BARCELONA SCHOOL OF ECONOMICS

DATA SCIENCE METHODOLOGY PROGRAM

21D009 NETWORKS: CONCEPTS AND ALGORITHMS

Movie recommendations using networks

Authors:
CODD, Jonny
CHEN, Joshua
GALLEGOS, Rafael
PÉREZ, Carlos

Professors:
MILÁN, Pau
KOMANDER, Björn

December 20th, 2023

Contents

List of Figures

List of Tables

1   Introduction

The travelling salesman problem (TSP) is an archetypal problem in linear programming. The setup tells the reader to imagine a salesman who must visit n cities, then asks them to find the shortest path the salesman can take, visiting each city exactly once before returning to the starting city given the distance between all pairs of cities. Its deceptively simple formulation lies in stark contrast to the vast computational expense required to solve it; it is NP-hard, and despite nearly one-hundred years of literature, we rely on heuristic algorithms to find good solutions.

2   Literature Review

3   Dataset

MovieLens is popular movie recommender system dataset developed by GroupLens, a computer science research lab at the University of Minnesota. The goal of this challenge is to recommend movies to its users based on their movie ratings. Group Lens offers datasets of different sizes and their datasets are widely used in research and teaching contexts.

The selected dataset consists mainly on two files: movies.csv and ratings.csv. Movies dataset has 9,742 unique films and a column indicating the genres of the film. All possible genres are: 'Romance', 'Musical', 'Children', 'Documentary', 'Sci-Fi', 'Film-Noir', '(no genres listed)', 'Crime', 'Mystery', 'Drama', 'Western', 'Fantasy', 'Animation', 'Thriller', 'War', 'Action', 'Adventure', 'IMAX', 'Comedy', 'Horror'. The number of movies per genre is represented in Figure 3.4.

Ratings dataset consists of 100,836 ratings with 610 unique users that rated 9,724 movies. As it can be observed in Figure ??, the ratings from users are right-skewed, which suggests that users tend to enter their rating on movies that they probably have liked. Ratings from users have been registered from 1996-03-29 until 2018-09-24. The most popular movies among users have been: Shawshank Redemption, The (1994), Godfather, The (1972), Fight Club (1999), Godfather: Part II, The (1974) and Goodfellas (1990).

The median user has watched 70 films, whereas the user with the lowest number of watched films was 20 movies and the user with the highest number of rated films is 2698.
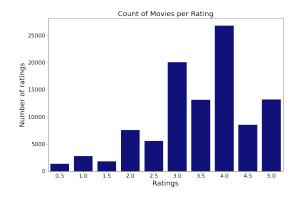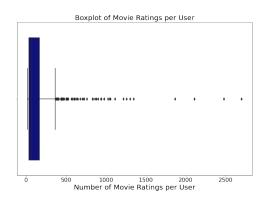
Figure 3.1: Count of Movies per Rating
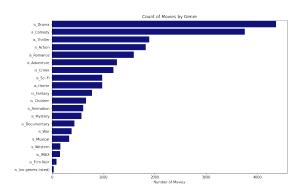

Figure 3.2: Count of Movies per Genre

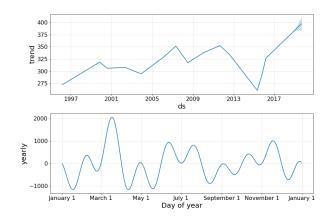

Figure 3.3: Count of Movies per User

Figure 3.4: Time Series Analysis of Ratings with Prophet

4 Network analysis

4.1 User-to-user network

4.2 Movie-to-Movie network

4.3 User and movie bipartite network

4.4 User-to-Genre network

5 Collaborative Filtering

6 Graphical Neural Networks

7   Conclusions