

学 位 授 与 申 請 書

2023 年 2 月 7 日

大阪府立大学長 様

大阪府立大学大学院 工学研究科 電気・情報系専攻

氏 名 KIM HAERANG

大阪府立大学学位規程第5条第1項の規定により
修士（工 学）の学位の授与を申請します。

（注意）

1. この申請書は、2通提出すること。

右肩の日付けは西暦年表示。

（A4）

目的関数ベースの Rough Membership C-Means クラスタリング

に基づく協調フィルタリング

Collaborative Filtering Based on Objective Function-Based

Rough Membership C-Means Clustering

| | | | |
|------------|--|------|-------------|
| 分野名 | 知能情報工学分野 | 氏名 | KIM HAERANG |
| Department | Computer Science and Intelligent Systems | Name | Haerang KIM |

Collaborative filtering (CF) is a technique for recommending desirable items to each user based on the dataset of users' preferences. Since users' preferences are based on human sensitivity, methods that can deal with the uncertainty inherent in data can be effective. Rough clustering can deal with the uncertainty by introducing rough set theory to hard C-means clustering (HCM). In rough clustering, various methods are proposed including generalized rough C-means, rough set C-means, and rough membership C-means (RMCM). In this study, we focus on RMCM which does not need to set the parameters for determining cluster centers.

However, since the original RMCM is a heuristic method, it is hard to discuss the validity of clustering. Therefore, we introduce objective function-based RMCM (RMCM2) which can be a basis for the discussion of the validity of clustering and further theoretical developments by introducing an objective function, which is designed to produce the same update rule for cluster center as RMCM.

In this study, we propose two types of CF methods based on RMCM2, namely, user-based RMCM2-CF and item-based RMCM2-CF. User-based RMCM2-CF extracts user-clusters by RMCM2 and recommends items that have high preference degrees in each user-cluster. On the other hand, item-based RMCM2-CF extracts item-clusters by RMCM2 and recommends each item to users that have high preference degrees in the item-cluster to which the item belongs.

Furthermore, we verified the recommendation performances of the proposed methods by applying them to real-world datasets, namely, NEEDS-SCAN/PANEL dataset and MovieLens-100k dataset, and comparing other CF methods, namely, HCM-CF, RMCM-CF, and memory-

based CF. We used ROC-AUC as an evaluation indicator of the recommendation performance.

From the result of numerical experiments, we confirmed that the performance of RMCM2-CF is the same or better than RMCM-CF, and better than HCM-CF by adjusting the roughness parameter τ and it maintains high performance regardless of the initial number of clusters, C . Also, we found that item-based methods have better performance than user-based methods if the data has enough number of items like MovieLens-100k dataset.

In summary, the results show that the consideration of the uncertainty by rough clustering is effective for CF tasks. We expect that our proposed methods can be a basis for better clustering-based CF algorithms by modifying the objective function. As a future task, we plan to introduce noise rejection scheme to RMCM2-CF to reduce the influence of outliers.

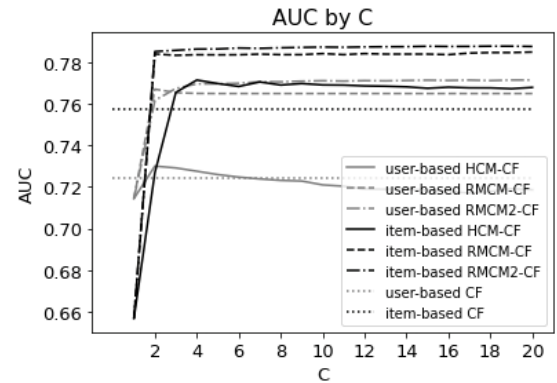


Figure 1: Changes in AUC by the initial number of clusters, C , in various methods in MovieLens-100k dataset.