ISSN:2231-0711
Arun Kumar R et al | IJCSET(www.ijcset.net) | November 2016 | Vol 6, Issue 11, 387-391

Ranking Based Techniques Using Optimised Recommendation Diversity

Arun Kumar R[#1], Buchibabu B[#2] , T.Venkat Narayana Rao[#3]

[#1]Asst. Professor , SNIST, [#2]Asst. Professor, CMREC, [#3]Professor, SNIST, Hyderabad, T.S, INDIA.

Abstract— A **recommender system is a system performing information filtering to fetch information items such as movies, music, books, news, images, web pages and tools for a user. This information is filtered so that it is likely to be used by the user. The aim of a recommender system is often to "help consumers learn about new products and desirable ones among myriad of choices" this would enable in improving the recommendation quality, accuracy and diversity. In this paper a item popularity based ranking algorithm is utilized. By using this algorithm the best selling items are recommended to each user. In this paper different types of ranking techniques have been considered. The techniques used are Item Popularity-Based Ranking, Reverse Predicted Rating Value, Item Average Rating, Item Absolute Likeability and Item Relative Likeability. These techniques can generate recommendations that have substantially and higher aggregate diversity across all users while maintaining optimal levels of recommended accuracy.**

Keywords— Recommender systems, Recommendation accuracy, Recommendation diversity, Ranking functions, collaborative filtering.

I. INTRODUCTION

In the age of information surplus load, 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 systems are used by Amazon and Netflix etc[1][2].

The most common formulation of the recommendation problem depends on the notion of ratings i.e., recommender systems guess 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 strange items for every user, frequently using other users' ratings, and recommend top *N* items with the highest predicted ratings values. Accordingly, there have been numerous studies on developing novel algorithms that can advance the predictive accurateness of the recommendations. However, the quality of recommendations can be evaluated along a number of dimensions, and relying on the precision of recommendations alone may not be enough to uncover the most relevant items for each user. These studies argue that one of the goals of recommender systems is to offer 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 anticipated new recommendation methods that can augment the diversity of recommendation sets for a given individual user, often measured by an average variation between all pairs of recommended items, while maintaining an suitable level of accuracy. These studies measure recommendation diversity from an individual user's perspective (i.e., individual diversity)[3][4].

In contrast to individual diversity, which has been explored in a number of papers, some recent studies started probing the impact of recommender systems on sales diversity. This was performed on aggregate diversity of recommendations across all the users and noted that high individual diversity of recommendations does not essentially imply high aggregate diversity. For example, if the system recommends to every users the similar five best-selling items that are not related to each other, the recommendation list for each user is diverse (i.e., high individual diversity), but only five separate items are recommended to every user who ever purchased them (i.e., resulting inflow aggregate diversity or big sales concentration).

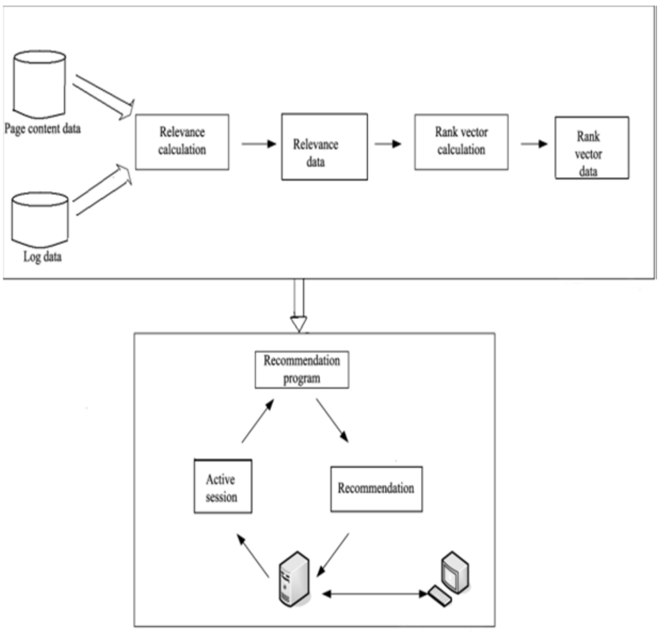

Fig. 1 Architecture of the recommendation diversity

The Figure 1 describes the main participating parties in A Recommender System. Assume that the information concerning a data provider is stored in a single tuple and database is kept confidential at the server. The users can be given their opinion and feedback for the particular products that they access on database. The admin can provide ranking to the products based on the recommendations. While the benefits of recommender systems that provide higher aggregate diversity would be apparent to many users (because such systems focus on providing wider range of items in their recommendations and not mostly bestsellers, which users are often capable of discovering by themselves), such systems could be beneficial for some business models as well. For example, it would be profitable to Netflix if the recommender systems can encourage users to rent "long-tail" type of movies. This is due to the fact that they are less costly to license and acquire from distributors than new release or highly-popular movies of big studios. However, the impact of recommender systems on aggregate diversity in real-world e-commerce applications has not been well understood. For example, one study, using data from online clothing retailer, confirms the "long tail" phenomenon that refers to the increase in the tail of the sales distribution (i.e., the increase in aggregate diversity) attributable to the usage of the recommender system. On the other hand, another study shows a contradictory finding that recommender systems actually can decrease the aggregate diversity in sales. This can be explained by the fact that the idiosyncratic items often have limited historical data and, thus, are more difficult to recommend[5][6].

The remainder of the paper is organized as follows: Section II reviews relevant literature on traditional recommendation algorithms and the evaluation of recommendation quality. Section III describes our motivations for alternate recommendation ranking techniques, such as item popularity. The main empirical results follow in section IV. Lastly, section V concludes the paper by summarizing the contributions with future directions.

II. RELATED WORK

A. Recommender Systems

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 users who have similar preferences (i.e., "neighbours") or liked in the past. Finally, hybrid approaches can combine content-based and collaborative methods in several diverse ways[7][8][9].

B. Collaborative Filtering

This family of algorithms is widely used in recommender systems which deals with collaborative filtering. Collaborative filtering methods are based on collecting and analysing a large amount of information on users' behaviour, 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].

C. Content Based Filtering

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].

D. Hybrid Recommender Systems

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].

E. Recommendation Algorithms Accuracy of Recommendations

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. As examples of statistical accuracy metrics, mean absolute error (MAE) and root mean squared error (RMSE) metrics measure how better a system can predict an precise rating value for a exact item[15][16].

F. Diversity of Recommendations:

As mentioned the diversity of recommendations can be measured in two ways: individual and aggregate. Most of recent studies have focused on increasing the individual diversity, which can be calculated from every user recommendation list (e.g., an average dissimilarity between all pairs of items recommended to a specified user). These techniques intend to avoid providing too similar recommendations for the identical user. On the other hand, except for some work that examine sales diversity across all users of the system by measuring a statistical dispersion of sales there have been few studies that explore *aggregate diversity* in recommender systems with good results.

III. MOTIVATIONS FOR RECOMMENDATION RANKING

A. Standard Ranking Approach

The standard ranking approach is designed to aid improve recommendation accuracy, but not recommendation diversity. Therefore, new ranking criteria are needed in order to achieve diversity improvement. Since recommending best selling items to every user typically tends to diversity reduction, recommending less popular items intuitively should have an effect toward increasing recommendation diversity. Following this motivation, we discover the option to use item popularity as a recommendation ranking standard.

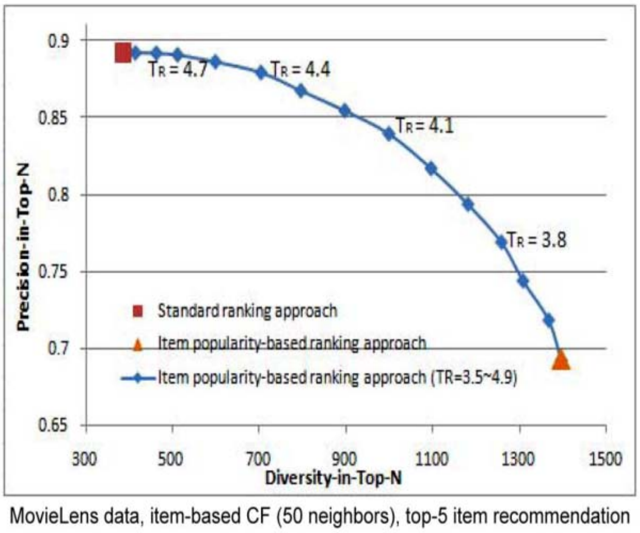

Fig. 2 Performance of the standard ranking approach and item popularity-based approach with its parameterized versions.

Figure 2 shows recommendations of the most highly predicted items selected by the standard ranking approach to help improvement in recommendation accuracy. Therefore, new ranking criterion is needed in order to attain diversity improvement. Since recommending best selling items to each user usually leads to diversity reduction, recommending less admired items intuitively should have an outcome toward increasing recommendation diversity.

IV. PROPOSED APPROACH: ITEM-POPULARITY-BASED RANKING

// Item Popularity-Based Ranking Algorithm
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 theRanking threshold T_R with respect to the threshold value T_H

   $$Rank(i, T_R) = \begin{cases} Rank_x(i), & \text{if } R^*(u, i) \epsilon [T_R, T_{max}] \\ \alpha_u + rank_{standard}(i), & \text{if } R^*(u, i) \epsilon [T_H, T_R] \end{cases}$$

   where $I_u^*(T_R) = \{i \epsilon I | R^*(u, i) \geq T_R\}$.
   $\alpha_u = max rank_x(i)$
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 a threshold choosen 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:

$$rank_{ItemPop(i)=|U(i)|, where\ U(i)=}\{|u \epsilon U \mid \exists R(u,i)\}.$$

There exist multiple variations of neighbourhood-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 neighbours 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 neighbourhood 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 neighbourhood-based CF calculations[17][18].

In this work compared the performance of the item popularity based ranking approach with the standard ranking approach using Movie Lens dataset and item-based CF, and presents a comparison using the accuracy-diversity as plotted in the Figure 1. The presented results show that, as compared to the standard ranking approach, the item popularity-based ranking approach increased recommendation diversity from 385 to 1395 (i.e., 3.6 times). However, the recommendation accuracy dropped from 89% to 69%. Here, despite the significant diversity gain, such a significant accuracy loss (20%) is not acceptable in most real-life personalization applications. Therefore, it is suggested to introduce a general technique to parameterize recommendation ranking approaches, which permits achieving significant diversity gains while controlling the precision losses.

A. *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 number of additional ranking functions $rank_x(i)$.

The first step, shown in Figure 3, 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 i.e. as long as they are above the highly-predicted rating threshold *TH*. 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).

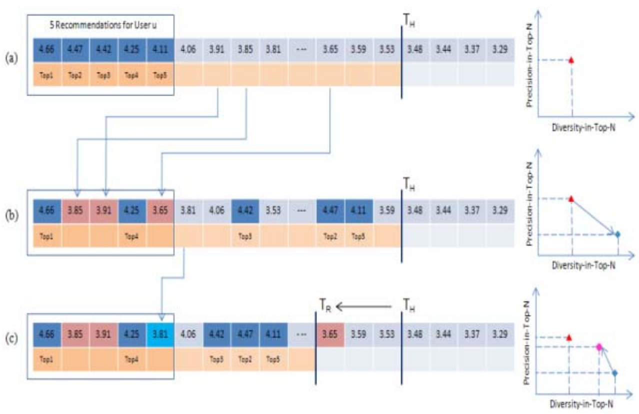

Fig. 3 General overview of ranking-based approaches for improving recommendation diversity

The second step, depicted in Figure 3, showcases the recommendations provided by applying one of the proposed ranking functions, $rank_x(i)$, where several items (that are not necessarily among *N* most highly predicted, but are still above *TH*) are recommended to the user. In this way, a user can get recommendations of more idiosyncratic, long-tail, less frequently recommended items that may not be widely popular, but can still be very pertinent to the user (as indicated by relatively high predicted rating). Therefore, re-ranking the candidate items can considerably improve the recommendation diversity although, and this usually comes at some loss of recommendation accuracy. The performance graph of the second step (b) demonstrates this accuracy-diversity trade off.

The third step, shown in Fig. 3 can significantly minimize accuracy loss by confining the re-ranked recommendations to the items above newly introduced ranking threshold *TR*. In this particular illustration, note that the amplified ranking threshold makes the fifth recommended item in the second step (b) filtered out and the next possible item above the new ranking threshold is recommended to user "*u*". Averaged across all users, this parameterization helps to make the level of accuracy loss moderately 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). We now introduce several additional item ranking functions, and offer empirical evidence that supports our impetus of using these item criteria for diversity improvement.

V. RESULTS

TABLE I depicts 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.

TABLE I Comparison Results with Different Ranking Techniques

Items	Itemp 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
CCleaner	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

VI. CONCLUSIONS AND FUTURE WORK

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 trade off 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.

Recommender systems have made significant progress in recent years and many techniques have been proposed to perk up the recommendation worth. However, in most cases, novel techniques are designed to improve the accuracy of recommendations, whereas the recommendation diversity has often been overlooked. It tends to perform poorly with respect to recommendation diversity. Therefore, 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[19][20][21].

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. Hu, 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, "*Evaluating 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 NetflixPrize.*,"http://www.research.att.com/~volinsky/netflix/ProgressPrize2008 Bell KorSolution.pdf, 2008

[16] R. Garfinkel, R. Gopal, A. Tripathi, and F. Yin, "*Design of a Shopbot and Recommender System for Bundle Purchases*," Decision Support Systems, vol. 42, no. 3, pp. 1974-1986, 2006.

[17] S.Funk, "*Netflix Update*:" http://sifter.org/simon/journal/20061211.html, 2006.

[18] X. Su and T.M. Khoshgoftaar, "*Collaborative Filtering for Multi-Class Data Using Belief Nets Algorithms*," Proc. Eighth IEEE Int'lConf. Tools with Artificial Intelligence, pp. 497-504, 2006

[19] Y. Koren, "*Tutorial on Recent Progress in Collaborative Filtering*," Proc. ACM Conf. Recommender Systems, pp. 333-334, 2008.

[20] Y. Koren, "*Collaborative Filtering with Temporal Dynamics*," Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 447-456, 2009.

[21] Z. Zheng and B. Padmanabhan, "*Selectively Acquiring Customer Information: A New Data Acquisition Problem and an Active Learning-Based Solution*," Management Science, vol. 50, no. 5, pp. 697-712, 2006.