Article can be accessed online at http://www.publishingindia.com

The Design and Implementation of an Intelligent Online Recommender System

Monika Arora*, Uma Kanjilal, Dinesh Varshney*****

Abstract

Recommender systems (RS) are intelligent applications in the field of information retrieval. Information retrieval assists users to take part in decision making process. It assists them in choosing one item from a vast set of alternative products or services. The scope of recommender systems has expanded gradually over the time from 1990s. As the user provides the inputs, these inputs are recorded and used as recommendations. These inputs are further used in the recommender system tool. The inputs received by the user are aggregated by the other people's inputs and then the system sends them directly to the appropriate recipients (Dean etal., 1995). RS is basically a technology which is based on the important aspects such as collaborative or social filtering. There are many researches already in the area for collaborative filtering and social filtering (Bigus,1996; Rennieand Srebro, 2005). The RS can be used in intelligent information retrieval in the field of artificial intelligence.

From information retrieval, recommendation technology explores and derives the vision that users are searching. In the process of recommendations, the user is engaged in information searching and the RS automates and collects the content required for matching purposes. Typically the search results are arranged in the form of a ranked list. One of the important phases of artificial intelligence is learning process. This will view the past knowledge, buying behaviour, and interest. The RS are primarily based on two phases that are search phase and user-based interaction model. These phases can be identified as user-model construction and recommendation generation. This paper defines the importance of both the models. The interaction model describes the user needs and preferences. Based on the interactions, during the session they are connected to the same interest group. This paper attempts to define a proposed model for considering the factors of intelligent retrieval.

Keywords: Recommender Systems, Intelligent Applications, Retrieval, Filtration

Introduction

An intelligent retrieval consists of sensing elements or agents. These agents receive events, and recognizer or classifier classifies them and that determines which event occurred. Also a set of logic programs to rule-based inference and a mechanism have to take the action accordingly (Bigus, 1996;Rennie and Srebro,2005). Other factors or attributes that are important for retrieval model include mobility and learning. For mobility, an information retrieval searches navigates through a network and performs various tasks on remote machines. A learning retrieval adapts to the requirements of user and automatically changes its behaviour in the features of environmental changes. Learning is an intelligent retrieval, where event-condition-action paradigm can be defined (Bigus,1998). In the context of intelligent retrievals, an event is defined as anything that happens to change the environment or anything of which the retrieval should be aware. At the occurrence of any event, the retrieval has to recognise and evaluate the event means and then respond to it. It also determines what the condition or state of the world is. It could be simple or extremely complex

* Department of IT, Apeejay School of Management, Dwarka Institutional Area, New Delhi, India.
** Department of Library and Information Science, Indira Gandhi National Open University Maidan Garhi, New Delhi, India.
*** School of Physics, Devi Ahilya University, Khandwa Road Campus, Indore, M. P., India.
*** Multimedia Regional Centre, Madhya Pradesh Bhoj (Open) University, Khandwa Road Campus, Indore, M. P., India.

depending on the situation. Recommenders system is considered to be an intelligent retrieval application which facilitates information retrieval.

Recommender system technique is a software application that aims to support the users in their decision-making. It is possible only by interacting with them in large information spaces. These techniques recommend items of interest to users and based on preferences they have expressed (either explicitly or implicitly). The expansion and increase of the data and information on the WWW (World Wide Web) therefore require a system which uses tools for users for information seeking and doing e-commerce activities. A recommender system is a tool to help and overcome the information overload problem also. As we know that it exposes users to the most interesting items, by offering innovation and significance. Recommender technology is the central piece of the information seeking puzzle. Mostly e-commerce sites such as Amazon and Yahoo are using recommendation technology in universal ways. As it insists on the focus to the users, it will give rise to socially connected system i.e. Social Recommender Systems (SRS).

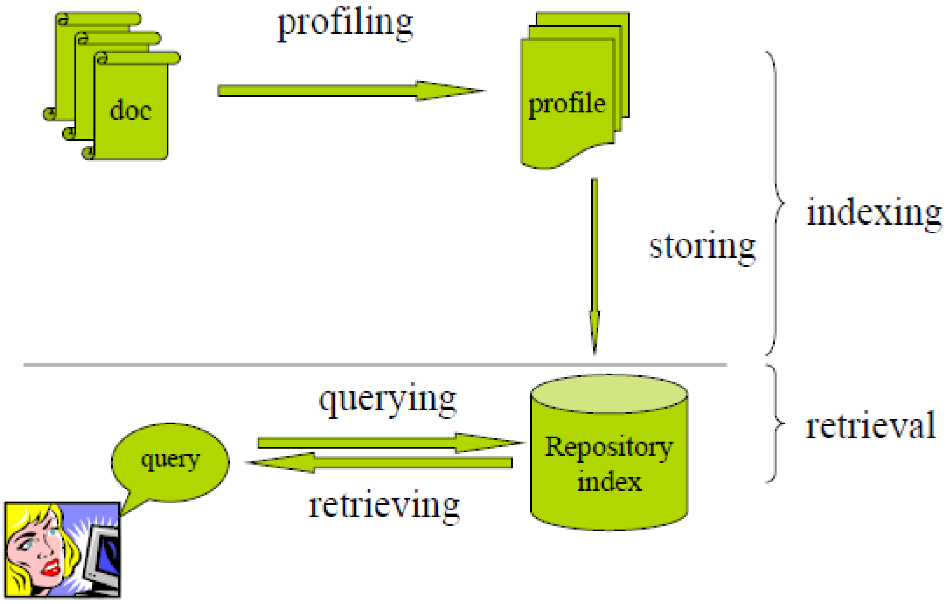
The Social Recommender Systems (SRS) aim to improve upon the information overloading. Over the years social media users are using most attractive and relevant contents at WWW. Also, specific user has to adapt these contents using personalization techniques. Recommender systems and social media can mutually benefit from one other. On one hand, recommender systems can widely influence the triumph of social media, ensuring each user is obtainable in the most attractive and suitable content on a personal basis. On the other hand, social media also introduces new types of public data(in the form of metadata, such as tags, comments, votes). Also it enables to define the explicit people relationships, which can be utilized to enhance and used as recommendations. The retrieval works for the recommender systems in order to: (1) share research documents and techniques that are used to develop effective information retrieval recommenders. It starts from algorithms through interfaces taking the data and evaluating it; (2) identify next key challenges in the specific area, and (3) recognize new cross-topic and the opportunities that take place. To take the advantage of the WWW setting and its broad and wide-spread diverse audience, particularly recommenders system and media are used as either rising applications for recommender systems on the Web or new sources of knowledge particularly Big Data by community and appliance to enhance current techniques, and also to develop new methods for recommender systems on the social websites

Literature Review

The World Wide Web has become the primary source of data for all work activities. WWW has a huge content and would be wasted if that information could not be found, analyzed, and exploited correctly. The relevant and comprehensive information about the user is retrieved quickly. RS techniques have become a principal driver of modernism and a variety of new techniques. It has also been introduced to grow and develop the information based on the base content. For finding and selecting products, services, or information, the tools of recommender systems are becoming popular. Basic concepts in information retrieval are those which deal with the representation, storage and organisation of unstructured data. Information retrieval is the process of searching within a document collection for a particular information need (a query). Its mission is to assist in information search. There are two main search paradigms, retrieval and browsing. The user task includes search for particular information, usually focused and purposeful, and browsing and general looking around for information.

There are many approaches which use recommendation technology for research purposes. In many cases a system designer is employed to set up a recommendation system and also it must choose the best practices between a set of best candidate or best item approaches or practices. The first step for selecting suitable algorithm is to make a decision upon the properties of the request in an application to focus upon when making different alternatives or choices. Recommendation system is diverse in its properties that may affect user experience. The user may check the retrieval in terms of accuracy, robustness, scalability, and so forth. These properties are based on the implementation of Collaborative Filtering algorithm. This also describes the experimental settings for making choices between algorithms. It discusses how to draw trustworthy conclusions from experiment and explains how to evaluate the systems with given relevant properties. The survey sample set of evaluation criteria in the context helps to evaluate the properties of recommender system.

Figure 1: Typical IR System



The flow of the IR systems is described in Figure 1.To compare the documents the similarity metrics for the taxonomy of web search is as follows. In the web context the "need behind the query" is often not informational in nature. It classifies web queries according to their intent in three types i.e. it aids in consistency among judges. The benchmark for identification of user was based on the taxonomy of web search as offered by Broder (Broder, 2002), and further investigated by Schneiderman (Schneiderman *et al*., 1997). Informational searches intent to locate content in terms of keywords concerning to a particular topic in order to address the information need of the searcher. The content is available in variety of forms, including data items, text items, documents, and multimedia items,for example child labour laws, cake recipes etc. Navigational classifications depend upon the content available on the specific location at the website. The searcher have in mind to look for a particular website at particular location. For example search website exists such as www.aol.com, Google, Capital One, YouTube. Transactional classification is used to locate a website with the goal to obtain some products or services. For examples it includes the online purchase and search of a product, execution of an online application (i.e. a game), and downloading multimedia or video for watching movies.

Recommender Systems users

As the user expects the minimum input and the maximum output, most of the users are impatient to get results providing just minimal input. That is why users' preferences are constructive and context dependent. The users want to make accurate choices, i.e., get relevant information items at the first time. The purpose of a Recommender System is to produce or create the meaningful likes or recommendations to collect from users with common preferences for items or products. The recommender systems is used for the operation and giving their strength in the specific domain where the RS concept is implemented. The design of these recommendation engines depends on the domain to domain. Also, it records the items characteristics for which the data/documents are available. The ratings affect the movie watchers, for example the reviews on Netflix regularly offer ratings on a range of 1to 5 where 1 is disliked and to 5 is mainly liked. The quality interaction stake place between users and items. The system uses the records as per requirement it may be user-specific or item-specific. These attributes are based on demographics and product descriptions. Recommender systems vary in the method they analyze these statistics sources to expand notions of empathy, the relation between users and items, in which data can be used to identify well-matched pairs. Collaborative Filtering systems can analyze the historical interactions.

But Content-based filtering systems are based on profile attributes and Hybrid techniques attempt to combine both of these designs. The evaluation on real-world problems is based on the architecture of recommender systems which is an active area of research. Obtaining recommendations from trusted sources is a critical component of the natural process of human decision making which is useful in obtaining recommendations. Based on the growing market scenario, there is a challenge for both buyer (in their more and more choices) and sellers (for personalizing advertising efforts). Also the enterprises have to collect the large transactional data and analyse it deeper to interact the customer with special product offerings.The automated technique using Recommender Systems fulfills the need of both buyers and sellers. The first commercial recommender system was introduced as Collaborative filtering by Tapestry (Linden *et al*., 2003). The system was designed for group of documents in related groups for users. The concept behind this is to facilitate user in only important and relevant documents.To find the user-item pairs which are matched by the criteria provided by the user and the documents can be shared across the users, Collaborative filtering works on this basics The another concept of content filtering originates from the field of information retrieval. Collaboration filtering and content filtering can also work in parallel. Content filtering works for user but collaborative filtering works across the entire user base. Some of the researches using collaborative filtering are related to domain as well (Resnick *et al*., 1994;Dean *et al*.,1995).The initial calculations in recommender systems are based on correlation statistics. Also some are using predictive modeling in support vector machine. The classification classifies the documents and collaborative filtering mapped the classification as per user requirements to achieve the quality retrieval of documents. The RS can directly be implemented in e-commerce applications. The linear algebra and statistical matrix analysis emerges a state of the art technique called matrix factorization (Bell *et al*.,2009).

Learning System: Collaborative Filtering

The user under current consideration for recommendations is referred to as the active user. The categorization can be Collaborative Filtering (CF), Content-based Filtering (CBF) and Hybrid Filtering.A user recommended based on the history ratings records of all users is called Collaborative Filtering. In Content-based recommending, the approaches recommend items that are comparable to items the user has liked in the past, or corresponding to attributes used by the user. The approach which uses both collaborative and content filtering is called hybrid approach.

The user feedback was collected and rated for items in a particular domain. The similarity in their rating recommends the item. This is called the Collaborative Filtering (CF). The neighborhood-based and model-based approaches are two further categories of CF methods. The memory-based approachis referred to neighborhood-based method(Dean*etal.*, 1995).

Neighborhood-based Collaborative Filtering

The subset of users is chosen based on the similar behaviour to active user and the weighted combination of their rating is useful to produce similarity pattern for that user.This prediction based filtering is called Neighborhood-based Collaborative Filtering. Most of these approaches can be generalized by the algorithm summarized in the following steps: 1. Assign a weight to all users with respect to similarity with the active user.2. Suppose we select user ‘k’ this forms a group showing a similarity pattern to the active user referred to neighborhood. 3.By selecting the neighboring ratings, calculate the predictions for active user based on weighted combination. In step 1, the weight x, y is a measure of similarity between the user u and the active user a. Pearson correlation coefficient is used to measure the similarity rating between two users(Broder, 2002). If there are two users x and y the Pearson correlation is defined as

$$sim(x,y) = \frac{\sum\limits_{i \in I_{xy}} (r_{x,i} - \bar{r_x})(r_{y,i} - \bar{r_y})}{\sqrt{\sum\limits_{i \in I_{xy}} (r_{x,i} - \bar{r_x})^2 \sum\limits_{i \in I_{xy}} (r_{y,i} - \bar{r_y})^2}}$$

wherer_{xy} is the set of items rated by both user x and user y.

The cosine-based approach defines the cosine-similarity between two users x and y as:

$$sim(x,y) = cos(\vec{x},\vec{y}) = \frac{\vec{x} \cdot \vec{y}}{||\vec{x}||_2 \times ||\vec{y}||_2} = \frac{\sum\limits_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum\limits_{i \in I_{xy}} r_{x,i}^2}\sqrt{\sum\limits_{i \in I_{xy}} r_{y,i}^2}}$$

wherethe set of items rated by both users,x and y is the rating given to item by theuser.In step 3, based on the neighboring mean the predictions are calculated to see the deviations. Similarity based on measures, the extent to which there is the concept of linear dependence,is measured by Pearson correlation between two variables. Also for m-dimensional space, we plot the ratingsof two users, and calculate using cosine angles between them.The negative ratings are untreated or ignored.It is found that t Pearson correlation performs better in such situations (Resnick *et al.*, 1994). There are similarity measures such as Spearman rank correlation andKendall's τ correlation to calculate the mean squared differences (Su *et al.*, 2007). When applied to millions of users and items, conventional neighborhood-based CF algorithms do not scale well, because of the computational complexity of the search for similar users. As an alternative (Linden *et al.*, 2003), the item-to-item based Collaborative Filtering is called Item-based Collaborative Filtering, in which the similarity is based on matching with items rather than users. This approach results to be faster online systems search applications (Bell *et al.*, 2009; Broder, 2002).

Pearson correlation is calculated as follows:

$$r_{xy} = \frac{n\Sigma XY - \Sigma X\Sigma Y)}{\sqrt{\left[n\Sigma XY^2 - (\Sigma X)^2\right]\left[n\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

where r is the set of all users who have rated both items x and y, r_{xy} is the rating of user r on item x, and r_x is the average rating of the xth item across users. The simple weighted average w, for the item I is calculated as

$$rr_{xy,w} = \frac{\sum w_i x_i y_i}{\sqrt{(\sum w_i x_i^2)(\sum w_i y_i^2)}}.$$

The weighted version of the reflective correlation is where K is the neighborhood set of the k items rated by a, that are most similar to i.

Significant weighing is important for item-based Collaborative Filtering. The comparison of the parameters of item-based methods is found by Su *et al.*(2008).For the active user having highly correlated neighbors,there may also be a probability of overlapping items. This concept is known as significant weighting The classifications should not be too small to have more overlapping, in that case it will be considered as bad predictors. Here significance weighing factor can be used for many correlated items (Herlocker *et al.*,1999).

To implement the constraints the matrix factorization techniques can be used. The distinguish ratings calculate the loss function. The SVM classification approach is based on the training data and it also handles multiple ordered rating in different categories (Rennie and Srebro, 2005).

The implicit preferences take care of likes as they are always considered, but explicit preferences like dislike ratings are always ignored. In large business organizations the transactional records of the product purchases are maintained but they do not maintain the client dislikes like why the customer is not having interest in buying a product. This information is not available explicitly. This information is difficult to gather and used fruitfully in the changing business environment. Also for recommending TV shows it may be based on watching habits of users where the preferences are implicitly used without using the explicit ratings.

Evaluation Metrics

The recommender system is evaluated by comparing the recommendations of items and users. The predictive accuracy metrics (Herlocker *et al.*, 2004), uses the predicted ratings that are directly compared to actual user ratings. Predictive accuracy metrics treat all items equally. The prediction of items a user will like is the primarily concern of the recommender systems. As such, researchers often view recommending as predicting good, i.e. items with high ratings versus bad or poorly-rated items. In Information Retrieval (IR), we define "relevant" and "irrelevant" items. The practices that can be used are IR measures, like Precision, Recall. Also there are other measures, such as F1-measure that can be used. Pearson's product-moment correlation is defined in terms of mean average precision and is calculated based on the normalized distance-based(Herlocker *et al.*, 2004).

$$r^*_{uj} = r_u + K\sum_{v \in N_j(u)} w_{uv}(r_{vj} - r_v)$$

A set of neighbours of u that have rated j

where, r_u is the average rating of user u, K is anormalization factor such that the absolute values of w_{uv} sumto 1, and

$$w_{uv} = \frac{\sum_{j \in I_{uv}} (r_{uj} - r_u)(r_{vj} - r_v)}{\sqrt{\sum_{j \in I_{uv}} (r_{uj} - r_u)^2 \sum_{j \in I_{uv}} (r_{vj} - r_v)^2}}$$

Pearson Correlation of users u and v

Data Interpretation and Analysis

The most general setting in which recommender systems are studied is presented in Table 1. The n users and m items represent the matrix to show the user preferences. Each cell corresponds to a $R_{u,I}$ which represents each item for a user. This user ratings matrix is typically sparse, as most users do not rate most items. The recommendation task analyzes the previous rated tasks and predicts for all the items that are observed by the user. The priority list will be generated based on the highest rated items using recommendations. The recommendations as inputs, Alice enter the keywords and weights for the results. The dataset describes the uses and explores the item 1-5.There are the history data which help to search for item 5.The numbers indicate the priority of a particular item based on the random number assigned to each item with respect to their users.

Table 1: User with Item

	Item1	*Item2*	*Item3*	*Item4*	*Item5*
Alice	5	3	4	4	
Rob	3	1	2	3	3
Carol	4	3	4	3	5
Mark	3	3	1	5	4
Bob	1	5	5	2	1

We use the Pearson correlation method, which helps in correlation. The recommender system technology is known as content repository defined as follows. Based on the correlation data the mark should be the nearest neighbour and then Carol. The interests of Rob and Carol are somewhat close to Alice. This means that when Alice is searching for the item 5 the collections of matching items is typically in the form of a ranked list. From an artificial intelligence point of view this can be viewed as a learning problem that exploits past knowledge about users, such as the search or buying behaviours. First things first,wanted a data structures to hold the recommendations of everyproduct:

Step1: Define a public class recommendations with attributes (Name , Rating)

For simplicity, the products in a vocabulary with the product name as the key and a List<Recommendation> as the value.

Step 2: Call the list recommendations

In a real system, data is existing in a database, but for this example, it should hard code our data into the dictionary:

Step 3: Adding items to the list recommendations as Row1 for item1.Repeat the steps for all 5 items for different 5 users.

Here, for a set collection of items, each product has a collection of reviews. Also the review is on the basis from 1 to 5 describing how the user liked the items where 1 is least liked and 5 is most liked product. This value defines the profile of the product. For example, you could use a value of 1 to represent a user (that a user had bought a particular product with a 0 showing users who did not purchase a product). As long as the numeric data value is input, we have to convert the data to a numerical value for data analysis.

Recommender systems have the capability to recommend products. These products are being able to find similarities between them. The Pearson Correlation Score finds the similarity between the related products. The Correlation Score is a measure by plotting them in the straight line charts. The two sets of data fit on a straight line showing the deviations. The interesting characteristic of the Pearson Score is that it corrects for grade inflation. If one product has consistently higher scores than another, there can still be a perfect correlation i.e. the difference between the ratings is consistent.

The algorithm for the Pearson score is as follows 1) Finding the reviewers that reviewed both products.2) It computes the sums and the squared sums of the ratings for the two products. 3) Then it computes the sum of the reviews of the products. 4) It uses these results to compute the Pearson Correlation Score.

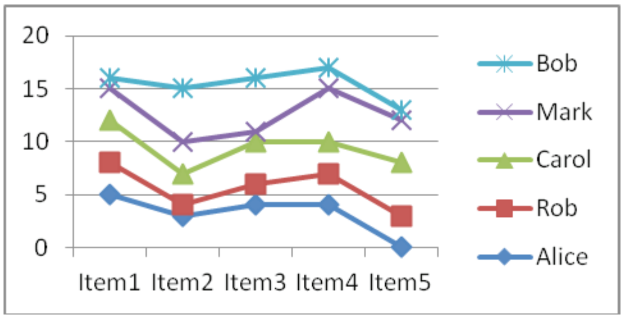
The code written for the algorithm is as follows:

Step 4:List the shared items in a temporary list .For each items have a review in common for users. Also sum up all the preferences.

Step 5: Sum up the squares. Calculate the rates for each user.

The values with respect to users are between 1 and -1 that are returned by the algorithm where 1 means the two users have exactly the same ratings. This algorithm only compares the values for two users, so to calculate scores for entire catalog, the loop through each user will continue until it reaches to all users. The score is calculated as follows:

Step6: list of products that *excludes* the item searching for the products



This is how the Karl Pearson method is applied to form the relation. As based on the given scores, users who were looking at the Alice Item Pack would also be interested in a pair of likes Rob and Carol.

Results and Conclusion

The IR-RS technique is proposed and implemented as domain-specific specifications. This technique is the combination of Ranked-IR and RS method. Many experiments are simulated for searching. This will definitely improve the retrieved results. The test experiments are positive, so this can be implemented in realistic environment. There should be relevance feedback that will be developed further on RS. The scenario we are exploring here considers a searcher exploring a new domain of interest. Thus the researcher has to be exploring the number of items. The use of relevance feedback, will explore the items in a personalized manner. Also, if they are learning about a new topic, there will often be a preferable order in which information should be viewed. The results of the study is to retrieve the relevant items to achieve the maximize efficiency.The emerging evaluation strategies will measure the effectivenessfor relevant items to get the efficient sequence.The RS should be designed based on information search and human interaction using computer based application.

References

Andrew I. S., Alexandrin, P., Lyle H. U., & David, M. P. (2002).Methods and metrics for cold-start recommendations.*In SIGIR '02: Proceedings Of The 25th Annual International ACM SIGIR Conference On Research And Development In Information retrieval*, 253–260, New York, NY, USA, ACM.

Bell, Robert, Yehuda Koren & Chris Volinsky (2009). Matrix factorization techniques for recommender systems. In IEEE Computer, *42*(8), pages 30–37.

Bigus, J.P. (1996).Data Mining with Neural Networks - Solving Business Problems – from Application Development to Decision Support, McGraw-Hill.

Bigus, J.P. (1998). Constructing Intelligent Agents with Java - A Programmer's Guide to Smarter Applications, John Wiley & Sons Inc.,.

Breese, J.S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering.*In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence,* Madison, WI.

Broder, A. (2002).A taxonomy of web search. *SIGIR Forum, 36*(2), 3–10, 2002.Search

Dean, T., Allen, J., & .Aloimonos, Y. (1995). Artificial Intelligence: Theory and Practice, The Benjamin/ Cummings Publishing Co. Inc.,.

Herlocker, J.L., Konstan, J.A., Borchers, A., & Riedl, J.T. (1999).An algorithmic framework for performing collaborative filtering.*In Proceedings of 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval,*230–237, Berkeley, CA, 1999. ACM Press.

Herlocker, J.L., Konstan, J.A., Terveen, L.G., & Riedl, J.T. (2004).Evaluating collaborative filtering recommender systems.*ACM Trans. Inf. Syst., 22*(1), 5–53.

Linden, G., Smith, B., & York, J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing, 7*(1), 76–80, 2003.

Melville, P., Mooney, R.J., & Nagarajan, R. (2002). Contentboosted collaborative filtering for improved recommendations.*In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI- 02),*187–192, Edmonton, Alberta.

Raymond J. M., &Loriene R. (2000). Content-based book recommending using learning for text categorization.*In Proceedings of the Fifth ACM Conference on Digital Libraries*,195–204, San Antonio, TX.

Rennie, J., & Nathan, S. (2005). Fast maximum margin matrix factorization for collaborative prediction. In International Conference on Machine Learning.

Resnick, P., Neophytos, I., Bergstrom, P., Mitesh, S., & Riedl, J. (1994).Grouplens: An open architecture for collaborative filtering of netnews. In CSCW94 - Conference on Computer Supported Cooperative Work, pages 175–186.Chapel Hill, Addison-Wesley.

Schneiderman, B., Byrd, D., & Croft, W. B. (1997). Clarifying search: A user-interface framework for text searches. D-Lib Magazine.Retrieved from http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html.

Xiaoyuan,Su., Russell, G., Taghi, M. K., & Xingquan, Z. (2007).Hybrid collaborative filtering algorithms using a mixture of experts.*In Web Intelligence,*645–649.

Xiaoyuan, Su., & Taghi, M. K. A survey of collaborative filtering techniques.*Advances in Artificial Intelligence*, 1–20.

Xiaoyuan Su, Taghi, M. K., Xingquan, Zhu., & Russell, G. (2008). Imputation-boosted collaborative filtering using machine learning classifiers.*In SAC '08: Proceedings of the 2008 ACM Symposium On Applied Computing*, 949–950, New York, NY, USA, ACM.