Ethics Pledge

Consistent with the above statements, all homework exercises, tests and exams that are designated as individual assignments MUST contain the following signed statement before they can be accepted for grading.

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature: Haodong Zhao Date: Apr 7th. 2019

Please note that assignments in this class may be submitted to

www.turnitin.com, a web-based anti-plagiarism system, for an evaluation of their originality.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Reading review**

**A Comparative Study of Collaborative Filtering Algorithms**

This article compares several classic or recent collaborative filtering techniques in various experimental environments and uses this comparison to determine which algorithms work well under what conditions.

Collaborative filtering is a rapidly evolving research area including:

1. Classic method: neighborhood method
2. The latest method: decomposition around the matrix

This paper compares the classic are the latest methods by controlling the number of users, the number of projects, the degree of sparsity, and considering multiple evaluation indicators and calculations costs. The conclusions are as follows:

1. In general, matrix decomposition-based methods perform best in terms of prediction accuracy.
2. The prediction accuracy of different algorithms depends on the number of users, the number of items and the density, and the nature and extent of the dependencies vary from algorithm to algorithm.
3. There is a complex relationship between prediction accuracy, variance, computation time, and memory consumption, which is important for choosing the right algorithm.

There are two types of collaborative filtering systems, memory-based and model-based.

1. The most popular memory-based CF method is a neighborhood-based approach, and by rating users that are similar to the querying user or similar items with similar items, they will have similar ratings for the remaining items.
2. The model-based approach adapts the parametric model to the training data. Includes cluster-based CF, Bayesian classifiers, and regression-based methods.

Through the design experiment, the author has reached the following conclusions:

1. Matrix decomposition based-methods usually have the highest accuracy. Memory-based methods have no special advantages other than simplicity.
2. The accuracy of all algorithms varies according to the number of users, the number of items and the density.
3. There is a trade-off between higher accuracy and other factors, and more accurate algorithms tend to be highly dependent on data and size and density, with higher accuracy variance. Lower computational efficiency and more suitable when the computational efficiency is not so important. Shen the computational efficiency is important, the slope-one may be a better choice.