Normalization based Multi-Criteria Collaborative Filtering Approach for Recommendation System

Noor Ifada 1,*, Nur Fitriani Dwi Putri 2 dan Mochammad Kautsar Sophan 3

1) Department of Informatics, Faculty of Engineering Trunojoyo University of Madura
2) Rekayasa

DOI: https://doi.org/10.21107/rekayasa.v13i3.8545

ABSTRACT

A multi-criteria collaborative filtering recommendation system allows its users to rate items based on several criteria. Users instinctively have different tendencies in rating items that some of them are quite generous while others tend to be pretty stingy. Given the diverse rating patterns, implementing a normalization technique in the system is beneficial to reveal the latent relationship within the multi-criteria rating data. This paper analyses and compares the performances of two methods that implement the normalization based multi-criteria collaborative filtering approach. The framework of the method development consists of three main processes, i.e.: multi-criteria rating representation, multi-criteria rating normalization, and rating prediction using a multi-criteria collaborative filtering approach. The developed methods are labelled based on the implemented normalization technique and multi-criteria collaborative filtering approaches, i.e., Decoupling normalization and Multi-Criteria User-based approach (DMCUser) and Decoupling normalization and Multi-Criteria User-based approach (DMCItem). Experiment results using the real-world Yelp Dataset show that DMCItem outperforms DMCUser at most Top-N in terms of Precision and Normalized Discounted Cumulative Gain (NDCG). Though DMCUser can perform better than DMCItem at large Top-N, it is still more practical to implement DMCItem rather than DMCUser in a multi-criteria recommendation system since users tend to show more interest to items at the top list.

Keywords: Decoupling normalization, Item-based approach, Multi-criteria collaborative filtering.

INTRODUCTION

The multi-criteria collaborative filtering recommendation systems give their users preferences to rate items of their interest according to several criteria. In this case, they have the opportunity to express a detailed judgement towards items. Instinctively, users have diverse tendencies in rating items that some of them are quite generous while others tend to be pretty stingy. Given the dissimilar rating patterns, implementing a normalization technique in the recommendation system is advantageous to reveal the latent relationship within the multi-criteria rating data.

Researchers have conducted studies regarding the benefit of the normalization technique. Jin and Si (2004) have shown that the decoupling normalization technique performs better than the Gaussian when implemented on a single-criterion collaborative filtering approach. Meanwhile, Bilge and Yargıç (2017) also confirm the outperformance of the decoupling normalization technique when implemented on the multi-criteria collaborative filtering approach. However, existing studies were limited to the implementation of the user-based collaborative filtering approach only. This arouses a research gap since the collaborative filtering approach is categorized as two models, i.e., user-based and item-based (Aggarwal, 2016).

This paper studies two methods that implement the normalization based multi-criteria collaborative filtering approach. The framework of the method development consists of three main processes, i.e.: multi-criteria rating representation, multi-criteria rating normalization, and rating prediction using a multi-criteria collaborative filtering approach. Meanwhile, Bilge and Yargıç (2017) also confirm the outperformance of the decoupling normalization technique when implemented on the multi-criteria collaborative filtering approach. However, existing studies were limited to the implementation of the user-based collaborative filtering approach only. This arouses a research gap since the collaborative filtering approach is categorized as two models, i.e., user-based and item-based (Aggarwal, 2016).
multi-criteria collaborative filtering approach. The developed methods are labelled based on the implemented normalization technique and multi-criteria collaborative filtering approaches, i.e., Decoupling normalization and Multi-Criteria User-based approach (DMCUser) and Decoupling normalization and Multi-Criteria User-based approach (DMCItem). Series of experiments are conducted using the Yelp Dataset that contains the multi-criteria hotel ratings.

The contributions of this paper are: (1) A framework of normalization based multi-criteria collaborative filtering approach for recommendation system; and (2) Performance comparison between two multi-criteria recommendation methods that implement the decoupling normalization technique on the user-based and item-based models.

**METHOD**

The framework of the normalization based multi-criteria collaborative filtering approach for recommendation system consists of three main processes (as illustrates in Figure 1), i.e.: multi-criteria rating representation, multi-criteria rating normalization, and rating prediction using a multi-criteria collaborative filtering approach.

**Multi-criteria Rating Representation**

Let $U = \{u_1, u_2, ..., u_n\}$ be the set of $n$ users, $I = \{i_1, i_2, ..., i_l\}$ be the set of $l$ items, $C = \{c_1, c_2, ..., c_k\}$ be the set of $k$ criteria, and $Y = \{y_1, y_2, ..., y_t\}$ be the set of $t$ rating category. A multi-criteria rating data is formed when user $u$ rates item $i$ as rating category $y$ based on criterion $c$. In this case, the multi-criteria rating data can be represented as a multi-criteria rating matrix $R \in \mathbb{R}^{n \times l \times k}$ where $r_{uci}$ denotes the rating category given by user $u$ to item $i$ based on criterion $c$. Figure 2 shows an example of a multi-criteria rating matrix $R \in \mathbb{R}^{4 \times 3 \times 3}$ where $U = \{u_1, u_2, u_3, u_4\}$, $I = \{i_1, i_2, i_3\}$, $C = \{c_1, c_2, c_3\}$, and $Y = \{1, 2, 3, 4, 5\}$. Hence, we can observe that user $u_4$ has rated item $i_2$ as rating category $y_5$ based on criterion $c_1$, i.e.: $r_{111} = 3$.

**Multi-criteria Rating Normalization**

From the example in Figure 2, we can observe that the users have different tendencies in rating items. $u_1$ and $u_3$ are the type of users that tend to rate items quite generously while $u_2$ and $u_4$ are those that tend to rate pretty stingily. Given the diverse rating patterns, the purpose of implementing a normalization technique is to reveal the latent relationship within the multi-criteria rating data by transforming it into another range of rating scales (Bilge & Yargić, 2017).

This paper implements the decoupling normalization technique that had previously been proven to perform better than the Gaussian (Jin & Si, 2004) and Z-Score (Bilge & Yargić, 2017). The decoupling normalization is a technique that treats the ratings as ordinal numbers and transforms them into probabilistic scores. The transformation is generated based on the assumptions that (Jin, Si, Zhai, & Callan, 2003):

1. A user is more likely to prefer items rated as rating category $y$ when the user has rated many items as no more than rating category $y$.

**Figure 1. Framework of Normalization based Multi-Criteria Collaborative Filtering Approach for Recommendation System**

**Figure 2. Example of multi-criteria rating matrix $R \in \mathbb{R}^{4 \times 3 \times 3}$**
2. A user is less likely to prefer items rated as rating category \( y \) when the user has rated many items as rating category \( y \).

The decoupling normalized multi-criteria rating of user \( u \) towards item \( i \) according to criterion \( c \) is calculated as follows (Bilge & Yargıç, 2017):

\[
P(\text{\( y \) is preferred}) = P(r_{uic} \leq y) - \frac{P(r_{uic} = y)}{2}
\]

where \( P(r_{uic} \leq y) \) represents the proportion of items that are rated as no more than rating category \( y \):

\[
P(r_{uic} \leq y) = \frac{\#(r_{uic} \leq y)}{\#(r_{uic} = \emptyset) + \#(r_{uic} = y)} \text{ where } \forall i \in I \text{ and } \forall c \in C
\]

and \( P(r_{uic} = y) \) represents the proportion of items that are exactly rated with rating category \( y \):

\[
P(r_{uic} = y) = \frac{\#(r_{uic} = y)}{\#(r_{uic} = \emptyset)} \text{ where } \forall i \in I \text{ and } \forall c \in C
\]

These latent relationships cannot be revealed from the original multi-criteria rating.

**Rating Prediction using Multi-Criteria Collaborative Filtering Approach**

The multi-criteria collaborative filtering approach consists of 5 processes: (1) similarity per criterion, (2) similarity selection overall criteria, (3) rating prediction per criterion, (4) aggregation of rating prediction overall criteria, and (5) recommendation generation. This paper implements two well-known multi-criteria collaborative filtering approaches: Multi-Criteria User-based and Multi-Criteria Item-based. For ease of explanation, we name the method that implements the combination of Decoupling normalization and Multi-Criteria User-based approach as DMCUser; where the method that implements the combination of Decoupling normalization and Multi-Criteria User-based approach is labelled as DMCItem.

**Similarity per Criterion**

In DMCUser, the similarity between user \( u \) and user \( v \) per criterion \( c \) is calculated as follows:

\[
Sim(u, v)_c = \frac{\sum_{i \in I_u \cap I_v} \sigma_j d_{uic} \cdot (\mu_{uc} - \mu_{vc})}{\sqrt{\sum_{i \in I_u} \sigma_j d_{uic}^2} \cdot \sqrt{\sum_{i \in I_v} \sigma_j d_{vic}^2}}
\]

where \( I_u \) is the set of items that have been rated by user \( u \) and \( \mu_{uc} \) is the average of the normalized rating of user \( u \) based on criterion \( c \).

On the other hand, DMCItem calculates the similarity between item \( i \) and item \( j \) per criterion \( c \) as follows:

\[
Sim(i, j)_c = \frac{\sum_{u \in U_i \cap U_j} \sigma_k d_{uic} \cdot (\mu_{uc} - \mu_{uc})}{\sqrt{\sum_{u \in U_i} \sigma_k d_{uic}^2} \cdot \sqrt{\sum_{u \in U_j} \sigma_k d_{uic}^2}}
\]

where \( U_i \) is the set of users that have rated by item \( i \).

**Similarity Selection Overall Criteria**

The similarity selection is the process of determining which similarity is considered to be most significant to use given the list of the multi-criteria similarities. In this paper, we assume that the most significant similarity is the similarity of a criterion that has the lowest values overall criteria:

\[
Sim_{\min}(x_1, x_2) = \min_{1 \leq c \leq k} Sim(x_1, x_2)_c
\]

\[
Sim(x_1, x_2)_c = \begin{cases} Sim(u, v)_c, & \text{case DMCUser} \\ Sim(i, j)_c, & \text{case DMCItem} \end{cases}
\]
Once then similarities overall criteria are calculated, we may find the top-\(h\) similar users or items to form the user or item neighbourhood, depending on which collaborative filtering approach is implemented. For the case of DMCUser, \(Q_u(j)\) is formed as the set of top-\(h\) similar users to the target user \(u\) that have rated item \(j\). For the case of DMCItem, \(X_u(j)\) is formed as the set of top-\(h\) similar items to item \(j\) that have previously been rated by target user \(u\). Note that, in the rest of this paper, we denote \(h\) as the neighbourhood size where \(h \geq Q_u(j)\) in DMCUser or \(h \geq X_u(j)\) in DMCItem.

**Rating Prediction per Criterion**

The rating prediction per criterion in DMCUser is calculated as follows:

\[
\hat{r}_{uic} = \mu_{uc} + \frac{\sum_{v \in Q_u(j)} \text{Sim}(u,v) \cdot (\gamma_{uc} - \mu_{vc})}{\sum_{v \in Q_u(j)} |\text{Sim}(u,v)|}
\]

Meanwhile, the rating prediction per criterion in DMCItem is formulated as:

\[
\hat{r}_{uic} = \frac{\sum_{v \in X_u(j)} \text{Sim}(u,v) \cdot (\gamma_{uc} - \mu_{vc})}{\sum_{v \in X_u(j)} |\text{Sim}(u,v)|}
\]

**Aggregation of rating prediction overall Criteria**

The aggregation of rating prediction overall criteria is the process of taking into account the rating prediction of each criterion in a way such that the multiple rating scores are combined to form a single rating score. This paper implements the Weighted Linear Sum (WLS) technique (Barredo & Bosque-Senda, 1998) that uses the criteria weighted scoring to distinguish the importance of each criterion. Let \(W = \{w_1, w_2, ..., w_k\}\) be the set of \(k\) criteria weighted scoring that satisfies \(1 = \sum_{c \in C} w_c\) and \(|W| = |C|\). The aggregation of rating predictions is calculated as:

\[
\hat{r}_{ui} = \sum_{c \in C} w_c \cdot Y_{uic}
\]

where

\[
Y_{uic} = \begin{cases} \hat{r}_{uic} & \text{, case DMCUser} \\ \hat{r}_{uic} & \text{, case DMCItem} \end{cases}
\]

**Recommendation Generation**

The system generates the list of recommendations by offering the target user to the Top-\(N\) items that have the highest aggregated predictions scores.

**RESULTS AND DISCUSSION**

**Experiment Setup**

This paper uses the Yelp Dataset that contains the ratings of hotels based on four criteria: Overall, Useful, Funny, and Cool. We filter the dataset such that we only use data of users that have rated at least 3 hotels, results in 25,000 ratings from 2,595 users on 5,209 hotels. The rating category of each criterion is from 1 to 5 for Overall, from 0 to 110 for Useful, from 0 to 59 for Funny, and from 0 to 103 for Cool. We apply the 5-fold cross-validation as the evaluation method and use both the Precision and Normalized Discounted Cumulative Gain (NDCG) as the evaluation metrics.

**Sensitivity Analysis**

**Impact of Neighborhood Size**

The neighbourhood size \((h)\) is the parameter that controls the per criterion rating prediction formulated in Equation (8) for the DMCUser and Equation (9) for the DMCItem. In this paper, we vary \(h = \{5,10,20,30,40,50\}\).

Figure 5 shows the impact of \(h\) to DMCUser and DMCItem in terms of Precision and NDCG. We can observe that DMCUser performs the best when \(h = 5\) and deteriorates on larger \(h\). Meanwhile, DMCItems achieves its best performance at the largest \(h\), i.e., \(h = 50\). These
results indicate that there is a size of neighbourhood difference that should be used for tuning DMCUser and DMCItem to their best performances. In this case, DMCUser significantly needs less $h$ than DMCItem.

**Impact of Criteria Weighted Scoring**

The criteria weighted scoring ($W$) is the parameter that controls the aggregation of rating prediction overall criteria formulated in Equation (10). This paper tries a variation of $W$ (listed in Table 1) such that either each criterion has the same importance or only one of them is considered as the most important criterion while the others are set as equally essential. Note that we give a label to each variation of $W$ for ease of explanation.

Table 1. Variation of Criteria Weighted Scoring ($W$)

| Label | Weighted Scoring per criterion |
|-------|--------------------------------|
|       | $w_1$ | $w_2$ | $w_3$ | $w_4$ |
| $WLS1$ | 0.25  | 0.25  | 0.25  | 0.25  |
| $WLS2$ | 0.40  | 0.20  | 0.20  | 0.20  |
| $WLS3$ | 0.20  | 0.40  | 0.20  | 0.20  |
| $WLS4$ | 0.20  | 0.20  | 0.40  | 0.20  |
| $WLS5$ | 0.20  | 0.20  | 0.20  | 0.40  |

The impacts of $W$ to DMCUser and DMCItem in terms of Precision and NDCG are shown in Figure 6. We can observe that DMCUser and DMCItem achieve their best performances respectively on $WLS2$ and $WLS5$. These results indicate that DMCUser considers that the first criterion (i.e., Overall) is the most important one compared to the other three criteria. On the contrary, DMCItem assumes that the last criterion (i.e., Cool) is the most significant among others.

**Performance Comparison**

The performances of DMCUser and DMCItem are compared and fine-tuned based on their best parameters from the sensitivity analysis. In this case, DMCUser uses $h = 5$ and $WLS2$ while DMCItem uses $h = 50$ and $WLS5$. We analyze the comparisons based on a variation of Top-$N = \{1, 2, 3, \ldots, 20\}$.

Figure 7 shows the performance comparisons in terms of Precision and NDCG. We can observe that DMCItem outperforms DMCUser at most Top-$N$, i.e., until Top-$N = 16$ in terms of Precision and Top-$N = 17$ in terms of NDCG. In other words, DMCUser outperforms DMCItem.
only when Top-\(N\) is large. However, knowing that a target user usually tends to choose items that are at the top of the recommended list (Mohan, Chen, & Weinberger, 2011; Wang, Wang, Li, & He, 2013), it makes it more practical to implement DMCItem instead of DMCUser in the recommendation system. Additionally, these findings confirm those of the single-criteria collaborative filtering studies, i.e., that the item-based approach performs better than the user-based (Aggarwal, 2016; Gong, 2009; Ifada, Susanti, & Mula‘ab, 2019; Sarwar, Karypis, Konstan, & Riedl, 2001).

**CONCLUSION**

This paper analyses and compares the performances of two methods that implement the normalization based multi-criteria collaborative filtering approach for a recommendation system. The developed methods are labelled based on the implemented normalization technique and the multi-criteria collaborative filtering approaches, i.e., Decoupling normalization and Multi-Criteria User-based approach (DMCUser) and Decoupling normalization and Multi-Criteria User-based approach (DMCItem). Experiment results using the Yelp Dataset that contains the multi-criteria hotel ratings show that DMCItem outperforms DMCUser at most Top-\(N\) in terms of Precision and NDCG. Though DMCUser can perform better than DMCItem at large Top-\(N\), it is still more practical to implement DMCItem rather than DMCUser in a multi-criteria recommendation system since users tend to show more interest to items at the top list.

For further study, a possible plan is to implement other rating normalization techniques. Alternatively, we can also try to implement a fusion of the user-based and item-based multi-criteria collaborative filtering approaches to improve the methods developed in this paper.

**REFERENCES**

Aggarwal, C. C. (2016). *Recommender Systems: The Textbook*. Switzerland: Springer International Publishing.

Barredo, J. I., & Bosque-Sendra, J. (1998). Multi-Criteria Evaluation Methods for Ordinal Data In a GIS Environment. *Geographical Systems*, 5, 313-328.

Bilge, A., & Yargiç, A. (2017). Improving Accuracy of Multi-Criteria Collaborative Filtering By Normalizing User Ratings. *Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi A-Uygulamalı Bilimler ve Mühendislik*, 18(1), 225-237.

Gong, S. J. (2009). Employing User Attribute Item Attribute to Enhance The Collaborative Filtering Recommendation. *Journal of Software*, 4(8), 883-890.

Ifada, N., Susanti, S., & Mula‘ab. (2019). Impact of Imputation on Cluster-based Collaborative Filtering Approach for Recommendation System. *Kursor*, 10(1), 13-20.

Jin, R., & Si, L. (2004). *A study of methods for normalizing user ratings in collaborative filtering*. Paper presented at the The 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, United Kingdom.

Jin, R., Si, L., Zhai, C., & Callan, J. (2003). *Collaborative Filtering With Decoupled Models for Preferences and Ratings*. Paper presented at the The 12th International Conference on Information and Knowledge Management, New Orleans, Louisiana, USA.

Mohan, A., Chen, Z., & Weinberger, K. (2011). Web-Search Ranking with Initialized Gradient Boosted Regression Trees. *Journal of Machine Learning Research*, 14, 77-89.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). *Item-based Collaborative Filtering Recommendation Algorithms*. Paper presented at the The 10th International Conference on World Wide Web, Hong Kong.

Wang, Y., Wang, L., Li, Y., & He, D. (2013). A Theoretical Analysis of NDCG Ranking Measures. Paper presented at the The 26th Annual Conference on Learning Theory (COLT 2013), Princeton, NJ, USA.
