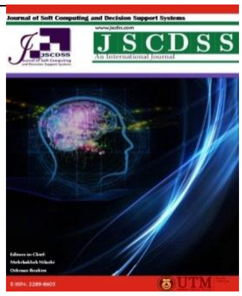
An International Journal
JSCDSS
http://www.jscdss.com
Vol.3 No.1 February 2016: 1-7
Article history:
Accepted 17 October 2015
Published online 17 October 2015

Journal of Soft Computing and Decision Support Systems



E-ISSN: 2289-8603

Entropy-based Ranking Approach for Enhancing Diversity in Tag-based Community Recommendation

Morteza Rashidi Koochi [a,*], Ab. Razak Che Hussin [a], Mehrbakhsh Nilashi [a]
[a] University Teknologi Malaysia, Faculty of Computing, Skudai, Johor, Malaysia

*** Corresponding author email address**: rashidi.morteza@gmail.com

Abstract

Accuracy is the dominant performance evaluation in recommender systems. However, user satisfaction in recommendation includes other factors such diversity and novelty. Some solutions for improving diversity in recommendation, use re-ranking as a post-process on recommendation list achieved from accuracy-aware algorithms. In this work, we propose a method to involve entropy of communities as a diversity factor into the predicted weights from HOSVD method, helping to improve diversity in recommendation list without re-ranking. Experiments on Last.FM dataset, for the case of community recommendation with multi-mode data including users and tags for each community, proves the benefit on introducing diversity factor into the accuracy-based recommendation solutions to improve diversity.

Keywords: Diversity, Recommender Systems, Entropy, Community Recommendation, HOSVD, Multi-Mode data

1. Introduction

Accuracy measurement is the main goal of most of the recommender system evaluation metrics. However, these evaluation metrics operate on narrow of whole collection of data. In the other words, they only evaluate the proportion of items which user has interacted with, and forsake the rest of items. The current evaluation metrics are unable to involve the diversity and coverage of recommended items.

Recommender systems refer to diversity as "how accumulate dissimilarity are between pairs of items in a recommendation list for specific user (intra-list diversity), between two recommendation lists for different users (inter diversity), or whole recommendations of system (aggregate diversity)" (Adomavicius and Kwon, 2012). Whereas novelty of an item is defined as "how different it is with previously seen/known items" (Castells et al., 2011).

In addition to the diversity, there are other concepts of recommendation quality which are in contrast with the accuracy. One of the most obvious questions about RS quality is whether its recommendations can satisfy users' information needs (Herlocker et al., 2004).

Quality, as a concept of measuring, has been discussed and different definitions have released. While the main goal of recommender systems is minimizing the prediction error, redundancy and obviousness as the most shortcomings of current accuracy-aware solutions are not considered. In recommendation context, diversity and novelty are mostly discussed as quality measures (Ge et al., 2010).

2. Previous Works

Recommendation quality of a recommender system is defined as "How many correct recommendations it can generate." However, correct recommendation is a broad concept and need to be investigated.

According to (Cao et al., 2012), there are some reasons to consider factors in addition to accuracy into consideration; humans tend to like variety, discovery, and change, hence pure accuracy oriented solutions may result a boring and ineffective recommendation list. Moreover, over-personalization based on past preferences harms user's growth and experience. However, it needs to balance the conflict between accuracy and non-accuracy factors like diversity, novelty and serendipity.

Auray and Nationale (2007) propose using folksonomy for the recommendation approaches with attention to discovery (relevant items which user wouldn't have found by himself). An intuitive representation of novelty and serendipity in recommendation is mentioned in (Vargas and Castells, 2011) by dividing the related items into three categories for user, including seen, chosen, and relevant items.

Diversity in a set of items is related to how different the items are, comparing with the others. In a specific set of items, when item set is diverse, each item tends to be novel comparing to the rest of the items. When talking about diversity, individual diversity refers to the difference between items in a recommendation for an specific user, and aggregate diversity refers to the total ratio of different item lists which a recommender system can provide for over all users (Adomavicius and Kwon, 2012).

Diversity and coverage is relevant to the concept called Long-Tail effect, which a few items are more popular and the rest are usually unknown for the majority of users. some researches state that, systems can benefit from introducing less-known items to users, even if this leads to lose the chance of introducing the most popular and maybe discovered ones (Adomavicius and Kwon, 2012; Oestreicher-singer and Sundararajan, 2010; Steck et al., 2011).

3. Community Recommendation using HOSVD

In the context of this work, recommender system is using membership data and annotated tag corpus of communities, as two sources of data to be used for community recommendation. It is addressed that, using tags as implicit preferences of users, tackles the problem of sparsity and improves the accuracy (Milicevic et al., 2010). However, extending the traditional user-item relation in recommender systems to user-tag-item, comes with the expense of multiway-array and multi-dimensional computation.

Tensor factorization methods are taken into consideration, when recommender system tries to use more than two sources of data (Symeonidis et al., 2008). For example once using tags as auxiliary information source to traditional dyadic user-item relations in recommender system, the multiway-array structure (tensor) and multi-dimensional decomposition, looks inevitable (Rendle et al., 2010).

Higher Order Singular Value Decomposition (HOSVD) is a popular tensor decomposition method in recommender systems (Nilashi et al., 2014b; Nilashi et al., 2014c). HOSVD de-composes 3-order tensor to components of users, communities and tags. Dense tensor including the predicted ratings, can be estimated using relative components.

In recommender systems, reviewing the process of making recommendation list using HOSVD, Fig. 1 shows the process of prediction of the ratings. Firstly, three primary matrices of User-Community, User-Tag, and Tag-Community is used to make sparse tensor, then Tensor Decomposition is used to predict all missed ratings in a form of dense tensor.

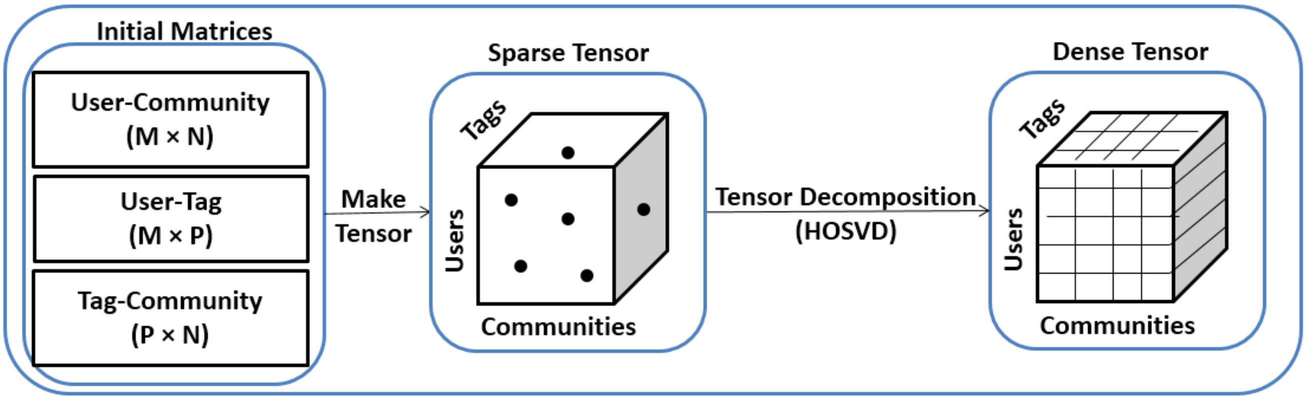


Fig.1. Rating Prediction using HOSVD

In tensor calculation, the result of collapsing a tensor $X_{U\times C\times T}$ on specific dimension T, is a matrix $R_{U\times C}$ which $r_{i,j} = \sum_{k=1}^{T} x_{i,j,k}$. Similar to what is shown on Fig. 2, collapsing the resulting HOSVD tensor on tag dimension provides User-Community weighting matrix. Basically, descending sort of each row of rating matrix and considering n highest values of predicted items, extracts the top-N recommendation for each user.

The recommendation list achieved form this solution, tends to be personalized, meaning that each predicted weight using HOSVD is based on the minimizing an error function with considering just the available ratings of the current user. Therefore, the resulted ranked list is also likely to be fully personalized and will not include the public factors of community such as entropy. Entropy of an item shows the heterogeneity of the attributes of the item. For instance, user entropy of a community shows how different the member users of the community are. And tag/topic entropy shows is a factor showing the diversity value of covered topics in the community.

4. Proposed Solution for Diversity Improvement

Diverse community recommendation is defined as: recommending a list of communities to user and satisfy him with groups from different genres, topics, people, even places, to join. This task can be formulated as ranking problem and predicting total order over communities. Usually it includes top-N communities (according to the rankings) proposed to the user as a recommending list, and user usually give more attention to the most top items of list rather than the end of list. The aforementioned satisfactions are defined as: Maximized diversity, limited to acceptable level of accuracy-loss.

HOSVD works based on Latent Semantic Indexing and solves several issues such as synonymy and ambiguity of words in text mining. In this work it is fitted and adopted to the tag-based recommendation area. This method handles prediction of the missed values of any row, which represents the potential preferences of the user. However, HOSVD predicts each value individually, and doesn't take

An International Journal JSCDSS
E-ISSN: 2289-8603

into account the global parameters of the data such as entropy of items.

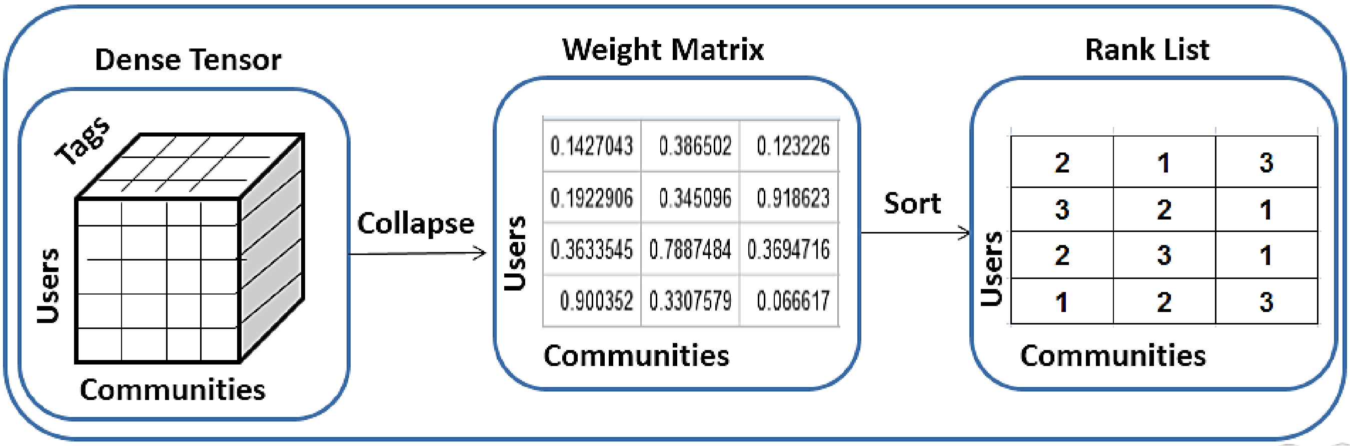


Fig. 2. Rank List Generation

Each community as an item to recommend, has two entropy value, shows its weight in terms of being different from the other items. Let's consider E_U and E_T as the user-entropy value and topic-entropy value respectively. E_U and E_T for each community c is calculated as below.

$$E_U(c) = -\sum_{u=1}^{m} \left(\frac{|member(c,u)|}{|m|}\right) \ln\left(\frac{|member(c,u)|}{|m|}\right) \tag{1}$$

$$E_T(c) = -\sum_{t=1}^{p} \left(\frac{|annotate(c,t)|}{|p|}\right) \ln\left(\frac{|annotate(c,t)|}{|p|}\right) \tag{2}$$

HOSVD-predicted weights of communities are personalized values which is different from user to user, whereas EU and ET are considered as impersonalized parameters of each community. The former, as a naïve accuracy-based parameter is exploited to provide the personalized accurate recommendation, while the latter can be used to prioritize the communities with higher entropy value which improves entropy-base diversity.

Based on this intuition, the new weighting scheme for communities including two facets of users and topics are proposed. It is worth to mention that descending sort of weights of communities for each user and excerption of top N highest ones, provides a personalized top-N recommendation.

$$entWeight(u,c) = Weight(u,c) + \alpha \times E_U(c) + \beta \times E_T(c) \tag{3}$$

In Eq. (3), α and β as tuning parameters handle the magnitude of user-entropy and topic-entropy weights which affect the original weights of communities. However this tuning parameter leads to different results for different datasets. Fig. 3 depicts the proposed entropy-weight including primary weight plus user-based and topic-based entropy values of communities.

5. Research Method

To evaluate the proposed method, we use a dataset from Last.FM music social network. Last.FM is a good instance of social media which use social tagging for its items, and also manage big amount of user-generated communities. Table 1 shows the statistics of selected Last.FM dataset used in this work.

Table 1
Last.FM dataset

	$\mathbf{A}_{\|User\|\|Community\|}$	$\mathbf{B}_{\|User\|\|Tag\|}$	$\mathbf{P}_{\|Tag\|\|Community\|}$
Rows	380	380	550
Columns	550	480	480
Non-Zeros	4820	37947	181765

For experiments, we select 30% of the known ratings in the dataset as the test set and 70% remaining ratings for training issue. Random sub sampling makes the training method similar to k-fold cross validation. However, we train the model for five times and make average for the results.

For accuracy measurement, we use the most common measure set including Precision (the rate of correct recommended items to all recommended items), and Recall (the rate of correct recommendation items to all relevant items). F-measure, which is defined as a harmonic mean of Precision and Recall is also used (Nilashi et al., 2014a; Nilashi et al., 2015a; Nilashi et al., 2015b). We consider F1-measure as a weighted average of the precision and recall. F1-measure is computed as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

An International Journal JSCDSS
E-ISSN: 2289-8603

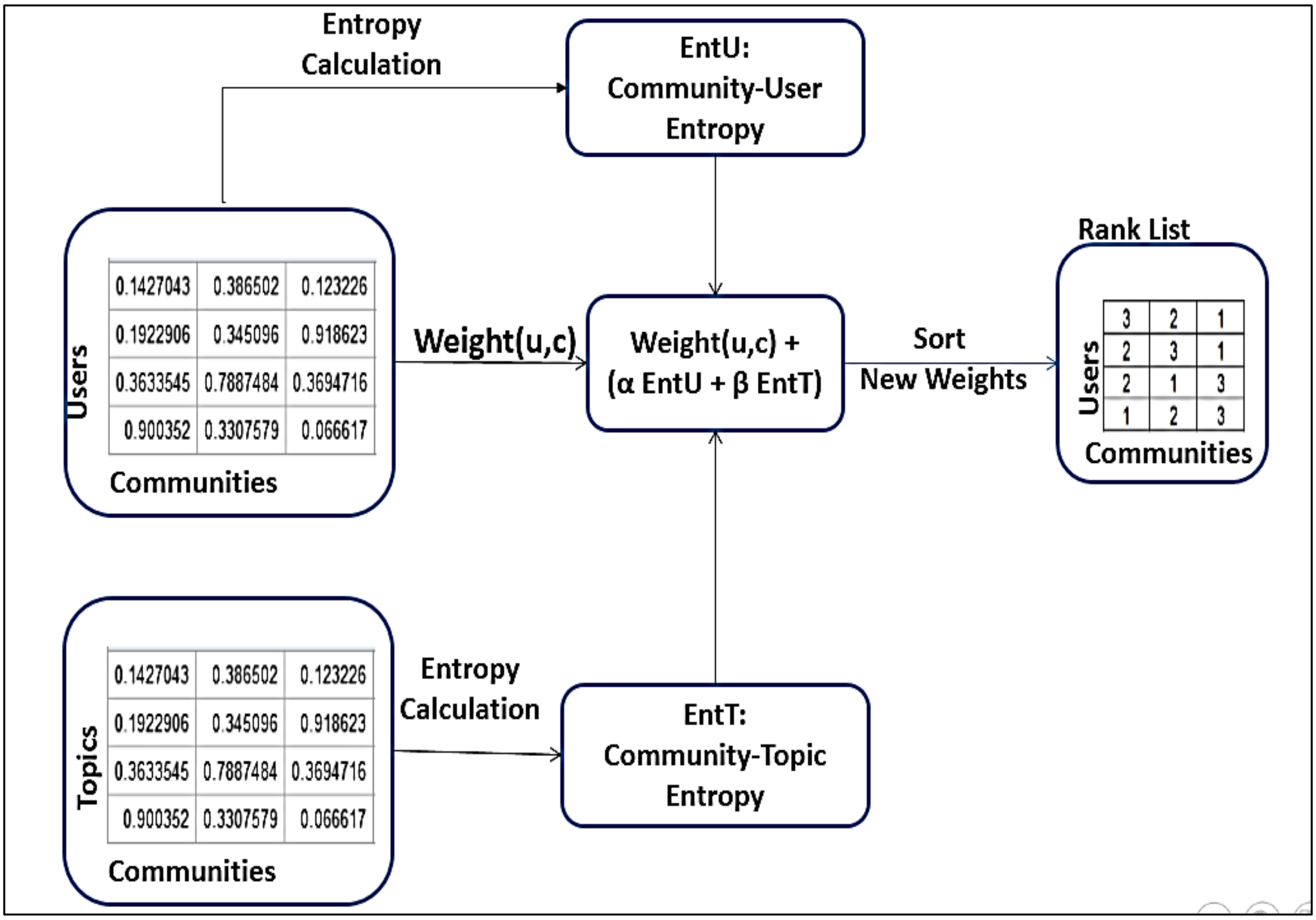


Fig. 3. Proposed Entropy-base weighting for Ranked List Generation

To study the effect of the proposed ranking method on recommendation diversity, we use a diversity metric, called Intra-List Diversity (ILD). ILD is defined as the inverse of a metric which is known as Intra-List Similarity (ILS) in the literature (Ziegler et al., 2005). ILD measures the dissimilarity of each pair of items in a list of specific user. Improvement in intra-list diversity helps user to receive diverse and heterogeneous list of recommendations. ILD is defined as

$$ILD(L) = 1 - \frac{2}{n|L|(|L|-1)} \sum_{i=1}^{n} \sum_{j,k \in L_i} 1 - s(j,k) \qquad (5)$$

where L is the recommendation list, n the number of users, and s similarity function.

As a complementary diversity metric, we use Shannon's Entropy value of recommended communities (Patil et al., 1982). Entropy for community membership shows how much each user behaved differently in joining communities. Higher entropy index means the heterogeneity of membership patterns is high, and lower entropy index means membership patterns for most of users are more homogeneous.

6. Results and Discussions

Standard recommendation process using HOSVD, starts with initializing the sparse tensor $\underline{\boldsymbol{X}}$ from three primary user-community, user-tag, and tag-community matrices. Then using HOSVD on the sparse data, provides low-rank approximation of the original tensor $\underline{\tilde{\boldsymbol{X}}}$. The usual type of recommendation includes N items in a ranked list, called top-N recommendation. Mapping the predicted ratings of user-tag-community in approximated tensor into two-dimensional user-community matrix, descending sort of weights, and cutting top N highest communities for each user provides a top-N community recommendation list.

To study the effect of entropy weighting scheme on accuracy and diversity of top-N recommendation, we included the user and topic entropy values of community into the predicted weights, as Fig. 3, and made another top-N list. Precession, recall, F1 measure, and Intra List Diversity index of Standard and Entropy-weight recommendation is depicted in Table 2.

JSCDSS E-ISSN: 2289-8603

Table 2

Accuracy and Diversity for Standard and Entropy-based Weighting Scheme

		Measure			
		ILD	Precision	Recall	F-1
Weight Scheme	Standard	0.4780	0.4491	0.6904	0.4631
	Entropy	0.4795	0.4305	0.6676	0.4504

The values of the above table is the average of factors for recommendation list from length=1 to 20. Fig. 4 compares the intra list diversity for different length of recommendation. As it is shown, the trend of ILD for the proposed method is like the standard method and stands above it. This is because of the effect of entropy as a public factor for each community on the personalized predicted weight of community. It means, the entropy acts as a coefficient on the standard weights, and try to leverage items with higher entropy value.

In the most of diversification methods, diversity improvement is gained at the expense of some accuracy-loss. The proposed method is not exceptional from this rule. Therefore, we have improvement in intra list diversity, but lose a part of accuracy. The point is to minimise the accuracy-loss while trying to achieve maximum diversity.

Fig. 5 shows the value of accuracy-loss during diversification for recommendation list from length=1 to 20. In fact the proposed method has tuneable parameters to select the casual threshold of diversity gain (and accuracy loss).

The proposed entropy-based weighting scheme in Eq. (3) is parameterized with factors α and β. These factors help user to set the desired threshold of diversification, as one likes to tune for higher diversity and doesn't care about losing the accuracy of recommender system.

As the last experiment to study the effect of entropy-based weighting on diversity, we compare the entropy of covered users and topics in total communities in a recommendation list. Higher entropy of users, shows the heterogeneity of recommended communities in term of dissimilarity between member users, while entropy of topics shows the diversity of covered topic of recommended communities. These two metrics are studied with different tuning parameters as reported in Fig. 6.

The algorithm with tuning weights of α=β=2 shows a smooth improvement in entropy-based diversity metrics. It also is at expense of losing average accuracy around 0.01 for all recommendations.

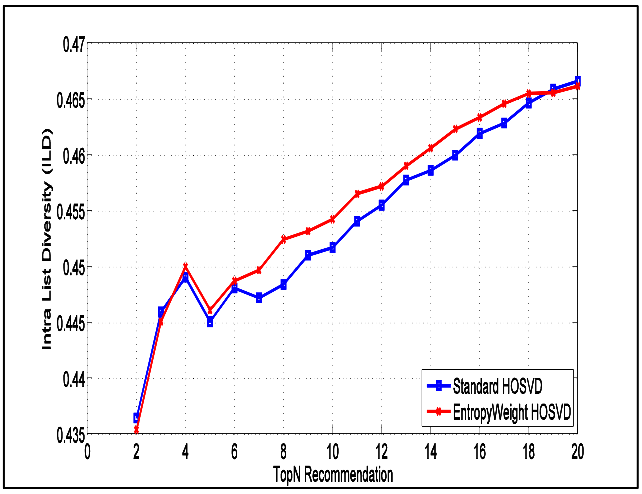


Fig. 4. Diversity-Gain Using Entropy-Weigh (α=β=2)

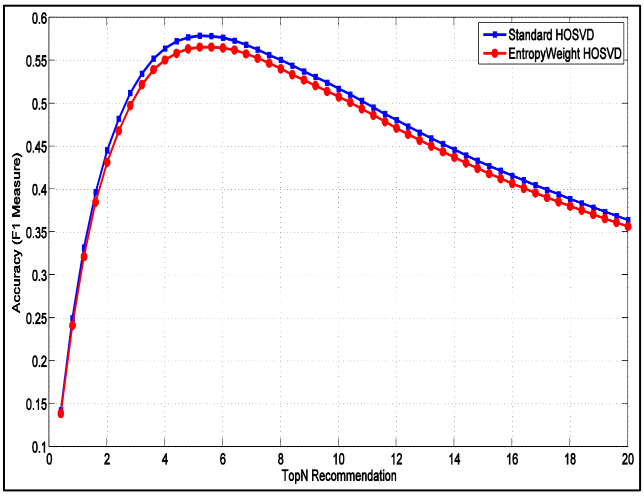
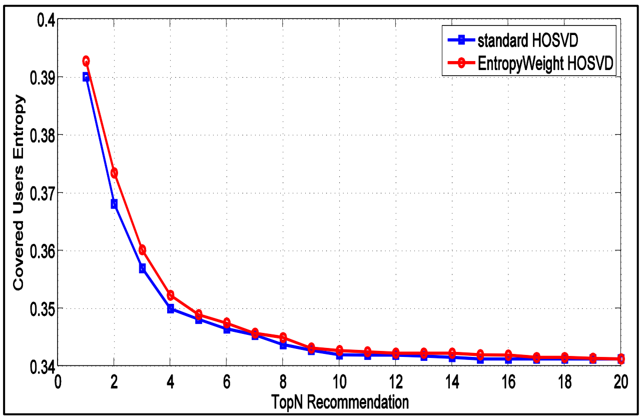


Fig. 5. Accuracy-Loss Using Entropy-Weigh (α=β=2)

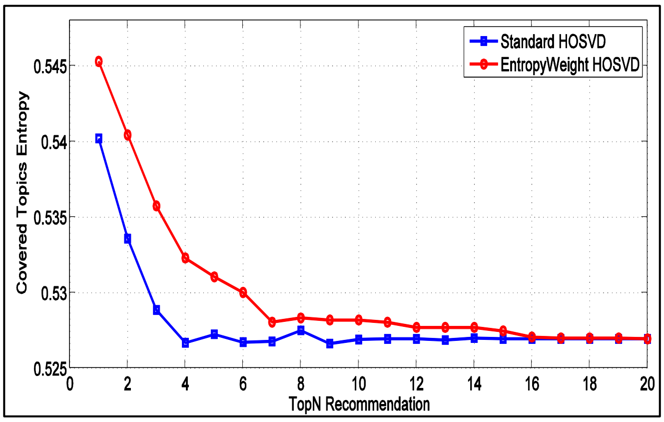
With tuning the parameter to higher values, we expect a higher diversity gain. As expected for α=β=5, diversity of entropy-based method shows much higher improvement compared to the standard method. As it is deducted from manipulating the tuning parameters, when giving more value to entropy of communities in weighting scheme, an acceptable improvement in entropy-based diversity is achieved. However, the accuracy-loss for the new experiment is about .04.

As a final conclusion for this work, we analysed the effect of including entropy of items as a public value of diversity, into the personalised predicted ratings. Unlike many other diversification approaches which re-rank the recommendation list for diversity improvement, this solution helps to improve diversity in the phase of weighting and before ranking top-N recommendation. Thus, the cost of re-ranking as a post-processing phase is reduced.

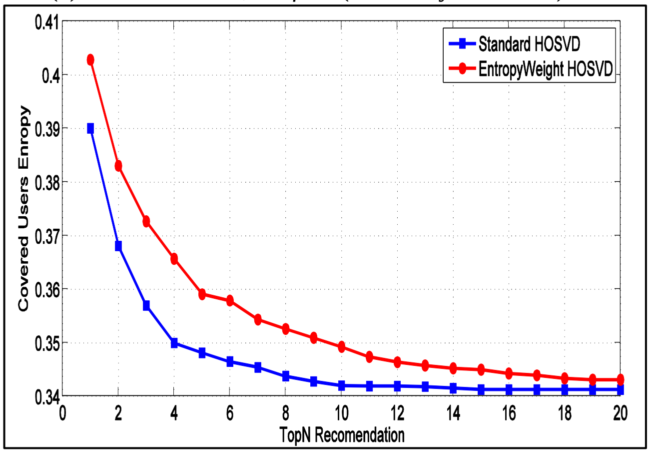
An International Journal JSCDSS
E-ISSN: 2289-8603



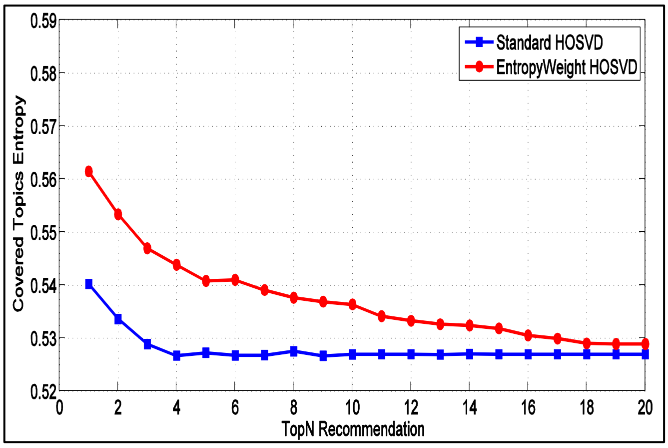
(a) covered users, α=β=2 (Accuracy Loss=.01)



(b) covered topics, α=β=2 (Accuracy Loss=.01)



(c) covered users, α=β=5 (Accuracy Loss=.04)



(d) covered topics, α=β=5 (Accuracy Loss=.04)

Fig. 6. Entropy of covered users and covered topics using Entropy-Weigh

References

Adomavicius, G., & Kwon, Y. (2012). Improving Aggregate Recommendation Diversity using Ranking-Based Techniques. IEEE Transactions on Knowledge and Data Engineering, 24(5), 896–911.

Auray, N., & Nationale, E. (2007). Folksonomy: the New Way to Serendipity. Available at SSRN 1009113, 65(65), 1–23.

Cao, Y., Diarmuid, Z., Daniele, Ó. S., Tamas, Q., Zhang, Y. C., Séaghdha, D. Ó., … Jambor, T. (2012). Auralist: introducing serendipity into music recommendation. In Proceedings of the fifth ACM international conference on Web search and data mining (pp. 13–22).

Castells, P., Vargas, S., & Wang, J. (2011). Novelty and diversity metrics for recommender systems: choice, discovery and relevance. In International Workshop on Diversity in Document Retrieval (DDR 2011) at the 33rd European Conference on Information Retrieval (ECIR 2011).

Ge, M., Delgado-battenfeld, C., & Jannach, D. (2010). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In Proceedings of the fourth ACM conference on Recommender systems (pp. 257–260).

Herlocker, J. L., Konstan, J. a., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems, 22(1), 5–53.

Milicevic, A. K., Nanopoulos, A., & Ivanovic, M. (2010). Social tagging in recommender systems: A survey of the state-of-the-art and possible extensions. Artificial Intelligence Review, 33(3), 187–209.

Nilashi, M., bin Ibrahim, O., & Ithnin, N. (2014a). Hybrid recommendation approaches for multi-criteria collaborative filtering. Expert Systems with Applications, 41(8), 3879-3900.

Nilashi, M., bin Ibrahim, O., & Ithnin, N. (2014b). Multi-criteria collaborative filtering with high accuracy using higher order singular value decomposition and Neuro-Fuzzy system. Knowledge-Based Systems, 60, 82-101.

Nilashi, M., bin Ibrahim, O., Ithnin, N., & Sarmin, N. H. (2015b). A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS. Electronic Commerce Research and Applications.

Nilashi, M., Ibrahim, O. B., Ithnin, N., & Zakaria, R. (2014c). A multi-criteria recommendation system using dimensionality reduction and Neuro-Fuzzy techniques. Soft Computing, 1-35.

Nilashi, M., Jannach, D., bin Ibrahim, O., & Ithnin, N. (2015a). Clustering-and regression-based multi-criteria collaborative filtering with incremental updates. Information Sciences, 293, 235-250.

Oestreicher-singer, G., & Sundararajan, A. (2010). Recommendation networks and the long tail of electronic commerce. Available at SSRN 1324064, 36(1), 65–83.

Patil, G. P., Taillie, C., & G. P. Patil and Taillie C. (1982). Diversity as a concept and its measurement. Journal of the American Statistical Association, 77(379), 548–561.

An International Journal
JSCDSS
E-ISSN: 2289-8603

Rendle, S., Schmidt-thieme, L., & Recommender, P. T. (2010). Pairwise interaction tensor factorization for personalized tag recommendation. In Proceedings of the third ACM international conference on Web search and data mining (pp. 81–90).

Steck, H., Labs, B., Hill, M., Management, H. D., Applications, D., & Bias, P. (2011). Item Popularity and Recommendation Accuracy. Proceedings of the 5th ACM Conference on Recommender Systems - RecSys '11, 125–132.

Symeonidis, P., Nanopoulos, A., & Manolopoulos, Y. (2008). Tag recommendations based on tensor dimensionality reduction. Proceedings of the 2008 ACM Conference on Recommender Systems - RecSys '08, 43.

Vargas, S., & Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. Proceedings of the Fifth ACM Conference on Recommender Systems - RecSys '11, 109.

Ziegler, C.-N., McNee, S. M., Konstan, J. a., & Lausen, G. (2005). Improving recommendation lists through topic diversification. Proceedings of the 14th International Conference on World Wide Web - WWW '05, 22.

An International Journal
JSCDSS
E-ISSN: 2289-8603