

# Clustering based approach for balancing accuracy and diversity in collaborative filtering

Apurva Pathak<sup>1</sup>, Bidyut Kr. Patra<sup>2</sup>, Ville Ollikainen<sup>2</sup>, and Raimo Launonen<sup>2</sup>

<sup>1</sup> Microsoft Research and Development Pvt. Ltd., Hyderabad, India.

<sup>2</sup> VTT Technical Research Centre of Finland, Finland.

E-mail: appatha@microsoft.com, {ext-bidyut.patra, ville.ollikainen, raimo.launonen}@vtt.fi

**Abstract.** In today's world of e-commerce, recommender system plays an important role for both individual user and business. Traditional recommender systems mainly emphasize on individual user satisfaction by recommending high quality items. However, recent study shows that diversity in recommendations is equally important factor from both user and business view points. Diversity in recommended items (individual diversity) provides a user to discover new (novel) items. On the other hand, recommending diverse range of items (aggregate diversity) prevent any item from becoming obscure in a large item space. While there are numerous works done on improving the recommendation accuracy. However, diversity in recommendation is often overlooked. In this paper, we explore clustering technique to generate diverse recommendations at both individual user and aggregate level while maintaining comparable level of accuracy. The results show high diversity gains while applying proposed approach on a real-world rating dataset (MovieLens).

**Keywords:** Collaborative Filtering, Accuracy, Individual diversity, Aggregate diversity, Clustering.

## 1 Introduction

Recommender system (RS) is an important tool to cope with information overload problem in the present internet era. Primary task of RS is to provide personalized suggestions of products or items to individual user so that user can select desired products or items directly without surfing over large products space. Collaborative Filtering (CF) is the most successful and widely used recommendation system [1, 2]. In CF, item recommendations to an user are performed by analyzing rating information of the other users and other items in the system. Traditional CF algorithms emphasize on predicting accurate ratings of the items suggested to individual user. Recent studies show that diversity in recommendations can be beneficial to not only users but also to business models. Diversity in recommendation can be viewed from two different perspectives *i.e.*, *user perspective (individual diversity)* and *business perspective (aggregate diversity)* [3]. Increasing in individual diversity helps an user obtain more idiosyncratic items in her recommended list while aggregate diversity of an RS increases selling of obscure items. However, both types of diversity in recommendation are achieved at the expense of accuracy.

In this paper, we propose an approach which explores hierarchical clustering to increase diversity (both individual and aggregate) of recommender systems. Proposed

**Table 1.** Results obtained from MovieLens dataset with density index 4.66%.

	RMSE	Aggregate Diversity (AD)	Individual Diversity (ID)	Increase in RMSE	Gain in AD	Gain in ID
RSVD	<b>0.7546</b>	32	0.0310	NA	NA	NA
Ranking Approach	0.7734	654	0.0562	2.49 %	<b>1943.75%</b>	80.82%
Proposed Approach	0.7571	643	0.0747	<b>0.39%</b>	1909.09%	<b>140.47%</b>

approach is tested with popular MovieLens data. Experimental results show that proposed approach gains significant diversity which is very close to the diversity obtained by ranking based approach, however, our approach suffers little accuracy loss compared to that of the ranking based approach [3].

## 2 Methodology

The proposed approach has two steps. In the first step, a standard matrix factorization technique such as regularized SVD (RSVD) is employed to obtain predictions of unknown ratings of items and we sort them in descending order in their predicted rating for each user in a system. It can be noted that this step is same as the step performed by many traditional RS which results in low diversity in the recommendation. Therefore, in the next step, we use hierarchical average-link clustering method to recommend top  $N$  items to individual user as follows.

We select a set of items  $S_u$  with predicted ratings more than a predefined threshold  $T$  for an active user  $u$  obtained in the first step. Subsequently, we apply hierarchical clustering technique to the set of selected items  $S_u$ . We reuse the feature vectors of items obtained from regularized SVD for computing similarity between a pair of items in  $S_u$ . We choose  $N$  clusters from the dendrogram formed by the hierarchical average-link clustering method and select one item with minimum predicted rating in each cluster to recommend  $N$  items to the active user  $u$ . We choose a threshold  $T$  to ensure minimal accuracy loss in our approach. Selecting items from each of the  $N$  clusters provide high individual diversity and reuse of global information (feature vectors) ensure high aggregate diversity of the recommender system.

## 3 Experimental Results

We used top 3000 users (rated maximum number of movies) and top 2000 movies (received ratings from maximum number of users) from MovieLens 1  $M$  dataset for experimental evaluation of the proposed approach. We compared our approach with the ranking based approach [3] and results are reported in Table 1. Experimental results show that our approach outperforms rank based approach in individual diversity (140.47%) and accuracy loss (0.39%) while it produces aggregate diversity very close to the diversity produced by that of the rank based approach (Table 1).

## 4 Conclusion and Future Work

We proposed an approach which provide significant individual and aggregate diversity maintaining high accuracy in a collaborative filtering system. Proposed approach can be used with traditional CF approach in order to provide high diversity. The work can be extended in utilizing state-of-art clustering technique in the second step of the approach.

## References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* **17**(6) (2005) 734–749
2. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. *Advances in Artificial Intelligence* **2009** (2009) 4:2–4:2
3. Adomavicius, G., Kwon, Y.: Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques. *IEEE Transactions on Knowledge and Data Engineering* **24**(5) (2012) 896–911