

A COMPARATIVE STUDY OF VARIOUS RECOMMENDER SYSTEMS

MANJIT ULLAL, ARSH IRFAN MODAK, OMKAR WAGHMARE

ABSTRACT

Recommender systems are widely employed in the industry and are ubiquitous in our daily lives. The major aim of recommender systems is to help users discover new and relevant items such as movies to watch, text to read or products to buy, find compelling content, so as to create a delightful user experience.

There are three traditional methods (content, context, hybrid) to implement recommender systems. Our goal is to present a comparative study between these traditional recommendation systems and their variants, along with neural networks; and hope to ameliorate some of the shortcomings of the traditional systems.

PROJECT / PROBLEM DESCRIPTION

We can classify Recommendation systems into two broad groups:

- I. Content-based systems, which uses item features to recommend other items similar to what the user likes, based on their previous actions or explicit feedback.
- II. Collaborative filtering systems, which recommend items based on similarity measures between users.

Some of the problems with these techniques are:

1. Cold-start problem where the system cannot generate recommendations for items not seen in training,
2. Hard to include side features (features beyond the query) for query/item,
3. Feature representation of the items are hand-engineered and requires domain knowledge,
4. Model has limited ability to expand on the users' existing interests

Neural Network (NN) models can address these limitations. Neural Networks can easily incorporate side features due to the flexibility of the input layer of the network.

Graph NN, Knowledge-Aware NN and RNN (Recurrent Neural Network) have shown promising results in solving cold start problems.

Contrary to linear models, NN is capable of modelling the non-linearity in data with nonlinear activations such as ReLU, Sigmoid, Tanh, etc.

We aim to implement these traditional models and compare them with their hypertuned variants and Neural Nets.

DATASET

The Movie Dataset

These files contain metadata for all 45,000 movies listed in the Full MovieLens Dataset. The dataset consists of movies released on or before July 2017. Data points include cast, crew, plot keywords, budget, posters, revenue, release dates, languages, production companies, countries, TMDb vote counts and vote averages.

This dataset also has files containing 26 million ratings from 270,000 users for all 45,000 movies. Ratings are on a scale of 1-5 and have been obtained from the official GroupLens website.

ALGORITHMS

1. Baseline Algorithms (NormalPredictor, BaselineOnly)
2. Matrix Factorization (SVD, SVD++, NMF, Co-Clustering, SlopeOne)
3. Neural Networks (Neural Collaborative Filtering)
4. If time permits, we will also implement RNN, Hybrid Autoencoder and CNN to create our Recommender System.

LIBRARIES AND TOOLS

Programming Language: Python (Numpy, Pandas, Matplotlib, Seaborn, Surprise, Scikit-Learn, Pytorch / Keras / Tensorflow.)

IDE: Spyder and/or Jupyter Notebook and/or Visual Studio

RESULTS

We aim to evaluate and compare the recommender systems that we will implement using the aforementioned methods. Moreover, we intend to provide a detailed explanation of these algorithms and their results.

The metrics we plan on using include but are not limited to RMSE, MAE, nDCG etc.

REFERENCES

- [1] <https://www.kaggle.com/rounakbanik/the-movies-dataset>
- [2] <https://arxiv.org/pdf/1606.07659v3.pdf>
- [3] <https://arxiv.org/pdf/1708.05031v2.pdf>