Multi-Aspect Collaborative Filtering based on Linked Data for Personalized Recommendation

Han-Gyu Ko
School of Computing, KAIST
291 Daehak-ro, Yuseong-gu
Daejeon, Republic of Korea
kohangyu@kaist.ac.kr

Joo-Sik Son
School of Computing, KAIST
291 Daehak-ro, Yuseong-gu
Daejeon, Republic of Korea
sonjoosik@kaist.ac.kr

In-Young Ko
School of Computing, KAIST
291 Daehak-ro, Yuseong-gu
Daejeon, Republic of Korea
iko@kaist.ac.kr

ABSTRACT

Since users often consider more than one aspect when they choose an item, relevant researches introduced multi-criteria recommender systems and showed that multi-criteria ratings add values to the existing collaborative-filtering-based recommender systems to provide more accurate recommendation results to users. However, all the previous works require multi-criteria ratings given by users explicitly while most of the existing datasets such as Netflix and MovieLens include only single-criterion ratings. Therefore, to take advantage of multi-criteria recommendation, there must be a way to extract necessary aspects and analyze users' preferences on those aspects from the given single-criterion type of dataset. In this paper, we propose an approach of utilizing semantic information of items to extracting essential aspects for performing multi-aspect collaborative filtering to recommend users with items in a personalized manner.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Selection process
http://www.acm.org/class/1998/

Keywords

Recommender System, Collaborative Filtering, Matrix Localization, Linked Data

1. INTRODUCTION

Collaborative Filtering (CF) based recommender systems utilize the feedback information that users provide for the items that they consumed or purchased to predict their preferences on items, and to recommend the users with new items based on the predicted preferences. The most popular examples of using CF-based recommendation include Netflix, which recommends movie contents, and Amazon.com, which recommends books, CDs, and various products to their users.

Most of the existing CF-based recommender systems use a single criterion, such as users' general ratings on items, to represent and analyze users' preferences on items in a two-dimensional user-item matrix. However, users often consider more than one aspect when they choose an item [1, 2]. Recently, many industries have begun employing multi-criteria ratings. For example, Zagat's Guide provides three criteria for restaurant ratings such as food, décor, and service. Online shopping malls such as Circuitcity.com and Buy.com also use multiple criteria such as display, performance, battery life, and cost for rating consumer electronics. Research

results [2-4] show that multi-criteria ratings add values to the existing CF-based recommender systems since they provide more accurate recommendation results to users. They all require multi-criteria ratings given by users explicitly. However, it is usually a burden for a user to provide more than one explicit feedback about an item. Even user feedback datasets are usually sparse even for the case of using a single criterion. We argue that this is the main obstacle of using multi-criteria ratings for the existing CF-based recommender systems.

To overcome this obstacle and take the advantage of multi-criteria recommendation, we propose an approach of utilizing semantic information of items to extracting essential aspects that can be used to perform multi-aspect collaborative filtering to recommend users with items in a personalized manner. The intuition is that we will be able to measure users' preferences on multiple aspects, which a user mostly considers in choosing items, by analyzing the common characteristics of the items that the user and other users of similar preferences consumed.

2. MULTI-ASPECT COLLABORATIVE FILTERING BASED RECOMMENDATION

As shown in Figure 1, the proposed approach provides four phases. Firstly, out of a given user-item matrix, we generate a set of concept groups by using keywords such as the words in the title of each item. We use Linked Data to extract semantic information about the keywords. A concept group is a group of semantically relevant terms, and the relationships among them that represent the essential semantics of a keyword. After generating concept groups, we analyze the common characteristics of them for each user by measuring their similarity in each aspect. We assume that domain aspects, such as story, acting and directing, which should be considered, are given. According to the given aspects, we identify users' preferences on items by counting the overlapped concepts in their concept groups for each of the aspects. Then, we find
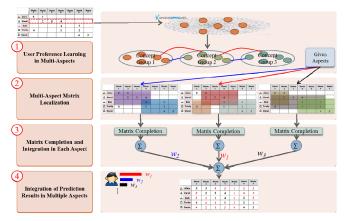

Figure 1. Overall process of the proposed approach

correlated users by measuring aspect similarity between users. Based on the correlated user groups, we divide the given user-item matrix into a set of sub-matrices that represent more focused groups of users and items in each aspect.

To predict users' ratings on new items, we perform matrix completion using regularized singular value decomposition (SVD) for each sub-matrix generated in the previous phase. We then integrate the results from the sub-matrices for each aspect. Lastly, we integrate the results generated for each aspect by assigning different weight values for different aspects to recommend users with new items.

3. EXPERIMENTAL RESULTS

In order to prove the effectiveness of the proposed approach, we conducted an experiment. We compared the prediction accuracy of the proposed approach against other CF-based recommendation methods to prove whether the proposed approach improves the prediction accuracy. Various CF-based methods such as user-based (UserBsd) and item-based (ItemBsd) methods which are the memory-based CF methods as well as regularized SVD (RegSVD), non-negative matrix factorization (NMF), probabilistic matrix factorization (PMF), Bayesian probabilistic matrix factorization (BPMF), and local low rank matrix approximation (LLORMA) which are model-based CF methods were selected as baselines.

For the experiment, we used the MovieLens dataset [1]. The MovieLens dataset that we used contains 1,000,209 ratings given by 6,040 users about 3,883 items (its rating density is 4.26%). In addition, we used Linked Movie Database and DBpedia in Linked Data to enrich the item information with appropriate semantics. The DBpedia dataset was selected as the primary knowledge base because it is highly interlinked with Linked Movie Database. We assumed that three aspects, genre, actor, and director, are given in this movie domain and obtained relevant data from the datasets. We split the MovieLens dataset into a training set and a testing set with the ratio of 8:2 respectively. To avoid bias in the experiment, we performed 5 fold cross validation.

We use mean absolute error (MAE) and root mean squared error (RMSE) as the evaluation criteria because they are the most popular methods to measure prediction accuracy. We also measured and compared the average precision and normalized discounted cumulative gain (NDCG) in the experiment. They are widely used to evaluate how accurate the ranking or ordering of recommendation results is. NDCG reflects the importance of considering the top-ranked items.

Table 1. Comparison of recommendation accuracy

Metric Dataset/Method		MAE	RMSE	Avg. Precision	NDCG @10
MovieLens	UserBsd	0.6941	0.8911	0.8340	0.9134
	ItemBsd	0.6725	0.8583	0.8420	0.9165
	RegSVD	0.6922	0.8686	0.8435	0.9174
	NMF	0.6925	0.8837	0.8271	0.9069
	PMF	0.8730	1.0494	0.7938	0.8886
	BPMF	0.6562	0.8409	0.8489	0.9196
	LLORMA	0.6941	0.8911	0.8340	0.9134
	MultiAspect	**0.5081**	**0.7182**	**0.8945**	**0.9601**

Table 1 shows recommendation performance of the baseline approaches and the proposed multi-aspect recommendation

---

¹ http://files.grouplens.org/datasets/movielens/ml-1m.zip

approach tested with the MovieLens dataset. Overall, the proposed approach, 'MultiAspect', shows better accuracy in predicting users' ratings on new items than other approaches. More specifically, the proposed approach outperforms the baseline approaches by around 0.1481 (22.57%) to 0.3649 (41.80%) in terms of MAE, and around 0.1227 (14.59%) to 0.3312 (31.56%) in terms of RMSE. The proposed approach also shows significant improvements in average precision and NDCG. More specifically, the proposed approach outperforms the baseline approaches in terms of average precision and NDCG around 0.0456 (5.37%) to 0.1007 (12.69%) and around 0.0405(4.40%) to 0.0715(8.05%) respectively. This implies that our approach is much more accurate at generating rankings of top *n* predicted ratings (in this experiment, *n*=10) than the baselines.

4. CONCLUSION

In this paper, to enable the multi-aspect collaborative filtering with the datasets that include only ratings on a single-criterion, we proposed an approach of enriching item information with concept groups that are generated by associating the items with relevant concepts retrieved from Linked Data.

The main contributions of our work can be summarized as follows. First, we proposed a framework of utilizing Linked Data to expand keywords extracted from item metadata to understand users' preferences on multiple aspects. Secondly, we proposed a way of identifying similar interests of users by comparing concept groups generated from a user-item matrix with given aspects. Lastly, we developed an effective way of aggregating the prediction results from sub-matrices in multiple aspects as a weighted sum, so that we personalize the recommendation results.

In our future research, we will make our approach more scalable. The proposed approach performs matrix completion, which time complexity increases exponentially as the number of users and items increases. We are currently investigating using cluster machines to run our approach in a concurrent manner.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Adomavicius, Gediminas, and YoungOk Kwon. "New recommendation techniques for multicriteria rating systems." IEEE Intelligent Systems 22.3 (2007). 48-55.

[2] Manouselis, Nikos, and Constantina Costopoulou. "Experimental analysis of design choices in multiattribute utility collaborative filtering." International Journal of Pattern Recognition and Artificial Intelligence 21.02 (2007): 311-331.

[3] Adomavicius, Gediminas, Nikos Manouselis, and YoungOk Kwon. "Multi-criteria recommender systems." Recommender systems handbook. Springer US, 2011. 769-803.

[4] Sahoo, Nachiketa, et al. "Collaborative filtering with multi-component rating for recommender systems." Proceedings of the Sixteenth Workshop on Information Technologies and Systems. 2006.