A Hybrid Neural Model for Recommender System

Yuanhang Cui[1] **Weijie Sun[1]** **Shiyu Wang[1]**

[1] *Electrical and Computer Engineering, University of Michigan*

EECS 598-008, W19
Advanced Data Mining

EECS

Problem Definition

- We implemented a CB-CF Hybrid Movie Recommendation system by using Neural Networks.
- **INPUT**: A specific user's watching history, her or his information (History) and information about those movies (Tastes).
- **OUTPUT**: A list of movies that one particular user may get interested.

OUTPUT: A list of Recommended movies

CB-CF Hybrid Movie Recommendation system by using Neural Networks

INPUT: User A's History & Tastes

Motivation

		Scalability?	Cold-start?	Hyperparameter tuning?
Traditional RS	CB: NA, KNN, etc.	×	✓	×
	CF: MF, SVD, etc.	×	×	×
Neural Net RS	NeuMF	✓	×	✓
K PyTorch	Approximating a CB-CF RS function via Neural Networks	Combined all those advantages above		
Hybrid RS	CB-CF Hybrid Neural RS	✓	✓	✓

Datasets

- We used the dataset of MovieLens of the version of ml-1m, and the dataset of The Movie Database(TMDb).
- Files in ml-1m of Movielens contain 1,000,209 anonymous ratings of approximately 3,900 movies made by 6,040 MovieLens users; We also request and parse data such like movie plot overview from TMDb via open API as a complement to the dataset of MovieLens.

movielens
Non-commercial, personalized movie recommendations.

THE MOVIE DB

References

- Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-SengChua. 2017. Neural collaborative filtering. (2017), 173–182
- Yehuda Koren. 2008. Factorization meets the neighborhood
- Wenxing Hong Youchun Ji. 2016. Regularized singular value decomposition innews recommendation system

Our Approach

Feature Selection & Embedding

- Raw Data:

```
1::F::1::10::48067
2::M::56::16::70072
3::M::25::15::55117
4::M::45::7::02460
5::M::25::20::55455
6::F::50::9::55117
7::M::35::1::06810
8::M::25::12::11413
9::M::25::17::61614
10::F::35::1::95370
```

```
adult                                                           0
budget                                                    1.5e+08
credits                                                Chris Buck
imdbId                                                2.29463e+06
movieId                                                    106696
overview             Young princess Anna of Arendelle dreams about ...
popularity                                                 28.575
production_countries                     United States of America
release_date                                           2013-11-27
revenue                                               1.27422e+09
runtime                                                       102
tmdbId                                                     109445
vote_average                                                  7.3
title                                               Frozen (2013)
genres               Adventure|Animation|Comedy|Fantasy|Musical|Rom...
userId                                                         96
user_tag             animation,beautiful,characters,Disney,feminist...
rating                                                        3.5
Gender                                                          F
Age                                                            25
Occupation                                                     16
Zip-code                                                    78028
```

- For Users:
  Treat their attributes as additional id-vocabulary
  Use One-Hot encoder

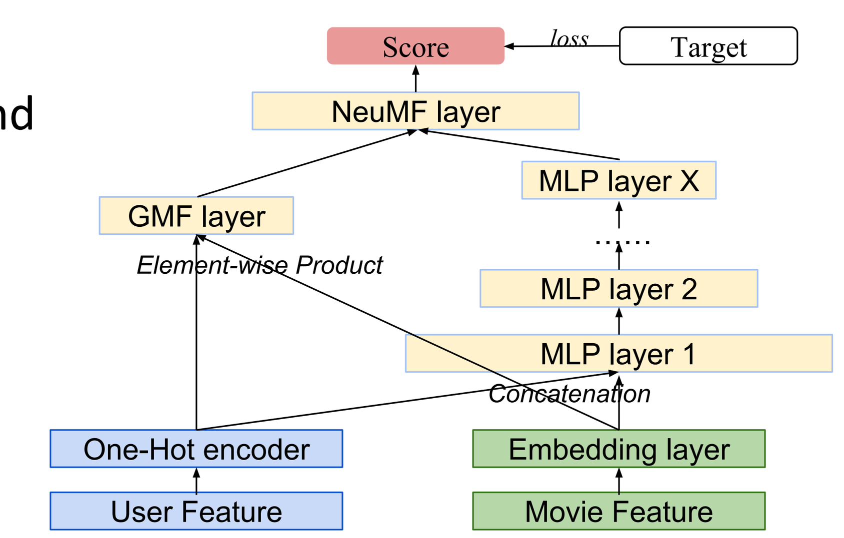
User ID	Gender ID	Age ID	Job ID	ZipCode ID
853	1099	2158	5179	13154

- For Movies:
  Build word dictionary for every word appears
  Use embedding layer before NeuMF model

Title	Genre	Director	Overview
Toy Story	Animation	John Lasseter	Led by woody, Andy's toys..

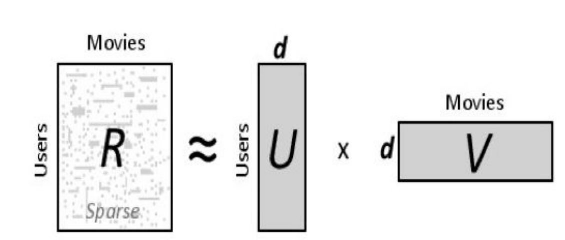
User/Movie Feature Embedding

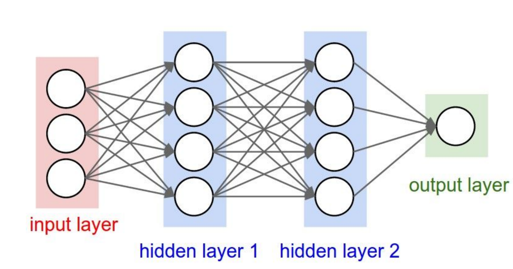
- Mainly two parts: general matrix factorization (GMF) and multi-layer perceptron (MLP)
- Input contains itemized info
- Concatenation & element-wise product keep interactive info



Model Pretrain

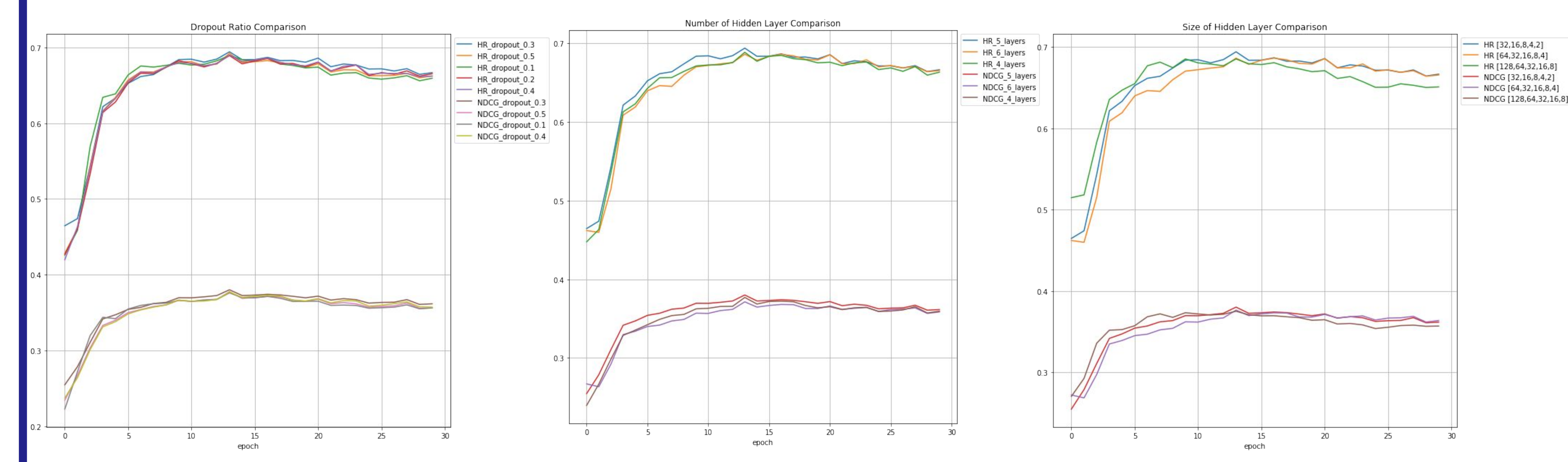
- GMF layers modeling the potential interaction between users and movies based on collaborative filtering
- MLP layers view the problem in a "black-box" way
- Significant improvement observed using pre-trained model!





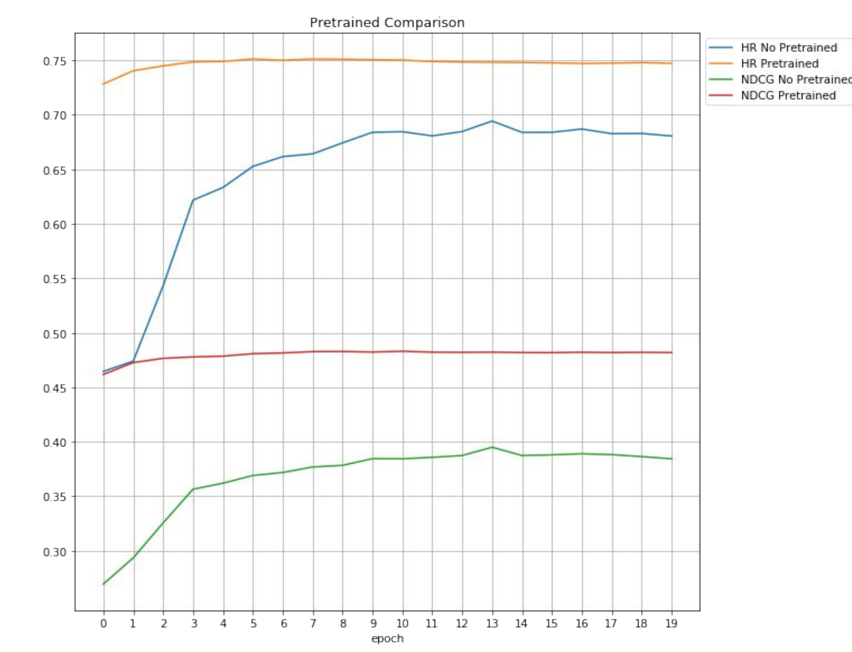
Experimental Results

Experiment 1: Hyper-parameter Tuning



Best Dropout Ratio: 0.3 Best number of layers: 5 Best layer dimensions: [32,16,8,4,2]

Experiment 2: Pretrained vs. No Pretrained



- With pre-trained parameters, both HR and NDCG outperformed the model without pretrained parameters
- Low variation of HR and NDCG in pretrained model, but fluctuated in model without pretrain.
- Improved in terms of both performance and stability!

Experiment 3: Performance Comparison

Model	HR	NDCG
Traditional Content Based	0.2630	0.1757
Traditional Collaborative-filtering	0.3345	0.1908
NeuMF	0.6015	0.3272
NeuMF With User Feature	0.6190	0.3230
NeuMF with User and Movie Feature	0.6991	0.3795
NeuMF with User and Movie Feature with Pretrain	**0.7510**	**0.4827**

Conclusions

- By combining itemized user and movie features, our model outperformed basic NeuMF model by He et al.
- Significant improvement observed by stacking pretrained models
- More data visualization works could be expected for better understanding