A Collaborative Filtering Framework Based on Both Local User Similarity and Global User Similarity⋆

Heng Luo, Changyong Niu, Ruimin Shen, and Carsten Ullrich

Department of Computer Science and Technology, Shanghai Jiao Tong University
1954 Huashan Road Shanghai 200030, China
{hengluo, cyniu, rmshen, ullrich_c}@sjtu.edu.cn

Collaborative filtering as a classical method of information retrieval is widely used in helping people to deal with information overload. In this paper, we introduce the concept of local user similarity and global user similarity, based on surprisal-based vector similarity and the application of the concept of maximin distance in graph theory. Surprisal-based vector similarity expresses the relationship between any two users based on the quantities of information (called surprisal) contained in their ratings, which is based on the intuition that less common ratings for a specific item tend to provide more discriminative information than the most common ratings. Furthermore, traditional methods of computing user similarity can not work if two users have not rated any identical item. To solve this problem, the global user similarity is introduced to define two users being similar if they can be connected through their locally similar neighbors. A weighted user graph is first constructed by using local similarity of any two users as the weight of the edge connecting them. Then the global similarity can be calculated as the maximin distance of any two nodes in the graph. Based on both of Local User Similarity and Global User Similarity, we develop a collaborative filtering framework called LS&GS. An empirical study using the MovieLens dataset shows that the proposed framework outperforms other state-of-the-art collaborative filtering algorithms.

References

1. Luo, H., Niu, C., Shen, R., Ullrich, C.: A collaborative ltering framework based on both local user similarity and global user similarity. Machine Learning 72(3), 231–245 (September 2008)

⋆ This is an extended abstract of an article published in the Machine Learning journal [1].