2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) 978-1-7281-4142-8/20/$31.00 ©2020 IEEE 10.1109/ic-ETITE47903.2020.453

Analysis of Movie Recommendation Systems; with and without considering the low rated movies

[1]Muppana Mahesh Reddy
[1]maheshreddy.nit@gmail.com

[2]R. Sujithra @ Kanmani
[2]sujithrakanmani@gmail.com

[3]Dr. B. Surendiran
[3]surendiran@nitpy.ac.in

Dept. of Computer Science and Engineering, National Institute of Technology, Puducherry
Karaikal, India

Abstract*—Movie recommendation system is one of the top research areas, currently. Due to the impact of high internet speeds, multimedia has become one of the best entertainments. Recommendation system has its applications like movie recommendations, course recommendations, e-commerce etc.. Movie recommendation system scope is not limited to entertainment, but also in information sharing. Movie recommendation systems suffer from problems like Cold-start problem, Sparsity, Long-tail problem, Grey sheep problem etc.. Some of these problems can be solved or at least be minimized if we take the right decisions on what kind of movies to ignore, what movies to consider. This paper examines the recommendations that are obtained with and without considering the movies that have never got an above-average rating, where average rating is defined here as the mid-value between 0 and maximum rating used, for example, 2.5 in 1 to 5 rating scale. The technique used is "collaborative filtering" and the similarity measure used is the "Pearson correlation coefficient". Dataset considered is Movie-Lens-100k. This experiment result shows that low rated movies are not significant in finding the movie predictions. So it's suggestable to ignore them while calculating movie predictions.

***Keywords*—Movie Recommendation Systems**

I. INTRODUCTION

Movie recommendation systems provide the user the movie suggestions that are more likely to be watched by him using some means like, using the user's past behavior, or user's profile, or user's demographic data etc.. It's difficult for the user to find out or chose a movie to watch out of thousands of available movies and Movie Recommendation Systems provide the movie suggestions and hence save a lot of time and also introduces interesting movies that the user may not be aware of. The Paper is organized as follows: section II discusses Literature survey, Section III discusses Existing work, Section IV discusses Proposed model and Algorithm. Sections V and VI show results and conclusion respectively.

II. LITERATURE SURVEY

Lot of research discusses on Recommendation System issues like; evolution of recommender system over the time[2], Performance comparison of different recommendation algorithms[3], Comparison of various metrics used in collaborative filtering for recommendation system[4], Movie Recommendations Using the Deep Learning Approach [5], Item-Based Collaborative Filtering in Movie Recommendation in Real time [6], Recommendation Using Frequent Itemset Mining in Big Data [7], Graph-based Recommendation System[8], Employing Sparsity Removal Approach and Fuzzy C-Means Clustering Technique on a Movie Recommendation System [9], Similarity measures for collaborative filtering recommender systems[10], Film recommendation systems using matrix factorization and collaborative filtering[11], Performance analysis of recommendation system based on collaborative filtering and demographics[12].

Problems that usually arise in Recommendation systems are Cold start problem, Long tail problem, Sparsity, Shared account problem, Grey sheep problem, Scalability etc.. The Adaptive Clustering Method for the Long Tail Problem of Recommender Systems[13], Recommending Long-Tail Items Using Extended Tripartite Graphs[14], Enhancing Long Tail Recommendation Based on User's Experience Evolution[15] gives some idea on the Long Tail Problem of Recommendation Systems.

Analysis of similarity measures like Pearson correlation coefficient, Jaccard similarity, Cosine-

978-1-7281-4142-8/$31.00 ©2020 IEEE

similarity etc. has been done. Analysis of the Recommendation system's performance when different classifiers like Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbor., Artificial Neural Networks/Deep Learning, Support Vector Machine etc. are used has been good research area. Problems Recommendation system problems like Cold start problem, Long tail problem, Sparsity, Shared account problem, Grey sheep problem, Scalability etc. has been a good research area.

III. PROPOSED WORK

This Paper shows the effect of ignoring the movies that have never got a rating above average using Movie-Lens-100k dataset[1]. First, we find the predictions of the movies of the user under test taking into account all the movies, then we find the predictions of the movies of the user under test ignoring the movies that have never got an above -average rating and compares these predictions with the predictions we got previously when all the movies are considered. Consider the value '2.5' as average rating since the dataset we use has a rating range of 1 to 5. Comparison of the predictions we get when all the movies are considered, with the predictions that we get when movies that are always rated below average is done. Graphs of few random users are shown in the figures, figure 2 through figure 5, which gives a clear view of changes of predictions when movies that are always rated below average are ignored, these changes in the predictions are found to be negligible. As such movies do not have a significant contribution in rating prediction and therefore suggesting to ignore such movies.

Proposed Algorithm is shown below:

Step 1: Find the similarities using the Pearson correlation coefficient taking into account all the movies.

Step 2: Predict the ratings of the movies (that are kept for testing, in the dataset) of the user under test and store it as actual_predictions.

Step 3: Find the similarities using the Pearson correlation coefficient ignoring the movies that have never got a rating above 2.5.

Step 4: Predict the ratings of the movies (that are kept for testing, in the dataset) of the user under test using similarities obtained in step 3 and store it as predictions_without_movies_always_rated_low.

Step 5: Plot graph between actual_predictions and predictions_without_movies_always_rated_low

The flow diagram of the proposed work is shown below

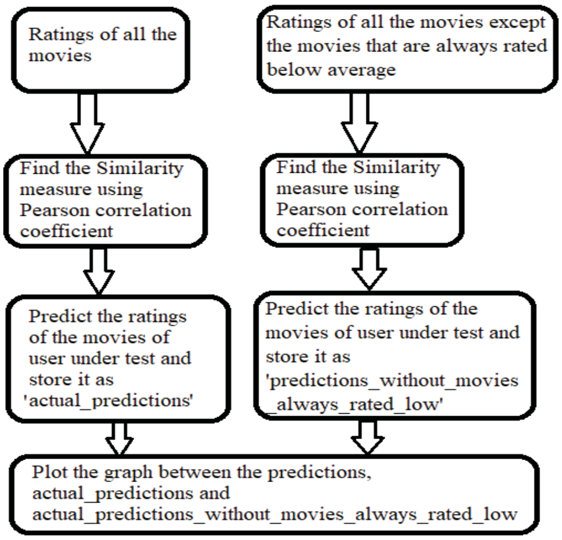

Fig. 1. Proposed model

IV. RESULTS

The dataset used is Movie-Lens-100K. MovieLens datasets were collected by the GroupLens Research Project at the University of Minnesota. This data set consists of 100,000 ratings (1-5) from 943 users on 1682 movies. The data was collected through the MovieLens web site (movielens.umn.edu[1]). In the graphs below, 'actual_predictions' is the predictions obtained by considering all movies, 'predictions_without_movies_always_rated_low' is the predictions obtained by ignoring movies that have never got a rating above 2.5. On x-axis: Movies (that are kept for testing, in the dataset) of the user under test. On y-axis: ratings. The results of four random users are shown below.

Movie predictions for the user with user-Id 254:

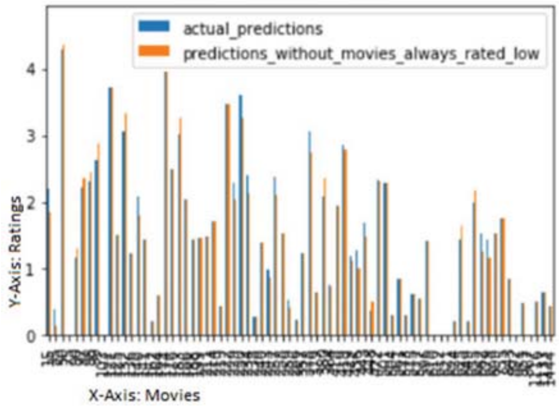

Fig. 2.Results of movie predictions of user-Id 254

From Fig. 2, it is observed that there is no significant difference between the predictions.

Movie predictions for the user with user-Id 77:

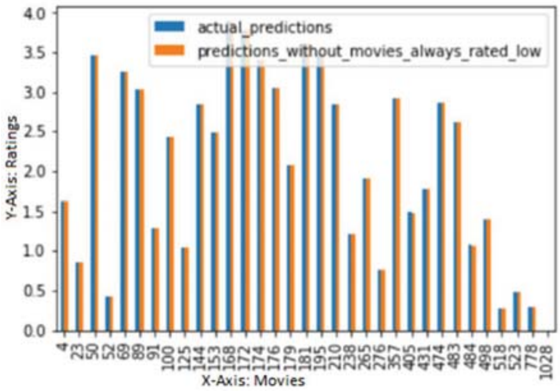

Fig. 3. Results of movie predictions of user-Id 77

From Fig. 3, it is observed that the predictions are almost same just like happened in the previous graph (Fig.2).

Movie predictions for the user with user-Id 125

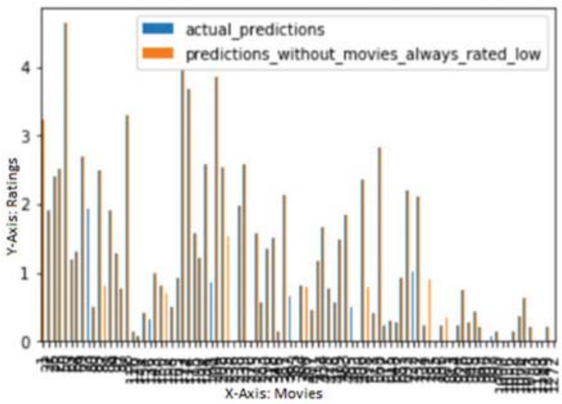

Fig. 4. Results of movie predictions of user-Id 125

From Fig. 4, it is observed that only for a very few movies the predictions are slightly varied. There is no significant difference between the predictions.

Movie predictions for the user with user-Id 457

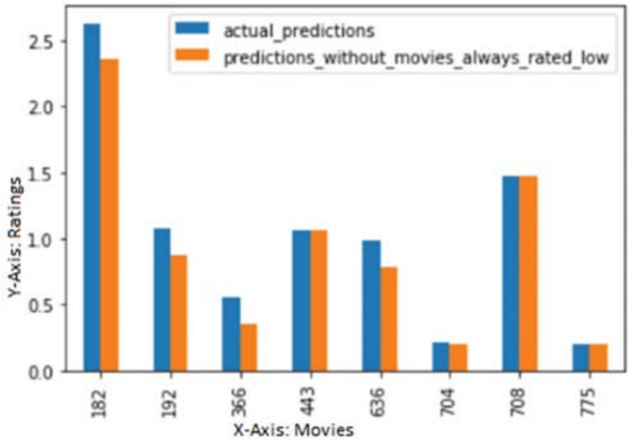

Fig.5. Results of movie predictions of user-Id 457

From Fig. 5, it is observed that there is no significant difference between the predictions. The negligible difference between the predictions shows that the effect of removing low-rated movies is negligible and hence can be removed.

V. CONCLUSIONS

The number of movies that never got a rating above 2.5 is 102 which is 6.06% of all the movies given in the Movie-Lens-100K dataset[1] which is significant but the difference between the predictions, actual_predictions and the predictions_without_movies_always_rated_low is negligible. Hence this proves that, movies that have never got an above average rating does not have significant contribution in movie recommendations and it's suggested to ignore such movies.

REFERENCES

[1] https://grouplens.org/datasets/movielens/

[2] Sinha, B.B., Dhanalakshmi, R. Evolution of recommender system over the time. Soft Comput 23, 12169–12188 (2019) doi:10.1007/s00500-019-04143-8

[3] G. Mustafa and I. Frommholz, "Performance comparison of top N recommendation algorithms," 2015 Fourth International Conference on Future Generation Communication Technology (FGCT), Luton, 2015, pp. 1-6. Publisher: IEEE

[4] A. Kumar, S. Gupta, S. K. Singh and K. K. Shukla, "Comparison of various metrics used in collaborative filtering for recommendation system," 2015 Eighth International Conference on Contemporary Computing (IC3), Noida, 2015, pp. 150-154. Publisher: IEEE

[5] J. Lund and Y. Ng, "Movie Recommendations Using the Deep Learning Approach," 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, 2018, pp. 47-54. Publisher: IEEE

[6] M. K. Kharita, A. Kumar and P. Singh, "Item-Based Collaborative Filtering in Movie Recommendation in Real time," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 340-342. Publisher: IEEE

[7] H. Kunjachan, M. Hareesh and K. Sreedevi, "Recommendation Using Frequent Itemset Mining in Big Data," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 561-566. Publisher: IEEE

[8] K. Yang and L. Toni, "GRAPH-BASED RECOMMENDATION SYSTEM," 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Anaheim, CA, USA, 2018, pp. 798-802. Publisher: IEEE

[9] N. Ifada, E. H. Prasetyo and Mula'ab, "Employing Sparsity Removal Approach and Fuzzy C-Means Clustering Technique on a Movie Recommendation System," 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM), Surabaya, Indonesia, 2018, pp. 329-334. Publisher: IEEE

[10] L. A. Hassanieh, C. A. Jaoudeh, J. B. Abdo and J. Demerjian, "Similarity measures for collaborative filtering recommender systems," 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM), Jounieh, 2018, pp. 1-5. Publisher: IEEE

[11] M. Ilhami and Suharjito, "Film recommendation systems using matrix factorization and collaborative filtering," 2014 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, 2014, pp. 1-6. Publisher: IEEE

[12] J. Gupta and J. Gadge, "Performance analysis of recommendation system based on collaborative filtering and demographics," 2015 International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, 2015, pp. 1-6. Publisher: IEEE

[13] Y. Park, "The Adaptive Clustering Method for the Long Tail Problem of Recommender Systems," in IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 8, pp. 1904-1915, Aug. 2013. Publisher: IEEE

[14] A. Luke, J. Johnson and Y. Ng, "Recommending Long-Tail Items Using Extended Tripartite Graphs," 2018 IEEE International Conference on Big Knowledge (ICBK), Singapore, 2018, pp. 123-130. Publisher: IEEE

[15] Y. Wang, J. Wang and L. Li, "Enhancing Long Tail Recommendation Based on User's Experience Evolution," 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD)), Nanjing, 2018, pp. 25-30. Publisher: IEEE