Group Project Proposal
A Learning-based Movie Recommender System

张皓天	张润博	赵尚澄	李佳璇	包叶琳
元培学院	元培学院	心理与认知科学学院	经济学院	心理与认知科学学院

Date: October 7, 2022

1 Project Descriptions

Recommender systems, also known as recommendation systems, is a subclass of information filtering system that provide suggestions for items that are most pertinent to a particular user. They are particularly useful when an individual needs to choose an item from a potentially overwhelming number of items that a service may offer. Recent years have seen the great progress of recommender systems and it is widely applied for business use such as video sharing platforms like Youtube and Bilibili and online purchasing platforms like Amazon and Taobao. For our project, we want to specify on movie recommendations and try to implement learning-based algorithms in order to get a glimpse of situations and prospects of this technique.

2 Datasets to Use

2.1 MovieLens 20M dataset

2.1.1 Source

kaggle: https://www.kaggle.com/datasets/grouplens/movielens-20m-dataset

2.1.2 Description

The datasets describe ratings and free-text tagging activities from MovieLens, a movie recommendation service. It contains 20000263 ratings and 465564 tag applications across 27278 movies. These data were created by 138493 users between January 09, 1995 and March 31, 2015. These data includes brief descriptions, tags by users and officials and user ratings of movies.

3   Potential Methodologies

The main issue studied in the field of recommender system is to learn a mapping from the set of users to a variety of commodities based on users' appetite. Basically, all the algorithms for recommender system can be involved into content-based filtering, collaborative filtering and the hybrid of these two. We would like to apply both traditional ML methods and deep learning methods to realize a recommender system.

3.1   Traditional Machine Learning Methods

Traditional machine learning methods in recommender systems, as far as we know, includes baseline algorithms, k-NN inspired algorithms, Matrix Factorization-based algorithms and Slope One algorithms. We plan to implement it both on movie datasets and compare the recommend result. A useful toolkit to apply is Surprise, an easy-to-use Python scikit for recommender systems.

3.2   Deep Learning Methods

Content-based and collaborative filtering models are relatively linear systems that cannot deliver such deep user insights. However, deep learning models have the capability of learning from multiple levels of data representations and abstractions. Commonly used DL models includes DNN, CNN including attention mechanisms, Restricted Boltzmann machine, RNN, LSTM, Variational Autoencoders and so on. For us, we are interest in LSTM and VAE.We plan to implement the model on MovieLens and compare the result of traditional and DL methods, for there are some papers indicating that traditional ones can reach or even surpass the performance of DL ones with careful fine-tuning.

3.3   Measurement

We may use recall, precision and coverage to measure the performance of the model.

4   Work Division

We divide our group into 2 small groups. The first group, involving 李佳璇 and 包叶琳, study for traditional algorithms and the second group, involving 张皓天, 张润博 and 赵尚澄, endeavor to apply deep learning algorithms. The two groups will study separately until some day at November when the whole group discuss the results and compare the capability of traditional and DL algorithms.