
Movie Recommendation System

Shamit Bhargava, Nikhil Suri, Harshard Naghware, Niraj Jaiswal

Abstract - Filtering systems often want to remove unnecessary information from an outsized amount of knowledge. Recommender systems want to seek and predict meaningful and informative items that a user might put into the info. The system emphasizes reusing the knowledge and preferences of the users which will be utilized in the calculation of future recommendations.

This paper proposes a recommender system which provides recommendation supported the knowledge given by the users.

It is achieved by analysing user's psychological profile, their watch history, and movie scores from additional websites. It is based on aggregate similarity conditions. This system uses both content as well as collaborative filtering. Both are often explained as follows: Collaborative filtering means building systems from the user's past behavior (ie. Items that have already been selected or rated) Afterwards the model is employed to predict outcomes that the user could be curious about.

Content-based filtering uses a series of distinct and discrete characteristics of an item to recommend more items with the same properties.

Both of those systems combine to form a hybrid recommender system.

This system which may be a hybrid of both filtering systems can recommend movies using analysis of the profiles.

1. INTRODUCTION

A recommendation system may be a model which is employed to filter information and predict the output supported the preferences of the user. These models became extremely popular that they're getting used in movies, books, television, restaurants, food etc. These systems help in improving the longer term suggestion of the corporate .

A large number of companies are taking advantage of the advice system in improving customer satisfaction and knowledge. Massive chunks of revenue can be attained from this that is why most of them are turning to a recommendation system.

One goal of this technique is to supply a system that considers the past ratings given by the user to supply suggestions to the user. It is done by using the collaborative filtering and Apache Mahout framework. The second goal is to match the performance and efficiency of user-based recommender system and item-based recommender system.

1.1 OVERVIEW:

Collaborative filtering: It implies building a system from user's past behaviour. Afterwards the model is employed to predict outcomes that the user could be curious about .

Content based filtering: It uses a series of distinct and discrete characteristics of an item so as to recommend more items with same properties.

Hybrid System: Hybrid recommendation systems are a mixture of both collaborative and content-based filtering methods. In these sort of systems, collaborative and content-based predictions are performed separately then the results of both techniques are combined to supply recommendations.

Mahout: It aims to supply free and distributed and scalable implementations of advanced machine learning algorithms utilized in the sector of clustering, classification, collaborative filtering

Over the years, many recommendation systems are developed using either collaborative, content-based, or hybrid filtering methods. These systems are implemented using various big data and machine learning algorithms.

The authors propose a collaborative recommendation system that is meant to figure on the Hadoop platform, using the Map-Reduce framework. The authors have used the set-similarity join method to create this technique, employing both user-based and item-based collaborative filtering techniques.

They proposed a movie recommendation system using collaborative filtering that focuses on the ratings given by the users to supply recommendations. The proposed system is made using the K-means algorithm to sort the films consistent with the ratings.

In one paper the authors propose a content-based movie recommendation system to recommend movies. The proposed system makes use of a neural network with the Content based filtering uses a series of distinct and discrete characteristics of an item so as to recommend more items with same properties. Both of those systems combine to form a hybrid recommender system.

This system which may be a hybrid of both filtering systems is capable of recommending movies using analysis of the profiles.

content information of the films to get features and learn the similarities between movies.

1.2 EXISTING MODELS

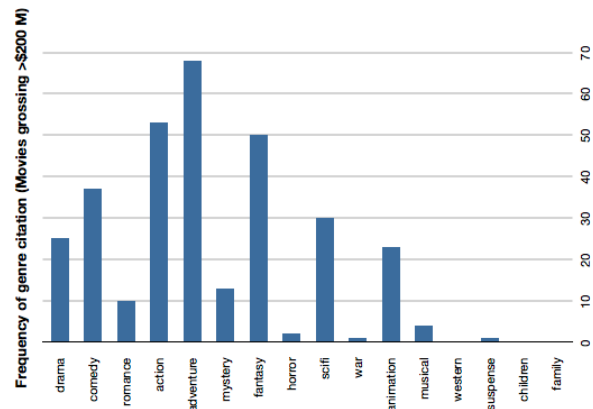
The movies are recommended supported the similarity between them.

The authors implement a recommendation system that mixes both user-based and item-based collaborative filtering approach. The system is made using nearest neighbours machine learning technique and develop a replacement algorithm that unifies used based and item-based recommendations. supported the research we conducted, collaborative filtering was found to be one among the popularly used approaches to create recommendation systems. Many of the systems used machine learning algorithms like clustering using K-means, neural networks then on to recommend items.

2. IMPLEMENTATION

1. Dataset: The dataset utilized in this paper is obtained from Yahoo Research Webscope database. It provides two files - Yahoo! Movies User Ratings and Yahoo! Descriptive Content Information. The Yahoo! Movies Users Ratings file contains 211231 records and contains User ID, Movie ID and Ratings. There are a total of 54058 records in The Yahoo! Movies Descriptive Content Information file and contains Genre, Actors, Movie ID, Title, Directors and etc.

2. Data Cleaning: the films Descriptive Content Information file contained about 40 columns. Most of those columns weren't required for our experiments and hence were removed. The dataset also contained tons of blank values and duplicate values which needed to be resolved. additionally , there



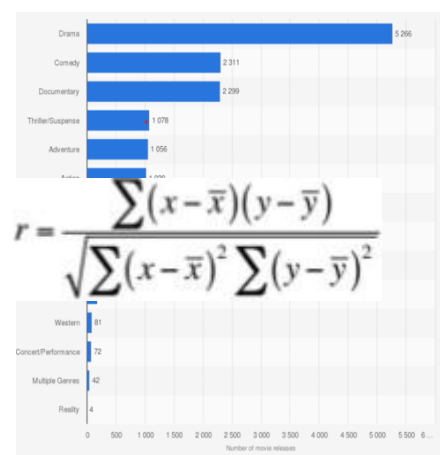
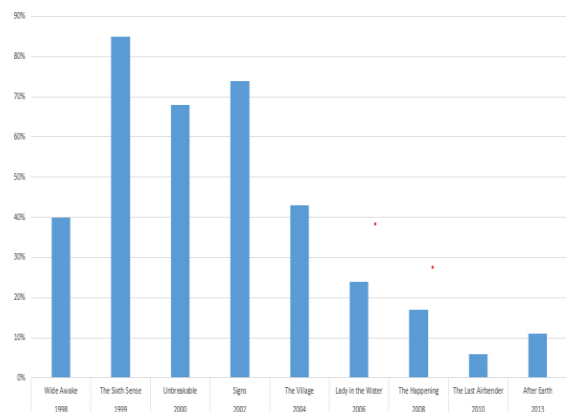
have been some entries for movies within the Movies Users Ratings files that did not correspond to any movie within the Movies Descriptive Content Information file. The entries were removed so that processing could be easy.

Data Analysis: We used the mahout library to create the recommender system. For User-based filtering we used the User-Similarity class additionally to the Pearson Correlation Similarity which uses the Pearson coefficient of correlation to work out the similarity between users' ratings. Below is that the mathematical formula for Pearson Correlation. the upper the coefficient is, the more correlated the 2 users' choices are.

movies in each genre and the number of movies rated in each rating category.

3. Model building: We used the mahout library to create the recommender system. For User-based filtering we used the User-Similarity class additionally to the Pearson Correlation Similarity which uses the Pearson coefficient of correlation to work out the similarity between users' ratings. Below is that the mathematical formula for Pearson Correlation. the upper the coefficient is, the more correlated the 2 users' choices are. .

The User-neighbour-hood is computed by using



$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

items are static, their similarities based on the user ratings are not going to change over time, we can pre-compute them and store them offline. The results from the Item Based recommender is loaded to the Hadoop Distributed File System (HDFS) to have a scalable and fault-resistant storage. The User Based recommender results have to be computed every time during a recommendation since unlike items, ratings provided by users.

3. MODEL EVALUATION

Qualitative evaluation: The movie recommender system inbuilt this paper facilitates the understanding of how a recommender system works. To gauge the accuracy and relevancy of the results produced by our system, we analyze both the approaches differently.

Movie 1	Movie 2	Similarity
1800421139	1800379216	0.99959636
1800061638	1800111258	0.99959064
1800121659	1800379216	0.99955463
1807537463	1804738128	0.9995903
180283191	1807858489	0.9995346
1800121659	1800111258	0.9995051
1800061638	1800121659	0.9994775
1800421139	1800121659	0.99944425
1800111258	1800379216	0.99939984
1800421139	1800111258	0.99938726
1800061638	1800379216	0.9993557
1800421139	1800061638	0.999335
1800080788	1800080795	0.9992829
180743259	1807428853	0.9992593

We compare the Item based similarity coefficient results as given within the above figure by mapping the Movie ID of Movie 1 and Movie 2 to their titles. As evident from the table, movies which are similar are given a better similarity metric.

The Lord of the Rings: The Fellowship of the Ring (2001)	The Lord of the Rings: The Two Towers (2002)	0.999948
The Empire Strikes Back (1980)	Star Wars (1977)	0.9994775
I diana Jones and the Last Crusade (1989)	I diana Jones and the Temple of Doom (1984)	0.9992829
ET: The Extra-Terrestrial (1982)	S...rV...n (1977)	0.9990453
The Godfather (1972)	The Godfather Part II (1974)	0.9989032
Pirates of the Caribbean: The Curse of the Black Pearl (2003)	The League of Extraordinary Gentlemen (2003)	0.99869
The Matrix Reloaded (2003)	Brace with Me (2003)	0.998663
Jeepers Creepers 2 (2003)	Heddy's Jam (2003)	0.99858
Harry Potter and the Chamber of Secrets (2002)	The Lord of the Rings: The Fellowship of the Ring (2001)	0.99873
Shrek (2001)	Shrek (2001)	0.998355
Shrek 2 (2004)	Shrek (2001)	0.99832
Shrek (2001)	Shrek (2001)	0.998276
The Texas Chainsaw Massacre (2003)	Scary Movie 3 (2003)	0.998155
The League of Extraordinary Gentlemen (2003)	Bad Boys II (2003)	0.998154
Iron Man (2008)	S...rV...n (2001)	0.99798
Iron Man 2 (2010)	Underworld (2003)	0.997915
Austin Powers: International Man of Mystery (2002)	Biggie (2002)	0.997732
Daddy Day Care (2003)	Brace with Me (2003)	0.997585
The Mummy (1999)	The Mummy Returns (2001)	0.997426
2 Fast 2 Furious (2003)	The Matrix Reloaded (2003)	0.997373
The Fast and the Furious (2001)	The Texas Chainsaw Massacre (2003)	0.997123
How to Lose a Guy in 10 Days (2003)	Bad Boys II (2003)	0.99708
Down with Love (2003)	How to Lose a Guy in 10 Days (2003)	0.99706
The Hot Chick (2002)	The Mummy Returns (2001)	0.99707
Derail (2002)	The Order (2001)	0.99659
Re: A Sinatra - 3 P...ck (2002)	Great Music, Is - Vol. 2 (1951)	0.952113
Biggie (2002)	The Betty Davis Collection (1993)	0.952113
Someone Like You (1991)	Someone Like You: The Movie (2002)	0.951782
Someone Like You (2002)	Someone Like You: The Movie (2002)	0.951386

For user-based recommender system, we evaluate the model using the typical Absolute Difference Recommender Evaluator. We divide the training data into a two different sets of test and train samples. Next, we evaluate the rating predictions on test data against the particular ratings as laid out in the training data.

The following figure shows the raw data output from the user-based filtering system.

The system recommends 10 movies to user and returns the closest neighbors which have most similar taste preference as him. For every movie recommended, it also predicts the ratings by that user. We get a mean absolute difference of 0 which proves that the predictions made on the ratings of the recommended items are 100% accurate. The below figure shows the output that's obtained and kept before the user.

4. CONCLUSION:

In this paper we've implemented a movie recommendation system using collaborative filtering. It's implemented using Apache Mahout and takes the ratings given to movies to supply movie suggestions. The system considers the recorded user ratings as reference to recommend movies according to the applied algorithm. Within the future, more features like the genre of the movie, the administrators, the actors and shortly might be considered also to supply suggestions. Additionally, a replacement framework called Apache Prediction 10 might be looked into to develop the system rather than Mahout.

5. REFERENCES

- [1] A. V. Dev and A. Mohan, "Recommendation system for big data applications based on set similarity of the user preferences" 2016 International Conference of on Next Generation Intelligent Systems(ICNGIS), Kottayam. 2016, pp.1-6. Doi: 10.1109/ICNGIS.2016.7854058
- [2] Kumar, Manoj & Yadav, D.K. & Singh, Ankur & Kr, Vijay, "A Movie Recommender System: MOVREC", 2015 International Journal of Computer Applications. 124. 7- 11. 10.5120/ijca2015904111

The Lord of the Rings: The Two Towers (2002)	5
Freaky Friday (2003)	5
The Lord of the Rings: The Fellowship of the Ring (2001)	5
Bad Boys II (2003)	5
How to Lose a Guy in 10 Days (2003)	4.5

-
- [3] S. G. Walunj, K. Sadafale, "An online recommendation system for e-commerce based on Apache Mahout framework", 2017 ACM SIGMIS International Conference on Computers and People Research, pp. 153- 158,2013.
- [4] H. W. Chen, Y. L. Wu, M. K. Hor and C. Y. Tang, "Fully content-based movie recommender system with feature extraction using neural network," 2017 International Conference on Machine Learning and Cybernetics (ICMLC), Ningbo, China, 2017, pp. 504-509. doi: 10.1109/ICMLC.2017.8108968
- [5] Z. D. Zhao, M. S. Shang, "User-Based collaborative filtering recommendation algorithms on Hadoop", Proc. of Third International Workshop on Knowledge Discovery and Data Mining, pp. 478-481, 2016.
- [6] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Item-based collaborative filtering recommendation algorithms", Proceedings of the 10th international conference on World Wide Web, pp. 285-295, 2001
- [7] Koen Verstrepen, Bart Goethals, "Unifying nearest neighbors collaborative filtering", Proceedings of the 8th ACM Conference on Recommender systems, October 06- 10,2014, Foster City, Silicon Valley, California, USA doi:10.1145/2645710.2645731
- [8] Jain, A., & Vishwakarma, S. K., "Collaborative Filtering for Movie Recommendation using Rapid Miner" International Journal of Computer Applications (0975 - 8887) Volume 169 - No.6, July 2017
- [9] M. Gardener, "Statistics for Ecologists Using R and Excel (Edition 2) ". [Online]. Available: <http://www.dataanalytics.org.uk> [Accessed: 05-May- 2018]
- [10] Community Shared, "Apache Mahout". [Online]. Available: <https://mahout.apache.org/> [Accessed:02- April-2018]
-