Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2002 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2002

RECOMMENDER SYSTEMS: A FRAMEWORK AND RESEARCH ISSUES

Yi Zhang
New Jersey Institute of Technology

Il Im
New Jersey Institute of Technology

RECOMMENDER SYSTEMS: A FRAMEWORK AND RESEARCH ISSUES

Yi Zhang and Il Im
Information Systems Department
College of Computing Sciences
New Jersey Institute of Technology
yxz1847@njit.edu ilim@homer.njit.edu

Abstract

Advances in IT development make it possible for us to access virtually all kinds of information everyday. However this also brings the problem of "information overload". Recommendation system is one of the solutions for information overload and it has been attracting more and more attentions recently. This paper reviews different approaches to recommender systems, mainly focused on collaborative filtering and content filtering approaches, and develops a general framework for recommender systems. We also discuss main research challenges and issues in the field of recommender systems in three areas: algorithm, human computer interaction (HCI), and social impacts.

Keywords: Rrecommender system, collaborative filtering, social filtering, content filtering, information filtering

Introduction

As the Internet serves as a huge storage of information, it is getting increasingly difficult to extract useful information for individual interests. One possible solution for this problem is to employ *recommender* systems. Recommender systems are an example of adaptive filters that use inferences drawn from users' known behavior to recommend documents they have not yet seen (Pemberton 2000). User interests can be drawn from their explicit expression or evaluations or from their implicit behavior. The most commonly used two technologies in building recommender systems are *content filtering/information filtering* and *collaborative/social filtering*.

Information filtering (IF) or Content Filtering is a technique that aims to reduce a person's information overload with respect to his or her interest. It comes from the field of information retrieval. It derives recommendations for a particular user from that user's profile or the knowledge of that user's past behavior (Schafer 1999; Good et al. 1999; Pemberton et. al. 2000). The user profile is based on explicit interests and the past behavior of the user. For example, an IF based system would recommend a book to a user based on the user's expressed interests about books in the profile or based on the user's previous book purchase history. An IF system usually has a dynamic content base, and a static information needs (Good et. al. 1999). The dynamic content base means that the data set from which recommendations are generated are changing from time to time. The static information needs means that the criterion/preference that used to extract recommendations do not change. More advanced IF systems use relevance feedback to update the user profile to get better performance. Relevance feedback is the feedback about whether the user likes the item recommended or not. There are many existing filtering systems based on user profile model, such as WebWatcher (Joachims et al. 1995), Syskill & Webert (Pazzani et al. 1996), and SAMURAI (Leong et al. 1997). The advantage of IF system is that it can work alone without requiring other user's information. However, this advantage also brings some limitations: First, it requires a long time and many item instances to generate an accurate user profile for a given information domain; second, the system will overspecialize as it will necessarily rank documents highly to similar ones seen before; finally, it cannot use other user's experience to help making recommendations. (Vel and Nesbitt 1998).

Collaborative/Social filtering (CF) is a technology that aims to reduce a person's information overload based on other person's preferences. It derives recommendations based on evaluations of other users who share similar interests with the particular user. It is a computerized process of 'Word of Mouth'. (Konstan et al. 1997; Shardanand and Maes 1995). For example, a CF based system would recommend a book to a user because other users who have similar interests rated the book high. As CF systems are based on other users' opinions about the item, they provide a degree of quality of the item based on human's judgment not on the item's attributes, and also because of this characteristic of CF systems, they are generally perceived to be more useful than IF based systems (Herlock et al.1999, Resnick 1997). Providing recommendations based on like-minded people makes CF based system more accurate. At the same time, CF's dependence on human ratings brings two problem: *sparsity* and *first-rater* effects. The sparsity problem is that for a CF system to work well, several users must evaluate each item. The first-rater problem refers to the fact that new items cannot be recommended until some users take the time to evaluate them (Good et al. 1999).

From the above introduction, it is clear that both IF and CF technologies have their advantages and disadvantages. Researchers are trying to find different ways to build more accurate, more acceptable systems over the years. Some of them try to combine these two technologies together to make a better system, some of them try to use filter agents to improve accuracy, others try to find new algorithms for calculating similarities, etc. However, how to address the advantages, and overcome disadvantages to generate more accurate recommendations with less user involvement, are still research challenges in this field. In this paper, we are going to give a framework to summarize and explain fundamental components and accuracy determinants of recommender systems. We will also propose research issues and challenges in recommender systems.

A Framework for Recommender Systems

The main components of a recommender system can be draw as the following diagram:

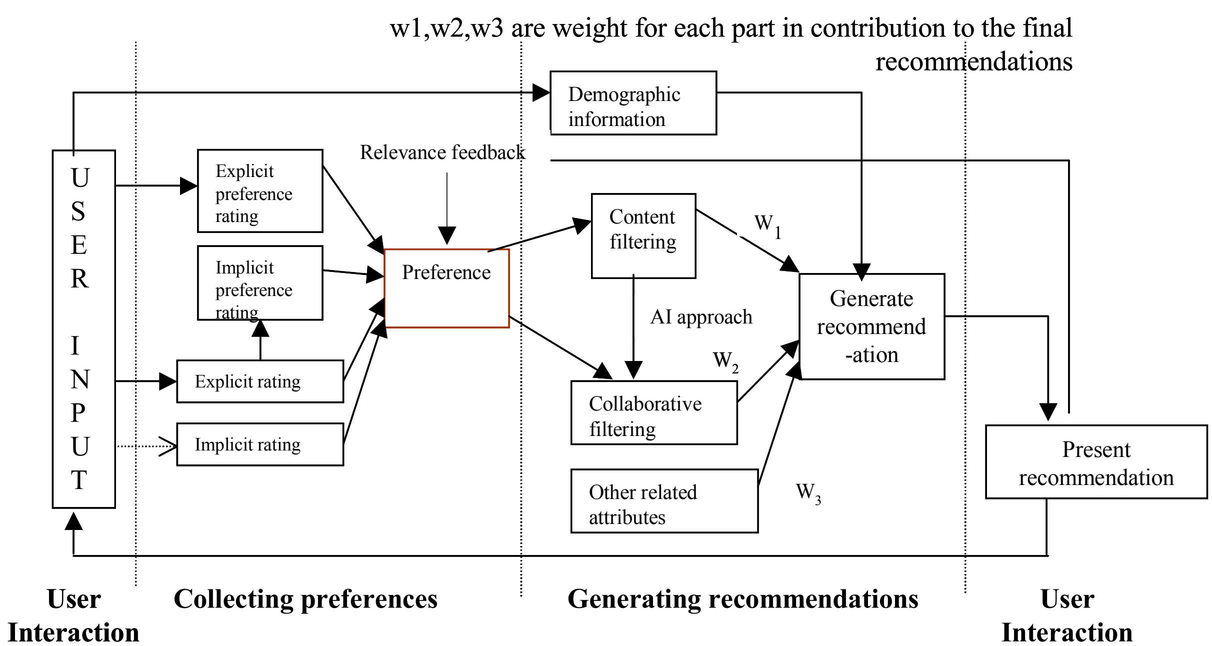

Figure 1. Framework of a Recommender System

From the above framework, we can divide the whole system into three parts: User interaction, collecting preferences and generating recommendations.

User input is the direct and initial source for a system to know the user's preferences. Except for implicit preference, all other sources of preference need direct or indirect user input. Implicit preference is the information that indirectly represents user's preference such as users browsing behavior and user history file.

From the user input and user behavior, four types of information are extracted – explicit preference, implicit preference, explicit rating, and implicit rating. These sources are either combined or directly generate the user preference. The feedback that user provides toward the recommended item (relevance feedback) could also be used to better predict user's preference.

Based on the preferences collected, content filtering will generate a list of recommendations based on matching of user's preference and document's content. Collaborative filtering will first find preference/interest neighbors for each user by calculating the similarities between the ratings provided by users. After finding neighbors of a user, collaborative filtering systems will generate a list of recommendations based on those neighbors' ratings. Besides recommendations provided by content filtering and collaborative filtering method, other attributes, such as item attributes and/or expert judgment, can also contribute to generate recommendations. The recommendations that we mentioned above can be used separately as the final recommendations or they can be integrated into a mathematical model to generate the final recommendations for the user.

After recommendations are generated, they have been presented to users. The method of interaction with users: How to present to users, when to present to users, etc - will all influence the users' perceptions and perspective toward the system.

In the following section, the three parts of the framework are discussed separately. Pervious studies and examples of different systems are reviewed which will be the basis for the discussions on research issues following:

User Interaction: User interaction part includes user input and recommendation presentation. For systems that use explicit preference and explicit rating, user input for preference or ratings are necessary as the input for recommendation generations. This is a very tedious part because users who are using the system usually do not want to spend time and effort entering their interests or ratings on the items they know. The amount of effort for signing up and entering ratings will greatly influence user satisfaction. It will also affect the accuracy of the final recommendations. The other part of user interaction is the presentation of generated recommendations. Systems can display the recommendations with only estimated ratings, like Movielens, GroupLens, or they can display ratings with additional information. Herlocker et al. (2000) compared display with or without the explanations on how recommendations are generated. Their study's result demonstrated that most users valued the explanations and would like to see the explanation features to their Automatic Collaborative Filtering (ACF) system (86% of survey respondents).

There have been several other systems that tried to combine additional information with recommendations. Tapestry system provides annotation together with the messages to show recommendations (Goldberg 1992). Pointer system contains a hypertext links to the source documents as well as contextual information to help recipients determine the interests and relevance of the documents prior to accessing them (Maltz 1995). There are not many studies on user interaction with recommender systems. This area is promising and has a lot of important research issues to be explored in the future, which will be discussed later in the research issues section.

Collecting Preference: Collecting user preference is one of the most important parts for a recommender system. User preference determines both matching of the items in content filtering and matching of similar user groups in collaborative filtering. There are many studies on how to collect preference information. Early systems, such as GroupLens and Fab use explicit rating for preferences; ReferralWeb, PHOAKS and Siteseer use mining technologies to get preference information from public data sources such as Usenet postings or existing bookmark folders (Resnick 1997). RAAP (Research Assistant Agent Project) system asks users to select their research area when they register. This information is used as initial user profile to match the items in the database to give recommendations. This user profile is modified each time the user rejects, accepts, or reviews the recommended items. User preference (profile) changes with user behavior to catch their interest more accurately (Delgado 1999). A system called GroupMark is totally based on implicit information to collect user preference; it does not need users' direct input. GroupMark system is a system to recommend bookmark to users. They use users' existing bookmarks as the interpretation of their preference to give recommendations (Pemberton et al. 2000). Another study was trying to build user profile by collecting user access patterns (Ahmad et al. 1999). They built an autonomous agent to learn users' preferences by analyzing their access pattern to web pages. There are also systems that generate preference based on user's personal history (Terveen 2002). In MOVIES2GO system, voting theory was used to help multiple individuals with conflicting preferences arrive at an acceptable compromise by collecting preferences in multiple dimensions (Mukherjee 2001).

Generate Recommendations: After getting user preferences, these preferences are sent to content filtering or collaborative filtering systems as the input for recommendation generation. Content based IF systems calculate similarities between a user's preferences and document contents. Then they generate recommendations based on these similarities. Collaborative filtering systems first generate a neighbor group for a particular user by calculating similarities of users based on their ratings. Then they generate recommendations based on the rating of the neighbor group.

There are different ways of calculating the similarity, and past studies have explored different algorithms and compared their result (Shardanand 1995; Breese 1998; Herloker et al. 1999). Some researchers are working on the mathematical model for generating recommendations. There are Baysian networks approaches (Breese 1998); dimensionality reduction (Goldberg 2000; Sarwar 2000); clustering techniques (Ungar 1998) and horting technique (Aggarwal 2000). Researchers in this field are trying to explore new mathematical model to calculate similarities in order to generate more accurate recommendations. As both content filtering and collaborative filtering have their own drawbacks when used alone, recent applications are trying to combine these two technologies. Among the systems mentioned above, RAAP, PHOAKS, and GroupMark all combine content filtering and collaborative filtering, and Referral Web is a system combining social networks and collaborative filtering (Kautz 1997). With the development of AI techniques, agent approaches and machine learning are now being broadly used in recommender systems. The GroupLens project implemented agents to help overcome the problems in collaborative filtering. They built several filter bots based on the content of the Usenet messages and combined the results given by filter bots and collaborative filtering to generate final recommendations (Sarwar 1998). Researchers are also trying to apply other theories to help generating recommendations. For example, Decision Theory has been tried in DIVA project. DIVA represents user preferences using pairwise comparisons among items rather than ratings. (Nguyen 1998).

Besides user preference, researchers are trying to take into account other attributes that might influence recommendation results. A previous study has shown that the accuracy of collaborative filtering systems is affected by domain, user characteristics, and purpose of use of the users (Im and Hars 2001). Attributes of recommended items and relationship between person and items have been used to help improve the effectiveness and efficiency of recommender systems (Sarwar 2001). Ansari and others built a Bayesian preference model that allows statistical integration of five types of information useful for making recommendations: a person's expressed preferences, preferences of other consumers, expert evaluations, item characteristics, and individual characteristics (Ansari 2000). This model performed well in generating recommendations.

Research Issues

Based on the framework and discussions in the previous section, we further discuss research issues for further research. Here, we divide research issues into three areas based on their objectives: Algorithm issues, human computer interaction issues, and social impact issues.

Algorithm issues: It is important to find ways to generate more accurate and satisfactory recommendations with less user involvement, and to generate recommendations more efficiently when there is a huge volume of data. More specific research issues in this area include:

- How to combine content filtering techniques, collaborative filtering techniques, and also software agents?
- What are the new attributes that are effective in recommender generation?
- How to refine algorithms for collaborative filtering, content filtering, or software agents?
- How many neighbors are enough for making recommendations? Does it depend on the application domain?
- How to motivate users to provide explicit ratings?
- How to support multi-domain recommendations – how to combine preference data across domains?
- Now, most recommender systems are designed to give recommendations to individual user, there might also be opportunity to study how to generate recommendations for groups of users. For example, people often go to restaurants or movies in groups. Now, the system has to consider how to combine the different tastes of different users in a group and generate recommendations for them all. There are some studies already started to looking at this direction such as PolyLens system (O'Connor 2001).

Human computer interaction (HCI) issues: HCI issues in recommender systems are about how systems interact with users, how it presents the final result to the users. Important issues in this direction include:

- What is the best way and best timing to show the recommendations? When should the recommendations be represented actively (displayed regardless of user's request) and when represented passively (displayed only upon user's request)?
- What content should be included when representing the recommendations?
- How to get accurate inputs from users while requiring minimum efforts from user side?
- How to implement Desktop based systems on other pervasive devices like PDAs, cell-phones, etc.

Social impact issues: The last direction is to look at the social impacts of recommender systems. This direction should be seen from both user side and the recommendation system side.

- What will be the individual benefits of recommender systems bring to users?
- Under what circumstances do users accept recommendations from the system better? We bring up this issue because sometimes user might need recommendation more than other cases and this is when they accept the system better than other cases.
- As the collaborative filtering methods allocate users into different groups according to their preference, people with similar interests are assigned to a same group. This technology provides a way to help people recognize other people who share similar interests. Will this property of the system actually help people form virtual communities? How to implement it?
- Researchers need to think beyond the system itself, to find out the essence of the recommender systems' social impact. How will it help people to help each other (Terveen 2001), how does the system help people to improve their life or their work, what is the social consequences of implementing the system, and will it help people to establish their personal social networks? Research in this area will help us better understand the needs for recommender systems and also give us the information for improving the system design.

Conclusion

Recommender systems are to help people deal with the abundant information they face everyday. In this paper, we developed a general framework for recommender systems and also identified major research issues for recommender systems in three areas:

(1) Combining different technologies, different sources of preference information and different algorithms to generate more accurate recommendations.
(2) Applying different interface designs for better user inputs and recommendation presentations.
(3) The social consequences of implementing recommender systems and how people accept the technology.

Reference

Aggarwal, C., Wolf, J., Wu, K. & Yu, P.S. "Horting hatches an egg: A new graph-theoretic approach to collaborative filtering," In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, 1999, pp.201-212.

Ahmad M. Ahmad Wasfi "Collecting User Accessing Patterns for Building User Profiles and Collaborative Filtering," In proceedings of the International conference on Intelligent User Interface Communications of ACM, 1999. pp 57-64.

Ansari, A. Skander Essegaier and Rajeev Kohli, "Internet Recommendation Systems," Journal of Marketing Research Vol XXXVII, August 2000, pp.363-375

Breese, J.S., David Heckerman and Carl Kadie "Empirical Analysis of Predictive Algorithm for Collaborative Filtering," In proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, July 1998.

De Vel, O., Nesbitt, S. "A Collaborative Filtering Agent System for Dynamic Virtual Communities on the Web," In Proceedings of the Conference on Automated Learning and Discovery CONALD-98, 1998.

Goldberg, D., David Nichols, Brian M. Oki and Douglas Terry, "Using Collaborative Filtering to Weave an Information Tapestry," Communication of ACM, Vol35, Dec. 1992. pp61-70

Goldberg, K., Roeder, T., Gupta, D. & Perkins, C. "Eigentaste: A Constant Time Collaborative Filtering Algorithm," UCB ERL Technical Report M00/41, August 2000.

Good, N., Schafer, J.B., Konstan, J., Borchers, A., Sarwar, B., Herlocker, J., and Riedl, J. "Combining Collaborative Filtering with Personal Agents for Better Recommendations," In Proceedings of the 1999 Conference of the American Association of Artifical Intelligence (AAAI-99). pp 439-446

Herlocker, J., Konstan, J., Borchers, A., Riedl, J. "An Algorithmic Framework for Performing Collaborative Filtering," In Proceedings of the 1999 Conference on Research and Development in Information Retrieval. Aug. 1999.

Herlocker, J., Konstan, J., and Riedl, J. "Explaining Collaborative Filtering Recommendations," In proceedings of ACM 2000 Conference on Computer Supported Cooperative Work , December 2-6, 2000, pp. 241-250.

Im, I. and Hars, A., "Finding information just for you – Knowledge reuse using collaborative filtering systems," Proceedings of International Conference in Information Systems (ICIS), Dec. 2001, New Orleans, LA.

Kautz, H. Bart Selman, and Mehul Shah. "Combining Social Networks and Collaborative Filtering," Communication of ACM. (40:3), March 1997

Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J. GroupLens: Applying Collaborative Filtering to Usenet News. Communications of the ACM (40:3), March 1997, pp.77-87.

Joachims, T., Mitchell, T., Feitag, D., Armstrong, R. "WebWatcher: Machine Learning and Hypertext," Fachgruppentreffen Maschinelles Lernen, Dortmund, Germany, August, 1995.

Leong, H., Kapur, S., de Vel, O. "Text Summarization for Knowledge Filtering Agents in Distributed Heterogeneous Environments," AAAI'97 Spring Symposium on Natural Language Processing for the World Wide Web, Stanford, 1997.

Maltz, D. and Kate Ehrlich "Pointing the way: active collaborative filtering" In proceeding CHI conference 1995, Denver, Colorado, May 7-11, 1995.

Mukherjee, R., Partha Sarathi Dutta, Gerdur Jonsdottir, & Sandip Sen "MOVIES2GO – An Obline Voting Based Movie Recommender System," AGENTS'01, May, 2001.

Nguyen, H. and Peter Haddawy "DIVA: Applying Decision Theory to Collaborative Filteirng," AAAI Workshop on Recommender Systems, AAAI Technical Report, Madison, 1998.

O'Connor, M., Cosley, D., Konstan, J. A., & Riedl, J. "PolyLens: A Recommender System for Groups of Users," In Proceedings of ECSCW 2001, Bonn, Germany, pp. 199-218.

Pazzani, M., Muramatsu, J., Billsus, D. Syskill & Webert "Identifying Interesting Web Sites," available from http://www.ics.uci.edu/~pazzani/Syskill.html, 1996

Pemberton, D., Tom Rodden and Rob Procter, GroupMark: A WWW Recommender System Combining Collaborative and Information Filtering. 6th ERCIM Workshop, Oct. 2000.

Resnick, P. and Varian,, H. "Recommender Systems," Communication of the ACM, (40:3). ACM Press, New York, 1997, pp.56-58.

Sarwar, B.M., Joseph A. Konstan, Al Borchers, Jon Herlocker, Brad Miller, and John Riedl. "Using Filteirng Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System," In proceeding of CSCW98, Seattle, Washington, November 14-18, 1998.

Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl, J. "Application of Dimensionality Reduction in Recommender System -- A Case Study". Full length paper at ACM WebKDD 2000 Web Mining for E-Commerce Workshop, 2000.

Sarwar, B., Geoge Karpis, Joseph Konstan, and John Riedl "Item-Based Collaborative Filtering Recommendation Algorithms," In Proceedings of the 10th International World Wide Web Conference (WWW10), Hong Kong, May 2001.

Schafer, J.B. and Good, N. "Talk to my Agent: Research Issues in Combining Inforamtion Filtering Agents with Collaborative Filtering" CHI'99 Workshop on Interacting with Recommender Systems, May Pittsburgh, 1999.

Shardanand, U., and Maes, P. (1995). "Social Informaiton Filtering: Algorithms for Automating 'Word of Mouth'," In Proceedings of CHI'95. Denver, Colorado, May 7-11, 1995.

Terveen, L.G and Hill, W. "Human-Computer Collaboration in Recommender Systems," in Carroll, J. (ed.), HCI in the New Millennium, Addison Wesley (2001).

Terveen, L., Jessica McMackin, Brian Amento, Will Hill "Specifying Preferences Based On User History," In proceeding of CHI2002 conference, Minneapolis, Minnesota, April 20-25, 2002.

Ungar, L.H. & Foster, D.P. "Clustering Methods for Collaborative Filtering," in AAAI Workshop on Recommendation Systems, Menlo Park, CA, 1998.