A Survey of E-Commerce Recommender Systems

Kangning Wei[1], Jinghua Huang[2], Shaohong Fu[3]

[1 2]School of Economics and Management, Tsinghua University, Beijing, 100084, China
[3]School of Management,Beijing Normal University, Beijing, 100088, China

ABSTRACT

Various personal services in business play important roles in the success of current marketing field. The personalized recommendation technique in recommender systems, one of the most important tools of personal service in websites, makes great significance in Internet marketing activities of e-Commerce. Through summarizing and analyzing personalized recommendation research, this paper presents an overview of personalized recommendation technique and proposes future research topics. The research content of this paper mainly includes the following three aspects, (1) the input of recommender systems, such as the acquisition and presentation of customers' interest profile as well as items profiles; (2) the typical methods of various recommendation techniques; and (3) based on current research and application situations, we finally discuss the future research hot topics and give some suggestions for the research on future recommendation technique.

***Keywords*:** personal services, recommender system, collaborative filtering, content-based filtering

1. INTRODUCTION

Increasing people have declined to purchase interesting items from Internet, but the boom of information relevant to customers, products and transactions, has lead to information overload problem in E-Commerce (Huang, et al 2004). Meanwhile in order to supply customers with various personal services, personalized recommender systems with recommendation techniques have been widely applied, which has already been considered one of the most important methods of personal service in websites. Basically, the personalized recommender systems present different items for customers according to different tastes. Moreover, personal recommender systems could not only reduce the searching time for interesting items, but also enhance e-Commerce sales by converting browsers into buyers, increasing cross-sell and building consumers' loyalty (Scharer, et al 2001).

Although people have researched e-commerce for a long time since the 1990s, when it became an important academic research filed of computer science, the industrial application is not as well as people expect. The important reasons for affecting the acceptance and wide use, include the difficulty of setting up customer's preference portfolio, the recommendation technologies and the persuasion of recommendation results. For these reasons, we present this paper and expect that this paper facilitates our finding better solutions to these problems.
The paper is organized as follows. In section 2, by reviewing the literature of recommender systems, we summarize the input data of recommender systems, such as the acquisition and presentation of customers' interest profile as well as items profiles. In section 3, we demonstrate the typical methods of various recommend techniques. In section 4, based on the research and application situations, we discuss the future research hot topics and give some suggestions for the research on future recommend technique
.

2. INPUT DATA

From the appearance of first e-commerce recommendation system, how to set up customer's preference portfolio exactly has always been the first task which various recommender systems have to face (Resnick, et al, 1994). Learning customers' preference from input data, when customers were browsing webs, is the most popular approach to build customers' portfolio. These recommender systems, such as Movielens, Fab, Entree restaurant recommender and so on, adopt various input data from customers. Thus, we summarize various input data from customers and then classify it into two classes, the customer data, and production data.

The customer data include four types of data, demographic data, rating data, behavior pattern data and transaction data. The customer data in early research period refer to customers' demographic data and rating data to responding products. Such a system as Movielens recommendation system has applied these two types of data. Nevertheless, recently more researchers introduce web mining and other web technologies to catch implicit feedback from customers when they are browsing web, which promotes the application of behavior data and purchase data introduced into organizing customers' portfolio. Behavior pattern data mean the recorders of customer instant behaviors when they are browsing webs. The transaction data indicate the information about customer purchasing items and usually are saved into transaction database. The production data refer mainly to the attributions of products for recommendation. For example, for Movielens recommender system, the products for recommendation are various movies. The productions data mean the topic, directors, actors, release years and so on. While in the Fab recommender system, which recommends news webs and documents, the production data mean the topic of news, key words of news and so on.

Besides the common input data discussed above, some researchers proposed more complex input data indicators. For instance, Shapira (2006) presented some new input

1-4244-0885-7/07/$20.00 ©2007 IEEE.

data indicators, such as mouse movement relative to reading time; scrolling time relative to reading time; reading time normalized by page size, number of links visited on a page; number of links visited on a page relative to the number of existing links on the page; and level of interaction on a page.

We summarized all the input data in the Table 1 as below.

3.TYPICAL METHOD OF RECOMMENDATION TECHIQUE

A recommender system includes usually three steps, which are acquiring preference from customers' input data, computing the recommendation using proper techniques, and finally presenting the recommendation results to customers. In order to evaluate the efficiency and satisfaction of result, one evaluating model of recommender system is also needed. In this process, computing recommendation via proper and various techniques is the most important part, and plays a significant role in the quality of recommendation result. The typical method of recommend technique, according to most popular classified standard, which has been widely accepted by most researchers, is sorted into collaborative filtering, contend-based filtering and hybrid approach.

Table 1 A taxonomy of input data

Data type	Explanation
Demographic data	name, age, gender, profession, birth date, telephone, address, hobbies, salary, education experience and so on.
Rating data	rating scores, such as discrete multi-levels ratings and continuous rating; and latent comments, such as best, good, bad, worse and so on.
Behavior pattern data	duration of browsing, click times, the links of webs; save, print, scroll, delete, open, close, refresh of webs; selection, edition, search, copy, paste, bookmark and even download of web content and so on.
Transaction data	purchasing date, purchase quantity, price, discounting and so on.
Production data	for movies or music, it means actor or singer, topic, release time, price, brand and so on, While for webs or documents, it means content description using key words, the links to others, the viewed times, the topic and so on.

3.1Collaborative filtering approach

As one of the most successful and earliest recommendation technology, collaborative filtering approach works on building a customer dataset from customers and present recommendation by collaborative algorithm. The critical step of collaborative filtering approach lies in searching the similar preference customers with the active customer, that is, find the similar customers. After finding similar customers, it then presents recommendation for active customer according to the preference of similar ones. As a popular reason, we also sort collaborative filtering methods into two types, heuristic-based method and model-based method (Breese, 1998; Adomavicius & Tuzhilin, 2005).

Heuristic-based collaborative filtering usually takes the rating data, duration time, purchasing binary data and others click dream data as input and mainly calculate recommendation result on the entire customer database. The traditional heuristic-based collaborative filtering approach mainly includes checking every customer in the dataset whether he/she is the active customers' neighbors by similarity measures, which includes personal correlation coefficient, cousin metric and even jaccard coefficient for binary data. And then through the KNN formulation the prediction value of the product is computed, which the active custom have not seen or rated but his/her neighbor customers have, and then recommendation is made of the product that has the highest predicted score.

Model-based collaborative filtering method builds a model using various techniques based on the training data, check the model validity by using testing data and then with model compute the production list or the prediction rating value of active customer for no-rating products. The main difference between the heuristic-based and model-based methods are that the heuristic-based approach will use the formulations to compute results on entire database for each customers, while the model-based approach just inputs some data from active customer into the model, and the model will give the prediction value and furthermore give presents.

Because of the disadvantages of collaborative filtering approaches, such as the cold starting problems, sparity problems, gray sheep problems and scalability problems, various techniques have appeared, such as clustering, graph theory, web mining and Support vector machine. According the usual techniques of collaborative approaches, a summary of collaborative filtering approach is table 2, as below.

Table 2 A taxonomy of collaborative filtering approach

Collaborative	Techniques	Typical papers
Heuristic-based method	KNN algorithm and improved one	Resnick,1994; Mohan, et al 2006
	Web mining	Choa, et al, 2004; Cho, et al, 2002
	Decision tree	Cho, et al, 2002
	Graph theory	Aggarwal, 1999
	Support vector machine	Min & Han, 2005
	Bayesian model	Breese, et al, 1998; Chien, et al, 1999
	Clustering	Breese, et al, 1998; Goldberg, et al, 2001
	Association rule mining	Kim, et al, 2004

Model-based method	Artificial neural network	Pazzani, et al, 1997
	Linear-regression	Vucetic, et al, 2005
	Maximum entropy	Pavlov, et al, 2002
	Latent semantic analysis	Hofmann, 2004; Cheung, et al, 2004
	Markov process	Shani, et al, 2002

3.2 Content-based filtering approach

The content-based filtering approach has its origins in information retrieval and information filtering. The item recommended by content-based filtering often indicates textual information, such as news webs and documents. And these items usually describe with keywords and its weights. Nearest neighbor functions or clustering method is used to analyze and cluster the textual feature content of items and recommend suitable content based on items characteristics and the user's preference. The challenge of this approach includes limited content analysis because of limited keywords, overspecialization problems and new user problems.

The techniques usually used in content-based approaches are TF/IDF measure, KNN algorithm, clustering methods, the artificial neural network and association rule mining. Similarly, we divided the content-based filtering approach into two classes, heuristic-based and model-based method. Table 3 shows a summary of content-based filtering approach.

Table 3 A taxonomy of content-based filtering approach

Content-based	Techniques	Typical papers
Heuristic-based Method	KNN	Balabanovic, et al , 1997
	Clustering	Xu, et al, 2005
Model-based Method	Bayesian	Mooney & Roy, 1999
	Clustering	Billsus & Pazzni, 2000
	Artificial neural network	Zhang, et al, 2002

3.3Hybrid filtering approach

To avoid the disadvantages of existing approaches, some researchers have combined these two methods and introduced the hybrid filtering approach.

All the hybrid filtering approaches can also be classified into three methods. Firstly, one class is introducing some component or features of one approach into the other approach. For instance, Melville (2002) introduced content-based predictor into a collaborative approach to compute the additional rating. Secondly, another type is combing the result of recommendation of these two approaches. Claypool (1999) linearly combined the separate rating of collaborative and content-based filtering approach. Finally, the last class of hybrid filtering approach is presenting a comprehensive and unique model depending on other information.

A classification of hybrid filtering approach using various techniques is summarized in table 4 as below.

Table 4 A taxonomy of hybrid filtering approach

Hybrid	Techniques	Typical papers
Feature combining	Bayesian	Mooney & Roy, 1999 Condliff, et al, 1999
	Clustering	Kim, et al, 2006
Recommendation result combining	Linear combination	Claypool, et al, 1999
Comprehensive and unique approach	Probabilistic model	Popescul, et al, 2001
	Maximum entropy	Jin, et al, 2005

4. TOPICS FOR FUTURE RESEARCH

With the appearance of more recommender systems in e-commerce companies, we think some directions discussed as below will become future research topics.

The first one is the construction of consumers profiles. How to capture efficiently and accurately consumers' preference by various data, displayed in Table 1, will be a key point for designing a novel recommender system. Kim (2005) proposed one approach to capture and analyze data from navigational and behavioral patterns of customers. Shapira (2006) analyzed the relationship between implicit indicators and explicit ratings and found that a certain combination of implicit indicators achieved higher correlation with the explicit ratings than any of the individual indicators.

Secondly, in order to overcome some shortages of existing problems, such as the cold-starting, sparsity, scalability problem and so on, another topic is applying new techniques to basic filtering approach and then designing better hybrid approach. Ahn (2006) presented a novel approach to supply automated product recommendation based on the popularity characteristics of products and weaken the influence of data sparsity and cold-starting. Lorraine and Barry (2006) introduced adaptive selection technique into recommender system and dramatically improve the recommendation.

Another topic is the persuasion of recommender system. Since not all the recommendation is suitable for customers' preference, many customers do not give credibility to recommendation result. The persuasion of recommendation result depends on various factors rather than techniques. Ulrike and Daniel (2006) indicated that the relevance, transparency, duration and required effort of the elicitation process are important cues for value of recommendation result, which in turn influences the user's enjoyment of the process and the perceived fit of the recommendation with the user's preferences. Ralitza and Sriram(2006) demonstrated that the marginal value of recommendation depends on the preference structure of the recipient, the attributes of the product on which the recommendation is based, and the characteristics of the population of consumers.

Finally, it is the relationship between customers and recommender systems. The recommender systems could not only present recommendation for customers, but also influence the customers' opinion to recommendation results. Cosley (2003) studied the rating scale and display pattern when customer rate product, to show the influence. How to design a comfortable system interface and make customers enjoy the production-select process could be understood after we identify the factors that affect customers' opinions.

To sum up, some efforts have been made from a computer science prospective, but there are comparably few efforts to study the application value of recommender system from other views which impacts acceptance of the business application by customers. It is the research from management, market, and even psychology views rather than only computer science that promotes business abroad applications of recommender systems in e-commerce.

ACKNOWLEDGEMENT

This work was partly supported by the National Science Foundation of China (70472007, 70231010, and 70621061), and MOE Project of Key Research Institute of Humanity and Social Sciences at Universities (06JJD630014).

REFERENCES

[1] Ahn, H. J., "Utilizing popularity characteristics for product recommendation", International Journal of Electronic Commerce, Vol. 1 No. 2, pp59-80, 2006.

[2] Aggarwal, C.C., J.L. Wolf, et al, "Horting hatches an egg: a new graph-theoretic approach to collaborative filtering", Proceeding of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp20l-2l2, 1999.

[3] Adomavicius, G., A.Tuzhilin, "Toward the next generation of recommender system: a survey of the state-of-the-art and possible extensions", IEEE Transactions on Knowledge and Data Engineering, Vol. 17, NO.6, pp734-749, 2005.

[4] Breese, J., D. Heckerman, et al, "Empirical analysis of predictive algorithms for collaborative filtering", Proceedings of the 14th Conference on Uncertaintly in Artificial Intelligence, pp43-52, 1998.

[5] Balabanovic, M., Y.Shoham, "Fab: content-based, collaborative recommendation", Communications of the ACM, Vol. 40, No. 3, pp66-72, 1997.

[6] Billsus, D., M.Pazzani, "User modeling for adaptive news access," User Modeling and User-Adapted Interaction, Vol.10, No.2-3, pp147-180, 2000.

[7] Cosley, D., S. K Lam, et al, "Is seeing believing? How recommender system interfaces affect users' opinions", Proceeding of the SIGCHI Conference on Human Factors in Computing Systems, pp585-592, 2003.

[8] Condliff, M., D. Lewis, et al, "Bayesian mixed-effects models for recommender systems", Proceeding of ACM SIGIR Workshop Recommender Systems: Algorithms and Evaluation, 1999.

[9] Claypool, M., et al, "Combining content-based and collaborative filters in an online newspaper", Proceeding of ACM Workshop Recommender Systems: Algorithms and Evaluation, 1999.

[10]Cheung, K.W., K.Ch. Tsui, et al, "Extended latent class models for collaborative recommendation", IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol. 34. No. 1, pp. 143-148, 2004.

[11]Choa, Y.H., J.K. Kimb, "Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce", Expert Systems with Applications, Vol. 26, No.2, pp233-246. 2004.

[12]Cho, Y.H., J.K. Kim, et al, "A personalized recommender system based on web usage mining and decision tree induction", Expert Systems with Applications.Vol.23, No. 3, pp329-342, 2002.

[13]Chien, Y.H., E.I. George, "A Bayesian model for collaborative filtering", Proceeding of 7th-Workshop Artificial Intelligence and Statistics, 1999.

[14]Goldberg, K., T. Roeder, et al, "Eigentaste: a constant time collaborative filtering algorithm", Journal of Information Retrieval, Vol.4, No.1, pp133-51, 2001.

[15]Hofmann,T., "Latent semantic models for collaborative filtering", ACM Trans. Information Systems, Vol. 22, No. 1, pp89-115, 2004.

[16]Huang, Z., W. Chung, et al, "A graph model for E-Commerce recommender systems", Journal of the American Society for Information Science and Technology, Vol.55. No.3, pp259-74, 2004.

[17]Jin, X., Y. Zhou, et al, "A maximum entropy web recommendation system: combining collaborative and content features", Proceeding of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp612-17, 2005.

[18]D.H. Kim, V. Atluri, et al, "A click stream-based collaborative filtering personalization model: towards a better performance", Proceedings of the 6th Annual ACM International Workshop on Web Information and Data Management, pp88-95, 2004.

[19]Kim, B.M., Q. Li, et al, "A new approach for combining content-based and collaborative filters" Journal of Intelligent Information Systems.Vol.27, No.1, pp79-91, 2006.

[20]Kim, Y.S., B.J. Yuan, et al, "Development of a recommender system based on navigational & behavioral patterns of customers in e-commerce sites", Expert System With Applications, Vol. 28, No. 2, pp381-93, 2005.

[21]Lorraine, M., S. Barry, "Adaptive selection: an analysis of critiquing and preference-based feedback in conversational recommender systems",

International Journal of Electronic Commerce, Vol. 11, No. 2, pp35-57, 2006.
[22]Mooney, R.J., L. Roy, "Content-based book recommending using learning for text categorization", Proceeding of ACM SIGIR Workshop Recommender Systems: Algorithms and Evaluation, pp195-204, 1999.
[23]Melville, P., R.J. Mooney, et al, "Content-boosted collaborative filtering for improved recommendations", Proceeding of Conference on Artificial Intelligence, pp187-92, Edmonton, 2002.
[24]Min, S.H., I. Han, "Recommender systems using support vector machines", Proceeding of International Conference on Web Engineering, pp387-93, 2005.
[25]Mohan, B.K., J. Keller, et al, "Scouts, promoters, and connectors: the roles of ratings in nearest neighbor collaborative filtering", Proceeding of ACM Conference on Electronic commerce, pp250-59, 2006.
[26]Pavlov, D., D. Pennock, "A maximum entropy approach to collaborative filtering in dynamic, sparse, high-dimensional domains", Proceeding of 16th Ann. Conf. Neural Information Processing Systems, pp1441-48, 2002.
[27]Pazzani, M., D. Billsus, "Learning and revising user profiles: the identification of Interesting web sites," Machine Learning, vol. 27, pp313-31, 1997.
[28]Popescul A, L.H. Ungar, et al, "Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments", Proceeding of 17th Conference on Uncertainty in Artificial Intelligence, pp437-44, 2001.
[29]Ulrike, G., FR. Daniel, "Persuasion in recommender systems", International Journal of Electronic Commerce, Vol. 11 No. 2, pp81-100, 2006.
[30]Ralitza, N., S. Sriram, "The moderating role of consumer and product characteristics on the value of customized on-line recommendations", International Journal of Electronic Commerce, Vol. 11 No. 2, pp101-23, 2006.
[31]Resnick P, Iacovou N, et al, "Grouplens: an open architecture for collaborative filtering of netnews", Proceedings of the ACM Conference on Computer Supported Cooperative Work, pp175-86, 1994.
[32]Scharer, J.B., J.A. Konstan, et al, "E-commerce recommendation applications", Data Mining and Knowledge Discovery, Vol.5, No.1,pp115-53,2001.
[33]Shani G., R. Brafman, et al, "An MDP-based recommender system", Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, pp453-60, 2002.
[34]Shapira, B., M. Taieb-Maimon, et al, "Study of the usefulness of known and new implicit indicators and their optimal combination for accurate inference of users interests", Proceeding of ACM Symposium on Applied Computing, pp1118-19, 2006.
[35]Vucetic, S., et al, "Collaborative filtering using a regression-based approach", Journal of Knowledge and Information Systems, Vol.7 No.1, pp1-22, 2005.
[36]Xu, B., M. Zhang, et al, "Content-based recommendation in E-Commerce", Proceeding of ICCSA, pp946-55, 2005.
[37]Zhang, Y., J. Callan, et al, "Novelty and redundancy detection in adaptive filtering", Proceeding of ACM SIGIR, pp.81-88, 2002.