Movie Recommender System

Ajla Sokol
Emine Yaman

Faculty of Engineering and Natural Sciences,
International University of Sarajevo,
Hrasnicka Cesta 15, Ilidža 71210 Sarajevo,
Bosnia and Herzegovina
ajla.sokol@hotmail.com

Article Info

ABSTRACT: Recommender systems are necessary in current time, since the information available online can be overloading to a user. These systems are used everywhere, starting from the online shops to the websites that are focused on recommending particular item, such as videos to watch or songs to listen to. Recommender system that predicts the likings of a user based on their previous behavior is very popular when it comes to picking up the movies to watch. This paper talks more about the movie recommender systems, and explains the way that different types of recommendations can be used in order to test datasets and provide good recommendations for variety of users.

1. INTRODUCTION

Growth of the Internet together with all of the available digital information can be overwhelming for users in certain period of times. This meant that the need for a certain kind of personalization of these information, as well as prioritization, was needed. More than ever, the demand has increased for systems that would recommend things user searches for, hence the making of recommender systems has come up. These systems are information filtering systems that solve the problem of overwhelming amount of information. They have the ability to predict the likings of particular user based on their profile and preferred items [2]. The way that recommender system works is that there should be some inputs involved, such as data collected from the users, and the output is the suggestions or predictions based on the data available on user [6].

In recent period, recommender systems have been present in many online stores of all types, but also in websites that serve as media to certain videos, songs and similar. These systems have helped the customers choose items that they are more likely to find interesting by guiding them towards that product more. Not just that they helped in the shopping

field, but also in guiding user towards the videos that could be of his or hers interest, as well as songs or movies [1]. Users are more likely to watch the movie that has been suggested by the recommender system, or see a video on YouTube that was placed below the "Recommended" label.

Since of the overwhelming information and constant movie releases, people also try to use recommender systems in order to find the movie that could be similar to those of their likings, or the movies that presents cast and plots that would be of their interest. This paper talks more about movie recommender systems and their types. It is also focused on testing out simple recommender system for movies that gives the user different types of recommendations based on the data available.

2. LITERATURE REVIEW

There has been a variety of recommender systems tested out, and many research papers that have been talking about certain types of recommender systems. Miller et al [1] have done research in the field of recommender systems provided on mobile devices. They have made a movie recommender systems which works in two different ways. In one scenario, this recommender system provides the video that user can rent or buy, and in the other scenario, it recommends the movies from the theater that are currently available. After the testing done in this field, they concluded that this type of recommender system could be very valuable service inside mobile phones. In their research paper, authors [3] have talked about the MovieTweetings dataset, which are collected automatically from social media posts, one of them being Twitter. They talk about how the social media can provide more natural and more realistic data, and how to incorporate that data into recommender systems in order to make them more accurate and up-to-date.

Authors Christakou et al [4] developed a hybrid recommender system. This recommender system combines content-based filtering with collaborative filtering, and training of the system has been done through neural network classifiers. They have tested the system and concluded that this type of the recommender system could be suitable for use in home entertainment center, or as a web recommender that could provide movies recommendations for users that have rated several of them previously. In the research paper of Golbeck et al [5], authors have developed a website called FilmTrust. This website integrates the Semantic Web-based social networking into a movie recommender system. They used TidalTrust algorithm as the basis of their personalization of recommendations for each user. Through the testing that was made on their system, they proved that the accuracy of their version of movie recommender system was significantly higher than of the average recommender system.

The topic about recommender system in E-commerce has been done in the research paper of the authors Schafer et al [7]. This paper talks about the importance of recommender system in E-commerce, and how their need has increased in time. Authors have provided several examples of websites that use recommender systems, such as Amazon, eBay, CDNOW, Levis and Reel. They also talk about the opportunities in this industry that could provide financial benefits to industry, and have touched upon the different types of recommender systems available in E-commerce business. This research paper was done in order to provide more opportunities for E-commerce websites, and showcase different ways of benefiting for both customers and industry.

3. PROJECT GUIDLINES

A. Types of Filtering Techniques Used

In order to provide accurate recommendations, the techniques of filtering the data are important. Content-based filtering is one of those filtering techniques. This technique is based on the data taken from the users past behavior. This means that the recommendations are based off of the positively rated items [2]. Content-based filtering is filtering out the data based on a different tags inside the movies that user liked. There can be different ways to make a content-based filtering for movie recommender systems. Some of those can be based on the taglines of movies and general description of plots, while others can go more in depth by also filtering out the data, such as the cast of the movies, as well as their genre and some keywords presented inside.

Other type of filtering technique that was tested out in this paper is called collaborative filtering. This recommender system filtering is more complex-type filtering, since this has more data to filter out and base recommendations on. Collaborative filtering matches users with the content that they may like by comparing their profile and behavior to other users and filtering out the probable recommendations that were previously liked by the users of the similar taste. In other words, the recommended movies are based on the ratings of other users of similar profile. Collaborative filtering can be model-based and memory-based filtering, where model-based filtering is related to data mining techniques, such as clustering, decision tree and association rule, while the memory-based filtering uses more of a cosine similarity in order to define similarities between users and items [2].

B. Tools and Dataset

The programming language that was used for this particular research was Python. Python is a programming language that is growing in use on a daily basis, since it provides a huge amount of libraries that could be used in many different purposes. This programming language is also used in data-mining, by providing different libraries that could perform data analysis as well as predictions based off of the training models provided. The libraries that were used in this particular project were Pandas, Numpy and SciPy.

Pandas is one of the libraries that is used for data analysis in python. It is open source, and provides a lot of tools in the data mining field. This library takes the different file types of data, and creates an object called data frame. NumPy is more of the library with mathematical approach, where it uses linear algebra, Fourier transform and can go through N-dimensional array objects, which can be useful when going through multidimensional dataset. SciPy is a library that can operate on the arrays taken from the NumPy, and accordingly manipulate with it. This library is used in order to go through the datasets taken from the NumPy, and try to

optimize the results. Another tool that is built upon SciPy and NumPy is called SciKit. This tool is used for different methods of analyzing data when it comes to data mining techniques. SciKit library uses regression, clustering and other tools of data mining that could help in preprocessing data, and can apply different algorithms in order to perfect the final prediction.

Dataset that is used when conducting this research is the dataset from Group Lens. Group Lens is a research lab based in the University of Minnesota. This research group is based on using different data mining techniques and testing them out in order to provide the best possible recommender systems and predictions that will make the advances in this field. One of their projects is called Movie Lens. This project is a web-based movie recommender system that is made in order to take in the data from all kinds of users on their preferences when it comes to the movies. The dataset that is made from this website is provided free in order for researchers to conduct different experiments in the field of recommender systems. This dataset is updated every several months in order to try to keep it as accurate as possible. They provide different sizes of datasets, from dataset with 20 million ratings by 138000 users, to smaller types of datasets with 100000 ratings by 600 users. The smaller dataset is one that is used generally in education purposes, and that dataset is used in this research paper as well. This dataset contains ratings of users under the ID of users, movies descriptions and tags, together with credits and keywords of movies, as well as their cast.

4. RESULTS AND DISCUSSION

This project test out two different types of filtering techniques, both of which are described in previous section. The first filtering technique that will be talked about is content-based filtering technique. This type of filtering technique is tested by using cosine similarity algorithm that generates the output by comparing the relation between the data by looking at the angle of their correlation. To test out how this filtering technique works, two different approaches are used.

In the first approach, data tested is only data taken from the descriptions of the movies provided. In this case, the input taken will be the movie that one user likes, and the output will be different movies that are predicted that a user will like. Several movie inputs were placed, such as The Dark Knight and Interstellar. The recommendations for The Dark Knight are following:

1. The Dark Knight Rises
2. Batman Forever
3. Batman Returns

4. Batman: The Dark Knight Returns, Part 2
5. Batman: Under the Red Hood

Recommendations for Interstellar are following:

1. Suburban Commando
2. Gattaca
3. Final Destination 2
4. Space Cowboys
5. All Good Things

Since this recommendation algorithm is based only on description of movies, it is assumed that the recommendations will all be with very similar plots inside of them. With the first movie tested, we can see that all of the recommended movies are those with the same plot.

Since The Dark Knight is movie based on the character called Batman, all of the recommended movies are about the Batman as the main character. The Interstellar is the movie about space travelling and exploring. All of the recommended movies have plots that are based purely on the space-related topics. The next testing was also done in content-based filtering, but this time trying to make it more personalized by, not only preprocessing data based on movie descriptions, but also based on the genre of the movie, cast and keywords from that movie. Using this technique, the recommendation should now be more personalized, and not only based on the pure plots of movies. In order to try to compare both recommendations, the movies tested are the same. For The Dark Knight, the movies recommended are:

1. The Dark Knight Rises
2. Batman Begins
3. The Prestige
4. Following
5. Inception

Recommendations for Interstellar are following:

1. Inception
2. Following
3. The Martian
4. Memento
5. Insomnia

When comparing these recommendations to the previous ones, we can see big difference. In this case, the movies that

are recommended are of a more various base, so in The Dark Knight recommendation we don't have only recommendations based on the pure description on the Batman character, but also based on the cast members and genre. This is also the case with Interstellar, where the movies recommended are also those of different plots, but with similar cast or genre. This recommendation is more likely to be better accepted from the variety of people, rather than only that of a pure plot comparison.

The other type of filtering technique that was used in order to be tested for the research purposes was the collaborative filtering technique. In this case, tools for collaborative filtering were provided from the libraries, where the main algorithm used was Singular Value Decomposition (SVD). This algorithm is based on the rectangular matrix calculations in order to minimize the square root error, or RMSE.

In this research paper, this type of filtering techniques takes the input of the user ID and the movie ID, and as the output predicts the possible rating of this movie from that user ID that is placed as an input. In order to test this, different user ID's are placed for the same movie ID, where the different estimated ratings were gotten. For the user with ID 10, the estimated rating was 3.48; user with ID of 1 had the estimated rating of 2.47, while the user of ID 15 had the estimated rating of 2.51, all for the movie with ID 45. This filtering technique gives out the result based on all of the data that was available, since it predicts the movies that user will like based on the profile of similar users.

5. CONCLUSION

Recommender systems in general are highly beneficial and highly used nowadays, especially with the rise of social media. This is not only used in shopping purposes, but also as a tool for people to find the movie or a song that they may like, as well as a new person to follow or send a friend request to. In this research paper, we touched upon the movie recommender system. This research was conducted in order to understand more about the different techniques that could be applied for searching the best possible prediction in the field of entertainment. Both content-based filtering and collaborative-based filtering were tested, and the results were presented. Overall, this research was done in the educational purposes in order to better understand data mining tools and techniques, and their applications in real world.

REFERENCES

[1] B. Miller, I. Albert, S. Lam, J. Konstant and J. Rield, "MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System", IUI, Florida, USA, January, 2003

[2] F. Isinkaye, Y. Folajimi and B. Ojokoh, "Recommendation systems: Principles, methods and evaluation", Egyptian Informatics Journal, June, 2015

[3] S. Dooms, T. De Pessemier and L. Martens, "Movie Tweetings: a Movie Rating Dataset Collected From Twitter", Ghent, Belgium, January, 2013

[4] C. Christakou, S. Vrettos and A. Stafylopatis, "A Hybrid Movie Recommender System Based On Neural Networks", International Journal on Artificial Intelligence Tools, vol. 16, pp. 771-792, 2007

[5] J. Golbeck and J. Hendler, "FilmTrust: Movie Recommendations using Trust in Web-based Social Networks", IEEE Consumer Communications and Networking Conference, 2006

[6] P. Resnick and H. Varian, "Recommender Systems", Communications of the ACM, vol. 40, March, 1997

[7] J. Schafer, J. Konstant and J. Rield, "Recommender Systems in E-Commerce", ACM Conference on Electronic Commerce, Colorado, USA, November, 19