University of Nottingham

Improving end-system recommender systems using cross-platform personal information

by

Sultan Alanazi, PhD

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, January 2017

Abstract

Today, the web is constantly growing, expanding global information space and more and more data is being processed and sourced online. The amount of electronically accessible and available online information is overwhelming.

Increasingly, recommendation systems, which engage in some form of automated personalisation, are hugely prevalent on the web and have been extensively studied in the research literature. Several issues still remain unsolved including high sparsity situation and cold starts (how to recommend content to users who have had little or no prior interaction with the system). Recent work has demonstrated a potential solution in the form of cross-domain user modeling.

This thesis will explore the design, implementation and testing of a cross-domain approach using social media data to model rich and effective user preferences and provide empirical evidence of the effectiveness of the approach based on direct real-world user feedback, deconstructing a cross-system news recommendation service where user models are generated via social media data.

This will be accomplished by identifying the availability of a source domain from which to draw resources for recommendations and the availability of user profiles that capture a wide range of user interests from different domains.

This thesis also demonstrates the viability of generating user models from social media data and evidences that the automated cross-domain approach can be superior to explicit filtering using self-declared preferences and can be further augmented when placing the user with the ability to maintain control over such models.

The reasons for these results are qualitatively examined in order to understand why such effects occur, indicating that different models are capturing widely different areas within a user's preference space.

Acknowledgements

I would like to express appreciation to my supervisor Professor Derek McAuley and my co-supervisor Dr James Goulding, for their guidance and valuable advice during my research. This thesis would not have been completed without their continued supervision and Support.

I am therefore indebted and grateful for the indispensable advice, cooperation and valuable information provided on various aspects of my research.

I would like to acknowledge my gratitude to Dr Richard Mortier from the University of Cambridge for his regular encouragement and supervision during the first year of my research.

I take this opportunity to thank the members of the Networked Systems Group for their help, encouragement. Their assistance, cooperation, and feedback were essential for the completion of the work.

I also place on record, my gratitude to all who, directly or indirectly, have supported me during this work.

Content

List of Figures

List of Tables

Glossary of Abbreviations

CF	Collaborative Filtering
CDF	Cross-Domain Filtering
TF-IDF	Term Frequency-Inverse Document Frequency
CBF	Content-Based Filtering
VSM	Vector Space Model
RSS	Rich Site Summary - Really Simple Syndication
CB	Content-Based
SVD	Singular Value Decomposition
LSI	Latent Semantic Indexing
VRM	Vendor Relationship Management
PDS	Personal Data Storage
UPDS	Sersonal Data Storage System
kNN	k-nearest Neighbours
CPS	Content Pushing Service
XML	eXtensible Mark-up Language
API	Application Programming Interface

1. Introduction

Recommendation Systems are ever-present on the internet, covering fields as diverse as movies, books, music and even academic references [1][2]. Such systems have become so entrenched in large online businesses within the digital economy[1] they may now be reasonably viewed as a barrier to entry for new competitors entering the market. Underpinning these systems, Collaborative Filtering (CF) techniques are now used extensively as a means of identifying and serving relevant content to any given user. While such systems have been widely researched, several issues remain unresolved - issues of handling sparsity [3] and cold starts [4][5] still remain active research topics. Furthermore, personalization of this form can quickly distill into a task of monitoring and tracking across users, raising potential issues of data control and privacy. It is not as yet clear how best to maintain user privacy within this context.

While CF systems have shown empirical effectiveness in practical settings, several research issues remain, not least the problem of "cold-start" [4][6][5] detailed in Section 2.4.2. Cross-System Filtering, and in particular Cross-Domain filtering (CDF) has been proposed as a potential solution. CDF functions by examining a user's transactional history in some external system or domain, determining commonalities between that domain and the new system in question and inferring informational content that can be transferred between the two [7].

Due to the increasing popularity of social networks, a huge amount of data and personal information are being continuously posted and shared via different online providers and contains implicit user knowledge which can be used in improving recommendation accuracy to meet the user's interests and enrich their experience.
Therefore, personal data created by and about people can be used to generate 'a new wave of opportunity for economic and societal value creation', says the WEF [8].

'Personal data is the new oil of the Internet and the new currency of the digital world'. Meglena Kuneva, European Consumer Commissioner, March 2009

[1] Primarily within mass online services (e.g. Amazon, Netflix, YouTube, Spotify, Google, Expedia, eBay, etc.)

Recent research has proposed a potential solution to these issues in the form of cross-system and cross-domain user modeling [9][1] and this area will be the research focus of this thesis. Here features generated within one system or domain can be used to effectively bootstrap recommendations in another, providing a promising line of attack to handle both sparsity and cold-starts. As such cross-system collaborative filtering approaches are receiving increasing interest. While results are highly encouraging, testing of these techniques has, so far, been restricted to internal cross-validation, assessing whether cross-domain recommendations appear in holdout data.

The combination of content-based filtering and the cross-system approach that I will be investigating already has some precedent [3][10][11][12] - but while such cross-system content-based approaches hold much promise little research has been performed on practical user-feedback studies, nor investigated whether such an approach can compete with users explicitly stating their preferences. I augment such research by focusing on analysis of direct, real world user-feedback rather than in-sample analysis.

In this work, I explicly extend this approach (i.e cross-system content-based approach) by 1. using social media data to model user preferences and 2. providing further empirical evidence of the effectiveness of the approach based on direct real-world user feedback.

This is achieved via implementation of a cross-system news recommendation service, with user model's being automatically generated from their social media data. Experiments will demonstrate not only the viability of harnessing linguistic vector-space user models generated from social media data (in this case mined from available Twitter and Facebook streams), but also that automated cross-domain modeling can be superior to explicit filtering using self-declared preferences.

As demonstrated via the statistically significant results of our user experiments, such approaches pave the way for scalable mechanisms that allow users to collate dynamic behavioural models, but which also enable third parties to generate personalised recommendations from them while respecting a user-defined privacy boundary.

Via the development of an experimental cross-system platform, I show that this is indeed possible, even when using a relatively naïve Term Frequency-Inverse Document Frequency (TF-IDF) vector space approach. I specifically investigate content-based recommendations due to practical issues of logistics and user-control that affect real world systems. Most previous work in cross-system recommendation has relied on models generated in one highly siloed domain being projected onto another, which requires the coordination of vast datasets and is dependent on the economies of scale which allow their construction.

I therefore investigate in this work a discussion of user-controlled cross-system recommendation, where models are generated through independent and private processing of a user's own data[2].

In depth, qualitative analysis of why such effects occur will show that the different models are capturing widely different areas within a user's preferences space, and that hybrid models present productive ground for future research.

In the next section, the main statement of this thesis is outlined, followed by the document structure of the main aspects of the different work carried out to develop a framework that enables the provision of cross-system user modeling using social media data.

1.1 Thesis statement and contributions

A framework will be developed to enable Recommendation Systems to benefit from *a cross-system user model* mined from *social media* in building a rich and effective user profile. Moreover, and as a consequence of their cross-domain nature, such systems can potentially *generate* recommendations that are of high quality and geared solely towards meeting user's interests and enriching user experience. This includes understanding the impact of using personal information to personalize delivery of news content and the extent to which personal information can improve recommendation accuracy to satisfying user needs.

The various contributions made to the state–of–the–art in support of this thesis are outlined here.

The result of the process described above would construct a user profile based on the source of a social media stream. Online social media services and social network sites have become a mainstream cultural phenomenon [13] and people are living in an age of sharing [14]. For example, users are sharing different aspects of their lives on Facebook [15], and sharing their thoughts on Twitter

[2] Currently, in most regulatory frameworks the personal data it is generated via our interactions with a company is owned by that company and not necessarily under our control, even when publicly available.

[16]. This represents a user with a profile reflecting their interests which can be applied to generate passive-user models using their Twitter or Facebook data stream.

The system is then required to assess whether using social network data can play an important role in cross-system recommendations and to assess which social media domain is performing better in generating more accurate profile reflecting and satisfying user needs. Also, this would assess the accuracy of reflecting user interests with the ability for a user to maintain control over such a profile. Thus, different user models constructed based on implicitly or explicitly defined user preferences, play a role in cross-domain recommendations using the relatively straightforward VSM technique where each document in the corpus is transformed into a corresponding vector space model which iterates through every n-gram it contains. The BBC news RSS feed (as our output documents) was used to test our modules by assessing recommendation performance of each of our user models.

Exploration of Cross-System User profiles:

Recommendation Systems are typically designed to improve the user experience by helping users to ease the searching of information and resources in rich environments. Thus, user profiles in the system are the only source to contain information that would help in filtering and in personalising the resources. According to this assumption, the opportunities for matching the right information will be challenged, and it limits the suitability of the recommendations that are promoted. Instead, this thesis will show how a wide range of user interests from different domains (i.e. social media data) can be automatically captured and compiled into a user profile to create a sufficient representation through finding accurate information in a satisfactory and efficient way.

In this these, I investigate the effectiveness of constructing and building user profile by mining different social network data streams. Individually, user's Twitter timeline and Facebook posts were used to generate user models. This thesis shows that user-models automatically generate via social media data provide statistically superior performance than models inferred with explicit user input. Moreover, this work shows statically that giving users an ability to maintain control over such models can be further augmented.

Interleaving of both Qualitative and Quantitative valuations:

The framework presented above was evaluated via different experiments that have been designed and executed. In order to support comparisons, each one of these experiments constructed different models to assess the feasibility of such a task. Once these models are constructed for each user, a comparative study is carried out to assess changes in performance of recommended items to appear in different models. So, the assessment of the model's performance in producing recommendations from the same source domain for each individual is evaluated. The value of using Twitter or Facebook data stream was examined in producing cross-system recommendations with a high quality of matching user preferences. Moreover, this thesis examined that whether given the ability for a user to maintain control over such a model to sculpt, edit and reflect upon that profile is sufficient to capture more accurate user interests. This is carried out by assessing differences in the relevance of recommendations occurring when the user manually updates his or her profile.

1.2 Document structure

The next chapter gives an overview of background literature. It first provides a review of the different recommendation system approaches, identifying the prominent characteristics of each approach. Also, the grouped recommendation techniques along with metrics and methods used for recommendation purposes are considered. Following, is a discussion on a selection of research challenges for recommendation systems.

Chapter 3 then provides a brief introduction to the definitions of *personal information* and *user expectations*. The current studies are then used to give a description of the studies conducted on personal data.

Chapter 4 goes onto provide an overview of background cross-domain literature for the concepts and functions that are used throughout the recommendation process. Then, levels of domains along with their distributions in the literature and goals of cross-domain recommendation were revised. Finally, a classification of existing cross-domain recommendation approaches is provided in accordance with the identified techniques and studies in each technique.

Chapter 5 develops the methodology of our studies for the cross-domain recommendation task, starting from a background and motivation grounded in the notion of a 'domain' and examining how these become potential sources of modeling user-preferences. The system design and system architecture is presented in this chapter along with how the documents would be modeled.

The use of Twitter as a cross-domain data source, generated by the user and which are subsequently mined passively, is first provided in Chapter 6. In this chapter, a brief introduction to our experimental platform is provided before detailing how its models can be used to test the recommendation of BBC news items. Followed by a statistical test to evaluate the performance of our approach before exploring a more in-depth analysis of how the system is functioning.

As such, Chapter 7 presents the model developed to utilize Facebook data for recommendations. In this section four different models for obtaining BBC news articles representative of domain resources are presented and evaluated, with one of these models is based on the use of Facebook posts. Each model applies to a different set of user-preference modeling characteristics, providing comprehensive data points that are sufficient to be applicable to the statistical tests. In this chapter, different statistically tests are performed with more in-depth discussion and post-analysis.

Chapter 8 proposes a combination between models in chapter 6 and 7 in a single study based on more rich data sources. This is the case where user-based models must first be established via cross-domain data mined passively from data stream sources used in chapter 6 and 7. The experimental platform developed to implement this task is also detailed in this chapter, followed by discussion of the different content sources that were used to accomplish this study. This chapter also details experiments and their results before discussing their significance via a quantitative analysis based on real user study and a qualitative analysis based on a post-study questionnaire.

Finally, Chapter 9 provides a summary and analysis of all experimental results with respect to the central thesis of the dissertation. Limitations and other directions for future work from the various aspects of this study are highlighted and outlined.

1.3 Publications resulting from the Thesis

During the development of this thesis the following work has been executed and published:

- **Published and presented:**

The first paper discusses and investigates the spread and the importance of social networks used in this thesis.

- Alqahtani, S. M., Alanazi, S., & McAuley, D. (2014). The role of Enterprise Social Networking (ESN) on business: Five effective recommendations for ESN. In *Proceedings of the Ninth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX. June 30–July 4, 2014, Brunów, Poland* (pp. 23-36). Springer International Publishing.

This paper published the work in Chapter 6 which discusses the use of Twitter as a cross-domain data source.

- Sultan Alanazi, James Goulding, and Derek McAuley. 2016. Cross-system Recommendation: User-modelling via Social Media versus Self-Declared Preferences. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media* (HT '16). ACM, New York, NY, USA, 183-188.

2. Literature Review: Recommendation Systems

2.1 Background

Recommendation systems have become a very common technology since first appearing in the mid-1990s when the first papers were published in this area [17][18][19]. The field has since continually expanded as a research area, with many studies focusing on this type of technology.

Clearly, recommendation systems have played a vital role in a range of areas from Netflix to social media. In many ways, the experience of interacting with digital services has been completely changed as these systems have provided new strategies to help users search or make decisions within large amounts of information [20]. Moreover, recommendation systems are recognised as one of the best available technologies used to support individuals finding relevant and suitable information, which meet their requirements. Recommendation systems aim to give the user satisfaction by suggesting items to the user [21].

Lampropoulou et al. [22] point out that the social process of recommendation is represented similarly in the recommendation systems by helping users (in the pressure of information overload) who feel overwhelmed with information. Also, they added that recommendation systems provide information in the most appropriate and valuable way to users by taking into account their personal needs and interests. Perugini et al. [23] states that, based on users' preferences, recommendation systems try to not only reduce information overload but also retain users by choosing a subset of items or products from a universal set.

It is clear that the benefits of recommendation systems are most obvious in numerous domains such as news due to a 'predictive utility', "which refers to the value of having predictions for an item before deciding whether to invest time or money in consuming that item" [24]. Undeniably, the value of a recommendation is to advise the user in advance of making a decision which can be useful in investing money, energy and time in consuming an item [24][23].

Normally, in ordinary life, we trust recommendations made by others since those recommendations appear as recommendation letters or book reviews in newspapers and magazines. Similarly, in the digital world, people use recommendation systems as trusted information from people's opinions (reputation) about other items including users, services or products they have used [25].

According to Resnick and Varian [25], recommendation systems are applications providing personalized advice for their users about different items (such as services, products, or even other users) that users might like or be interested in. Traditional recommender systems are mainly used to recommend products, services or people. Also, Resnick and Varian define recommendation systems as "systems where people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients".

In general, recommendation systems only work as a centralized service, especially on the web. On Amazon, for example, a well-known pioneer of recommendation technologies, users can rank products - and even rank other users. While this can obviously provide recommended products that meet users' interests, perhaps of equal importance is that it can in addition increases shared experience between individuals.

In some literature the focus is on meeting needs, rather than novel recommendation. Burke [26], for example, states that recommendation systems have become common in electronic communities and information access, because they aid users in finding items that match the users' needs.

Sarwar et al. [27] have proposed a recommender system architecture, where the system collects all ratings from different users and provides recommendations based on the past collected ratings as illustrated in Figure 1.

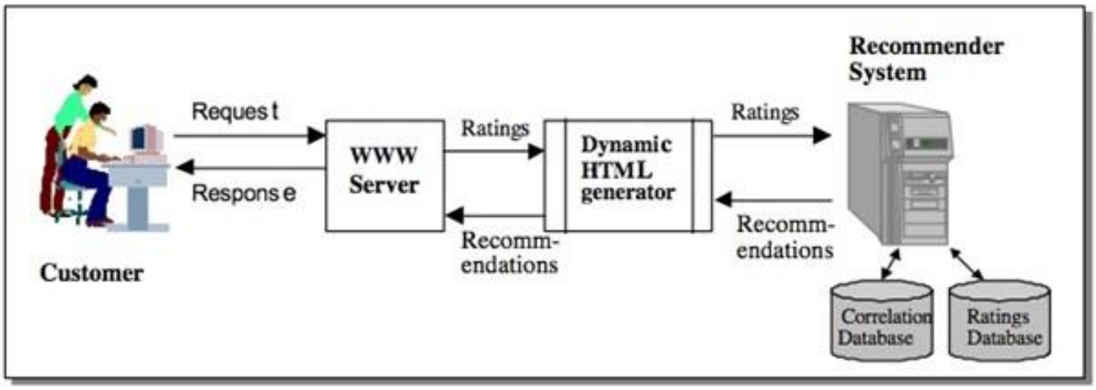


Figure 1: Recommender system architecture [27]

➢ Implementation of a recommendation system

Certainly, nowadays, a number of the most common companies and websites use recommendation systems to recommend products, services or people for their customers and visitors. Recommender systems like those deployed by *Amazon*[3] for books [28], *MovieLens*[4] for movies [29][30] or *Last*.fm[5] for music. Also, such systems are deployed to recommend training courses at *Emagister*[6] and *Coursera*[7], recommend vendors and products at *eBay*[8] [31], news and friends at *Twitter*[9] and *Facebook*[10] [32][33][34] and videos and channels on *YouTube*[11] [35], as well as studies of using *LinkedIn*[12] data to build recommendation systems [36].

Recently, there has been increased implementation of recommendation systems via the Internet in many different domain including movies, books, music and others [1] [2]. According to Park et al. [2], movie recommendation studies have gained the most research articles [37][38]; however, a great number of papers are based on different areas, such as books [39][40][41][42], music [43][44][45][46][47], web search [48], e-learning [49][50][51][52], documents [53][54][55][56][57], and e-commerce [58][26][59] among others.

---

3 www.amazon.com
4 www.movielens.org
5 www.last.fm
6 www.emagister.com
7 www.coursera.org
8 www.ebay.com
9 www.twitter.com
10 www.facebook.com
11 www.youtube.com
12 www.linkedin.com

2.2 Different Approaches for Recommendation Systems

The scale of implementation of Recommendation Systems is clear. There are, however, several different approaches used in recommendation systems to aid users in choosing relevant and suitable information, services, people or items for their requirements. Well-known recommendation systems techniques include collaborative filtering, content-based filtering, knowledge-based filtering, demographic-based and others. In some cases, in order to achieve better performance, different approaches are combined together as a hybrid recommendation system [26][60].

Despite the number of approaches, all of these built-in recommendation systems simplify searches based on sorted data in the system database.

The following approaches for recommendation systems will be investigated [26] [61]:

- Non-personalised
- Demographic-based
- Collaborative filtering
- Content-based
- Knowledge-based
- Utility-based
- Hybrid

Also, there are two extended approaches for each recommendation system:

- Emotional Intelligence
- Personality traits based

➢ Non-personalised:

One of the simplest recommendation approaches is the non-personalised approach. This approach depends on the top voted and most popular items, so it gives different users the same recommendations, as represented in Figure 2. Many websites recommend their most popular items or products, and YouTube presents the most viewed videos. Fan et al. [58] point out that the recommendations here depend on average ratings or total visits, and may not be appropriate for all users because of the lack of personalisation. This approach has the advantage of updating and continuously changing recommendations [61].

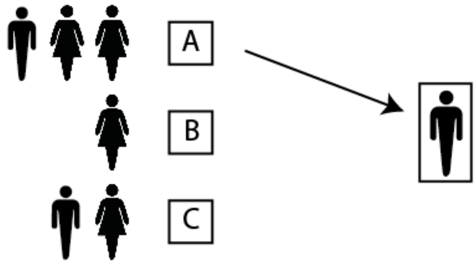


Figure 2: Non-personalised approach [61]

➢ Demographic-based

The demographic-based recommendation approach is based on customers' demographics; all their personal attributes are collected to categorise them in classes based on personal attributes. When a number of customers in one class prefer a product, the system will recommend the product to any customer in that class. So, recommendations are made by systems when there is a match between a user's rating history and another's rating history when they have similar attributes. Therefore, such systems need to know a user's personal attributes and identify users who are demographically similar to the active user [62][63][64][65]. Hiralall [61] points out that one of the advantages of this approach is that it does not need to take user-item ratings into account to recommend products to different users, as indicated in Figure 3.

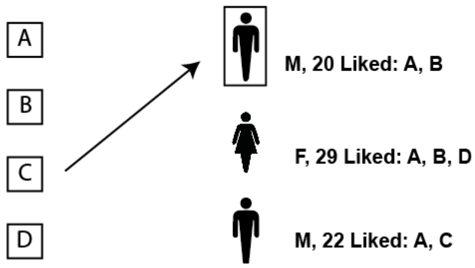


Figure 3: Demographic-based approach [61]

➢ Collaborative Filtering

This approach is probably the most common and successfully used approach in recommended systems. Collaborative recommendation systems collect all users' opinions about the products and then filter and analyse this data. Based on the result of analysing, in addition to past users' ratings, systems can recommend products based on the users' similarities [61] as illustrated Figure 4. In this sort of system, the system recommends to a user the items that other users with similar interests and preferences liked in the past. This means that the system identifies users who share the same tastes and preferences (e.g., based on rating pattern) with the active user. Then, the system finds products that users liked before and recommends these products to the active user [62]. Hence, the collaborative filtering approach plays a vital role in generating recommendations since it is used with other recommendation filtering approaches [66].

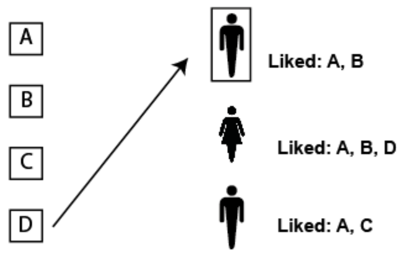


Figure 4: Collaborative filtering-based approach [61]

➢ Content-based

Content-Based techniques derive from a different assumption than other approaches. They assume that items will be rated similarly if they have similar *features*, whereas collaborative filtering assumes people will rate items similarly to other people with similar preferences, and neither filtering nor demographic-based approaches depend on any additional product details. The main factors in generating recommendations with the content-based approach are the item's description and the user's profile [61]. Therefore, recommender systems provide recommendations automatically to users by matching their interests with items' contents. Recommendations are based only on the users' profiles and items' contents without relying on information and ratings made by others.

There are two sources from which systems generate recommendations. The first source is features associated with items and the second source is ratings that a user has given to these items. Systems learn a classifier that uses user's rating behaviour and applies it to items [63]. However, in this approach only very similar products to previous consumed products by the user have the chance to be recommended. This leads to a problem of overspecialisation because the system will not recommend other products that are relevant and suitable to the user since they have not been rated before, resulting in recommendations that become impossible [67].

Figure 5 illustrates that, since product "A" and "B" are already preferred by the user, the system recommends product "C", because it is similar to what a user previously liked.

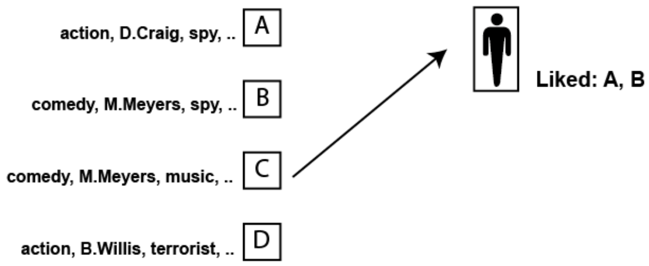


Figure 5: Content-based approach [61]

➢ Knowledge-based

The knowledge-based recommendation approach is based on different sources; it uses knowledge about items and users, and reasons that items match to the users' interests. Such systems have functional knowledge of users' requirements and past purchases. Burke [26] points out those systems applying this approach need knowledge engineering to figure out which items can match the user's interests.

Hiralall [61] points out that the advantage of this system is that it does not have to keep information about people for a long period of time. Any user can get recommendations immediately, as soon as they provide their preference [26][61]. Xia et al. [62] clarify that knowledge-based recommendation systems aggregate knowledge about users and items and use this knowledge in generating recommendations. Furthermore, these systems attempt to not make long-term generalisations about a system's users. knowledge-based recommendation systems need to have features of items and knowledge of how these items meet a user's preferences and needs as a background before inferring a match between users' needs and items [63][62]. Figure 5 illustrates how the knowledge –based approach works.

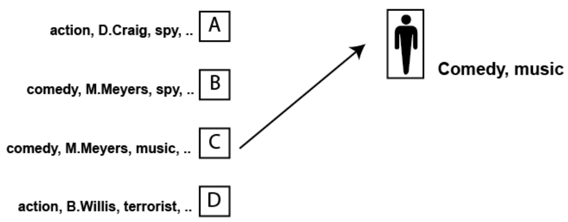


Figure 6: Knowledge-based approach [61]

➢ Utility-Based Recommendation

This approach applies a complete utility function over all items. The recommender system then makes a computation of this utility for a specific user. Also, this approach takes all features of the items into consideration. Therefore, non-product attributes can be considered such as the delivery schedule, vendor reliability, and availability to contribute the value of the product. Utility-based systems aim to calculate a utility for all products for each user. Also, utility-based recommender systems acquire the user profile for a user from the utility function which describes a user's preferences [26].

➢ Hybrid Recommendation

The recommendation systems described above have advantages and disadvantages. Much research has been done on the limitations of using only one of them, and hybrid recommendation approaches have emerged in order to solve the resulting issues, combining two or more different approaches [61][26]. Burke [63] explained the different hybridisation methods and techniques for hybrid recommender systems including (weighted, switching, mixed, feature combination, cascade, feature augmentation and meta-level). Also, two different kinds of filtering algorithms can be used for the same approach to generate a hybrid recommendation system. For example, Lu et al. [68] discussed both collaborative filtering based on user and collaborative filtering based on item; then they present a hybrid collaborative filtering method where the proposed collaborative based on both item and user algorithm to improve the recommendation accuracy.

Adomavicius and Tuzhilin [69] address four ways of how the hybrid recommendation system can be designed [69]:

1. **Combining Separate Recommenders:** Implementing collaborative filtering (CF) and content-based (CB) separately and combine their predictions.
2. **Adding Content-Based Characteristics to Collaborative Models:** Incorporating some CB characteristics into CF approach.
3. **Adding Collaborative Characteristics to Content-Based Models:** Incorporating some CF characteristics into CB approach.
4. **Developing a Single Unifying Recommendation Model:** Constructing a general unifying model that incorporates both content-based and collaborative characteristics.

The authors of [2] have studied the distribution of the research paper by application field from 2001 to 2010 in recommendation systems research, as represented in Figure 7. They found that the most research articles were focused on movie recommendation studies. However, as a summary of well-known recommendation systems techniques, Table 1 provides the advantages, disadvantages, and requirements for each recommendation system technique includes collaborative filtering, content-based, knowledge-based, utility-based, demographic-based and hybrid. Also, this table includes the studies, which have been published in each of these techniques.

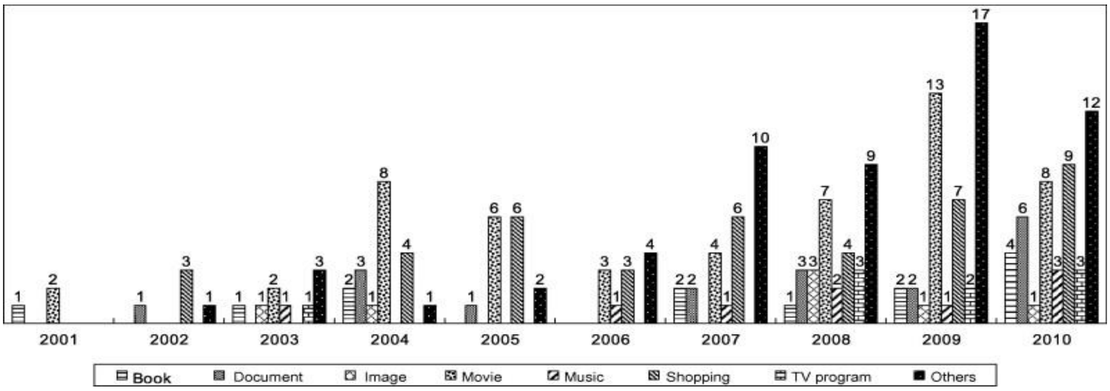


Figure 7: Distribution of research papers by publication year and application fields [2]

Table 1: Classification of recommendation systems research

Recommender System	Advantages	Disadvantages	Requirements / Background	Studies
Collaborative Filtering	○ Domain knowledge not required. ○ Quality improves over time.	○ Sparsity. [70][71][69] ○ New user/ "Cold-Start". [4][70] ○ New item / "First rater". [71] ○ "Gray sheep" problem. [71] ○ Stability.[26] ○ Quality dependent on large historical data set. ○ Insensitive to preference change.	○ A set of users with requirements and preferences. ○ A set users' ratings about products. ○ Large historical data set. PROCESS: Identify users in users similar to active user, and hypothesise from their ratings of item.	[18][72][73][24][65][28][27][62][60][17][30][35][26][61][74] [22] [19][75][76] [77] [78] [79] [80] [81] [82] [83][84] [85][86] [87] [88] [89] [90][70]
Content-Based Filtering	○ Domain knowledge not required. ○ Quality improves over time.	○ Overspecialization[69]. ○ New user Cold-Start. ○ Quality dependent on large historical. ○ Stability.	○ Well knowledge about user's preferences and interests. ○ Well knowledge of items' features and contents. ○ User's ratings of items. PROCESS: Generates a classifier that uses user's rating behaviour and applies it on items.	[65][60][26][62][61][67][21][91][22][67][69][74][92] [93] [94] [95][96] [17] [97] [98]
Knowledge-Based	○ Sensitive to interests / preference and changes. ○ Can map from user needs to products. ○ Can be initialised with no need to a database of user preferences.	○ Required knowledge engineering.	○ Knowledge of users' requirements. ○ Features of items. ○ Knowledge of how these items meet a user's requirements. PROCESS: Aggregate knowledge about users and items and matching between item and user's need.	[63][26][62][61][99][100][101][74][69]
Utility-Based	○ Can include non-product attributes, such as vendor reliability and product availability. ○ Sensitive to interests / preference and changes.	○ Must Creating a utility function for each user.	○ Features of items. ○ Understanding the utility functions of the user to describe and build a user' preferences and profile. PROCESS: Apply the utility functions to the items to determine each item's ranking and make recommendations based on computation of each item for the user.	[26][62] [74][102]
Demographic	○ Domain knowledge not required. ○ Quality improves over time. ○ Not necessarily to require a history of user's ratings.	○ New user Cold-Start. ○ "Gray sheep" problem. ○ Quality dependent on large history. ○ Stability. ○ Must gather demographic information which can vary greatly.	○ Personal attributes and demographic information about users such as gender and age. PROCESS: Identify users with similar demographical and personal attributes, and generate recommendations on their demographic classes.	[65][26][62][64][61] [74]
Hybrid	○ Combining Separate Recommenders. ○ Adding Content-Based Characteristics to Collaborative Models. ○ Adding Collaborative Characteristics to Content-Based Models. ○ Developing a Single Unifying Recommendation Model.			[26][63][71][65][68][103][65][95][104][105][106][107][108]

➢ Recommendation system extensions:

○ Emotional Intelligence

As illustrated in Figure 8, Burke [26] proposed five recommendation approaches; knowledge-based, utility-based, content-based, collaborative-based, and demographic recommender systems. Gonzalez et al. [109] proposed an extension for Burke' approaches. So, in order to improve the recommendations, he considered also the user's emotional intelligence aspects.

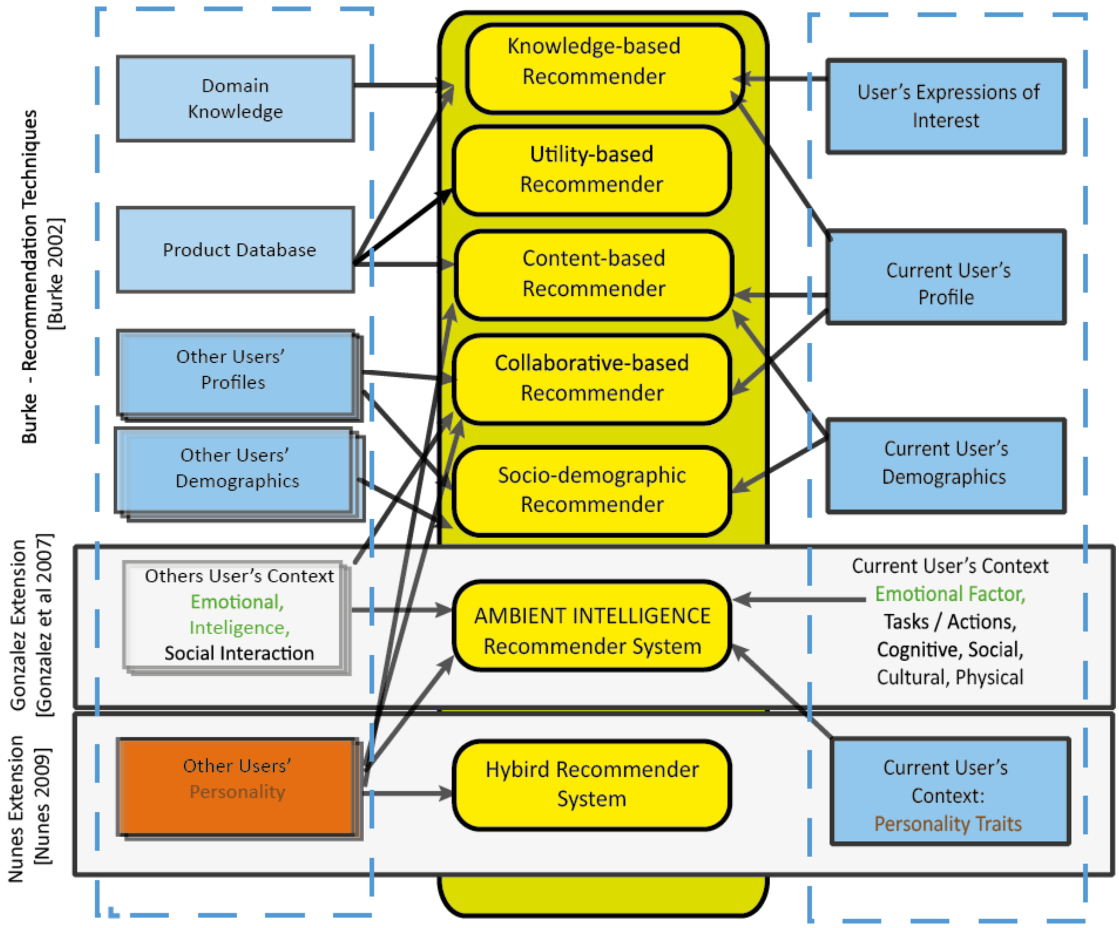


Figure 8: Emotional Intelligence and personality traits based extensions [109][110]

○ Personality-based

Nunes [110] also supposed that if a recommender system considers a user's personality traits it would provide a better experience for the user. Therefore, he proposed an extension to Burke' approaches and called it the personality traits technique with traits about users as an input and background to identify users who have similar traits to the current user.

2.3 Recommendation Algorithm Techniques

Now core concepts and approaches have been discussed, the algorithmic techniques underpinning those approaches are considered. Recall that the core function of recommendation systems is identifying the useful items for the user; recommendation systems therefore have a single, clear goal - they must be able to predict whether an item is to be recommended or not. In order to implement this, an algorithm must know the utility of items, users and ratings and how to compare them and then, based on this comparison system, decide which items are worth recommending. According to [66], a widely accepted taxonomy, recommendation techniques (methods) can be grouped into two classes (memory-based methods and model-based methods):

- **Memory-based (or heuristic-based) methods:** use only the matrix of user ratings for products and use any rating generated before the prediction process. So, the heuristic-based method makes ratings predictions by using the entire collection of ratings that was previously provided by users through a similarity function and its predictions are always updated. Therefore, it usually uses similarity metrics to be able to differentiate between levels of users and find the "*nearest neighbours*" for each user based on each of their ratios [69] [66][111]. The two most popular memory-based methods are:

  - **Pearson Correlation- based:**

This measures the extent to which users or items are similar or linearly related to each other. Therefore, it is a heuristic artefact used for finding the differences and similarities between levels of user; a set of "closest users". The main idea in this method is based on a rating database and an active user system to identify other users (i.e., *nearest neighbours*) who have similar preferences to the active user. Then, predictions will be computed for each item that the active user has not yet seen. So, it simplifies the rating estimation procedure.

The Formula (1) shows how to compute the similarity $sim(x, y)$ between user *x* and *y*, given the rating matrix *R.* The similarity between *x* and *y* is based on their ratings of items that have been rated by both of them. To present this, let $P_{x,y}$ be the set of all items rated by both users *x* and *y*. This means that$P_{x,y} = \{p \in P \,|\, r_{x,p} \neq \emptyset \,\&\, r_{y,p} \neq \emptyset\}$. Where the symbol $\overline{r_x}$ correspond to the average rating of user$x$.

$$sim(x,y) = \frac{\sum_{p \in P_{x,y}} (r_{x,p} - \overline{r_x})(r_{y,p} - \overline{r_y})}{\sqrt{\sum_{p \in P_{x,b}} (r_{x,p} - \overline{r_x})^2} \sqrt{\sum_{p \in P_{x,b}} (r_{y,p} - \overline{r_y})^2}} \qquad (1)$$

By calculating the similarities between users, the "*nearest neighbours*" for x will be taken into account to make a prediction for a new item that user x has not yet seen. Formula (2) can be used for computing a prediction for the rating of user x for item p will take the "*nearest neighbours*" N and x's average rating $\overline{r_x}$ is the following [19][24] [111].

$$pred(x,p) = \overline{r_x} + \frac{\sum_{y \in N} sim(x,y) * (r_{y,p} - \overline{r_y})}{\sum_{y \in N} sim(x,y)} \qquad (2)$$

- **Cosine-based**

This very prevalent technique is based on determining the angle between two n-dimensional vectors, and from this a similarity is computed. This method is commonly used in the information retrieval and text mining field in which documents are represented as a vector of terms or words. Therefore, systems can compare the similarity between two text documents by treating each one as a vector of word frequencies. In a recommendation system, this method is adopted to produce more accurate results by using the users' or items' ratings instead of word frequencies.

In cosine-based, the two users x and y are represented as vectors in n-dimensional matrix, where $n = |P_{x,y}|$. Therefore, in order to measure the similarity between two users x and y, the cosine similarity of the angle between them should be computed as follows:

$$sim(x,y) = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} . \vec{y}}{\|\vec{x}\|_2 \times \|\vec{y}\|_2} = \frac{\sum_{p \in P_{x,y}} r_{x,p} r_{y,p}}{\sqrt{\sum_{p \in P_{x,y}} {r_{x,p}}^2} \sqrt{\sum_{p \in P_{x,y}} {r_{y,p}}^2}} \qquad (3)$$

Where the "." symbol denotes the dot product between both vectors $\vec{x}$ and $\vec{y}$ [69] [66][111][76].

- **Model-based methods:** use the collection of information to create and learn a model to generate the recommendations through a similarity function and probability methods. There are a number of different widely used methods such as:

  - Bayesian classifiers [112]

- Neural networks [113]
- Fuzzy systems [114]
- Genetic algorithms [115]
- Matrix factorization [116]
- Latent semantic analysis [117]
- Latent semantic indexing (LSI) [118][27]
- Singular Value Decomposition (SVD) [119][120][27]

Park et al. [2], in their research they have classified the reviewed recommendation systems based on application fields and data mining techniques used in number of recommendation systems presented in Figure 9.

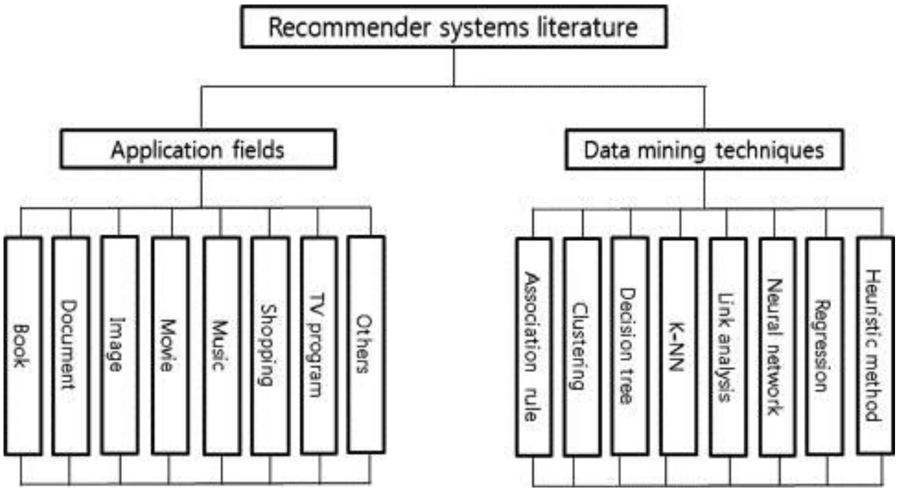


Figure 9: Classification framework for application fields and Data Mining Techniques [2]

2.4 Research Challenges for Recommendation Systems

2.4.1 Data Sparsity

In practice, many recommendation systems act upon immensely large data sets (e.g., Amazon) and they need to evaluate these datasets in order to produce recommendations. In any recommendation system, however, users provide only a very small number of ratings compared to what needs to be predicted. Therefore, one of the major challenges that appeared and is encountered by recommendation systems is *sparsity*, which has a great influence on the quality of generated recommendations. For example, in the movie recommendation system, most movies would not be rated by most of the users and their available ratings are sparse in the system. Hence, the lack of ratings leads to poor recommendations because many movies that have been rated by only a few users would have an unlikely chance of being recommended even if they had high ratings from users [96][121][69][27].

Collaborative filtering systems face the data-sparsity challenge and have difficulty when the rating is sparse since they depend on overlap in the rating matrix. Much research has been done to reduce the sparsity problem. Pazzani [65] attempted to overcome this rating-sparsity problem by using user profile information during calculating user similarity. This can raise the chance of two users to be considered similar even if they do not rate the same items similarly because they fall in the same demographic section. In the restaurant recommendation application, he takes into account different user's data such as age, gender, area code, employment and education information.

A different technique, which is called Singular Value Decomposition (SVD), was explored by Sarwar et al. [27] to reduce the dimensionality of recommender system databases. This technique has become a well-known method for matrix factorisation. In addition, a single unified framework, which incorporates user ratings, user features, and item features, has been proposed to solve the sparse ratings [122].

One of the examples of reducing *sparsity* is promoting users to rate items in order to improve their experience. For example, Amazon prompts their customers to rate items recently purchased as well as prompting users with different items to check whether the user likes them or not. This can help Amazon to get an idea about a user and build effective user profiles which can be applied as an input to a recommendation engine to help the customer find other items that a user might like. Therefore, customers are asked to invest effort in giving their opinions about a number of items, in exchange for which they get more personalised recommendations.

Figure 10: The Amazon.com rating page

❖ Amazon Betterizer: Improve Your Recommendations

Amazon tries to provide user with a rich interface that can help the user to adjust purchased and liked items. Therefore, Amazon tends to help the recommendation engine to provide users more personalised product recommendations.

Also, Amazon *Betterizer* which is a tool located in each user' Amazon store, can quickly let Amazon know which items and products the user likes, resulting in Amazon being able to personalise the user's shopping experience.

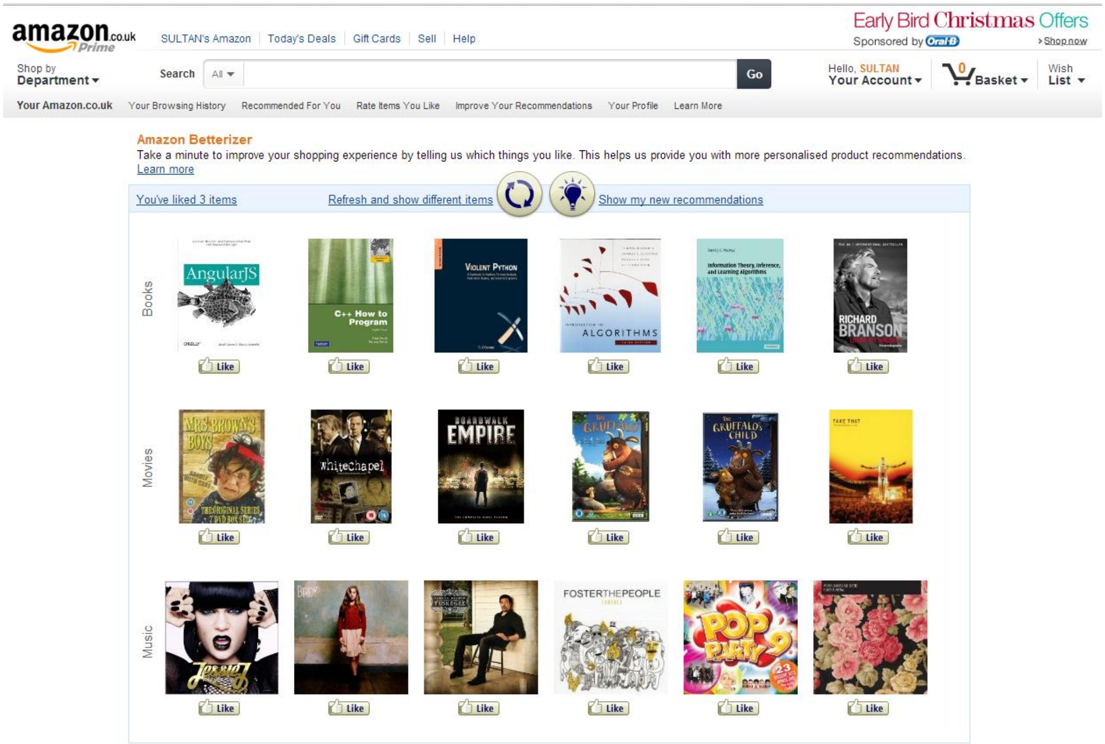


Figure 11: The Amazon Betterizer

2.4.2 Cold Start

Recommendation systems leverage their techniques to drive their mechanisms for identifying and serving relevant and personalised content to users. As such techniques however face the issue of cold-start where systems attempt to recommend content to users who have had little or no prior interaction with the system.

- Content Based (CB) :

In the content-based technique, it is crucial that users have to give sufficient ratings of items before the systems can build a user's preference. This can help a system based on user's preference in presenting reliable recommendations for the user. Therefore, a new user would face a problem of getting satisfactory and reliable recommendations since they do have no/few ratings in the system.

- Collaborative Filtering (CF):

Similarly, collaborative filtering systems have the same problem that the system must know the user's preferences in order to make good recommendations. Basically, a user's rating of items is the main technique that helps systems learn the user's preferences. Thus, in a new user case, the system will struggle to find other users who have similar tastes to the new user because of the lack of the new user's ratings. It is known that users share their opinion in the form of ratings which helps the system to predict their interests and preferences. This problem appears in collaborative systems since they rely on users' preferences which the system can build from users' ratings to make recommendations. Therefore, the recommendation system faces this problem when new items are added to the system and they do not have ratings. This results in not recommending them to users.

This challenge can be addressed by using a hybrid system since the content-based approach does not depend on other previously provided users' rating of items. Different techniques have been proposed to solve this problem; most of the proposed studies combine both collaborative and content-based techniques, known as the hybrid recommendation system, in order to avoid this challenge. Claypool et al. [71] proposed combining content-based with collaborative filters that can mitigate this short-coming. Also, other studies tried to solve this problem differently by using strategies such as: user personalisation, item popularity, item entropy and random strategy [123].

In addition, Zhang et al. [5] proposed bi-clustering and fusion (BiFu), a newly-fashioned schema, to overcome the cold-start problem based on the BiFu techniques under cloud computing setting.

2.4.3 Scalability

Recommendation systems deal with enormous growth in information of both the number of items and the number of users. Such systems suffer from serious scalability challenges. So, it is a great challenge for such systems to handle this vast amount of growing data in a refined way. Systems with millions of customers demand scalability with large datasets and can react to any change in online requirements and be able to make reliable recommendations. There are some techniques used to deal with the scalability problem such as Singular Value Decomposition (SVD) which is a dimensionality reduction technique [124]. Also, Condliff et al. [122] proposed a Bayesian methodology in which user ratings, user features and item features can be incorporated into a single unified framework to overcome scalability issues.

2.4.4 Limited Content Analysis (CB)

It is one of the shortcomings in content-based techniques that they are limited by features associated with items that they recommend. In such systems, it is important to have a sufficient set of features assigned to items in order to have satisfactory recommendations that meet a user's profile. Therefore, the system can extract items' features from text documents using information retrieval techniques. However, the features must be in a form that systems can parse them automatically, for example, a text form. Alternatively, they should be assigned manually by users in order to help the system have enough information about items to match them with the user's profile. Another challenge here is that automatic feature extraction methods are not that easy to apply to different types of data such as audio, video streams, and graphical images. Moreover, limited content analysis can lead to another problem when the same set of features is assigned to two different items. This can result in a content-based system that cannot sufficiently distinguish between different items.

2.4.5 Overspecialisation (CB)

Overspecialisation refers to the phenomenon when users are only recommended items that are very similar to those already rated and what users have consumed or liked. Also, content-based systems may suffer from this problem when they can only suggest items that score highly against a user's profile or his rating history. As a result, items that are relevant to a user cannot be recommended since the user has not rated them before. It is important that in some cases, such as a news article system, that the system doesn't recommend a news article if they are very similar to the event that the active user has already seen. This means that such systems should not only take into account items that are different from items the user has rated before but also, not recommend items that are too similar to what the user has already seen [95].

2.4.6 Gray Sheep

In some recommendation techniques, e.g. demographic filtering, the system may face the problem called "gray sheep" when a user falls on a border between two different cliques of users. This is a problem for demographic systems since they build their filtering on categorising users on demographic information and personal characteristics. Also, in collaborative systems a user, who can be fitted in one of the groups with many neighbours of similar preferences, can have a high quality of recommendations. Therefore, users, in the "gray sheep" situation would not benefit from collaborative filtering since this technique does not work well with such users. Using both a weighted average of the content- based prediction and the collaborative perfection can help in solving the "gray sheep" challenge [71][70].

2.4.7 Synonymy

This is when very similar items are represented in the system with different names. In most collaborative recommender systems, this latent association cannot be discovered, resulting in treating these items differently. So, the system, in this case, would not find a match between these similar products since they seem different to the system which will accordingly affect computing similarity between them. For example*, "Action film"* and *"Action movie"* appear seemingly different to the system while they are in fact the same. As a result, the system will find no match between these terms to compute correlation. Sarwar et al. [27] addressed the synonymy problem and proposed the Latent Semantic Indexing (LSI) method in LSI/SVD techniques to overcome this issue.

2.4.8 Privacy

While all previous challenges have centred on technical issues, there are also important considerations to be made in the nature and exposure of data being collected. It is clear that recommendation systems collect a user's ratings and combines them with other users' ratings in order to answer questions such as "Would user A like product P?" based on some statistical processes and predictions. Therefore, a user will receive anonymous recommendations which are generated from other users with similar tastes and preferences [125]. Seemingly, this process is innocuous since it is important to filter the available information in order to provide better personalised recommendations and enhance user experience by aggregating and processing ratings and preferences of multiple users through statistical database queries [126].

On the flip side, in recommendation systems, this can highlight the conflict between privacy and personalisation. Thus, the privacy protection problem is one of the concerns raised in using recommendation systems [25][127]. As mentioned before, insufficient data about users or items is a challenge for recommendation systems since they will suffer from different issues such as sparsity and cold-start shortcomings. It is more effective for such systems to collect as much data as possible about both users and items in order to build increasingly reliable recommendations [128]. Also, it is always desirable to design recommendation systems that can use not just large sets of data but rich data about users including their ratings and behaviour to help them receive more accurate recommendations. Nonetheless, users may not want their behaviours or views widely known. Accordingly, people need assurance that their personal data and privacy are being carefully protected.

Some recommendation systems allow a user to use anonymous participants which can give users more freedom of not caring about monitoring their behaviours and collecting data about their habits. This is not an acceptable solution since some users may desire to use their bank account details such as credit card information. Also, recommendation system sites can share information about their customers with other sites to have more information about their customers resulting in making better recommendations faster. But sites that are based on a wide range of customer data prefer not to share their customers' data because they feel that when other sites have this data can it would give others a more competitive advantage. "User control" has become an investigated issue in which a user should control his profile and ratings information.

Miller. et al., [129] defined the "user control" as:

> "The ability for a user to choose which of their ratings they would like to share with the community, and perhaps even which individuals they are willing to share their ratings with"[129]

They investigated a variety of available algorithms and concluded that there are two main challenges involved with user control. Firstly, to develop algorithms for recommendation systems that can provide control choices. Secondly, to develop interfaces that allow individuals operate those choices in current recommendation systems.

Indeed, recommendation systems must know some information about the users in order to generate more personal recommendations. Therefore, the more the recommendation systems know, the better personalised recommendations they can generate. As a result, E-commerce sites can have a great deal of information they have collected about their customers even without properly informed consent or awareness of the customers [130]. Generally, people everywhere are generating a huge amount of data about themselves and their behaviours, often without being fully aware of what sort of data they generate and how important these data are [131]. Skatova et al. [132] found that some people consider some of their personal data as risky to share and would pay to secure such personal data (for example digital communication history). About 80% of the participants in the study would pay up to £20 for protecting their social networking profiles and activities; 60% require medium protection online purchase history [132], as represented in Figure 12.

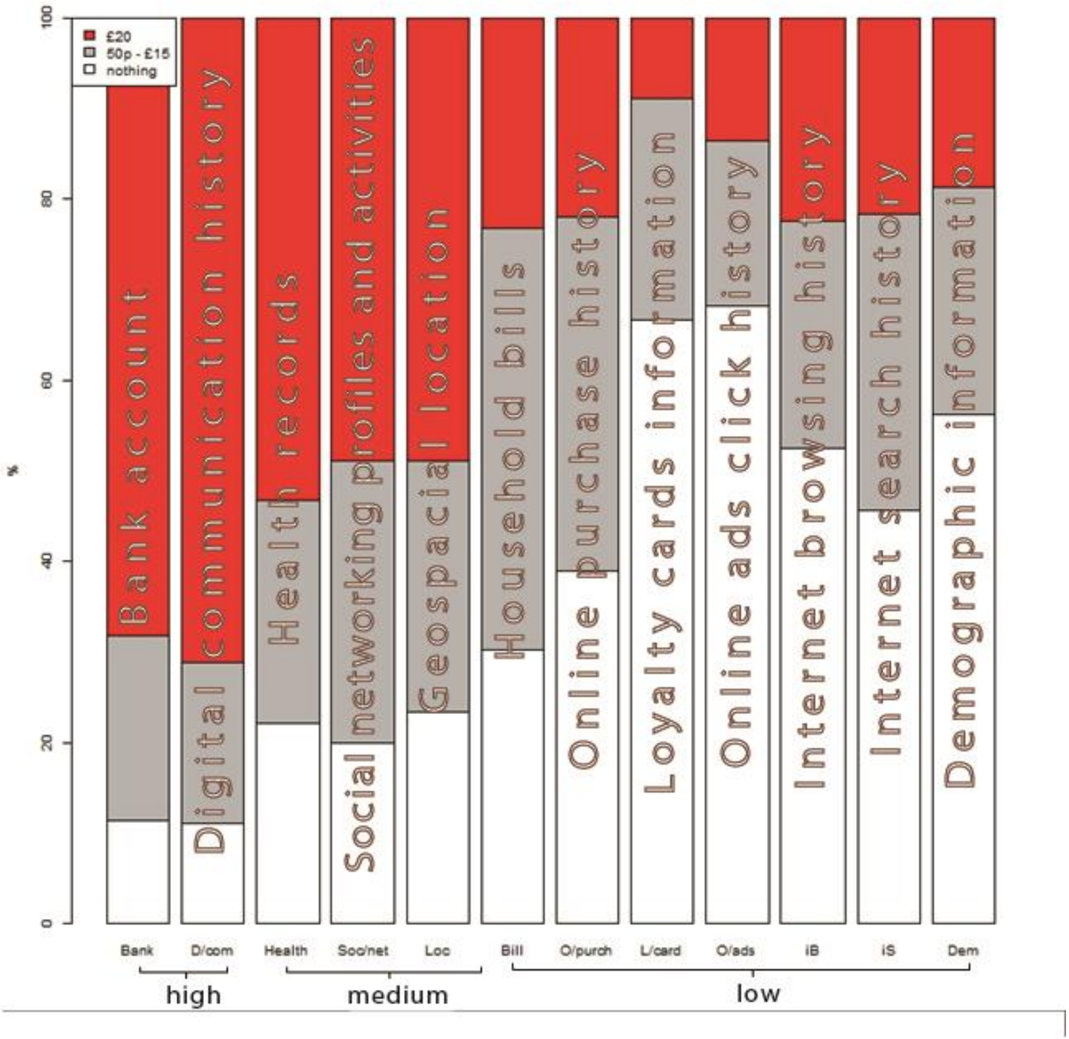


Figure 12: Percentage of participants binned by how much they were ready to pay (£20, red bars; from 50 pence to £15, grey bars; nothing, white bars) to secure different types of personal data [132]

Schafer et al. [130] examined **the types of personal data**, which customers make available to the sites while they are shopping.

- **Explicit preference information:** product ratings, comments, attributes of interest.
- **Implicit preference information:** products viewed, time spent viewing, searches performed.
- **Transactional information:** purchased products, payment information, account numbers, shipping addresses.
- **Explicitly identification information:** name, phone number, e-mail address, address.
- **Implicitly identification information:** IP address, machine name.

There are a number of studies in this area focusing on ways to preserve users' privacy in recommended algorithms [133][134][135][126][136] and techniques [137][138][139][140]. Miller et al. [129] argue that one of the limitations of recommendation systems is that they require the customer to trust the owner of the recommender with personal preference information. As a result, whether users' personal data is stored locally or shared in encrypted form, user's privacy should be protected. Therefore, *PocketLens* was proposed as a new collaborative filtering algorithm to run in a peer-to-peer environment and enable individuals to choose some of their ratings to share with others. [141] proposed a *sparse fac*tor analysis which can also protect user privacy since it supports computation on encrypted individual data [141]. McSherry and Mironov [126] show how the Netflix Prize[13] algorithms can be modified to the framework of differential privacy which can provide privacy guarantees. Also, a context based personalised privacy settings recommender system, which can measure user's privacy with respect of his friends, has been proposed to help users set their privacy settings [142].

2.5 Summary

This chapter reviewed the recommendations systems in the literature in terms of the background of the recombination systems as well as the common approaches and algorithms techniques. As pointed out in this chapter, there are many challenges in this field and people generate a huge amount of data about themselves and their behaviours, often without being fully aware of what sort of data they generate and how important these data are. In addition, we use several systems and websites when handling our public and private information, but we do not always consider who is recording this information. Therefore, very large and rich data silos and information about our lives have been made available online (e.g., Facebook)—whether we realise it or not. More opportunities for mutually beneficial exploitation of digital personal data exist. The following chapter will discuss personal information and social networks used in this thesis.

[13] www.netflixprize.com

3. Literature Review: Use of Personal Information for User Modelling

3.1 Overview

During the past few years, the use of cloud computing has increased along with the use of online social networks. Online social networks have become a mainstream cultural phenomenon and have gained a huge user base and have become phenomenal technologies in the 21st century [13][143].

As a result, thousands of companies are basing their interests and investments on social networks because these networks allow them to easily target their customers by providing services and building applications that attract and reach those users. Many of these applications and services expose users' information. Therefore, the rising use of online social networks has led to an increasing number of privacy issues of which users are not fully aware. After all, users share their personal and private information on different social network websites.

This has become an increasingly active field of research, with much literature written on the subject. Cavoukian [144] states that companies and websites that provide such services—along with other third parties who help us to deal with different online social networks—are in charge of controlling most of our data.

Therefore, during the past few years, many studies and projects have been done on the concept of personal data stores or personal data 'vaults'. These projects aim to help individuals to take control of their personal data. This means that a shift will occur, where the company will no longer be in charge of controlling individuals' data; rather, individuals will be empowered to collect and to control their own data as well as to share them for their own purposes. A number of companies are working on such projects, such as Paoga and Mydex in the United Kingdom (UK).

3.2 Social Networks

During the past few years, social networks have become widespread, gaining increasing acceptance on the Internet. These online social networks have witnessed phenomenal development and a clearly explosive increase [16].

Globally, social network sites are one of the most visited websites and remain a remarkable technological phenomenon in the 21st century. For example, more than a billion daily active users on average for September 2016 share different aspects of their lives on Facebook [145] ; also Foursquare has gained 50 million users during its lifetime of 8 years [146]. Furthermore, 29.57 million users arrange 486,309 monthly meetups at Meetup [147]. More than 313 million monthly active users share their thoughts on Twitter [148]. In addition, YouTube has over a billion users, and users watch hundreds of millions of hours and generate billions of views [149]. A significant amount of user-generated content is stored on such sites.

According to Jamie Bartlett [14], in the data dialogue, people are living in an age of sharing. For reasons of convenience, necessity, security or choice, we are either required to share or choose to share our personal information more than we ever have before. Bartlett also states that the information we share tends to be one of two types:

- Personal information: This information is provided as we shop and subscribe to services both online and offline; this sort of information directly identifies us, such as by our telephone numbers, bank details, home/work addresses and so on.

- Behavioural data: These data are generated as we spend more time connected to the Internet; we create more information over time, which may or may not be anonymised and aggregated when stored and analysed. Examples of these data are locations and purchasing or browsing histories [14].

Although the online social networks have proved to be useful in many ways, such as for keeping in touch with friends, for sharing ideas and interests and for research collaboration, they have acquired a number of privacy threats as well. Users who generate data on these networks are unaware of who is recording what information about them. Also, people everywhere are generating such data without being fully aware of the degree to which their privacy depends on third-party applications [131].

Therefore, the increase in private data online has led to growing privacy concerns for online social networking users, and some users have experienced the negative effects of sharing their private information on these online networking sites. For example, students, after posting some photos on their Facebook walls, have been fined for their online social behaviour [150].

Giles [143] in the paper 'Security Issues and Recommendations for Online Social Networks' has listed 15 privacy-related threats, including targeted scams, cyber-bullying, unwanted stalking, reconstruction of users' identities, and child solicitation.

A growing number of companies have been building databases on users' details in order to track users online. *The Wall Street Journal* (*WSJ*), in its 'What They Know' series, studied the privacy policies of the 50 most-visited websites and analysed the tracking files installed on users' computers. Also, the journal built an 'exposure index' in order to determine the exposure of each site's users to monitoring by exploring the installed tracking technologies and by studying the privacy policies that guide their use. This study analysed three types of tracking methods commonly used online: HTML cookies, Flash cookies and beacons [151].

3.2.1 User Expectations

Users have strong expectations in terms of privacy on online networking sites, which results in their using and trusting such sites with no privacy concerns. Still, many users wonder why they see advertisements on their private pages that address some of their interests, such as advertisements that appear on their Facebook pages. They do not realise that this means that users' private personal profiles have been exposed and revealed publicly. In some cases, private user data may be sold to advertisers and marketers [152][153]. In addition, some websites track users' behaviour and gather information about users. Therefore, some websites can collect extremely detailed users' profiles about users' shopping habits by gathering data from Facebook partners' advertising sites. This means, users' information in recommendation systems can be exposed and shared with other sites. Doubtless, sharing information about customers between sites can result in making more accurate recommendations. However, it is considered as a disclosure of a users' privacy since users' personal information is leaked without permission [154][130].

3.3 Personal Information

Basically, we are enriching our footprints by generating a huge amount of data about ourselves and our behaviours without realising it. Therefore, users do not generally know what sorts of data are generated about them and how important this data can be. By using social networks, users essentially hand out public and private information to different parties but do not think about who is recording that information[131] .

Therefore, modern life involves individuals who play roles in the generation of data, whether this data is generated by individuals or about them. For instance, we can create our own data through e-mail accounts and social network accounts, or companies or governments can generate this data via health records and bank transactions, for example [155]. These data, no matter how they are generated and whether we control them or not, are still considered our data. Basically, users' online data are presented through Web pages on many platforms (phones, tablets or computers). However, users still do not have full control of the data collected or created about them because the data holders are the ones who control how these data are presented to individuals.

Cavoukian [144] states that service providers and other third parties control most of our data, so they can create value in what is being called 'the age of Big Data'. Therefore, personal data created by and about people can be used to generate 'a new wave of opportunity for economic and societal value creation', says the WEF [8].

Now, companies and organisations collect and disclose personal data for their own purposes. This means that individuals are not fully in charge of controlling their data; also, they often do not know who can share and access this data. All transactions with individuals and the access to their data happen according to the organisations' privacy policies. Therefore, organisations can profit from individuals' personal data. The current technology has facilitated 'business to business' (B2B) transactions, allowing for the automation of daily transactions and operations on behalf of businesses while they involve users only occasionally.

3.3.1 Recent research on Personal Information

Many studies have been conducted on personal data. Cavoukian [144] discusses the personal data ecosystem (PDE), which is the emerging of organisations that believe individuals should control their personal data and believe in developing various technologies and tools to make this happen (Figure 13 contrasts this PDE to the current *status quo*). He also compares the current use of personal data with PDE. The starting point of the PDE is that individuals collect and share information only for their own purposes, thus resulting in individuals being the central point of control and data integration. This means that individuals can establish and define rules regarding the use of their personal information according to their personal privacy settings. Moreover, individuals always have the ability and the right to extract their data and to take it wherever they wish [144].

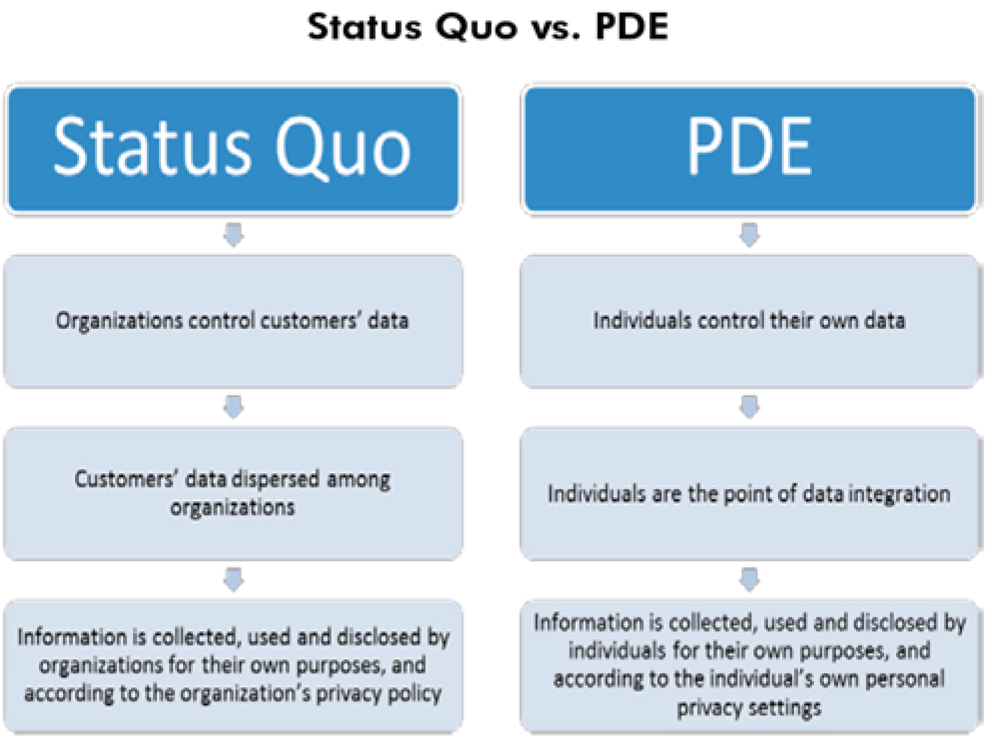


Figure 13: Status Quo vs. PDE [144]

The *control shift*[14], which is a series of interconnected shifts in the relationship between individuals and organisations, is associated with personal data in several ways:

- The organisation as a manager of customer data as well as individuals who manage their own data.
- Information as a tool in the hands of the organisation as well as in the hands of the individual. In [156], they outlined different types of data and showed the individual as a data manager where the personal information management services help the individual gather, collate and use information about their lives.

The project '*midata*' is about putting individuals in charge of controlling their personal data, which may help to inform their decisions on purchasing and other lifestyle choices.

Therefore, developers and researchers have looked for an alternative architecture that would put individuals in charge of controlling their data

[14] www.ctrl-shift.co.uk

because the centralised services have many privacy threats. Therefore, they have provided different solutions, such as vendor relationship management (VRM) systems, personal data storage (PDS) and distributed social networks. These approaches have faced some challenges, and many papers have discussed this critically [157] [158] [159]. Mortier et al. [160] proposed a system called the 'Personal Container' to help individuals to collect and to manage data collected by and about them. They investigated how to build an ecosystem around an individual's data while allowing the ecosystem to support current and new applications and services. Figure 14 illustrates how the Personal Container fits into the ecosystem of users, data sources and third parties.

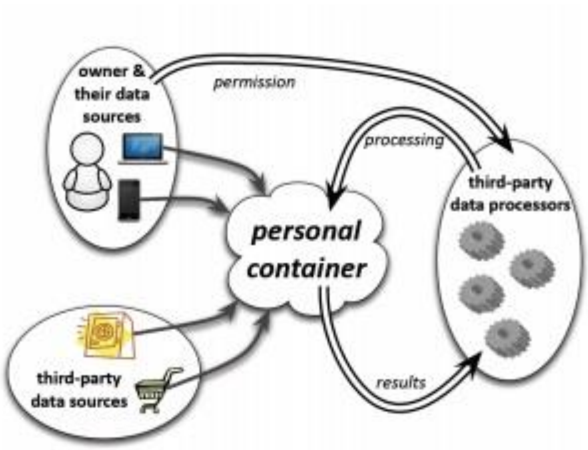


Figure 14: The Personal Container in context [160]

It is clear that we are increasing our footprint online because we all have a large amount of data spread throughout different online websites and services; some of them are stored on Hotmail, eBay, Amazon, Twitter, Google Docs and other online services. As I discussed above, regarding the importance of online data, we need an effective and secure way to protect these data. Liu et al. [161] proposed a personal data storage system, UPDS, to help individuals to protect and to improve their data's availability and durability. The first step to building a personal data storage system is to understand the system's design space; three fundamental questions must be considered in system design:

- Where does personal data storage store data?
- What does personal data storage store?
- How does personal data storage store data? [161].

3.4 Summary

As pointed out in this chapter, social networks and personal data has become an increasingly active field. The concept personal data have been actively studied in the literature and the personal data generated about individuals and their behaviours has been involved in many studies and projects. Both the recommendation systems and the use of personal information have been reviewed in the previous chapters. In the following section, cross-domain filtering in the literature will be covered.

4. Literature Review: Cross-Domain Filtering

4.1 Overview

As a final section in this extensive literature review, and having contextualized both the current situation in recommender systems and the handling of personal information, I move on to the key topic for this thesis – the use of information that crosses multiple systems or domains. As detailed, recommendation systems have been extensively studied in the literature, with a multitude of distinct approaches emerging, all with the aim of identifying apposite information, services, people or items relevant to a given user's requirements (common techniques discussed in chapter 2 included collaborative filtering [14], content-based filtering [162], knowledge-based filtering [61], demographic-based filtering [62], and hybrid approaches that combine several of these techniques are also prevalent [57] [22]).

Recently, a new approach has been introduced - cross-domain recommendation systems have been gaining popularity due to the overwhelming growth of available cross-domain data in e-commerce sites and social networks. This could play a significant role in personalising different types of items of many domains with no need to seek users to provide separate feedback in every single domain they have experienced with. Nowadays, users provide feedback for different types of items for instance; in Amazon, they can rate movies, books, and DVDs. Moreover, users can also express their opinions publicly on different social media networks and different providers such as Facebook, Twitter and TripAdvisor. This raises a need for an approach providing personalised services taken the advantages of all the available information and help in providing cross-sell products, services and recommendation to new users [3]. Cross-domain recommender systems may have been taken into account as an effective solution due to the results of market research (e.g. [163]) that demonstrates the effectiveness of promoting a user with products from different domains if the products fit the user's shopping patterns.

4.2 Users provide - problems

Cross-domain recommendation systems have been researched from different perspectives with a diversity of research areas. According to [164] the term "cross-domain recommender" appeared for the first time in a patent for Triplehop Technologies (now Oracle) and some other papers suggested "cross-domain" as an interesting topic such as [165]. However, the first papers with a contribution on "cross-domain" were in 2007 by [166] [167]. Also, Loizou's Doctoral Thesis [168] tried to classify the problems and approaches of cross-domain recommendations. Some recent papers also present a review of the state-of-the-art of cross-domain system and established a general taxonomy of problems and techniques for better categorisation [9][169][170]. For example, Winoto and Tang [171] examine the similarity of domains by conducting a study that allows users to evaluate some recommendations in twelve domains. They addressed three important issues a) examining the correlation of user interest in items in different domains, b) constructing models that are able to exploit user interests, and c) developing evaluations for recommendation systems. Other studies, however, aim at designing a unified user model across domains such as [140][172][173]. Another line of research focuses on applying machine learning methods to enable cross-domain applications without overlapping users between domains [174][175].

4.3 Domain levels and goals

Domains differ because of the different types of items that they are concerned (e.g. movies vs. books) or the types of users who interact with them. In [164], Ivan and Paolo [164] define four levels of domains along with the percentage in the literature 1) Attribute level: where same type of items with different values of certain attributes (e.g. comedy – thriller) 12% , 2) Type level: where similar types sharing some attributes (e.g. movies – book) 9%, 3) Item level: when there are distinct types with differences in most attributes (e.g. movies – restaurants) 55%, and 4) System level: with almost the same items collected in different ways or from different providers (e.g. Netflix – MovieLens) 24%.

According to [164], cross-domain recommendation systems aim to achieve a set of goals based on the literature as following:

- Addressing the cold-start problem including generating a recommendation to new users and cross-selling products.
- Improving accuracy of recommendations by reducing sparsity.

- Offering added value to recommendations such as novelty, diversity and chance.
- Enhancing user models by discovering new user preferences.

As stated in [164], the priority of these goals across the literature is predominantly focussed on improving accuracy, however, as illustrated in Figure 15.

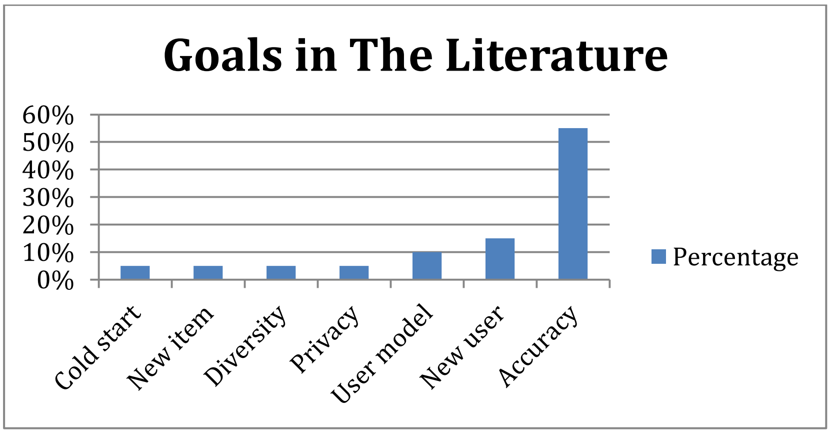


Figure 15: Goals in the literature

4.4 Approaches

Research in cross-domain recommendation systems is relatively new and there are a number of studies that identify a main approach forward and best practices for implementation. Loizou [168] categorised the development of cross-domain recommendations in three main research trends:

- Compiling user profiles into a unified cross-domain user profile;
- Deploying a web user agent for profiling user's interests by monitoring their transactions and behaviours;
- Improving the performance by generating recommendations from multiple domains using cross-domain profiling information.

The authors of [164] identified two main cross domain approaches including Linking/aggregating knowledge and Sharing/Transferring knowledge as illustrated in Figure 16, which are broken down in the next sections:

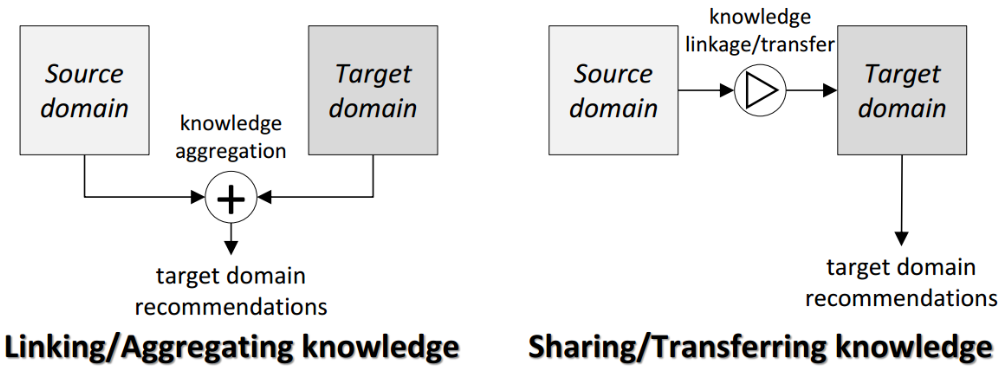


Figure 16: The main cross-domain approaches

4.4.1 Linking /aggregation knowledge

- Merging user preferences

This approach aggregates user preferences using ratings, tags, logs or click through data. It works well for the new-user problem but requires user-overlap between the source and target domains. Different techniques can be applied on the aggregate matrix such as user-based k-nearest neighbours (kNN) [171][166][3], graph-based techniques such as those used in other studies [176][169][177] or matrix factorization (e.g. as applied by Loni et al. [178]).

- Mediating user modeling data

Based on using user similarities and user neighbourhoods, this approach aggregates models from different domains which can overcome the new-user problem and improve accuracy. However, user overlap or item-overlap is still required between the source and target domains. Latent features [179], user neighbourhood [166][3][180] and collaborative or content similarities [172][166] can be aggregated in this approach.

- Combining recommendations

This approach is easy to implement since it aggregates single-domain recommendations using ratings and probability distribution. Also, it is independent of content and increases the diversity, but it faces difficulties in tuning weight assigned to recommendations coming from different domains in addition to the need of overlapping of users. This approach can be implemented by aggregating estimated values of ratings [181] or combining estimation of rating distribution [182].

❖ Linking domains

Another approach of cross-domains is linking domains by a common knowledge using item or user attributes or semantic networks. This approach can be implemented without user or item overlapping but it is difficult to generalise. Another disadvantage is that this approach is designed for particular cross-domain scenarios. Fernandez et al. [183] published a semantic-based framework for the cross-domain recommendation. Other studies such as [184][12][185] overlapped social tags while [167] used an overlap of user/item attributes.

4.4.2 Sharing /transferring knowledge

❖ Sharing latent features

Unlike linking knowledge, sharing latent features is used when the source and target domains are related by means of shared latent features. Based on common latent features shared and learned from the data over all domains, user and item latent features of the source domain are used to regularize latent features in the target domain. This approach works well to reduce sparsity and increase accuracy. However, there is a need to overlap users or items and it is computationally expensive [186][187].

❖ Transferring rating patterns

In this approach, rating patterns need to be transferred between domains with apparently no need for user or item overlap. This approach is also computationally expensive. The authors of [188] and [189] proposed a cross-domain recommendation by transferring cluster-level rating patterns.

4.5 Summary

This literature review has shown that Recommendation Systems are virtually ubiquitous on the Web and there has been increased implementation of such systems in fields as diverse as movies, books, music and others. While such systems have been widely researched, several issues remain unresolved. Recent researches, however, have proposed a potential solution to these issues in the form of cross-system and cross-domain user modelling. This requires the coordination of vast datasets and is contingent on the economies of scale which allow their construction.

There is therefore a gap in research that focuses on user-controlled cross-domain recommendation, where models are generated through a user's own data. This thesis will address and fill in the remainder of this thesis by extending prior work via a cross-system approach using social media data to model user preferences and providing empirical evidence of the effectiveness of the approach based on direct real-world user feedback. Having established this research focus, the next section will introduce the methodology used to investigate the issues involved.

5. Experimental Methodology

5.1 Background and Motivation

In order to explore the effectiveness of using external communities in providing more accurate recommendations, the methodology employed in this thesis to help me implement and test different types of modeling is now presented. Underpinning this methodology is the aim of supporting comparisons, and therefore our experimental system must be able to generate multiple user models for the same individual. In the following experiments, our experimental recommendation system is able to consider the cross-system process by retrieving data from other domains to draw resources from. Moreover, the profiles are automatically constructed through social network APIs and then used in presenting user's interests for finding best match recommendations.

Many recommendation systems are based on different approaches that collect insufficient data about individuals, which results in poor recommendations. The aim of this research is to assist in the improvement of the efficiency and accuracy of recommendation systems. It seems likely that cross-system recommendation can only help the traditional recommendation approaches, both in accuracy but also because it that it is preferable to generate recommendations by using external information, and not only that which is centralised in a recommendation system database.

5.2 Experimental platform

An experimental platform was developed to investigate the effectiveness of constructing a cross-system model mined from social media data to model user preferences and to evidence the potential value of passive mining of web behaviour. So, this thesis will explore the design, implementation and testing of a cross-system approach using social media data to model user preferences and provide empirical evidence of the effectiveness of the approach based on direct real-world user feedback.

The platform was setup to allow construction of different user-models. As illustrated in Figure 17, the user modeling module will define different user models based on the study. Therefore, these models can be constructed in different ways. For example, it can be constructed through an n-gram vector space representation derived from social media streams or explicitly defined declaration of categorical user interests.

Application of these models to dynamically rate the relevance of articles supplied to it via any RSS feed. Based on the model, the most relevant articles will be delivered and presented to a user via a web interface that allowed for relevance feedback ratings.

5.2.1 System Architecture

The architecture of our platform is illustrated in Figure 17. Its functionality is split into five modules. The **content module** is used to retrieve our item sets via external RSS feeds, collating results. For each document in the item set the module extracts appropriate features, and from these generates and stores a compressed model for that item (more details of this process are provided in 5.2.3). In parallel, the **user modeling module** is responsible for hosting potentially multiple user models for the current participant and, in the case of this research, the module is also tasked with constructing these models (as described in more detail in 5.2.4). With this data in place, the **recommendation module** becomes a core module in the process - it is responsible for computing similarity calculations between output items and user models across the two domains. The recommendation module is further responsible for sorting the resulting output feeds, and for selecting a particular user model when it is requested to make some recommendations. It then serves feeds to a user via a **presentation module**, which also allows user-evaluation to occur. Once this evaluation is provided an anonymised log is recorded in the **logging module** ready for post analysis.

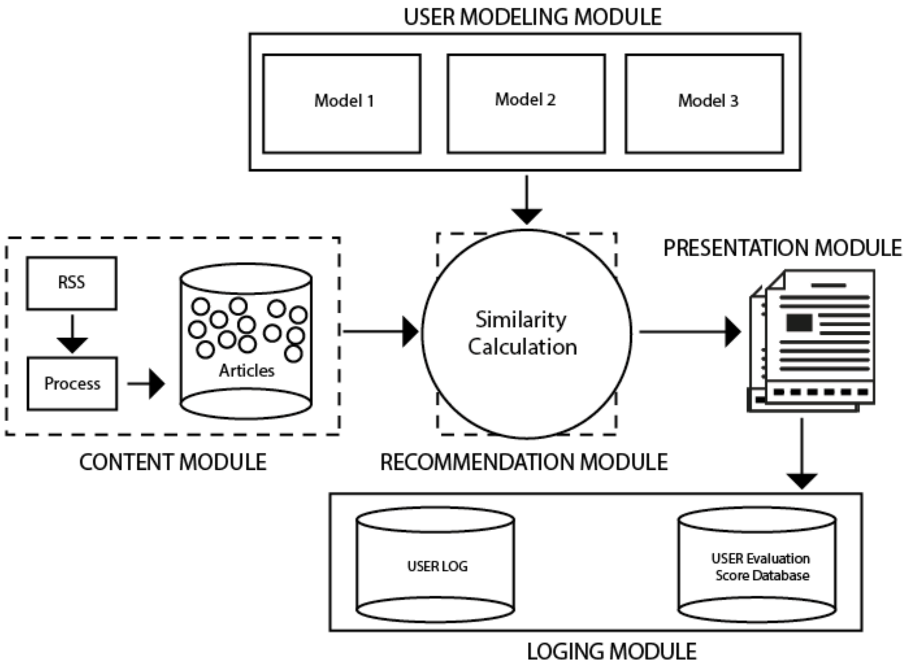


Figure 17: Architecture of experimental platform

5.2.2 System Design

To classify any of the items that I may or may not recommend, all content items were stored in our experimental system's database according to their category label extracted from the RSS feed. These items and categories were retrieved from all the external sources and feeds in advance in order to prepare the datasets for other processes. The system needs to be ready with all data items before running the processes. The system consists of four layers:

- Database layer: includes basic user database, user evaluation ratings and RSS news feed storage.
- Processing layer: performs the processing of datasets and accessing user's profile of an external community including computing the TF-IDF weights and similarity calculation.
- Application layer: allows the user to sign up with the system and to authenticate the system to access the external resource via API.
- Interface layer: helps the user to connect with the system and receive a number of recommendations via a web interface.

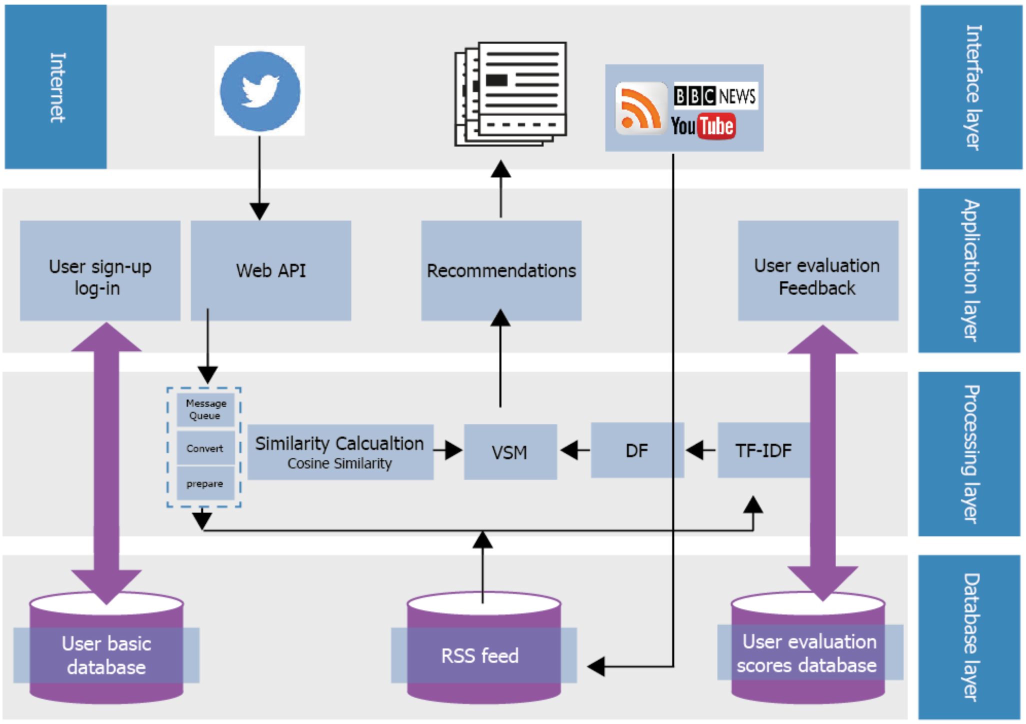


Figure 18: Design of the system

5.2.3 Document Modelling

Based as it is on content recommendation, our experimental system currently leverages a traditional vector space model (VSM) to represent documents. The content module, therefore first extracts a series of documents from the content source via an RSS feed endpoint. The XML documents which are delivered via the RSS feed are parsed and processed. Each document, *d*, is systematically transformed into a corresponding vector space model, $\vec{d}$ before subsequently being stored along with the URL for its source content.

The VSM represents textual information within each document by first constructing a dictionary of feature n-grams, F that are present across the whole corpus of documents, D. For any given document, d I iterate through every n-gram t within F and assign it a weighted value, $W_{t,d}$. This value reflects the term's importance within that document. The simplest approach to assign such a value would seem to be to use the term's frequency - however, due to the need to represent a document by its most identifying features a range of weighting schemes have been developed [190]. Due to its popularity in the literature, in this work I employ the traditional Term Frequency-Inverse Document Frequency (TF-IDF) statistic to measure term expression. TF-IDF is commonly used within text mining and information retrieval systems, and has proven an effective empirical method for term weighting from a probabilistic point of view [190].

Within this statistic term-frequency (TF) is simply a counting function, representing how many times each term appears in the document. The inverse document frequency (IDF) is a measure of how much information the word provides, that is, whether the term is common or rare across all documents. This is formulated by finding the percentage of documents in which the term appears, inverting it and logging. Thus:

$$\boldsymbol{TF - IDF(t) = TF(t, d) \times IDF(t, D)}$$

$$= \sum_{x \in d} \boldsymbol{f(x, t)} \log \frac{\boldsymbol{N}}{\boldsymbol{1 + |\{d \in D : t \in d\}|}} \quad \textbf{(1)}$$

where d is the document under consideration, N is the total number of documents in the corpus, |{d ∈ D : t ∈ d}| reflects the number of documents where term t appears in document x, and $f(x,t)$ is simply the function:

$$f(x,t) = \begin{cases} 1, & if\ x = t \\ 0, & othewise \end{cases} \qquad (2)$$

Thus it can be said that the TF-IDF score for an n-gram, t, takes: 1. a higher value, if t occurs numerous times within a small number of documents; 2. a low value, if t appears fewer times in a document or occurs in many documents and 3. a low value, if it occurs in almost all documents. By transforming a document, in practical applications the set of features, F, can become extremely large. This leads not only to issues of sparsity and computational efficiency, but also reduces effectiveness of similar comparisons due to the curse of dimensionality. It is therefore desirable to reduce the dimensions of the vector space by removing irrelevant and redundant features [191]. Responsibility for modelling of the document dataset in this manner is handled by the system's content module, as detailed in Figure 17[15] [192].

5.2.4 User-Preference Modelling

In order to support comparisons, the experimental system must be able to generate multiple user models for the same individual. While produced via different techniques, all models for any given user are constructed simultaneously when he/she first registers with the system (as represented by the user modelling module in Figure 17). Some require explicit input on the part

[15] In practice this is achieved via the functionality available in the python scikit-learn libraries (available at http://scikitlearn.org).

of the user, and if so this is facilitated via the web interfaces during registration. Other models require access to cross-domain data streams generated by the user and which are subsequently mined passively. Currently, the models available to the system are as follows:

Passive user-modelling: On registration, the user is asked to provide controlled access to an active social media account, which is then mined to generate a linguistic preference model. The user's social media posts are collated and parsed into a bag-of-words representation (an n-gram frequency model) of their term usage. This bag-of-words is then cleaned in exactly the same manner as with output documents described in 5.2.3), before being encoded as a vector-space model using TF-IDF weightings. I henceforth refer to the resulting model as PASSIVE.

Manual user-modelling: Here a user's preferences are modelled via the presentation of a pre-constructed set of document categories and sub-category labels, and asking the user to rate their level of interest in each. These categories might be selected didactically to reflect the document categorisation approach used by the RSS source, or generated dynamically by parsing the corpus and modeling the k most common topics that emerge [193]. Note that, unlike the other two user models, in this case data must be actively supplied by the user through a web interface when they register onto the system for the first time. I denote this model as MANUAL within our experiments.

Random user-modelling: In order to provide a baseline, the system will also generate a random preference model for each user. This is essentially a VSM, containing the same dimensionality as the passive user model, but with TF-IDF scores set randomly for each feature in F. Use of this model to assess the relevance of any particular article should, therefore, produce random recommendation results. I refer to this model as RANDOM.

Once the models have been established, the documents in dataset D are ranked according to their relevance to the user. For the three models detailed above, this will yield three document rankings. The mechanism by which this is achieved is detailed in the following section.

5.2.5 Determining Document Relevance

Recommendation occurs when users begin interacting with the system. Participants are presented with a stream of n documents, each in turn and each generated by one of the available user models (which model is selected for each recommendation is specified by the testing regime). To achieve this, prior to presentation the recommendation module must generate multiple rankings of all documents in D , one for each user model that may be used during the experiment.

For categorical models, documents are ranked according to the number of labels each individual document is tagged with that match the user's explicitly declared categories of interest. For vector space models (i.e. the Passive and Random user-modelling approaches detailed in §3.3), each document, d is ranked by calculating the similarity between its VSM, $\vec{d}$, and the user's preference-model, $\vec{p}$. The relevance score used here is the traditional cosine similarity measure:

$$\text{similarity}(\vec{d}.\vec{p}) = \cos(\vec{d}.\vec{p}) = \frac{\vec{d}.\vec{p}}{\|\vec{d}\|\|\vec{p}\|} = \frac{\sum_{i=1}^{|F|} w_{d,i} w_{p,i}}{\sqrt{\sum_{i=1}^{|F|} w_{d,i}^{2}}\sqrt{\sum_{i=1}^{|F|} w_{p,i}^{2}}} \qquad (5)$$

5.3 Content Sources

Each study has different sources in order to implement its experiments and support comparisons. These resources used for constructing multiple user models. Some of these sources retrieved user's data streams such as Twitter or Facebook posts. However, BBC news feeds, which used as main source of output documents in the following studies.

- **RSS**

The content pushing service (CPS) delivers subscribed messages for users; one of the popular CPS applications is RSS. RSS (Rich Site Summary, Really Simple Syndication) uses a family of standard formats to publish frequently updated information. It is also called an RSS feeds which contain full or summarised text. The eXtensible Mark-up Language (XML) is emerging as a standard for the description, storage and exchange of semi-structured data published on the World Wide Web and organised into XML documents [191][194].Therefore, RSS is represented in an XML format and enables publishers to automatically syndicate data and also benefits users by helping them in receiving frequently updated data from their favourite websites or to aggregate data from many sites, instead of checking the website for new content. In order to use RSS, an RSS reader was developed to retrieve BBC feeds in different categories and store them in our database as illustrated in Figure 19.

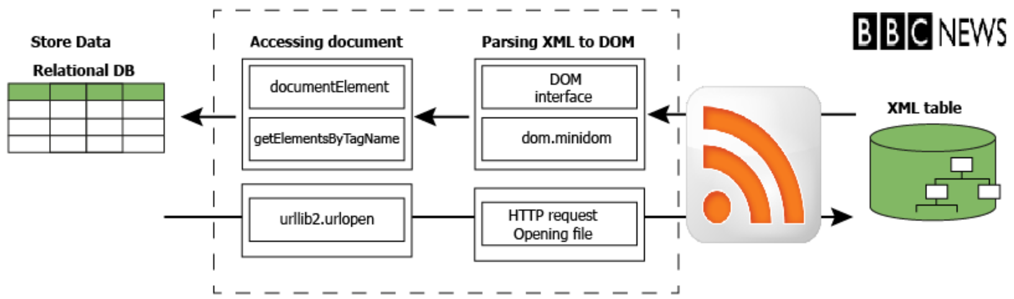


Figure 19: Retrieving a BBC newsfeed Python API

5.4 Recommendation process (Experimental Method)

The recommendation process can be seen as consisting of four stages, as shown in Figure 20. First, the complete RSS news feed, the domain from which to generate resources for the recommendations, needs to be retrieved. In this experiments, the BBC News RSS feed[16] was used as output documents to assess recommendation performance of each of our user models. In this stage also, the profiles of an external community (or social network), used to generate passive-user models, are authenticated and processed. Then, the whole retrieved data is cleaned and aggregated in order to be used in the analysing stage. The process of analysing and matching will then take place in order to compute the similarity of recommendations. Based on the user modelling module, recommendations will be filtered. As a result, users will be served a number of feeds in the presentation stage allowing user-evaluation to occur.

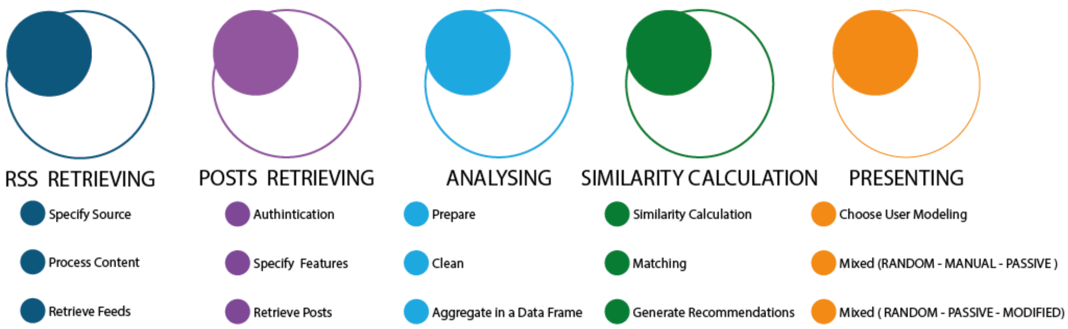


Figure 20: The main process of the recommendation

[16] specifically via the BBC's RSS feed endpoint, which is available at http://feeds.bbci.co.uk/news/rss.xml

5.5 Methods for evaluation:

This section provides a brief description of the recruitment strategy and the evaluation of each study and outlines the methods used to carry it out.

- Recruitment strategy:

Participants were recruited through printed advertisements on notice boards at various places at the University of Nottingham. Also, the following studies were advertised via *Call For Participants*[17] which is online service to promote research and recruit participants. After contacting the experimenter, participants received the participant information sheet explaining the procedure and the goal of the study (see the Appendix B). Participants were informed that the study focused on using social network data (i.e. Twitter and Facebook) in personalising delivery of BBC news feed. Participants who are actively engaged in social media were recruited in this study. When they were eligible and still willing to participate, they were invited to participate in a lab-experimental study. Participants were compensated with a £10 Amazon voucher for their participation.

Ethical approval of research was required to follow the University of Nottingham Code of Research Conduct and Research Ethics. It was obtained by the School of Computer Science research ethics committee at the University of Nottingham.

5.5.1 Study 1

In this study, Twitter was used to construct passive-user models via a user's Twitter stream. In order to support comparisons, the experimental system in this study must have the ability to conduct different user modules for the same user. Therefore, three different models were constructed including random, manual and linguistic models (passive-user models). The BBC News RSS feed as our output documents, from which to draw resources for recommendations, was used to assess recommendation performance of each of our user models.

The application of this study then was run involving 40 participants, and which generated linguistic user-models through mining of participants' Twitter streams. Testing of content-based recommendation accuracy was performed using BBC News RSS streams and produced a total of 1,800 recommendation evaluations. The application of the experimental approach is described in more detail in Section 6.

[17] https://www.callforparticipants.com/

5.5.2 Study 2

In this study, user profiles used in building passive-user models were conducted via a different social network source. For those joining this study, Facebook was the source used to generate each user profile. This time more models were added to this experimental platform by adding the top stories model where the sources tagged as top stories are nominated by the BBC. As a result, more news feeds are added to the recommendations source by retrieving the top stories feeds and updating the database of our output documents. This study extended the cross-domain user modelling approach using social media data (Twitter streams) in the previous study by concentrating on social connections and the text generated therein, rather than microblogging.

In this study, 23 participants were involved in experimental runs. Testing of content-based recommendation accuracy was performed using BBC News RSS streams and produced a total of 1,380 recommendation evaluations in the experiment. More details of this experiment are provided in Chapter 7.

5.5.3 Study 3

In order to compare both passive user models used in Study 1 and Study 2, 20 participants were involved to generate two different personal passive-user models using different social networks which generated linguistic user-models through the mining of participants' Twitter streams and Facebook posts. This study also investigates the feasibility of utilising available personal information from different social streams for producing well-matched recommendations and reflecting users' interests.

To achieve this, I put users in control of their profiles generated from different sources with the ability to evaluate and change their profiles followed by a post study. So, I provide the ability for a user to maintain control over such a model to investigate user responses when allowed to sculpt, edit and reflect upon that profile.

As a result, a new updated profile was generated which was then used to produce recommendations for each user to evaluate the effectiveness of giving the user the ability to update his/her passive-model. Testing of content-based recommendation accuracy was performed using BBC News RSS streams and produced a total of 1,800 recommendation scores. The experiment discussed above is carried out in Chapter 8.

5.6 Summary

This chapter detailed the methodology used to implement the different experiments. The experimental platform and the recommendation process to assess the various aspects during running our studies have been outlined. The methods of evaluation and the recruitment strategy also were provided in this chapter. The experiments discussed above are outlined with more details in the following chapters.

6. Study 1: Microblogging

6.1 Overview

It is increasingly rare to encounter a Web service that does not engage in some form of automated recommendation, with Collaborative Filtering (CF) techniques being virtually universal as the means for delivering relevant content. Yet several key issues still remain unresolved, including optimal handling of cold-starts and how best to maintain user privacy within that context. Recent work has demonstrated a potentially fruitful line of attack in the form of cross-system user modelling, which incorporates features generated from one domain to bootstrap recommendations in another.

This study evidences the effectiveness of this approach through direct real-world user feedback, deconstructing a cross-system news recommendation service where user models are generated via social media data (in this case mined from publicly available Twitter stream).

However, in this study I also ask the question - can a user's interactions with the Web and social media be leveraged in order to produce a cross-system user model that actually out-performs explicit filtering using self-declared preferences? And, if so, why?.

It is shown that even when a relatively naive vector-space approach is used; it is possible to automatically generate user-models that provide statistically superior performance than when items are explicitly filtered based on a user's self-declared preferences. Detailed qualitative analysis of why such effects occur, indicate that different models are capturing widely different areas within a user's preference space, and that hybrid models represent fertile ground for future research.

6.2 Experimental Platform

This section details the experimental platform developed to investigate different forms of cross-systems user modelling and recommendation. This work is focused on the specific research question:

- *"Can a cross-system user model mined from social media generate more accurate article recommendations than a model of user preferences that has been explicitly self-declared?"*

If true, this would provide contributing evidence as to 1. the effectiveness of cross-system approaches; and 2. the potential value of passive mining of web behaviour to generate user-preference models. In order to answer this question the platform was set-up so that it could:

1. Construct a model of user interest either:
    - via an n-gram vector-space representation derived from social media streams.
    - through an explicitly defined declaration of categorical user interests.
    - using random parameterisation (to serve as a baseline for our testing procedure).
2. Use one of these models to dynamically rate the relevance of articles supplied to it via any RSS feed.
3. Deliver the most relevant articles to a user based on one of the above models, presenting items via a web interface that allowed for relevant feedback ratings.

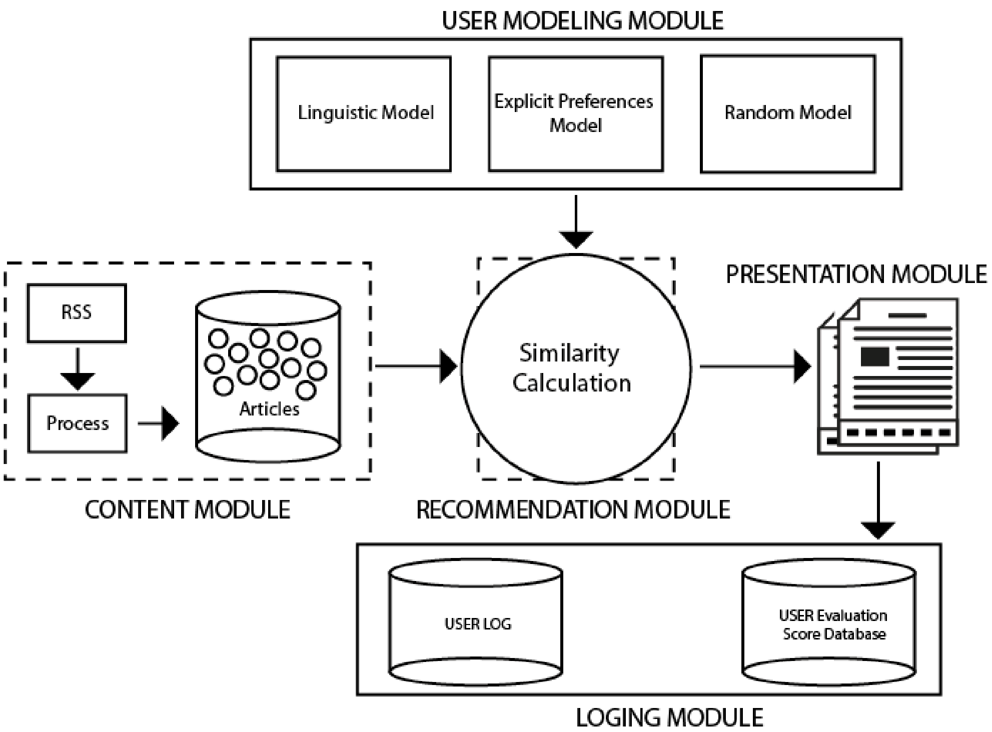


Figure 21: Architecture of experimental platform presenting the three user-preference models

6.2.1 User-Preference Modelling

In this study, multiple user models for the same individual would be constructed by the experimental system in order to support comparisons. The main user models described in section 5.2.4 are used. While produced via different techniques, the following models are constructed simultaneously when he/she first registers with the system (as represented by the user-modelling module in Figure 21). Currently, the models available to the system in this study are as follows:

- **Passive user-modelling:** User's Twitter posts are collated, and parsed into a bag-of- words representation (an n-gram frequency model) of their term usage, cleaned and finally encoded as a vector-space model, again using TF-IDF. For more details see section 5.2.4
- **Manual user-modelling**: here a pre-constructed set of document categories and sub-category labels are used to model user preferences based on the user rating level of interest in each as with user-modelling described in section 5.2.4.
- **Random user-modelling:** For providing a baseline, a random preference-model will be generated for each user as with random-modelling described in section 5.2.4.

Once models have been established, the documents in dataset D are ranked according to their relevance to the user. For the three models detailed above, this will yield three document rankings. The mechanism by which this is achieved is detailed in section 5.2.5.

6.2.2 Presentation and Logging

When a user interacts with the system, articles are presented to them via a web interface. For each article delivered the following process occurs:

1. First, one of the system's user-models (as detailed in section 6.2.1) is selected, the distribution of these selections being stochastic and/or defined by the particular experimental regime being used.
2. For the model selected, its highest recommended document that has not yet been viewed is identified.
3. That document is retrieved using its associated URL and presented to the user as illustrated in Figure 22. The document is accompanied by a 7-point Likert scale that serves as a mechanism for the user to evaluate its relevance.
4. The user's assessment rating is then collected and stored in a database for analysis.
5. The process then iterates, with a new user-model option being chosen, and a new article presented.

Figure 22: An example news article being presented

6.3 Experimental Method

In our experiments, passive-user models were generated using a user's Twitter stream. The BBC News RSS feed provided our output documents, and were used to assess the recommendation performance of each of our user models. In order to set-up our experiments, four stages of preparation took place, as indicated in Figure 24. The first task was to retrieve data from the BBC News feed. By running its reader over the BBC News feeds, the system's content module extracted 2,180 articles. Each document was transformed into a corresponding VSM, and stored along with category metadata, and the source article's URL. Simultaneously, the explicit labels that each document could be tagged with by the BBC were extracted - these were: Technology, Science, Environment, Entertainment, Arts, Education, Family, Health, Politics, Business, UK, and World (Figure 23 illustrates the distribution of articles that formed our document set, D).

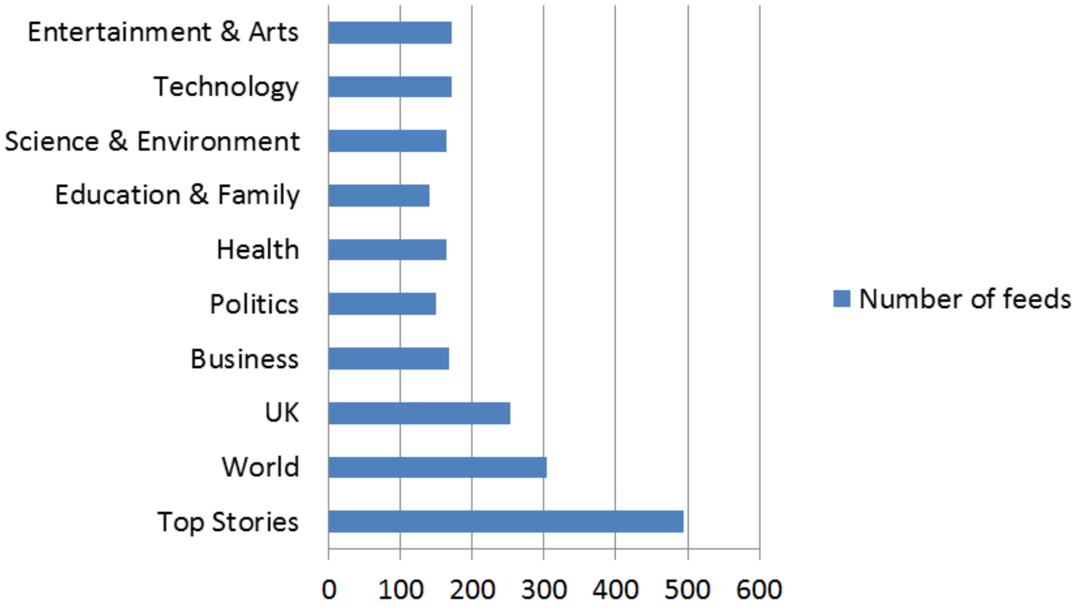


Figure 23: Distribution of the article tags as retrieved

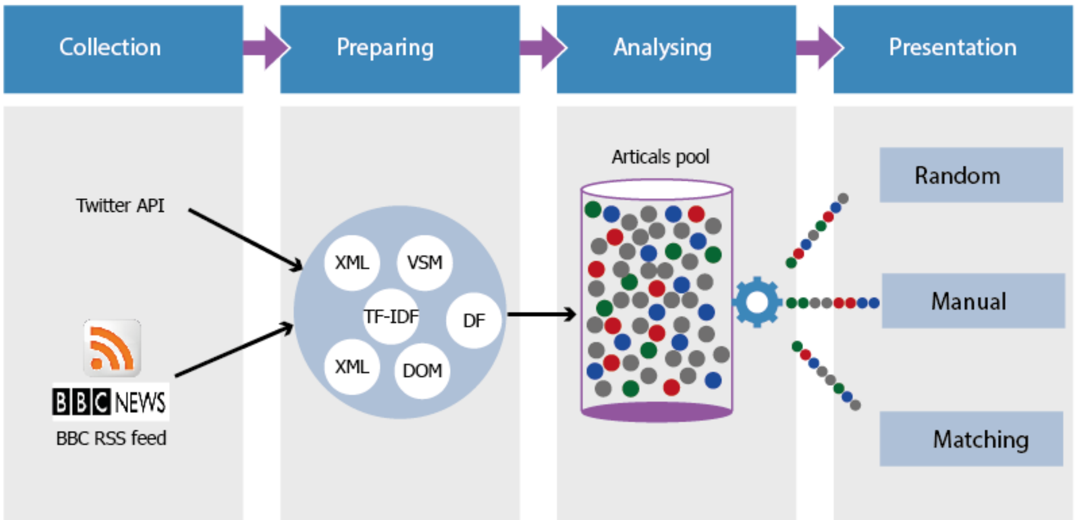


Figure 24: Stages of preparation for a user experiment

40 participants from a wide range of backgrounds were registered with the system. Out of the 40 participants 23 were male and 17 were female. Most participants were students and spanned different majors from geospatial tech, user-centred design, computer science, software engineering, human factors, chemistry, geography and other fields. Each participant was required to be an active Twitter user and to have posted a minimum of 150 tweets. They joined between March 2009 and November 2013. The maximum number of tweets of any user since signing up to the service was 46,700; the mean was 6,479 and the standard deviation 10,332 (note however that in these experiments each user's last 150 tweets only were considered when constructing user models).

In this study, I tried to avoid the selection biases that can occur including non-representative sample and non-response bias. It is difficult to conduct experiments that are completely representative of the population. However, the participants were not selected from specific groups and do not exclude certain groups from the research. In order to minimize the non-representative biases, participants from different ages, fields and backgrounds were recruited in this study. In terms of a non-response bias which is about the members of the sample that choose not to join the study who may have certain characteristics that will prevent me from inferring parameters about the whole population and creating a representative sample. Unfortunately, this is something that it is unlikely to be controlled. In addition, participants rated the same number of items in the available user-models. These items were randomly generated from the three user-models. This can be useful in minimizing and decreasing the reasons that can cause responses bias.

I used the University of Nottingham Code of Research Conduct and Research Ethics provided by the School of Computer Science research ethics committee at the University of Nottingham to recruit participants. Participants were invited to participate in a lab-experimental study and were paid a nominal amount for their time.

On registration, each user authenticated the system's Twitter application in their Twitter account, and then explicitly declared their categorical interests via the web interface (as selected from the tags used by the BBC News feed). This allowed Manual, Passive, and Random models to be automatically then constructed for each user. In order to construct the passive model, the Twitter API [18] was interrogated with the user's recent history of available tweets (statuses) being extracted, cleansed and transformed into a VSM model (as detailed in section 5.2.3). Our own experimentation showed that exceeding the most recent 150 tweets showed rapidly diminishing marginal returns.

All participants then engaged in a lab-based task experiment where the system presented them with a total of 45 articles. The participants had to evaluate in sequence the relevance of these articles to themselves. To choose the next article to be presented, the system selected a random user-model from the three available, determined that model's highest ranked unseen document, and presented it to the user along with a Likert evaluation scale. In cases of a tie between a set of documents in any ranking, one was selected at random. Each

[18] https://dev.twitter.com/rest/public

user was ultimately presented with an equal number of articles for each model. Thus, while ranking is deterministic for all models used, each experimental run would still be stochastic in nature in terms of the ordering of articles presented. For each document presented, the system recorded the participants' user-id, the presented article's document-id of the article presented, the user's evaluation score and an identifier of the model used to select that article (either PASSIVE, MANUAL or RANDOM). This data collection process allowed me to test the following hypotheses:

H1: That the mean rating recorded for articles selected using the MANUAL model would be the same as that of the RANDOM model used as our baseline (i.e. a hypothesis aimed at determining that any sort of filtering is better than simply serving random BBC News articles).

H2: That articles recommended by the PASSIVE model would not record higher mean ratings than those recommended by the RANDOM model.

H3: That articles recommended by the PASSIVE model would not produce higher ratings than those selected by the MANUAL model (i.e. a hypothesis aimed at determining if a managed, implicit filtering approach is superior to a filter based on explicit statements of preference).

At the end of this experiment, I was able to test these hypotheses against the 1,800 data points produced (40 people × 45 ratings).

6.4 Experimental Results

6.4.1 Summary

Results for our user experiments indicated that the PASSIVE generated the most relevant news item recommendations for users in comparison to both MANUAL and RANDOM models. The mean relevance scores recorded for our baseline model (RANDOM) was 3.81 points. Recommendations generated via automated preference models (MANUAL) were rated at an average of 4.13 across 600 evaluations. The mean rating of implicit/linguistic filtering model (PASSIVE) was 4.30. In 85% of rating cases, a user-model improved over the baseline random recommendation. Standard deviation of evaluations was relatively low for all models at 0.912, 0.850 and 0.814 for RANDOM, MANUAL, and PASSIVE models respectively.

For an overall view, Figure 25 shows the mean average scores for random, explicit and implicit filtering. It can also be seen that the majority of the articles selected by the PASSIVE model were favoured most by participants - however this was not the case across the board (and in rare cases random selection was favoured [19]). Also of note, was a surprising lack of correlation (r = 0.05) between model performance and the number of tweets the participant had posted over their whole lifetime. Long term Twitter activity (which I view as a proxy for Twitter experience) did not seem to have any impact on the performance of the PASSIVE model.

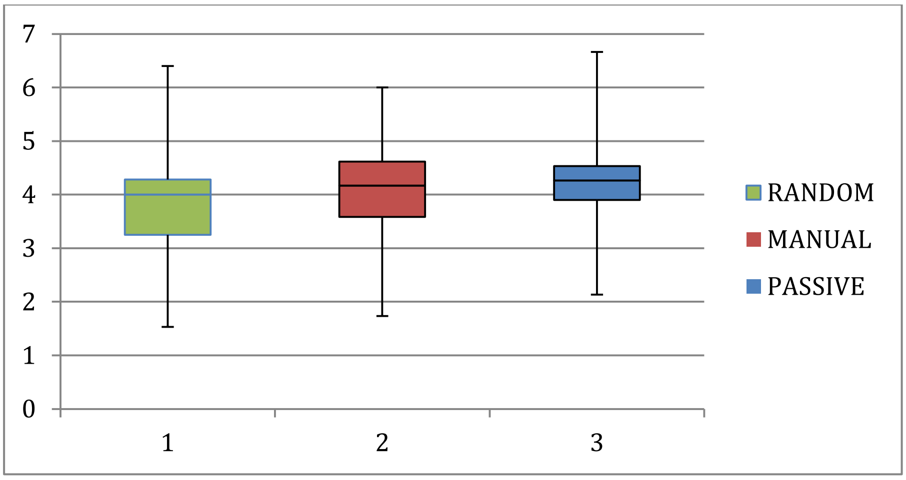


Figure 25: Participants' mean evaluation scores across PASSIVE, EXPLICIT and RANDOM models

6.4.2 Wilcoxon Signed Ranks Test

So, a Wilcoxon Signed Ranks Test [195] was run on the 40 Twitter user participants to determine whether there was a statistically significant mean difference between the average score when explicit information was used in filtering (i.e. manual filtering) compared to random recommendations (this study measured user response using 7-likert-scale, so Wilcoxon tests are preferred to a Bonferroni corrected paired t-tests due to the ordinal nature of the scale).

Participants logged a mean evaluation of 4.13 for MANUAL and 3.81 for RANDOM. Test results produced a value of p = 0.002, indicating that the difference in means of 0.32 represented a statistically significant increase (Z = -3.127, p = 0.002). Thus I was able to reject hypothesis H1, and conclude that the MANUAL model was producing superior performance.

Similarly, for hypothesis H2, participants expressed preferences for PASSIVE models (4.30) as opposed to the RANDOM baseline (3.81); a statistically significant increase of 0.486 (Z = -4.098, p = 0.000042, p < .05) indicates

[19] It is noted that this was expected in some instances, due to both the stochastic nature of the experiment and the pre- filtering of all BBC news stories for interest by their editorial staff.

again that there is a statistically significant difference between our two variable scores.

Finally, I was able to show a statistically significant preference for the use of passively mined personal information via the PASSIVE model (4.30) as opposed to the MANUAL filtering (4.13); a statistically significant increase of 0.17 ($Z = -2.045$, $p = 0.041$), $p < .05$. This indicates that use of passively mined personal information generates further statistically significant improvements and that the PASSIVE model was the most effective in our tests. Full details are provided in Table 3.

As detailed in Table 4, there are 28 cases where the PASSIVE filtering recorded higher than RANDOM filtering and only 9 cases recorded negative ranks while only 3 cases have no changes. Also, the PASSIVE Model recorded 26 positive cases better than the MANUAL filtering and the sum of ranks resulted in 258.00 negative ranks and 562.00 for the positive ranks.

Table 2: Wilcoxon Signed Ranks Test Statistics

	M - R	**P - R**	**M - P**
Z	-3.127	-4.098	-2.045
Asymp. Sig. (2-tailed)	0.002	0.000042	0.041

Model R = RANDOM, M = MANUAL, P = PASSIVE

Table 3: Ranking Results

		N	Mean Rank	Sum of Ranks
M - R	Negative Ranks	**11**	16.14	177.50
	Positive Ranks	**29**	22.16	642.50
	Ties	**0**		
	Total	**40**		
P - R	Negative Ranks	**9**	8.89	80.00
	Positive Ranks	**28**	22.25	623.00
	Ties	**3**		
	Total	**40**		
P - M	Negative Ranks	**14**	21.44	262.50
	Positive Ranks	**26**	18.75	557.50
	Ties	**0**		
	Total	**40**		

6.5 Discussion and Post-Analysis

Results correspond to the intuition that generating a user model, whether based on implicitly or explicitly defined user preferences, can play an important role in cross-domain recommendations - even using the relatively straight-forward VSM techniques that formed the basis of our initial experiments. These improvements illustrate the ability of a model that is using just 20,000 characters and drawn from a completely different domain to obtain positive results. This section qualitatively investigates the reasons for this apparent effectiveness in further detail - why is it producing these results?

With the PASSIVE model having produced the highest evaluation scores, this study first examined how its own particular model would have considered articles that were recommended by its rivals (in order to shed light on its selections). On average the cosine similarity between the Twitter extracted VSM and PASSIVE recommendations was 0.09642, for MANUAL recommendations, it was 0.00881, and for RANDOM it was 0.00155 (an example of these differences is illustrated in Figure 26). Therefore, as far as the Twitter generated VSM model was concerned, MANUAL and RANDOM models both performed similar, poor recommendations at an order of magnitude worse than its own recommendations.

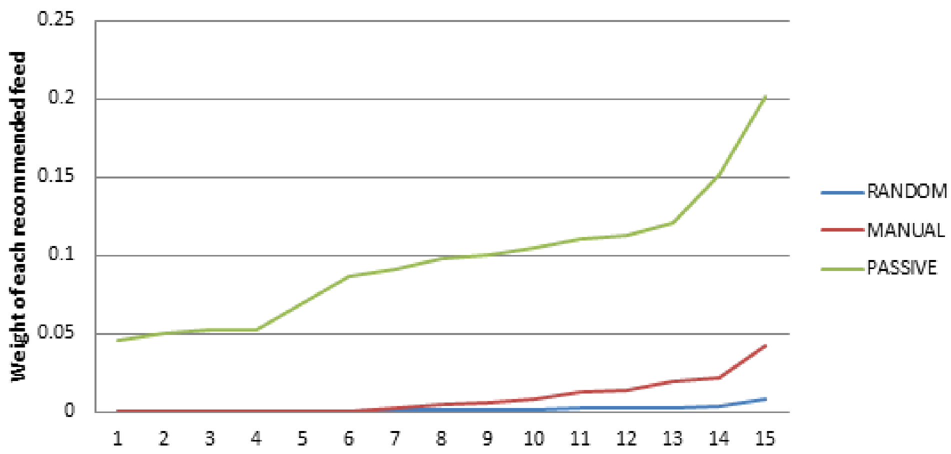


Figure 26: Interpretation of recommendation relevance for all models, based upon the Twitter extracted VSM for Participant 38 (n.b. articles have been ordered with respect to their cosine value).

Given that MANUAL did, in fact, achieve statistically significant improvements over RANDOM, it can be inferred from this that the preference information that the PASSIVE model is able to capture via Twitter is wholly distinct in nature to that established via an explicit statement of categorical preferences. It appears that the Twitter-based model is indeed capturing the intricate cross-category preferences that one would expect it to achieve. However, this also suggests that improved recommendations could be achieved by some combination explicitly declared and passively mined preferences.

It was noted that there were five cases where participants exhibited an overall preference for recommendations made by the MANUAL model. This could be for several reasons: perhaps the labels offered to them during explicit user modelling better expressed their interests than was the case for other participants. It may have been the converse though, and was due to their tweets not sufficiently expressing the diversity of their interests. Alternatively, it could simply be because their preferences are very focused (to one BBC News feed tag for example).

Closer investigation of the five users who most favoured explicitly generated recommendations, however, showed not only multiple label selections for the MANUAL model but also that their ratings for its recommendations were not in fact statistically different from their ratings for the PASSIVE model (this proximity is illustrated in Figure 26). From this it seems most likely that these are indeed cases of participants being well matched to the ontology of labels supplied (and it is not the case that passive models were completely failing them).

This contrasts, however, with the five users who most favoured passively generated recommendations, where the difference between that model and their MANUAL evaluations was indeed significant (again see Figure 27). I infer from this that even though, in general, Twitter modelling is giving us different informational value, the information it is producing is still sufficient to overcome anything lost by not incorporating explicitly stated preferences. Again a combination of approaches seems likely to enhance user response.

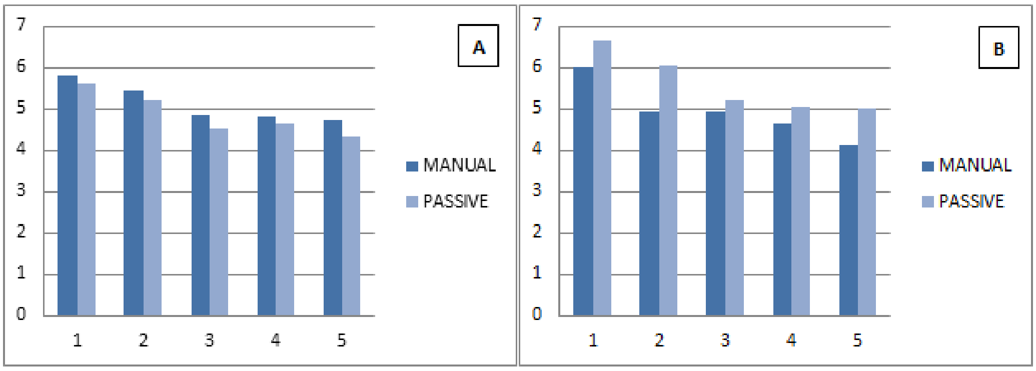


Figure 27: Mean evaluation scores for the [A] Top 5 users favouring MANUAL models and [B] Top 5 users favouring PASSIVE models (along with the closest rival model)

Investigation of which articles were being missed by some models compared to others indicated that in particular, the Twitter model was able to make extremely niche recommendations that cut across broad categorisations and tags. For example, participant four declared an interest in Science & Environment, Politics and Technology tags, which evaluated at 4.1 for MANUAL recommendations (compared to 3.1 for RANDOM recommendations). However, a third of his/her PASSIVE model's recommendations did not include these labels - and yet were still awarded an even higher evaluation of 4.5. I infer from this that unlike explicit filtering, which is necessarily bound to some pre-defined ontology of labels and categories, passive filtering was detecting more personalised, specific article recommendations that cut across categories.

For many participants, MANUAL preference selection appears to have been consistently too coarse to reflect the subtleties of individual user preferences. Participant 19 serves as an example of this. He/she was presented with several news articles tagged with the 'UK' label by both PASSIVE and MANUAL models. Yet, those served by the MANUAL model received ratings of only 3.66, compared to 5.50 for those served by the PASSIVE model. This represented a commonly identified theme where a participant was indeed accurately identifying an interest in UK articles but was unable to specify that it was a specific subset of these that held the most interest for them.

Missing articles due to not manually selecting the superset that a label represented was another common theme. An example of this occurred in participant 30, who highly rated two medical articles that were recommended by his/her PASSIVE model (evaluating them both with a score of 6), despite stating no general interest in health tagged items when they had been explicitly asked for their user preferences. Because this pattern appeared to occur so frequently, now some specific examples are presented in closer depth. Participant 22 stated that items tagged as 'World' (i.e. non-UK) events were not of interest to him/her. However, when the PASSIVE model created a VSM via his social media posts, it found several hits with news items with the 'World' tag. Two illustrative examples, which the user evaluated as having high relevance to them, were the news items:

News item 1: "The moment Nepal's earthquake hit my home"

News item 2: "The day my generation will talk about for the rest of our lives"

Investigating the participant's VSM indicated that the similarity was being expressed due to a high TF-IDF score for the features 'Nepal' and 'aid', and this was corroborated by the detection of posts in his Twitter timeline expressing sympathy for the people of that region following the recent natural disasters it faced[20]. A similar situation occurred for a user who indicated a high evaluation for a 'UK' tagged news item (which referenced the UK based soccer team, Chelsea), despite stating that he had no favoured preference for UK specific stories. The title of the article was:

News item 3: "Why Chelsea won the league, by Alan Shearer"

Exploration of the participant's VSM initially drew a blank, showing no indication of a high expression for any terms related to Sport, Football or Soccer. However, further investigation identified a tweet on the participant's timeline that referenced "Mourinho" (the coach of the Chelsea Football team at the time of the experiment). This term did indeed have a high TF-IDF expression in the document's vector-space model as well as a high activation in the participant's PASSIVE model, consequently resulting in the articles recommendation.

It is noted that this level of granularity would be impossible to achieve in any explicitly stated set of preferences. For some participants, however, it was also noted that their Twitter hashtags (which were parsed as n-grams within the PASSIVE modelling process) themselves served as forms of folksonomic tagging and expressions of highly granular categorical interest. An example of this was participant 32 who used the following hashtags with high frequency: #bigdata, #datascience, #statistics, #analytics and #IoT. As a result, the participant was recommended the following article via the PASSIVE model:

News item 4: "Why measure feet with iPads?"

This article discussed how a shoe retailer had introduced tablet devices in order to automate the measuring and capturing of invaluable data about their customers' feet. Despite no apparent relevance to any of their other interests, the user evaluated the article with a score of 6 (it is assumed due to its data collation component). Because the story was tagged with the label 'Health' it was overlooked by the MANUAL model.

These fine-grained investigations drew me to several conclusions concerning the sensitivity and behaviour of the Twitter-based passive model approach, and its divergence from self-declared preferences: 1. Participants did indeed Tweet about things they were interested in reading about, and thus we were able to pick up true positives via the PASSIVE model; 2. The dynamic nature of both domains under consideration (social media posting and topical news articles) meant that collation of data from a constrained time window was appropriate

[20] Tweets are paraphrased in order to prevent re-identification of participants)

(decaying historical Tweets seems a promising next step for improving the PASSIVE model); 3. Highly relevant content-based recommendations that are picked up by the PASSIVE model can be easily missed by the MANUAL model if they are tagged with an over-generalised label; 4. It is probably not possible to provide a universal taxonomy for an explicit statement of favoured topics due to the highly granularised nature of user preferences; 5. While PASSIVE modelling via Twitter produced statistically superior results to MANUAL models, the two approaches appear to be capturing different forms of preference information. From this it can be concluded not only that a hybrid model would produce improved results, but that generating a user model from numerous combined domains (web search logs, Twitter posts, Facebook usage, etc.) would likely produce even more effective functionality, each providing a slightly different window into a user's preferences.

6.6 Conclusion

Via direct user-feedback, this study has investigated the effectiveness of a recommendation system based on personal data stream information that combines the advantages of both cross-system and content-based filtering. A cross-system user model was constructed by mining Twitter data streams, and its performance corroborated via real world user assessments of BBC News recommendations. Firstly this study showed not only the viability of harnessing linguistic vector-space user models generated from social media data (in this case mined from publicly available Twitter streams) but also that this automated cross-domain approach can actually be superior to explicit filtering using self-declared preferences, as demonstrated via the statistically significant results of our user experiments. However, post-analysis indicates that these two approaches were capturing different information and there is fertile ground in combining the two mechanisms.

Therefore, it would seem that not only user "profile" (e.g. interest) information from outside a recommender system might be useful in a recommender system but also that such profiles could be inferred with no explicit user input from Twitter posts.

In the following study I will examine the possibility of opening different windows into generating user's interests and preferences from more personal and private information by concentrating more on social connections and the text generated therein, rather than microblogging.

7. Study 2: Social Connections

7.1 Overview

The rise of social networks and the enormous amount of data that they collect about their users make them a valuable source of personal information about users. In the last chapter it was shown that for recommendation systems, this sort of personal information can be very useful for improving the quality of personalised recommendations.

However, this study extended the cross-domain user modelling approach using social media data (Twitter streams) in the previous study. In this study, the feasibility and effectiveness of utilising available personal information from social networks (specifically from Facebook) for the recommendation process were investigated. This study examined the possibility of opening different windows into generating user's interests and preferences from his/ her Facebook activities. To do so, content published by users on their personal account about their different interests and aspects of their lives on Facebook was extracted.

I studied the integration of Facebook data with the recommendation process and constructed several other models to compare the performance of these models with that of previous models used in the earlier work. I conducted a field study in which different experiments were run. These included an online experiment where recommendations generated using Facebook data were tested and compared for 23 users.

7.2 User-Profile Modelling

The experimental system must be able to generate multiple user models for the same individual. As in the previous study, **PASSIVE user-model**, **MANUAL user-model** and **RANDOM user-model** were used in order to support comparisons as represented by the user-modelling module in Figure 17. Currently, the models available to the system include the three main models described in section 5.2.4 in addition to a new user model named **TOP-Feeds model**.

- **TOP-Feeds modelling:** In order to have a different level of comparisons, the system will also generate a random preference-model from the top feeds nominated by the BBC with a different random one generated for each user. This is a VSM containing the same dimensionality as the passive user model, but with TF-IDF scores set randomly for each feature in F. Similarly, each feed will be assigned with TF-IDF scores indicating its similarity to a user profile. I refer to this model as TOP in our experiments. The relevance of random recommendation results produced from the top stories article will be assessed. This model is denoted as TOP within our experiments.

However, in this study, Facebook is the cross-system source used to construct this passive model. In order to represent documents, the experimental system currently leverages a traditional vector space model (VSM) as in the previous study.

Recommendation occurs when users begin interacting with the system. Participants are presented with a stream of n documents, each in turn and each generated by one of the available user models (which model is selected for each recommendation is specified by the testing regime) as detailed in the methodology Chapter (see 5.2.5).

7.3 Content Sources

- **Facebook API**

Basically, the application programming interface (API) is a set of tools, protocols and routines for building applications. Facebook provides a programmatic access allowing reading and writing Facebook data which was used for this study. API queries were sent over a HTTP connection to receive the latest Facebook posts matching a search query and remain in sync with a user's posts and updates. API calls were never used to write Facebook data to the user's profile, and access was only ever used in order to access their accessible public timeline and to retrieve post content.

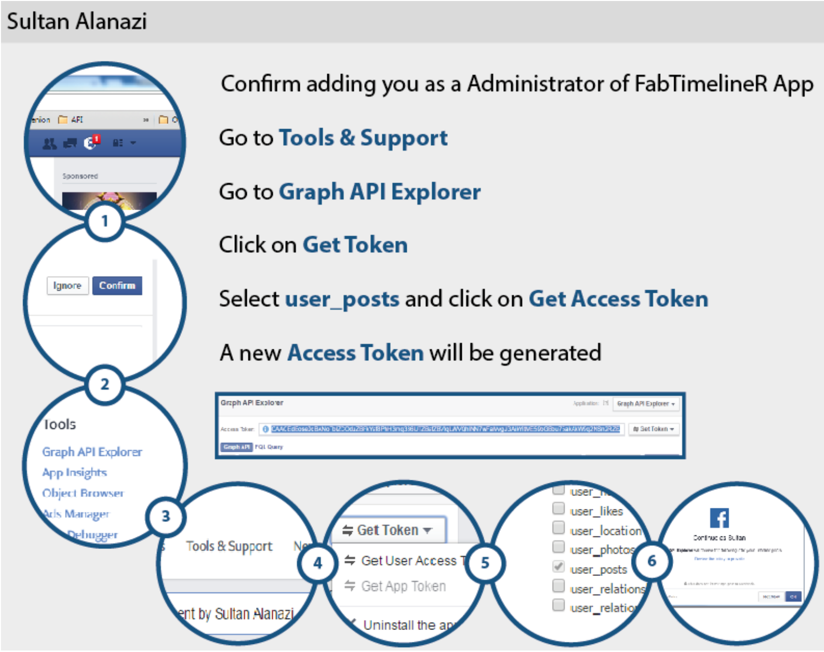


Figure 28: Steps for generating an access token for Graph API explorer

The study also saw the creation of the "Fabtimeliner"[21] app which is a Facebook Graph API[22] app to get data in and out of Facebook's social graph. The Graph API is a low-level HTTP-based API that can be used to do different tasks including querying data, posting new stories, uploading photos and other tasks. So, the Fabtimeliner was developed to retrieve data from Facebook using the Graph API. As illustrated in Figure 28, each participant was added as an administrator to the Fabtimeliner app in order to access user's timeline. By this step, it was able perform the task of accessing user activities on each users wall on Facebook. In this study, the main part needed for constructing user' Facebook-based model was users' Facebook posts. Each user had to generate an access token in order to allow the Fabtimeliner app to access users' timeline and to retrieve post content.

7.4 Experimental Result and Evaluation

7.4.1 Summary

First, an RSS reader was run over the BBC News feeds, allowing the system's content module to extract 3,488 articles. All documents were transformed into a corresponding VSM, and stored along with category meta data including Technology, Science, Environment, Entertainment, Arts, Education, Family, Health, Politics, Business, UK, and World. Data from the BBC News feed was

[21] https://www.facebook.com/Fabtimeliner-425383837630083
[22] https://developers.facebook.com/docs/graph-api

retrieved. In this study, a study consisting of 23 participants was performed. Each participant generates 15 rating scores for each user-profile modelling. This experiment was applied using four models including RANDOM, TOP, MANUAL and PASSIVE modelling (i.e. cross-system model).

Participants were recruited to participate in a lab-experimental study on using social network data (i.e. Facebook) in personalising delivery of BBC news feed. Participants were compensated with a £10 Amazon voucher for their time. Tasks assigned to participants were explained in details before starting the experiment (see information sheet in the Appendix B).

7.4.2 Experimental Method

Twenty three participants in the rating procedure owned Facebook accounts (as of April 2016 when the collection took place). Out of the 23 participants 14 were male and 9 were female. Also, 16 were students: 4 pursuing a bachelors degree, 2 masters, and 10 doctorate, ranging 9 different majors from information sciences, engineering, computer science, human factors, data science, international affairs and other fields. The other 5 participants were employed and working in the fields of the human factors, software engineer, human rights, computer science and technical architecture.

Thus, it was able to access their Facebook posts. It should be noted that on the time the described experiment was conducted users tended to have enough data on their time line. In this experiment the participants had on average 472.11 friends. They had 57.74 check-ins and 609.14 likes on average in different categories such as books, TV programmes, music, restaurants, games and sports teams.

In order to avoid the potential bias that can occur including non-representative sample and responses bias, this study recruited participants with different fields and backgrounds and they holding different degrees. Also, items were recommended to each participant are equally and randomly selected from the available user-models in order to minimize the responses bias.

I used the University of Nottingham Code of Research Conduct and Research Ethics provided by the School of Computer Science research ethics committee at the University of Nottingham to recruit participants. Participants were recruited through a lab-experimental study and were paid a nominal amount for their time.

In this experiment, each participant engaged in a lab-based task experiment where the system presented a total of 60 articles undertaken in four rounds for all user-profile modelling described in Section 7.2. To choose the next article to be shown, the system selected a random user-profile model from the four available user-models and presented an article to the user with a 7-likert-scale as illustrated in Figure 22.

All 23 participants are engaged in this experiment, and each user generated 60 rating scores divided into four models. Therefore, this study performed a controlled user study (N=23) with the objective of testing the following hypotheses:

H1: That the rating scores mean recorded for articles selected using the TOP model would be not higher ratings than those chosen by the RANDOM model used as our baseline.

H2: That articles recommended by the MANUAL model would not record a higher mean than those recommended by the RANDOM model.

H3: That articles recommended by the PASSIVE model would not produce higher ratings than those selected by the RANDOM model.

H4: That the rating mean recorded for articles selected using the MANUAL model would not have higher ratings than those recommended by the TOP model.

H5: That articles recommended by the PASSIVE model would not record higher mean ratings than those recommended by the TOP model.

H6: That articles recommended by the PASSIVE model would not produce higher ratings than those selected by the MANUAL model.

At the end of this experiment, I was able to test these hypotheses against the 1,380 data points produced (23 people × 60 rating scores).

7.4.3 Experimental Result

- **Summary**

23 participants from a wide range of backgrounds were registered with the system. Each participant was required to be an active Facebook user and to have posted a minimum of 150 posts.

Results for our user experiments indicated that the PASSIVE model generated the most relevant news item recommendations for users in comparison to

MANUAL, TOP, and RANDOM models. The mean relevance scores recorded for our baseline model (RANDOM) was 3.568 points and 3.855 for the TOP model. Recommendations generated via manual preferences models (MANUAL) were rated at an average of 4.093 across 345 evaluations. The mean rating of the implicit/linguistic filtering model (PASSIVE) was 4.5.

While only a limited set of participants, results were extremely positive. In 91% of cases, a user-model improved over the baseline random recommendation. Also, a user-model recorded a better score over the TOP and MANUAL models within 86% and 65% of the cases respectively. Standard deviation of evaluations was relatively low for all models at 0.789, 0.624, 0.834 and 0.787 for RANDOM, TOP, MANUAL and PASSIVE models respectively.

○ Friedman Test

In order to compare the performance of the models, the Friedman test was run [196] on the null hypothesis, that all models provide the same results. The null-hypothesis, against which all four methods for which user response was measured, was rejected using the Friedman test with (3) = 30.066, P < 0.000. There was a statistically significant difference in perceived articles depending on which type of filtering was used to filter the resources served to a user, $\boldsymbol{X}^2(\mathbf{3}) = \mathbf{30.066}, \mathbf{p} < \mathbf{0.000}$.

Table 4: Friedman Test Statistics

N	23
Chi-Square	30.066
df	3
Asymp. Sig.	.000

The table above provides the test statistic (χ2) value ("Chi-square"), degrees of freedom ("df") and the significance level ("Asymp. Sig."). Therefore, it can be stated that there is an overall statistically significant difference between the mean ranks of the related models. This test can identify whether there are overall differences but does not show which models, in particular, differ from each other.

○ Wilcoxon Signed Ranks Test

Once the null hypothesis was rejected using the Friedman test, the post-hoc test was applied in order to compare the models with each other. I was interested to see if there were statistically significant differences between the tested models. Therefore, a Wilcoxon Signed Ranks Test was run on 23 Facebook user participants to determine whether there was a statistically significant mean difference between the models. This study measured user responses using 7-likert-scale, so Wilcoxon tests are preferred to a Bonferroni corrected paired t-tests due to the ordinal nature of the scale.

Participants logged a mean evaluation of 3.568 for RANDOM and 3.855 for TOP. Test results produced a value of $p = 0.03$, indicating that the difference in means of 0.286 represented a statistically significant increase ($Z = -2.175$, $p = 0.03$, $< .05$). Thus I was able to reject hypothesis H1, and conclude that the TOP model was producing a superior performance. Similarly, for hypothesis H2, participants expressed preferences for MANUAL models (4.093) as opposed to the RANDOM baseline (3.568); there is a statistically significant increase ($Z = -3.378$, $p = 0.001$) which indicates again that there is a statistically significant difference between our two variable scores. Also, for hypothesis H3, participants expressed preferences for PASSIVE models (4.501) as opposed to the RANDOM baseline (3.568); there is a statistically significant increase ($Z = -4.016$, $p = 0.00006$, < 0.05) indicating again that there is a statistically significant difference between our two variable scores.

For hypothesis H4, participants expressed preferences for MANUAL models (4.093) as opposed to the TOP model (3.855); there is a statistically significant increase ($Z = -2.587$, $p = 0.01$) indicating again that there is a statistically significant difference between our two variable scores. Also, for hypothesis H5, participants expressed preferences for PASSIVE models (4.501) as opposed to the TOP model (3.855); there is a statistically significant increase ($Z = -3.316$, $p = 0.001$, $< .05$) indicating again that there is a statistically significant difference between our two variable scores.

Finally, I was able to show a statistically significant preference for the use of PASSIVE mined personal information (4.501) as opposed to the MANUAL filtering (4.093); a statistically significant increase of 0.408 ($Z = -1.979$, $p = 0.048$), $p < .05$. This indicates that the use of passively mined personal information generates further statistically significant improvements than RANDOM, TOP and MANUAL models. Full details are provided in Table 6 and 7.

Table 5: Wilcoxon Signed Ranks Test Statistics

	TOP - R	M - R	P - R	M - TOP	P - TOP	M - P
Z	-2.175	-3.378	-4.016	-2.587	-3.316	-1.979
Asymp. Sig. (2-tailed)	.030	.001	.000	.010	.001	.048

Model R = RANDOM, TOP = TOP_FEEDS M = MANUAL, P = PASSIVE

Table 6: Ranking Results

		N	Mean Rank	Sum of Ranks
TOP - RANDOM	Negative Ranks	6	8.83	53.00
	Positive Ranks	15	11.87	178.00
	Ties	2		
	Total	23		
MANUAL - RANDOM	Negative Ranks	3	7.50	22.50
	Positive Ranks	19	12.13	230.50
	Ties	1		
	Total	23		
PASSIVE - RANDOM	Negative Ranks	3	2.00	6.00
	Positive Ranks	20	13.50	270.00
	Ties	0		
	Total	23		
MANUAL - TOP	Negative Ranks	4	13.38	53.50
	Positive Ranks	19	11.71	222.50
	Ties	0l		
	Total	23		
PASSIVE - TOP	Negative Ranks	3	9.83	29.50
	Positive Ranks	20	12.33	246.50
	Ties	0		
	Total	23		
PASSIVE - MANUAL	Negative Ranks	8	9.06	72.50
	Positive Ranks	15	13.57	203.50
	Ties	0		
	Total	**23**		

7.4.4 Discussion and Post-Analysis

For a general overview, the mean average scores for random, top, manual and passive filtering for each participant in turn is illustrarted in Figure 29. The majority of the articles selected by the PASSIVE model were favoured the most by participants, however, this was not the case across the board. It is also noted that in some cases the MANUAL model performed better than the PASSIVE model.

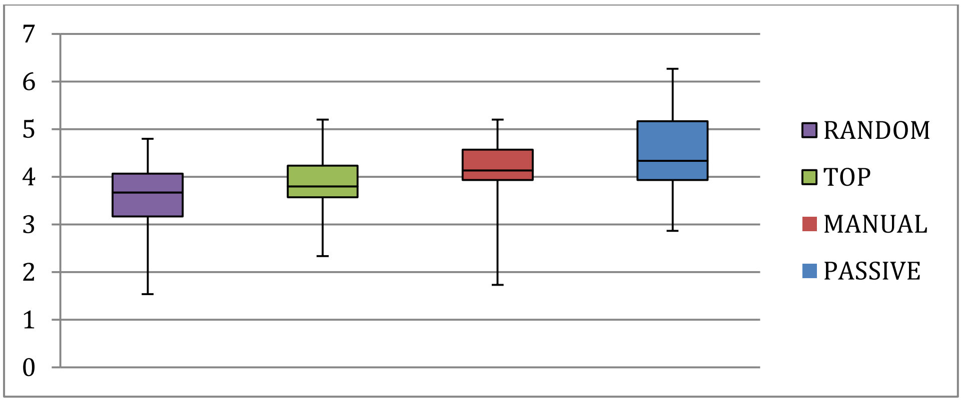


Figure 29: Distribution of mean evaluation scores across PASSIVE, EXPLICIT, TOP and RANDOM models

It can be seen that in Figure 29 more than half of the articles served by the PASSIVE model scored over four on the Likert-scale. There were 187 out of 345 resources recorded as interesting articles while only 224 articles were not matched against user's interests and considered as not interesting articles in the RANDOM model and nearly 200 articles in the top model. Figure 30 shows the articles recorded over four on the scale in each model. However, there were seven users who scored more than 10 out of 15 recommendations, served by the PASSIVE model. These were rated as interesting articles which were personalised according to the users' posts retrieved from the users' social media source. This gives a clear image about how much the PASSIVE model can help the user be served with the resources that match his/her interests.

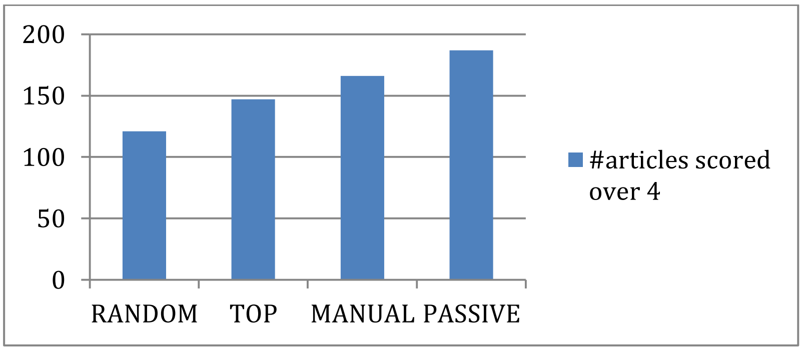


Figure 30: The articles scored with an interesting and strongly interesting score

Figure 31 shows the number of articles nominated as interested or strongly interested versus not interested and strongly not interested for the random, top manual and passive models for each user. Only 59 articles served by the

PASSIVE model were nominated as not or strongly not interested. This means that only 17% of the cases, where recommendations were generated via user-model were not interesting for the users while over 54% of the cases were scored as one of the top three scores on the scale out of 7. It can be seen that six participants strongly liked articles generated by the PASSIVE model and they did not record any recommendation as 'not an interesting' article. From the same figure, it can be clearly noted that, for many participants, the TOP preferences selection has more interested recommendations than the RANDOM model. For more details, Figure 32 shows the distribution of the most interesting and not interesting articles served by each model for all users.

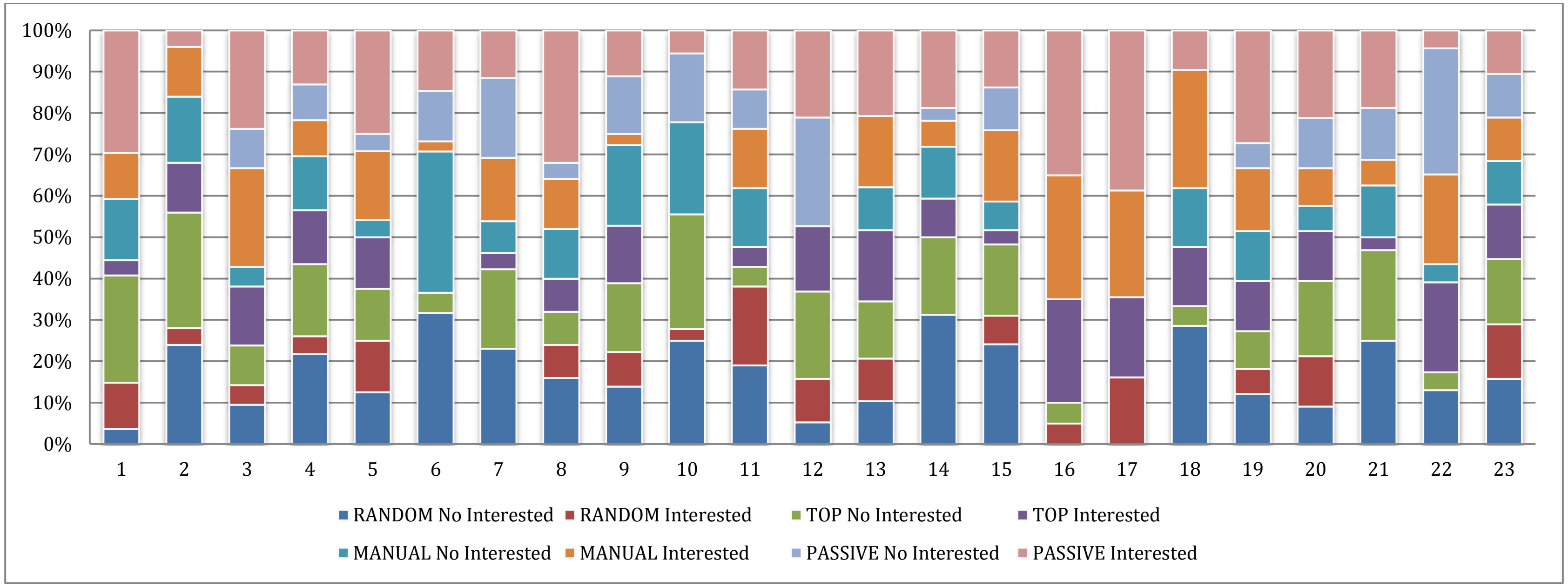


Figure 31: Distribution of the participants' evaluation scores over four on the evaluation scale across PASSIVE, EXPLICIT, TOP and RANDOM models

For example, participant 13, did not nominate any of the articles served by the PASSIVE model as ‘not an interested’ article while at least three articles that were generated randomly, from the entire BBC articles pool, including those tagged as top story articles, were judged with a very low score as not or strongly not interested recommendations. However, recommendations which were generated based on the user’s social media posts recorded an average rating score of 5.53 while the RANDOM and TOP models logged a mean evaluation of less than 4.3. Also, she found that three news articles, personalised by the PASSIVE model, were strongly interesting articles while she also found three other articles as interesting and scored them six out of seven on the scale. In addition, the user-model generated articles that were nominated as interesting and strongly interesting recommendations were twice as those nominated as interesting articles in the RANDOM model.

The lowest average score, assigned by the user to the resources generated by the PASSIVE model, was 5 compared to 3.3 for those filtered based on self-declared user interests. The participant declared her interests in four different categories via the MANUAL model including ’Science and Environment’, ’World’, ‘UK’, ’Entertainment and Arts’ and ’Technology’ with 5.3, 4, 3.3, 4 and 5.3 averages score respectively. However, the news articles tagged with the ’Health’ label, which were overlooked by the user, recorded an average of six and the user found those four ‘Health’ labelled recommendations generated by the Passive model as interested. Also, articles tagged with the ‘Education’ label were overlooked in the MANUAL model and logged a high average score with 5.5, exceeding articles self-declared manually. For more details, Figure 32 compares the most interesting and not interesting articles served by each model and the average score for participant 13.

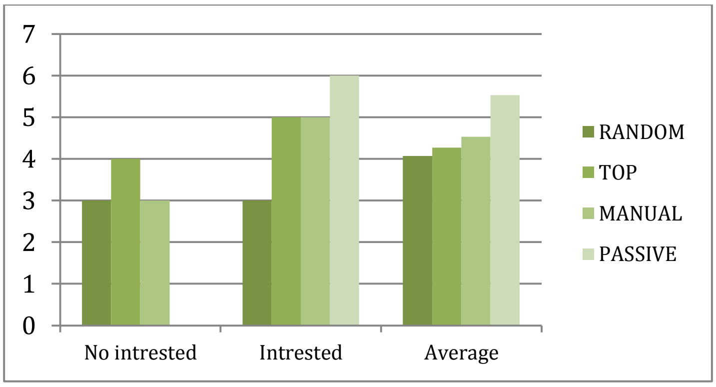


Figure 32: Distribution of the interesting, not interesting articles and the average score rating across PASSIVE, Explicit, TOP and RANDOM models

For many participants, the MANUAL preference selection performs better than the TOP model (82% of the cases) and much better than the RANDOM model (86% of the cases). This means that, explicit predefined preferences can play an important role in improving the quality of recommendations. This would be expected since there is a direct connection between the user and the system. However, it was obviously clear that using implicit preferences in constructing a user model performed significantly in generating more accurate recommendations and would be effective in cross-system recommendations. The user model implicitly generated a higher evaluation score than those produced explicitly in 65% of the cases and 91% of the cases better than the TOP model. Participant 19 is an example of this. She was presented with several news articles tagged with the 'World' label by both the PASSIVE and the MANUAL models. On average the average cosine score for those served by the PASSIVE model was 0.364 compared to 0.135 for the news articles with the 'World' label served by the Manual model. Also, articles have received high average rating of 5 in the PASSIVE model, compared to 4.6 for those served by the MANUAL model.

These examples would draw me to a number of conclusions in terms of applying explicit and implicit user preferences in order to build a user model. First, it would be possible to use a user's posts from Facebook to build a user preferences model. Additionally, users shared on their online social network accounts sufficient information which could be used to help the user in filtering a number of resources in order to find the most personalised ones. This can be useful in saving user time by finding their interesting items and enriching the user's experiences.

7.5 Conclusion

In this study, there was an attempt to achieve an effective and practical content recommendation system based on using personal data stream information. This study makes contribution in supporting similarity-based rankings for more accurate recommendations, content-based recommendation system that achieve better recommendation efficiency using user's private data (i.e. Facebook timeline) and in constructing user profile from a personal social media. For accuracy, the popular similarity measure, using the vector space model with cosine measures, was used to produce the correct recommendations.

In this study, it was explored that the possibility of utilising user preferences, derived from Facebook, for content-based filtering. I examined data mined from a user's data stream that could then integrated as input to the cross-system recommendation process during the similarity and the analysing stages. The derived data from Facebook can also enrich user's data to improve the performance of the recommendation system.

Our results show the potential value of using the available data in a user's Facebook timeline towards improving personalised recommendations that require data about users in order to provide useful results. Our findings reveal that recommendations based on Facebook data are more accurate than those based on explicitly defined user preferences and when there is no data available for generating recommendations. It can be concluded that when cross-system data is available, users would be provided with more relevant items. This result is important as it enables different recommendation systems to generated well match recommendations when a user's data is sparse or non-existent.

Through analysis of real-world document sets from the BBC News feed, the effectiveness and efficiency of our system have been shown by running experiments on user personal data from the user's Facebook timeline. This experiment has shown there is an impact of using personal information in personalising delivery of news articles and revealed a significant statistic in improving recommendation accuracy and enriching the user's experience. Therefore, it can be concluded that using personal information such as user's posts elicits a statistically significant increase.

In the next section, this work will be extended further by placing user in a position of control over a user model generated from personal information such as a user's tweets or posts allowing him/her making any changes to end up with a new preferences profile.

8. Study 3: User in Control

8.1 Overview

This study allows a user to be shown both generated preference profiles from different social media streams with user control permissions. This would allow the user to end up with a new preferences profile according to user updates in the profiles mined from his/her social media data. Also, this study is trying to identify which social media data streams of those used in study one and study two can provide users with more personalised recommendations. This would allow me to model different effective users' profiles which can play a major role in generating accurate recommendations. Therefore, recommendation systems take into account a user's preferences coming from different datasets to generate recommendations resulting in improvement of the personalisation process.

Also, it is taken in account that users should be in charge of controlling their generated profiles modelled from different social media accounts. Therefore, users should have the ability to decide which elements of their information reflects their interests and what they wish to retain in their updated profiles. Also, a user would be able to update his/her profile by deleting data from the generated profile or feeding the profile with other words that are not represented. This could potentially offer a different window in delivering new services and applications that can put individuals in charge of controlling their data. This is unlike the current situation, where governments and companies are sharing our data widely with third parties because they are in charge of controlling our data.

This could enrich the user's experience with more personalised recommendations by adding more data and extra information from outside the generated profiles modelled from the user's social media accounts.

8.2 System Design and Interaction

In this study, all models constructed via **user modeling module** used in both study one and study two were used in terms of using more than one source to construct user profiles as illustrated in Figure 33. In this instance, two personal social media accounts need to be authenticated before the user can engage with this experiment. As illustrated in the following flow chart (see Figure 34), retrieving the process sources needs to be accessed in order to construct user models mined passively from social data streams. User preferences would be modelled based on user activities on his/her social media source. Furthermore, resources that would be served to the target user, based on the constructed models, will also be retrieved as a required process to the following steps. The user's profiles mined from active social media accounts will be used to generate linguistic preference models for each data stream source.

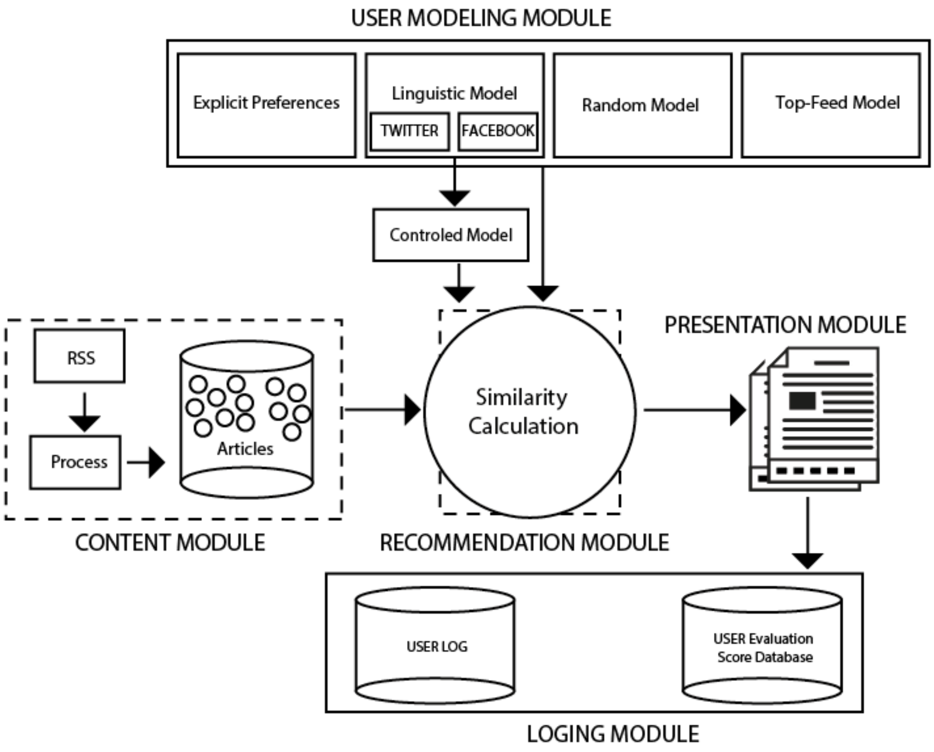


Figure 33: Architecture of experimental platform

The content of the model will be analysed and prepared along with the RSS feeds resources to be ready for the similarity calculation process. This process is responsible for computing similarity calculations between RSS resources and user models. Then, output feeds will be sorted and made ready for making recommendations based on the requested model. These recommendations will be presented via a web interface with the ability to recode a user's responses to the suggested recommendations. As it was mentioned above, the user can update his/her linguistic preference profile in order to generate new recommendations.

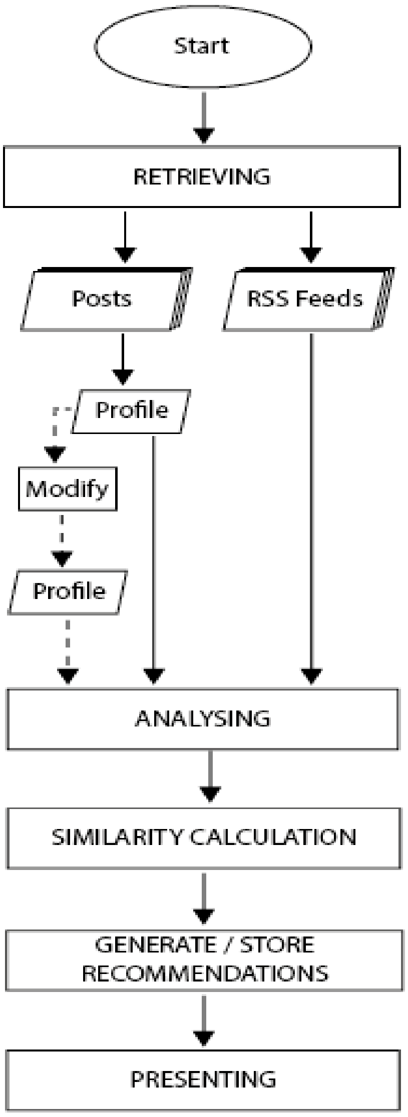


Figure 34: Flow chart of the experiment processes

8.3 Experimental Platform

8.3.1 Research question

This study is trying to investigate the effectiveness of placing the user in a position of control over a cross-system user model mined from different social media streams, with the ability to make any updates, in generating more accurate recommendations.

Specifically, it is trying to answer the following research questions:

- *Can a useful user profile (e.g. interests) be inferred with no explicit user input (social media data streams: Twitter and Facebook)? Which one is better?*

- *Can a cross-system user model from social media (i.e. Facebook posts or Twitter tweets) generate an identifiable profile?*
- *Can a modified user profile perform better in terms of recommendation accuracy? (i.e. would giving the ability for a user to maintain control over such a profile (generated from social media) perform better)*

According to [197][198] as a way of controlling the user's privacy strategy, they delete or edit their content posted in the past. Therefore, I allowed the user to maintain control over such a model; which enabled me to investigate the value of giving a user full control of sculpting, editing and reflecting on that profile.

To answer these questions, I have developed a system that constructs a model of user interests via:

1. A vector-space representation derived from social media streams (i.e. Facebook and Twitter).
2. An explicitly defined declaration of categorical user interests.
3. Random parameterisation (to serve as a baseline for our testing procedure).
4. Random parameterisation of the articles labelled as top stories.
5. An updated profile retrieved from a modified social media stream.

Then, using one of these models to dynamically rate the relevance of articles retrieved from RSS feed source, the most relevant articles will be delivered to a user via a web interface that allowed for relevance feedback ratings.

8.3.2 Content Sources

In contrast to previous chapters, there will be different sources used in this experiment in order to answer our research questions. The Twitter context resource used in this study was similar to that used in study one, where a user's data stream was used to construct a passive model from a user's tweets. This is now combined with the Facebook posts, as used in study two, in order to provide a combined source to build a user passive model.

BBC News feeds were again used as a source of output documents by extracting 8,555 articles across different categories, as indicated in Figure 35. As before, each article was transformed into a corresponding VSM, and stored along with category metadata, and the source article's URL. Each document in this source was tagged with one of the following labels: *Technology, Science, Environment, Entertainment, Arts, Education, Family, Health, Politics, Business, UK, World and*

Top Stories (Table 8 shows the distribution of articles that formed our document set).

Table 7: Distribution of articles in different categories

Category	Total
Top Stories	1401
World	1020
UK	979
Business	804
Politics	705
Health	723
Education & Family	604
Science & Environment	783
Technology	784
Entertainment & Arts	752
	8555

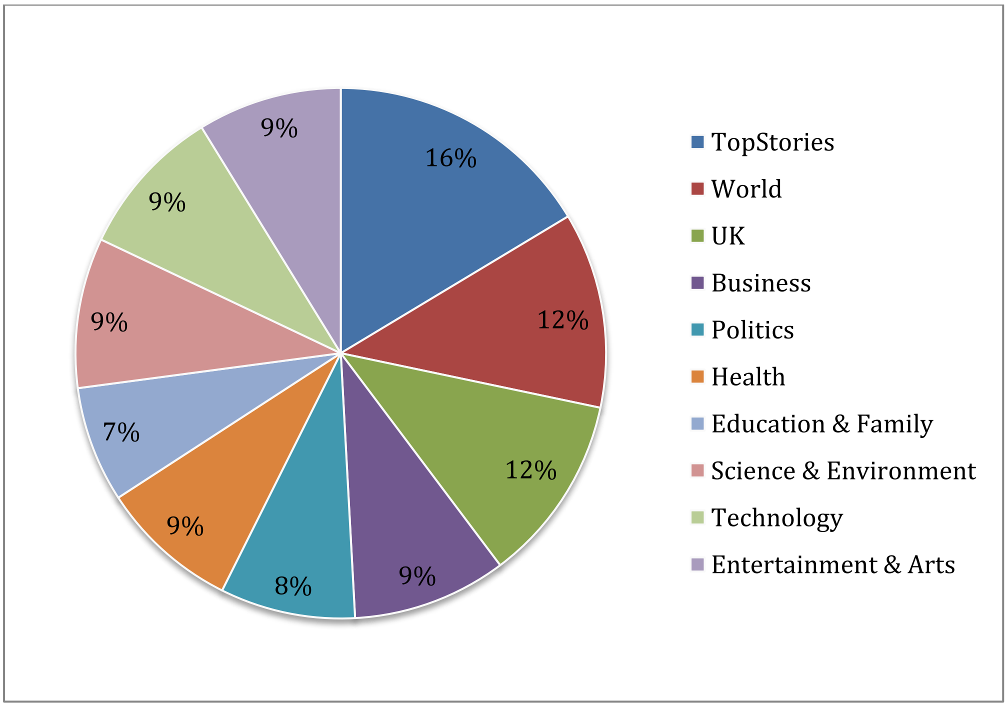


Figure 35: Distribution of the article tags as retrieved

However, *controlled context* is another source of data and will be generated according to user updates in the profiles mined from his/her social media streams. This highly distinct from previous studies in the thesis, allowing the user to 1. be shown both generated preference profiles; but also 2. allowing him/her to make any changes in order to end up with a new preferences profile. This profile will be used to construct a new user-model with the purpose of generating a number of recommendations that meet user interest.

8.3.3 User-Profile Modelling

In order to build this experimental system, I must be able to produce multiple user models for the same individual where these models cover the models in this work's previous studies. Therefore, the main three models - PASSIVE, MANUAL, and RANDOM - are constructed for each user when he/she first registers with the system. Also, the TOP model, which was used in Chapter 7, is considered as one of the models in our experimental system. In this study, the PASSIVE models are generated from different cross-system sources resulting in two separate passive models which are generated from Twitter and Facebook data streams. There will be a new user-preference model constructed, allowing the user to have full control over the social media-based model. The models available to the system are as follows:

- **Twitter-Passive user-modelling:** The user must provide controlled access to an active Twitter social media account to generate a linguistic preference model. Users' Twitter data streams are collated, parsed, cleaned and finally encoded as a vector-space model, again using TF-IDF [190]. This model was denoted as TWITTER_PASSIVE within our experiment.
- **Facebook user-modelling:** The user must provide controlled access to an active Facebook social media account to generate a linguistic preference model. Users' Facebook posts are retrieved and then cleansed in the same manner as with output documents described in Section 7.2. This model was referred as FACEBOOK_PASSIVE within our experiments.
- **MANUAL user-modelling:** This model is described in Section 5.2.4.
- **RANDOM user-modelling:** As detailed in Section 5.2.4.
- **TOP-Feeds modelling:** As detailed in Section 7.2.
- **Controlled user-modelling:** Here a user's preferences are modelled when the user is given the ability to maintain control over a model; allowing him to sculpt, edit and reflect upon that profile. The new modified model will be then cleaned in the same manner as with output documents described in Section 5.2.3. This model was denoted as CONTROLLED.

8.4 Methodology

In this study, user profiles were again initialised through mining a user's social data stream via APIs (albeit multiple ones in this case). These profiles were applied to generate a passive-user model for each social data stream. Therefore, one of the passive-user models would be constructed from a user’s Facebook account and another one from a user’s Twitter account. Each user must have an active account on both social networks in order to join this experiment.

Twenty participants from a wide range of backgrounds were registered with the system. Each participant was required to be an active Facebook and Twitter user and to have posted a minimum of 150 posts.

On registration, each user authenticated with their Facebook and Twitter accounts to allow Facebook-Passive Twitter-Passive models to be automatically constructed for each user. Then each user explicitly declared their categorical interests via the web interface (as selected from the tags used by the BBC News feed). This also allowed for MANUAL, TOP, and RANDOM Models to be constructed for each user. In order to construct both the passive models, the Facebook [199] [23] and Twitter APIs were interrogated with the user’s recent history of available posts and tweets (statuses) being extracted, cleansed and transformed into a VSM model.

This experiment is completely distinct from studies 1 and 2 in that it was conducted in three sessions with sharply differing objectives:

- **First Session:**

At the beginning of this session, a user is asked to identify (via a web form) the BBC RSS channels that interest them (e.g. I would select the '*Technology*' and '*Science and Environment*' channels) to identify the most categories meeting his/her interests. Once this is completed, the users are given 75 randomly selected news feeds.

This 75 news feed articles are divided into:

- 15 articles from the *subset* of the articles pool that is formed when it is filtered using these nominated channels.
- 15 random news articles from the articles pool.
- 15 news articles from the articles tagged with a ’Top-Story’ label.
- 15 'personalised' articles from the pool. These articles will have the highest cosine similarity and the best match to the user's Twitter data stream.

[23] https://developers.facebook.com/blog/post/2016/10/05/graph-api-2.8/

- 15 'personalised' articles from the pool. These articles will have the highest cosine similarity and the best match to the user's Facebook data stream.

Users log into a detected web site and engage in an on-line session during which they will be asked to score the news feed articles that have been randomly drawn from the articles pool using a 7-point Likert scale. The score they award will reflect their interest in the article and this will be displayed in a web page, plus associated images.

- **Second Session:**

After having rated all recommendations, users were asked to do the following tasks:

- Join a card sorting task where the participants will be shown a mix of profiles generated from both social media networks. In this task, the participants will be asked to identify their profile from the list of profiles by ranking them from the 'least' to 'most likely' match to their profile (provided in the Appendix C). This will indicate whether the retrieved posts can generate an identifiable profile for each user and give a general idea of the model accuracy in generating an identical user profile;
- Complete a questionnaire (detailed in the Appendix C) related to the user preferences model used in generating user profiles in order to provide feedback on their satisfaction;
- Modify profiles generated from social data streams. The user's profile will be presented in a way that shows their list of terms in the profile with a score indicating the importance of each word in the list see Figure 36. Then, the user will be able to edit/remove words that do not reflect his/her interests. As a result, the modified profile will be updated based on the new changes and then adopted for the next session.


```
Top words in document 1
good, 12.0
night, 11.0
league, 10.0
great, 10.0
team, 8.0
today, 8.0
game, 7.0
lads, 6.0
england, 6.0
tomorrow, 6.0
time, 6.0
champions, 6.0
football, 6.0
uefa, 6.0
```

WORD CLOUD TERM FREQUENCIES

Figure 36: An example of word cloud and most frequent words in one of the participants

- **Third Session:**

Once the second session is completed, there will be a modified profile generated from the previous session. Next, the best articles matching the modified profile will be computed and served to the user. In this session, the user will be shown 30 randomly selected news articles from the pool.

This 30 news feed articles are divided into:

- 15 random news feeds from the articles pool.
- 15 'personalised' articles from the pool. These articles will have the best match to the user's modified profile.

Similarly to session one, participants will be asked to score each article using a 7-point Likert scale. At the end of this experiment, it will be able to generate 90 data points produced for each participant (75 ratings from session one and 15 from the third session, as I will only consider the personalised articles).

As a result, all participants are engaged in a lab-based task experiment where the system presents them with a total of 90 articles, whose relevance to them they were asked to evaluate in sequence. To choose the next article to be presented, the system selected a random user-model from the three available, determined that model's highest ranked unseen document, and presented it to the user along with a Likert evaluation scale. When ties occur between a set of documents in any ranking, one was selected at random. Each user was ultimately presented with an equal number of articles for each model. Thus, while ranking is deterministic for all models used, each experimental run would still be stochastic in nature in terms of the ordering of articles presented.

For each document presented, the system recorded the participants user-id, the presented article's documented of the article presented, the user's evaluation score, and an identifier of the model used to select that article (either TWITTER_PASSIVE, FACEBOOK_PASSIVE, MANUAL, TOP or RANDOM). Next, each user is allowed to evaluate his /her profiles used in PASSIVE models and finish with a new modified profile.

This data collection process allowed me to test the following hypotheses:

H1: That the mean rating recorded for articles selected using the MANUAL model would not produce higher ratings than those selected by the TOP or RANDOM models used as our baseline (i.e. a hypothesis aimed at determining that any sort of filtering is better than simply serving random BBC News articles).

H2: Those articles recommended by the TWITTER_PASSIVE or FACEBOOK_PASSIVE model would not record higher mean ratings than those recommended by the RANDOM or TOP model.

H3: That articles recommended by any PASSIVE model would not produce higher ratings than those selected by the MANUAL model (i.e. a hypothesis aimed at determining if a managed, implicit filtering approach is superior to a filter based on explicit statements of preference).

H4: That the mean rating recorded for articles selected using the TWITTER_PASSIVE model would be the same as that of the FACEBOOK_PASSIVE models.

H5: That articles recommended by the CONTROLLED model would not produce higher ratings than all other models (i.e. a hypothesis aimed at determining that if a user was put in control of his/her profile it would be better than any other sort of filtering whether accomplished explicitly or implicitly by the user and even better than simply serving random BBC News articles).

Based on th questionare, the following haypotheses will be tested:

H6: That both profiles generated from Twitter and Facebook user's data streams provide the same scores in terms of generating more identifiable profiles.

H7: That the scores recorded for profiles generated from Twitter would be the same as that generated from Facebook in reflecting users' interests.

H8: That the scores rcorded for Twitter in generating more relevant recommendations would be the same as that for Facebook.

Evaluation of these results will be described in the following sections, first via a user study and then via statistical testing.

8.5 Evaluation via a user study:

8.5.1 Set-up

A controlled user study (N=20) was performed with the objective of answering the research questions posed in our earlier discussion - to assess the effects of mining social media data streams to construct passive user models on user experience and understanding of the recommendation process.

A quantitative analysis was also performed on the performance of the six recommendation models. I ran the Friedman test on the null hypothesis, where all models provide the same results in order to compare the performance of the models. Then, the post-hoc test was applied in order to compare the models with each other. I was interested to see if there were statistically significant differences between the tested models by running a Wilcoxon Signed Ranks Test.

Finally, a qualitative analysis was performed based on a questionnaire completed by each participant. I asked questions on how useful mining was on their social activities in generating profiles and how users experienced different aspects of the experiment. In addition, each user was asked to evaluate his/her profile in terms of reflecting their interests and score them in order to see which of the social data streams better represented a user's preferences. Each user was able to update his profile at the end of this questionnaire resulting in generating a new controlled profile which was used to generate a new list of recommendations.

As an addendum, I also performed a quantitative analysis on the performance of the CONTROLLED model. The Wilcoxon Signed Ranks Test was used to compare the performance of the CONTROLLED model with other models. Each of the six models produced a ranked list of recommendations.

8.5.2 Participants

A total of 20 users participated in the user study - 12 were male, and 8 were female. Most participants were students and spanned different majors from computer science, engineering, human factors, physical, analytical chemistry and other fields.

The participants recruited in this experiment were active on Twitter and joined Twitter between 2009 and 2013. The earliest participant was signed up to the service in February 2009. On average, since the creation of their Twitter accounts, participants had recorded 6,838.4 tweets per user. The maximum number of tweets of any user since signing up to Twitter was over 50,000 tweets.

On average, participants had 357 friends on Facebook while two participants had hidden their number of friends. They had different interests which spanned several categories such as films, music, restaurants, and games. They had 630 likes on average in different categories.

It is important to consider that this experiment was designed to be as unbiased as possible by avoiding the potential bias that can occur including non-representative sample and responses bias. This study recruited participants spanning different fields and backgrounds. Also, the number items were recommended to each participant are equally and randomly selected from the available user-models in order to minimize the responses bias.

Our analyses showed that participants were keeping a different group of friends on Facebook and on Twitter; even a different number of friends. According to [200], who studied how users manage and maintain friendships across multiple social networks, users prefer to maintain different friendships in different online social networks while keeping only a small clique of common friends across these online social networks.

Participants were familiar with recommendation systems such as Amazon and Netflix. They were recruited through a lab-experimental study and were compensated with a £10 Amazon voucher for their time.

8.5.3 Procedure

The study had three parts: a pre lab-experiment, a post-study questionnaire and a post lab-experiment. Each participant was required to have a Facebook and Twitter account with at least 150 posts. Each participant was asked to approve his/her account in order to gather their profile data.

Participants were then asked to engage in the task of scoring each recommendation using a 7-point Likert scale of the different models, which were presented in a random order.

Next, users were asked to join a post-study questionnaire including a profile sorting task. The purpose of this task was for participants to sort profiles based on the possibility of having their profiles generated from their posts. By the end of this task, I considered the order of each of the sorted profiles out of 20 as this was the number of profiles issued to each user (see Appendix C for details).

Each user then completed the rest of the questionnaire relating to the quality of recommendations produced by each profile on a 5 star scale, 1 being the lowest and 5 being the highest (as included in the Appendix C). The user was shown profiles generated by each social media accounts, scored them and updated the content by eliminating and adding content. Finally, a new list of recommendations was provided to the user based on his/her changes on the profile.

8.6 Experimental results

8.6.1 Friedman Test

All six models were tested to generate recommendations in order to see if there is a statistically significant difference in perceived relevance evaluation scores, based on which type of model was used to filter the resources. As input for the test, I used the results from six methods using RANDOM as the baseline, TOP, MANUAL, FACEBOOK_PASSIVE, TWITTER_PASSIVE and CONTROLLED models.

To achieve this I performed a non-parametric Friedman test [196] to detect if there are differences between the tested models for the different users' responses in order to reject the null-hypothesis that all six models provided the same result in terms of mean evaluation scores. The null hypothesis was rejected, with $\mathrm{X}^2(4) = 39.929$, $p < 0.05$ for the first five models excluding the CONTROLED model and with $\mathrm{X}^2(5) = 58.628$, $p < 0.05$ for all six models. This means that there was a statistically significant difference in recorded evaluation scores depending on which type of model used to generate recommendations (for more details see Table 9).

Table 8: Friedman Statistics Test

N	20	20
Chi-Square	39.929	58.628
df	4	5
Asymp. Sig.	.000	.000

It can be concluded that the null hypothesis (that there is no difference between models) is rejected for a mean evaluation score and there is an overall statistically significant difference between the mean evaluation scores of the tested models. However, it does not show which models, in particular, differ from each other.

8.6.2 Twitter vs Facebook

In this section, I was interested in comparing the personal models generated from different user personal data streams to find which of these streams would perform better in serving the user with more accurate recommendations that meet a user's interests. In addition, by applying this experiment, I would be able to calculate whether the source of personal data used in construing a user-passive model would make any difference in terms of helping the user find the most relevant items. I would be able to identify which social data stream would record a better performance than that generating a user model based on explicitly defined user preferences.

In order to test our hypotheses, 20 participants joined a lab-experiment who were were served with 75 news feeds which generated 1,500 data points. Two different tests were applied in order to compare the performance of each model. A quantitative analysis was performed by running the Friedman Test in order to evaluate whether there is a significant difference between all the models used in this session.

This was followed by a post-hoc test to compare the models with each other. I also performed a qualitative analysis via a questionnaire completed by each user to see which one of the social media data streams shows a good representation of user interests. I showed the user 20 profiles have been generated from users' social media streams and asked the user to sort them to give a final score for his/her profile generated from the user's Facebook and Twitter accounts. This helped me to run a Wilcoxon Signed Ranks Test to see if there is a statistically significant difference.

Based on our result, both TWITTER_PASSIVE and FACEBOOK_PASSIVE recorded the highest mean relevance scores for users in comparison to MANUAL, TOP, and RANDOM models. The mean relevance scores recorded for our baseline model (RANDOM) was 3.61 points and 3.87 for the TOP model. Recommendations generated via automated preference models (MANUAL) were rated at an average of 4.11 across 300 evaluations. The mean rating of the implicit/linguistic filtering model constructed from a user's Facebook posts (FACEBOOK_PASSIVE) was 4.62. However, the passive model constructed from a user's Twitter timeline (TWITTER_PASSIVE) generated the most relevant new articles with an average

4.73 mean evaluation score. Figure 37 shows the mean relevance scores for RANDOM, TOP, MANUAL, FACEBOOK_PASSIVE and TWITTER_PASSIVE models.

It was noted that the user-passive models improved over the baseline RANDOM and TOP recommendations in most cases. For example, in over 90% of cases, the TWITTER_PASSIVE model performed better than the RANDOM and TOP models. Also, the FACEBOOK_PASSIVE model improved over the baseline recommendation in 90% of cases. In addition, the TWITTER_PASSSIVE and FACEBOOK_PASSIVE were better than the MANUAL model respectively in 75% and in 85% of cases. Another important point is that the mean average of a user's response to the recommendations generated, based on the user's Twitter data stream, improved over those generated based on the user's Facebook data in 65% of cases. Standard deviation of evaluations was relatively low for all models at .81140, .60752, 0.88, 0.74 and .87 for the RANDOM, TOP, MANUAL, FACEBOOK_PASSIVE and TWITTER_PASSIVE models respectively, see Table 10. In order to test our hypotheses, and to determine whether there was a statistically significant difference between models, the Wilcoxon Signed Ranks Test was run.

Table 9: Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
RANDOM	20	3.6100	.81140	1.53	4.80
TOP	20	3.8733	.60752	2.33	5.20
MANUAL	20	4.1100	.88106	1.73	5.20
FACEBOOK_P	20	4.6233	.73581	3.33	6.27
TWITTER_P	20	4.7367	.86693	3.47	6.20
CONTROLED	20	5.2100	.84252	2.80	6.33

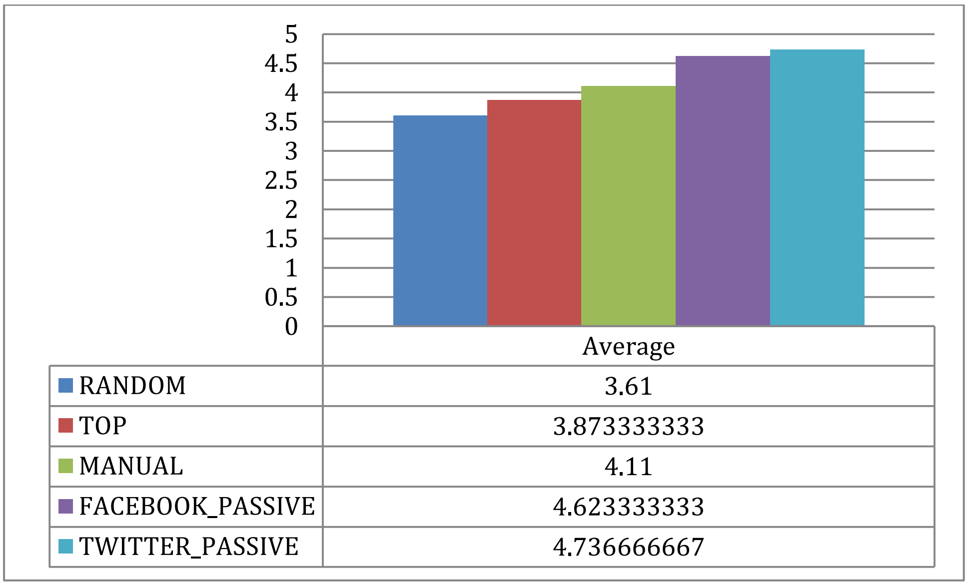


Figure 37: Distribution for each models based on users' average rating score

Wilcoxon Signed Ranks Test

Based on quantity analysis - Summary

Based on the null hypothesis using the Freidman test, the post-hoc test in order was applied to compare the models with each other. I am interested to see if there were statistically significant differences between the tested models. At this stage I am trying to test the first four hypotheses to detect if there are differences between the RANDOM, TOP, MANUAL, FACEBOOK_PASSIVE and TWITTER_PASSIVE models for the different users' responses.

Therefore, a Wilcoxon Signed Ranks Test [195] was run on the 20 Twitter and Facebook user participants to determine whether there was a statistically significant mean difference between the models by measuring user responses using the 7 Likert Scale.

Participants logged a mean evaluation of 3.61 for RANDOM and 4.11 for MANUAL models. Test results produced a value of $p = 0.002$, indicating that the difference in means of 0.5 represented a statistically significant increase ($Z = -3.082$, $p = 0.002$, $< .05$). Also, participants logged a mean evaluation of 3.87 for the TOP model indicating a difference in means of 0.23 from the MANUAL model, showing a statistically significant increase ($Z = -2.391$, $p = 0.017$, $< .05$). Thus, for hypothesis H1, a Wilcoxon signed-rank test revealed that generating a user model based on explicitly defined user preferences did elicit a statistically significant change in logged mean evaluation scores.

Similarly, for hypothesis H2, participants expressed preferences for FACEBOOK_PASSIVE models (4.62) and 4.74 for TWITTER_PASSIVE as opposed to the RANDOM baseline (3.61) and 3.87 for the TOP model. There is a statistically significant increase ($Z = -3.884$, $p = 0.0001$) which indicates again that there is a statistically significant difference between FACEBOOK_PASSIVE and RANDOM variable scores. Test results produced a value of $p=0.0002$ indicating that there was a statistically significant difference between FACEBOOK_PASSIVE and TOP variables ($Z = -3.697$, $p = 0.0002$). Also, results showed a statistically significant preference for the use of passively mined user's personal information via the TWITTER_PASSIVE model with a difference in means of (1.12) as opposed to the RANDOM filtering ($Z = -3.678$, $p = 0.0002$) and with the difference in means of (0.86) as opposed to the TOP filtering ($Z = -3.660$, $p = 0.0002$). As a result, it can be concluded that the PASSIVE models were producing superior performance over the RANDOM and TOP models.

For hypothesis H3, our results revealed that a participant's responses to recommendations, generated passively by mining a user's social media data, recorded a higher mean score than those generated explicitly. A Wilcoxon Signed Ranks Test showed a statistically significant difference between FACEBOOK_PASSIVE and MANUAL variable scores with ($Z = -2.318$, $p = 0.02$, < 0.05). In addition, it showed a statistically significant increase difference between TWITTER_PASSIVE and MANUAL variable scores with ($Z = -2.390$, $p = 0.17$, < 0.05). This indicates that the use of passively mined personal information generates further statistically significant improvements than the MANUAL model.

Finally, I was not able to show a statistically significant preference for the use of TWITTER_PASSIVE mined personal information from a user's Twitter account (4.73) as opposed to the FACEBOOK_PASSIVE filtering mined personal information from a user's Facebook account (4.62). Our test showed that there was no statistically significant change between mining personal information from Twitter and mining from FACEBOOK ($Z = -.503$, $p = 0.615$, $p > .05$). Full details are provided in Tables 11 and 12.

Table 10: Wilcoxon Signed Ranks Test Statistics

	TOP - R	M - R	F_P - R	T_P - R	M - TOP	F_P - TOP	T_P - TOP	F_P - M	T_P - M	T_P - F_P
Z	-2.049[b]	-3.082[b]	-3.884[b]	-3.678[b]	-2.391[b]	-3.697[b]	-3.660[b]	-2.318[b]	-2.390[b]	-.725[b]
Asymp. Sig. (2-tailed)	.041	.002	.000	.000	.017	.000	.000	.020	.017	.469

Model R = RANDOM, TOP = TOP_FEEDS M = MANUAL, F_P = FACEBOOK_PASSIVE, T_P = PASSIVE_TWITTER

Table 11: Ranking Results

		N	Mean Rank	Sum of Ranks
TOP - RANDOM	Negative Ranks	5[a]	7.70	38.50
	Positive Ranks	13[b]	10.19	132.50
	Ties	2[c]		
	Total	20		
MANUAL - RANDOM	Negative Ranks	3[d]	7.50	22.50
	Positive Ranks	17[e]	11.03	187.50
	Ties	0[f]		
	Total	20		
FACEBOOK_P - RANDOM	Negative Ranks	1[g]	1.00	1.00
	Positive Ranks	19[h]	11.00	209.00
	Ties	0[i]		
	Total	20		
TWITTER_P - RANDOM	Negative Ranks	2[j]	3.25	6.50
	Positive Ranks	18[k]	11.31	203.50
	Ties	0[l]		
	Total	20		
MANUAL - TOP	Negative Ranks	3[p]	13.67	41.00
	Positive Ranks	17[q]	9.94	169.00
	Ties	0[r]		
	Total	20		
FACEBOOK_P - TOP	Negative Ranks	2[s]	3.00	6.00
	Positive Ranks	18[t]	11.33	204.00
	Ties	0[u]		
	Total	20		
TWITTER_P - TOP	Negative Ranks	3[v]	2.33	7.00
	Positive Ranks	17[w]	11.94	203.00
	Ties	0[x]		
	Total	20		
FACEBOOK_P - MANUAL	Negative Ranks	6[ab]	7.17	43.00
	Positive Ranks	14[ac]	11.93	167.00
	Ties	0[ad]		
	Total	20		
TWITTER_P - MANUAL	Negative Ranks	5[ae]	8.20	41.00
	Positive Ranks	15[af]	11.27	169.00
	Ties	0[ag]		
	Total	20		
TWITTER_P - FACEBOOK_P	Negative Ranks	7[ak]	11.00	77.00
	Positive Ranks	12[al]	9.42	113.00
	Ties	1[am]		
	Total	20		

Wilcoxon Signed Ranks Test based on questionnaire

Also a Wilcoxon Signed Ranks Test was run, based on the results of our questionnaire. This focused on the profile sorting task which was assigned to each participant. Each participant had to arrange a list of 20 profiles to identify their profiles, generated from their personal social media data streams (provided in the Appendix C). At the end of this task, I finished with 20 scores representing profiles order based on the sorting task. As a result, each user scored their two profiles which were generated from Twitter and Facebook accounts. I compared the performance of the TWITTER_PASSIVE and FACEBOOK_PASSIVE models in constructing identifiable profiles and which one of the social networks is better in representing user's interests.

On average, Twitter profiles registered a high average score of 18.6 as opposed to 17.2 for Facebook profiles. This illustrates that users recognised their Twitter profiles listed them with a higher score in 68% of cases when compared to their Facebook profiles.

Each user was provided their profiles (i.e. Twitter and Facebook) along with the best matched items to each profile. User was able to evaluate whether each profile reflecting their interests based on a scale of 5 (see the Appendix C). Also, based on the most relevant items presented to the participant, he/she were asked to judge which social data stream generate better recommendations.

Based on the questionnaire score, specific results on the Identifiability of user profiles, the extent to which they reflect user interests, and how far recommendations can be improved off them are broken down below:

Identifiable profile — the null-hypothesis, that both user's data streams, from which profiles generated to join the sorting task, were measured, provide the same result, was rejected using a Friedman test with $F(1, 20) = 5.000$, $P = 0.025$, as detailed in Table 13 . The post-hoc Wilcoxon Signed Ranks Test with $P < 0.05$ was then run to compare both models in which they were sufficient in generating a more identifiable profile. I was able to show a statistical preference increase for the profile generated based on mining a user's Twitter data stream (18.55) as opposed to a generated profile based on mining a user's Facebook posts (17.15); a statistically significant increase of (1.4) (($Z = -2.260$, $p = 0.024$, $p < .05$). Full details are provided in Table 14.

Reflecting user interests – The two generated profiles in TWITTER_PASSIVE and FACEBOOK_PASSIVE models were compared. The Friedman test was used to reject the null-hypothesis, that both models generated the same matching interests score with $F(1, 20) = 5.400$, $P = 0.02$. Using the post-hoc Wilcoxon Signed Ranks Test with $P < 0.05$, it can be conclude (as observed in Table 14) that the TWITTER_PASSIVE model generated a sufficient profile reflecting user' interests and significantly better than the profile generated via FACEBOOK_PASSIVE model with (Z = -2.504, p = 0.012, p < .05).

Generating better recommendations – The null-hypothesis, that both social media profiles used to construct passive models, would provide the same result, was rejected using a Friedman test with F(1, 20) = 9.941, P = 0.002. The post-hoc Wilcoxon Signed Ranks Test with P < 0.05 was then applied to compare the generated profiles (in terms of generating more relevant newsfeeds). Results reveal that there was a statistically significant difference with (Z = -2.802, p = 0.005, p < .05).

Table 12: Friedman Statistics Test

N	20	20	20
Chi-Square	5.000	5.400	9.941
df	1	1	1
Asymp. Sig.	.025	.020	.002

Table 13: Wilcoxon Signed Ranks Test Statistics

	TWITTER - FACEBOOK		
	Identifiable profile	**Reflecting user's interests**	**Generating better recommendations**
Z	-2.260^{b}	-2.504^{b}	-2.802^{b}
Asymp. Sig. (2-tailed)	.024	.012	.005

Table 14: Ranking Results

TWITTER – FACEBOOK		N	Mean Rank	Sum of Ranks
Identifiable profile	Negative Ranks	5^{a}	9.20	46.00
	Positive Ranks	15^{b}	10.93	164.00
	Ties	0^{c}		
	Total	20		
Reflecting user' interests	Negative Ranks	3^{a}	6.00	18.00
	Positive Ranks	12^{b}	8.50	102.00
	Ties	5^{c}		
	Total	20		
Generating better recommendations	Negative Ranks	2^{a}	9.50	19.00
	Positive Ranks	15^{b}	8.93	134.00
	Ties	3^{c}		
	Total	20		

8.6.3 User in control

It seems reasonable that users should be in charge of controlling their data and personal information. This translates to users having the ability to decide which elements of their information they would like to share with the system, and what should be included in their generated profiles.

This section aims to examine how one might let users decide which type of data they would like to share in constructing a new updated profile, according to their interests and preferences. This can then provide control choices that allow individuals to operate those choices within their generated profile. Each individual is unique, and if services are to be personalised to improve his or her experience the fact that many such services collect insufficient data about individuals (which results in poor recommendations) must be accounted for. The aim of ameliorating this sparsity is, of course, to improve the efficiency and accuracy of recommendation systems.

To address this task, at the end of the post-questionnaire, each individual was given two profiles, generated separately from his/her personal social media accounts. Each profile was modelled according to his/her activities on a particular, targeted social data stream. Users were then able to see their profiles represented as a word cloud profile. This means that each user was shown a word cloud profile, modelled from his/her Facebook posts and another word cloud constructed from his/her Twitter data stream. Each user at this stage would be able to examine his/her profiles and have a general view of his/her activities on a particular social media platform.

Also, a list of the top words mentioned in a user's profile was given to each user. This list consisted of the top words in a user's profile with their weights representing the word weight in the profile. In order to have a more accurate profile matching user's interests, each user was put in charge of controlling his/her profile. Items could be eliminated from a user's profile, based on user preferences, if he/she finds they do not represent what they like or what they wish to use to build a more personalised profile. In addition, the user was able to update his/her profile with new items which they would like to receive.

By finishing this task, I would conclude with 20 controlled profiles reflecting a user's interests as a result of allowing users to update their own profiles. In order to evaluate the effectiveness of allowing a user to play a role in updating their own profile, I used theses profiles in constructing yet another new user-model to compare it with other models. In order to support comparisons, the new user-model, named the CONTROLLED preference user-model, was applied to generate the new recommendation results. The CONTROLLED model then, as ever, attempted to generate the most relevant articles to the user, based on his/her updated profile, presented via a web interface that allowed for relevance feedback ratings.

This model produced 15 news feed articles and received 15 evaluation scores by each individual. At the end of this task, 20 participants had therefore generated 300 relevance rating scores. Based on these scores, I was able to compare the 300 users' responses produced from the CONTROLLED user-model with the 1,500 responses recorded from the RANDOM, TOP, MANUAL, FACEBOOK_PASSIVE and TWITTER_PASSIVE models. This allowed me to perform a Wilcoxon Signed Ranks Test to see if there was a statistically significant difference between the models.

Wilcoxon Signed Ranks Test

Based on the null hypothesis using the Freidman test, the post-hoc test was applied in order to compare the CONTROLLED model with other models. I was interested to see if there were statistically significant differences between the tested models. This part tested the last hypothesis to detect if there were differences between the RANDOM, TOP, MANUAL, FACEBOOK_PASSIVE, TWITTER_PASSIVE and CONTROLLED models for the different users' responses.

Therefore, a Wilcoxon Signed Ranks Test was run on the rating scores generated from the same 20 participants to determine whether there was a statistically significant mean difference between the models by measuring user responses using the 7-likert-scale.

Accordingly, I was able to show a statistical preference for the use of the CONTROLLED model where the users had the ability to update their profiles mined passively from social media data streams (5.21) as opposed to the RANDOM model (3.61); a statistically significant increase of 1.6 (($Z = -3.922$, $p = 0.000$, $p > .05$), see Table 16.

Also, participants expressed preferences compared to the TOP models (3.87) which means a statistical significant increase of 1.34 in favour of the CONTROL model ($Z = -3.808$, $p = 0.000$, $p > .05$).

For the users' ratings recorded via self-declared preferences (i.e. MANUAL models), the post-hoc Wilcoxon Signed Ranks Test with P < 0.05 was applied. Test results produced a value of p= 0.001, indicating a difference in means of 1.1, representing a statistically significant increase ($Z = -3.416$, $p = 0.001$, $p > .05$).

A Wilcoxon Signed Ranks Test confirmed that using the CONTROLLED model generated higher scores than the FACEBOOK_PASSIVE model (with an average evaluation score of 5.21 vs. 4.62 respectively) and higher score than the TWITTER model (with an average evaluation score of 5.21 vs. 4.74 respectively). Again, this proved a statistically significant improvement over both the FACEBOOK_PASSIVE (0.0048) and the TWITTER_PASSIVE (0.0054) filtering. This evidences the fact that direct user input into user models can but improve their results.

Therefore, **for hypothesis H5,** I was able to conclude that by allowing the user to update their passively mined personal information from different data sources elicits a statistically significant increase in user' interests scores. This indicates that the CONTROLLED model was the most effective in our test and shows a statically significant to other tested models. Full details are provided in Table 16 and 17.

Table 15: Wilcoxon Signed Ranks Test Statistics

	CONTROLLED - R	CONTROLLED - TOP	CONTROLLED - M	CONTROLLED - F_P	CONTROLLED - T_P
Z	-3.922^b	-3.808^b	-3.416^b	-2.819^b	-2.778^b
Asymp. Sig. (2-tailed)	.000	.000	.001	0.0048	0.0054

Model R = RANDOM, TOP = TOP_FEEDS M = MANUAL, P = PASSIVE

Table 16: Ranking Results

		N	Mean Rank	Sum of Ranks
CONTROLLED - RANDOM	Negative Ranks	0[m]	0.00	0.00
	Positive Ranks	20[n]	10.50	210.00
	Ties	0[o]		
	Total	20		
CONTROLLED - TOP	Negative Ranks	1[y]	3.00	3.00
	Positive Ranks	19[z]	10.89	207.00
	Ties	0[aa]		
	Total	20		
CONTROLLED - MANUAL	Negative Ranks	3[ah]	4.50	13.50
	Positive Ranks	17[ai]	11.56	196.50
	Ties	0[aj]		
	Total	20		
CONTROLLED - FACEBOOK_P	Negative Ranks	3[an]	9.83	29.50
	Positive Ranks	17[ao]	10.62	180.50
	Ties	0[ap]		
	Total	20		
CONTROLLED - TWITTER_P	Negative Ranks	4[aq]	6.50	26.00
	Positive Ranks	15[ar]	10.93	164.00
	Ties	1[as]		
	Total	20		

8.7 Discussion and Post-Analysis

It would seem that constructing a user model either via passively mined personal information or via explicitly defined user preferences can be very effective in cross-system recommendations. Moreover, putting the user in charge of controlling the user model built from personal information, which retrieved passively through social network data streams, can play a significant role in improving recommendation accuracy. Also, this would help the recommendation process to enrich the user experience by finding the most relevant items which match user interests. This part explores the potential reasons for this improvement in more detail.

Results from our user experiments revealed that generating a user model based on implicitly or explicitly defined preferences played a significant role in representing the most personalised items for users compared to those with no preferences. Also, it indicated that any type of passive filtering generated the most relevant news item recommendations for users in compassion to all other types of filtering. This means that applying implicit information in constructing a user model would be significantly effective in enriching user experience with the most suitable recommendations.

For an overview, Figure 38 shows the average mean scores for RANDOM, TOP, explicit and implicit filtering and the CONTROLLED model for each participant; it would appear that most of the articles selected by the CONTROLLED model were favoured by the majority of the participants. In 70% of cases, CONTROLLED model improved over all other models. In comparison to some models, the CONTROLLED model was favoured by all participants over the RANDOM and the TOP models while it recorded improvement over the MANUAL model in 90% of cases. In addition, there were only three instances where participants exhibited an overall preference for recommendations made by the FACEBOOK_PASSIVE model and only four cases, for recommendations generated via the Twitter-user model.

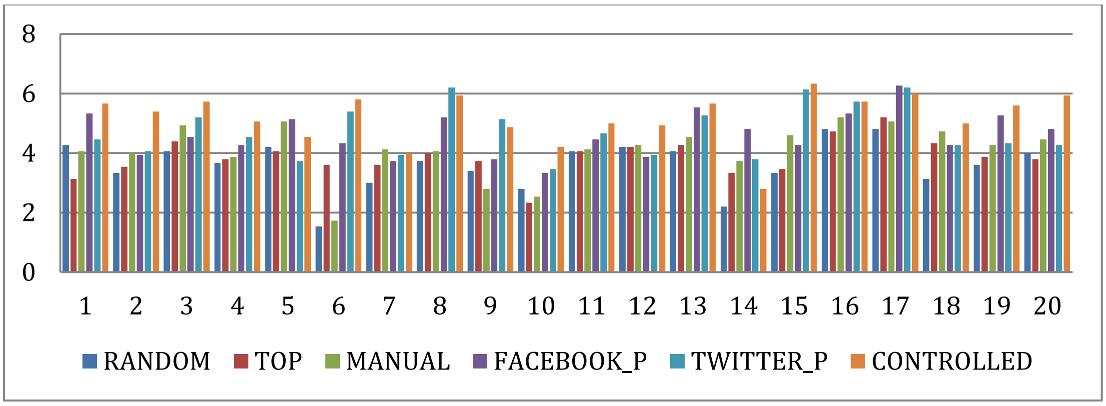
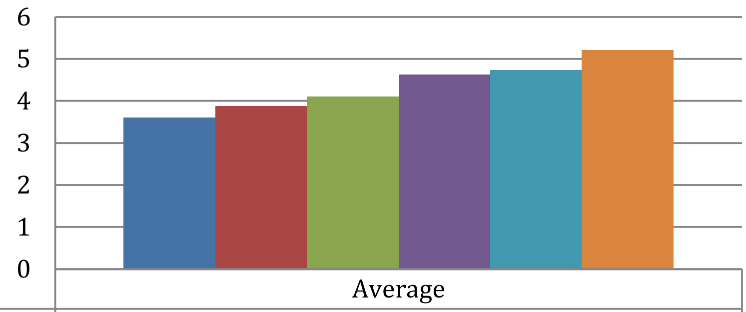


Figure 38: Participants' mean evaluation scores across CONTROLLED, FACEBOOK_PASSIVE, TWITTER_PASSIVE, MANUAL, TOP and RANDOM models

By looking at the average evaluation score, it can be stated that the CONTROLLED model, where participants were able to see their profiles generated from the different social data streams, generated the most relevant recommendations to users with an average rating score of 5.21 in comparison to other models. Models which were constructed based on implicit or explicit user preferences, recorded mean relevance scores over 4.1. For example, the passive models mined from users' Facebook and Twitter activities logged mean rating scores with 4.62 and 4.74 respectively, which is higher than the relevance rating score of 4.11 recorded via a self-declared preference. However, the mean relevance score recorded for our baseline model was 3.61 and 3.87 for the recommendations nominated from the top stories articles pool, as illustrated in Figure 39.



	Average
RANDOM	3.6100
TOP	3.8733
MANUAL	4.1100
FACEBOOK_P	4.6233
TWITTER_P	4.7367
CONTROLLED	5.2100

Figure 39: Mean evaluation scores for the CONTROLLED, FACEBOOK_PASSIVE, TWITTER_PASSIVE, MANUAL, TOP and RANDOM models

It was surprising that users preferred the recommendations generated via the Twitter user model, where profiles have been constructed based on user's tweets, more than these filtered based on the profile constructed from user Facebook data stream. In 60% of cases, the recommendations served to users, based on a user's Twitter timeline, recorded a higher relevance score than those generated via the Facebook user model. Based on the average of users' rating scores in each model, the lowest mean average score recorded in the Twitter-based model was 52, while 50 was recorded for the Facebook-based model as the lowest mean relevance score.

There were only seven cases where participants found that the passive model constructed from Facebook posts gave more accurate articles than the one based on Twitter data. In order to understand the reasons for this difference, it was founded that some users were allowing their tweets to be automatically posted on their Facebook timeline which made their Facebook profile richer with personal data. This led to a profile which was automatically built with content that represented more of the user's preferences. An example of this was participant 6 who used his/her tweets to be posted automatically on their Facebook wall. After closer inspection, he or she didn't use the word "Church" to a great extent on their Twitter timeline, resulting in this term not being considered as important in building the user's profile. However, since the same term was also mentioned many times on the user's Facebook posts, in addition to taking into account this term shared from the user's tweets, this assigned a high weight to this user in the Facebook-based model. As a result, the participant was recommended the following article via the FACEBOOK_PASSIVE model:

News item 1: "Call for fewer Church bishops in Lords"

This article was missed by the Twitter-based model and was not listed in the top recommendations to a user; it was ranked sixth on the evaluation scale which meant that the user found it an interested article.

Therefore, it would seem that Twitter reflects user preferences much better than Facebook since most of the cases assured that users shared most of their interests on the Twitter account. In fact, the Twitter-based model achieves statistically significant results over the Facebook-based model according to representing users' interests and generating recommendation matching their interests.

By looking at the statistics from the CONTROLLED model with Facebook and Twitter, it can be seen that the CONTROLLED model performed better than the FACEBOOK model with 0.0048 while it was 0.0053 with the TWITTER model (see p-value in Table 16). This means that the CONTROLLED model was able to surpass the FACEBOOK model more easily than the TWITTER model. This small difference would be considered in favour of the TWITTER model. This could also explain why the TWITTER model performs better than the FACEBOOK model in 60% of cases.

For all users, on average the cosine similarity score was .02 based on a user TWITTER_PASSIVE profile, but .019 for FACEBOOK_PASSIVE while the cosine similarity to the CONTROLLED model was 0.3 (Figure 40 shows these difference for all users).

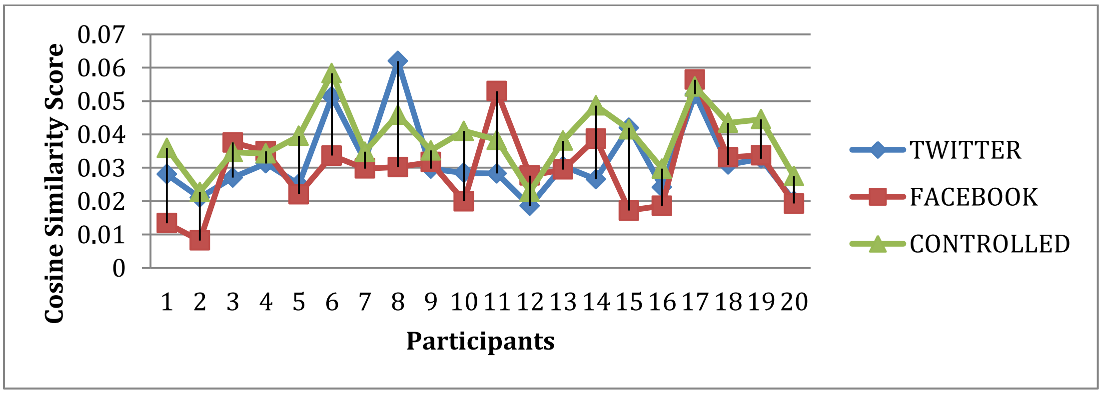


Figure 40: Plot of means of recommendation relevance cosine similarity for all users, based upon the Twitter, Facebook and modified profiles extracted VSM for all participants

The articles recommended to each user were investigated, based on the modified profiles by users. As I expected, on average, the highest cosine similarity average score recorded for the CONTROLLED model, was 0.104. This would be expected, since these recommendations were generated according to the user's updates to the generated profiles. However, I compared the cosine similarities of these recommendations with the user's Twitter and Facebook profiles, and it was founded that the recommendations recorded were more similar to the user's Twitter profile than to their Facebook profile with an average of 0.06 and 0.04 respectively. This means that a user's interests, mined from the Twitter data stream, are more likely to have a better match to the user's updates than the one mined from the Facebook data stream.

An example of this occurred with participant 1, who was served with articles recommended by the CONTROLLED model where these recommendations recorded a cosine score of .08 for the user's Twitter profile and 0.01 for the user's Facebook profile. There was a difference in cosine similarity mean of 0.07 indicating that the Twitter posts could be more accurate in matching the user's preferences than Facebook posts. By looking at the articles which were rated highly by participant 1 (evaluating them with a score of 7), recommended by the CONTROLLED model, it was noticed that these articles matched best to the user's Twitter profile with a 0.09 cosine score and 0.002 for the user's Facebook profile. Three articles were found as interesting recommendations by participant 1 and were rated with scores of six, seven and seven despite having no similarity to the user's Facebook profile with 0 cosine similarity score . However, these articles recorded cosine similarity scores of 0.07838, 0.079511 and 0.084 to the user's Twitter profile and 0.11, 0.096 and 0.093 16 to the CONTROLLED profile.

The following articles were recommended to the participants:

News item 2: "VIDEO: May: Digital society has 'challenges'"

News item 3: "Hunt 'misrepresented deaths data'"

These articles were rated highly with an average score of six, despite the user stating no general interest in politics and health categorised items. In addition, there were six articles overlooked by the MANUAL model but recommended, based on the user's updates that were applied to the generated profile.

In another example, participant 11 was recommended by two articles tagged with different labels and he or she scored the first one with a score of five and the second one with a score of three. By looking at these articles, it was noticed that the user assigned a high score to the article which was more similar to his/her Twitter profile with a cosine score of 0.022 while it had no similarity at all to his/her Facebook profile. However, the second recommendation was rated with a low score although it had a 0.0463 cosine score to the user's Facebook profile. On investigation of its similarity to the user's Twitter interests, I found that there was no similarity to his or her Twitter profile. This would indicate that the VSM created via Twitter data posts would representing a user's interests more accurately than the VSM created via Facebook media posts. The title of the news items was:

News item 4: "What is name-blind recruitment?"

News item 5: "Queen sends message to Lomu family"

Figure 41 illustrates the average of the cosine similarity of the recommendations, served in the CONTROLLED model, calculated based on its similarities to Twitter, Facebook and Controlled profiles for each participant.

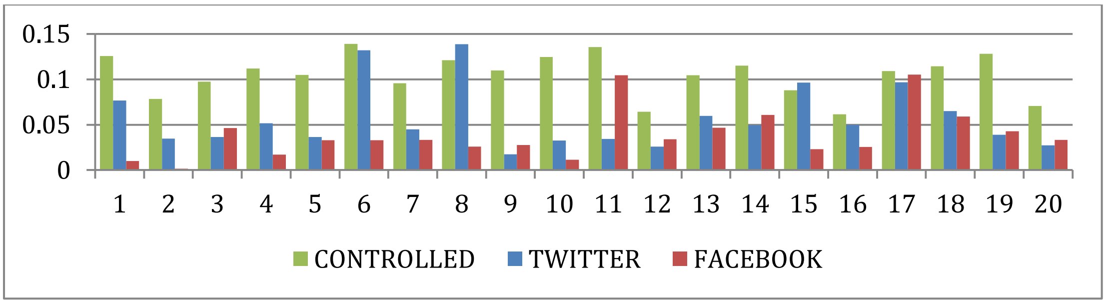


Figure 41: The mean cosine similarity score based on the Twitter and Facebook extracted VSM for the recommendations served via the CONTROLLED model

User Experience

To assess the quality of generating a user model based on implicit user preferences mined from social network data, a questionnaire was also completed by all participants. Figure 42 presents the results from the study questionnaire. The goal of the questionnaire was to examine the perceived accuracy of profiles generated in each of the passive models in terms of reflecting the user's interests and the quality of these profiles in producing better recommendations matching user's preferences. It analysed the effectiveness of the social media platform type on the role of representing user interests and matching his/her interests. Also, I investigated the level to which users believed that constructing a model via mining user's social data streams was useful.

All positively phrased questions received scores on a 1-5 scale. The graph shows results for both Twitter and Facebook generated profiles in representing user interests. Users in ten of the cases expressed that the Twitter profile is more accurate in reflecting their interests with a score of 5. On average, Twitter recorded a score of 4.6 matching user interests as opposed to 3.8 for the Facebook profiles. Also, Twitter was reported as the most useful source for generating better recommendations (4.3) followed by 3.15 for Facebook.

For the more general questions, the highest score of five, for example "can you see value in combining different sources to get better recommendations?" and "do you understand how you get the recommendations?" (as presented in the Appendix C). It was clear that users found that by applying a different data source from a different system this could be helpful and effective, with an average score of 4.7. In terms of understanding why users perceived their recommendations, users recorded an average score of (4.45).

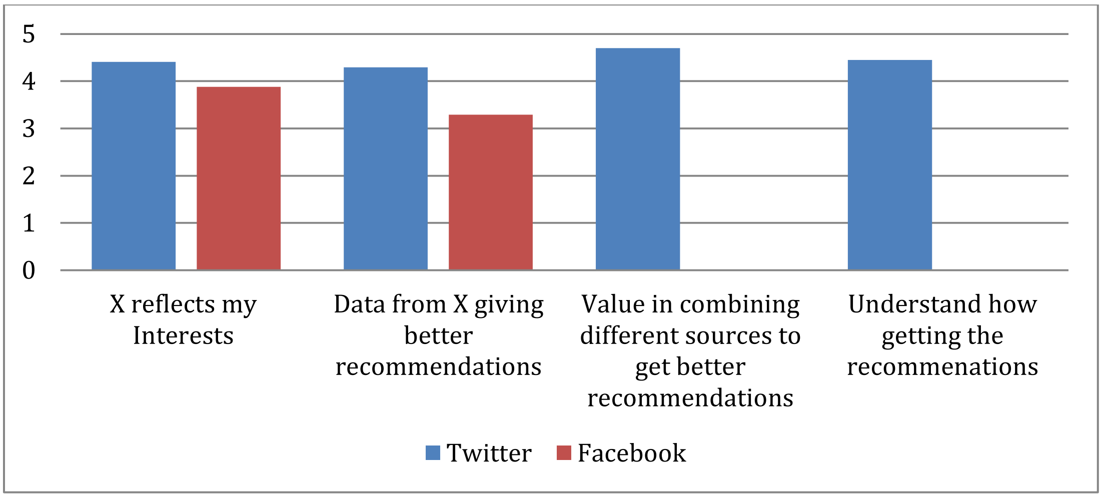


Figure 42: Post-study questionnaire results

Based on the results from the post-study questionnaire, it can be acknowledged that user's use Twitter mainly for their academic, business or work activities, resulting in their interests being presented. Also, this leads Twitter to represent users' preferences in a way that allows any model that is constructed from Twitter activities to be well-matched interests in reflecting well-matched interests. An example of this is participant 14 who uses Twitter to share and talk about what is going on in the media and the news. He or she stated that Twitter was the service they turned to first in order to discuss and express themselves when sharing their thoughts about any topic and then this may be also be shared on their Facebook wall.

Also, they confirmed that Facebook would not be the best place to share controversial topics, despite sharing some of their tweets on Facebook. I found that there was a quite a similarity in the interests on both of the user's profiles as a result of sharing some of his/her tweets on their Facebook wall. Another reason to this similarity was that they were active on both of the platforms and opened both of the social media accounts at the same time. This user rated their Twitter generated profile from their timeline by four out of five and three out of five for their Facebook profile, in terms of reflecting his/her interests.

Another example was participant 6 who claimed that Twitter was the right place to share and discuss topics in which he/she was interested. Also, they shared ideas and topics they were experienced in and joined hashtags that mentioned those topics. The reason behind this was that it would represent him/her publicly as having the necessary skills to find a job via the social media platform. Likewise, this could also help companies or organisations to find people who matched their requirements for job vacancies. Many news articles and papers which match the participant's interests were shared on their Twitter profile. However, the participant's Facebook account was used mainly for his personal *social life*, commenting on his friend's posts and representing his/her daily life activities which did not represent their actual interests, and thus did not express their preferences. This could explain why the participant declared a high average interest of 5.4 in news articles served via the Twitter-based model, while the average was 4.3 for the Facebook-based articles. It was unsurprising that the user found his profile generated from the Twitter timeline as a much better match to their preferences when compared to the profile generated from Facebook. Therefore, the Twitter mined profile was rated highly by the participant (evaluating it with a score of five out of five) while the score assigned to the mined Facebook profile was three.

There were also other participants who shared some thoughts about using Twitter and Facebook. Participants two and four also pointed out that Twitter was a platform for them to share topics related to their academic, science and work which mainly represented their jobs and academic fields. Participant 4 joined a number of hashtags that talked about their research field as well as hashtags related to conferences he or she joined such as #circlesq and #RSscf. Participant 2, who is a PhD student, used Twitter as a target for his work and research interests while the Facebook account was avoided for sharing work details and used mainly for personal and family topics.

Participant 7 was surprised because the model generated a profile that they thought did not reflect their interests because of some words were appeared in the generated word cloud form of his/she profile. The participant claimed they did not join any of the places that appeared in the word cloud of the profile mined from their Facebook data stream. By looking closer at the participant's profile, I could see that the user liked and sometimes also commented on club events that he never joined. He or she only shared information about these events in order to win free vouchers or discounts. Most of his/her activities on Twitter were to share thoughts about events and topics related to his/her country. Also, they found that Twitter was the best place to share information about his/her new business and to try to attract more customers and advertise their business. This would clarify why the participant expressed a preference to the TWITTER_PASSIVE model (5.1) as opposed to the FACEBOOK_PASSIVE model (3.8).

Another common theme was that articles were missing via the MANUAL model. This was due to the articles not manually selected while the user stating his interests in the label which they were tagged with. Because this pattern was so frequent, I present some specific examples in closer detail.

One such instance occurred with participant 15, who rated political articles recommended by the PASSIVE models highly (evaluating them with an average score of 6.6), despite stating no general interest in political categorised items. He or she nominated three subsets, including the "UK", "World" and "Entertainment and Arts", explicitly while the MANUAL model was constructed. However, when the PASSIVE model created a VSM via his/her Twitter social media posts, it found eight hits to items with the 'Politics' tag and they were evaluated with a six or a seven. Also, based on the user's Facebook social media posts, it found an article with the 'Politics' tag which scored a six. From the "Politics" subset, the CONTROLLED model served several recommendations which the user evaluated as being highly relevant. Two illustrative examples were:

News item 6: "Brexit: Lib Dems 9,000 new members after EU vote"

News item 7: "Brexit: No EU compromise on freedom of movement"

Participant 8 stated that items tagged as 'Technology', 'World', 'UK', 'Politics' and 'Science and Environment' events were not of interest to him/her. The participant evaluated the articles recommended from his or her subset with an average score of four. However, when the PASSIVE model created a VSM via his/her Facebook social media posts, it found several hits to items with the 'Health' and 'Education and Family' tags which they scored at an average relevance of 5.3. Similarly, when the two 'Business' and 'Education and Family' categorised items were recommended to the user, the recorded average rating score was 5.5. Also, the CONTROLLED model constructed based on user updates, found several items with the 'Education and Family', 'Entertainment and Arts' and 'Business' tags. Two illustrative examples, which the user evaluated as having a high relevance to them, from the CONTROLLED model, were:

News item 8: "Google 'collects children's data'"

News item 9: "Mail loses Paul Weller privacy appeal"

Investigating the participant's VSM indicated that the similarity was being expressed due to a high *TF-IDF* score for the features 'privacy' and 'data'. This was corroborated by the detection of posts in his/her Twitter and Facebook timelines expressing how much Google knows about users, phone logs revealing deeply personal information and other privacy issues.

The CONTROLLED models were constructed based on user updates to their generated profiles, mined passively from different social media data streams. Each participant was able to eliminate and/or add new items to the generated profiles. Therefore the profile was built under the user's control with an ability to reflect their interests by updating their profiles. Participants eliminated items that were not representing their preferences due to the locality or generality and other reasons.

Participant 19 eliminated some highly specific words such as "VICTORIOUSGRACE" and "Fulani" because they are so local and he/she thought they would not be useful in providing interesting recommendations. However, participant 6 removed words which were more general and didn't reflect his/her interests, which hence may have lead to poor recommendations. So, he/she eliminated words such as "Time", "Hope" and "Part" and added words such as "Technology". As a result, the MODIFIED model recorded the highest relevance rating score (5.8) among the other five models including RANDOM, TOP, FACEBOOK_PASSIVE and TWITTER_PASSIVE models.

Examples of the articles recommended to the participant included two news articles from the "health" and "Education and family" categories which were overlooked by the manual selection in the MANUAL model. Although they were also missed by the passive models mined from Twitter and Facebook data streams, the user evaluated them as interesting articles with a score of six. These articles produced a cosine similarity score of 0.186 and 0.1003 to the profile modelled in the CONTROLLED model. However, their recommendations were more similar to the Twitter profile with scores of 0.175540236 and 0.0946 than the Facebook profile with 0.04 for scored for both of them.

Participant 7 added items to his/her profile which related to their new business in order to find items that could help improve their own business. They updated their profile with words related to business and film topics since as he/she had created a new website for this type of work. As a result, the new articles recommended, based on his/her updates to the profiles, recorded a superior cosine similarity score of 0.11 with an increase of .07 and 0.02 to the recommendations generated from Twitter and Facebook based models. Accordingly, the user was recommended the following articles:

News item 10: "The rise of the student entrepreneurs"

News item 11: "Will students get 'value for money'?"

These articles were recommended to the user via the CONTROLLED model based on his/her changes (these articles are tagged with the "education and family" label). Although the user did not state a general interest in this category, they were highly rated with a score of 7.

In order to evaluate the quality of recommendations based only on the items scored as "interesting somehow", "interesting" or "strongly interesting", I computed the recommendations that matched this requirement. I found that the CONTROLLED model was the best model for generating 224 out of 300 articles evaluated with a score over four followed by the TWITTER_PASSIVE model with 179. The FACEBOOK_PASSIVE model served 170 which were rated over four while the RANDOM, TOP, and MANUAL models recorded less than 155 articles.

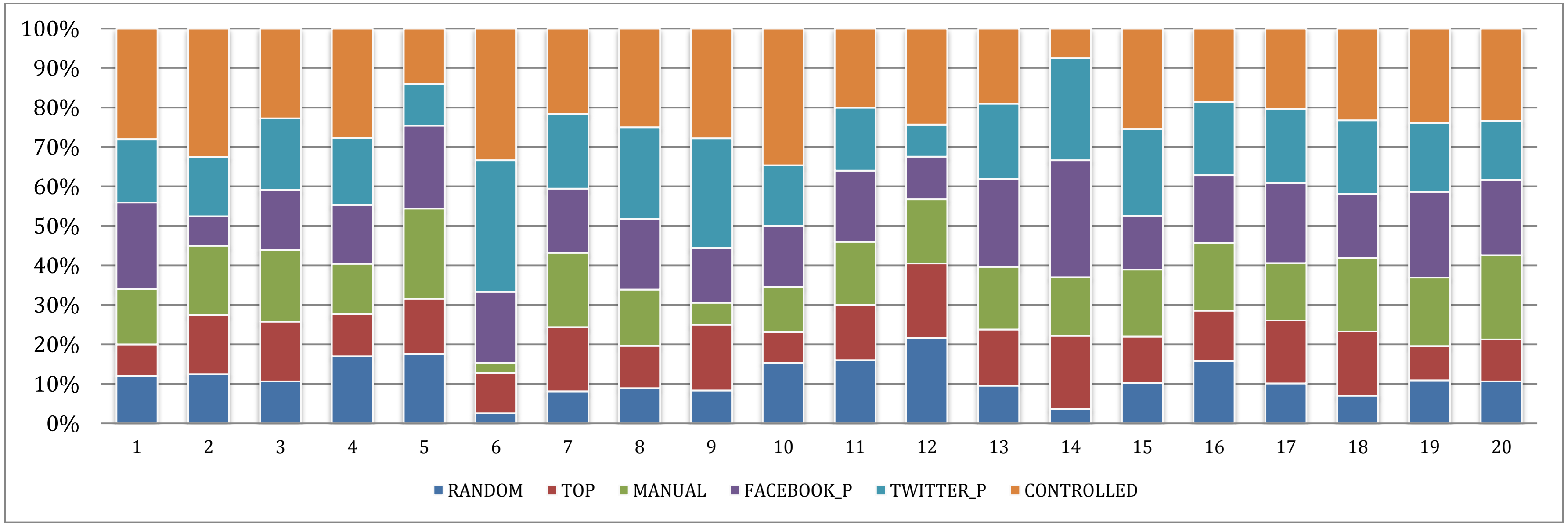


Figure 43: Distribution of the recommendations scored over 4 out of 7 on scale for all models for each user in turn

Fifteen participants found that more than ten out of the 15 recommendations, served to each one of them via the CONTROLLED model, were interesting items. This provides clear evidence that by permitting the user with the ability to update his/her profile would be superior in reflecting his/her inserts and enriching the user' experience. Figure 43 shows the distribution of articles nominated as interesting articles for all users in each model.

Based on user rating score, the following figure shows that the distribution for the RANDOM, TOP, MANUAL, FACEBOOK_PASSIVE, TWITTER_PASSIVE and CONTROLLED models.

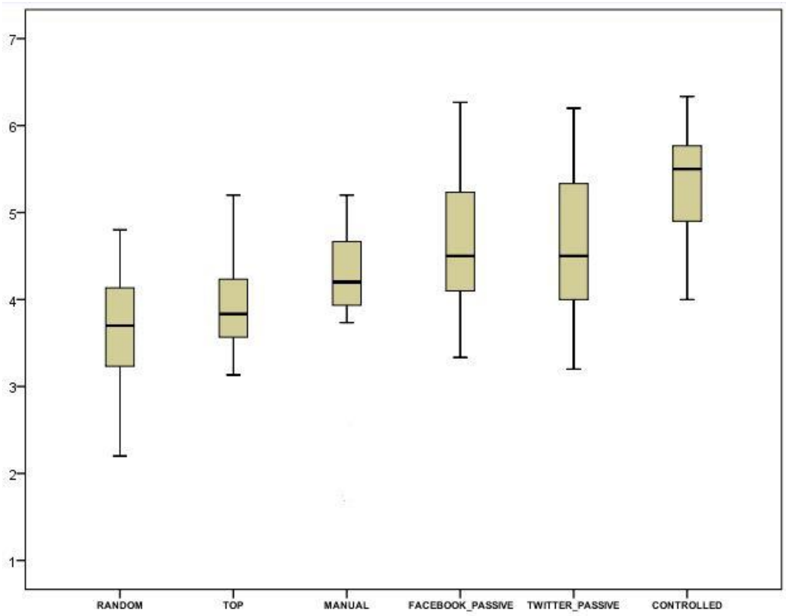


Figure 44: Distribution for each model based on the user's rating score

According to these investigations, this study can draw several conclusions concerning the behaviour of using different social media streams in a PASSIVE approach, and the effectiveness of assigning the ability to change to the user:

1. Participants did use their social network accounts to express things in which they were interested reading about, playing an important role in allowing the PASSIVE model to pick up true positives;

3. The passive model constructed via mining social media data streams did overcome missing highly relevant content-based recommendations in the MANUAL model, if tagged with an over-generalised label;

4. While any PASSIVE modelling via Twitter or Facebook produced statistically superior results to MANUAL models, different forms of preference information can be captured by the two approaches;

5. Profiles modelled from a user's Twitter activities seems to be more accurate in reflecting user interests than those built via a Facebook platform;

6. Allowing users to be in charge of controlling their data and giving them the ability to decide what information they would like to use and to add in constructing their profiles, would help in building an effective users' profile which can play a vital role in generating accurate recommendations and producing statistically superior results to other models.

These outcomes led me to conclude that not only could an effective users' profile be constructed from a user's social media data stream, but also a hybrid model, constructed via cross-system recommendation, would be very useful in improving recommendation results. Also, even moreso, effective functionality could be accomplished by generating a user model from numerous combined domains (web search logs, Twitter posts, Facebook usage, etc.).

8.8 Conclusion

This study explored the possibility of using derived user preferences from public social media networks, specifically from Twitter and Facebook, for a content-based recommendation system. By combining the advantages of both cross-system and content-based filtering, I was able to achieve a practical and effective recommendation system based on personal data stream information mined from different social media platforms (i.e. Twitter and Facebook).

In order to improve the performance of the recommendation process, this information can then be integrated as input into the content recommendation system during the similarity calculation stage.

In order to further personalised recommendations, the recommendation process was fed with extra information from outside the system by giving users the ability to update their profiles, resulting in building an effective users' profile. All these actions were evaluated by recruiting 20 Twitter and Facebook users to join the study which produced 1,800 data points via real world user assessments of BBC News recommendations.

The results demonstrate the potential values of utilising available data in a social media network, particularly in Twitter and Facebook for personalising services that require more data about users in order to provide effective results.

The study findings also reveal that recommendations based on Twitter data are at least as good as those based on Facebook data. While qualitatively they appear even more accurate in certain cases, no statistically significant difference was isolated in these experiments. Based on our questionnaire, profiles modelled from users' Twitter data streams, were more identifiable and reflected a user's preferences more accurately than those modelled via a user's Facebook data. However, this type of model (i.e. cross-system recommendation) would be useful in certain situations such as the cold-start situation when there is no data available about users for generating recommendations or in a high sparsity situation.

By examining different models, which included cross-system data from Twitter and Facebook to improve recommendations, we were able to state that cross-system models were superior to explicit filtering using self-declared preferences and random selection in accuracy by finding more relevant items matching a user's interests. Moreover, when the user was in charge of controlling their own data and personal information, it allowed generating an effective users-based model which was superior to all other models.

Therefore, it can be concluded that not only does a user model generated by using personal information such as a user's tweets or posts elicits a statistically significant increase in performance, but this can be further augmented when placing the user in a position of control over that model.

9. Conclusion

As I have discussed in the previous chapters, in order to develop my experimental framework, a number of requirements have been identified:

- The availability of a source domain from which to draw resources for recommendations. This could be conducted from any source of items, services, etc. and prepared for the process of recommendation, but in our studies, I focussed on items obtained the BBC News RSS streams. This provided the output documents upon which recommendation performance was assessed.
- The availability of user profiles that capture a wide range of user interests from different domains (i.e. social media data). In Chapter six, capturing user interests by mining Twitter data streams was provided. Chapter seven presented a fully implemented model of user profiles through the mining Facebook posts, contrasting the differences with those found in Chapter six. Implementation of modified profiles updated by users themselves to reflect their interests was examined with in Chapter eight. In all case captured information was used to assess cross-system approaches to content recommendations.

Generally, I found that recommendation systems can benefit from information from outside the system in constructing users' profiles. In addition, no explicit user input from Twitter and Facebook platform is an effective source to infer user profile.

Statistically, it was found that the cross-system approach significantly outperforms explicit filtering using self-declared preferences. This result is important as it enables different recommendation systems to generate well-matched recommendations when user's data is sparse or non-existent. This also can enrich user's data to improve the performance of the recommendation system. A user model generated by using personal information shows a statistically significant increase in performance. It can be further augmented when placing the user in a position of control over that model.

In this final Chapter of this thesis, a summary of the experimental results conducted in Chapters six, seven and eight is presented with their contributions to the central thesis.

Additionally, in the last section of this Chapter, an overview of the research questions that remain open, and an outline of directions for future work are provided.

9.1 Study 1

Twitter was successfully used in the role of producing a cross-system recommendation. This was demonstrated in Chapter six, by implementing a cross-domain user modelling approach using Twitter media data via real world user assessments of items recommendations.

Three different models were constructed for generating items for users joining this experiment. Each model was applied, based on different techniques, in order to support comparisons. These models were constructed via:

- an n-gram vector-space representation derived from social media streams,
- An explicitly defined declaration of categorical user interests.
- Using random parameterisation (to serve as a baseline for our testing procedure).

In addition, the evaluative methodology applied to assess the benefit of using Twitter data streams and dynamically rate the relevance of articles delivered, based on one of the models above, presented items via a web interface that allowed for relevance feedback ratings.

Therefore, I was able to empirically show that a user "profile" (e.g. interest) information from outside a recommender system can be useful in a recommender system, and a useful user profile of this nature (e.g. interests) could be inferred with no explicit user input from the Twitter data timeline.

9.2 Study 2

Chapter seven extended the cross-domain user modelling approach by integrating passive models generated from different data streams. In this instance, I concentrated more on social connections and the text generated therein, rather than microblogging. Facebook data streams were examined to produce cross-system recommendations from another social media domain. This opened different windows into a user's interests and preferences. First, it provided a model constructed from Facebook posts instead of the model built from the Twitter data stream in Chapter six. In addition, a new user preferences models were generated from a pre-constructed set of documents labelled as top stories from the same resources in Study One.

In order to provide strong evidence, this Chapter argued not only the feasibility of using the Facebook activities of cross-domain approaches to content recommendations, but also the potential value of passively mining *a range* of web behaviour in order to generate user-preference models (rather than just one lens into the user's interests). This Chapter showed the possibility of implementing cross–domain recommendations using the Facebook activities and recorded results of similar quality to those produced by using Twitter data streams in Chapter six.

Further, as shown by our evaluation of recommendations drawn from the BBC News RSS feed, the cross–domain recommendations generated via Facebook based passive model approach, proved its beneficial effectiveness in terms of improving the quality of recommendations. Recommendations via Facebook-based were superior to all other models conducted in the study.

9.3 Study 3

Chapter eight addressed the questions of: 1. which social media platform would be more accurate in generating a profile that reflects user' preferences; and 2. how profiles generated passively from social media data streams would perform in reflecting user' interests with the ability for a user to maintain control over such a model. Motivated by data generated from different social media networks to build a united model with the ability to record user updates over the model to generate a high quality recommendation, the profiles in this study were implemented to collect and integrate user interests from different social media data (i.e. Twitter and Facebook posts).

A variety of information sources were merged, including information generated from two different online data stream accounts held by the same user. Further, information can be added or eliminated from information derived from these different social media accounts according to the user's interest. In this Chapter, a user's profile captured a wide range of user interests integrated from his or her different social media data streams and captured from a user's updates. A wide range of models was constructed in this study (including models used in both Chapter six and Chapter seven). As a result, six models were assessed, with results being augmented via analysis of the relevance of these recommendations to the user's interests.

Chapter eight also provided an evaluation task where the user judges the interest's reflected in the profiles generated from the user's social media accounts. This allowed identification via a statistical analysis, of the social media platforms which was better as reflecting user preferences. The 'why' of these results was then also explored by performing a post-study questionnaire completed individually for each user.

By giving the user the ability to maintain control over such a model, it allowed me to investigate user responses when that user was allowed to sculpt, edit and reflect upon that profile. This played an important role in constructing a user-based model to produce recommendations. The recommendations generated in this study were superior in meeting user preferences and recorded a higher relevance score over all the other models. This is seen as evidence to the claim that, when given the ability to maintain control over a generated profile mined passively from a social media data stream, a user will see a corresponding significant improvement in the recommendations produced for them.

Overall, this thesis has provided evidence towards the claims that each of the two main requirements identified at the beginning of this Chapter have been fulfilled by the proposed framework.

9.4 Directions for future work

As is the nature of research, when various aspects of the work presented were carried out during implementing and constructing our studies there were a number of further interesting questions raised. It has not been possible to address all of these questions within the scope of this work. As such, this section provides an overview of the issues that warrant further examination - and possible extensions in any future research.

One limitation of this thesis, and a recommendation for the future, relates to the users' privacy regulations. Any type of personal information, whether extracted from social networks or other personal resources, should necessarily be handled with care and maintain the users' privacy where privacy risks may arise. For example, in Facebook a rich user profile may be generated for various purposes, but even with the user's permission there is a strong argument for the user having control (especially during the extraction of data for building a user model to be used in a cross-system application).

By using social networks, users essentially hand out public and private information to different parties without always being conscious of who is recording what about them [131]. For privacy concerns, therefore, delivering new services and applications that can put individuals in charge of controlling their data would tackle such issues. McAuley et al. [155] discussed that becoming 'Dataware', a project that considers the services required to enable individuals to combine data generated and held about them, would allow individuals to control their data by taking a federated approach to controlling their data's use. This is possible by using a distributed system with interfaces and other software mechanisms. This system can be an appropriate approach because it helps an individual to control his/her data without necessarily taking sole ownership of this data. Research integrating cross-domain user modelling along this track could prove useful.

Several technical extensions to this work naturally lend themselves – first is the opportunity to improve the complexity of the linguistic model used to represent user preferences, for which extensive literature exists (for example the integration of more involved features such as word2vec [10]). Second is to both compare and integrate passive models generated from different data streams (e.g. web search logs, product purchase descriptions, etc.), which potentially offer different windows into a user's interests and preferences.

Another avenue which could be explored in order to enrich a user's profile, the data analysed and retrieved from the social media network could be extended to include different metadata. For example I might consider contextual data (e.g., location and time) or information about likes and posts on Facebook or even the favourite tweets on a user's Twitter account. For a better understanding of the semantic meaning of short posts, it would be useful to enrich the posts or tweets by taking into account different sources (e.g., embedded links in the posts [201][202]). In addition, the bag-of-concepts approach might be used as a background knowledge of concepts from a knowledge base, which is the combination of an ontology and instances of the classes in ontology [203], leveraged for extending a user's interests [204].

Finally, future work in this area could focus on presenting an interactive cross-system recommendation generated from multiple social and semantic web resources [205]. It would be of particular interest to allow user exploring this data in the form of an interactive visual recommender system that combines social and semantic knowledge to personalise any sort of resources (see for example [206]). Such a system could potentially help the user see changes to their profiles interactively and in real time, an affordance which could well be more effective in contacting with user preferences.

It is hoped that this work, with its encouraging statistical results and qualitative analysis, can serve as a stimulus for further research in this expanding field.

References

[1] Y. Noh, Y.-H. Oh, and S.-B. Park, "A location-based personalized news recommendation," *Big Data and Smart Computing (BIGCOMP), 2014 International Conference on*. pp. 99–104, 2014.

[2] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, "A literature review and classification of recommender systems research," *Expert Syst. Appl.*, vol. 39, pp. 10072–10059, 2012.

[3] B. Shapira, L. Rokach, and S. Freilikhman, "Facebook single and cross domain data for recommendation systems," *User Model. User-adapt. Interact.*, vol. 23, no. 2–3, pp. 211–247, 2013.

[4] N. Chekkai, S. Chikhi, and H. Kheddouci, "A weighted-graph based approach for solving the cold start problem in collaborative recommender systems," *2012 IEEE Symp. Comput. Commun.*, pp. 000759–000764, 2012.

[5] D. Zhang, C.-H. Hsu, M. Chen, Q. Chen, N. Xiong, and J. Lloret, "Cold-Start Recommendation Using Bi-Clustering and Fusion for Large-Scale Social Recommender Systems," *Emerging Topics in Computing, IEEE Transactions on*, vol. 2, no. 2. pp. 239–250, 2014.

[6] Y. Rong, X. Wen, and H. Cheng, "A Monte Carlo algorithm for cold start recommendation," in *Proceedings of the 23rd international conference on World wide web*, 2014, pp. 327–336.

[7] A. Elkahky, "A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems," *WWW 2015 Proc. 24th Int. Conf. World Wide Web*, pp. 278–288, 2015.

[8] M. Henderson, D. Lindsay, M. Phillips, and M. De Zwart, "Personal Data : The Emergence of a New Asset Class An Initiative of the World Economic Forum," 2011.

[9] B. Li, "Cross-domain collaborative filtering: A brief survey," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2011, pp. 1085–1086.

[10] A. Ahmed, A. Das, and A. J. Smola, "Scalable hierarchical multitask learning algorithms for conversion optimization in display advertising," *Proc. 7th ACM Int. Conf. Web search data Min.*, pp. 153–162, 2014.

[11] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Twitter-Based User Modeling for News Recommendations.," in *IJCAI*, 2013, vol. 13, pp. 2962–2966.

[12] F. Abel, E. Herder, G. J. Houben, N. Henze, and D. Krause, "Cross-system user modeling and personalization on the Social Web," *User Model. User-adapt. Interact.*, vol. 23, no. 2–3, pp. 169–209, 2013.

[13] D. M. Boyd and N. B. Ellison, "Social network sites: definition, history, and scholarship," *Engineering Management Review, IEEE*, vol. 38, no. 3. pp. 16–31, 2010.

[14] J. Bartlett, "THE DATA DIALOGUE," 2012.

[15] B. Debatin, J. P. Lovejoy, A.-K. Horn, and B. N. Hughes, "Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences," *J. Comput. Commun.*, vol. 15, no. 1, pp. 83–108, 2009.

[16] S. Mahmood, "New Privacy Threats for Facebook and Twitter Users," *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2012 Seventh International Conference on*. pp. 164–169, 2012.

[17] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, "Recommending and evaluating choices in a virtual community of use," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1995, pp. 194–201.

[18] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," in *Proceedings of the 1994 ACM conference on Computer supported cooperative work - CSCW '94*, 1994, pp. 175–186.

[19] U. Shardanand and P. Maes, "Social Information Filtering: Algorithms for Automating 'Word of Mouth,'" *CHI 1995*, 1995.

[20] P.-Y. P. P.-Y. Pan, C.-H. W. C.-H. Wang, G.-J. H. G.-J. Horng, and S.-T. C. S.-T. Cheng, "The development of an Ontology-Based Adaptive Personalized Recommender System," *Electron. Inf. Eng. (ICEIE), 2010 Int. Conf.*, vol. 1, 2010.

[21] R. A. Gotardo, C. Teixeira, and S. D. Zorzo, "IP2 Model - Content Recommendation in Web-Based Educational Systems Using User's Interests and Preferences and Resources' Popularity," *2008 32nd Annu. IEEE Int. Comput. Softw. Appl. Conf.*, 2008.

[22] P. S. Lampropoulou, A. S. Lampropoulos, and G. A. Tsihrintzis, "A Mobile Music Recommender System based on a Two-level Genre-Rating SVM Classifier enhanced by Collaborative Filtering," in *NEW DIRECTIONS IN INTELLIGENT INTERACTIVE MULTIMEDIA SYSTEMS AND SERVICES - 2*, 2009, vol. 226, pp. 361–368.

[23] S. Perugini, M. A. Gonçalves, and E. A. Fox, "Recommender Systems Research: A Connection-Centric Survey," *J. Intell. Inf. Syst.*, vol. 23, pp. 107–143, 2004.

[24] L. R. Gordon, B. N. Miller, J. Riedl, J. A. Konstan, D. Maltz, and J. L. Herlocker, "GroupLens: applying collaborative filtering to Usenet news," *Communications of the ACM*, vol. 40. pp. 77–87, 1997.

[25] P. Resnick and H. R. Varian, "Recommender systems," *Commun. ACM*, vol. 40, pp. 56–58, 1997.

[26] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Model. UserAdapted Interact.*, vol. 12, pp. 331–370, 2002.

[27] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, "Application of Dimensionality Reduction in Recommender System -- A Case Study," *Architecture*, vol. 1625, pp. 264–8, 2000.

[28] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Comput.*, vol. 7, 2003.

[29] "MovieLens." [Online]. Available: http://www.movielens.org/.

[30] B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan, and J. T. Riedl, "MovieLens unplugged: experiences with an occasionally connected recommender," in *Proceedings of the 8th international conference on Intelligent user interfaces - IUI '03*, 2003, p. 263.

[31] R. Zeckhauser, J. Swanson, K. Lockwood, and P. Resnick, "The value of reputation on eBay: A controlled experiment," *Experimental Economics*, vol. 9. pp. 79–101, 2006.

[32] X. Zhao and K. Tajima, "Online Retweet Recommendation with Item Count Limits," *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on*, vol. 1. pp. 282–289, 2014.

[33] FaceBook, "Key-Facts," 2013. [Online]. Available: http://newsroom.fb.com/Key-Facts.

[34] O. Phelan, K. McCarthy, and B. Smyth, "Using twitter to recommend real-time topical news," in *ACM Conference On Recommender Systems*, 2009, pp. 385–388.

[35] J. Davidson, B. Livingston, D. Sampath, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, and M. Lambert, "The YouTube video recommendation system," *Proc. fourth ACM Conf. Recomm. Syst. - RecSys '10*, p. 293, 2010.

[36] A. Sharma and B. Yan, "Pairwise learning in recommendation: experiments with community recommendation on linkedin," *Proc. 7th ACM Conf. Recomm. Syst. - RecSys '13*, pp. 193–200, 2013.

[37] T. Y. Tang and P. Winoto, "The role of user mood in movie recommendations," *Expert Systems with Applications*, vol. 37. pp. 6086–6092, 2010.

[38] R. Valencia-García, M. L. Hernández-Alcaraz, W. Carrer-Neto, and F. García-Sánchez, "Social knowledge-based recommender system. Application to the movies domain," *Expert Systems with Applications*, vol. 39. pp. 10990–11000, 2012.

[39] P. Jomsri, "Book recommendation system for digital library based on user profiles by using association rule," *Innovative Computing Technology (INTECH), 2014 Fourth International Conference on*. pp. 130–134, 2014.

[40] V. García-Díaz, J. M. Cueva Lovelle, P. Ordoñez de Pablos, C. E. Montenegro Marín, O. Sanjuán Martínez, and E. R. Núñez-Valdéz, "Implicit feedback techniques on recommender systems applied to electronic books," *Computers in Human Behavior*, vol. 28. pp. 1186–1193, 2012.

[41] P. O. de Pablos, R. G. Crespo, O. S. Martínez, J. E. L. Gayo, B. C. P. García-Bustelo, and J. M. C. Lovelle, "Recommendation System based on user interaction data applied to intelligent electronic books," *Computers in Human Behavior*, vol. 27. pp. 1445–1449, 2011.

[42] D.-R. Liu, W.-H. Chen, and P.-H. Chiu, "Recommending quality book reviews from heterogeneous websites," *Internet Res.*, vol. 23, pp. 27–46, 2013.

[43] K. Lee and K. Lee, "Using Dynamically Promoted Experts for Music Recommendation," *Multimedia, IEEE Transactions on*, vol. 16, no. 5. pp. 1201–1210, 2014.

[44] M. Barthet, A. Anglade, G. Fazekas, S. Kolozali, and R. Macrae, "Music recommendation for music learning: Hotttabs, a multimedia guitar tutor," *Processing*, p. 7, 2011.

[45] S. K. Lee, S. H. Kim, and Y. H. Cho, "Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations," *Information Sciences*, vol. 180. pp. 2142–2155, 2010.

[46] A. Nanopoulos, D. Rafailidis, P. Symeonidis, and Y. Manolopoulos, "MusicBox: Personalized Music Recommendation Based on Cubic Analysis of Social Tags," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 18, 2010.

[47] J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, L. Zhang, and X. He, "Music recommendation by unified hypergraph: combining social media information and music content," in *Proceedings of the international conference on Multimedia*, 2010, pp. 391–400.

[48] M. P. O'Mahony, M. Coyle, B. Smyth, P. Briggs, and K. McNally, "A Case Study of Collaboration and Reputation in Social Web Search," *ACM Transactions on Intelligent Systems and Technology*, vol. 3. pp. 1–29, 2011.

[49] O. R. Zaiane, "Building a recommender agent for e-learning systems," *Int. Conf. Comput. Educ. 2002. Proceedings.*, 2002.

[50] D. Zakrzewska, "Building Group Recommendations in E-Learning Systems," in *Agent and Multi-Agent Systems: Technologies and Applications, Pt I, Proceedings*, vol. 6070, 2010, pp. 391–400.

[51] X. Wan and T. Okamoto, "Building Learner Profile for Group Learning Recommender System from Learning Process," *2009 Int. Work. Intell. Syst. Appl.*, 2009.

[52] J. Bobadilla, F. Serradilla, and A. Hernando, "Collaborative filtering adapted to recommender systems of e-learning," *Knowledge-Based Systems*, vol. 22. pp. 261–265, 2009.

[53] F. P. Romero, J. A. Olivas, A. Cerezo, J. Serrano-Guerrero, and E. Herrera-Viedma, "A google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0," *Information Sciences*, vol. 181. pp. 1503–1516, 2011.

[54] M. A. Martínez, A. Tejeda-Lorente, C. Porcel, and E. Herrera-Viedma, "A hybrid recommender system for the selective dissemination of research resources in a Technology Transfer Office," *Information Sciences*, vol. 184. pp. 1–19, 2012.

[55] Z. Huang, W. Chung, T.-H. Ong, and H. Chen, "A graph-based recommender system for digital library," in *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '02)*, 2002, pp. 65–73.

[56] E. Herrera-Viedma and C. Porcel, "Dealing with incomplete information in

a fuzzy linguistic recommender system to disseminate information in university digital libraries," *Knowledge-Based Systems*, vol. 23. pp. 32–39, 2010.

[57] E. Herrera-Viedma, J. M. Moreno, and C. Porcel, "A multi-disciplinar recommender system to advice research resources in University Digital Libraries," *Expert Systems with Applications*, vol. 36. pp. 12520–12528, 2009.

[58] Y. Fan, Y. Shen, and J. Mai, "Study of the Model of E-commerce Personalized Recommendation System Based on Data Mining," *Electronic Commerce and Security, 2008 International Symposium on*. pp. 647–651, 2008.

[59] L. M. López-López, R. Miguel, J. J. Castro-Schez, and D. Vallejo, "A highly adaptive recommender system based on fuzzy logic for B2C e-commerce portals," *Expert Systems with Applications*, vol. 38. pp. 2441–2454, 2011.

[60] H. Drachsler, H. G. K. Hummel, and R. Koper, "Recommendations for learners are different : Applying memory-based recommender system techniques to lifelong learning," *Inf. Syst. J.*, vol. 235, pp. 18–26, 2007.

[61] Manisha Hiralall, "Recommender systems for e-shops," Vrije Universiteit, 2011.

[62] F. Xia, N. Y. Asabere, A. M. Ahmed, J. Li, and X. Kong, "Mobile Multimedia Recommendation in Smart Communities: A Survey," *Access, IEEE*, vol. 1. pp. 606–624, 2013.

[63] R. Burke, "Hybrid Web Recommender Systems," in *Adaptive Web*, 2007, pp. 377–408.

[64] B. Krulwich, "Lifestyle finder: Intelligent user profiling using large-scale demographic data," *AI Mag.*, vol. 18, pp. 37–45, 1997.

[65] M. J. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artif. Intell. Rev.*, vol. 13, pp. 393–408, 1999.

[66] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Syst.*, vol. 46, pp. 109–132, 2013.

[67] M. Pazzani and D. Billsus, "Content-Based Recommendation Systems The Adaptive Web," in *The Adaptive Web*, vol. 4321, 2007, pp. 325–341.

[68] L. Lu, Y. Li, and L. Xuefeng, "A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce," *Expert Systems with Applications*, vol. 28. pp. 67–77, 2005.

[69] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, pp. 734–749, 2005.

[70] X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Artif. Intell.*, pp. 1–19, 2009.

[71] M. Claypool, T. Miranda, A. Gokhale, P. Murnikov, D. Netes, and M. Sartin, "Combining content-based and collaborative filters in an online newspaper," in *Proceedings of Recommender Systems Workshop at ACM*

SIGIR, 1999, pp. 40–48.

[72] B. Heitmann and C. Hayes, "Using Linked Data to Build Open , Collaborative Recommender Systems ∗," in *AAAI Spring Symposium: Linked Data Meets Artificial Intelligence*, 2010, pp. 76–81.

[73] P. Formulation, "Recommendation System Based on Collaborative Filtering," no. 1, pp. 1–10, 2008.

[74] M. A. S. N. Nunes, "Towards to Psychological-based Recommenders Systems: A survey on Recommender Systems," *Sci. Plena*, vol. 6, p. 1028, 2010.

[75] J. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," in *The Adaptive Web*, 2007, pp. 291–324.

[76] J. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," *UAI'98 Proc. Fourteenth Conf. Uncertain. Artif. Intell.*, pp. 43–52, 1998.

[77] D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles, "Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach," in *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, 2000, vol. 64, no. 10, pp. 473–480.

[78] C. C. Aggarwal, J. L. Wolf, K.-L. Wu, and P. S. Yu, "Horting Hatches an Egg : A New Graph-Theoretic Collaborative Filtering Approach to Collaborative Filtering," in *Proceedings of the fifth ACM SIGKDD (KDD'99)*, 1999, pp. 201–212.

[79] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," *Proc. tenth Int. Conf. World Wide Web WWW 01*, vol. 1, pp. 285–295, 2001.

[80] D. Billsus and M. J. Pazzani, "Learning Collaborative Information Filters," in *ICML '98: Proceedings of the Fifteenth International Conference onMachine Learning*, 1998, pp. 46–54.

[81] L. Ungar and D. Foster, "Clustering Methods For Collaborative Filtering," in *Proc. the Workshop on Recommendation Systems*, 1998.

[82] R. Jin and L. Si, "A Bayesian approach toward active learning for collaborative filtering," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, 2004.

[83] L. Getoor and M. Sahami, "Using probabilistic relational models for collaborative filtering," *Work. Web Usage Anal. User Profiling*, 1999.

[84] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: A Constant Time Collaborative Filtering Algorithm," *Inf. Retr. Boston.*, vol. 4, pp. 133–151, 2001.

[85] D. Y. Pavlov and D. M. Pennock, "A Maximum Entropy Approach to Collaborative Filtering in Dynamic, Sparse, High-Dimensional Domains," *Adv. Neural Inf. Process. Syst. 15*, vol. 15, pp. 1441–1448, 2003.

[86] G. Shani, D. Heckerman, and R. I. Brafman, "An MDP-Based

Recommender System," *J. Mach. Learn. Res.*, vol. 6, pp. 1265–1295, 2005.

[87] K. Y. K. Yu, A. Schwaighofer, V. Tresp, X. X. X. Xu, and H.-P. Kriegel, "Probabilistic memory-based collaborative filtering," *IEEE Trans. Knowl. Data Eng.*, vol. 16, 2004.

[88] T. Hofmann, "Latent semantic models for collaborative filtering," *ACM Transactions on Information Systems*, vol. 22. pp. 89–115, 2004.

[89] B. Marlin, "Modeling User Rating Profiles For Collaborative Filtering," *Science (80-.).*, vol. 16, pp. 627–634, 2004.

[90] L. Si and R. Jin, "Flexible Mixture Model for Collaborative Filtering," *Mach. Learn. Work.*, vol. 20, p. 704, 2003.

[91] T. P. T. Peng, W. W. W. Wang, X. G. X. Gong, Y. T. Y. Tian, X. Y. X. Yang, and J. M. J. Ma, "A Graph Indexing Approach for Content-Based Recommendation System," *Multimed. Inf. Technol. (MMIT), 2010 Second Int. Conf.*, vol. 1, 2010.

[92] M. Pazzani and D. Billsus, "Learning and revising user profiles: The identification of interesting web sites," *Mach. Learn.*, vol. 27, pp. 313–331, 1997.

[93] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in *Proceedings of the fifth ACM conference on Digital libraries*, 2000, pp. 195–204.

[94] D. Billsus, D. Billsus, and M. J. Pazzani, "A personal news agent that talks, learns and explains," in *Proceedings of the third annual conference on Autonomous Agents*, 1999, pp. 268–275.

[95] D. Billsus and M. J. Pazzani, "User Modeling for Adaptive News Access," *User Model. UserAdapted Interact.*, vol. 10, pp. 147–180, 2000.

[96] M. Balabanović and Y. Shoham, "Fab: content-based, collaborative recommendation," *Commun. ACM*, vol. 40, no. 3, pp. 66–72, Mar. 1997.

[97] M. Pazzani, J. Muramatsu, and D. Billsus, "Syskill & Webert: Identifying interesting web sites," in *Proceedings of the thirteenth national conference on Artificial intelligence - Volume 1*, 1996, pp. 54–61.

[98] Y. Zhang, J. Callan, and T. Minka, "Novelty and redundancy detection in adaptive filtering," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '02*, 2002, pp. 81–88.

[99] R. Burke, "Knowledge-based recommender systems," *Encycl. Libr. Inf. Syst.*, vol. 69, pp. 175–186, 2000.

[100] N. Suksom and M. Buranarach, "A Knowledge-based Framework for Development of Personalized Food Recommender System," *… Support Syst.*, p. 2, 2010.

[101] M. Espinilla, L. G. Perez, L. Martinez, and M. J. Barranco, "A Knowledge Based Recommender System with Multigranular Linguistic Information," *International Journal of Computational Intelligence Systems*, vol. 1. p. 225,

2008.

[102]S. Huang, "Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods," *Electronic Commerce Research and Applications*, vol. 10. pp. 398–407, 2011.

[103] N. Good, J. Ben Schafer, J. a Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl, "Combining collaborative filtering with personal agents for better recommendations," *Artif. Intell.*, vol. vol35pp, pp. 439–446, 1999.

[104] W. X. Hong, Y. Weng, S. Z. Zhu, and M. Q. Li, "Hybrid recommender system for vertical e-commerce website," *Xitong Gongcheng Lilun yu Shijian/System Eng. Theory Pract.*, vol. 30, pp. 928–935, 2010.

[105] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations," *Mach. Learn.*, vol. telligence, pp. 187–192, 2002.

[106] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as Classification: Using Social and Content-Based Information in Recommendation," in *In Proceedings of the Fifteenth National Conference on Artificial Intelligence*, 1998, pp. 714–720.

[107] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," *Proc. 25th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. SIGIR 02*, vol. 46, p. 253, 2002.

[108] A. Popescul, L. H. Ungar, D. M. Pennock, and S. Lawrence, "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments," *Artif. Intell.*, vol. 2001, pp. 437–444, 2001.

[109] G. Gonzalez, J. L. de la Rosa, M. Montaner, and S. Delfin, "Embedding Emotional Context in Recommender Systems," *2007 IEEE 23rd Int. Conf. Data Eng. Work.*, 2007.

[110] M. Nunes, *Recommender Systems based on Personality Traits:Could human psychological aspects influence the computer decision-making process?*, 1st ed. Berlin, 2009.

[111] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems: An Introduction*, vol. 40. 2011.

[112] M.-H. Park, J.-H. Hong, and S.-B. Cho, "Location-Based Recommendation System Using Bayesian User's Preference Model in Mobile Devices," in *Ubiquitous Intelligence and Computing*, vol. 4611, 2007, pp. 1130–1139.

[113] T. Roh, "The collaborative filtering recommendation based on SOM cluster-indexing CBR," *Expert Systems with Applications*, vol. 25. pp. 413–423, 2003.

[114] R. R. Yager, "Fuzzy logic methods in recommender systems," *Fuzzy Sets and Systems*, vol. 136. pp. 133–149, 2003.

[115] G. L. G. Linqi and L. C. L. Congdong, "Hybrid Personalized Recommended Model Based on Genetic Algorithm," *2008 4th Int. Conf. Wirel. Commun. Netw. Mob. Comput.*, 2008.

[116] Y. Xia, Q. Zhu, and X. Luo, "Incremental Collaborative Filtering recommender based on Regularized Matrix Factorization," *Knowledge-Based Systems*, vol. 27. pp. 271–280, 2012.

[117] T. Hofmann, "Collaborative Filtering Via Gaussian Probabilistic Latent Semantic Analysis," *ACM Int. Conf. Res. Dev. Inf. Retr. (SIGIR 2003)*, pp. 259–266, 2003.

[118] I. Soboroff and C. Nicholas, "Combining content and collaboration in text filtering," in *Proceedings of the IJCAI*, 1999, vol. 99, pp. 86–91.

[119] K. MARGARITIS and M. VOZALIS, "Using SVD and demographic data for the enhancement of generalized Collaborative Filtering," *Information Sciences*, vol. 177. pp. 3017–3037, 2007.

[120] S. Z. S. Zhang, W. W. W. Wang, J. Ford, F. Makedon, and J. Pearlman, "Using singular value decomposition approximation for collaborative filtering," *Seventh IEEE Int. Conf. E-Commerce Technol.*, 2005.

[121] L. Sharma and A. Gera, "A Survey of Recommendation System: Research Challenges," *Int. J. Eng. Trends Technol.*, vol. 4, no. 5, pp. 1989–1992, 2013.

[122] M. Condliff, D. Lewis, D. Madigan, and C. Posse, "Bayesian mixed-effects models for recommender systems," *Proc. ACM SIGIR*, 1999.

[123] A. M. Rashid, I. Albert, D. Cosley, S. K. Lam, S. M. McNee, J. A. Konstan, and J. Riedl, "Getting to know you: learning new user preferences in recommender systems," in *International Conference on Intelligent User Interfaces*, 2002, pp. 127–134.

[124] B. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Incremental SVD-based algorithms for highly scaleable recommender," in *{P}roceedings of the {F}ifth {I}nternational {C}onference on {C}omputer and {I}nformation {T}echnology ({ICCIT} 2002)*, 2002.

[125] N. Ramakrishnan, B. J. Keller, B. J. Mirza, A. Y. Grama, and G. Karypis, "Privacy risks in recommender systems," *IEEE Internet Comput.*, vol. 5, 2001.

[126] F. McSherry and I. Mironov, "Differentially private recommender systems: building privacy into the net," *Int. Conf. Knowl. Discov. Data Min.*, p. 9, 2009.

[127] S. Lam, D. Frankowski, and J. Riedl, "Do You Trust Your Recommendations? An Exploration of Security and Privacy Issues in Recommender Systems Emerging Trends in Information and Communication Security," in *Emerging Trends in Information and Communication Security*, vol. 3995, 2006, pp. 14–29.

[128] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, 2011, pp. 1–35.

[129] B. N. Miller, J. A. Konstan, and J. Riedl, "PocketLens: Toward a Personal Recommender System," *ACM Trans. Inf. Syst.*, vol. 22, no. 3, pp. 437–476, Jul. 2004.

[130] J. B. Schafer, J. Konstan, and J. Riedl, "E-Commerce Recommendation Applications," in *Applications of Data Mining to Electronic Commerce SE - 6*, R. Kohavi and F. Provost, Eds. Springer US, 2001, pp. 115–153.

[131] H. Haddadi, R. Mortier, S. Hand, I. Brown, E. Yoneki, D. McAuley, and J. Crowcroft, "Privacy Analytics," *CCR*, no. 0146–4833, pp. 94–98, 2012.

[132] A. Skatova, J. Johal, R. Houghton, R. Mortier, N. Bhandari, T. Lodge, C. Wagner, J. Goulding, J. Crowcroft, and A. Madhavapeddy, "Perceived risks of personal data sharing," in *The Fourth Digital Economy All-hands Meeting: Open Digital (DE)*, 2013.

[133] K. Li, Z. Zhong, and L. Ramaswamy, "Privacy-Aware Collaborative Spam Filtering," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, 2009.

[134] H. Polat and W. D. W. Du, "Privacy-preserving top-N recommendation on horizontally partitioned data," *2005 IEEE/WIC/ACM Int. Conf. Web Intell.*, 2005.

[135] Y. Wang and A. Kobsa, "Performance evaluation of a privacy-enhancing framework for personalized websites," in *User Modeling, Adaptation, and Personalization: 17th International Conference, UMAP 2009, Springer, Berlin (2009a), vol. 5535*, 2009, vol. 2009, pp. 78–89.

[136] Z. Cheng and N. Hurley, "Effective diverse and obfuscated attacks on model-based recommender systems," in *ACM conference on Recommender systems*, 2009, pp. 141–148.

[137] H. Polat and W. D. W. Du, "Privacy-preserving collaborative filtering using randomized perturbation techniques," *Third IEEE Int. Conf. Data Min.*, 2003.

[138] H. Polat and W. Du, "Privacy-Preserving Collaborative Filtering," *Int. J. Electron. Commer.*, vol. 9, pp. 9–35, 2003.

[139] W. Shi, M. Xie, and Y. Huang, "Collaborative spam filtering technique based On MIME fingerprints," *2011 9th World Congr. Intell. Control Autom.*, pp. 225–230, 2011.

[140] S. Berkovsky, Y. Eytani, T. Kuflik, and F. Ricci, "Enhancing privacy and preserving accuracy of a distributed collaborative filtering," in *Proceedings of the 2007 ACM conference on Recommender systems*, 2007, pp. 9–16.

[141] J. Canny, "Collaborative filtering with privacy via factor analysis," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR 02*, 2002, vol. 14, p. 238.

[142] A. Srivastava and G. Geethakumari, "A privacy settings recommender system for Online Social Networks," *Recent Advances and Innovations in Engineering (ICRAIE), 2014*. pp. 1–6, 2014.

[143] E. Position and P. No, "Security Issues and Recommendations for Online Social Networks," *Soc. Networks*, vol. 34, no. 1, pp. 487–502, 2007.

[144] A. Cavoukian, "Privacy by Design and the Emerging Personal Data Ecosystem," *Inf. Priv. Comm.*, 2012.

[145] Facebook, “Facebook Statistics,” 2017. [Online]. Available: http://newsroom.fb.com/content/default.aspx?NewsAreaId=22. [Accessed: 20-Jan-2017].

[146] Foursquare, “About Foursquare,” 2017. [Online]. Available: https://foursquare.com/about/. [Accessed: 20-Jan-2017].

[147] Meetup, “About Meetup,” 2017. [Online]. Available: http://www.meetup.com/about/. [Accessed: 20-Jan-2017].

[148] Twitter, “About Twitter,” *2017*. [Online]. Available: https://about.twitter.com/company. [Accessed: 20-Jan-2017].

[149] YouTube, “YouTube Statistics,” 2017. [Online]. Available: http://www.youtube.com/t/press statistics. [Accessed: 20-Jan-2017].

[150] T. Panja, “Oxford using Facebook to snoop,” 2007. [Online]. Available: http://www.nbcnews.com/id/19813092/. [Accessed: 20-Dec-2016].

[151] “What They Know,” 2013. [Online]. Available: http://blogs.wsj.com/wtk/.

[152] E. Steel and G. A. Fowler, “Facebook in Privacy Breach,” *wsjcom*, vol. 35, no. 35, pp. 1–4, 2010.

[153] Fahmida Y. Rashid, “Facebook Leaks Access Tokens, Exposes Private User Data to Advertisers,” 2011. [Online]. Available: http://www.eweek.com/c/a/Security/Facebook-Leaks-Access-Tokens-Exposes-Private-Data-to-Advertisers-416736/.

[154] J. Bonneau, J. Anderson, and G. Danezis, “Prying Data out of a Social Network,” *2009 Int. Conf. Adv. Soc. Netw. Anal. Min.*, pp. 249–254, 2009.

[155] D. McAuley, R. Mortier, and J. Goulding, “The Dataware manifesto,” *Communication Systems and Networks (COMSNETS), 2011 Third International Conference on*. pp. 1–6, 2011.

[156] “WHAT IS THE CONTROL SHIFT?.” [Online]. Available: https://www.ctrl-shift.co.uk/about_us/.

[157] A. Mitchell, I. Henderson, and D. Searls, “Reinventing direct marketing | with vrm inside,” *J. Direct Data Digit. Mark. Pract.*, vol. 10, no. 1, 2008.

[158] D. Reed, “The Personal Cloud,” 2011. [Online]. Available: http://equalsdrummond.name/2011/02/07/the-personal-cloud/.

[159] A. Narayanan, S. Barocas, V. Toubiana, H. Nissenbaum, and D. Boneh, “A Critical Look at Decentralized Personal Data Architectures,” in *DUMW: Data Usage Management on the Web*, 2012.

[160] R. Mortier, C. Greenhalgh, D. Mcauley, A. Spence, N. Ng, A. Madhavapeddy, J. Crowcroft, and S. Hand, “The Personal Container , or Your Life in Bits,” 2010.

[161] H. Liu, J. Cui, T. Liu, and W. Li, “UPDS: Reliable Storage for Personal Data in Online Services,” *Networking and Distributed Computing (ICNDC), 2010 First International Conference on*. pp. 357–361, 2010.

[162] R. Van Meteren and M. Van Someren, “Using Content-Based Filtering for Recommendation,” *ECML/MLNET Work. Mach. Learn. New Inf. Age*, pp.

47–56, 2000.

[163] B. Larivière and D. den Poel, "Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services," *Expert Syst. Appl.*, vol. 27, no. 2, pp. 277–285, 2004.

[164] I. Cantador and P. Cremonesi, "Tutorial on cross-domain recommender systems," *Proc. 8th ACM Conf. Recomm. Syst. - RecSys '14*, pp. 401–402, 2014.

[165] G. González, B. López, and J. L. de la Rosa, "A multi-agent smart user model for cross-domain recommender systems," *Proc. Beyond Pers.*, 2005.

[166] S. Berkovsky, T. Kuflik, and F. Ricci, "Cross-Domain Mediation in Collaborative Filtering," *User Model. 2007*, vol. 4511, pp. 355–359, 2007.

[167] R. Chung, D. Sundaram, and A. Srinivasan, "Integrated personal recommender systems," in *Proceedings of the ninth international conference on Electronic commerce - ICEC '07*, 2007, pp. 65–74.

[168] A. Loizou, "How to recommend music to film buffs: enabling the provision of recommendations from multiple domains," unpublished, May 2009.

[169] P. Cremonesi, A. Tripodi, and R. Turrin, "Cross-domain recommender systems," in *2011 IEEE 11th International Conference on Data Mining Workshops*, 2011, pp. 496–503.

[170] I. Fernández-Tob'\ias, I. Cantador, M. Kaminskas, and F. Ricci, "Cross-domain recommender systems: A survey of the state of the art," in *Spanish Conference on Information Retrieval*, 2012.

[171] P. Winoto and T. Tang, "If you like the devil wears prada the book, will you also enjoy the devil wears prada the movie? a study of cross-domain recommendations," *New Gener. Comput.*, vol. 26, no. 3, pp. 209–225, 2008.

[172] S. Berkovsky, T. Kuflik, and F. Ricci, "Mediation of user models for enhanced personalization in recommender systems," *User Model. User-adapt. Interact.*, vol. 18, no. 3, pp. 245–286, 2008.

[173] T. Tsunoda and M. Hoshino, "Automatic metadata expansion and indirect collaborative filtering for TV program recommendation system," *Multimed. Tools Appl.*, vol. 36, no. 1–2, pp. 37–54, 2008.

[174] B. Cao, N. N. Liu, and Q. Yang, "Transfer learning for collective link prediction in multiple heterogenous domains," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 159–166.

[175] B. Li, Q. Yang, and X. Xue, "Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction.," in *IJCAI*, 2009, vol. 9, pp. 2052–2057.

[176] M. Nakatsuji, Y. Fujiwara, A. Tanaka, T. Uchiyama, and T. Ishida, "Recommendations over domain specific user graphs," in *Proceedings of the 2010 conference on ECAI 2010: 19th European Conference on Artificial Intelligence*, 2010, pp. 607–612.

[177] A. Tiroshi, S. Berkovsky, M. A. Kaafar, T. Chen, and T. Kuflik, "Cross social networks interests predictions based ongraph features," in *Proceedings of the 7th ACM conference on Recommender systems*, 2013, pp. 319–322.

[178] B. Loni, Y. Shi, M. Larson, and A. Hanjalic, "Cross-domain collaborative filtering with factorization machines," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8416 LNCS, pp. 656–661.

[179] Y. Low, D. Agarwal, and A. J. Smola, "Multiple domain user personalization," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 123–131.

[180] A. Tiroshi and T. Kuflik, "Domain ranking for cross domain collaborative filtering," in *International Conference on User Modeling, Adaptation, and Personalization*, 2012, pp. 328–333.

[181] S. Givon and V. Lavrenko, "Predicting social-tags for cold start book recommendations," in *Proceedings of the third ACM conference on Recommender systems - RecSys '09*, 2009, pp. 333–336.

[182] F. Zhuang, P. Luo, H. Xiong, Y. Xiong, Q. He, and Z. Shi, "Cross-domain learning from multiple sources: A consensus regularization perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 12. pp. 1664–1678, 2010.

[183] I. Fernández-Tobías, I. Cantador, M. Kaminskas, and F. Ricci, "A generic semantic-based framework for cross-domain recommendation," *2nd Int. Work. Inf. Heterog. Fusion Recomm. Syst. - HetRec '11*, pp. 25–32, 2011.

[184] F. Abel, S. Araujo, Q. Gao, and G.-J. Houben, "Analyzing Cross-System User Modeling on the Social Web," in *Proceedings of the 11th international conference on Web engineering*, 2011, pp. 28–43.

[185] M. Szomszor, H. Alani, I. Cantador, K. O'Hara, and N. Shadbolt, "Semantic modelling of user interests based on cross-folksonomy analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, vol. 5318 LNCS, pp. 632–648.

[186] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu, "Personalized Recommendation via Cross-Domain Triadic Factorization," *Proc. 22nd Int. World Wide Web Conf.*, pp. 595–606, 2013.

[187] W. Pan, N. N. Liu, E. W. Xiang, and Q. Yang, "Transfer learning to predict missing ratings via heterogeneous user feedbacks," in *IJCAI International Joint Conference on Artificial Intelligence*, 2011, pp. 2318–2323.

[188] S. Gao, H. Luo, D. Chen, S. Li, P. Gallinari, and J. Guo, "Cross-domain recommendation via cluster-level latent factor model," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2013, pp. 161–176.

[189] O. Moreno, B. Shapira, L. Rokach, and G. Shani, "Talmud: transfer learning for multiple domains," in *Proceedings of the 21st ACM international conference on Information and knowledge management*, 2012, pp. 425–434.

[190] A. Aizawa, "An information-theoretic perspective of tf--idf measures," *Inf. Process. Manag.*, vol. 39, no. 1, pp. 45–65, 2003.

[191] G. Costa and R. Ortale, "XML Document Co-clustering via Non-negative Matrix Tri-factorization," in *Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on*, 2014, pp. 607–614.

[192] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in {P}ython," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[193] J. Tang, S. Wu, J. Sun, and H. Su, "Cross-domain collaboration recommendation," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12*, 2012, p. 1285.

[194] E. Wilde and R. J. Glushko, "XML fever," *Queue*, vol. 6, no. 6, pp. 46–53, 2008.

[195] L. M. LaVange and G. G. Koch, "Rank score tests," *Circulation*, vol. 114, no. 23, pp. 2528–2533, 2006.

[196] V. Bewick, L. Cheek, and J. Ball, "Statistics review 10: further nonparametric methods.," *Crit. Care*, vol. 8, no. 3, pp. 196–9, 2004.

[197] Y. Feng and W. Xie, "Teens' concern for privacy when using social networking sites: An analysis of socialization agents and relationships with privacy-protecting behaviors," *Comput. Human Behav.*, vol. 33, pp. 153–162, 2014.

[198] W. Xie and C. Kang, "See you, see me: Teenagers' self-disclosure and regret of posting on social network site," *Comput. Human Behav.*, vol. 52, pp. 398–407, 2015.

[199] Facebook, "The Graph API," 2016. [Online]. Available: https://developers.facebook.com/docs/graph-api.

[200] K.-W. R. Lee and E.-P. Lim, "Friendship Maintenance and Prediction in Multiple Social Networks," in *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, 2016, pp. 83–92.

[201] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Semantic enrichment of twitter posts for user profile construction on the social web," in *Extended Semantic Web Conference*, 2011, pp. 375–389.

[202] S. Kinsella, M. Wang, J. G. Breslin, and C. Hayes, "Improving categorisation in social media using hyperlinks to structured data sources," in *Extended Semantic Web Conference*, 2011, pp. 390–404.

[203] S. Staab and R. Studer, *Handbook on ontologies*. Springer Science & Business Media, 2013.

[204] G. Piao, "Towards Comprehensive User Modeling on the Social Web for Personalized Link Recommendations," in *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 2016, pp. 333–336.

[205] S. Bostandjiev, J. O. Donovan, and T. Höllerer, “TasteWeights: A visual interactive hybrid recommender system,” in *Proceedings of the 6th ACM conference on Recommender systems*, 2012, pp. 35–42.

[206] S. A. Bostandjiev, J. O. Donovan, and H. Tobias, “LinkedVis : Exploring Social and Semantic Career Recommendations,” *Design*, pp. 107–115, 2013.

Appendix

A. Consent Form

The University of Nottingham

Personalised recommendation system study
Consent form

Thank you for participating in this study. In this session, you will be participating in a lab-based task experiment. For full details, please refer to the information sheet.

Please tick each box to confirm you have understood the agreement.

I confirm that I have read and understood the information sheet provided which I may keep for my records. I confirm that I have had the opportunity to ask the researcher any questions that I have about the study. ☐

I understand that the results of my participation may be disseminated through academic publications and conferences but that my data will be anonymised. ☐

I understand that my participation is entirely voluntary and that I am free to withdraw from the study at any time without giving any reason and without being disadvantaged in any way. ☐

Name of Participant..

Signature of participant...

Date..

B. Information Sheets

Personalised Recommendation System Study
Information Sheet

OVERVIEW

Thank you for expressing interest in the study. This study is designed to measure the impact of using personal information to personalize delivery of BBC news article and YouTube services. This includes understanding the extent to which personal information can improve recommendation accuracy to meet user's interests and enrich their experience. The whole session should not take more 20 minutes.

The session will be conducted by Sultan Alanazi who is a PhD student at the University of Nottingham.

WHAT IS INVOLVED?

Users' Twitter feeds will be parsed in order to access their public accessible timelines and to retrieve tweet content. Then, the similarity between tweets and news feed articles will be computed in order to determine the relevance of an article to that person. Once completed, this is followed by a lab-based task experiment, where the user's feedback and ratings for articles recommended by our system.

The experiment will be conducted as follows:

In this experiment, you will be shown a number of news articles (BBC news feed) drawn from our article pool. You will be asked to identify (via a web form) the categories that interest you. (e.g. you would select the 'Sport' and 'Education and Family' channels). Once this is done, you will be shown a web page that displays a single article. By using 7-point Likert scale, you will be asked to give a score reflecting your interest in this article. This task will be repeated 45 times.

WHAT DATA WILL BE STORED?

The researcher will collect the results from all two parts. Your response (recommendation ratings) will be stored in a database. NO PERSONAL INFORMATION such as personal phone numbers will be stored. Your data will be stored in a secure, password-protected server. It will only be accessible by those directly involved in the research. Users' tweets will be retrieved temporally during the experiment but not stored in the results database.

WHAT ELSE DO I NEED TO KNOW?

The results of the study may be disseminated through academic publications and conferences but your data will be anonymised. Your participation is entirely voluntary and you are free to withdraw from the study at any time without giving any reason and without being disadvantaged in any way. Should you choose to withdraw; any data relating to you will be deleted.

If you have any questions or require further information, please contact sultan Alanazi at psxsa16@nottingham.ac.uk

The University of Nottingham

Personalised Recommendation System Study
Information Sheet

OVERVIEW

Thank you for expressing interest in the study. This study is designed to measure the impact of using personal information to personalize delivery of BBC news article. This includes understanding the extent to which personal information can improve recommendation accuracy to meet user's interests and enrich their experience. The whole session should not take more 20 minutes.

The session will be conducted by Sultan Alanazi who is a PhD student at the University of Nottingham.

WHAT IS INVOLVED?

Users' Facebook and Twitter feeds will be parsed in order to access their timelines and to retrieve post content. Then, the similarity between posts and news feed articles will be computed in order to determine the relevance of an article to that person. Once completed, this is followed by a lab-based task experiment, where the user's feedback and ratings for articles recommended by our system.

The experiment will be conducted in three parts:

- PART 1: In the first part, you will be asked to identify (via a web form) the categories that interest you (e.g. you would select the 'Sport' and 'Education and Family' channels). Once this is done, you will be given 45 articles randomly selected from the article pool. You will be shown a web page that displays a single article. Using 7-point Likert scale, you will be asked to give a score reflecting your interest in this article. This task will be repeated 45 times. This task will be carried out twice.
- PART 2: In the second part, you will be shown number of profiles generated from participants joined session one. You will be asked to identify your profile from the list of profiles and complete a questionnaire.
- PART 3: In the third part, you will be asked to modify the generated profile that includes list of words. Once this is done, you will be given 30 articles randomly selected and asked to give a score reflecting your interest in this article. This task will be repeated 30 times.

WHAT DATA WILL BE STORED?

The researcher will collect the results from all two parts. Your response (i.e. ratings) will be stored in a database. NO PERSONAL INFORMATION such as personal phone numbers will be stored. Your data will be stored in a secure, password-protected server. It will only be accessible by those directly involved in the research. Users' posts will be retrieved temporally during the experiment but not stored in the results database.

WHAT ELSE DO I NEED TO KNOW?

The results of the study may be disseminated through academic publications and conferences but your data will be anonymised. Your participation is entirely voluntary and you are free to withdraw from the study at any time without giving any reason and without being disadvantaged in any way. Should you choose to withdraw; any data relating to you will be deleted.

If you have any questions or require further information, please contact sultan Alanazi at psxsa16@nottingham.ac.uk

C. Questionnaire for participants

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Personalised Recommendation System study

- Please sort the following word cloud cards based on the probability of being your own X social media profile:

Order	Profile number
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

My profile generated from X reflects my interests:	1 Strongly Disagree	2 Disagree	3 Undecided	4 Agree	5 Strongly Agree
• *Twitter*					
• *Facebook*					

Data from X is useful for generating better recommendations	1 Strongly Disagree	2 Disagree	3 Undecided	4 Agree	5 Strongly Agree
• *Twitter*					
• *Facebook*					

	1 Strongly Disagree	2 Disagree	3 Undecided	4 Agree	5 Strongly Agree
I see value in combining different data sources to get better recommendations					
I can understand how I got the recommendations					

Participant name: ……………………………………………………………………………………

Signature: …………………………………………………………………………………………