Movie Recommendation System Using Deep Neural Network Based on Historical Rating Records

Team Members: Jiawei Yu, Yuqi Wang, Dantong Chen

Introduction

A recommendation system is a type of information filtering system which attempts to predict the preferences of a user, and make suggests based on these preferences. [1]

In this project, we built a movie recommendation system using Deep Learning method. We began with building a deep neural network to predict the most possible rating given by a specific user on a movie. The inputs are user matrix, movie matrix, and movie genre matrix. Then, for a given user as the input, we found a set of movies which haven't rated by this user and predicted the most possible ratings on these movies by this user. Finally, we will recommend top 5 movies with 5 highest predicted ratings to the input user.

In addition, we compared the performance of our model with one of the models in state of art on the MovieLens dataset and found out our model slightly decreased the mean squared error by adding one input layer -- movie genre.

Data

The data we used describes 5-star rating and free-text tagging activity from MovieLens. Since we train the model locally, we used the smaller dataset including 100836 ratings and 3683 tag applications across 9742 movies. These data were created by 610 users from March 29, 1996 to September 24, 2018. This dataset was updated on September 26, 2018.

In the datafile, users were randomly selected. All selected users had rated at least 20 movies. No demographic information is included.

We used the following 3 datasets:

- ratings.csv: userId, movieId, rating, timestamp
- tags.csv: userId, movieId, tag, timestamp
- movies.csv: movieId, title, genres

Methodology

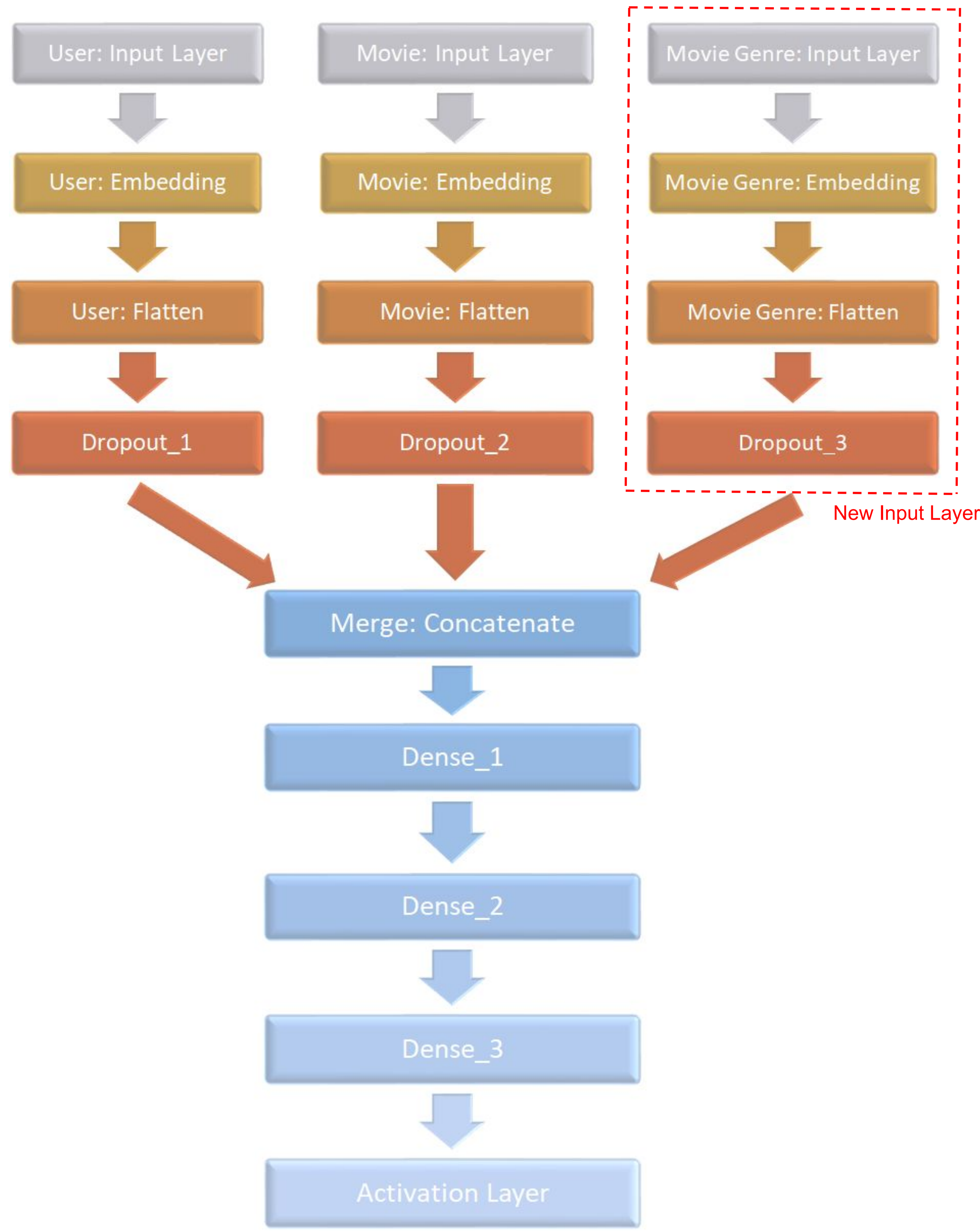

Fig 1. Our neural network

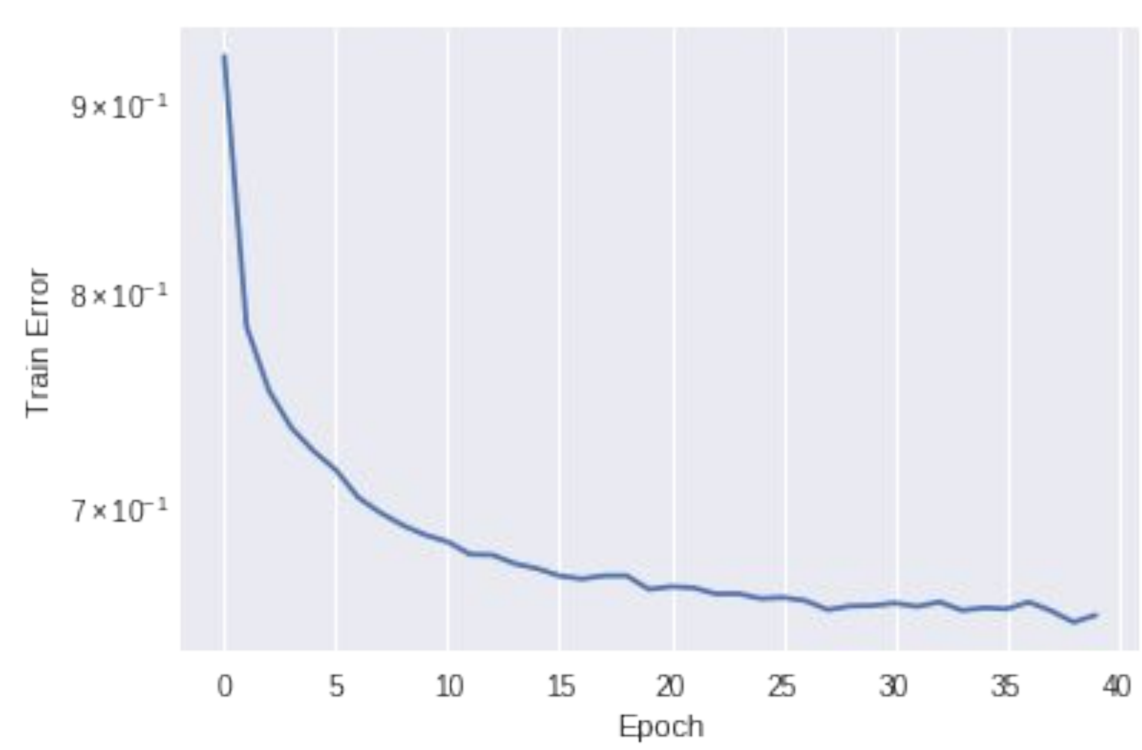

Fig 2. Change in loss while training our new model

Result

Old Model:
Mean Absolute Error: 0.66702
Mean Squared Error: 0.76486

New Model:
Mean Absolute Error: 0.65754
Mean Squared Error: 0.75325

	userId	movieId	rating	timestamp	title	genres	Predicted
60347	18	1355	3.0	965706276	Zero Effect (1998)	731	3.20989
87019	198	73	3.0	940544216	Things to Do in Denver When You're Dead (1995)	769	2.53266
23354	302	2096	5.0	1053302810	Airplane! (1980)	634	4.01753
18496	560	1053	3.0	1491095552	Star Trek: The Motion Picture (1979)	508	2.70716
89535	367	1191	2.0	975828061	Cop Land (1997)	237	2.84386

Fig 3. Prediction result on the test data

	userId	movieId	timestamp	title	genres	Predicted
18812	55	4039	1306463708	Galaxy of Terror (Quest) (1981)	312	5.13407
94255	55	3500	1000193988	Phantom of the Paradise (1974)	699	5.06762
86917	55	9078	1520409581	L.A. Slasher (2015)	651	4.99024
86631	55	7383	1520409088	Get Low (2009)	684	4.96813
86805	55	8605	1520409188	National Lampoon's Bag Boy (2007)	634	4.93645

Fig 4. Top 5 movies recommendation for the user 55

	userId	movieId	rating	timestamp	title	genres
388	55	43	5.0	835799219	Seven (a.k.a. Se7en) (1995)	937
2458	55	314	5.0	835799274	Forrest Gump (1994)	690
26001	55	302	5.0	835799081	Ace Ventura: Pet Detective (1994)	634
27757	55	18	5.0	835799219	Ace Ventura: When Nature Calls (1995)	634
4336	55	510	5.0	835799139	Silence of the Lambs, The (1991)	786

Fig 5. Top 5 ratings movies of the user 55

Discussion

- The mean absolute error and the mean squared error are both decreased with the model we added the genres of movies feature, compared to the model in state of art.
- The result shows that the model will recommend some movies with the same genres to the specific user.
- In the future, we could gather the user information like gender and age as other input embedding features to build a new recommendation model.

Reference

[1] https://pdfs.semanticscholar.org/767e/ed55d61e3aba4e1d0e175d61f65ec0dd6c08.pdf
Datasets: https://grouplens.org/datasets/movielens/
Model: https://nipunbatra.github.io/blog/2017/recommend-keras.html