International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com, Volume 5, Issue 2 (March - April 2017), PP. 95-99

AN APPROACH FOR RECOMMENDATION SYSTEM BASED ON CF

[1] Manoj Chaudhary, [2] Dr Anshu Srivastava, [3] Dr S.Q.Abbas

[1]Research Scholar , [2]Asst Professor ,[3] Professor and Director

[1,2] Shri Venkateshwara University, Gajraula, UP, India.

[3] AIMT Lucknow UP, India,

[1] mchaudhary@birlacorp.com, [2] anshu_qrat@yahoo.co.in, [3] qrat_abbas@yahoo.com

***Abstract* — Personalized online systems for Web search, news recommendation, and e-commerce, are developed. The process of personalization of online systems consists of three main steps: determining a user's needs, classifying products or services, and matching the user's needs with suitable products or services. A multi-feature based method to automatically classify Web pages into categories of topics hierarchically representing the Web pages is proposed. An approach to modeling and quantifying a user's interests and preferences using the user's Web navigational data is presented. The approach is based on the premise that frequently visiting certain types of content or Web sites indicates that the user is interested in related content or retrieving information from those sites. This paper approaches a line in which we intend to design a model for integrating users interest based on Collaborative filtering for a personalized recommender system.**

***Index Terms* — Approach for recommendation, hierarchically representing.**

I. INTRODUCTION

The rapid development of electronics and the World Wide Web (WWW) have significantly impacted our society and daily lives. Currently, the Web is widely used for individuals and organizations in various fields, such as e-banking, education, e-commerce, research, news distribution, entertainment, and communication [1, 2]. Over the past few of years, the amount of online information brought and distributed by online services has been rapidly expanding. Moreover, due to the increasing popularity of the WWW, the acceleration of this information explosion on the Web keeps growing. Accordingly, Web users are requesting more and more retrieval tasks with their increasing familiarity with utilizing the Web. However, the Web is well known as a poorly organized and indexed information repository due to the extreme openness and diversity of the Web's structure and information.

Currently, a combination of using Web search engines and manual Web navigation is one of the most commonly used methods for searching and retrieving contents on the Web [3]. Although search engines have become the most popular tool to retrieve information from the Web, they do not take the initiative in providing information to a user without the process of a query. Most people frequently retrieve information from Web sites they are familiar with, whose contents are updated from time to time; news sites, product information pages, and personal blog sites for instance. It is often tedious and time consuming to repeatedly check each Web site for any new content. Therefore, a Web feed format called Really Simple Syndication (RSS) was designed and used to automatically distribute frequently updated Web files to users [4]. RSS is one of the solutions used to notify subscribers of any new updated content on the Web. By using the RSS based Web content notification and downloading systems, a user can manage the updated files from multiple Web sites in a much simpler way. The resulting files are grouped based on their Web sources and presented in an organized email structure. However, RSS cannot take charge of most individuals' navigation on the Web due to its inherent limitations. Obviously, one limitation of RSS is that not every Web site integrates RSS into its online service. In addition, updated contents are simply organized by their information sources (Web sites) to RSS subscribers, which is non-flexible and unintelligent. Therefore, significant attention has been paid to recommender systems as the other approach to active information distribution [5]. This approach is focused on delivering information of interest to Web users with little manual effort. Accordingly, the main objective of this chapter is to develop a Web content recommender system. A user's long term interest and preference models are constructed based on the user's navigational history and integrated with the recommender system. The similarity between Web content and the user's models is used to determine whether the content will be provided to the user. A user collaboration method is designed to improve the effectiveness of the proposed recommender system.

II. THE PROPOSED RECOMMENDER SYSTEM

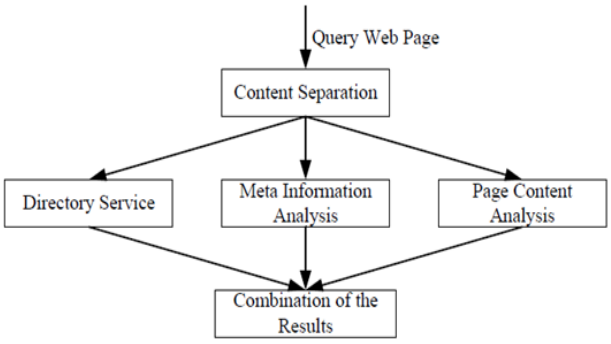
The personalization of online services usually requires the identification of users based on their interests, preferences, navigational data, purchased products, etc. The content based recommendation method is available if user modeling is feasible. Many techniques of modeling a Web user's interests and preferences have been presented [8]. There are two types of approaches to user modeling: explicit and implicit approaches. In an explicit approach, a user is asked to present his or her preferences directly, while in an implicit approach, the system usually builds up a user's models by analyzing the

user's tracked information. A news recommender system can be integrated with either of the two user modelling methods. For example, a news subscription system is a typical recommender system that uses the explicit user modeling method. This work employs the implicit approach to user modeling that is presented in this paper for news recommendation.

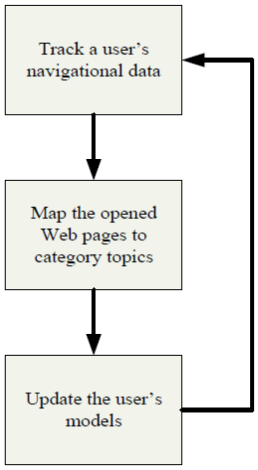
A. User Modelling for the Recommender System

In the proposed recommender system, the construction of a user's interest and preference models is based on analyzing the user's navigational data. In order to identify a Web page browsed by a user, a method of auto-classifying Web pages based on various features (Wen et al., 2008b) is applied. The schematic representation of the classification solution used in the system is illustrated in Figure 2. The classification method includes three steps: content separation, parallel feature classification, and a combination of the results. Initially, a Web page's hyperlink information, meta information, and content information are analyzed for classification, respectively. Then, a fusion algorithm generates the final classification result. In this approach, in order to classify text information such as a Web page's meta and content information, weights of terms in categories are used for computing the cumulative weights of terms in the target text.

In this work, the user interest model indicates the degree of a user's interest in each topic category, while the preference model represents the user's degree of preference for a Web source (Web site) to acquire information. Based on the premise that frequently visiting Web pages belonging to a certain category indicates that the user is interested in that topic, a user's interest model can be constructed by analyzing the user's navigational data. After applying the method of Web page classification, the Web pages browsed by a user can be categorized into different topic categories. Therefore, a quantitative measurement for creating and updating a user's interest model is available by utilizing the Bayes theorem. The process of modeling a user's interests and preferences for the personalized recommender is illustrated in Figure 3



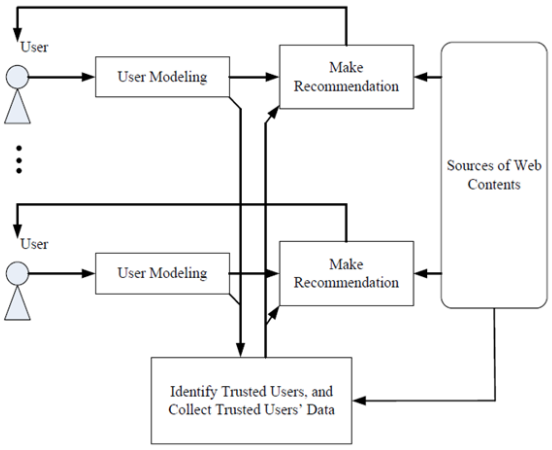
The schematic of the Web page automatic classification method.



The process of user modelling for the recommender system

B. Architecture of the Proposed Recommender System

The architecture of the proposed hybrid recommender system for news recommendation on the Web is shown in Figure 1. In the system, a Web user is distinguished by identifying his or her interest and preference models. A user's navigational data is monitored and analyzed to conduct user modeling. The automatic classification method presented in Chapter 3 is utilized to categorize the Web contents browsed by a user. In the proposed Web page classification method, the ontology base WordNet determines the terms [6], and the weights of terms are calculated by the tf–idf (term frequency–inverse document frequency) method. Three components of a Web page, meta information, effective content area, and the Web address, are extracted and classified separately. The cumulative weights of the terms are utilized for the classification of meta information and effective content areas, while a small number of Web pages on the Internet can be directly classified through a Web directory service. A fusion method is used to combine the classification results for the three components.



The architecture of the proposed recommender system

It is assumed that a user is interested in a certain type of content if the user frequently visits that type of content. The user modeling method presented in Chapter 4 is utilized in the proposed recommender system. It consists of two steps: determining the content of a Web page using the Web page classification method; and utilizing the Naïve Bayes model for updating the user's interest and preference models. In this work, a user's preference model scores a Web site based on the degree to which the user prefers to retrieve information from that Web site.

The recommendation rating of the proposed system can be divided into two steps. First, a content-based algorithm is utilized to determine the probability of recommending Web content to a user, considering the factors of the user's interest and preference models, the Web content, and the time limitation. Second, the method of collaborative filtering is used to modify the probability of recommending Web content. The system will distribute some test Web content, which has been well classified and identified by users. The users who send back positive responses are considered as the trusted users. Additionally, the Web content browsed by more trusted users will obtain higher scores in the recommending process. A preliminary version of the recommender system is reported by Wen[7].

III. Methodology for Recommender system using Collaborative Filtering

This work attempts to recommend Web news content to a user based on their interest and preference models, Web contents, and the behaviours of other users. After a user's interests and preferences have been identified, they are used to determine if certain Web content will be recommended to a user. For an item of Web news, three components of the content need to be quantified: the degree of the Web content belonging to the categories, the source of the Web content, and the time factor. The quantification of mapping the Web content to Web sites is straightforwardly determined by its hyperlink.

Considering that inevitable errors may arise by the current techniques of user modeling and Web page auto-classification, a method of collaborative filtering (CF) is utilized to adjust the recommender system. The process of adjustment based on CF is as follows.

(a) The classified Web pages that are either clustered manually or acquired from well classified Web sites are considered as test pages.

(b) Based on interest models, test Web pages are provided to the corresponding users. In other words, the test Web pages belonging to a certain category topic are delivered to the users who are likely interested in that topic according to the users' models.

(c) Users' feedback on the test Web pages is collected by the system. The users who positively respond to the test pages are regarded as trusted users on the related topic until the next round of testing.

(d) Navigational data from the trusted users is collected by the system. The likelihood of an item belonging to a certain category is increased if trusted users give many positive responses to the item. Moreover, the degree of popularity for news items is calculated based on the click rate of the trusted users.

The degree of popularity for a Web page can be determined by whether or not the users who are interested in a certain topic category recommend the Web page.

IV. Experimental Results and Discussion

In order to evaluate the performance of the proposed Web news recommender system, evaluation experiments and results are presented in this section.

For the Web news recommender system presented in this paper, the weights of terms have been computed by utilizing approximately 7000 Web pages from well-categorized Web news sources, which are used to automatically classify Web pages into twelve topic categories. Test users' interest and preference models are constructed and updated by tracking and analyzing their navigational data. Example interest and preference models of a test are shown in Table 1 and Table 2, respectively. The Web news captured from several news Web sites are provided to test users based on their user models and the content and location of the news.

	Title of Recommended News
1	Buildings iPhone app has a solid foundation for architecture lovers (Yahoo)
2	Djokovic returns for Davis Cup semi-final (Yahoo)
3	France reach Davis Cup final (Yahoo)
4	Chivas USA wins on penalty kicks (Yahoo)
5	Schools use video games as teaching tools (CBC)

Table 1 Example of recommended news for a user at a certain time. 4.3 Evaluation of the Proposed News Recommender System.

In the experiments, news content was collected from Web sources of categorized news, such as CNN, CBC, and Yahoo News. Although news providers have classified the news items obtained from these Web sites, they are re-classified by the Web page classification method in order to consider the scalability issue of news sources including well-categorized and non-categorized content, except when the news items are used as the test information for identifying the trusted users. After the collected news is classified, the recommendation process presented in Section 3.2 is conducted based on a user's interest and preference models. Generally, a recommender system will present a user with only the part of the content that is determined to be relevant to the user. However, in the experiments, all collected news arranged according to the probability of recommendation scoring from high to low are presented to a user, which is for effectively evaluating the

recommender system discussed in Table 1 shows an example of the top 5 recommended news items for a Web user at a certain time.

A. Evaluation of Recommender Systems

In order to evaluate the proposed Web news recommender system, 12 participants were asked to surf the Internet for the purpose of building their interest and preference models. Then the news collected from news Web sites was organized by the probability of recommendation based on the participants' models. Over a period of 15 to 20 days, the 12 participants were asked to use the prototype system for rating the relevance of the recommended news items at least once a day. Since it was not practical for the participants to rate the relevance of all news items by reading through them, the participants were asked to view all the titles of the news items in the list, and then decide whether they were willing to read through an item of news. Therefore, the precision and recall rates can be calculated by analyzing the participants' rating. Based on the average precision and recall rates obtained from the participants' rating, the precision-recall curve of the system is drawn in Figure 4, where the particular pair of precision and recall rates is indicated by a small filled square for each number of recommended items. It shows that the system performs at a good precision rate if the number of recommended items is set to 30 or below. The precision rate drops when the number of recommended items is increased, which also implies that the news content of interest to a user is more likely to appear at the top of the list.

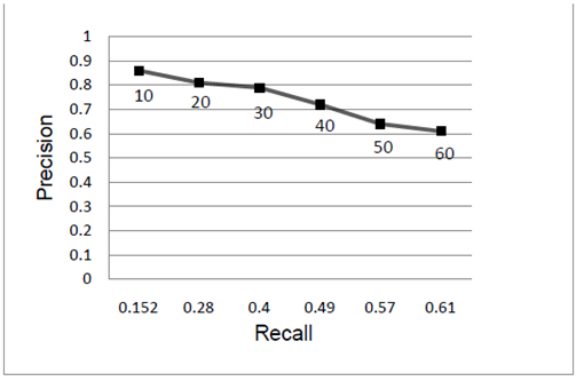


Figure 4 The precision-recall curve of the system obtained through the participants' rating.

In order to evaluate the effectiveness of the proposed CF integrated in the system, the precision rates of the system using two types of recommendation algorithms are calculated for a comparison study. In other words, based on the same rating data from the participants, the recommendation results generated are used to calculate the precision rate of the system, respectively. Figure 5 shows the comparison results of the precision rates with and without the proposed CF method. It can be noticed that the precision rate of the hybrid system is several percentage points higher than the content-based recommender system. Since the proposed CF method is developed for a large-scale recommender system, the performance of the CF method could be fairly evaluated if a large number of users interact with the system.

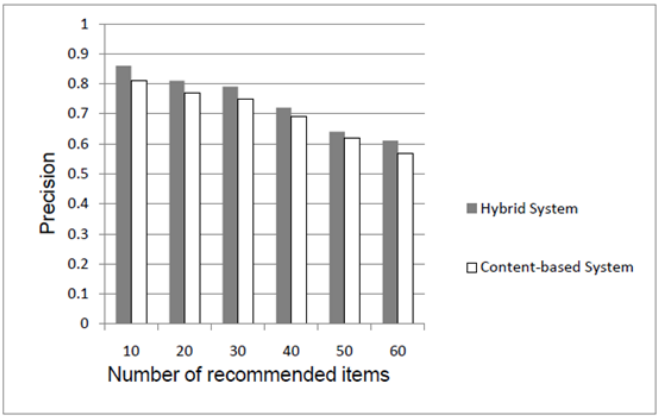


Figure 5 Precision rates with and without the proposed CF method.

Because of intrinsic features of recommender systems, the statistical-based comparison of recommender systems is difficult to apply. Several reasons why quantitative evaluations of different recommender systems are usually incomparable is discussed. First, different recommender systems usually use different data sets. For example, it is not practical that retail recommender systems have identical products. News recommender systems often have different news sources, collection rules, and languages. Advertisement recommender systems have different clients and targets. Second, evaluation of recommender systems is difficult because of the differences in rating properties, such as density and scale. In other words, each recommender system may use a dedicated criteria of rating, which keeps the systems from a fair comparison. Third, the goals for evaluation of recommender systems may differ. Some evaluations concentrate on judging the accuracy of their systems, which emphasizes the capacity of prediction.

However, some researchers argued that user satisfaction should be the eventual measuring criterion for recommender systems. Based on this point of view, commercial systems usually measure user satisfaction by sale numbers, while non-commercial systems make inquiries about users' satisfaction upon using these systems. Some methods of measurement have been proposed to evaluate the performance of Web news recommender systems with various motivations and goals. The number of times that users access recommended news menus is used to evaluate their news recommender system for the mobile Web. Their recommender system provides three types of news menu to users: categories, recommended, and current menus. The read article ratio by menu is used to evaluate the effectiveness of the system.

Instead of using the click rate of menus, [9] use a click through rate (CTR) to evaluate their personalized news article

recommender system. A CTR is the ratio of the number of users who click on an item and the number of times the item is delivered to users. They believe that an effective recommender system is able to improve the overall CTR. Chen et al. [10] consider that the evaluation of the top 10, 20, and 30 recommended news items is more valuable than the average evaluation of overall news items. Therefore, they only provide news listening rates of the top 10, 20, and 30 recommended items to evaluate their phonic Web news recommender system. Bomhardt [11] proposed a Support Vector Machine (SVM) driven personal recommender system for news Web sites, in which recall and precision rates are used to evaluate the overall prediction quality of the system. The SVM-driven recommender system can reach a precision rate up to 61.69% with the consideration of the top 30 recommended news items.

CONCLUSION

A hybrid recommender system for personalized recommendation of Web news to users has been presented. A user's navigational data is tracked and analyzed through a Web page auto-classifying process, which is used to construct and update the user's interest model. The hyperlinks extracted from the user's navigational history are used to build the user's preference model. Web news collected from various news sites is classified by the Web page classification method, and then the probability of recommendation is calculated for a user by matching the contents of news and the user's models. In addition, test information is sent to users in order to identify the trusted users, which is an attempt to improve the performance of the recommender system. The experimental results show the effectiveness of the hybrid recommender system for personalized recommendation of news from the Web for users. A solution of integrating user modelling, architecture, content-based filtering, and CF for a personalized recommender system has been proposed.

REFERENCES

[1] Van Den Heuvel, W.-J. and M.P. Papazoglou (2010). Toward business transactionmanagement in smart service networks. IEEE Internet Computing, vol. 14, no. 4, pp. 71-75.

[2] Von Borstel, F.D. and J.L. Gordillo (2010). Model-based development of virtual laboratories for robotics over the Internet. IEEE Transactions on Systems, Man, and Cybernetics, Part A, vol. 40, no. 3, pp. 623-634.

[3] Trestian, I., Ranjan, S., Kuzmanovic, A., & Nucci, A. (2010). Googling the Internet: Profiling Internet endpoints via the World Wide Web. IEEE/ACM Transactions on Networking, vol. 18, no. 2, pp. 666-679.

[4] Pera, M. S., & Ng, Y. (2008). Utilizing phrase-similarity measures for detecting and clustering informative RSS news articles. Integrated Computer-Aided Engineering, vol. 15, no. 4, pp. 331-350.

[5] Adomavicius, G. and A. Tuzhilin (2005a). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734-749.

[6] Miller, G.A. (2009). WordNet – about us. WordNet, Princeton University, http://wordnet.princeton.edu.

[7] Wen, H., L. Fang, and L. Guan (2010). Classifying customers using navigational history for developing personalized Web services. Proceedings of the 7th International Conference on Service Systems and Service Management, Tokyo, Japan, pp. 1-6.

[8] Tyler, S.K. and J. Teevan (2010). Large scale query log analysis of re-finding. Proceedings of the 3rd ACM International Conference on Web Search and Data Mining, New York, USA, pp. 191-200.

[9] Li, L., W. Chu, J. Langford, and R.E. Schapire (2010). A contextual-bandit approach to personalized news article recommendation. Proceedings of the 19th International Conference on World Wide Web, Raleigh, USA, pp. 661-670.

[10] Chen, W., Zhang, L., Chen, C., & Bu, J. (2009). A hybrid phonic Web news recommender system for pervasive access. Proceedings of the 2009 International Conference on Communications and Mobile Computing, Kunming, China, pp. 122-126.

[11] Bomhardt, C. (2004). NewsRec, A SVM-driven personal recommendation system for news Websites. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Beijing, China, pp. 545-548.