

学位授与申請書

2024年 2月 9日

大阪公立大学長 様

大阪公立大学大学院 情報学研究科 基幹情報学専攻

氏 名 毛利 憲竜

大阪公立大学学位規程第5条第1項の規定により

修士（情報学）の学位の授与を申請します。

（注意）

- この申請書は、1通提出すること。

（A4）

注： 右肩の日付は西暦年表示。

粒状性を考慮したラフ集合ベースの混合多項分布型

共クラスタリングに基づく協調フィルタリング

Collaborative Filtering Based on Rough Set-Based Co-clustering Induced by Multinomial Mixture Models Considering Granularity

専攻名	基幹情報学専攻	氏名	毛利 憲竜
Department	Department of Core Informatics	Name	Kenryu MORI

In clustering-based collaborative filtering (CF), clusters of users with similar preference patterns are extracted, and items with high preferences within the cluster are recommended. Since data in CF tasks contain uncertainties arising from human sensibilities, represented as co-occurrence relationships between users and items, approaches such as rough clustering and co-clustering can be effective. Thus, rough co-clustering induced by multinomial mixture models (RCCMM) and its application to CF (RCCMM-CF) have been proposed.

However, RCCMM does not consider the granularity, which is an important viewpoint in rough set theory. Therefore, rough set CCMM (RSCCMM) was proposed as a rough co-clustering method that considers the granularity of the data.

In this study, we propose a CF method based on RSCCMM (RSCCMM-CF) that considers the granularity of the user space in CF tasks. Furthermore, we verified the recommendation performance and the effectiveness of the proposed method by comparing with conventional CF based on HCCMM (HCCMM-CF) and RCCMM-CF in real-world datasets, namely, NEEDS-SCAN/PANEL dataset and MovieLens-100k dataset. We used ROC-AUC as an evaluation indicator of the recommendation performances.

Figure 1 shows the changes in AUC by the initial number of clusters, C , for the three methods in the NEEDS-SCAN/PANEL dataset. It can be

confirmed that the results of the proposed method RSCCMM-CF are better than the conventional method HCCMM-CF and RCCMM-CF by adjusting the roughness parameter δ .

Also, in the MovieLens-100k dataset that underwent two types of preprocessing, we confirmed that RSCCMM-CF produced better recommendation performances than HCCMM-CF and RCCMM-CF and the recommendation performance was better when the data was binarized and the missing values were imputed with 0.5 than when the missing values were imputed with the average evaluation value of each user.

In summary, the results show that considering cluster overlaps and handling uncertainty based on granularity are effective in CF tasks.

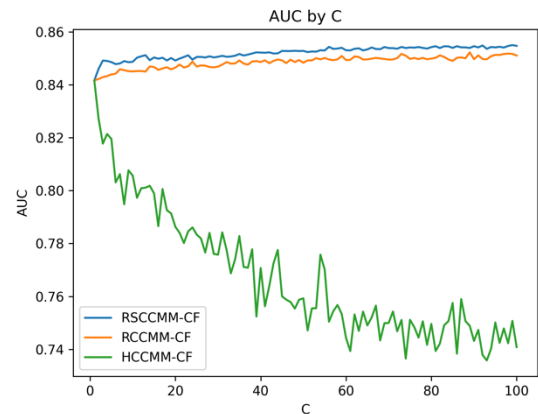


Figure 1: Changes in AUC by the initial number of clusters, C , for the three methods in the NEEDS-SCAN/PANEL dataset