ISSN:0975-9646

Diversification of User Recommendations Using Entropy Redistribution

Abhinav Pachauri
Indian Institute of Technology,Kharagpur
India

***Abstract*—Different users may have different ranges of interests - the preference of a highly focused user might include only few topics, whereas that of the user with broad interests may encompass a wide range of topics. In the recommendation systems most of the algorithms mostly focus on the accuracy and recall of results. The most accurate result according the defined metric may not be in the interest range of the user. The diversity of the result interests users while the diversification must be within the range of interest of the users. In this paper a method for diversification has been proposed which includes both individual and aggregate diversity using the entropy of topics and users who are being recommended by this diversification task. Whenever a particular user will have high entropy, a topic having low entropy should be recommended to him and vice-versa for redistributing entropy and so eventually the diversification of recommendations will be achieved.**

***Index Terms*—Recommendation systems, Information retrieval, Entropy.**

I. INTRODUCTION

Recommendation systems are becoming more and more popular to help users find relevant information on the Web according to their interest. The successes of Netflix and Amazon are good testaments to this trend. However, while users enjoy receiving relevant content items, they also tend to loose interest quickly if the recommended items are too similar to each other. Recommender systems intend to provide people with recommendations of products they will appreciate based on their past preferences, history of purchase, and demographic information. In this paper a entropy based approach which will take both entropies (entropies of the topics and users dynamically) and the re-distribution of the entropy will used to diversify the recommendations.

II. LITERATURE REVIEW

Many of the most successful systems make use of collaborative filtering [1,2,3], and numerous commercial systems, e.g., Amazon.com' s recommender [4], uses these techniques to give the similar personalized results. These variety of collaborative filtering algorithms includes algorithms like probabilistic latent semantic analysis [5] and the adaptation of text retrieval models for recommendation [6,7] .Though the accuracy of state-of- the-art collaborative filtering systems, i.e., the probability that the active user1 will appreciate the products recommended ,is very high. But after showing the similar results the repeatedly, user interest will eventually decrease with a great factor. The diversification work has also been done using attribute analysis [8] and explanation based analysis [9]. Retrieving attributes in many cases like videos and images can be very costly. The explanation construction will have to be made at a with a lot of internal processing of the latent features. Instead of these approaches , in this paper entropy normalization and recommendation according to the entropy distribution for a particular user has been proposed.

III. ENTROPY ESTIMATION AND REDISTRIBUTION

Diversity index is the mathematical measure of topic divesity in a given network of interests of people. Based on the topic richness(the number of topics present) and topic abundance (the number of people liking that particular topic).The more topics a particular user likes more diverse he is. The two types of indices which can be used in this context are dominance index and information statistics indices.The equation of the two indices can be written as :

Shannon Index $= \sum_{i=1}^{T} p(i) \ln p(i)$

Simpson Index $= \sum_{i=1}^{T} 1/p(i)^2$

In the Shannon index, p is the proportion (n/N) of individuals of one particular topic (n) divided by the total number of individuals (all the active users) N, *ln* is the natural *log*, is the sum of the calculations, and *T* is the total number of topics.

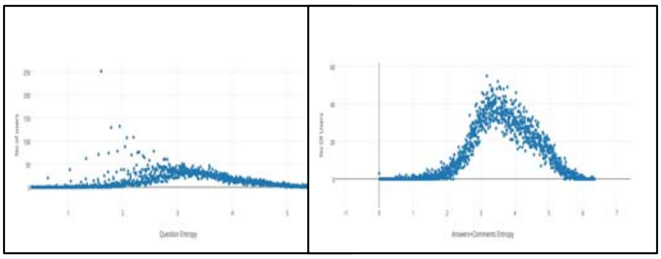
The Simpson index is a dominance index because it gives more weight to common or dominant topics(more popular topics).In this case, a few rare topics with only a few representatives will not affect the diversity. In this paper, the Shanon index will be used because topics having less popularity cannot be neglected directly and have to be taken into consideration during diversification.

Clustering is used to put the users of similar interests together and by this the entropy corresponding to a particular user can be estimated. For this standard clustering algorithm K-means has been used. Value of K will be chosen adaptively. The dimensions of the vectors are the number of topics. The total number of points on the graph corresponds to the users. After estimating the entropies of the topics and users, The redistribution scheme can be applied in which the users corresponding to lower entropy(less active) can be assigned to higher entropy tags/topics and users having higher entropy (more active) can be assigned to the topics which have less entropy (less number of people have shown their interests in that particular topic).By this scheme users will be assigned to more diverse topics than the ones, they were previously liking or showing interests in.

IV. Testing And Results

This scheme is tested on stack-overflow data in which a particular user asks questions, makes comments and answers. Each Question has some tags related to a specific topic on which it has been asked. In our case we have taken the different types of tags as different topics.

First all the tags of questions were indexed according to the userids (a particular userid contain the list of tags).Similarly, tags of answers and comments were also being indexed. In the answer list,a particular userid will contain the tags of those posts on which that user answered .In the comment list, a particular user will have tags of the posts on which he commented and then these lists were merged with respect to the user ids. The final answer+comment list will contain the list of user ids and tags corresponding to answer and comment posts. Using these lists , entropy of the users are being calculated for questions and answer+comments tags. The graphs of this distribution have been shown in fig-1.



(a) b)

Fig. 1: (a)Questions Entropy (b)Answer+ Comments Entropy

Users having entropy value less than a particular threshold (In our case less than 2) will be considered as passive and these users will be recommended higher entropy topics. For diversification of the recommender system, the greedy approach is being used where the recommendation returned by a baseline recommender system is re-ranked for a balance between preference-matching (seeking item relevance and recommendation accuracy) and diversity. This re-ranking is often performed in the light of some kind of categorization of the items in the recommendation domain (such as genres, tags, etc), as the basis to assess the diversity of recommended items. A table showing the effect of the recommender system is being show in Table-1.In this table the Effect of recommendation (according to which user will decide whether it was diversified for him or not) with respect to the number of best item recommended by the system.

TABLE I: Recommendation Model effect on N-best Item

Recommendation Model Effect (%)			
N	**No model**	**Model-1**	**Model-2**
1	15.2	16.2	**19.2**
2	24.6	26.7	**32.5**
3	37.5	40.5	**48.3**

No-Model: Baseline recommendation
Model-1 : Attribute based model used for recommendation.
Model-2 : Our model of entropy redistribution

V. Conclusion And Future Work

We have proposed a way of exploiting the diversity of user preferences to enhance recommendations in terms of accuracy and diversity. Our experiments show that this approach is competitive against diversification based on explicit aspects, particularly so with aspect spaces of high cardinality which raise scalability problems. The diversification may face several problems in which the user having less entropy will be showing more diverse interests. The selection of the users must be done carefully otherwise users showing diverse interests will not be considered and then the diversification will become faulty. The more robust approach will be the future work which should take all the cases into the consideration.

References

[1] Schafer, B., Konstan , J., and Riedl, J. Meta-recommendation systems: User-controlled integration of di- verse recommendations. InProceedings of the 2002 Interna- tional ACM CIKM Conference on Information and Knowl- edge Management(2002), ACM Press, pp. 43–51

[2] Goldberg, D., Nichols, D., Oki, B., and Terry, D. Using collaborative filtering to weave an information tapestry. Communications of the ACM 35, 12 (1992), 61–70.

[3] Herlocker, J., Konstan, J., Borchers, A., and Riedl, J. An algorithmic framework for performing collaborative filter- ing. InProceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Infor- mation Retrieval(Berkeley, CA, USA, 1999), ACM Press,pp. 230–237

[4] Linden, G., Smith, B., and York, J. Amazon.com recommendations: Item- to-item collaborative filtering. IEEE Internet Computing 4, 1 (January 2003)

[5] T. Hofmann. Latent semantic models for collaborative filtering.ACM Trans. Inf. Syst., 22(1):89115, Jan.2004.

[6] A. Bellogin, J. Wang, and P. Castells. Text retrieval methods for item ranking in collaborative filtering. In Proceedings of the 33rd European conference on Advances in information retrieval, ECIR'11, pages301-306, 2011.

[7] J. Wang, S. Robertson, A. P. Vries, and M. J.Reinders. Probabilistic rele- vance ranking for collaborative filtering.Inf. Retr.,(6):477-497, Dec.2008

[8] S. McNee, J. Riedl, and J. A. Konstan. Being Accurate IsNot Enough: How Accuracy Metrics Have Hurt Recommender Systems. In CHI, 2006

[9] Cong Yu,Laks Lakshmanan,Sihem Amer-Yahia. It Takes Variety to Make a World:Diversification in Recommender Systems