NTDS Project Summary - Team 7

Jinhui Guo, Hiroki Hayakawa, Kuan Tung, Chun-Hung Yeh

1 Introduction

It is always great to spend a couple of hours watching a good entertaining movie. However, finding a good movie is not an easy task. There are many other users' ratings on the internet, but they are not always reliable. Thus, it will be great if there is a system which recommends movies to you according to your preference.

The purpose of our project is to design a personalized movie recommendation system and evaluate its performance. To be more specific, we would like to use graph neural networks to build the system, which will predict users' rating to a unseen movie based on their ratings to watched movies and pick up new movies with high rating score as recommendation to the user.

2 Dataset

We will use MovieLens Latest [1]. It contains 27753444 ratings and 1108997 tag applications across 58098 movies created by 283228 users between January 09, 1995 and September 26, 2018. Movie titles, year of release and genres are included as well. It also includes tag genome, which is a data structure that contains tag relevance scores for movies. The score indicates how strongly a tag apply to a movies.

Table 1: Dataset Summary

Dataset	Users	Movies	Ratings	Density	Rating Levels	Tags
MovieLens 20M	2.8×10^5	5.8×10^4	2.8×10^6	1.7×10^{-4}	0.5, 1, ..., 5	1.1×10^6

3 Plans

Create following three graphs.

Table 2: Graphs

Graph Movies	Nodes titles	Edges similarity	Features name, release date, genre
User-Movie	user / movie	Bipartite ratings	-

- Explore some properties of the graphs (e.g. degree distribution, connected components, diameter)
- Build the recommendation system with graph neural network, predicting the existence of edges with a certain value of ratings in the bipartite graph between users and movies based on unsupervised learning methods. Specifically, we exploit two methods [2] and [3] to do this. For the first one, we apply multi-graph convolutional neural network and LSTM network for matrix factorization. For the other one, utilizing graph convolutional network, we design graph encoder and decoder to make predictions.
- Combine the features of users and movies with the recommendation system to refine its accuracy (Scrape more data (e.g. casts, crews) from IMDb or TMDb to enrich the features as necessary).
- Perform dimensionality reduction, visualize the resulting embeddings and observe if there are any meaningful clusters.
- Compare our result of the recommendation system to other methods (e.g. matrix factorization)

4 Potential Problems

- Unlike movies, no information about users is provided.
- The dimension of casts or crews data could be very high.

5 References

[1] "Movielens latest datasets." [Online]. Available: http://files.grouplens.org/datasets/movielens/ml-latest-README.html

[2] F. Monti, M. M. Bronstein, and X. Bresson, "Geometric matrix completion with recurrent multi-graph neural networks," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. USA: Curran Associates Inc., 2017, pp. 3700–3710. [Online]. Available: http://dl.acm.org/citation.cfm?id=3294996.3295127

[3] R. van den Berg, T. N. Kipf, and M. Welling, "Graph convolutional matrix completion," 2017.