

# Project Proposal for 4995 PFP

Zhou Zhuang

## Introduction

Collaborative Filtering has been a popular choice among recommendation algorithms. The core idea of collaborative filtering is that if two users share similar interests historically there's higher chance that they would also like other items in each other's interest list that they are not aware of. One implementation of collaborative filtering is memory-based collaborative filtering, where a similarity matrix would be calculated for the users. When a user's preference towards an item he hasn't seen is being predicted, this value is averaged over the ratings on this item from his similar users.

## Data

I selected the MovieLens 1M dataset [1] for this project. It contains 1 million ratings from 6000 users on 4000 movies.

## Parallel Methodology

Several parallel methods have been developed. I chose Slope One [2] as my main reference. Slope One implements parallel collaborative filtering with a shared memory map and multiple threads, each calculating different parts of the similarity matrix and write to the shared map. This algorithm achieves 6x acceleration on MovieLens dataset when the number of threads is set to 8. I will refer to its main idea and implement it in Haskell for this project.

## Evaluation

The main metric for evaluation for this project would be the time cost to calculate the similarity matrix. If time allows I would also implement these features: prediction, evaluation with different similarity functions, scale to the MovieLens 20M dataset, etc.

## Reference

[1]. <https://grouplens.org/datasets/movielens/1m/>

[2]. Efthalia Karydi, Konstantinos Margaritis. Multithreaded Implementation of the Slope One Algorithm for Collaborative Filtering. 8th International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2012, Halkidiki, Greece. pp.117-125, ff10.1007/978-3-642-33409-2\_13ff. fhal01521419f