Chapter 5

Conclusion

To solve the cold-start and sparsity problems in the recommendation systems on the internet, we developed the recommendation systems based on the visual clustering methods. The objective of these methods is to cluster the users-items in a binary image which is created from the transaction records. There are two visual clustering methods. The first visual clustering method is based on the genetic algorithm. The second visual clustering method based on the memetic algorithm. In the genetic and memetic algorithms, we designed two fitness functions to optimize the clustering methods. We proposed four recommendation systems by using the visual clustering methods with the two fitness functions, i.e., VCR-GA1, VCR-GA2, VCR-MA1, and VCR-MA2. We also proposed the other eight hybrid methods between our proposed methods with the common methods, i.e., VCR-GA1-UB, VCR-GA1-IB, VCR-GA2-UB, VCR-GA2-IB, VCR-MA1-UB, VCR-MA1-IB, VCR-MA2-UB, and VCR-MA2-IB. We used three synthetic and five real-world data sets to evaluate the performance of the clustering methods. We also used five real-world data sets to evaluate the performance of the top-*N* recommendation systems. In the real-world data sets, the KDD and TTS are the transaction of purchasing collected by Blue Martini Software and Thaiherbs-Thaimassage shop, respectively. The RCM is the transaction of the rated restaurants collected by the Department of Computer Science, National Center for Research and Technological Development in Mexico. We divided these three data sets into training and test sets using 10-fold cross validation. The

ECR is the transaction of the visited and rated restaurants collected by Robin Burke at the University of California on the UCI machine learning repository. We randomly selected customers and divided this data set into 80% training set and 20% test set. The MovieLens data set is the rated movies collected by GropLens Research Project at the University of Minesota. In this data set, the rated movies are mapped in to the binary images. The data set were randomly selected customers and divided into 80% training set and 20%.

To evaluate the performance of the clustering methods, we used the three synthetic data sets and the five real-world data sets. The results of the clustering on the binary images clearly show that all of the proposed visual clustering methods are able to properly cluster the information on both the synthetic and real-world data sets.

To evaluate the performance of the recommendation systems, we used the five real-world data sets. We compared our proposed methods with the four common methods, i.e., VCM-KM, UB, IB, and FB. We show the results by using the F-measure, i.e., precision, recall, and F1. In the sparsity problem, we calculated the sparsity level of each data set. The sparsity level of the KDD, TTS, RCM, ECR, and MovieLens are 0.9855, 0.9692, 0.9502, 0.9832, and 0.9369, respectively. To test the cold-start problem, each item is randomly selected for addition into the basket of the active user. There are sparsity problems in the KDD, TTS, RCM, and ECR data sets because the frequencies of the purchased items are too small. Hence, to evaluate the performance of the recommendation systems under the cold-start and sparsity problems, we used the KDD, TTS, RCM, and ECR data sets. The results on the KDD, TTS, RCM, and ECR data sets show that our proposed methods yield better performance than that of the common methods. The VCR-MA2 yields the best performance. Hence, we recommend the VCR-MA2 to solve the cold-start and sparsity problems. In the MovieLens data set, however, the results show that the VCR-MA2-IB and VCR-MA2 yield better performance than the common methods. The performance results of the VCR-MA2-IB and VCR-MA2 are very similar. Hence, we recommend the VCR-MA2 to improve the performance of the recommendation system on the MovieLens data set as well.