x - v_i||)^2 \tag{4.1}$$

where $||x - v_i||$ is the Euclidean distance between points x and v_i, c_i is the set of data points in i-th cluster, c is the number of cluster centres.

The initial cluster means are placed randomly. The k-means algorithm is iterative. In every iteration each data point is assigned to the closest cluster mean, and then the new cluster mean positions are recalculated as follows:

$$c_i = \frac{1}{n_i} \sum_{x \to c_i} x \tag{4.2}$$

where for each cluster i, c_i is the new cluster mean of all points x assigned to the cluster from the previous iteration; n_i is the total number of points x that belong to the cluster i (Lavrenko, 2014). The optimal solution is found when the assignments no longer change the distance between the data points and the cluster mean.

The value of k should be carefully selected. There are many methods used to determine the optimal number of k. Pham et al. (2005) reviewed that a lot of reported studies did not contain explanation or justification for the chosen value of k. Furthermore, Hamerly and Elkan (2004) suggested that choosing k is often an *ad hoc* decision based on prior knowledge, assumptions and practical experience. Review of clustering models revealed that values of k were typically chosen based on the decision of the domain users, determined by statistical measures or through the visualisation of the data set.

4.4 Choosing the Optimal K Number

Hamerly and Elkan (2004) developed an algorithm for choosing k based on a statistical test for the hypothesis that a subset of data follows a Gaussian distribution. The proposed algorithm runs k-means with increasing k in a hierarchical fashion until the test accepts the hypothesis that the data assigned to each k-means centre are Gaussian. The developed algorithm improved k-means performance and proved to be successful in high dimensional data sets.

Wagstaff et al. (2001) tested algorithms for finding an optimal k and determined that combining the general similarity information, which is present in k-means clustering, and the domain specific information (constraints) provided best results compared to the use of these elements on their own. This method however, requires initial knowledge of the constraints, i.e. which instances should or should not be grouped together, which is not available in our domain.

In our case clustering problem, we first need to find the number of clusters k. There are many methods to suggest the optimal number of clusters in the data set. Some of them are designed for large data sets and are therefore not suitable for our case base. To find the optimal number of k we reviewed a range of clustering methods before applying the three clustering methods to our data set, including Partitioning Around Medoids (PAM), model-based and Calinski-Harabasz clustering approaches. These approaches have different clustering models and can therefore provide a good comprehensive overview of the case distribution. PAM is a robust centroid type of clustering that can be referred to as hard clustering, i.e. the data point belongs only to one cluster. Model-based clustering is a probabilistic type of clustering which is sometimes referred to as a soft clustering, i.e. a data point can belong to many clusters, to a certain degree (probability). The clusters belong to different models and the model with the highest probability is chosen. Calinski-Harabasz clustering does not specifically look for normally distributed clusters compared to the formerly introduced methods.

4.4.1 Partitioning Around Medoids

Partitioning Around Medoids (PAM) is a heuristic clustering method which is a variation of k-means clustering (Kaufman and Rousseeuw, 1990). Data is partitioned into clusters in such a way as to minimise the distance between points labelled to be in a cluster and a point designated as the centre of that cluster. In PAM a medoid is used as the centre of the cluster. The medoid of a finite dataset is a data point from this set whose average dissimilarity to all the data points is minimal. That is, PAM selects the most centrally located data point of the cluster (which is closest to the median value). PAM is based on medoids calculated by minimising the absolute distance between the points and the selected medoid. PAM uses Manhattan distance, based on absolute value distance as opposed to the squared error used by the Euclidean distance in the k-means algorithm. As a result, it should be more robust to the noise and outliers than the k-means. However, in practice it is reported that these two methods produce similar results most of the time. Velmurugan and Santhanam (2010) observed that the k-means algorithm performed better on smaller data sets and overall it took less computational time compared to k-medoids.

4.4.2 Model-Based Clustering

Fraley and Raftery (2002) concluded that heuristic clustering methods did not answer questions such as which clustering methods should be used, how many clusters are needed, or how outliers should be handled. They proposed a model-based clustering that belongs to the class of probability-based methods. In model-based clustering the data can belong to different clusters with a certain probability, for example, to cluster 1 with 70% and cluster 2 with 30% of confidence. Each cluster in model-based clustering is represented by a generative model, i.e. probability distribution. A mixture model represents the probability distribution of observations in the sub-population within an overall population. In the mixture model each sub-population corresponds to a different cluster. The process of determining the number of clusters is reduced to the model selection problem in the probability framework.

In model-based clustering the data x are viewed as coming from a mixture density $f(x)$:

$$f(x) = \sum_{k=1}^{G} \tau_k f_k(x) \tag{4.3}$$

where where $f_k(x)$ is the probability density function of the observations x in group k, and τ_k is the probability that an observation x belongs to the kth mixture cluster ($\tau_k \in (0,1)$ and $\sum_{k=1}^{G} \tau_k$=1) (Fraley and Raftery, 2007). Each cluster is usually modelled by the Gaussian distribution G.

In this work Gaussian mixture models are used, in which each known class is modelled by a single Gaussian term with the same covariance structure among classes. Cluster distributions are characterised by the mean μ_k and the covariance matrix $\sum_k$, and have the probability density function ϕ:

$$\phi(x_i; \mu_k, \sum_k) = \frac{exp{-\frac{1}{2}}(x_i - \mu_k)^T \sum_k^{-1}(x_i - \mu_k)}{\sqrt{det}(2\pi \sum_k)} \tag{4.4}$$

where x_i is the nth observation for $i = 1, ..., n$.

The likelihood for data consisting of n observations assuming a Gaussian mixture model with G mixture clusters is as below:

$$\prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k \phi(x_i; \mu_k, \sum_k) \tag{4.5}$$

For a fixed number of components G, the model parameters τ_k, μ_k, and $\sum_k$ can be estimated by using the *expectation - maximisation* (EM) algorithm (Fraley and Raftery, 2007).

The clusters in the mixture model have an ellipsoidal shape and are centred around μ_k. Geometric features (shape, volume, orientation) of the clusters are determined by the covariances $\sum_k$ (Fraley and Raftery, 2007).

In the *mclust* package available in *R* software that will be used here, 10 combinations of volume, shape and orientation are included. There is an identifier for each possible parametrisation of the covariance matrix that has three letters: E for 'equal', V for 'variable' and I for 'coordinate axes'. The first identifier refers to volume, the second to shape and the third to orientation:

- 'EII': spherical (equal volume)
- 'VII': spherical (unequal volume)
- 'EEI': diagonal (equal volume and shape)
- 'VEI': diagonal (varying volume, equal shape)
- 'EVI': diagonal (equal volume, varying shape)
- 'VVI': diagonal (varying volume and shape)
- 'EEE': ellipsoidal (equal volume, shape, and orientation)
- 'EEV': ellipsoidal (equal volume and equal shape)
- 'VEV': ellipsoidal (equal shape)
- 'VVV': ellipsoidal (varying volume, shape, and orientation)

Bayesian Information Criterion (BIC) is used to compare and select the models. BIC balances model fit by penalising the number of parameters in the model to avoid overfitting.

4.4.3 Calinski-Harabasz Clustering

The two methods described above look for normally distributed clusters. In contrast, Calinski-Harabasz clustering (Calinski and Harabasz, 1974) presented in this section, does not specifically search for normally distributed clusters.

Calinski-Harabasz is a heuristic type of clustering where the method searches for a local optimal solution. This clustering method introduces the variance ratio criterion (VRC), which is used to determine the correct number of clusters in a cluster analysis. VRC is a method for identifying clusters of points in a multidimensional Euclidean space. It ensures an effective reduction of the number of possible splits of points. VRC uses k-means to get clustering results for different values of k (Calinski and Harabasz, 1974). These k-means runs are randomly initialised and therefore have to be run multiple times to ensure an optimal clustering. When the data has been clustered in a few different clusters the best cluster is selected using VRC (Expression (4.6)). For a solution with n data points and k clusters, the criterion is as follows:

$$VRC_k = \frac{CC_B}{CC_W} \times \frac{n-k}{k-1} \tag{4.6}$$

where CC_B is the overall between-cluster variation and CC_W the overall within-cluster variation with regard to all clustering variables. CC_B and CC_W are computed using Euclidean distance. $(n-k)$ is the degree of freedom of the within-cluster variations, and $(k-1)$ is the degree of freedom of the between-cluster variations. Compact and separated clusters are expected to have small values of CC_W and large values of CC_B. Hence, the better the data partition, the greater the value of VRC (Zhang and Li, 2013). VRC provides an informative visual tool to gain some insight into the structure and spread of the points.

When the data has been clustered in a few different models, the best model is selected based on k and the number of data points. Evaluating the ratio for the different models with increasing k, the optimal clustering should be given by the first local maximum of the ratios.

4.4.4 Plotting of the Clusters

All of the clustering methods presented above provide a graphical overview in R in the form of plots that provide information on the clustering of the data.

When visualising a multi-dimensional data set that contains more than two variables the problem arises of how to plot it in an XY coordinate plot. Principal component analysis (PCA) is a dimension reduction technique that is useful for visualising the clustering. It summarises the information of all variables into new coordinates called components.

PCA is helpful in understanding the internal structure of the data in a way that best explains the variance in the data. If a multi-dimensional data set is visualised as a set of coordinates in a high-dimensional data space (1 axis per variable) the plot becomes complex and can be difficult to explain. PCA provides, with a lower-dimensional plot, a projection of the data set that is viewed from the most informative viewpoint. This is done by using the two principal components so that the dimensionality of the transformed data is reduced. The PCA plot visualises data points according to two principal component coordinates in a so called principal plane. Each component is responsible for explaining a certain percentage of the data points variability. The higher the percentage, the better the components captured the data information.

PCA is also used to identify the effectiveness of clustering. The clusters will be merged in the principal plane when clustering is unsuccessful. On the other hand, in successful clustering the clusters will be clearly separated in the principal plane.

4.5 Determining the Number of Clusters: Experimental Results

This section compares the results obtained by the chosen methods for determining the optimal number of case clusters. It is required to find the optimal number of clusters based on our data set. In order to define the number of clusters we use off-the-shelf techniques available in *R* packages. *R* is an open source computing environment for statistical analysis of a data set. In our study *fpc* package in *R* is used for PAM clustering analysis (Hennig, 2014), *mclust* package (Fraley et al., 2012) is used for model-based clustering and *vegan* package for Calinski-Harabasz clustering (Oksanen et al., 2007).

4.5.1 PAM Clustering Results

PAM provides a novel graphical display, the silhouette plot, which allows the user to select the optimal number of clusters.

The silhouette plot in Figure 4.2 shows the clustering measure for each case and how it fits into the assigned cluster. This is done by comparing how close the case is to other cases in its own cluster and cases in other clusters. Each line on the silhouette plot corresponds to a case. The Y axis represents cases and the X axis shows how well the cases fit the assigned cluster. For example, values near 1 mean that the corresponding case fits well into the cluster and values near 0 mean that the case would fit a different cluster better. In each cluster the value is displayed from smallest to largest. The information in the silhouette plot is used to determine the optimal number of clusters. The PAM method suggests 2 clusters for the current case base based on the tumour position data as illustrated in the silhouette plot in Figure 4.2. We can see that most of the cases belong to cluster 1, and a smaller number of cases belong to cluster 2. The average silhouette width of 0.9 was obtained, which suggests that 90% of cases clustered were fitted well by the suggested method.

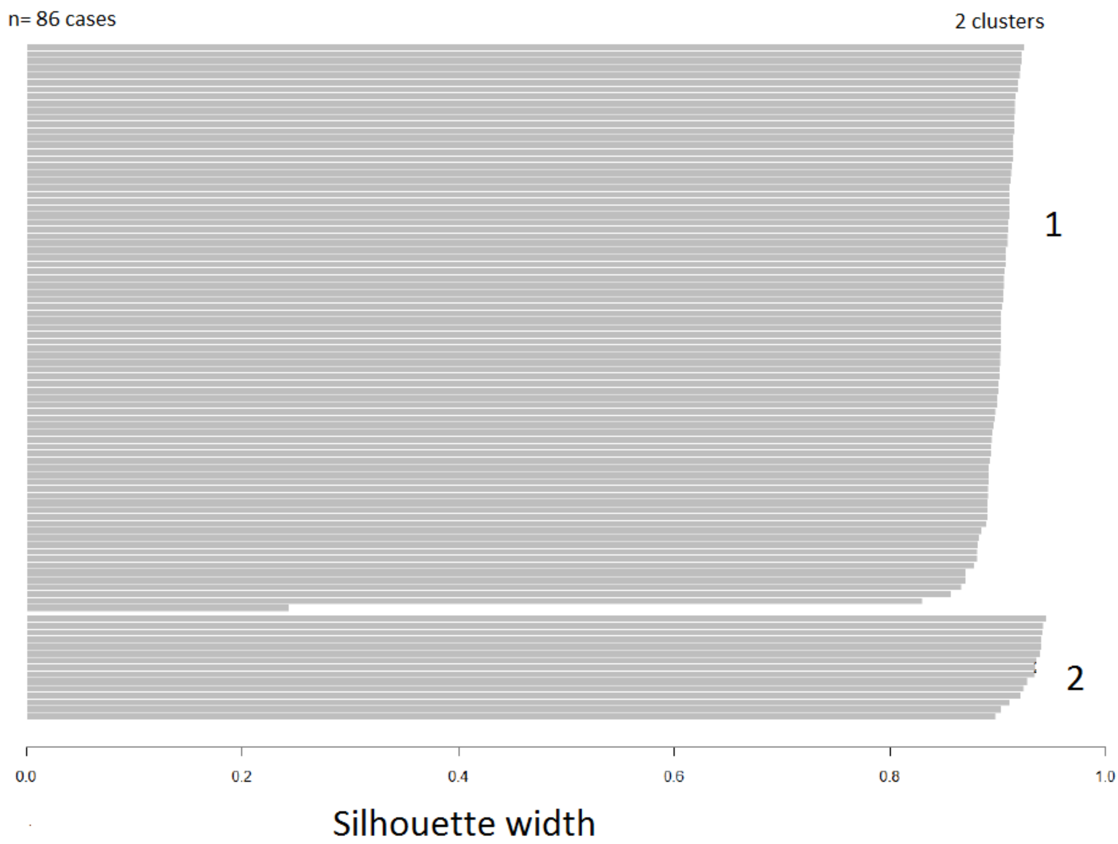

FIGURE 4.2: PAM silhouette obtained with cases from the case base

Figure 4.3 provides a plot of the clustered data observations in the principal plane.

In our experiments the two components explain 93.9% of the point variability. This means that components 1 and 2 capture the information on our data very well. From Figure 4.3 we can also see that in our experiments successful clustering was achieved, as the clusters are clearly separated.

4.5.2 Model-Based Clustering Results

Figure 4.4 presents models of case clustering for ten parametrised covariance structures (Fraley et al., 2012).

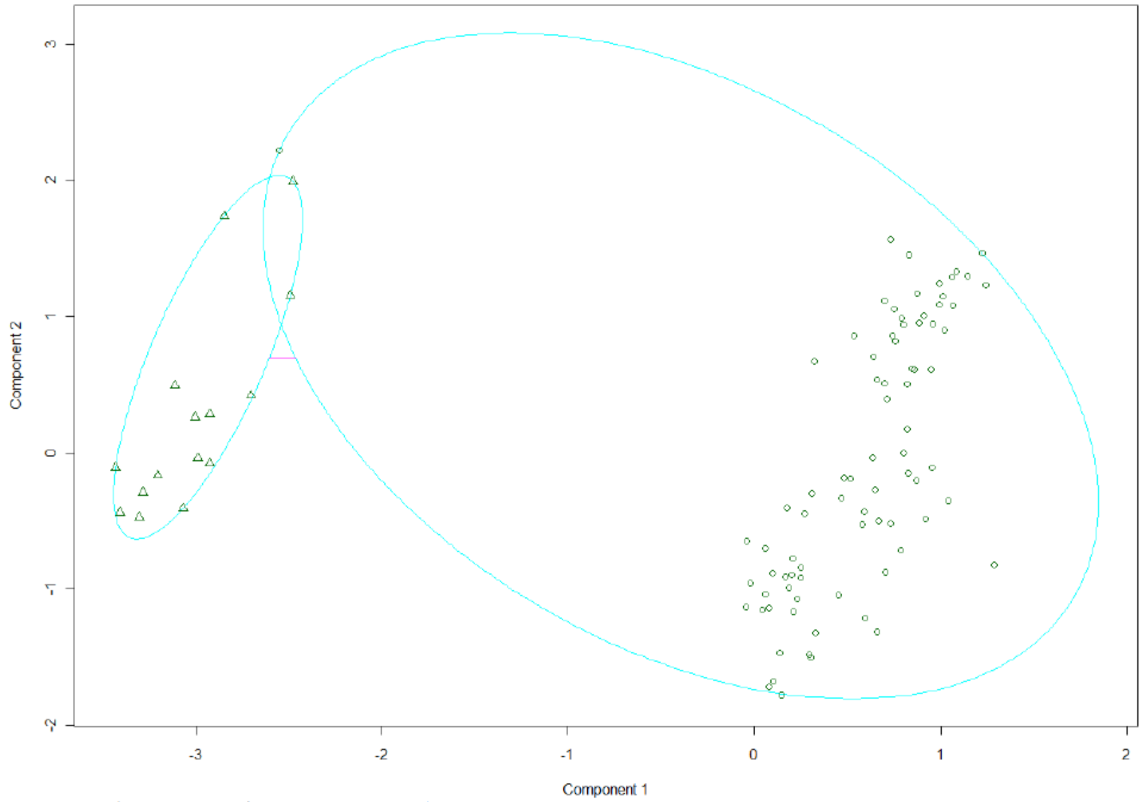

FIGURE 4.3: Cases clustered with PAM method

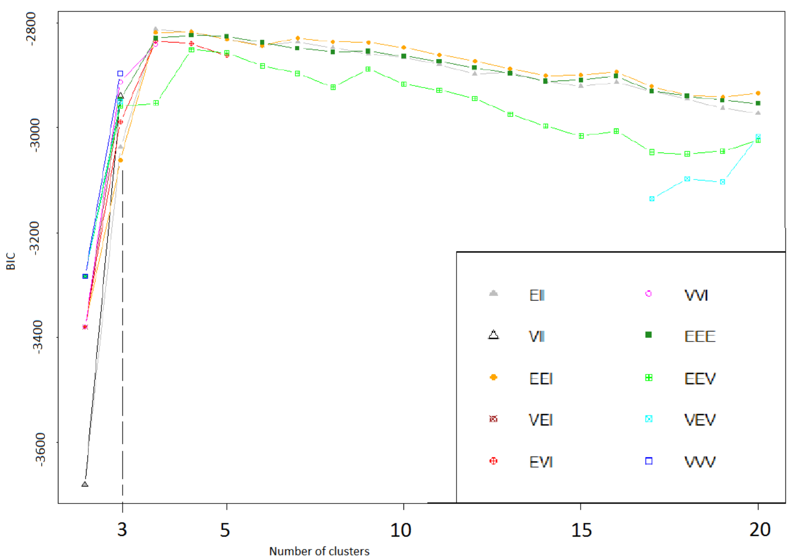

FIGURE 4.4: BIC for the 10 available model parametrisations and up to 20 clusters for our case base

The X axis represents the number of clusters and the Y axis their distribution in the space based on Bayesian probability using BIC.

All of the models suggested up to 20 clusters, however, Figure 4.4 shows that all models converge to a point, which suggests 3 to be the optimal number of clusters. Thus, model-based clustering with Bayesian probability has clustered cases into 3 different probability models.

4.5.3 Calinski-Harabasz Clustering Results

In Figure 4.5 the X axis represents the number of cases and the Y axis represents the number of clusters. Different colours represent different case clusters.

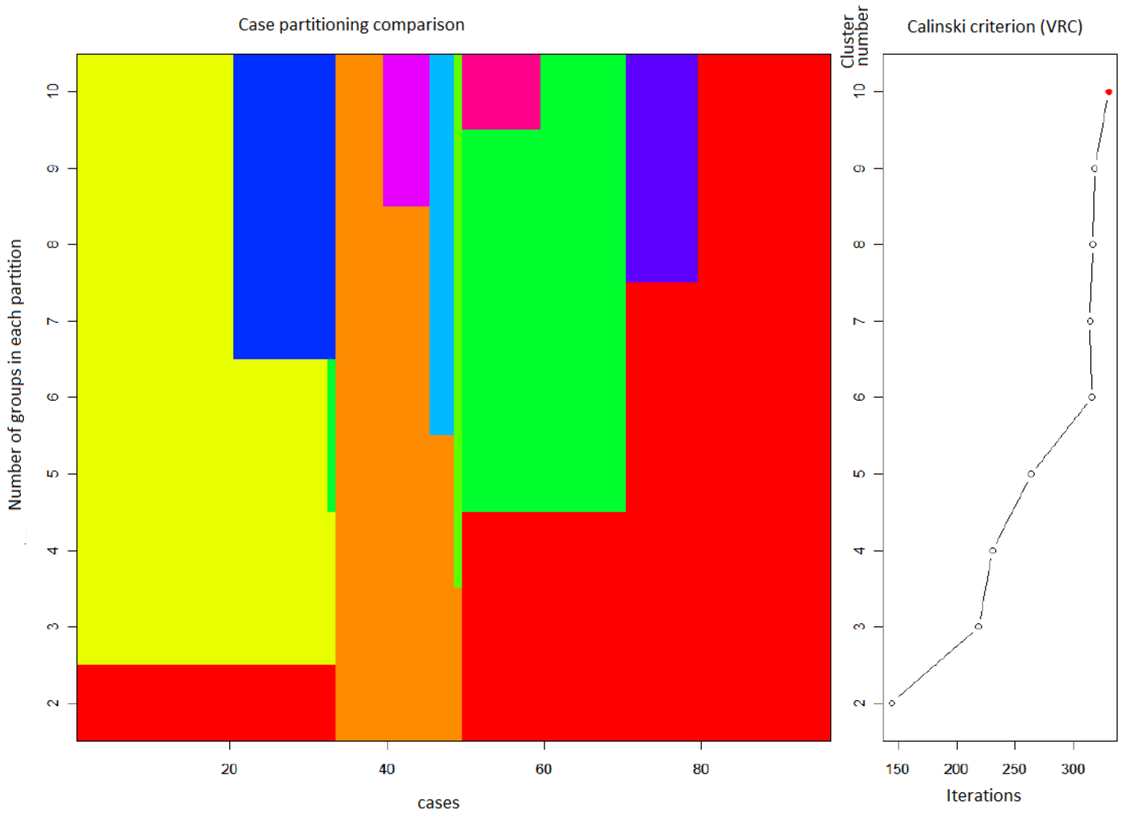

FIGURE 4.5: Calinski-Harabasz clustering analysis

Based on 300 runs the Calinski-Harabasz clustering analysis presented in Figure 4.5 suggests 10 clusters to be the optimal solution.

The plot of the cases grouped in 10 clusters can be seen in Figure 4.6. The principal components captured information on the data points well. The two components explain 93.9% of the point variability, which is the same as in the PAM method presented earlier.

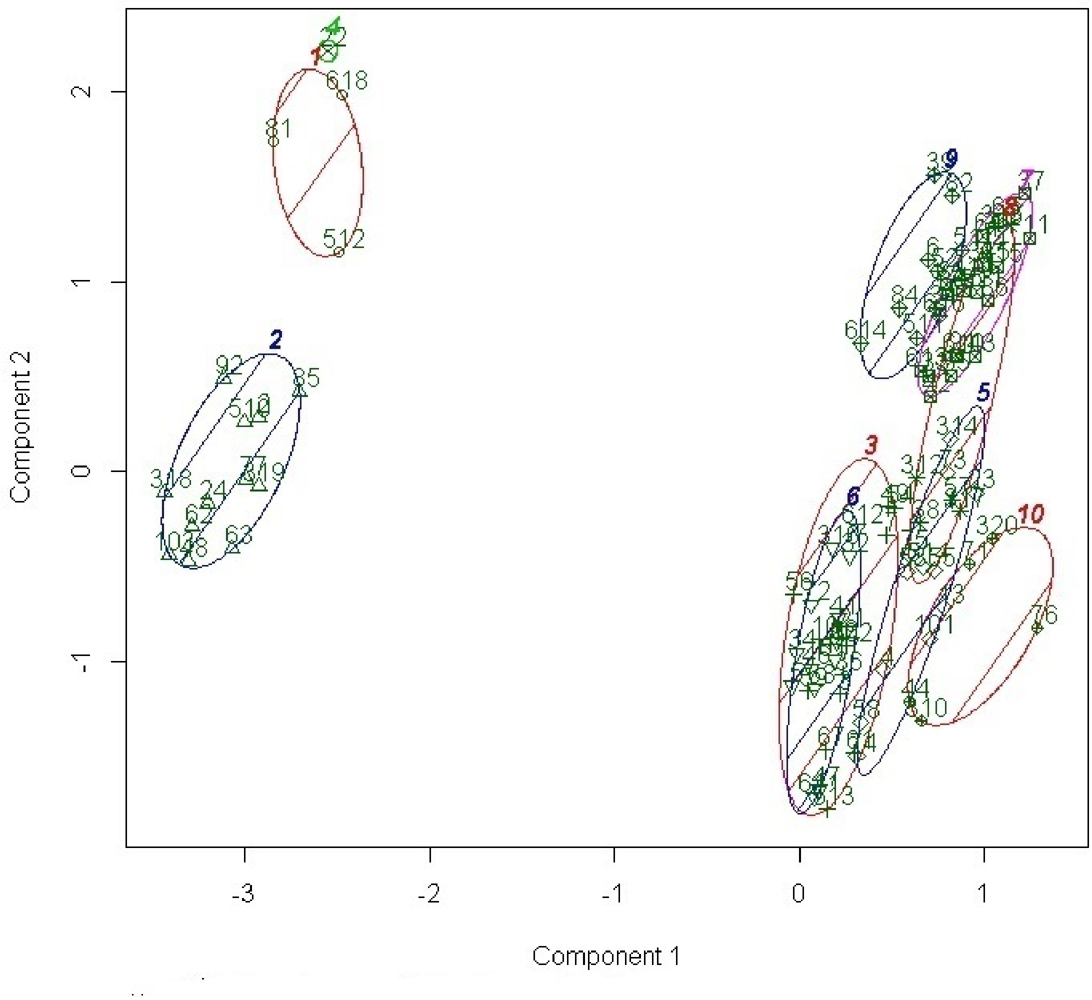

FIGURE 4.6: Cases clustered into 10 clusters suggested by Calinski-Harabasz presented in the principal plane

The suggested number of clusters from the three methods above, 2, 3 and 10, were used by the k-means algorithm to cluster cases by their tumour positions.

Additionally, based on the coordinate system described in Figure 4.1 we can identify 8 regions in the brain: (Left, posterior, superior), (Left, posterior, inferior), (Right, posterior, inferior), (Right, posterior, superior), (Left, anterior, inferior), (Right, anterior, superior), (Left, anterior, superior), (Right, anterior, inferior). Consequently, we can cluster our data set into 8 clusters. However, our data set might not be distributed evenly, e.g. in some clusters there might not be any cases, or all the cases might be grouped in only one cluster. Therefore, it is important to consider other numbers of k clusters. Clustering into 8 predefined clusters based on the geometry of a brain anatomical coordinate system given in Figure 4.1 was also considered.

4.6 Revised CBR Retrieval Results and Discussion

Experiments were carried out to assess the clustering of cases based on their tumour positions and the number of clusters, and how these affect the retrieval mechanism performance. The case base contains a total of 86 cases. All of these cases were clustered. However, in the CBR retrieval only 80 cases retrieve similar cases. The remaining 6 cases are different from other cases, e.g. have two right lens OAR defined or have all OARs missing and therefore none of the similar cases can be found and retrieved for them. Therefore, all experiments were run with these 80 cases. The experiments were run using 2, 3, 8 and 10 clusters. The leave-one-out cross validation method was used to measure the success rate of the CBR system.

The treatment planning software is not available to us, thus we cannot assess if the retrieved cases in the revised CBR system now have beams that are adaptable. The success rate of the CBR system is calculated by the beam number error (E_{BN}) and beam angle error (E_{BA}). Therefore, we evaluate the success of the revised CBR systems based on the improvement of the E_{BN} and E_{BA} that were achieved in the original CBR system.

Table 4.5 presents the comparison between the original and revised CBR systems. In addition to E_{BN} and E_{BA} we also measure the distance between the isocentre of the retrieved treatment plan and the known original treatment plan of the new case.

The Euclidean distance between the two isocentre points (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) is calculated using the formula below:

$$d(isocentre_1, isocentre_2) = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2} \qquad (4.7)$$

Table 4.5: Comparison between retrieval results of the original CBR system and the revised CBR system with clustering

CBR system	Average isocentre distances (mm)	Average E_{BN}	Average E_{BA}	Improvement of E_{BN} (absolute improvement)	Improvement of E_{BA} ($^\circ$, degree)
original CBR system	207.87	0.4	24.36	-	-
CBR with 2 clusters	107.5	0.4	24.1	0	0.26
CBR with 3 clusters	108.6	0.4	24.4	0	-0.04
CBR with 10 clusters	30.29	0.28	21.00	0.12	3.36
CBR with 8 clusters	31.26	0.35	26.99	0.05	-2.63

The revised CBR system with clustering retrieved treatment plans with decreased average isocentre distances between the retrieved treatment plans and the known treatment plans of the new cases, compared with the original CBR system.

The average $E_{BN} = 0.4$ and $E_{BA} = 24.36\,^\circ$ was obtained in the original CBR system. The revised CBR systems with 2 and 3 clusters retrieved the same cases as the original CBR system in 85% of cases. Both revised CBR systems had improved isocentres, from 207.87 mm in the original CBR system to 107.5 mm in the CBR system with 2 clusters, and 108.6 mm in the CBR system with 3 clusters. The average E_{BA} value has improved in the CBR system with 2 clusters, compared to the original CBR system, by $0.26\,^\circ$. However, while the revised CBR system with 3 clusters has improved the isocentre distance to the tumour position of the new case when compared to the original CBR system, the average E_{BA} has worsened by $0.04\,^\circ$. No improvement was noted in the E_{BN} value in the CBR system with 2 or 3 clusters compared to the original CBR system.

The revised CBR system with 8 clusters retrieved different cases from the original CBR system in 71% of cases. This CBR system has decreased isocentre distances between the retrieved case and the new case from 207.87 mm, in the original system, to 31.26 mm in the revised system. The average E_{BN} has improved by 0.05 (from 0.4 to 0.35). However, the average E_{BA} value has increased compared to the original CBR system by $2.63\,^\circ$ (from $24.36\,^\circ$ to $26.99\,^\circ$), which resulted in a lower success rate in the revised CBR system. From this we can observe that the improved E_{BN} in the revised CBR system with 8 clusters was obtained at the expense of E_{BA}. That is, retrieved treatment plans that suggested a better solution with respect to beam number had a worse solution with respect to beam angles compared to the original CBR system. In addition, another possible explanation for the worsened E_{BA} value is that it could be the trade-off between X, Y, Z coordinates; for example, a case is grouped into a cluster with a better position with respect to the Z coordinate rather than the X or Y coordinates.

The revised CBR system with 10 clusters retrieved different cases to the original CBR system in 80% of cases. It performed the best and retrieved cases with the isocentre distance decreased to an average of 30.29 mm. The average E_{BA} value has improved by 3.36 ° (from 24.36 ° to 21.0 °). The average E_{BN} has decreased by 0.12 (from 0.4 to 0.28).

The revised CBR system using clustering of cases retrieves cases whose treatment plans have more appropriate isocentres than the original system. However, while all of the revised CBR systems have improved isocentre distances compared to the original CBR system, the E_{BA} was not improved in all of them. Improvement in E_{BA} was noted in the CBR system with 2 and 10 clusters. The CBR system with 3 clusters did not improve on the original CBR system with respect to either E_{BN} or E_{BA}. In the CBR system with 8 clusters, improvement in E_{BN} was noted at the expense of E_{BA}. The revised CBR system with 10 clusters performed the best with respect to both the average E_{BN} and E_{BA}. Therefore, we determine 10 clusters to be an optimal number for clustering of the case base.

4.7 Conclusions

The experiments conducted have demonstrated that clustering of cases has improved the isocentre distances of the retrieved cases. However, the improved isocentre did not directly correlate with the improvement of the beam number or beam angle errors. The optimal number of clusters should be found, as it will identify the distribution of cases in the clusters and will affect the success rate of the whole system.

The revised CBR system with 10 clusters performed the best with respect to both the average beam angle and beam number error. The success rate of the revised CBR system with 10 clusters was improved by 0.05 with respect to the beam number error and by 3.36 ° with respect to beam angle error, compared to the original CBR system.

With the better retrieved treatment plan and with an appropriate isocentre, the beam angle adaptation should be easier to conduct.

In our future research the medical physicists will be involved in the validation of the CBR system with clustering.

The next chapter will introduce the adaptation approach for the CBR system using the knowledge available within the case base.

Chapter 5

Machine Learning-Based Adaptation

5.1 Introduction

In this chapter we investigate adaptation mechanisms to improve the results of the current CBR retrieval system. In Chapter 2 some of the existing approaches were presented that were used for adaptation phase. We learned that adaptation often requires a large amount of the domain specific knowledge, which can be difficult to acquire and often is not readily available. In this chapter, we investigate approaches to adaptation which do not require much of the domain knowledge. In particular, the knowledge already available in the case base is utilised for the adaptation of the beam number of the retrieved case.

The main contribution of the research presented in this chapter is the application of several machines learning tools for CBR adaptation. The presented methodology can be useful when there are no additional sources of the knowledge required for adaptation. The evaluation of the developed adaptation methods was performed using the cases from the available case base.

The chapter is organised as follows: first, an overview of the machine learning tools used for adaptation is presented. Next experimental results on real world brain cancer patient cases are reviewed, followed by conclusions.

5.2 Machine Learning Algorithms

Machine learning algorithms can resolve how to perform tasks by generalising from the examples, which are referred to as a training set. The training set consists of the input and output data. Machine learning algorithms can learn patterns and relations between the input and output and thus make predictions.

According to Domingos (2012) developing successful machine learning systems requires a lot of knowledge which is not readily available as the developed applications are often built particularly for the domain and the case base and thus, they are suitable only for the particular domains they were developed for. Hence, many machine learning projects take much longer than necessary or produce unsatisfactory results. Also, in the domains in which many independent input features correlate well with the output the learning is easy. However, if the output is a very complex function of the features, a machine learning application may not be able to learn from such data.

Machine learning is often used in the decision support systems and most commonly is used for classification applications. The advantages of using machine learning tools is their ease of application. There are a variety of machine learning tools available. Their performance varies depending on the domain. Therefore, when using machine learning tools it is suggested to choose and compare a few rather than basing evidence on one.

In the literature there are examples in which machine learning tools have been used in conjunction with CBR (Im and Park, 2007; Sarwar et al., 2010; Reyes et al., 2015). Bose and Mahapatra (2001) provided a review of comparison of several machine learning tools and CBR used in business data mining domain. CBR offered better explanation capability compared to the machine learning tools and in addition, was reported to be easy to operate. CBR adaptation usually takes into account the differences between a new case and the retrieved case. These differences are analysed to learn some knowledge or pattern that would help in adapting the retrieved case for the new case. Machine learning algorithms are used for pattern recognition, therefore they could be useful in learning the patterns or relations between the new case and retrieved case in our domain.

5.3 Machine Learning-based Adaptation

Prior to choosing machine learning algorithms we observe data available to us and identify the objective that needs to be achieved, e.g. classification of the data, regression analysis, density estimation. The choice of the algorithms is decided based on the objective. Adaptation usually considers the prominent differences between a new case and the retrieved similar case. Hence, we consider for our dataset case attribute values differences between the new and retrieved cases, used as input values, and respective beam number difference between them used as the output. For the input values total of seventeen attributes discussed in

Chapter 3 were used, therefore the dataset is highly dimensional as it is represented by many attributes.

In the adaptation we want to change the beam number of the retrieved case to match the new case. The cases in our case base have two, three or four number of beams. Therefore, an adaptation of the beam number suggested by the retrieved case is to add or subtract one or two beams or not to change it. Such an adaptation can be formulated as a classification problem. This motivated us to investigate machine learning algorithms as an adaptation tool in CBR. However, in our case base there is only one case with two beams. Experiments have shown that having only one case with two beams is not sufficient for training, leading to poor results. Therefore we consider only cases with three and four beams in the training phase. Consequently, output in the proposed adaptation methods contains three classes that correspond to three possible changes: 0 if no change is needed and 1 or -1 if the suggested beam number has to be increased or decreased by 1, respectively. However, the proposed adaptation methods are general and can handle any number of beams in the cases on the condition that there are a reasonable number of cases with that number of beams in the case base, so that enough cases are used in the training phase. In such settings, if we identify beam number change as (`BN_Change`), the output class will have the value `-BN_Change` or $+$`BN_Change` to denote the required change to the beam number of the retrieved case.

In this research, we considered three well-known machine learning tools: neural networks (NNs), the naive Bayes classifier and support vector machines (SVMs) for adaptation. The reason for choosing these particular algorithms is that they are popular choice for classification and reported to be successful in the literature (Kazmierska and Malicki, 2008; Almonayyes, 2006; Biswas et al., 2014). In addition, the chosen algorithms are different from each other, and therefore should provide more perspectives in learning the relations in the dataset. NNs estimate a function that depends on the large number of inputs. They consider all input attributes together versus the output. In radiotherapy planning the assumption is that all input attribute values interplay with each other in forming an output, therefore by using NNs we can test this assumption. Naive Bayes classifiers consider input values independently of each other and how they affect the output values. Medical physicists presumed that some attributes might be more important in defining an output than others in the radiotherapy planning. Therefore, by using the naive Bayes classifier we wanted to test this assumption and see if considering the attributes individually would affect the output results. SVMs work well in a high dimensional datasets which can be useful in our dataset as we have a fairly large number of inputs.

5.3.1 Neural Networks-based Adaptation Method

NNs are computational models that have been successfully used in machine learning in a variety of problem domains including classification. NNs were inspired by biological neural networks of the human brain where connections exchange messages between each other. Artificial NNs consist of interconnected input neurons, hidden neurons, which are organised in layers, and output neurons that serve as decision variables. Each of the layers can apply any function to the previous layer to produce an output. The hidden neurons transform input neurons using the provided function into the information that output neurons can use to define an output.

Connections between neurons are assigned adaptive weights, which are activated and modified during the training phase of the NN development. The learned weights are then used in the classification. The output of the network is computed by multiplying the input by the weight and passing the result through an activation function. The softmax activation function was used where each neuron received a vector of outputs from other neurons that activated, each connection with its own weighting. These are then linearly combined and used in the softmax function to determine if the next neuron activates or not. This way the softmax function can map a vector to a given output class. Neuron weights and the value of the activation function are adjusted during the training by using bias neurons, that are created by the NN during the training. These bias neurons provide a trainable constant value to the neurons and therefore help to control the layer behaviour. Such training models enable the NN to discover the patterns in the problem domain.

NNs cannot produce logic-like rules because it is impossible to interpret the distribution of connection weights in the network in the form of *if-then* rules (Liu and Yan, 1997). The classification function in an NN is buried within the network topology and the connection weights. Thus, the classification logic is not obvious to the end-user.

Bose and Mahapatra (2001) reviewed that NNs were less efficient in processing large datasets and required the user to possess substantial tool knowledge to set up and operate the system. However, the predictive accuracy of the NNs was reported to be better compared to CBR.

NNs have been used in the retrieval phase of CBR. For example, Shin et al. (2000) and Im and Park (2007) used NNs to determine the weights of the features in the CBR retrieval. They reported that the CBR retrieval with the feature weights generated with the NN performed better than the retrieval without the predefined weights. However, to our knowledge, there are only two approaches that employ NNs in adaptation. In the medical domain, Henriet et al. (2014) developed an adaptation tool in a CBR system for the numerical representation of human organs to make a phantom in 3D using an NN model. The retrieved solution was suggested to the experts, who decided whether to use it or whether a new solution was required. To adapt the case, experts passed the input descriptors to the CBR system, and the NN was used to adapt the retrieved solution using these descriptors. The research suggested that when the adaptation was guided by experts, the obtained results were well-suited for the new cases. Corchardo et al. (1998) retrieved a few similar cases during the CBR retrieval phase and then used NNs for weighting of these cases in the adaptation phase to determine the best solution in the oceanographic water temperature forecasting. They reported that the CBR system performed better with the NN-based adaptation compared to the original CBR system.

The aim of the developed NN is to learn how the differences in attribute values of the retrieved and new cases affect the outputs of these two cases. In the training phase, we use the beam number data recorded in the cases. Given a pair of training cases, the NN is trained to discover the pattern of differences between the attributes of two cases and the difference in their beam numbers. Once the NN is trained, it is used to suggest the adaptation of the beam number of the retrieved case based on the differences between the attributes of the new and retrieved cases. The input neurons contain the attribute differences between the new and the retrieved cases, while the output neuron is the class that denotes the required change to the beam number of the retrieved case.

The NNs can be designed with a single layer or multiple-layers. Single layer NNs have input neurons directly connected to the output neurons which can be too simple for many problems. The single layer models cannot produce the non-linearity that can be obtained by multiple layers, each layer in a multi-layer NN adds its own level of non-linearity, hence multiple-layer NNs have superior performance. For that reason, supervised multi-layer feed-forward NNs that are used for prediction are implemented here. In feed-forward NNs each subsequent layer has connection from the previous layer. This is a simple yet effective type of the NN.

A feed-forward network with one hidden layer and enough neurons, can fit any finite problem therefore, it is considered as the most popular architecture among NNs (Meyer et al., 2003). However, in this research we consider NNs with both one and two hidden layers, as the latter work well for complex domains.

5.3.2 Naive Bayes Classifier Adaptation Method

Naive Bayes classifier belongs to a group of probabilistic classifiers. A classifier can be defined by a function f that maps input attribute vectors $x \in X$ to the output class labels $y \in 1, ..., C$, where X denotes an attribute range and C is a class of an outcome. The probability p of y outcome based on the evidence x is calculated as follows (Murphy, 2006):

$$p(x|y = c) = \prod_{i=1}^{D} p(x_i|y = C) \tag{5.1}$$

Here, the attribute vector contains D real numbers.

Naive Bayes classifier is often used when attributes are assumed to be independent of each other, i.e., the value of a particular attribute determines a class independently of the presence or absence of any other attribute (Friedman et al., 1997).

Even though the independence assumption is often false, as in many domains attributes are either dependant or interact between each other in some way, the naive Bayes classifier works well in various fields such as data mining (Schultz et al., 2001), text classification (Kibriya et al., 2005) and clinical evaluation (Kazmierska and Malicki, 2008). Zhang (2004) proposed an explanation of the classification of the naive Bayes classifier, suggesting that dependence distribution between attributes, being consistent (support a certain classification) or inconsistent (cancel one another) plays a crucial role in determining the output. Therefore, even when the independence assumption is violated and there is a strong dependence between attributes the naive Bayes classifier can still be a good classifier.

Despite the simplicity of the naive Bayes classifier, there are examples of its successful use in practice and also in CBR systems. For example, Yeow et al. (2013) used it in a CBR system to learn weights of the case attributes in the retrieval phase. Almonayyes (2006) employed CBR and the naive Bayes classifier to analyse and classify Arabic texts. The CBR system consisted of the retrieval only where retrieved texts were classified with the naive Bayes classifier.

Experiments have shown that in RTP, some attributes play a more considerable role than others in determining the output. Therefore, we investigate if the naive Bayes classifier can detect such attributes, i.e. whether considering each attribute difference value separately allows the classifier to correctly suggest the change in the beam number. To our knowledge this approach is the first attempt to incorporate the naive Bayes classifier into the adaptation phase of the CBR.

The naive Bayes classifier method is different from NN, where all attributes differences are considered together as one input, and their overall effect on the outcome is observed. Furthermore, naive Bayes classifiers are considered to be particularly suitable when the range of the input attribute values is highly varied, which is characteristic of radiotherapy cases.

The naive Bayes classifier classifies each case in the input training set and then calculates the prior probabilities of class membership. The prior probabilities are recalculated with respect to the actual output, that is the beam number of the training set. A new set of testing cases is classified based on the probability values calculated by Expression (5.1), where the class with the highest probability is selected.

The method is applied in a similar experimental setting as described above for the NN. The naive Bayes classifier uses the input training set and classifies each case in the classes -1, 0 and $+1$, which have the same meaning as in NN.

5.3.3 Support Vector Machines Adaptation Method

SVM is a type of linear classifier that uses kernels to separate data in a high dimensional feature space to improve classification. A kernel is an algorithm for pattern analysis used in SVM. Opposite to the traditional machine learning algorithms such as NNs, which aim to minimise the empirical training error, an SVM constructs a hyperplane as a decision plane that separates classes with the margin and tries to minimise an upper bound of generalisation error between the hyperplane and data. The benefit of it is that it can capture complex relationships between the data points without having to perform difficult transformations. In addition to performing linear classification, SVM can efficiently perform a non-linear classification. In the case when data is not linearly separable the linear classifier fails to find a good decision function and has a low performance, thus a non-linear classifier is better in these circumstances. The disadvantage of the non-linear classifier is that it has a risk of poor generalisation (overfitting). SVM is considered one of the best in its predictive capability and therefore it is a popular machine learning algorithm (Byun and Lee, 2002).

SVM implementation requires the definition of a kernel function. The Gaussian radial-basis function (RBF) and Laplace kernels are general-purpose kernels used when there is no prior knowledge about the data and therefore they are appropriate choice for non-linear domains. The Gaussian RBF function is defined as follows:

$$k(x, x') = \exp(-\sigma||x - x'||^2) \tag{5.2}$$

and Laplace kernel as follows:

$$k(x, x') = \exp(-\sigma||x - x'||) \tag{5.3}$$

These kernels were used in our experiments. In Expression (5.2) and (5.3) $k(x, x')$ - is a kernel function that calculates the inner product of two vectors x and x', σ - is a tuning parameter identified by the system during the training phase which prevents data from overfitting or underfitting (Karatzoglou et al., 2006).

Apart from the kernel function, another parameter that has to be set is a cost parameter C, that is used in the model to penalise misclassification. If C is very small, the classifier will look for a large-boundary hyperplanes which can include misclassified points. On the contrary, larger value of C will lead to a smaller-boundary hyperplane that classifies all the training points correctly. The hyperplane is adjusted when incorrectly classified samples are found, and the algorithm stops when the hyperplane is optimal.

SVM-based adaptation is applied in the similar experimental settings as the two previously described methods.

5.4 Results

All the machine learning methodologies described earlier were validated. The case base used for the evaluation of the proposed adaptation methods consists of 80 cases collected from the Nottingham City Hospital. Out of 80 cases, 56 were treated with 4 beams, 23 cases with 3 beams and 1 case with 2 beams.

5.4.1 Validation Criteria

When getting results it is important to evaluate and generalise them beyond the examples of the training set. For this purpose, the machine learning-based system is evaluated based on how it performs on an unseen test set.

The case base is divided into a training and testing set by a ratio commonly used for a machine learning tool, typically 80%/20% or 70%/30%. The sizes of the training/test sets are adjusted during the experiments based on the observations of the results from the experiments. For example, the size of the training set is increased in some instances in order to improve the performance of the algorithm.

Each case is assigned randomly into a training or testing set. In every new run the dataset is again split randomly into training and testing sets. The result obtained would differ, and that is called a model variance. In order to reduce the variance of the random process the experiments are repeated multiple times. In smaller datasets, such as our case base, the results are collected from a fair number of runs, e.g. 50-100 runs per experiment, and the average result is calculated. This provides an estimate of the performance of the algorithm on the dataset (Brownlee, 2014). This approach is used for validation in our experiments.

In such random splits it is possible that some data instances are never included for training or testing, whereas others may be selected multiple times. The effect is that this may skew the results and may not give a precise idea of the accuracy of the algorithm. This is especially true in the small size datasets. In order to ensure that each instance is used for both training and testing an equal number of times a k-fold cross validation is carried out. In k-fold cross validation, k is the number of equally sized splits in the dataset. In each new run, one of the subsets k is made the test or validation set and the other (k-1) subsets constitute the training data. Each time the algorithm is run it changes k set of the data the algorithm was tested on previously. The results from each fold are then averaged over k times. This way we ensure all of the cases are included in both training and test sets. We use 10 folds cross validation, where our dataset is split into 10 folds of 8 cases each, thus, 9 folds, 90% of cases, are used for the training set and 1 fold, 10% of cases, are used for the testing set in every new run.

5.4.2 Random Guess

Prior to using machine-learning tools, random guessing was carried out to suggest the beam number change. In random guessing the suggested beam number change is randomly assigned from the dataset. Random guessing accuracy depends on distribution of the beam number in the dataset. The majority of cases in our case base, 70%, have four beams which means that our output classes will not be uniformly distributed suggesting no change, class 0, most of the time. An average success rate of 38% on 100 runs was achieved. A classifier can be considered accurate if it has an error rate better than random guessing on an unseen examples (Dietterich, 2000).

5.4.3 NN-based Adaptation

We make use of the *neuralnet* and *nnet* packages available in *R* for the experiments. Based on suggestions for the NN training found in the literature (Ripley, 1996), and also our empirical results, the NN was trained with 80% of cases (65 cases), while 20% of cases (15 cases) were used for testing. The larger training set ensures that the NN model can capture the generalisation more efficiently, thus making a better prediction for the testing set.

We ran experiments with one and two layers and a different number of neurons for each layer, ranging from 9 to 34 to determine an optimal NN combination for the best classification performance. This range was determined based on the number of input (seventeen attribute differences) and output (beam number change) neurons and should provide good results in the testing set (Heaton, 2008). Each of the NN models, with the defined number of layers and nodes was run 50 times. In each run, the training and testing set cases were chosen arbitrarily from the case base. This ensured that a large number of combinations of cases was considered.

The results obtained with different numbers of neurons in the NN models with the single hidden layer can be seen in Table 5.1. The results reported in the table were obtained by using *nnet* package which is designed specifically for a single hidden layer experiments. The experiments have shown that the single hidden layer with 15 neurons produced the best result. With 15 hidden neurons, the suggested beam number change was correct in 73.3% of the test cases.

NN hidden nodes	14	15	16	17	21
Correct *BN* change (in 15 test cases)	8	11	7	9	6
Average accuracy (%)	53.3	73.3	46.7	60	40

TABLE 5.1: NN results with different number of hidden neurons

Figure 5.1 provides a visualisation plot of the developed NN model. The developed model contains input neurons *I1 - I17*, hidden neurons *H1 - H15* and output neurons *O1 - O3* with the connections. *B1, B2* represent bias neurons. The weights in black lines are added and in grey lines are subtracted in the activation function; the thicker the line the more weight it has.

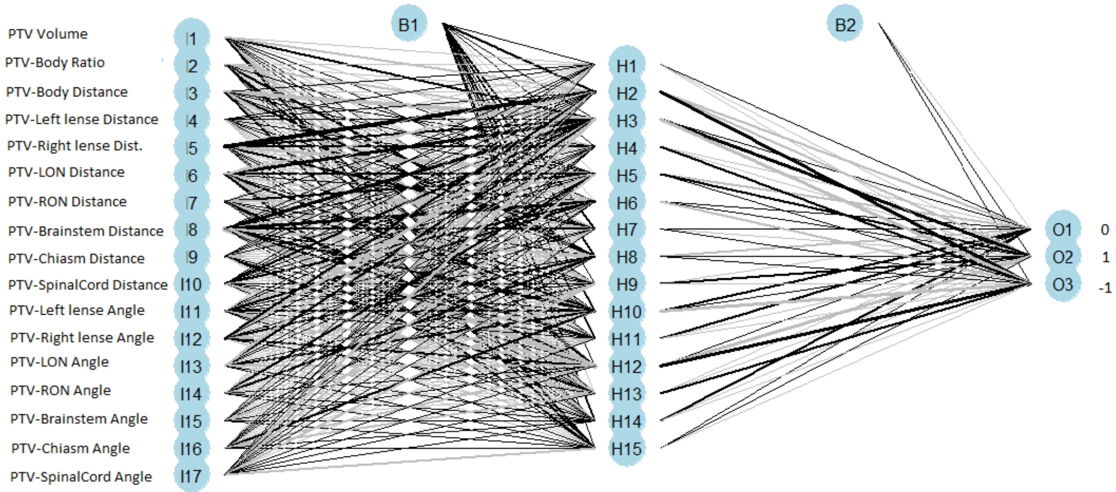

FIGURE 5.1: A plot of the developed neural network with 15 hidden nodes

The initial retrieval success rate of the CBR without adaptation was 61%. Including NNs increased it to 73.3%, which demonstrates that NNs can be useful for beam number adaptation.

5.4.4 Adaptation Based on the Naive Bayes Classifier

In these experiments we make use of the *e1071* package available in *R*. The naive Bayes classifier was trained with the three different training sets. First, it was trained on 2/3 of the cases, i.e. 66% of the case base (54 cases) and tested on the remaining 34% of the cases (26 cases) based on the generally accepted observation that an advantage of this classifier is that it only requires a small amount of training data (Bishop, 1995). Then the classifier was trained with a larger number of cases, i.e. 90% of the case base, in the hope of learning the pattern better. Finally, 10 fold cross validation was used. All experiments were run 100 times. Mean values, standard deviations and average success rates across all models were compared. The obtained results are shown in Table 5.2.

From Table 5.2 we can see that although the maximum success rate of the naive

	Training set		
Performance	**66%, 54 cases**	**90%, 72 cases**	**10 fold cross validation, 8 cases in each fold**
Standard deviation	16.7	20.6	22
Maximum success rate	76.9	100%	75%
Minimum success rate	15.3	12.5%	12%
Average success rate	47.3%	55.3%	56%

TABLE 5.2: Naive Bayes classifier-based adaptation success rate results

Bayes classifier is higher than that of the original CBR retrieval phase, 61%, the minimum success rate of the naive Bayes classifier, 12% is worse than of the original CBR retrieval. This shows that our data has high variance (imbalanced). Variance represents the spread of the data and high variance demonstrates that the values of the attributes are very spread out with respect to each other. The high value of the standard deviation value obtained in the Table 5.2 also indicates that our dataset is spread out over a wide range of values. The standard deviation also demonstrates how stable are the results. In all experimental settings obtained the standard deviation was large which shows that the variation between the obtained results is large, therefore the performance of the algorithm is not robust. A larger training set has improved the average success rate of the naive Bayes classifier as it makes the classification more representative and accurate. Naive Bayes classifier-based adaptation method with 10 fold cross validation has better average success rate (56%) compared to other models. However, there is still a big difference between the maximum and minimum success rates which shows that possibly naive Bayes classifier could not handle our data with high variance.

5.4.5 Adaptation Based on SVM

For SVM classification we use *Kernlab* package in *R*. Initially, the SVM adaptation model was trained with 80% training set and tested with the remaining 20%. We also experimented with using a larger training set, 90% of cases and 10 fold cross validation, similar to the naive Bayes classifier presented earlier.

SVM models based on two kernels were implemented. Table 5.3 presents results for the SVM with Gaussian RBF and $C = 10$. The value of C is chosen as an optimal value for our model based on a number of runs. The results obtained with the different values of C are presented in Table 5.4.

	Training set		
Performance	**80%, 64 cases**	**90%, 72 cases**	**10 fold cross validation, 8 cases in each fold**
Standard deviation	9.8	15.4	19.7
Maximum success rate	81.25	87.5%	87.5%
Minimum success rate	31.25	25%	25%
Average success rate	53.4%	55.25%	53.75%

Table 5.3: Results obtained by the SVM adaptation method with Gaussian RBF, C=10

It is can be noted that on average SVM performed similar to the naive Bayes classifier, 55.25% versus 56% by the naive Bayes classifier. However, it can be seen from the table that the minimum success rate in any SVM trained model (25%-31.25%) is better than in the naive Bayes classifier-based models (12%-15.3%), which shows that it handled better the dimensionality of the case attributes. In addition, the maximum success rate is more balanced and similar across the tested models in SVM compared to the naive Bayes classifier, where it is quite varied in different trained models. This could be credited to the better performance of the SVM in a high-dimensional dataset. The standard deviation value has slightly improved in the experimental settings with 64 and 72 cases, however is still high. The average success rate of 55.25% was obtained with the SVM-based adaptation which is lower compared to the original CBR system. Further from the table we learn that the SVM performance improves when the training set size increases. Training sets with 90% of data and 10 fold cross validation both performed better than the training set with 80% of data.

Table 5.4 presents results from setting C to different values: small and high.

	Training set			
Performance	**90%, 72 cases; C=0.1**	**90%, 72 cases; C=1**	**90%, 72 cases; C=0.05**	**90%, 72 cases; C=50**
Standard deviation	18.04	17.5	18.17	16.95
Maximum success rate	100	87.5%	87.5%	87.5%
Minimum success rate	12.5	12.5%	0%	12.5%
Average success rate	53%	52.75%	51.6%	51.125%

Table 5.4: Results obtained by the SVM adaptation method with Gaussian RBF, different values for C

It can be observed that with the smaller penalty parameter C, the success rate of the SVM classifier worsens, namely the minimum success rate is 0% for C = 0.05. The average success rate of both the small C ($C = 10$) and the large C ($C = 50$) are worse compared to the success rate with the $C = 10$ in all models.

SVM with Laplace kernel performed slightly worse compared to the SVM with Gaussian RBF kernel. Table 5.5 presents the results with the 80% and 90% training sets.

	Training set	
Performance	**80%, 64 cases**	**90%, 72 cases**
Standard deviation	9.7	17.6
Maximum success rate	87.25%	100%
Minimum success rate	31.25%	12.5%
Average success rate	50.43%	53.125%

TABLE 5.5: Results obtained by SVM adaptation method with Laplace kernel, C=10

In the similar experimental setting, the success rate of the SVM models with Laplace kernel are lower, 50.43% and 53.125%, compared to the 53.4% and 55.25% achieved in the SVM model with Gaussian RBF kernel.

5.5 Conclusions

In this chapter, we proposed a machine learning-based adaptation approach. The obtained empirical results demonstrate that the proposed adaptation method based on neural networks improved the performance of the existing CBR system.

The adaptation based on neural networks improved the success rate by 12%. This method could also adapt the beam number by stating the direction of the change, i.e. either increase or decrease the number of beams retrieved. The average success rates of the naive Bayes classifier-based and support vector machine-based adaptations were lower compared to the CBR retrieval, 56% and 55.25% respectively. The naive Bayes classifier did not improve the current retrieval results which indicates that the interplay among the attributes needs to be considered. From the experiments it was revealed that support vector machine-based adaptation handled better the dimensionality of the case attributes compared to the naive Bayes classifier-based adaptation. We learnt that the high dimensionality and variance of our dataset cannot be handled by some machine learning algorithms.

The beam angle adaptation is formulated as a more complex optimisation problem in terms of the decision variables to consider. Contrary to the beam number which is represented as a discrete integer number, the beam angles are in the range of 0° - 360° and can take any real value, which makes it difficult to define them as a classification problem. Therefore, in this part of the research we do not apply machine learning algorithms to beam angle adaptation.

Next chapter will introduce a knowledge-light adaptation-guided retrieval method which will be used for both beam number and beam angle adaptation.

Chapter 6

Adaptation-Guided Retrieval of Beam Number and Beam Angle

6.1 Introduction

The previous chapter introduced a machine learning-based adaptation approach that utilised knowledge available in the case base. An adaptation-guided retrieval approach for adaptation of the beam number and beam angle of the retrieved case will be presented in this chapter. Similarly to the method described in the previous chapter this method also acquires knowledge from the case base but without interaction with medical physicists. This method was chosen as it can provide useful information on the domain and the case base during adaptation, in particular, it utilises a pair of cases with known treatment plans to help adapt the retrieved case to be applicable to the new case. The proposed adaptation method intertwines with the retrieval phase. Furthermore, it turned out that that the most similar case is not necessarily the best to adapt. The evaluation of the developed adaptation method was performed using cases from the available case base as described in the previous chapter. The obtained empirical results demonstrate that the proposed adaptation method considerably improves the performance of the existing CBR system.

The implemented adaptation-guided retrieval method is described first followed by experimental results and conclusions.

6.2 Adaptation-Guided Retrieval Methodology

The adaptation-guided retrieval method was developed and implemented first for beam number adaptation. The experiments showed promising results; hence the method was further extended and reformulated for beam angle adaptation.

6.2.1 Adaptation-Guided Retrieval of Beam Number

Many real-world problems do not exhibit global linearity in the sense that differences in attribute values have different impacts on the output in different regions of the domain space. RTP belongs to such a class of non-linear domains. The knowledge-light approach to adaptation introduced by McDonnell and Cunningham (2006), which was developed for non-linear domains, has been modified to fit the RTP problem.

The standard CBR practice would be to retrieve a case from the case base that is the most similar to the new case and afterwards adapt the beam number in the retrieved case to address the differences between the retrieved and the new case. However, in our retrieval, we consider adaptation as well. The aim of the developed adaptation-guided retrieval of beam number (AGRBN) is to retrieve cases from the case base that suggest a correct change in the beam number of the retrieved case for treating a new patient. This approach assumes some degree of local linearity (i.e., in a small region within which local linearity is assumed), and also that the similar differences in attribute values result in similar differences in the output value. Apart from searching the case base for a case most similar to the new case, the retrieval algorithm searches for a pair of cases whose attribute differences have similar impacts on the beam number as the attribute differences between the new and retrieved case. The difference in the beam number of the treatment plans for that pair of cases is used to adapt the beam number suggested for the new case by the retrieved case.

Let C_n denote the new case and C_r denote the case retrieved by the CBR system. In order to retrieve an appropriate case C_r from the case base, which can be adapted to meet the requirements of the new case, the retrieval algorithm searches for three cases (C_r, C_1, C_2) such that the impact of the attribute differences between C_1 and C_2 on the beam number (output) of C_1 and C_2 is similar to the impact of the attribute differences between C_n and C_r on their beam numbers. This approach is different to the one presented in (Mishra et al., 2009), where the most similar case is retrieved, and the retrieval then searches for a pair of cases whose difference in attribute values is the most similar to the difference between the new and retrieved case.

Two vectors are defined to be used in the knowledge-light adaptation: the case differences vector, which represents the differences between the corresponding attributes of the new case and a retrieved case, and the case gradient vector, which represents the rate of change of the output with respect to the attribute differences. Their mathematical formulation is given below.

Each element of the case differences vector, $\triangle(C_n,\ C_r)$, represents the difference between the values of attribute l, l=1,...,6 for attributes A-the angle between PTV and OAR, E -the distance between the edge of the PTV and OAR, Vol-PTV volume, R-the ratio of PTV volume to the volume of the body, Dt-the distance between the body and PTV and P-the relative position of the PTV to OAR, respectively:

$$\triangle_l(C_n, C_r) = V_{n_l} - V_{r_l}, \quad l = 1, 2, 3, 4, 5, 6 \tag{6.1}$$

where V_{n_l} and V_{r_l} are values of attribute l in cases C_n and C_r, respectively.

The case differences vector is defined as the vector of all attribute differences:

$$\triangle(C_n, C_r) = [\triangle_1(C_n, C_r), \triangle_2(C_n, C_r), \triangle_3(C_n, C_r), \triangle_4(C_n, C_r), \triangle_5(C_n, C_r), \triangle_6(C_n, C_r)] \tag{6.2}$$

Let Y be the output of function f, which depends on the attributes $a_1,\ a_2,...,a_m$. The gradient of function $Y = f(a_1, a_2, ..., a_m)$ is defined as the rate of change of output Y with respect to input attributes $a_1,\ a_2,\ ...,a_m$, as shown in Expression (6.3):

$$\nabla f = (\frac{\partial Y}{\partial a_1}, \frac{\partial Y}{\partial a_2}, ... \frac{\partial Y}{\partial a_m}) \tag{6.3}$$

In our adaptation procedure, function f describes the change in beam number (BN), and its analytical form is not known. If we denote C_{sim} as the case that is the most similar to case C with beam number BN_{Csim}, retrieved using Expression (3.2), then the gradient vector of case C with beam number BN_C is calculated as follows:

$$\nabla C = [\frac{\partial BN}{\partial V_1}, ..., \frac{\partial BN}{\partial V_6}] = [\frac{BN_C - BN_{C_{sim}}}{V_{C_1} - V_{C_{sim1}}}, ..., \frac{BN_C - BN_{C_{sim}}}{V_{C_6} - V_{C_{sim6}}}] \tag{6.4}$$

This is a general formula for the gradient of case C. This will be used to calculate the gradient of cases C_r and C_2.

The purpose of using the gradient is to investigate how large the change in the attributes has to be in a given region of the case base to cause the beam number to change. In our domain, the majority of cases have three or four beams. Consequently, the most similar case retrieved can have the same beam number as the case for which the gradient is calculated, resulting in a value of 0 for the gradient. Therefore, the case retrieved for the gradient calculation should have a different beam number. If a small difference in attribute values triggers the change in the beam number, then the gradient is large; otherwise, it is small. The retrieval algorithm searches the cases sorted by their similarity with a case for which gradient is calculated until the most similar case with a different beam number is found.

The retrieval can be stated as an optimisation problem. The task is to find a set of three cases to be used in the adaptation, (C_r, C_1, C_2), such that the sum of the difference between C_r and C_n, the difference between the product of attribute differences between C_1 and C_2, the gradient of C_2, the product of attribute differences between C_n and C_r, and the gradient of C_r is minimised:

$$R_{C_r,C_1,C_2} = d(C_n, C_r) + d(R_{C_n,C_r}, R_{C_1,C_2}) \tag{6.5}$$

where

$$R_{C_n,C_r} = \triangle(C_n, C_r) * \nabla C_r \tag{6.6}$$

and

$$R_{C_1,C_2} = \triangle(C_1, C_2) * \nabla C_2. \tag{6.7}$$

The pseudo-code for the retrieval of the triplet of cases (C_r, C_1, C_2) is given in Algorithm 1.

Algorithm 1 Algorithm for retrieval of the set of three cases for beam number adaptation

BEGIN

$M_{NxN} = [\triangle(C_j, C_k)]$ holds the case differences vectors of cases C_j, and C_k in the case base

$m_{j,k} = \triangle(C_j, C_k)$ calculate using Expression (6.2)

$R_{min} = 0$

for $r = 1, ..., N$ **do**

  $d_1 = d(C_n, C_r)$ calculate using Expression (3.1)

  $R_{11} = \triangle(C_n, C_r)$ calculate using Expression (6.2)

  $R_{12} = \nabla C_r$ calculate using Expression (6.4)

  $R_1 = R_{11} * R_{12}$ calculate using Expression (6.6)

  for $r_1 = 1, ..., N \quad r_1 \neq r$ **do**

    for $r_2 = 1, ..., N \quad r_2 \neq r_1 \neq r$ **do**

      $R_{21} = \triangle(C_{r_1}, C_{r_2})$ (take from matrix M)

      $R_{22} = \nabla C_{r_2}$ calculate using Expression (6.4)

      $R_2 = R_{21} * R_{22}$ calculate using Expression (6.7)

      $d_2 = d(R_1, R_2)$ calculate using Expression (3.1)

      $R = d_1 + d_2$ calculate using Expression (6.5)

      if $R <= R_{min}$ **then**

        $R_{min} = R$

      end if

    end for

  end for

  Output triplet (C_r, C_{r_1}, C_{r_2}) for which $R_{C_r,C_{r_1},C_{r_2}}$ has minimum value R_{min}

  $C_1 = C_{r_1}, C_2 = C_{r_2}$

end for

END

First, the algorithm calculates the case differences vector $\triangle(C_j, C_k)$ for each pair of cases (C_j, C_k) and forms a matrix $M = [\triangle(C_j, C_k)]_{\mathrm{NxN}}$. The algorithm searches for a pair (C_1, C_2) for which the relation between the case differences vector $\triangle(C_1, C_2)$ and gradient of C_2 is similar to the case differences vector $\triangle(C_n, C_r)$ and gradient of C_r, respectively (similarity of these values is identified by the remaining lines of the algorithm).

Two different situations can occur in which the considered pair of cases (C_1,C_2) yields an appropriate change to the beam number of C_r. First, there is a good match between the case differences vectors $\triangle(C_n, C_r)$ and $\triangle(C_1, C_2)$; that is, both case differences vectors have small or large values, and both gradients ∇C_r and ∇C_2 have either large or small values. This implies that C_1 and C_2 are in the region of similar characteristics as C_n and C_r, and, consequently, the relation between the beam numbers of C_n and C_r should be the same as that between C_1 and C_2. The second situation occurs when the case differences vector $\triangle(C_1, C_2)$ does not match $\triangle(C_n, C_r)$ well; i.e., $\triangle(C_n, C_r)$ is small, while $\triangle(C_1, C_2)$ is large or vice versa. If the corresponding differences vector and gradient offset each other, the pair (C_1,C_2) can then be used to change the beam number of C_r. For example, if $\triangle(C_n, C_r)$ is small, and the gradient of C_r is large, a small difference in attribute values triggers the change in BN. If $\triangle(C_1, C_2)$ is large, and the gradient of C_2 is small, a large difference in attribute values is needed to cause a change in BN. This means that even, if the differences between the attribute values of C_n and C_r and between C_1 and C_2 do not match, if the gradients of C_r and C_2 are opposite (small versus large) they affect the BN change in the corresponding regions of the domain space accordingly.

In the case base, not all OARs are present in all patient cases. The absence of a certain OAR in a case means that no radiation risk is imposed for that particular OAR (i.e., it is too far from the tumour to consider). Note that in the search for C_r, only the cases that have the same OAR as the new case are considered. Similarly, the (C_1,C_2) pair should have the same OARs. However, (C_1,C_2) and (C_n,C_r) pairs do not need to have the same OARs.

The beam number of the retrieved case is adapted based on the beam numbers of cases C_1 and C_2 in the obtained triplet of cases. If cases C_1 and C_2 have the same beam number in their treatment plans, then the beam number in the treatment plan of the retrieved case C_r is not changed. If C_2 has a different number of beams than C_1, the beam number of the retrieved case is then changed by adding or subtracting a beam accordingly. The formula for adaptation is as follows:

$$BN_{C_r} = BN_{C_r} + (BN_{C_2} - BN_{C_1}) \tag{6.8}$$

Examples of the retrieved beam number adaptation for the new case based on the difference of beam numbers in C_1 and C_2 is given in Table 6.1.

In Adaptation 1, the retrieved case C_r suggests four beams. Since the difference in

Case	BN(adaptation 1)	BN (adaptation 2)	BN (adaptation 3)
C_r	4	3	4
C_1	3	4	3
C_2	3	3	4
Suggested change BN_{C_2}-BN_{C_1}	3-3=0	4-3=+1	3-4=-1
$BN_{C_n} = BN_{C_r} + (BN_{C_2}$-$BN_{C_1})$	$4 + 0 = 4$	$3 + 1 = 4$	4 - 1 = 3

TABLE 6.1: Three examples of the retrieved beam number adaptation

beam numbers between the retrieved pair of cases (C_1,C_2) is equal to 0, the retrieved beam number for the new case is applied without change ($4 + 0 = 4$). The remaining adaptations are to be interpreted similarly.

6.2.2 Adaptation-Guided Retrieval of Beam Angles

Experiments (to be reported in Section 6.3) have demonstrated that the adaptation of beam number performed by the adaptation-guided retrieval method was more successful than the adaptation by NN. Therefore, an adaptation-guided retrieval approach to beam angle adaptation was developed and investigated.

A beam angle consists of two angles that determine the position of the patient during radiation: patient couch and gantry. However, in the majority of cases, the patient couch angle is set to $0°$ or $90°$, which would make the calculation of gradient difficult. Therefore, we use only the gantry angle as beam angle. In contrast to the beam number retrieval, which has to suggest a single value for the beam number, the beam angle retrieval deals with multiple values (i.e., an angle value for each beam). Furthermore, the order in which the beams are applied is not fixed; therefore, in the comparison of two cases, the best correspondence between beam angles must be established. The matching of the beams of two cases with the minimum difference between the corresponding angles determines the correspondence.

The pseudo-code of the retrieval of the triplet of cases for beam angle adaptation is the same as for the beam number adaptation with a modification of the calculation of gradient given by Expression (6.4):

$$\nabla C = [\frac{\partial BA}{\partial V_1}, ..., \frac{\partial BA}{\partial V_6}] = [\frac{BA_C - BA_{C_{sim}}}{V_{C_1} - V_{C_{sim1}}}, ..., \frac{BA_C - BA_{C_{sim}}}{V_{C_6} - V_{C_{sim6}}}] \tag{6.4'}$$

where $BA_C = \{angle_{C_1},...,angle_{C_{BN}}\}$, and $BA_{C sim} = \{angle_{C_{sim1}},...,angle_{C_{simBN}}\}$.

The adaptation-guided retrieval of beam angles is illustrated using an example. Note that in our description, {} denotes a set, while [ ] denotes a vector.

Let us suppose that case C_n is described by the following attributes:

$C_n = [Vol = 0.0840, R = 0.4104, A = 0.4655, Dt = 0.5475, E = 0.1568, P = 0.0823]$.

The CBR system with adaptation-guided retrieval of beam number suggests four beams to be used in the treatment. Therefore, in the second phase of the retrieval, only cases with four beams are considered. We will illustrate one iteration in the adaptation-guided retrieval of beam angles. Let us suppose that in the current iteration, the retrieval considers case C_r with attributes $[0.0737, 0.1869, 0.2314, 0.1494, 0.1428, 0.0423]$ and beam angles [300 °, 306 °, 235 °, 90 °].

The difference between C_n and C_r, $d(C_n, C_r)$, is calculated to be 0.0654 using Expression 3.1. The case differences vector $\triangle(C_n, C_r) = [0.0103, 0.2235, 0.3161, 0.0465, 0.0140, 0.0400]$ is calculated using Expression (6.2).

To calculate the gradient of case C_r, the most similar case to C_r (C_{sim}, with the same OAR and four beams) is retrieved with beam angles [355 °, 315 ° , 270 °, 51 °],

The gradient of case C_r is calculated using Expression (6.4'):

$$\nabla C_r = \begin{bmatrix} 0.0178 & 0.0460 & 1.7730 & 1.4467 & 0.2582 & 0.0540 \\ -0.0773 & -0.1994 & -7.6832 & -6.2691 & -1.1188 & -0.2340 \\ 0.0694 & 0.1789 & 6.8952 & 5.6261 & 1.0041 & 0.2100 \\ 0.0109 & 0.0281 & 1.0835 & 0.8841 & 0.1578 & 0.0330 \end{bmatrix}$$

Using Expression (6.6), R_{C_n,C_r} is calculated as follows:

$$R_{C_n,C_r} = \begin{bmatrix} 0.1842 & 10.2812 & 560.5443 & 67.2692 & 3.6232 & 2.1600 \\ -0.0008 & -0.0446 & -2.4290 & -0.2915 & -0.0157 & -0.0094 \\ 0.0007 & 0.0400 & 2.1799 & 0.2616 & 0.0141 & 0.0084 \\ 0.0011 & 0.0628 & 3.4255 & 0.4111 & 0.0221 & 0.0132 \end{bmatrix}$$

All combinations of (C_1,C_2) from the filtered case base are considered for adaptation of the beam angles suggested by C_r.

Let us consider a pair C_1 $[0.2832, 0.4042, 0.5408, 0.5743, 0.0134, 0.3145]$ with beam angles [325 °, 270 °, 136 °, 59 °] and C_2 $[0.1000, 0.1800, 0.2700, 0.0743, 0.0140, 0.2311]$ with beam angles [270 °, 348 °, 312 °, 52 °].

The case differences vector $\triangle(\mathrm{C}_1,C_2) = [0.1832, 0.2242, 0.2708, 0.0653, 0.0040, 0.1111]$.

The most similar case to C_2 is case C_{sim2} with beam angles corresponding to C_2 [0 °, 88 °, 310 °, 330 °]. It is used to calculate the gradient of C_2 using Expression (6.4'):

$$\nabla C_2 = \begin{bmatrix} -0.1984 & -0.0888 & -1.2488 & -1.5966 & -0.2448 & -0.1566 \\ 0.2052 & 0.0918 & 1.2919 & 1.6517 & 0.2533 & 0.1620 \\ 0.2281 & 0.1020 & 1.4354 & 1.8352 & 0.2814 & 0.1800 \\ 102.6249 & 45.9147 & 645.9253 & 825.8396 & 126.6419 & 81.0001 \end{bmatrix}$$

Subsequently, R_{C_1,C_2} is calculated as follows:

$$R_{C_1,C_2} = \begin{bmatrix} -0.3636 & -0.1990 & -3.3818 & -1.0423 & -0.0098 & -0.1740 \\ 37.6116 & 20.5839 & 349.8448 & 107.8233 & 1.0139 & 18.0000 \\ 41.7907 & 22.8710 & 388.7164 & 119.8036 & 1.1266 & 20.0000 \\ 18.8058 & 10.2920 & 174.9224 & 53.9116 & 0.5069 & 9.0000 \end{bmatrix}$$

In order to calculate the difference d(R_{C_n,C_r},R_{C_1,C_2}), the correspondence between the beam angles of C_r [300 °, 306 °, 235 °, 90 °] and C_2 [270 °, 348 °, 312 °, 52 °] has to be established. The angles of C_r and C_2 are matched in pairs, and the sum of angle differences is calculated. The matching, which results in the minimum sum, establishes the following correspondence between the beam angles of C_r and C_2:

C_r [306 °, 90 °, 235 °, 300 °] and C_2 [348 °, 52 °, 270 °, 312 °]

Using Expression (6.5), we obtain:

d(R_{C_n,C_r},R_{C_1,C_2}) = 1.0935x10^{+03}

Let us suppose that this triplet of cases (C_r,C_1,C_2) has the minimum value of R_{C_r,C_1,C_2} among all other combinations of pairs (C_1, C_2) from the case base. Then, the angles in C_r will be adapted as follows:

C_n = C_r + C_1 - C_2

This will lead to the following beam angles suggested for the case C_n:

C_n = [306°, 90°, 235°, 300°] + [325°, 270°, 136°, 59°] - [348°, 52°, 270°, 312°] = [283°, 308°, 101°, 47°].

6.3 Experiments and Results

The results of experiments using the adaptation-guided retrieval method for beam number and beam angle adaptation are reported in this section.

The case base used for the evaluation of the proposed adaptation methods consists of 80 cases collected from the Nottingham City Hospital. Out of 80 cases, 56 were treated with 4 beams, 23 cases with 3 beams and 1 case with 2 beams.

6.3.1 Beam Number Adaptation

Comprehensive experiments were carried out to evaluate the performance of the adaptation-guided retrieval of beam number. The comparison of the results of the CBR with no adaptation with those of CBR with AGRBN are presented in Table 6.2. The leave-one-out cross validation method explained previously is used to measure the success rate. First, using AGRBN, in 80% of retrievals, the cases retrieved to suggest the beam number, C_r, are different from those retrieved in the original CBR system with no adaptation. The retrieved case in AGRBN is often not the most similar case to the new case. This suggests that the case most similar to the new case might not be the best to adapt. The experiments demonstrate that the performance of the CBR system with AGRBN is improved compared to the CBR with no adaptation. The success rate of CBR with AGRBN is 90%, which represents an improvement compared to the success rate of CBR with no adaptation (61%).

Methodology	CBR with no adaptation	CBR with AGRBN
Correct *BN*	49	72
Incorrect *BN*	31	8
Success rate	61%	90%

TABLE 6.2: Comparison of CBR retrieval with and without the proposed AGRBN adaptation

The detailed results of the AGRBN adaptation are presented in Table 6.3. AGRBN suggested the correct adaptation (i.e., the beam number is adapted correctly) in 72 out of 80 test cases; that is, the beam number was increased by one, decreased by one or not changed. In 13 cases in which the retrieved beam number had three beams, AGRBN correctly suggested to add one beam to the retrieved case. In eight cases, AGRBN correctly adapted the retrieved beam number (i.e., decreased it by one). In 51 cases, no change was performed, which was correct because the beam number in the retrieved case was equal to the actual beam number of the new case.

		CBR with AGRBN (80 cases)
Correct BN	No change to *BN*	51
	Add beam to *BN*	13
	Subtract beam to *BN*	8
	Number of cases	**72**
Incorrect *BN*	Add beam to *BN*	5
	Subtract beam to *BN*	3
	No change to *BN*	0
	Number of cases	**8**
	Success rate	90%

Table 6.3: Results obtained using the proposed AGRBN adaptation

The experiments have shown that the CBR system with the AGRBN adaptation performed better than the original CBR system and the CBR system with machine learning-based adaptation.

6.3.2 Beam Angle Adaptation

Initially, there were 80 cases in the case base. However, one case that had two beams had to be discarded because at least three cases with the same beam number have to be present in the case base to form a triplet of cases (C_r,C_1,C_2) required for AGRBA. Hence, 79 cases were used. AGRBA adaptation did not produce any results for five cases because the system could not find similar cases with the same OARs or cases that could be used to calculate the gradient. AGRBA produced results for 74 cases, among which 20 cases were treated by medical physicists with three beams, and 54 cases were treated with four beams. The success rate of the proposed adaptation method was measured by calculating the beam angle error, E_{BA}, as explained in Chapter 3.

The success rate of AGRBA was compared with the success rate obtained with the CBR system with no adaptation. The comparison of the performances of the two methods is presented in Table 7. The cases retrieved in the two systems may have different numbers of beams; namely, CBR with no adaptation retrieved 21 cases with three beams and 53 cases with four beams, while CBR with AGRBA retrieved 18 and 56 cases with three and four beams, respectively. The success rate of CBR without adaptation for beam angles is 39% (i.e., the suggested beam angles had beam angle error, E_{BA}, less than 30° in 29 out of 74 cases). However, the success rate of AGRBA is lower than that of CBR with no adaptation (i.e., the beam angle error, E_{BA}, was less than 30° in only 21 cases). AGRBA correctly performed the adaptation in only three retrievals of cases that suggested three beams. The success rate was higher (18 cases) when the retrieved case had four beams. This higher success rate for retrievals of cases with four beams is likely due to the larger number of cases with four beams in the case base. When a treatment requires three beams, the case base does not contain enough cases for the beam angle adaptation, which requires a triplet of cases as opposed to the retrieval of a single most similar case.

Case base	CBR with no adaptation	CBR with AGRBA
Retrieved cases with 3 beams	7/21	3/18
Retrieved cases with 4 beams	22/53	18/56
Total	29	21
Success rate	39%	29%

TABLE 6.4: Success rates of CBR with no adaptation and CBR with AGRBA

The next step in the analysis of the obtained results is to take a closer look at the beam angles suggested by CBR with no adaptation and CBR with AGRBA (i.e., examine the beam angle errors, E_{BA}, produced by the two methods). The following question arises: for how many new cases did CBR with AGRBA produce a smaller E_{BA} compared to CBR with no adaptation, even if the retrieval was not successful (i.e., E_{BA} was greater than 30°). The results are presented in Table 6.5.

In 6 cases with three beams and 19 cases with four beams, CBR with AGRBA resulted in a smaller E_{BA} than CBR with no adaptation. However, in the remaining majority of cases, CBR with AGRBA produced worse results.

Case base	Number of cases in which CBR with AGRBA performed better than CBR with no adaptation
20 cases with 3 beams	6
54 cases with 4 beams	19
Total	25

TABLE 6.5: Improvement in E_{BA} using CBR with AGRBA compared to using CBR with no adaptation

Later, in each retrieval, we measure the difference in E_{BA} produced by the two CBR systems and count the number of cases in which one CBR system outperformed the other. Figure 3 presents the distribution of beam angle error improvement by CBR with AGRBA over CBR with no adaptation and vice versa. The difference in E_{BA} between the two systems is shown in intervals of $10\,^{\circ}$. In higher ranges of E_{BA} difference, where the E_{BA} difference is greater than $60\,^{\circ}$, both CBR with AGRBA and CBR with no adaptation have similar performances. That is, in two cases, AGRBA performs better by improving the E_{BA} compared to that of CBR with no adaptation, while the opposite occurred in two other cases. In a lower range of E_{BA} difference, where the difference in E_{BA} is smaller than $10\,^{\circ}$, both systems perform similarly, with AGRBA exhibiting a better performance over CBR with no adaptation in seven cases and CBR with no adaptation having the better performance in eight cases. However, in the majority of the remaining cases, CBR with no adaptation had a smaller E_{BA} than CBR with AGRBA.

6.4 Conclusions

In this chapter we have described an adaptation-guided retrieval adaptation method to improve the results obtained by the retrieval phase of the previously developed CBR system.

The adaptation-guided retrieval approach for beam number improved the success rate of the CBR system by 29%. In most retrievals, the cases retrieved after implementing the adaptation-guided retrieval were different from the cases obtained by the CBR system with no adaptation because not only was the similarity of the retrieved case considered, but also how it can be adapted for the new case.

The computation time of the developed adaptation-guided retrieval approach de-

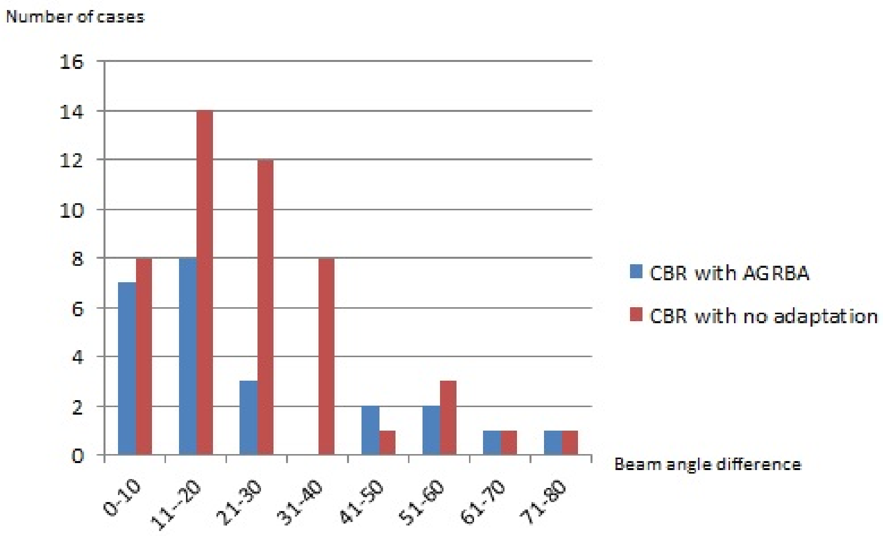

FIGURE 6.1: Difference in E_{BA} between CBR with AGRBA and CBR with no adaptation

.

pends on the size of the case base. The larger the case base the longer the system will take as it will consider all combinations of the cases when looking for a triplet to learn adaptation from. The adaptation-guided retrieval approach took 2 hours (using the leave-one-out approach) to adapt the beam number for the 80 new cases out of the case base of 80 cases. This results in the processing time of 1.5 minutes on average per case to adapt.

The beam angle adaptation is formulated as a more complex optimisation problem in terms of decision variables to consider. Its performance did not improve the CBR with no adaptation. In particular, this approach performed the worst for the cases with three beams as the relevant case base was too small.

Involvement of medical physicists in the evaluation of the developed CBR system is a possible future research direction. Medical physicists were not actively included in adaptation phase development in this research because the aim was to investigate a knowledge-light adaptation. Implementation of the knowledge-intensive adaptation with the help of medical physicists could be a part of a new project in future.

Chapter 7

Conclusions

7.1 General Summary

The main topic of this thesis is the adaptation phase in the CBR system. In this research knowledge acquisition process for implementing the adaptation phase and knowledge-light adaptation methods that make use of the domain knowledge available in the case base are investigated.

The retrieval mechanism of the previously developed CBR system was revised by incorporating a clustering of cases based on their tumour position. The revised CBR retrieval system is described in Chapter 4.

Two knowledge-light adaptation approaches that are developed are discussed further. Machine learning-based adaptation that was implemented for beam number adaptation is described in Chapter 5. Machine learning tools are used for pattern recognition, therefore they are combined here with CBR for adaptation purposes in the hope of enhancing and complementing each other. The idea of combining machine learning algorithms and CBR in designing a decision support system is already present in the literature but mostly for the CBR retrieval phase development. Three machine learning algorithms were used including neural networks, the naive Bayes classifier and support vector machines out of which only neural networks were reported previously to be used in the adaptation phase but in a different manner. To our knowledge this is the first attempt of applying the naive Bayes classifier and support vector machines in the CBR adaptation phase.

An adaptation-guided retrieval approach for adaptation of beam number and beam angles was described in Chapter 6. Traditionally, in the retrieval, the case most similar to the new case is retrieved. Adaptation is typically performed on the retrieved most similar case. However, it may be possible that the retrieved case is not a good starting point for adaptation. In the adaptation-guided retrieval approach, the adaptation phase is intertwined with the retrieval phase and the case is retrieved only if it can be adapted. In most retrievals, the cases retrieved by the adaptation-guided retrieval were different from the cases obtained by the original CBR system, which suggests that the most similar case is not necessarily the most adaptable.

The success rate of the developed methods was evaluated based on the beam number and beam angle error threshold defined by Jagannathan et al. (2012) that was set up based on consultation with medical physicists. From the obtained results we get insights into the complexity of the adaptation problem and the amount of knowledge required for adaptation. In particular, both developed knowledge-light adaptation approaches improved the success rate of the CBR system with no adaptation with respect to beam number. The beam number is represented as a discrete integer number and for its successful adaptation the knowledge acquired from the case base was sufficient. The beam angle adaptation is formulated as a more complex optimisation problem in terms of decision variables to consider: they are in the range of 0° - 360° and can take any real value. The proposed knowledge-light adaptation methods did not improve the beam angle performance of the original CBR system. It can be concluded from the experiments that beam angle adaptation requires more cases in the case base as the adaptation-guided retrieval approach requires a triplet of cases not just a single case, or it might be that additional knowledge that is not present in the case base is needed. Thus, alternative methods for beam angle adaptation are discussed.

7.2 Contributions of the Thesis

The main points of contribution that are the outcome of the research questions addressed in this thesis are as follows:

- **Clustering of cases:** When the medical physicists tested the previously developed CBR system they noted that some of the retrieved cases had beam angles that could not be adapted for the new cases. This discovery showed the requirement to include additional attributes to the CBR retrieval. Investigation of how and at which stage in the CBR retrieval this attribute could be added was conducted. A suitable clustering methodology was determined. Additional clustering of cases by the attribute improved the performance of the CBR retrieval.

In radiation therapy the beams are placed such that the radiation is focused at the point, known as the isocentre, where all beams meet and ideally this point should be placed at the centre of the tumour. It was found that in the cases which could not be adapted the CBR system retrieved cases with an isocentre that did not match the tumour position of the new case. In particular, for posterior tumours, located at the back of the brain, anterior tumour cases, located in the front part of the brain were retrieved. Therefore, the retrieved beam angles were located in the front part of the brain and were not aimed at the centre of the tumour and thus, could not be adapted. For this reason, cases in the case base are clustered based on their tumour centre position information. Filtering of the case base is performed prior to the retrieval phase so that cases to be retrieved are within the same cluster as the new case. It was shown that the revised CBR retrieval system retrieved cases with improved isocentre position, i.e. the isocentre of the retrieved treatment plan was closer to the tumour centre of the new case. The appropriate isocentre position of the retrieved case means that its beam angles are suitable for adaptation.

- **Machine learning-based adaptation:** Application of machine learning methods for adaptation showed promising results. The most successful machine learning-based adaptation was performed using neural networks. It was found that the implemented support vector machine and naive Bayes classifier-based adaptation did not improve the results of the CBR retrieval. The possible reason for naive Bayes classifier not improving the current retrieval results is that it did not consider the interplay between the attributes. Support vector machines-based adaptation did not improve results, possibly because our dataset consists of attribute values that are highly varied and therefore difficult to learn from. In our domain neural networks-based adaptation for beam number improved the CBR retrieval results by 12%.

- **Adaptation-guided retrieval approach:** This adaptation approach was the most promising. This approach considered only cases for retrieval that could be adapted, in particular if for the retrieved cases there existed a pair of cases in the case base similar to the target case - retrieved case pair that could be used to guide its adaptation. The implemented approach considerably improved on the CBR system with no adaptation. In particular, it improved the beam number performance by 29% compared to the original CBR system.

• **Adaptation approach applicable to classification problem:** The beam number adaptation approaches described in Chapters 5 and 6 are general in nature, i.e. they can handle any number of beams in the cases on the condition that there are a reasonable number of cases with that number of beams in the case base, so that enough cases can be used in the training phase. In addition, the developed methods can be used for any classification problems in other domains.

All developed CBR adaptation methods can be used in any other domains as well as radiotherapy treatment planning, both linear and non-linear. The developed knowledge-light adaptation methods are especially useful in the domains where additional knowledge for adaptation is hard to obtain.

7.3 Other Key Results

Other key points and insights derived from the results noted during the experiments are presented below:

• **Relationship between inputs and outputs:** The naive Bayes classifier-based adaptation method did not perform well. In this approach attributes are considered independently of each other. The neural network-based adaptation in which all input attributes versus beam number output are considered has improved the success rate of the CBR retrieval system (Chapter 5). It can be concluded from these results that none of the individual attributes considerably affect the output on their own. Attributes interplay between each other in deciding an output beam number or beam angle.

• **Computation time:** The computation time of the presented machine learning-based adaptation methods is fast, up to 4 minutes to process the whole case base. The adaptation-guided retrieval approach took 2 hours to get results for the 80 new cases in the leave-one-out fashion, based on the case base of 80 cases. However, the computation time of the developed adaptation-guided retrieval approach depends on the case base size. The larger the case base the longer it takes as the system will consider all possible combinations of a similar case and a similar cases pair from which to learn adaptation. The computation time could be important to consider if the case base is large and the domain requires short computation time.

• **The importance of knowledge acquisition:** It was learnt that the amount of knowledge available plays an important role in the CBR development process, particularly for adaptation. Chapter 4 could not be implemented unless additional information, on the importance of the isocentre and the tumour position match, was obtained from the medical physicists during the system testing. Adaptation is the knowledge-intensive phase of CBR, i.e. it requires a lot of knowledge, which is often obtained from the domain users and/or experts. However, they are not always available when needed especially in the medical field. Thus, knowledge-light adaptation methods that make use of the domain knowledge available in the case base are considered in this research. The domain knowledge present in the case base was sufficient for beam number adaptation; the knowledge-light adaptation methods successfully improved its performance compared to the CBR system with no adaptation. The beam angle adaptation is a more complex problem. Contrary to the beam number which is represented as a discrete integer number, beam angles are in the range of 0° - 360° and can take any real value. Each of the retrieved angles in the range of 0° - 360° need to be considered separately for adaptation, thus the problem is multiplied by the number of beams.

7.4 Extensions and Future Work

Potential directions for future research are listed below:

• **Involving medical physicists into the CBR system evaluation process:** In the current evaluation of the CBR system, we compare the beam number and beam angles suggested by the retrieval/adaptation process with the beam number and beam angles initially suggested by the medical physicists and stored in the cases in the case base. The disadvantage of this approach is that it only considers the retrieval to be successful when the retrieved cases have the same beam number or differ within a pre-defined threshold for beam angles as suggested by the medical physicists. However, it may be that the CBR system suggests a better treatment plan in terms of the radiation dose distribution with a different beam number and/or beam angles. In reality, medical physicists evaluate the treatment plan based on the homogeneity of dose administered to cancer cells and the dose administered to organs at risk and non-tumour tissue (i.e., healthy tissue) by looking at the dose distribution lines (isodose lines) visualised in the treatment planning system. The busy schedule of the medical physicists prevented them from becoming more involved in the evaluation of the system at this moment. The involvement of the medical physicists in the evaluation of the developed CBR system is a possible future research step.

- **Knowledge-intensive adaptation:** Medical physicists were not involved in the adaptation phase development in this research and adaptation approaches presented in this thesis are knowledge-light. The implementation of knowledge-intensive adaptation with the help of the medical physicists will be part of a new project in the future. Such knowledge-intensive adaptation would simulate the reasoning process used by medical physicists when adapting a retrieved treatment plan. The adaptation procedure will be documented, and a case base of adaptation cases will be built. This type of adaptation is known as case-based adaptation. Acquiring domain knowledge from the medical physicists could lead to a better learning process of the CBR adaptation and hence may improve the success rate of the CBR system, particularly of the beam angle adaptation.

- **Extension of the case base:** The number of cases we currently have in the case base is satisfactory, considering the difficulties in collecting and processing cases for the case base. Nottingham City Hospital gets approximately 100 patients a year with the particular type of brain tumour presented in this thesis. Therefore, our case base contains a reasonable number of patient cases treated in a year. On the other hand, this number of cases is insufficient for any profound statistical analyses. Therefore, more cases will be added to the case base as they become available, enabling us to evaluate the performance of the developed adaptation approaches using a larger case base. In addition, a larger case base could lead to better results in the current adaptation methods and further development of additional knowledge-light adaptation methods.

- **Application of the CBR system and the proposed adaptation methods to other radiotherapy treatment planning modalities:** 3D conformal radiotherapy treatment planning is a popular choice for treatment of the head and neck and brain cancers and is commonly used in many hospitals including Nottingham City Hospital. However, in the last meeting with the medical physicists it was mentioned that the replacement of 3D-conformal radiotherapy planning with IMRT is a future prospect at Nottingham City Hospital. Therefore, the application of the developed CBR system and adaptation methods to other radiotherapy treatment planning modalities, for example IMRT, will be investigated.

- **Additional information on attributes:** For brain cancer radiotherapy treatment no physical attributes such as e.g. female, male, physical fitness of the patient, etc. are considered currently at the Nottingham City Hospital. To our knowledge only the age of the patient is considered during the treatment planning. It would be interesting to see how the addition of the extra attributes would affect the CBR results.

References

A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI communications*, 7(1):39–59, 1994.

M. Abdul, A. M. Muhammad, N. Mustapha, S. Muhammad, and N. Ahmad. Database workload management through cbr and fuzzy based characterization. *Applied Soft Computing*, 22:605–621, 2014.

D. W. Aha. The omnipresence of case-based reasoning in science and application. *Knowledge-Based Systems*, 11(5):261–273, 1998.

A. Almonayyes. Multiple explanations driven naive bayes classifier. *Journal of Universal Computer Science*, 12(2):127–139, 2006.

MNE Athinoula A. Martinos Centre for Biomedical Imaging. The head and device coordinate systems.mne software. http://martinos.org/mne/dev/manual/source_localization/forward.html, 2014. Accessed online 02.12.2014.

J. Berger. Roentgen: Radiation therapy and case-based reasoning. In *Artificial Intelligence for Applications, 1994., Proceedings of the Tenth Conference on*, pages 171–177. IEEE, 1994.

I. Bichindaritz. Mnaomia: Improving case-based reasoning for an application in psychiatry. *Artificial intelligence in medicine: Applications of current technologies*, pages 14–20, 1996.

I. Bichindaritz and C. Marling. Case-based reasoning in the health sciences: Foundations and research directions. In *Computational Intelligence in Healthcare 4*, pages 127–157. Springer, 2010.

I. Bichindaritz and S. Montani. Advances in case-based reasoning in the health sciences. *Artificial Intelligence in Medicine*, 51(2):75 – 79, 2011.

C. M. Bishop. *Neural networks for pattern recognition.* Oxford university press, 1995.

S. Biswas, N. Sinha, B. Purakayastha, and L. Marbaniang. Hybrid expert system using case based reasoning and neural network for classification. *Biologically Inspired Cognitive Architectures*, 9:57–70, 2014.

A. Blansche, J. Cojan, V. Dufour-Lussier, J. Lieber, P. Molli, E. Nauer, H. Skaf-Molli, and Y. Toussaint. Taaable 3: Adaptation of ingredient quantities and of textual preparations. In *18th International Conference on Case-Based Reasoning Workshop Procs*, pages 189–198. Citeseer, 2010.

I. Bose and R. K. Mahapatra. Business data mining-a machine learning perspective. *Information and Management*, 39(3):211 – 225, 2001.

J. Brownlee. How to choose the right test options when evaluating machine learning algorithms. Online blog., 2014.

H. Bulu, A. Alpkocak, and P. Balci. Ontology-based mammography annotation and case-based retrieval of breast masses. *Expert Systems with Applications*, 39(12):11194–11202, 2012.

H. Byun and S.H. Lee. Applications of support vector machines for pattern recognition: A survey. In *Pattern recognition with support vector machines*, pages 213–236. Springer, 2002.

T. Calinski and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3:1–27, 1974.

CBR Cbrwiki. Case-based reasoning wiki. http://cbrwiki.fdi.ucm.es/mediawiki/index.php/, 2012. Accessed 15.11.2015.

B. Chebel-Morello, M. K. Haouchine, and N. Zerhouni. Reutilization of diagnostic cases by adaptation of knowledge models. *Engineering Applications of Artificial Intelligence*, 26 (10):2559–2573, 2013.

J. Cojan, V. Dufour-Lussier, E. Gaillard, J. Lieber, E. Nauer, and Y. Toussaint. Taaable 4: knowledge extraction for improving case retrieval and recipe adaptation. *Proceedings of the Computer Cooking Contest*, pages 197–206, 2011.

J.M. Corchardo, B. Lees, C. Fyfe, N. Rees, and J. Aiken. Neuro-adaptation method for a case-based reasoning system. In *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*, volume 1, pages 713–718. IEEE, 1998.

T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

S. Craw, N. Wiratunga, and R. C. Rowe. Learning adaptation knowledge to improve case-based reasoning. *Artificial Intelligence*, 170(16):1175 – 1192, 2006.

S. Craw, S. Massie, and N. Wiratunga. Informed case base maintenance: A complexity profiling approach. In *Proceedings of the National Conference on Artificial Intelligence*, volume 22, page 1618. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007.

M. D'Aquin, J. Lieber, and A. Napoli. Adaptation knowledge acquisition: a case study for case-based decision support in oncology. *Computational intelligence*, 22(3-4):161–176, 2006.

L. A. Dawson and M. B. Sharpe. Image-guided radiotherapy: rationale, benefits, and limitations. *The lancet oncology*, 7(10):848–858, 2006.

W. De Neve, C. De Wagter, K. De Jaeger, M. Thienpont, C. Colle, S. Derycke, and J. Schelfhout. Planning and delivering high doses to targets surrounding the spinal cord at the lower neck and upper mediastinal levels: static beam-segmentation technique executed with a multileaf collimator. *Radiotherapy and oncology*, 40(3):271–279, 1996.

J. F. De Paz, S. Rodriguez, J. Bajo, and J. M. Corchado. Cbr system with reinforce in the revision phase for the classification of cll leukemia. In Sigeru Omatu, MiguelP. Rocha, José Bravo, Florentino Fernández, Emilio Corchado, Andrés Bustillo, and JuanM. Corchado, editors, *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing,* *and Ambient Assisted Living*, volume 5518 of *Lecture Notes in Computer Science*, pages 964–971. Springer Berlin Heidelberg, 2009.

F. Diaz, F. Fdez-Riverola, and J. M. Corchado. gene-cbr: A case-based reasonig tool for cancer diagnosis using microarray data sets. *Computational Intelligence*, 22(3-4):254–268, 2006.

T. G. Dietterich. Ensemble methods in machine learning. In *Multiple classifier systems*, pages 1–15. Springer, 2000.

P. Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, October 2012.

M. Ehrgott and R. Johnston. Optimisation of beam directions in intensity modulated radiation therapy planning. *OR Spectrum*, 25(2):251–264, 2003.

M. Ehrgott and I. Winz. Interactive decision support in radiation therapy treatment planning. *OR Spectrum*, 30(2):311–329, 2008.

M. Ehrgott, C. Guler, H. W. Hamacher, and L. Shao. Mathematical optimization in intensity modulated radiation therapy. *4OR*, 6(3):199–262, 2008.

C. Elith, S. E. Dempsey, N. Findlay, and H. M. Warren-Forward. An introduction to the intensity-modulated radiation therapy (imrt) techniques, tomotherapy, and vmat. *Journal of Medical Imaging and Radiation Sciences*, 42(1):37–43, 2011.

A. V. Esch, J. Bohsung, P. Sorvari, M. Tenhunen, M. Paiusco, M. Iori, P. Engstrom, H. Nystrom, and D. Pierre Huyskens. Acceptance tests and quality control (qc) procedures for the clinical implementation of intensity modulated radiotherapy (imrt) using inverse planning and the sliding window technique: experience from five radiotherapy departments. *Radiotherapy and Oncology*, 65(1):53 – 70, 2002.

M. C. Ferris, J. Lim, and D. M. Shepard. Radiosurgery treatment planning via nonlinear programming. *Annals of Operations Research*, 119(1-4):247–260, 2003.

C. Fraley and A. E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631, 2002.

C. Fraley and A. E. Raftery. Model-based methods of classification: Using the mclust software in chemometrics. *Journal of Statistical Software*, 18(6):1–13, 2007.

C. Fraley, A. E. Raftery, T. B. Murphy, and Scrucca L. Mclust version 4 for r: Normal mixture modeling for model-based clustering, classification, and density estimation. Technical report, Department of Statistics, University of Washington, 2012.

N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

B. Fuchs, J. Lieber, A. Mille, and A. Napoli. Differential adaptation: An operational approach to adaptation for solving numerical problems with cbr. *Knowledge-Based Systems*, 68(0):103 – 114, 2014.

A. J. Gonzalez, X. Lingli, and U. M. Gupta. Validation techniques for case-based reasoning systems. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 28(4):465–477, Jul 1998.

P. A. Gonzalez-Calero, M. Gomez-Albarran, and B. Diaz-Agudo. A substitution-based adaptation model. In *ICCBR Workshops*, pages 17–26. Citeseer, 1999.

S. Guessoum, M. T. Laskri, and J. Lieber. Respidiag: A case-based reasoning system for the diagnosis of chronic obstructive pulmonary disease. *Expert Systems with Applications*, 41(2):267 – 273, 2014.

G. Hamerly and C. Elkan. Learning the k in k- means. *Advances in neural information processing systems*, 16:281 –288, 2004.

K. J. Hammond. Chef: A model of case-based planning. In *Proceedings of the fifth National Conference on Artificial Intelligence*, pages 267–271. Philadelphia, PA, Los Altos, CA : Morgan Kaufmann., 1986.

K. Hanney and M. Keane. Learning adaptation rules from a case-base. *Advances in Case-Based Reasoning*, pages 179–192, 1996.

J. A. Hartigan. Printer graphics for clustering. *Journal of Statistical Computation and Simulation*, 4(3):187–213, 1975.

J. Heaton. *Introduction to Neural Networks for Java, 2Nd Edition*. Heaton Research, Inc., 2nd edition, 2008.

D. Hennessy and D. Hinkle. Applying case-based reasoning to autoclave loading. *IEEE Expert: Intelligent Systems and Their Applications*, 7(5):21–26, 1992.

C. Hennig. Package 'fpc'. *Available: http://cran.r-project.org/web/packages/fpc/.*, 2014.

J. Henriet, P.-E. Leni, R. Laurent, and M. Salomon. Case-based reasoning adaptation of numerical representations of human organs by interpolation. *Expert Systems with Applications*, 41(2):260–266, 2014.

W. Hoegele, R. Loeschel, N. Merkle, and P. Zygmanski. An efficient inverse radiotherapy planning method for vmat using quadratic programming optimization. *Medical Physics*, 39(1):444–454, 2012.

A. Holder and B. Salter. A tutorial on radiation oncology and optimization. In *Emerging Methodologies and Applications in Operations Research, (Chapter 4)*, pages 4–1. H. Greenberg (Ed.), Kluwer Academic Press, Boston, MA, 2004.

A. Holt, I. Bichindaritz, R. Schmidt, and P. Perner. Medical applications in case-based reasoning. *The Knowledge Engineering Review*, 20(03):289–292, 2005.

M. L. Huang, Y. H. Hung, W. M. Lee, R. K. Li, and T. H. Wang. Usage of case-based reasoning, neural network and adaptive neuro-fuzzy inference system classification techniques in breast cancer dataset classification diagnosis. *Journal of medical systems*, 36 (2):407–414, 2012.

K. H. Im and S. C. Park. Case-based reasoning and neural network based expert system for personalization. *Expert Systems with Applications*, 32(1):77–85, 2007.

R. Jagannathan and S. Petrovic. *A Local Rule-Based Attribute Weighting Scheme for a Case-Based Reasoning System for Radiotherapy Treatment Planning*, volume 7466 of *Lecture Notes in Computer Science*, pages 167–181. Springer Berlin. Heidelberg, 2012.

R. Jagannathan, S. Petrovic, A. McKenna, and L. Newton. A novel two phase retrieval mechanism for a clinical case based reasoning system for radiotherapy treatment planning. *International Journal on Artificial Intelligence Tools*, 21(04):46–70, 2012.

B. Kang and P. Compton. A maintenance approach to case based reasoning. *Advances in Case-Based Reasoning*, pages 226–239, 1995.

A. Karatzoglou, D. Meyer, and K. Hornik. Support vector machines in r. *Journal of Statistical Software*, 15(i09), 2006.

L. Kaufman and P.J. Rousseeuw. Partitioning around medoids (program pam). *Finding groups in data: an introduction to cluster analysis*, pages 68–125, 1990.

J. Kazmierska and J. Malicki. Application of the naive bayesian classifier to optimize treatment decisions. *Radiotherapy and Oncology*, 86(2):211 – 216, 2008.

A. M. Kibriya, E. Frank, B. Pfahringer, and G. Holmes. Multinomial naive bayes for text categorization revisited. In *AI 2004: Advances in Artificial Intelligence*, pages 488–499. Springer, 2005.

J. L. Kolodner. An introduction to case-based reasoning. *Artificial Intelligence Review*, 6 (1):3–34, 1992.

J. L. Kolodner and R. M. Kolodner. Using experience in clinical problem solving: Introduction and framework. *Systems, Man and Cybernetics, IEEE Transactions on*, 17(3): 420–431, 1987.

P. Koton. Reasoning about evidence in causal explanations. In *In Proceedings AAAI-88*, pages 256–261, 1988.

V. Lavrenko. Clustering 4: K-means algorithm. `https://www.youtube.com/watch?v=_aWzGGNrcic`, 19 January 2014. Accessed online:2016-01-23.

D. B. Leake and D. C. Wilson. Categorizing case-base maintenance: Dimensions and directions. In *Advances in Case-Based Reasoning*, pages 196–207. Springer, 1998.

D. B. Leake, A. Kinley, and D. Wilson. Learning to improve case adaptation by introspective reasoning and cbr. In *In Proceedings of the First International Conference on Case-Based Reasoning*, pages 229–240. Springer-Verlag, 1995.

J. Legras, B. Legras, and J. Lambert. Software for linear and non-linear optimization in external radiotherapy. *Computer Programs in Biomedicine*, 15(3):233–242, 1982.

J. Liber and B. Bresson. Case-based reasoning for breast cancer treatment decision helping. *Lecture Notes in Computer Sceince*, pages 173–185, 1999.

Z. Liu and F. Yan. Fuzzy neural network in case-based diagnostic system. *Fuzzy Systems, IEEE Transactions on*, 5(2):209–222, 1997.

J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proceedings of the 5th Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. University of California Press, Berkeley, California, USA., 1967.

N. McDonnell and P. Cunningham. A knowledge-light approach to regression using case-based reasoning. In T.R. Roth-Berghofer, M.H. Göker, and H.A. Güvenir, editors, *Advances in Case-Based Reasoning*, volume 4106 of *Lecture Notes in Computer Science*, pages 91–105. Springer Berlin Heidelberg, 2006.

D. McSherry and C. Stretch. Learning more from experience in case-based reasoning. In *Case-Based Reasoning Research and Development*, pages 151–165. Springer, 2011.

D. Meyer, F. Leisch, and K. Hornik. The support vector machine under test. *Neurocomputing*, 55(1):169 – 186, 2003. Support Vector Machines.

N. Mishra, S. Petrovic, and S. Sundar. A knowledge-light nonlinear case-based reasoning approach to radiotherapy planning. In *21st International Conference on Tools with Artificial Intelligence, 2009.*, pages 776–783. IEEE, 2009.

N. Mishra, S. Petrovic, and S. Sundar. A self-adaptive case-based reasoning system for dose planning in prostate cancer radiotherapy. *Medical Physics*, 38(12):6528, 2011.

S. M. Morrill, R. G. Lane, G. Jacobson, and I. I. Rosen. Treatment planning optimization using constrained simulated annealing. *Physics in medicine and biology*, 36(10):1341, 1991.

S. G. Mota and B. D. Agudo. Acook: Recipe adaptation using ontologies, case-based reasoning systems and knowledge discovery. In *Proceedings of the Cooking With Computers workshop*, pages 41–45, 2012.

K. P. Murphy. Naive bayes classifiers. *University of British Columbia*, 2006.

NEMA. Dicom - digital imaging and communications in medicine. Technical report, National Electrical Manufacturers Association, 1998.

J. Oksanen, R. Kindt, P. Legendre, B. O'Hara, M. H. Stevens, M. J. Oksanen, and M. Suggests. The vegan package. *Community ecology package*, pages 631–637, 2007.

M. Oldham, V. S. Khoo, C. G. Rowbottom, J. L. Bedford, and S. Webb. 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, 1998.

S. Petrovic, N. Mishra, and S. Sundar. A novel case based reasoning approach to radiotherapy planning. *Expert Systems with Applications*, 38(9):10759–10769, 2011.

D. C. Pham, S. S. Dimov, and C. D. Nguyen. Selection of k in k-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1):103–119, 2005.

T. G. Purdie, R. E. Dinniwell, D. Letourneau, C. Hill, and M. B. Sharpe. Automated planning of tangential breast intensity-modulated radiotherapy using heuristic optimization. *International Journal of Radiation Oncology Biology Physics*, 81(2):575–583, 2011.

RadBlog RAD planning. Linear accelerator (linac) vault upgrades. http://radplanning.com/RADblog/2012/04/20/linear-accelerator-linac-vault-upgrades/, 2015. Accessed online 15.04.2015.

V. Raman, P. Then, and P. Sumari. Mammogram problem solving approach: Building cbr classifier for classification of masses in mammogram. In V. V. Das, J. Stephen, and Y. Chaba, editors, *Computer Networks and Information Technologies*, volume 142 of *Communications in Computer and Information Science*, pages 284–289. Springer Berlin Heidelberg, 2011.

A. T. Redpath, B. L. Vickery, and D. H. Wright. A new technique for radiotherapy planning using quadratic programming. *Physics in Medicine and Biology*, 21(5):781, 1976.

E. Roldan Reyes, S. Negny, G. Cortes Robles, and J. M. Le Lann. Improvement of online adaptation knowledge acquisition and reuse in case-based reasoning: Application to process engineering design. *Engineering Applications of Artificial Intelligence*, 41(0):1 – 16, 2015.

B. D. Ripley. *Pattern recognition and neural networks.* Cambridge university press, 1996.

H. Rocha, J. M. Dias, B. C. Ferreira, and M. C. Lopes. Discretization of optimal beamlet intensities in imrt: A binary integer programming approach. *Mathematical and Computer Modelling*, 55(7):1969 – 1980, 2012a.

H. Rocha, J. M. Dias, B. C. Ferreira, and M. C. Lopes. Combinatorial optimization for an improved transition from fluence optimization to fluence delivery in imrt treatment planning. *Optimization*, 61(8):969–987, 2012b.

C. Rorden. Neuroanatomy. http://www.mccauslandcenter.sc.edu/mricro/anatomy/, 10 2002. Accessed online 10.11.2015.

A. Sanchez-Ruiz, P. P. Gomez-Martin, B. Diaz-Agudo, and P. A. Gonzalez-Calero. Adaptation through planning in knowledge intensive cbr. In *Proceedings of the 9th European Conference on Advances in Case-Based Reasoning*, ECCBR 08, pages 503–517, Berlin, Heidelberg, 2008. Springer-Verlag.

S. Sarwar, Z. Ul-Qayyum, and O. Malik. Cbr and neural networks based technique for predictive prefetching. In G. Sidorov, A. Hernández Aguirre, and C.A. Reyes García, editors, *Advances in Soft Computing*, volume 6438 of *Lecture Notes in Computer Science*, pages 221–232. Springer Berlin Heidelberg, 2010.

R. C. Schank. *Dynamic Memory: A Theory of Reminding and Learning in Coumputers and People.* Cambridge University Press, 1982.

A. Schlaefer and S. Dieterich. Feasibility of case-based beam generation for robotic radiosurgery. *Artificial Intelligence in Medicine*, 2011.

R. Schmidt and L. Gierl. Case-based reasoning for antibiotics therapy advice: an investigation of retrieval algorithms and prototypes. *Artificial intelligence in Medicine*, 23(2): 171–186, 2001.

R. Schmidt, L. Gierl, and B. Pollwein. Case-based reasoning for antibiotics therapy advice. In K. D. Althoff, R. Bergmann, and L. K. Branting, editors, *Case-Based Reasoning*

Research and Development, volume 1650 of *Lecture Notes in Computer Science*, pages 550–560. Springer Berlin Heidelberg, 1999.

R. Schmidt, O. Vorobieva, and L. Gierl. Case-based adaptation problems in medicine. *Proceedings of WM2003: Professionelles Wissensmanagement, Kollen-Verlag Bonn*, 2003.

E. Schreibmann, M. Lahanas, L. Xing, and D. Baltas. Multiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy. *Physics in Medicine and Biology*, 49(5):747, 2004.

M. G. Schultz, E. Eskin, E. Zadok, and S. J. Stolfo. Data mining methods for detection of new malicious executables. In *Proceedings. 2001 IEEE Symposium on Security and Privacy, 2001. S&P 2001.*, pages 38–49. IEEE, 2001.

C. Shin, U. T. Yun, H. Kim, and S. C. Park. A hybrid approach of neural network and memory-based learning to data mining. *Neural Networks, IEEE Transactions on*, 11(3): 637–646, 2000.

B. Smyth and M. T. Keane. Adaptation-guided retrieval: questioning the similarity assumption in reasoning. *Artificial Intelligence*, 102(2):249 – 293, 1998.

B. Smyth and E. McKenna. Competence models and the maintenance problem. *Computational Intelligence*, 17(2):235–249, 2001.

X. Song, S. Petrovic, and S. Sundar. A case-based reasoning approach to dose planning in radiotherapy. In *Workshop Proceedings, The 7th International Conference on Case-Based Reasoning ICCBR*, volume 7, pages 13–16, 2007.

STMC St Rita's Medical Centre. Linear accelerator and a patient couch. http://www.ehealthconnection.com/regions/mercy_st_ritas/linear_accelerators.aspx, 11 2015. Accessed online 15.11.2015.

D. Teodorovic, M. Selmic, and L. Mijatovic-Teodorovic. Combining case-based reasoning with bee colony optimization for dose planning in well differentiated thyroid cancer treatment. *Expert Systems with Applications*, 40(6):2147–2155, 2013.

UPCI University of Pittsburg Cancer Institute. What is 3d conformal radiation therapy? http://www.upmccancercenter.com/radonc/conformal.cfm, 2015. Accessed online 15.11.2015.

VMS Varian Medical Systems. Multileaf collimators. millenium mlc image. http://newsroom.varian.com/imagegallery?view=category&mode=gallery&cat=2473, 2016. Accessed online 14.01.2016.

T. Velmurugan and T. Santhanam. Computational complexity between k-means and k-medoids clustering algorithms for normal and uniform distributions of data points. *Journal of Computer Science*, 6(3):363, 2010.

C. Vong and P. Wong. Case-based adaptation for automotive engine electronic control unit calibration. *Expert Systems with Applications*, 37(4):3184–3194, 2010.

C. Vong, T.P. Leung, and P.K. Wong. Case-based reasoning and adaptation in hydraulic production machine design. *Engineering Applications of Artificial Intelligence*, 15(6): 567–585, 2002.

K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl. Constrained k-means clustering with background knowledge. In *18th International Conference on Machine Learning*, volume 1, pages 577–584, 2001.

W. Wilke and R. Bergmann. Techniques and knowledge used for adaptation during case-based problem solving. *Tasks and Methods in Applied Artificial Intelligence*, pages 497–506, 1998.

W. Wilke, I. Vollrath, K. D. Althoff, and R. Bergmann. A framework for learning adaptation knowledge based on knowledge light approaches. In *Proceedings of the Fifth German Workshop on Case-Based Reasoning*, pages 235–242, 1997.

W. L. Yeow, R. Mahmud, and R. G. Raj. An application of case-based reasoning with machine learning for forensic autopsy. *Expert Systems with Applications*, 41(7):3497–3505, 2013.

H. Zhang. The optimality of naive bayes. In V. Barr and Z. Markov, editors, *Proceedings of the 17th International FLAIRS conference*, AAAI Press, Miami, Florida, 2004.

Y. Zhang and D. Li. Cluster analysis by variance ratio criterion and firefly algorithm. *International Journal of Digital Content Technology and its Applications*, 7(3):689, 2013.