

Extensive Recommendation System for movies Using user Ratings

Guide: M. Indumathy¹, Student: Tushar Kumar Tewary², Mukundan.S³, Somasekar.S⁴

*Department of Computer Science Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai

Abstract- The movies that are suggested in the websites are nowadays having more unrelated suggestions which are irrelevant to the user activity leading to unwanted or ignorant. The recommendation system incorporates content based filtering and collaborative filtering with matrix factorisation playing a part in the arrangement and calculation of data for recommendation. In collaborative filtering the suggestions are based on many patterns which are similar to user's views and the ratings and in content based filtering all the suggestions are based on the views and ratings of a single user's activity. Now the acquired reviews and ratings are processed by filtering techniques for recommendation. While processing the dataset for training and testing, the missing or incomplete data will slower the process which can lead to inconsistencies in the recommendation. In this we are mainly concerned on recommendation of movies with less user activity for better suggestion of movies to entertain based on the user's reviews and ratings.

I. INTRODUCTION

The creative industry has a huge influence in people's lives. People consuming the products of creative industry is increasing day to day. With the inclusion of popular media platforms along with the conventional movies and TV series, the data to be handled is growing at an exponential rate. This data was handled by humans previously but now seems impossible without automation.

III. OBJECTIVE AND SCOPE

The sole purpose of movie recommendation system is to recommend or predict the films which are in best interest for the end user according to their online movie streaming portal activity. However, the prediction does not always succeed because of lack of information. Due to the exponential rise of the online platforms they offer a pool of options for movies, online series, music, goods and etc, it has become crucial for these platforms to provide this service as they can save time and money of the user and also can introduce new commodities for recommendation.

The recommender system is a driving engine for these kinds of online platforms. These Platforms use recommendation system for helping them to search contents in websites like Netflix, Amazon prime, Zee5, Hotstar etc. The ability of the recommender system is a huge reason for the success of the online streaming platforms because the movies and the series provided by these online streaming websites are same, the deciding factor for the success is the amount of effort that it reduces for searching movies or series

This paper focuses on recommendations of movies using machine learning that is these movies will be recommended to the end user according to their tastes. The recommendation will be according to the genres the user has most watched movies from, according to the rating of the movie, and reviews of viewers.

Machine learning is used to make the background for the project, the machine will access the data, analyse them and find the best recommendation for the end user and provide them with options for the end user.

II. DOMAIN

The Machine learning is a subclass of Artificial Intelligence to acquire logic and recommend solutions of the problems in current scenario with the usage of demographic approach and data. Machine learning supports the systems to lucubrate from the data and situations given in the input in form of collection of data known as datasets. The usage of mathematical theories, methods, and applications are done in-order to verify the adaptability of the Machine learning to learn from datasets and situations which allows the system to recognise and provide the potential or applicable solutions which are mathematically derived and thus trustable. Machine learning helps in building the decisions with help of the analytical models, historical relationships and trends in the data which produces credible results with Iteration of the result.

for the end user. The other factors include the user interface, user options, subtitles and other such options. There are 2 fundamental modules that is present in all recommendation system contains Collaborative and content based filtering. They generally use a typical matrix factorization module which divides ratings and movie ID's in a matrix form and is represented in a user-item matrix fashion. The challenges faced by the recommender systems include:

1. Lot of data of the users are required for recommendation. Sometimes even user's personal data is used for evaluation for recommendation.
2. Unpredictable due to inadequate information from the user's steaming activity.

IV. ARCHITECTURE OF THE SYSTEM

The project focuses on using content based filtering and collaborative filtering algorithm to filter different features based reviews and rating to a give precise suggestions for every user. Also, the project prevents the incorrect reviews, comments and suggestions for every user.

The architecture diagram for the system developed is given as the following. The movie title and link is sent towards the pre-processing stage, which also takes ratings, suggestions, and reviews as its input. The pre-processing stage leads to training dataset which then is sent towards the testing stage. In the testing stage the output of the recommender system is matched with the response of the user. The predictions are made after the testing stage, where the recommendations are shown to the end user. A complete structure including all these steps forms the model of this project.

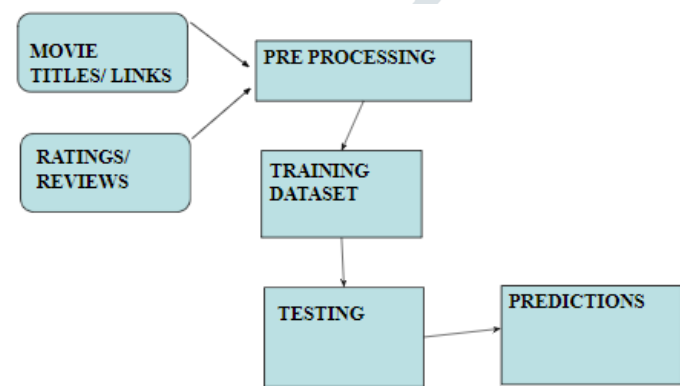


Fig. 1 - The architecture diagram for the system developed

The UML diagram used for the model starts with a feedback mechanism from the end user. The preferences are set and reviews are read, new index search are acquired. The models are trained in an interval of few days. Then monitor users gives reviews and feedbacks and also provide recommendation. The system interacts with the last four stages of the structure while the user interacts with the feedback and review and suggestion reading. Collaborative filtering is a technique of making automatic filtering about the needs of a user by adhering their requirements or taste information. Content based recommender system, is used to define the products and a profile of user is constructed to denote the kind of product this user's tastes. Alternatively, these algorithms try to provide recommendations that are similar to the likes and tastes of the user.

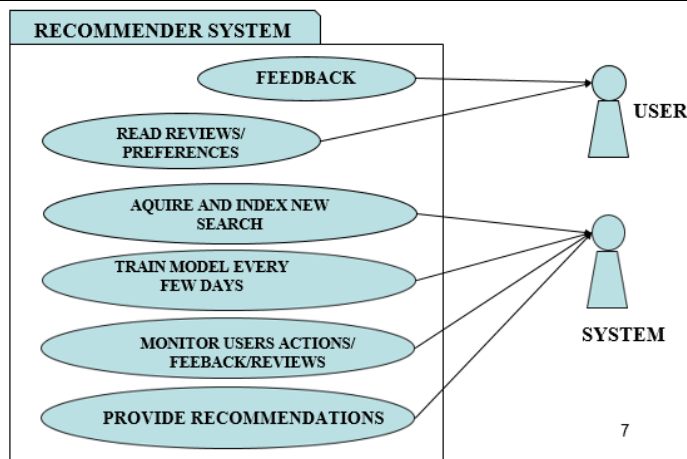
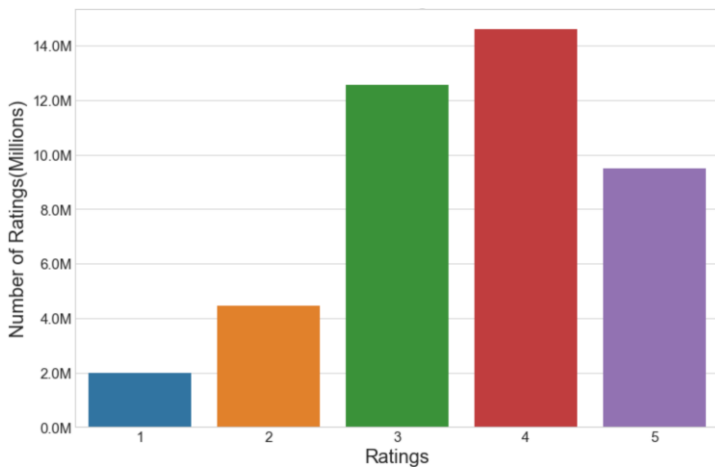


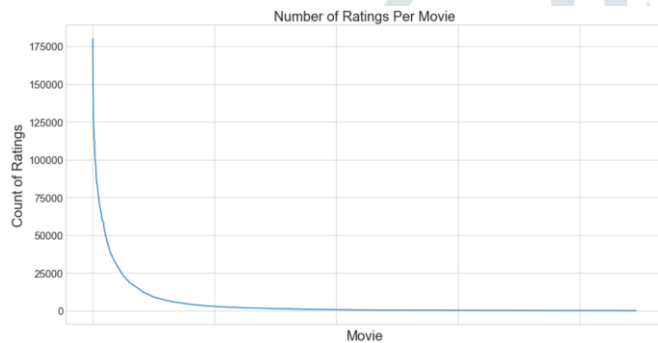
Fig. 2 - The UML diagram for the system developed

V. PROPOSED SYSTEM

The proposed system for recommendation system is collaborative filtering and content based filtering with each of modules executed together. The probabilistic matrix factorization (pmf) and Bayesian factorization (bpmf) uses 20 items for helping the system for fast process and . The pmf and bpmf helps in the matrix split in user-item for having accurate information from the dataset. The difference between the pmf and bpmf decides the type of matrix that builds up for having user-item matrix. First the dataset is pre processed for clearing incomplete or duplicate information from the dataset which done by comparing with the required datasets. Dataset is split into 2 parts mainly for training and testing where the outcome is already known for validating the algorithm. The algorithm is trained according to the given situation, here in this case the ratings are main criteria for judging a movie for recommendation for the particular user. In training and testing the algorithm uses both collaborative filtering and as well as content based filtering for recommendation. In collaborative filtering the pmf creates the user-item matrix creating a pattern for every user which can be used for comparison with other user's matrices due to this there will be similarity between users pattern leading to more accurate and precise selection of movies or series. In content based filtering the steaming portal activity of single user is viewed to get information about like and disliking of a particular genre or sub genre, where the matrix is zeroed on the basis of that particular user's history of search and previous ratings. When this is zeroed then the prediction occurs as the user is being studied. Here the average rating per week, month and year wise is also reviewed. The distribution of the rating in the training data is explored so the worst data (in this case they are the 1/5 rated movies) are not recommended to their user. The training is a reference for the testing case where algorithm's efficiencies are test for the testing purpose.

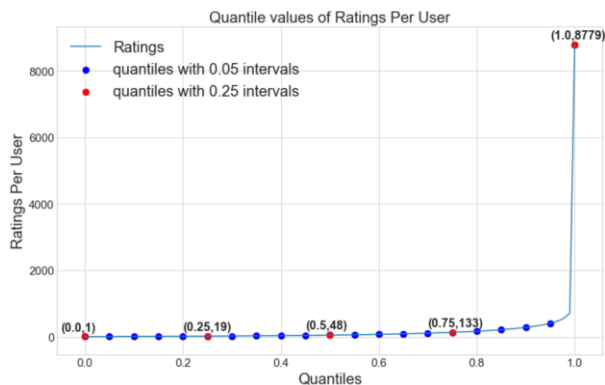


VI. RESULT



The number of ratings per movie decreases due to less rating
 fig.6.1 Number of ratings per movie

given by the users as the data frame increases. The count of ratings vs movie shows the decrement of movie as the count of rating decreases. The sparse User-item matrix forms data frame which is used for calculating the average rating per user, average rating per movie helps in finding the amount of rating an average person



gives for movies. These average calculations help in determine the pattern and the review activity in the streaming portal

fig.6.2 Quantile Values of ratings per user

which in turn shows the numerical values that is used as a reference in recommendation. The quantile graph shows the quantile value ratings per user with 0.05 and 0.25 intervals in between them. The information of rating of movies grouped by users are with count 401901.00000, mean 107.14104, Standard deviation 155.05350, Minimum rating 1.00000. We have average rating for each week and the round of is around 3.5 with variation in decimal values.

VII. CONCLUSION

There is a decrement in the Root Mean Squared Error(RMSE) in both pmf and bpmf with 0.02% and 0.03% respectively. This shows the credibility of the algorithm. The graphs represent the average rating given by users which decrease as the time increases. However the recommendation of movies are determined not only by the ratings but also depended on the user's activity and sparse user-item matrix which are calculated for having numerical values used for recommendation in the system.

REFERENCES

1. D.H. Park, H.K. Kim, I.Y. Choi, J.K. Kim, A literature review and classification of recommender systems research, *Expert Syst. Appl.* 39 (2012) 10059–10072. <http://linkinghub.elsevier.com/retrieve/pii/S0957417412002825> (accessed May 22, 2013).
2. L. Lü, M. Medo, C.H. Yeung, Y.-C. Zhang, Z.-K. Zhang, T. Zhou, *Recommender Systems*, (2012) 97. <http://arxiv.org/abs/1202.1112> (accessed August 22, 2014).
3. Z. Sun, L. Han, W. Huang, X. Wang, X. Zeng, M. Wang, et al., *Recommender systems based on social networks*, *J. Syst. Softw.* 99 (2015) 109–119. doi:10.1016/j.jss.2014.09.019.
4. J. Borràs, A. Moreno, A. Valls, *Intelligent tourism recommender systems: A survey*, *Expert Syst. Appl.* 41 (2014) 7370–7389. doi:10.1016/j.eswa.2014.06.007.
5. Z. Dehghani, S. Reza, S. Salwah, B. Salim, *A systematic review of scholar context-aware recommender systems*, *Expert Syst. Appl.* 42 (2015) 1743–1758. doi:10.1016/j.eswa.2014.09.01