Diversity in Recommender System

Vaishali Ghanghas[#1], Chhavi Rana[#2], Sunita Dhingra[#3]

Department of computer science engineering,UIET
M.D.University, Rohtak, Haryana- 124001, India ,

***Abstract*—Recommendation system is an activated area of research that helps to allow users to find the preferable items quickly and to avoid the possible information overloads. Recommender systems use data on past user preferences to predict possible future likes and interests. A better recommender system would offer less common papers that also draw the user's interest. Diversity is very much related to this aspect. It generally applies to a set of items and is related to how different the items are with respect to each other. Many diverse recommendation techniques have been developed, including collaborative filtering, content-based analysis. These techniques involve presenting different types of recommendation to the users which are similar in taste. The laid down work has concluded that the diversity in a set of items can be increased at a cost of reducing system accuracy. Though, the feature of diversity is contrasting to accuracy, many researchers have tried to bring in congruence between the two. This paper present objective functions that capture the trade-off between the goals and optimization problems associated with the maximization of these objectives.**

***Keywords*- Recommender system, collaborative filtering, diversity, dataset and control parameter.**

I. INTRODUCTION

The primary goal of recommenders is to provide personalized recommendations so as to satisfy users' interests. Recommendation tasks generally involve a large set of items –such as books, movies or songs– and a large set of users to which the system provides suggestions of items they may enjoy or benefit. A good recommender system would offer less common papers that also draw the user's interest. Let's take example of academic paper recommender system. Suppose all the recommendations were for the paper that the user has read somewhere. Even if the user is strongly interested in the papers written on a topic and the recommender system is very good at ranking them in order of preference, it is a poor recommender system because it shows similar pages repeatedly and not the diverse one[7]. Consider both Collaborative Filtering (CF) and case-based top-N recommenders whose goal is to predict a list of N products that a user will like or be interested in purchasing. The previous evaluations, mainly focused only on the accuracy of the generated predictions based, e.g., on the Mean Absolute Error. However, a few recent works states that accuracy is not the only metric for evaluating recommender systems and that there are other important aspects we need to focus on in future evaluations [4, 10]. In recommender system the property of dynamism itself is a combination of various parameters where each parameter can be the core ingredient in developing a standalone system. DRS are the systems that are able to register changes occurring in the user sphere, the system sphere as well as other environmental changes implicitly or explicitly and accordingly modify their recommendations to the users. Based on how the information of the user profile is employed, recommender systems can be divided in three categories:

• Content-based (CB): the recommender will retrieve items whose content is similar to those of the profile.
• Collaborative Filtering (CF): the recommender will retrieve items based on connections or similarities between user profiles.
• Hybrid approaches, combining CB and CF.

One of the most famous recommendation systems nowadays is the Amazon.com

A key challenge is that while the most useful individual or recommendations are to be found among diverse objects, the most reliably accurate results are obtained by methods that recommend objects based on user or object similarity [3]. Recommendation systems apply data mining techniques to determine the similarity among thousands or even millions of data. Hence, diversity is being identified as key dimension of recommendation utility in real scenarios, and a fundamental research direction to keep in making progress in this field. The problem of results diversity has been already addressed in IR, but from a different angle. The diversity dimension of search results is being researched in the IR field as a means to address the ambiguity and/or under specification involved in user queries. Current approaches to enhance and evaluate the diversity of search results use concepts such as query intents and document similarity. Query intents can be seen as the different meanings or purposes an underspecified query can represent. Taxonomies and query logs have been used for discovering and describing these intents. The identification of query intents and interpretations is then used to discover categories or refinements which may suit a query. Maximizing the range of categories covered by returned documents is a means to cope with the initial ambiguity of a query.

Diversity is highly desirable features for automatic recommendation. ones In most scenarios, the purpose of recommendation is inherently linked to a notion of discovery of new items. This is generally a good approach to enhance the chances that the user is pleased by at least some recommended item. Sales diversity may enhance businesses as well, leveraging revenues from market niches [4]. Diversity play a fundamental part as dimensions of recommendation utility, most authors have dealt with this property as opposing goals to accuracy, stating the problem as a multi-objective optimization issue, where an optimal trade-off between accuracy and diversity is sought. Novelty and diversity are different though related notions means if there is diversity in recommended data, novelty is certainly present there. The novelty of a piece of information generally refers to how different it is with respect to "what has been previously seen or known", by a specific user, or by a community as a whole. Diversity generally applies to a set of items, and is related to how different the items are with respect to each other. This is related to novelty in that when a set is diverse, each item is "novel" with respect to the rest of the set. Moreover, a system that promotes novel results tends to generate global diversity over time in the user experience; and also enhances the global "diversity of sales" from the system perspective. It is worth to make a distinction between individual diversity and aggregate diversity. The first case accounts for how different are items in a recommendation list for a only user, which is normally the notion of diversity employed in most works. Nevertheless, aggregate diversity –understood as the total amount of different items a recommendation algorithm can provide to the community of users is also a very interesting quality of a RS as a whole.

Recommending too similar items is less profitable for the user and the vendor– than offering a more varied experience. Information Retrieval Diversity for Recommender Systems includes both the notion of novelty and serendipity.

II. EXISTING TECHNIQUES

The best of these methods, greedy selection, adds cases one at a time to the retrieval set, according to a heuristic measure combining diversity and similarity. However, the authors do not examine the impact that additional diversity has on retrieval performance. We examine this issue by explicitly evaluating a system's ability to recommend novel relevant items also tackle the issue of diversification of recommendation lists [13]. The authors propose a similarity metric using a taxonomy-based classification and use it to compute an introits similarity metric to determine the overall diversity of the recommended list. Intra list similarity is analogous to notion of diversity, except that it is a decreasing rather than in- creasing function of diversity. Originating with the work of Smyth and Macclave [1], the issue of diversification of recommendation lists has been tackled.

Collaborative filtering techniques have been proven to provide satisfying recommendations to users [5]. Movie Lens is a movie recommendation system based on Group Lens technology. Chee [19] explained Recommendation Tree (Rec Tree) is one method using divide-and-conquer approach to improve correlation-based collaborative filtering and performing clustering on movie ratings from users. The ratings are extracted from MovieLens Dataset. Ringo Shardanand and Maes [23] provides music recommendations using a word of mouth recommendation mechanism. The terminology "social information filtering" was used instead of collaborative filtering in the paper. Ringo determines the similarity of users based on user rating profiles. Firefly and Gustos are two recommendation systems which employed the word-of-mouth recommendation mechanism to recommend products. Web Watcher has been designed for assisting information searches on the World Wide Web Armstrong et al. [18]. Web Watcher suggests users which hyperlinks would lead to the information that users want. The general function serving as the similarity model is generated by learning from a sample of training data logged from users. Yenta is a multi-agent matchmaking system implemented with the clustering algorithm and the referral mechanism Foner [11]. Jester is an online joke recommendation system. The clustering is based on continuous user ratings of jokes Goldberg et al. [10].

The authors provide a heuristic algorithm to increase the diversity of the recommendation list. Their results show that lists ordered for greater diversity perform worse on accuracy measures than unaltered lists, but nevertheless users preferred the altered lists. In the context of conversational recommender systems, Smith[4] propose the idea of presenting diverse compound critiques and evaluate the effectiveness of two alternative approaches in terms of their recommendation performance. Furthermore, it highlights the pitfalls of naive incorporation of current diversity enhancing techniques into existing recommender systems. Another paper Mcsherry [5] examines the conditions in which similarity can be increased without loss of diversity and presents an approach to deliver such similarity-preserving increases in diversity when possible.

Lathia et al. [16] deal with temporal diversity in CF recommender systems. In a realistic scenario, users interact with recommender systems iteratively over time, so new models must be trained regularly to adapt to new users, new items or updated user profiles. They carried out two experiments. An online experiment showed that user's perception of the recommendations lists degrades if the does not show diversity with respect to paste recommendations to the same user. The offline experiment compared the temporal diversity of some CF recommenders among time, reaching interesting conclusions. Adomavicius et al. [8] address diversity as the ability of a system to recommend as many different items as possible over the whole population. This form of aggregate diversity is measured as the size of the set of all items a recommender system is able to recommend to its users as a whole. As a diversity-enhancing approach, they proposed a parametric re-ranking method combining standard CF recommenders with other ranking criteria that promote aggregate diversity but have poor accuracy, so they compensate.

Finally, examine the effect of recommender systems on the diversity of sales. It uses a measure of statistical dispersion called the Gini coefficient to measure sales diversity. This work is to address top-N recommendation rather than rating prediction, that is, the focus is to recommend N products that the system predicts are likely to be relevant to the end-user, rather than to predict the rating that an end-user might give to any particular product. We view the problem as a classification problem, rather than a ranking problem; that is, our goal is to classify items as being either "relevant" or not relevant to a particular user and a good recommendation should contain as many relevant items as possible. Diversity involves presenting different types of recommendation to user which are similar in taste. Though, the feature of diversity is contrasting to accuracy, many researchers have tried to bring in congruence between the two[22]. C-N. Ziegler et al. [22] Present topic diversification, a novel methodology designed to balance and diversify and personalized recommendation lists in order to reflect user's complete spectrum of interests. This procedure is somewhat detrimental to accuracy and is being worked upon.

Furthermore, proposes a new approach that can improve the diversity of Top - N item selection by taking account rating variance which can work in conjunction with any existing recommendation technique [12]. Moreover, user can control; the balance between the accuracy and diversity of recommendation through an adjusted ranking and filtering combined approaches which adjust condition in selecting N items acc. to the user. Similarly, Adomavicius and Kwon [8] explore the advantage of variance in neighbourhood rating in the recommendation process to overcome the accuracy/diversity trade-off. Another researcher develops a model to maximize the diversity of received list while maintaining adequate similarity to the user query as a binary optimization problem Bridge [5].

Kwon [25] also shows that temporal diversity is an important facet of RS through a user study. They also examine the diversity of three CF overtime by defining diversity matrices. Moreover they provide several methods that could be used to improve the diversity of recommendation.

III. POSITIVE ASPECTS

As defined by Clarke [3] diversity is a quality of result lists that helps cope with ambiguity or under specification. Quite often a typical short textual query can represent more than one concept or interpretation (the case is clear, for example, with acronyms or polymeric words), in which case the query is called ambiguous. Consider the query "apple", which could refer to the fruit, the computer industry corporation, a record label, and other less common interpretations. Users interested in one interpretation would not usually be interested in the others. Even when the query does identify a unique concept or entity, it may still be underspecified in the sense that it may have different aspects. Consider a query like "Mallorca", which refers clearly to an island in the Mediterranean Sea, but still involves uncertainty about the actual specific user interest behind the query, which might relate to general information about the island, touristic deals, the football team, etc. In this case these aspects do not need to be mutually exclusive, that is, users may be interested in two or more of them. In this work we will refer to both interpretations and aspects as subtopics, since we shall deal with both in the same way –as generally do prior approaches in the state of the art literature. As a strategy to cope with ambiguity and under specification, several authors have researched approaches that aim to cover as many subtopics as possible.

IV. FUTURE SCOPE

A good recommendation system should be dynamic, in nature. It should help in the updates on profiles that can be performed approximately in real time. Although it is certainly true to the aspect that the innovations of hardware designs increase the computational speed, algorithms and techniques with low time computational complexity are expected in the recommendation system developments. User data flow in every second so we require memories to keep the profiles up-to-date. Therefore, it is important to maximize the offline computations. Two factors on which the computation time depends are: the number of items and the number of users in the database. The impact of the first factor, i.e. the number of items, may be reduced. The database is formed by adding data continuously (the opposite is the data stream, e.g. video stream). However, the decision on the frequency of updates on user profiles is more complex. How often should the updates be performed in order to keep track of the user preference trends? If the updates are required to be performed approximately in real time, an algorithm or a technique with low memory computational complexity is essential to reduce the system loading and eliminate the potential effects on the system synchronization. The comprehensive research on the core module increases by implementing dynamic nature of RS. We will improve various dimensions with respect to above and not just accuracy. Therefore, the collaborative filtering algorithms are being continuously modified in order to handle the dynamism of the whole process.

Increasing diversity is always considered a desirable feature of dynamic recommendation system. This issue has been addressed in some previous research by Smith and MClave [1]. But there is a trade off between diversity and accuracy. As increasing diversity decreases accuracy .So, we have to include a parameter named thetta, which allows explicit control of the weighing given to the concern described here. Amazon.com use the overlap strategies between customers' past purchases and browsing activity for recommendation of products to the users , on the other hand, TiVo digital video system recommends TV shows and movies on the basis of correlations in users' viewing patterns and ratings. The risk associated with such an approach is that, with the recommendations based on overlap rather than difference, more and more users will be exposed to a narrowing band of popular objects, while the items that are very relevant will be overlooked. To fulfil this purpose diverse sets are made to increase diversity of a specific topic recommended by the user. Making a diverse set of recommendations is easy but it is rather difficult to maintain this diverse set. Diversity basically leads to decrease similarity among items recommended by user. So, breaking barriers of similarities and stills maintains relevant results are one of the main challenges. Moreover research of recommended data is also a necessary work; the off-line research refers to availability of data sets. At present, the recommender systems community has a single widely-available data set—the Each Movie data set. The remaining challenge is to develop a set of data analysis tools, designed to generate community-accepted quality and performance metrics, that can be used to verify diversity.

V. CHALLENGES

The problem regarding diversity is to consider the importance of control parameter, which determines the importance given to diversity in recommendation set. Making a diverse set of recommendations is easy. It is difficult to ensure that this diverse set contains many items that are relevant to the user query. The focus on similarity is compounded by the metrics used to assess recommendation performance. A typical method of comparison is to consider an algorithm's accuracy in reproducing known user opinions that have been removed from a test dataset. Recommendation, however, is not necessarily a *useful* one: real value is found in the ability to suggest objects users would not readily discover for themselves, that is, in the novelty and diversity of recommendation. Despite this, most studies of recommender systems focus overwhelmingly on accuracy as the only important factor [for example, the Netflix Prize challenged researchers to increase accuracy without any reference to novelty or personalization of results.

VI. CONCLUSION

Different collaborative filtering techniques have been proposed to decrease the processing time and the data latency. The research presented here aims to contribute to understanding the basic element involved in recommendations named diversity. The proposed framework provides details that include different perspectives on diversity and deriving new ones. These aspects generally introduce the purpose why diversity is necessary in recommendation system. Further it tells about the benefits as well as the challenges associated with this feature of recommendation system. .

Diversity for RS and IR has attained a big interest in the last few years. This property is essential in RS for real-world scenarios and applications –example: online commerce–, where the aggregated relevance of individual items or document does not necessarily guarantee an optimal or even satisfactory user experience. Diversity is directly associated with avoiding the monotony of recommendations, thus improving the capacity of discovery and broadening and enriching the user experience. They worked on the concept of aspect space as a mean to translate two key notions of IR diversity, document similarity and query intents, to their correspondences to the RS field: item similarity and user profile aspects, respectively.

ACKNOWLEDGEMENT

The clear concern is that an algorithm that focuses too strongly on diversity rather than similarity is putting accuracy at risk. How this algorithm can be coupled in a highly efficient hybrid with a diffusion-based recommendation method recently introduced by our group. Using three different datasets from three distinct communities, we employ a combination of accuracy- and diversity-related metrics to perform a detailed study of recommendation performance and a comparison to well-known methods.

REFERENCES

1) B. Smyth and P. Mcclave (2001). Similarity vs. diversity, in the proceedings of communication of the ACM, 40(3): 47–56.
2) C. N. Ziegler, S. M. Mcnee, J. A. Konstan and G. Lausen (2005). Improving recommendation lists through topic diversification, In the Proceedings of the 14th international conference on World Wide Web (WWW'05), ACM, New York, USA, pp. 22-32.
3) C. Clarke, M. Kolla, G. Cormack, O.Vechtomova, A. Ashkan, S. Büttcher, et al. (2008). Novelty and Diversity in Information Retrieval Evaluation. 31st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'08), (pp. 659-666). Singapore.
4) D. Bridge, M. H. Goker, L. Mcginty, and B. Smyth (2005). Case-Based recommender systems, In the proceedings of 3rd International conference on Recommender System (RecSys'05), New York, NY, USA, pp. 361-387.
5) D. Fleder and K. Hosanagar (2007). Recommender systems and their impact on sales diversity, In Proceedings of the 8th ACM Conference on Electronic Commerce, Montreal, Canada, pp. 192–199.
6) D. Mcsherry (2002). Diversity-conscious retrieval, In Proceedings of the 3rd European Conference on Case-Based Reasoning, Chicago, Illinois, USA, pp. 219–233.
7) G. Karypis (2001). Evaluation of item-based top-N recommendation algorithms, In Proceedings of the 10th International Conference on Information and Knowledge Management of ACM SIGMIS, Chiba, Japan, pp. 247–254.
8) G Adomavicius and Y. Kwon (2007). Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques. IEEE Transactions on Knowledge and Data Engineering.
9) J. Breese, D. Heckerman and C. Kadie, (1998). Empirical analysis predictive algorithm for collaborative filtering. Microsoft Technical Report MSR-TR, 98(12):45-67.
10) K. Goldberg, T. Roeder, D. Gupta and C. Perkins, (2001). Eigentaste: A constant time collaborative filtering algorithm. Information Retrieval 4(2), pp. 133–151.
11) L. Foner, (1997). Yenta: A multi-agent, referral-based matchmaking system. Proceedings of The First International Conference on Autonomous Agents. ACM. 16(49): 21-27.
12) M. Balabanovic and Y. Shoham (1997). Fab: Content-based collaborative recommendation, Communications of the ACM, 40(3): 88-89.
13) M. Dias and P. J. Lisoba (2003). On the role of diversity in conversational recommender systems, In Workshop Proceedings of the 5th International Conference on Case-Based Reasoning (ICCBR'03), pp.65-98.
14) M. Zhang (2009). Enhancing diversity in Top-N recommendation, In Proceedings of the 3rd ACM conference on Recommender systems RecSys, Geneva, Switzerland, pp.397-400.
15) M. Zhang and N. Hurley (2010). Avoiding Monotony: Improving the Diversity of Recommendation Lists, In ACM RecSys, 40(6):86-98.
16) N. Lathia, S. Hailes, L. Capra and X. Amatriain (2010). Temporal diversity in recommender systems. In the proceedings of 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'10), Geneva, Switzerland, pp. 210-217.
17) P. Resnick and H. J. Varian (1997). Recommender Systems, In Proceedings of the Communications of the ACM, 40(3): 56–58.

18) R. Armstrong, D. Freitag, Joachims, T. Mitchell (1995). Webwatcher: A learning apprentice for the world wide web. AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments,pp. 6-12.

19) S. H. S. Chee, J. Han and K. Wang (2001). RecTree: An Effecient Collaborative Filtering Method, in the proceedings of ACM SIGCHI Conference on Information Retrieval, New York, USA, pp.141-151.

20) S. Sweeney, F. Crestani, and D. Losada (2008). "Show me more": Incremental Length Summarisation Using Novelty Detection. Information Processing & Management, pp. 663-686.

21) T. Chen and W. Han (2007). Content recommendation system base on private dynamic user profile, Machine Learning, 16 (4):12-18.

22) T. Sakai (2006). Evaluating evaluation metrics based on the bootstrap, in the proceedings of 29th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'06), Seattle, WA, USA, pp. 525-532.

23) U. Shardanand and P. Maes (1995). Social information filtering: Algorithms for automating 'Word of Mouth', In Proceedings of 1995 conference on recommender systems, Tokyo, Japan, pp. 210-217.

24) W. Hill, L. Stead, M. Rosenstein, and G. Furnas (1995). Recommending and evaluating choices in a virtual community of use, In Proceedings of the 1995 conference on recommender systems, Madrid, Spain, pp. 194-201.

25) Y. Kwon (2008). Improving top-n recommendation techniques using rating variance, Proceedings of the 2008 ACM conference on Recommender Systems, Geneva, Switzerland, pp. 456-487.

26) Y. H. Chien and E. I. George (1999). A Bayesian Model for Collaborative Filtering and Information Gathering from Heterogeneous Distributed. Environments, In Proceedings of 4th ACM conference on recommender systems, Vancouver, Canada, pp. 6-12.