**Recommender system using collaborative filtering**

COMP9417 Assignment

|  |  |
| --- | --- |
| Peiguo Guan | z5143964 |
| Taoran Sun | z5150998 |
| Wanze Liu | z5137189 |
| Yunhe Hu | Z3351805 |

1. **Introduction**

In the era of internet, individuals are overwhelmed with huge amount of information online. This common situation has brought an obvious problem that internet users have difficult to choose the most appropriate services or contents. Nowadays, individuals are seeking more personalized services and experiences from service and content providers. Amazon, widely known as an online book reseller, deployed a recommender system in 1998 which can recommend books to its customers based on purchasing history, past preference and demographic information. As at today, recommendation systems are used not only for books, but also almost all types of internet entertainment services including movie, music and news.

The purpose of this assignment is to implement a collaborate filter recommender system based on a movie rating dataset and predict the highest rated movies for users. More specifically, the 1k dataset from MovieLens will be used to implement and test the method.

1. **Related Work**

The MovieLens dataset has been used by a large number of researchers to develop collaborate filtering system since 1999. Herlocker, Konstan, Borchers and Riedl (1999) conducted a study to solve the collaborate filtering problem by using similarity analysis and neighborhoods selection. Herlocker et al. also introduced two different methods to compute similarity, namely Pearson correlation coefficient and Spearman rank correlation. In addition, the effective of parameter normalization methods has been examined. The result indicates that average z-score normalization method is the best method for personalized recommender system. However, the average deviation from mean normalization method performs similar to the average z-score normalization method and performs much better for non-personalized recommender system.

Howe and Forbes (2008) conducted a further examination on the parameter normalization based on Herlocker et al.’s study. The result indicates that parameter normalization has a significant role when applying collaborate filtering algorithm and cosine vector similarity could be better than Pearson correlation coefficient similarity in some cases. The study also mentioned that different datasets favor different parameter normalization methods.

1. **Implementation**

The dataset used in this assignment is the ML-100K one downloaded from MovieLens website at <https://grouplens.org/datasets/movielens/>. There are several different data files in this dataset including user rating, user demographic information, user occupation and movie information.

1. **Result**
2. **Conclusion**
3. **Future Work**

**Reference**

Herlocker, J. L., Konstan, J. A., Borchers, A. & Riedl, J. 1999, ‘An algorithmic framework for performing collaborative filtering’, in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 1999, pp. 230-237.

Howe, A. E. & Forbes, R. D. 2008, ‘Re-considering neighborhood-based collaborative filtering parameters in the context of new data’, in *Proceedings of the 17th ACM conference on Information and Knowledge management*, Napa Valley, California, USA, October 26-30, 2008, pp. 1481-1482.