International Journal for Modern Trends in Science and Technology
Volume: 03, Issue No: 07, July 2017
ISSN: 2455-3778
http://www.ijmtst.com

Web-Page Recommendation in Data Retrieval Exploitation Domain Information and Web Usage Mining

Kaviya J[1] | Meyyappan T[2]

[1]M.Phil., Scholar in Department of Computer Applications, Alagappa University, Karaikudi, Tamilnadu, India.
[2]Professor, Department of Computer Science, Alagappa University, Karaikudi, Tamilnadu, India.

To Cite this Article
Kaviya J and Meyyappan T, "Web-Page Recommendation in Data Retrieval Exploitation Domain Information and Web Usage Mining", *International Journal for Modern Trends in Science and Technology*, Vol. 03, Issue 07, July 2017, pp. 85-89.

ABSTRACT

Web-page recommendation plays an important role in intelligent internet systems. Helpful knowledge discovery from internet usage data and satisfactory knowledge illustration for effective Web-page recommendations unit crucial and challenging. This paper proposes a singular technique to expeditiously give higher Web-page recommendation through semantic-enhancement by human action the domain and Web usage knowledge of an online website. Two new models are proposed to represent the domain knowledge. The primary model uses ontology to represent the domain information. The second model uses one automatically generated linguistics network to represent domain terms, Web-pages, and the relations between them. Another new model, the abstract prediction model, is proposed to automatically generate a semantic network of the semantic Web usage knowledge, which is the mixing of domain knowledge and internet usage knowledge. A number of effective queries are developed to questing relating to these knowledge bases. Based on these queries, a set of recommendation strategies have been proposed to generate Web-page candidates. The recommendation results are compared with the results obtained from an advanced existing Web Usage Mining (WUM) method. The experimental results demonstrate that the proposed method produces significantly higher performance than the WUM method.

KEYWORDS: *semantic network, webpage recommendation, Domain knowledge knowledge representation, Web Usage Mining.*

I. INTRODUCTION

Web usage mining is one of the frequent usage areas of web mining. The awareness of Web mining lies in analyzing user's behavior on the web after exploring access logs and its popularity is increasing at a faster face especially in E-services areas. The applications in these web semantic search areas added its approval and made it as an inevitable part in computer and information sciences. Details like user log files demand for resources and maintain web servers, which is the core mining area of web usage. The semantic analysis gives the user browsing patterns utilized for target advertisement, development of web design, and fulfillment of users and making market analysis.

In the current age of information overload, it is becoming increasingly harder to find relevant content. This problem is not only widespread but also alarming. Over the last 10- 15 years, recommender systems technologies have been introduced to help people deal with these vast amounts of information, and they have been widely used in research as well as e-commerce applications, such as the ones used by Amazon and Netflix. The most common formulation of the recommendation problem relies on the notion of ratings, i.e., recommender systems estimate ratings of items (or products) that are yet to be consumed by users, based on the ratings of items already consumed. Recommender systems typically try to predict the ratings of unknown items for each user, often using other users' ratings, and recommend top *N* items with the highest predicted ratings. Accordingly, there have been many studies on developing new algorithms that can improve the predictive accuracy of recommendations. However, the quality of recommendations can be evaluated along a number of dimensions, and relying on the accuracy of recommendations alone may not be enough to find the most relevant items for each user.

In particular, the importance of *diverse* recommendations has been previously emphasized in several studies. These studies argue that one of the goals of recommender systems is to provide a user with highly idiosyncratic or personalized items, and more diverse recommendations result in more opportunities for users to get recommended such items. With this motivation, some studies proposed new recommendation methods that can increase the diversity of recommendation sets for a given *individual* user, often measured by an average dissimilarity between all pairs of recommended items, while maintaining an acceptable level of accuracy. These studies measure recommendation diversity from an individual user's perspective (i.e., *individual diversity*).

In contrast to individual diversity, which has been explored in a number of papers, some recent studies [10], [14] started examining the impact of recommender systems on sales diversity by considering *aggregate diversity* of recommendations across all users. Note that high individual diversity of recommendations does not necessarily imply high aggregate diversity. For example, if the system recommends to all users the same five best-selling items that are not similar to each other, the recommendation list for each user is diverse (i.e., high individual diversity), but only five distinct items are recommended to all users and purchased by them (i.e., resulting in low aggregate diversity or high sales concentration).

Different from the majority of the existing web recommendation techniques, we propose an intelligent web recommendation system that uses a term pattern frequency mining technique. Which is very suitable for predicting the next web pages? Different evaluation measures including time constraint, precision, satisfaction and applicability are proposed to measure the performance of the recommendation system.

II. RELATED WORK

Recommender systems are usually classified into three categories based on their approach to recommendation such as content-based, collaborative and hybrid approaches i.e. content-based recommender systems recommend items similar to the ones the user preferred in the past. Collaborative filtering recommender systems recommend items that user who have similar preferences (i.e., "neighbors") or liked in the past. Finally, hybrid approaches can combine content-based and collaborative methods in several diverse ways [7][8][9].

This family of algorithms is widely used in recommender systems which deal with collaborative filtering. Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behavior, activity or preferences and predicting what users would like based on their similarity to other users. One of the most common types of collaborative filtering is item-to-item collaborative filtering (people who buy x also buy y), an algorithm popularized by Amazon.com recommender system. User based collaborative filtering attempts to model the social process of asking a friend for a recommendation [10].

Content based filtering methods are based on the information about the items that are going to be recommended. In other words, these algorithms try to recommend the items similar to those that a user liked in the history In particular, various candidate items are compared with items earlier rated by the user and the best matched items are recommended. This approach has its roots in and information filtering research. Basically those methods utilize an item profile i.e. a set of attributes (features) characterizing the item within the system. The system creates a content based profile of users based on a weighted vector of item features. The weights specify the significance of each feature to the user and could be computed from individually rated content vectors using a variety of techniques. Simple approaches use the average values of the rated item vector while other sophisticated methods utilize Bayesian Classifiers (and other machine learning techniques, including clustering, decision trees, and artificial neural networks) in order to guess the probability that the user is going to like the item[11][12][13].

Recent research has demonstrated that a hybrid approach, combining collaborative filtering and content-based filtering could be more efficient in some cases. Hybrid approaches can be implemented in several ways such as by making content-based and collaborative-based predictions discretely and then combine them. By adding content-based capabilities to a collaborative-based approach (and vice versa); or by merging the approaches into one model. Numerous studies empirically compare the performance of the hybrid with the pure collaborative and content-based methods and demonstrate that the hybrid methods can provide more exact recommendations than pure approaches. Such methods can also be used to conquer some of the common problems in recommender systems such as cold start and the scarcity problem[14].

Numerous recommendation techniques have been developed over the last few years, and various metrics have been employed for measuring the correctness of recommendations, including statistical accuracy metrics and decision support measures.

III. PROPOSEDAPPROACH: ITEM POPULARITY-BASED RANKING

Input: -Number of Visitors and their Feedbacks.

Output:-Average of overall ratings.

1.Consider the threshold value T_H (T Scale [T_H,T_{max}] prediction T_{max}=5).

2.Choose the level of recommendation accuracy of users

3.Calculating the Ranking threshold T_R with respect to the threshold value T_H

$Rank_x(i)$, if $R^*(u, i)\epsilon[T_R, T_{max}]$

$Rank(i, T_R) = \alpha_u + rank_{standard}(i)$,

if $R^*(u, i)\epsilon[T_H, T_R]$

where$I_u^*(T_R) = \{i\epsilon I \mid R^*(u, i) \geq T_R\}$.

$\alpha_u = maxrank_x(i)$

4.Identify items above T_R get ranked and increase ranking threshold $T_R \varepsilon [T_H, T_{max}]$. It defines more accuracy and less diversity.

5.Choose T_R value between extreme limits which allow users to set the balance between accuracy and diversity.

The above algorithm is implemented for websites. In step1 we choose threshold value 'T_H' to Google as '3.5'. In step 2 the level of recommendation of accuracy is taken as 360 from Figure 3. In step 3 is meant for computing the rank threshold value 'T_R' is '3.8'. The value is set between the extreme limits to balance the accuracy and diversity. We tend to choose item popularity based ranking algorithm to solve this problem, thus the threshold chosen value (T_H) is '0' and the max value (T_{max}) is '5'.

Item popularity-based ranking approach ranks items directly based on their popularity, from lowest to highest, where popularity is represented by the number of known ratings that each item possess. More formally, item popularity-based ranking function can be written as follows:

$rankItemPop(i) = |U(i)|$,where $U(i) = \{|u \in U \mid \exists R(u,i)\}$.

There exist multiple variations of neighborhood-based CF techniques. In this paper, to estimate $R^*(u, i)$, i.e., the highest predict rating R^* that user "u" would give to an item "i", first calculate the similarity among user "u" and other users "u'" using a cosine similarity metric. Where $I(u, u')$ denotes the set of all items rated by both user "u"and user "u'". Based on the similarity calculation, set $N(u)$ of adjacent neighbors' of user "u" is obtained. The size of set $N(u)$ can range from 1 to $|U|$-1, i.e., all other users in the dataset.

Then, $R^*(u, i)$ is computed as the adjusted weighted sum of all known ratings $R(u', i)$. Here $R(u)$ denotes the average rating of user "u". A neighborhood based CF method can be user-based or item-based, basing on whether the similarity is calculated between users or items, the user-based approach and they can be also straightforwardly rewritten for the item-based approach because of the symmetry between users and items in all neighborhoods-based.

CF calculations[17][18].

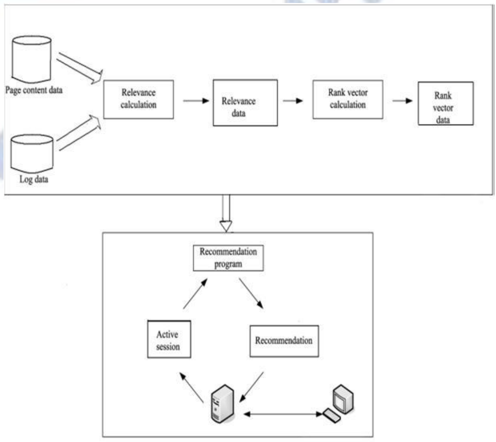

Fig. 1 Architecture of the recommendation diversity

3.1 *General Steps for Recommendation Re-ranking*

The item popularity-based ranking approach described above is just one example of possible ranking approaches for improving recommendation diversity, and a number of additional ranking functions, $rank_X(i)$. The summarize the general ideas behind the proposed ranking approaches, as illustrated by Fig. 2.

The first step, shown in Fig. 2a, represents the standard approach, which, for each user, ranks all the predicted items according to the predicted rating value and selects top-N candidate items, as long as they are above the highly-predicted rating threshold T_H. The recommendation quality of the overall recommendation technique is measured in terms of the precision-in-top-N and the diversity-in-top-N, as shown in the accuracy-diversity plot at the right side of the example (a).

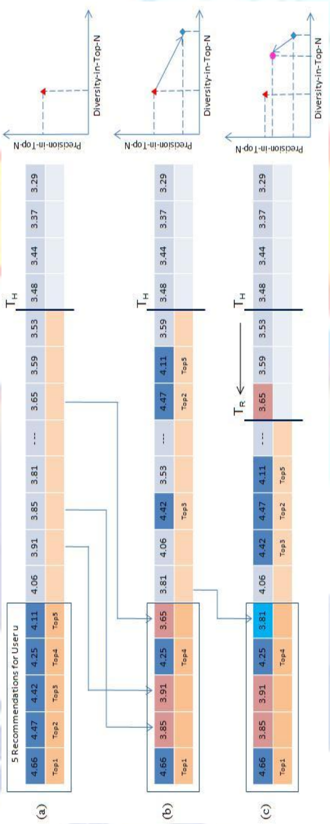

Fig. 2. General overview of ranking-based approaches for improving recommendation diversity

(a) Recommending top-N highly predicted items for user u, according to standard ranking approach

(b) Recommending top-N items, according to some other ranking approach for better diversity

(c) Confining re-ranked recommendations to the items above new ranking threshold T_R (e.g., ≥ 3.8) for better accuracy

The second step, illustrated in Fig. 2b, shows the recommendations provided by applying one of the proposed ranking functions, $rank_X(i)$, where several different items (that are not necessarily among N most highly predicted, but are still above T_H) are recommended to the user. This way, a user can get recommended more idiosyncratic, long-tail, less frequently recommended items that may not be as widely popular, but can still be very relevant to this user (as indicated by relatively high predicted rating). Therefore, re-ranking the candidate items can significantly improve the recommendation diversity although, this typically comes at some loss of recommendation accuracy. The performance graph of the second step (b) demonstrates this accuracy-diversity tradeoff.

The third step, shown in Fig. 2c, can significantly minimize accuracy loss by confining the re-ranked recommendations to the items above newly introduced ranking threshold T_R (e.g., 3.8 out of 5). In this particular illustration, note that the increased ranking threshold makes the fifth recommended item in the second step (b) (i.e., item with predicted rating value of 3.65) filtered out and the next possible item above the new ranking threshold (i.e. the item predicted as 3.81) is recommended to user u. Averaged across all users, this parameterization helps to make the level of accuracy loss fairly small with still a significant diversity gain (as compared to the standard ranking approach), as shown in the performance graph of the third step (c).

Now introduce several additional item ranking functions, and provide empirical evidence that supports our motivation of using these item criteria for diversity improvement

IV. RESULT

The comparison of products with different ranking techniques. In this table, item popularity based ranking is compare with the other ranking techniques (item average ranking, item absolute ranking techniques) and also maintain the accuracy and diversity of recommendation.

Items	Item op	Avg rating	Abslut	RP rating	Relat ive
Websites	14	34	1	5	20
Business	14	30	8	10	60
Products	5	9	2	7	13
Software's	3	5	0	7	0
Cleaner	1	2	1	10	7
Eye cleanr	0	0	0	0	0
Google	2	4	0	5	0
Yahoo	0	0	0	0	0
Orkut	0	0	0	0	0
Avg	0	0	0	0	0
Laptops	0	0	0	0	0
Cells	0	0	0	0	0
Books	1	2	0	5	0

V. CONCLUSION

This work appraises several interesting directions for future research. In particular, additional important item ranking criteria should be explored for possible diversity improvements. This may contain consumer-oriented or producer oriented ranking mechanisms, depending on the given application domain, as well as external factors, such as social networks. Also, as mentioned previously, optimization-based approaches could be used to achieve further improvements in recommendation diversity, although these improvements might come with a (probable significant) increase in computational intricacy. Moreover, since of the inherent tradeoff between the accuracy and diversity metrics, an interesting research direction would be to expand a new calculate that captures both of these aspects in a solitary metric.

In this project, proposed a number of recommendation ranking techniques that can present significant improvements in recommendation diversity with only a minute quantity of accuracy loss. In addition, these ranking techniques offer flexibility to system designers, since they are parameterizable and can be used in conjunction with diverse rating prediction algorithms (i.e., they do not require the designer to use only some specific algorithm). They are also based on scalable sorting based heuristics and, thus, are extremely efficient. In this work provide a comprehensive empirical evaluation of the proposed techniques and obtain consistent and healthy diversity improvements transversely numerous real-world datasets and using different rating prediction techniques.

REFERENCES

[1] A. Ghose and P. Ipeirotis, "*Designing Novel Review Ranking Systems: Predicting Usefulness and Impact of Reviews*," Proc. Ninth Int'l Conf. Electronic Commerce (ICEC), 2007.

[2] D.G. Goldstein and D.C. Goldstein, "Profiting *from the Long Tail*," Harvard Business Rev., vol. 84, no. 6, pp. 24-28, June 2006

[3] D. Billsus and M. Pazzani, "*Learning Collaborative Information Filters*," *Proc.Int'l Conf. Machine Learning*, 1998.

[4] D. Lemire, S. Downes, and S. Paquet, "*Diversity in Open Social Networks*," technical report, University of Quebec, Montreal, 2008.

[5] D. Fleder and K. Hosanagar, "*Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity*," Management Science, vol. 55, no. 5, pp. 697-712, 2009

[6] E. Brynjolfsson, Y.J. Hub, and D. Simester, "*Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales*," Management Science, vol. 57, no. 8, pp. 1373-1386, 2011.

[7] G. Adomavicius and A. Tuzhilin, "*Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*," *IEEE Trans. On Knowledge and Data Engineering*, 17(6), pp. 734-749, 2005.

[8] G. Shani, D. Heckerman, and R. Brafman, "*An MDP-Based Recommender System*," J. Machine Learning Research, vol. 6, pp. 1265-1295, 2005.

[9] J. Bennett, and S. Lanning, "*The Netflix Prize*," *Proc. of KDD-Cup and Workshopat the 13th ACM SIGKDD Int'l Conf. on Knowledge and Data Mining*, 2007.

[10] J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J. Riedl, "*Evaluat-ing Collaborative Filtering Recommender Systems*," ACM Trans. Information Systems, vol. 22, no. 1, pp. 5-53, 2004.

[11] K.Bradley and B. Smyth, "*Improving Recommendation Diversity*," Proc. 12th Irish Conf. Artificial Intelligence and Cognitive Science, 2001.

[12] M. Balabanovic and Y. Shoham, "*Fab: Content-Based, Collaborative Recommendation*," Comm. ACM, 40(3), pp. 66-72, 1997.

[13] P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl, "*GroupLens: An Open Architecture for Collaborative Filtering of Netnews*," Proc. Computer Supported Cooperative Work Conf., 1994.

[14] R. Bell and Y. Koren, and C. Volinsky, "*The BellKor solution to the Netflix Prize*," www.netflixprize.com/assets/ProgressPrize2007_KorBell.pdf, 2007.

[15] R. M. Bell, Y. Koren, and C. Volinsky, "*The Bellkor 2008 solution to the Netflix prizes*, "http://www.research.att.com/~volinsky/netflix/Progres sPrize2008 Bell KorSolution.pdf, 2008.