

Personalized Rough-Set-based Recommendation by Integrating Multiple Contents and Collaborative Information

Ja-Hwung Su, Bo-Wen Wang, Chin-Yuan Hsiao and Vincent S. Tseng*

Department of Computer Science and Information Engineering, College of Electrical Engineering and Computer Science, National Cheng Kung University

tsengsm@mail.ncku.edu.tw

Information Sciences, 180 (2010), pp. 113-131. (SCI, Impact Factor: 3.095, Ranking: 8/99, COMPUTER SCIENCE, INFORMATION SYSTEMS)

To help users clarify what they prefer more easily, in this work, we propose a novel recommender named *FRSA* (*Fusion of Rough-Set and Average-category-rating*) that integrates multiple contents and collaborative information to predict user's preference based on rough-set theory. The major contribution of this paper is that our proposed recommender can successfully solve the traditional problems occurring in recent work, such as cold-start, first-rater, sparsity and scalability problems. The empirical evaluations reveal that the proposed method can associate the recommended items with user's interest much more effectively than other existing well-known ones in terms of accuracy.



The major efforts in this work are: 1) User Partition: Due to the huge amount of users, the prediction cost for CF is very expensive. Therefore, we group the users into several clusters to decrease the prediction cost. Without considering irrelevant users per prediction, of course the scalability problem can be solved, 2) Unknown Values Imputation: Whether for new items or new users, the prediction can work smoothly through imputing the unknown values by the discovered rules. Accordingly, the problems for cold-start, first-rater and sparsity can be prevented, 3) Items Reduction: Similar to User Partition, the computation cost will be reduced very much since irrelevant items are all eliminated before predicting unknown values, 4) Statistical-based fusion: On the basis of the variance of user's behavior, if the user's behavior is consistent, ACR (Average-Category-Rating) is a good candidate to generate the accurate results. Otherwise, RS (Rough-Set) can help the recommender infer the correct values.

The critical switch point between any two approaches is determined by the variance of user's behavior as:

$$\alpha = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Consider a set of the candidate approaches $\{ap_1, ap_2, ap_x, \dots, ap_y\}$. For the active user u_i , the mixed rating PR_{mixed} derived by the fusion paradigm is defined as:

$$PR_{mixed} = \begin{cases} \text{result of } ap_x, & \text{if } \alpha_i > \alpha_{thold} \\ \text{result of ACR}, & \text{otherwise} \end{cases}$$

Figure 1 shows some important aspects. First, if the user's behavior is consistent enough, ACR performs better than RS. On the contrary, RS can generate the better results when $\alpha_{threshold}$ is varied from 0.6 to 1. That is, RS outperforms ACR in handling the unstable users. Second, the best switch point for $\alpha_{threshold}$ is 0.5. Third, the lowest prediction error can reach 0.69 if the users' behaviors are not extreme diverse. It indicates that the coefficient of variation α is very important for detecting the stable preferences. To sum up, the results indicate that the idea for switching the prediction between RS and ACR is very helpful to bridge user's preferences and items.

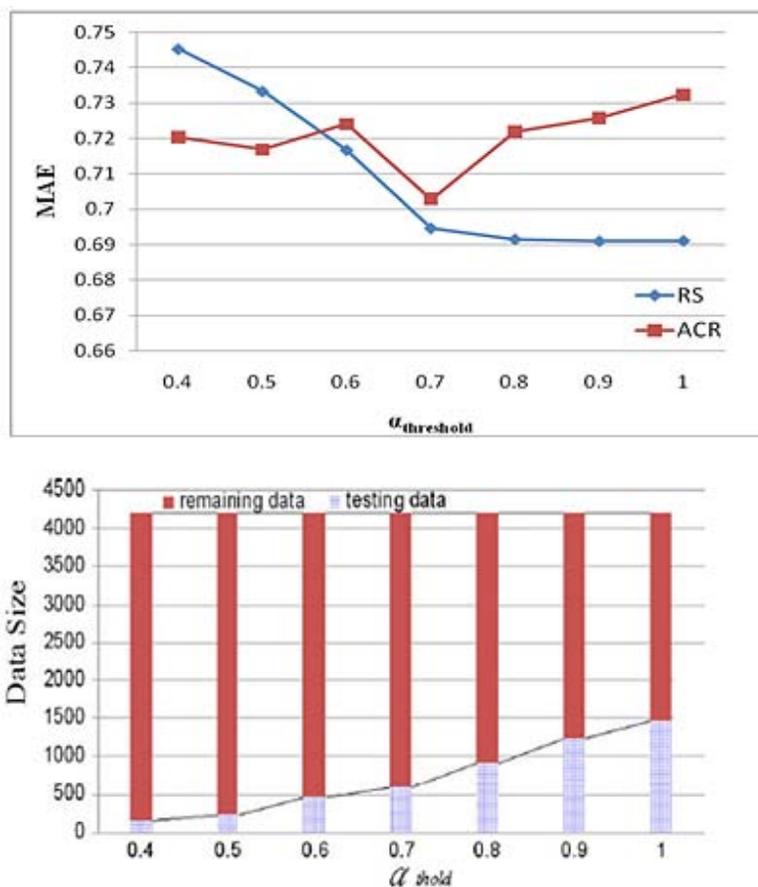


Figure 1. Comparisons between RS and ACR under different $\alpha_{threshold}$ and datasizes

Figure 2 shows the results if comparing: First, rule-based (AR) approach performs the worst. Second, the results of classification-based, probabilistic-based and statistics-based approaches are not better than those of KNN-based and social filtering-based approaches. In depth, SVM is the best classification-based approach and HMM is the best probabilistic-based approach. Furthermore, SVM is better than HMM. Similarly, item-based CF is better than both user-based CF and social filtering-based CF. In overall, set-based approach (RS) we propose is the best individual solution for personalized recommendation.

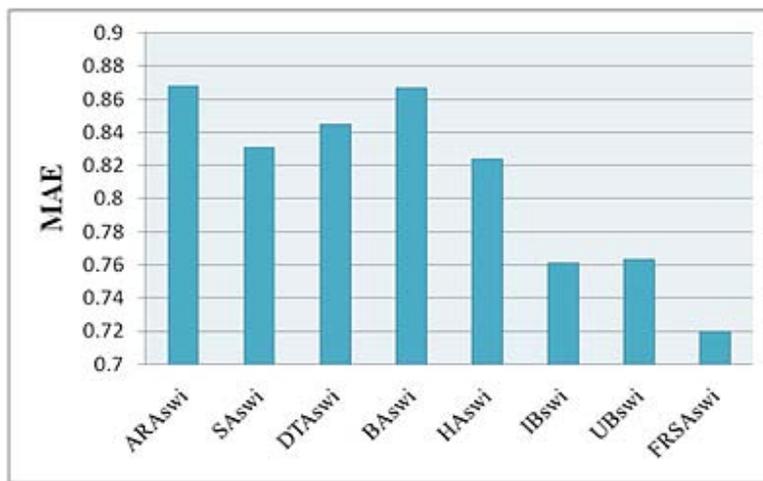


Figure 2. Comparisons between FRSA and other well-known recommenders

Figure 3 shows that, FRSA is the best one among the fusions. That is, whatever the fusion is, FRSA outperforms the other fusions if fusing association rules (AR) and average-category- rating (ACR).

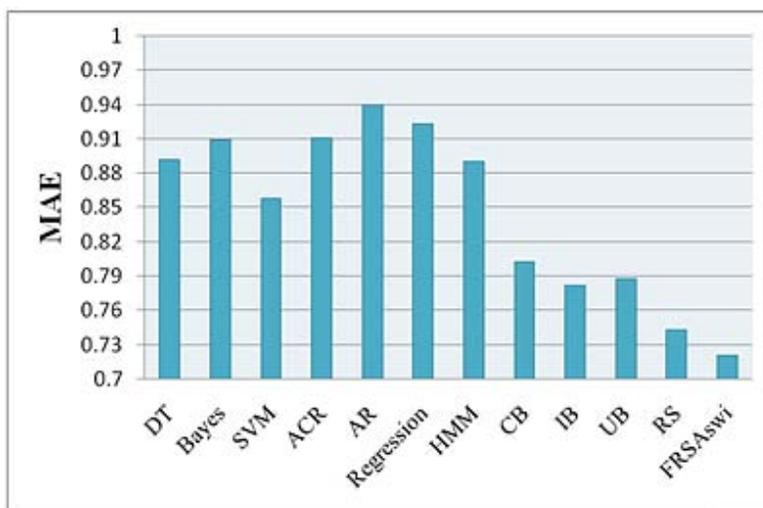


Figure 3. Comparisons between FRSA and other fusion approaches.

In the future, we will further investigate the optimal switch point by machine learning techniques.