Predicting Ratings for Netflix Users

Amiya Renavikar

Contents

Introduction

- Video streaming is an increasingly popular form of entertainment
- More and more competitors entering the market
  - Customers must be provided with high quantity of content and with content they actually want to watch
- Recommendation: one of the most important concepts of video streaming

Recommendation Systems

- One of the most commonly used recommendation algorithms is collaborative filtering
- Our dataset: Netflix Prize dataset
- Analyzed two approaches to collaborative filtering
  - “Traditional” approach using nearest neighbors in collaborative filtering
  - “Novel” approach using clustering to implement collaborative filtering
    - Implemented from “A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering” by Songjie Gong.
- Potentially reveals a more powerful method of video streaming recommendation

Why does this matter to us?

- We use services that utilize recommendation systems every day!
  - Netflix, Hulu, YouTube, Prime Video, etc.
- More powerful/accurate recommendation system = more likely to be suggested new content that you will truly enjoy

Related Work

- "Item-Based Collaborative Filtering Recommendation Algorithms" (Sarwar, Badrul Munir, et al)
  - Detailed look at collaborative filtering
- "Matrix Factorization and Neighbor Based Algorithms for the Netflix Prize Problem" (Takács, Gábor, et al.)
  - In-depth look at collaborative filtering as it relates to the Netflix dataset
- "Lessons from the Netflix prize challenge." (Bell, Robert M., and Yehuda Koren.)
  - Analysis of the original Netflix Prize competition and implementations made by participants

Method

Collaborative Filtering Algorithms based on User & Item Clustering

- Traditional approach: KNN filtering
- Novel approach: Clustering

JOURNAL OF SOFTWARE, VOL. 5, NO. 7, JULY 2010 745

A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering

SongJie Gong
Zhejiang Business Technology Institute, Ningbo 315012, China
Email: songjie_gong@sina.com

Novel Clustering-Based Algorithm

```
Input: clustering number k, user-item rating matrix
Output: smoothing rating matrix
Begin
   Select user set U={U1, U2, …, Um};
   Select item set I={I1, I2, …, In};
   Choose the top k rating users as the clustering
CU={CU1, CU2, …, CUk};
   The k clustering center is null as c={c1, c2, …, ck};
   do
     for each user Ui∈U
       for each cluster center CUj∈CU
         calculate the sim(Ui, CUj);
       end for
       sim(Ui, CUm)=max{sim(Ui, CU1), sim(Ui,
CU2), …, sim(Ui, CUk);
       cm=cm∪Ui
     end for
     for each cluster ci∈c
       for each user Uj∈U
         CUi=average(ci, Uj);
       end for
     end for
    while (C is not change)
End
```

Rationale

- Compare results for both algorithm approaches
- Several routes to evaluate different parameter settings
- Fine-tuning k-value value for traditional approach and clustering-based approach

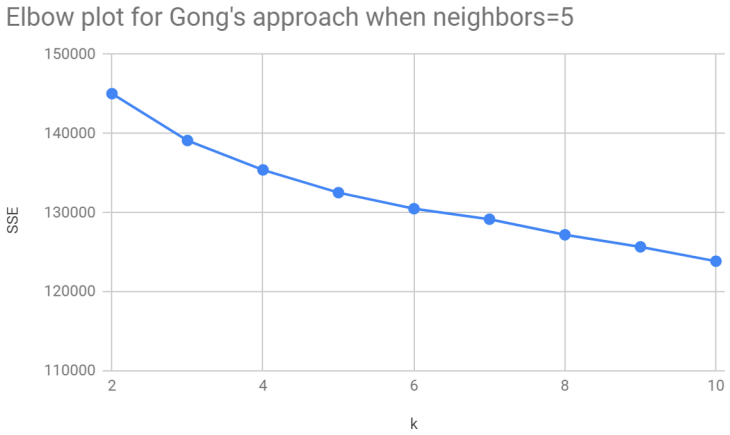
Experimental Design

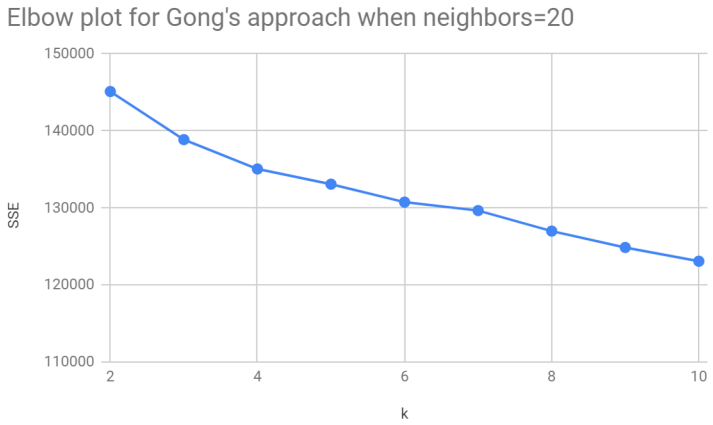
- For Gong's algorithm, we tested with k values from 2 to 10 and chose the best based on SSE.
- Gong's algorithm makes use of a "neighbors" parameter to find the *n* most similar movies to a movie in question.
- We tested the algorithm with values of this parameter of 5 and 20.
- For comparison, we tested the traditional collaborative filtering algorithm with user neighbors values of 5 and 20.
- Performance was evaluated by computing the mean absolute error (MAE) between the predicted ratings and actual ratings
- 90/10 random train/test split

Data

- We tested our algorithm on a subset of the Netflix Prize Dataset limited to data from users with id less than 4000.
- Data was preprocessed to remove the movies and users in the lowest 20th percentile by number of reviews. (https://www.kaggle.com/laowingkin/netflix-movie-recommendation)

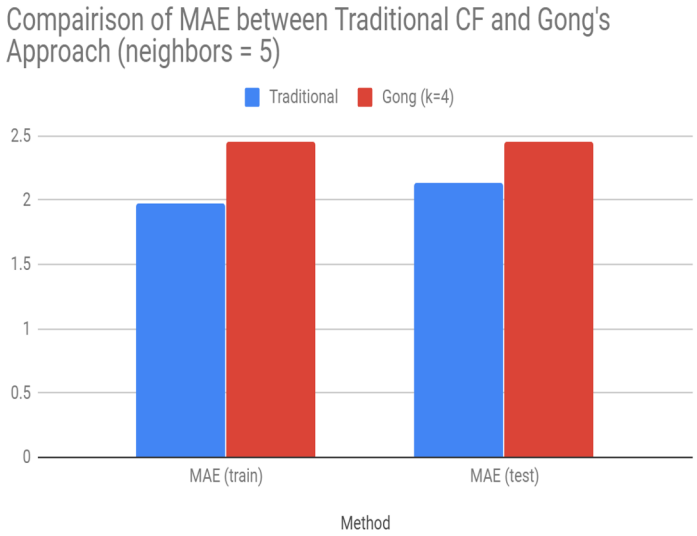
Results: Elbow Plots

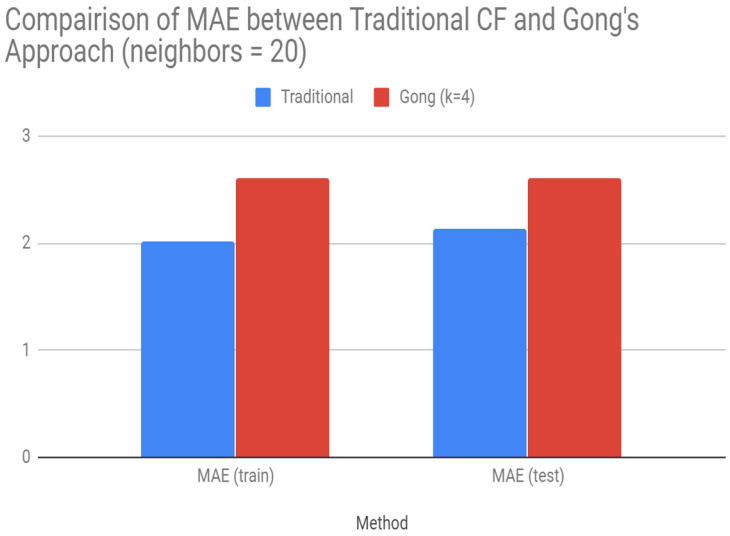




- The best value of k is somewhat unclear from the plots
- For both, it seems that 4 is close to an elbow

Results: Prediction accuracy





Conclusion

- Systems to predict user ratings provide a powerful tool to recommend new products to a customer.
- Our results contradicted those provided by Gong in the paper, suggesting that traditional collaborative filtering is the best fit for this problem.
- There were differences in data and evaluation methods between our tests and those in the paper, so more tests are needed.