Building a Book Recommender system using time based content filtering

CHHAVI RANA
Department of Computer Science Engineering, University Institute of Engineering and Technology, MD University, Rohtak, Haryana, 124001, INDIA.
chhavi1jan@yahoo.com
SANJAY KUMAR JAIN
Department of Computer Engineering, National Institute of Technology, Kurukshetra, Haryana,136119, INDIA.
skj_nith@yahoo.com

Abstract: - Recommender System are new generation internet tool that help user in navigating through information on the internet and receive information related to their preferences. Although most of the time recommender systems are applied in the area of online shopping and entertainment domains like movie and music, yet their applicability is being researched upon in other area as well. This paper presents an overview of the Recommender Systems which are currently working in the domain of online book shopping. This paper also proposes a new book recommender system that combines user choices with not only similar users but other users as well to give diverse recommendation that change over time. The overall architecture of the proposed system is presented and its implementation with a prototype design is described. Lastly, the paper presents empirical evaluation of the system based on a survey reflecting the impact of such diverse recommendations on the user choices.

Key-Words: - Recommender system; Collaborative filtering; Content filtering; Data mining; Time; Book

1 Introduction

Internet and World Wide Web is changing the way people live and communicate with each other. Together with this, there is lot of information bombardment on the user who wants to access some information on the internet. In the midst of this complex environment of web, Recommender System (RS) serves as an agent that helps user in getting the relevant information. With the growth of economy and advent of new technology, lots of people are using internet as a source of information for making comparative analysis of products that they would like to buy online. As such lot of websites are also offering books as one of the product and some sites are dedicated websites for online book shopping People are increasingly using internet to choose the best deal to spend their money with a widely available list of authors and genres.

In this competitive market, lots of vendors are employing different strategies to attract customers. The customers are bombarded with information on the internet, finding relevant among which is a dotting task. Recommendation systems is one such technology that help user by providing them with relevant information based on their profile. A user profile is generated on the basis of user navigation history and his similarity with other users. Recommender System gives a list of recommendations to the user which is an attempt of predicting user's preferences. A website using a recommendation system can more effectively provide a user with useful and relevant suggestion that could fulfil his current information requirement. As such, these websites have an edge over others in gaining customer loyalty as well as long term partnership.

The term recommendation system was first introduced by Resnick and Varion [13] to generalize the concept of collaborative filtering [7]. They implemented the first recommender system using collaborative filtering technique. Since then the term is being used by different researchers and is explained in various ways. The most common technique used for building recommendation system is collaborative filtering, so much so that both these terms are used interchangeably by many researchers. There are few other techniques also which are used for making recommendations such as content based filtering, demographic and knowledge based technique but they are not so widely applied. Collaborative filtering is more successful applied technique, prominent example of

which is amazon.com. In this technique, the user is given recommendation on the basis of similar user profiles which are calculated through various measures. Mostly Pearson correlation coefficient, Cosine similarity measure and Euclidian distance are taken as widely applied similarity measures.

A number of recommender systems are applied in various domains on the internet and each one of them tries to accurately predict user preference. Also, various researchers studied different facts of recommender system like semantics, scalability etc. Another dimension which is being highlighted lately is the temporal dimension that incorporates changes in the recommendation process over a period of time [8]. It is being observed that with time user preference changes as well as the system content also gets updated [3]. Thus, the recommendation models that predict user preference needs to be updated by refreshing the user profile at regular intervals to present relevant recommendation for users. The effect of temporal dynamics in collaborative filtering [8, 9] is being studied lately, but its effect on content based filtering [2] is yet to be taken up.

In this paper, we have developed a book recommendation system that is based on content based recommendation technique and takes into account the choices of not only similar user but all users to predict new recommendations for the user. This approach combines user choices with the most widely visited links over a period of time and thus gives a set of diverse as well as most widely used preferences. The paper also analyzes the effect of temporal factor on content based recommendation. Though a number of books Recommender System already exist, but none have so far implemented the time factor on content based recommendation. We implemented a system which will use a counter for each item that gets updated with time in relation to other items and combined it with content based recommendations. As such, the paper presents a completely new outlook on the effect of time factor on the recommendation process in a content based environment. Though collaborative filtering is widely used methodology in most of the recommendation systems, content based technique can have leverage where there are some characteristic values that represent item content and this is the case with books.

Our proposed system ensures that recommended items remains relevant to the changing user preferences. In Section 2 literature survey of existing book recommendation system is presented. Section 3 introduces the overall design and implementation of the content based RS. Section 4 describes the evaluation of the proposed system with respect to diversity and temporality. Finally, section 5 provides concluding remarks.

2 Current Recommender System

Now a days Recommendation systems are applied to a wide range of domains, though initially their focus primarily was on the entertainment domain and online shopping sites. Currently there use is being extended to social network sites, e-leaning and travel domain. The focus area of this paper is related to book and as such we present here an overview of the Book Recommender Systems. They help users in managing their reading list by learning their preference. There are two category of such system, one which gives a list of recommendation based on user profile in a library automation system and others which tells user specifically what should he read next according to the current requirements. There are sites like whichbook.net, what should I read next, lazy library, library think etc., each of which uses a particular strategy to fulfil user current requirement. A number of such recommender systems are described hereby.

Whichbook.net allow users to select a particular category, plot and settings for giving consequent recommendations. Thus, the focus of the system is on users current information need. Also this system allows user to select a particular mood like happy, funny, unusual and unpredictable and at the same time allows them to change their search criteria between moods and plot setting. The diversity quotient of such recommendation is narrowed within the unpredictable option and therefore user could find the system repeating similar books with passage of time.

WhatshouldIreadnext.com uses collaborative filtering technique and give recommendations to the user by using an author or title of a book and matching it with others reading list. Such technique generally makes an assumption that similar users have similar preferences [4]. Thus, the users having some books common in their reading list can be given recommendations of others books which are not in their reading list. Such preferences could get stale after a period of time as similar books will be repeatedly given as recommendations.

Lazylibrary.com search books with topic wise listing and have a limit on size of book. Thus the system uses content based filtering approach that chooses books with similar content [1]. The novelty about this system is the restriction on the number of

page of the book which can also be a stumbling block in giving relevant recommendations.

Librarything.com is a completely different system which is best known for cataloguing book into category for each user. It uses a hybrid technique that combines recommendation that are based on library of congress categories and what other similar uses have in their reading lists. It also includes links to a number of books stores as well [11,14, 5]. The major drawback is that it assumes that users have already looked at all the books by same author and thus select something completely different which could be irrelevant for the user.

Abebook Corporation which own the world largest online book store have also developed and launched a new online book recommender system called Bookhints.com based upon title owned, read and enjoyed by members of Librarything.com, the online cataloguing site. It uses the database of library thing and generates recommendations based upon titles found on the catalogues of similar users who also own the book originally sought. "It is the first time that unique information added by an online community has been directly applied to retailing", said Boris Watz, COO of Abebook.

Another book recommended system Booklamp.com uses a novel technique to predict user preference. It matches readers to a book by analyzing the writing style, perspective and description of the whole book together and thus allows users to find books with similar tones, action level and dialog level. It removes outside influences such as advertising, author marketing or social networking. As the system has grown, it has also incorporated human feedback and mixed it with computer analysis. It allows user to match books with far greater details than any other such system.

Goodsreads.com, a social network for book lovers was started in 2006. Initially it does not have a recommendation system embedded in it, but lately it acquired discoverread.com which has a strong recommendation algorithm. Discoverread.com analyzes book trends and give recommend on the bases of user rating by matching it with friends and similar users [13]. The major drawback with the system is that as the system becomes popular, more and more irrelevant ratings are added which could harm the accuracy of recommendations.

Bookexplorer.com is another recommendation engine in which the books are arranged category wise for user and there is also a search option for users. Once user selects a book from a category or from the search option, lists of recommended books are also presented to the user. The recommendation system is fairly simple as it implements content based filtering within the similar category of books to present recommendation. The drawback of the system lies in its ability to handle diverse recommendation.

3 Proposed System

The book recommender system implemented till now utilizes either similar user interest or item characteristic for matching it with user profile for giving future recommendation. Our recommender system adds another important dimension to this which is the temporal dimension. This dimension takes into account the number of time an item is liked by the user over a period of time. A counter is stored for each item which is updated whenever a user checks that items in its favourite links. The greater the counter for any particular book, the more are the chances of it being recommended. Moreover instead of using counter rating of similar user, our algorithm uses counter ratings of all users that improves diversity of the recommendation process. Over a period of time, the counter for different items will change in relation to other items which could either become more favourite or less favourite in comparison to each other. Thus, the overall system reflects an updated favourite list with respect to a particular category at any given point of time. User preferences are incorporated through content based filtering where only a particular category of book will be recommended. The user will be given a search option as well a category listing to choose a book and once it chooses a category or book title, recommendation will be enlisted. The timely updation of recommendation is a very important aspect as user need as well as the system content keeps changing with time [3]. Thus, there should be a mechanism to precisely incorporate these changes into the recommendation process. This will give user more relevant, diverse as well as updated information which establishes the effectiveness of Recommender System.

3.1 System Overview

The proposed approach is very simple and it effectively improve the recommendation process by incorporating diverse as well as updated recommendations by maintaining a simple counter for each item. The recommendation process is divided into four categories:

(a) Gathering of useful web data
(b) Pre-processing of the collected data
(c) Interpretation of the web data
(d) Decision making or Final Recommendation.

When user sends request for a particular book or author in the form of a query to the Search Engine, the Search Engine return the response by giving URL's of all the matching links depending upon the algorithm used by the Search Engine. The proposed system uses content based filtering algorithm over a categories of books to search similar books. Moreover, books which are listed as favourites by maximum number of users are given as recommendations. When user enter query for its particular search request, the Query is first processed to filter out irrelevant data and then analyzed to discover interesting navigation patterns and relation between books category and counter value of items in that category. A response repository is formed which serves as a knowledge base. After building Knowledge Base, system can give recommendations based on the navigation and most visited links by particular user in the featured domain. The Personalized data is then given as a input to the ranking algorithm for personalizing the result according to the most visited links or favourite links of the particular user by evaluating web mining results. Thus, the System helps the user to retrieve his/her interested domain by clicking on the list of his/her favourite links with more convenience. The current recommendation is based on the recent most last selected item by the user. The content are filtered according to this category and the items with highest number of favourite checks at this point of time are given as recommendations. Fig 1 shows the overall architecture of the proposed book recommender system.

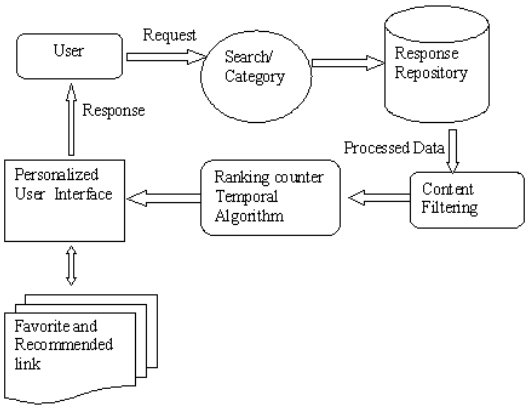

Fig. 1. Architecture of Book Recommender System

The proposed system tries to give diverse recommendations to the uses which are also relevant to his current information needs. In the area of certain domain like books, e-learning and Academic research content based filtering can be more useful than collaborative filtering [10]. This is due to the fact that such domains offer lots of features that can be easily used for matching users and items. As such we have focused on content based filtering and added a feature that could combine updated as well as diverse information into the system with passage of time. There is no need to communicate with other users in such systems [2]. The major strength of content filtering is that even if the system has fewer ratings or none at all which usually is in case of most of the new users, recommendations can be made. The only requirement is to have some information about each item which is easily available in case of book recommender system.

3.2 Process flow of the System

The personalized book recommender system consists of the following workflow -:

1) Firstly user casually uses the system without registration and collects raw information i.e. information which is available to the every visitor's of the website. After analyzing this information, user login to the system and generates its profile which contain user personal details and most visited link recommended by the system based on the URL ranking algorithm (collectively most visited favourite links of all users).
2) After the registration is completed and a set of general recommendation with most visited links provided, the system proceeds on to next level. Based on the user navigation pattern, system analyzes user behaviour and predict its interest by performing content filtering on the basis of this analysis. As a result, system provide favourite link to the user based on its profile via personalized web page and simultaneously update already exiting links on the basis of periodical temporal update.
3) In the last step, the system again perform analysis on the previous result and deliver top n recommendations of that category of books in which user is most interested in by filtering the contents available on the system by performing content filtering.

The result of the overall process workflow is that system deliver web page to user which contain its personal detail, most visited link of the site, user's favourite link and contents recommended by the system of his/her interested domain in single customized format.

This figure below depicts overall process workflow for book recommender system.

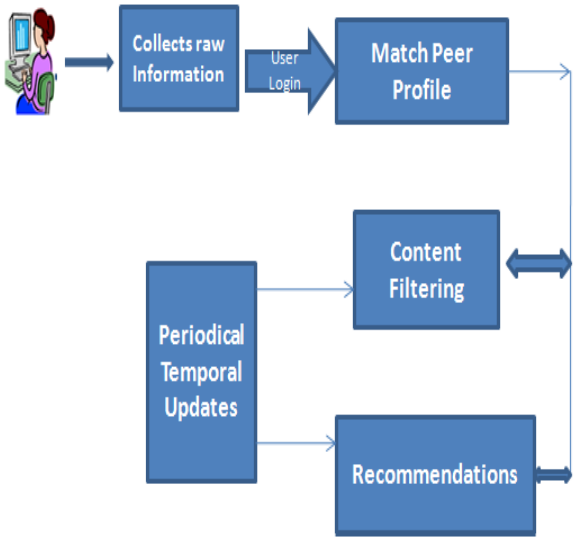

Fig. 2. Process workflow for Book Recommender System

3.3 Personalized Recommendation

The process of recommendation can enhance the overall user interface of a particular website. The user interface can further be enhanced if the recommendations are given in a particular manner that helps user in further navigation. In case of books, most of recommender system provides a mechanism through which user is presented with further links to buy recommended books from online stores. Other recommender system gives an option of borrowing such books from library which could be limited to particular country or city, as is case with librarything.com. The emphasis suggest that calculating recommendation list is as important as presenting them in certain manner. Finally some feedback mechanism could also be incorporated at the end of the recommendation list that will take suggestion from users and thereby improve the recommendation process. This will act as an unobtrusive way of getting suggestion from the users at the end of the whole process.

4 Research Design

The performance analysis of this system is done using an online survey to explore the effect of temporal updation on user satisfaction. The study also analyses how user reacts to a set of diverse recommendations. There was no identifying private data which was sought. There were 10 questions out of which five focuses on demographic data (gender and age) so that results could be categorized. The other questions were specifically focusing on recommendation relevancy to the user, accuracy with temporal changes and diversity.

4.1 Implementation

The system was implemented using ASP.net as front end and SQL server as backend on an Intel Core2duo processor with 2GB ram and 300GB hard disk.

4.2 Participants

Participants were college students who were already using some book recommendation system and are requested to use the current system. They volunteered to take the survey by clicking on the link on the website. Around 150 respondents participated in the survey among which 50 girls and 100 boys were there. Each participant belongs to age group 18-25 years.

4.3 Results

The results were inferred by using descriptive statistics to generate percentages and further analysis of the data gathered from the survey through qualitative methods to discover major themes [6].

4.3.1 Diversity Impact

In the first experiment shown in Figure 3, it is the evaluation of the recommendation preferences with respect to diversity. The users are asked to give a score to both types of recommendations. The higher score indicates the more useful the recommendations are. It can be seen that most of the recommendation that comes under diversity were more liked by the users. Participants were asked to explain their preferences regarding diverse recommendation and content based similar recommendation, and from the data five participant preference categories were developed: Diverse, Similar, both, neither, and no preference. Of the 74 participants that responded to the survey, the majority of the members, 54% (40) preferred diverse recommendations while only 22% (16) preferred similar recommendations. The other 24% (18) of the participants preferred neither, both or had no preference.

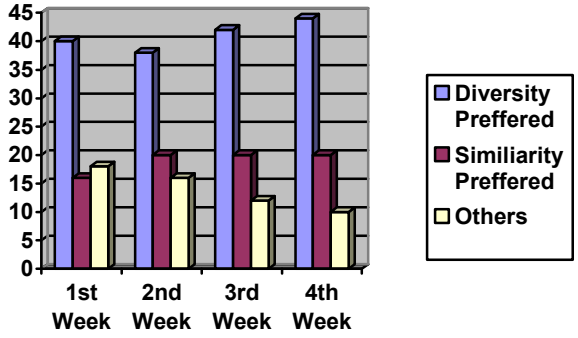

Fig. 3. Diversity Effect

4.3.2 Temporal Impact

In the second experiment, preferences of users over a period of time were analyzed. The users were asked to rate the recommendation after using the system for a period of time on a daily basis and then after a week they were asked to give score about the nature of recommendation. Participants were asked specific question about whether the recommendation were changing with time and were those changes helpful or not.

Users were also asked to rate the quality of recommendation on a scale of 1 to 10. Figure 4 shows that only 20 (26%) participants did not notice any changes in their recommendations while 54 (74%) participants see a noticeable change in their recommendations daily, weekly or periodically. Most of the participants 37 (70%) among the 54 who noticed the changes, find those changes useful and give a score 8 and above to the new recommendations which were changing with time. The above result shows that the newly developed RS adapt to the user changing requirement in a very effective manner.

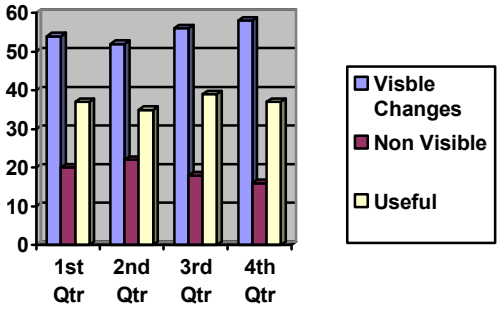

Fig. 4. Temporal Effect

4 Conclusion

In this paper, we propose a book recommender system that uses content based approach for recommending item to a particular uses. In this approach we embedded a new dimension called temporal dimension using a counter for each item which gets update with passage of time and thereby improving the whole recommendation process.

The major goal of the current system is to provide diverse recommendation to the user even if there are fewer ratings about an item for a particular user. The process also get updated with passage of time which makes different item less or more relevant according to their counter value which gets changes with time in relation to other items. The literature survey listed a number of recommender system but none takes into account this criteria for suggesting diverse as well as updated recommendation. After a certain amount of time, the accuracy of the recommendation process gets stuck and users keep on receiving similar items. Further accuracy can only be increased by taking into account temporal dimension which is what this paper addresses. Result shows that proposed architecture for the book recommendation engine provide diverse and temporally updated recommendation to the users which are more useful and relevant.

The current system applies this approach in the domain of book recommender system which could further be extended to other domains as well. Further the system can be extended in other dimension that includes incorporating contextual information, taking multiple ratings and providing a more flexible recommendation that could range into different domains as well. The performance of the system and its comparison with the other similar systems is scheduled for future ongoing work.

References:

[1] Baudisch, P., and Terveen, L. Interacting with recommender systems. *CHI 99 extended abstracts on Human factors in computing systems CHI 99*, 1999, pp. 164.

[2] Chen, T., and Han, W. Content recommendation system based on private dynamic user profile. *Machine Learning*, Vol. 4, 2007, pp. 2112-2118

[3] Chu, W. and Park, S. T. Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models. *18th International*

WWW Conference, Madrid, Spain, 2009, pp. 691-706.

[4] Davies, J. E. What shall I read next? Developing tools for reader support. *Proceedings of IFLA General Conference and Council*, Vol. 68, 2002.

[5] DeZelar-Tiedman, C. Doing the LibraryThing in an academic library catalog. *Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications*, 2008, pp. 211–211.

[6] Glaser, B.G. and Strauss, A.L. The constant comparative method of qualitative analysis, *In The discovery of grounded theory: Strategies for qualitative research.* 1967, Aldine de Gruyter: Hawthorne, New York. P. 101-115.

[7] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. Using collaborative filtering to weave an information tapestry. *Communication of the. ACM,* Vol. 35, No. 12, 1992, pp. 61-70.

[8] Koren, Y. Collaborative filtering with temporal dynamics. *Proceedings of the 15th ACM International Conference on Knowledge Discovery and Data Mining,* Vol. 53, No. 4, 2009, pp. 447.

[9] Lathia, N., Hailes, S., Capra, L., and Amatriain, X. Temporal Diversity in Recommender Systems. *Changes*, 2010, pp. 210-217.

[10] Garden, M., and Dudek, G. Mixed Collaborative and Content-Based Filtering with User-Contributed Semantic Features. *AAAI Conference On Artificial Intelligence*, Vol. 21, No. 2, 1999, pp.1307.

[11] Naughton, R., and Lin, X. Recommender Systems: Investigating the Impact of Recommendations on User Choices and Behaviors. *ACM RecSys UCERST*, Barcelona, Spain, 2010, pp. 9-13.

[12] OReilly, T. Goodreads vs Twitter: The Benefits of Asymmetric Follow. *OReilly Radar, 2009*.

[13] Resnick, P. and Varian, H.R. Recommender Systems. *Communications of the ACM*, Vol. 40, No. 3, 1997, pp. 56–58.

[14] Thomas, M., Caudle, D. M., and Schmitz, C. Trashy tags: problematic tags in LibraryThing. *New Library World*, Vol. 111, pp. 5/6, 2010, pp. 223-235