MovieLens Recommendations

19 February 2019

David Merten-Jones

Thinkful Unit 7 Capstone Proposal

Overview

Recommendation algorithms using collaborative filtering have become increasingly important in advertising, educational services, and in the creation of adaptive web applications. One such application in the entertainment industry is matching media to consumers based on how consumers rate other media, and on ratings given by other consumers with similar taste. Within the context of a platform such as Hulu or Netflix, it helps viewers get to the content they want to watch faster, which leads to better user experience (and possibly lower churn.)

Here, I will be using the MovieLens 100,000-entry dataset (ml-100k) to create a collaborative filtering model to simulate such a scenario.

Goals

1. Create several models that predict how viewers would rate movies.
2. Tune the models to minimize error, and use the best-performing model to provide recommendations for movies users would rate highly.

Process

The principal technique used for this purpose will be matrix factorization, implemented through simple collaborative neural network models. For purposes of comparison, other methods may include more traditional ensemble models such as random forest regression, or deep neural networks.

Challenges

I anticipate that the most challenging aspect of this project will be tuning the neural network models for peak performance. Hyperparameters such as learning rate, weight decay, momentum, and the number of latent factors included may have a significant impact on the performance of the models. Also, should I pursue the avenue of ensemble modeling, feature engineering may prove difficult due to the large, sparse dataset.

Sources:

https://towardsdatascience.com/collaborative-filtering-for-morphing-applications-af4f9f7cc259

https://grouplens.org/datasets/movielens/