Recommendation System for Movie Cast and Crew using Datamining Algorithm

Ranjitha B R

*Abstract: **In this article, the data mining algorithms like apriori algorithm is used to suggest the best cast and crew to make a particular genre of movies so that the movie is successful. The data of cast and crew is extracted for which the users have given high average ratings and apply apriori algorithm to give recommendation. There are recommendation systems which gives the recommendation to the users so that the users can watch movies in which they are interested in. But there are no recommendation systems which gives the right information which is helpful for movie making. It is important to make good movies so that the viewers and entertained by it due to which the profit of the producers increases. This research work can be used by the movie makers to select the best cast and crew for the particular genre of movie.***

Keywords : recommendation system, apriori algorithm, cast, crew, movies.

I. INTRODUCTION

Recommendation systems are important to give required information to the people. People in the world rely on internet to get information. The information which is available on the internet is too much so, the users find it difficult to get the right amount of information which is required for them. So, the recommendation systems are developed. Different types of recommendation system are built to give the desired information to the users like book recommendation, tourist places recommendation, movie recommendation, course recommendation system etc. This article is a director recommendation system which recommends directors to the producers. More than 1000 movies are released all around the world [1]. Release of good movies gives entertainment to people and also if large number of people like the movies released then the producer who produces these good movies gets profit. To make good movies the producer needs to plan. Different attributes like budget, actors, director, producer, set locations, story writer, movie release day, competing movie releases at the same time, music, release location and target audience should be considered to make a movie which attracts people. [2] Cast, crew and the interest of the people are very important for the success of movies. It is important for the producer to know which is the best suited director to direct the particular genre of movie so that the movie becomes successful. In this article, a method is proposed to recommend the movie directors to the producers so that good movies are released in the society and the producers get more profit from the movie they produced. This reduces the work of the producer and helps them to select the best cast and crew to make the particular genre of movie. Data mining algorithms like Apriori algorithm is used for suggesting the top cast and crew who can make the particular genre of movies so that the movie will be successful and the producer who puts the money for movie get the maximum profit out of it. The rest of the paper is organized as follows. Section 2 mentions the related works in this field. Section 3 gives a brief explanation of the algorithm; Apriori. Section 4 gives information about the data set used and the simulation results followed by conclusions and future scope in Section 5. Section 6 gives all the references for this research work.

II. LITERATURE REVIEW

Several movie recommendation systems are developed which gives the recommendation to the users based on the interest of the users but there is no recommendation system for the movie makers. Dr. Manoj Kuman in his research article A Movie Recommender System: MOVREC has used k-means algorithm sort movies based on ratings given by previous users, User can find movies of interest based on experience of other users from MOVREV Collaborative filtering and content-based filtering. [3]. Christina Christakou and Andreas Stafylopatis in their research article have used Java Parser to extract information and used k-means algorithm for recommending the movies to the users. [4]. Rahul Katarya and Om Prakash Verma used Data clustering and computational intelligence to recommend movies. [5]. Zan Wang, XueYu, NanFeng, ZhenhuaWang K-means clustering coupled with genetic algorithms to recommend movies to the users. [6]. Artificial immune system with user and item based collaborative filtering is used for recommending movies to the users by AlDen Chang, Jhen-Fu Liao, Pei-Chann Chang, Chin-Hung Teng, Meng-Hui Chen. [7]. Ningning Yi, Chunfang Li, Xin Feng, Minyong Shi has used graph database and based on node and edges the movies are recommended. [8]. The research articles Qiming Diao, Minghui Qiu, Chao-Yuan Wu used Collaborative filtering and topic modelling for recommending movies based on sentiment analysis. [9]. Kim, Mucheol and Sang Oh Park has used cognitive social network model to recommend movies to the users. [10]. Azaria, Amos et al has given algorithm PUMA which uses general recommendation system as black box and increases utility of business. [11].

The collaborative filtering system predicts which items a user is likely to rate highly, purchase, or click on by Miller, Bradley & Konstan, Joseph & Riedl, John in the article PocketLens: Toward a personal recommender system. [12]. Katarya, Rahul & Verma, Om has applied particle swarm optimization to k-means algorithm to improve the efficiency of recommendation to users. [13]. There are several different methods for recommending movies to the users. But there are no systems for recommending the directors to the producers of the movie. In this article, Apriori algorithm for recommending the best director to direct a particular genre of movie.

III. DATASET

Kaggle dataset of movies is used in this research for giving the recommendations. The information regarding the cast and crew of the movie and the votes given to the movie is used in the research. Movies_metadata, keywords, credits, links, ratings are the files which are present in the dataset. Movies_metadata consists of the information regarding posters, backdrops, budget, revenue, release dates, languages, production countries and companies. Keywords.csv contains the movie plot keywords of the movie. Credits contains the information regarding the cast and crew of the movies which is used in this research. Links contains the ids of the movies. The dataset is in form of JSON object. The data pre-processing is done by removing all the columns with null values and extracting the names of all the characters in the movie and the names of the crew members to build the system.

IV. APRIORI ALGORITHM

Apriori algorithm is used to find the frequently used items in the dataset. Apriori algorithm was proposed by Agarwl for mining association rule. It is a bottom-up and breadth first approach. Apriori's principle: If an itemset is frequent, then all of its subset must also be frequent. The support of an itemset never exceeds 0 of its subset support. This is known as anti-monotype property of support. The main idea of this algorithm is, it generates k-th candidate item sets from the (k-1)-th frequent item sets. It also finds the k-th frequent itemset from the k-th candidate item sets. The algorithm gets terminated when the frequent item sets cannot be extended further. The advantage is that multiple scans are generated for candidate sets. The disadvantage is that the execution time is more as wasted in producing candidates every time, it also needs more search space and computational cost is too high. [14] Support, confidence and lift are the measures used in apriori algorithm for finding the frequent itemset in the data. Support is the ratio of the number of transactions in which the particular item appears to the total number of transactions. Confidence shows the probability of an item being in the transaction if another item is present in the transaction. Lift gives the probability of an item being in the transaction if another item is present in the transaction while taking the popularity of the item into account.

V. SIMUATION RESULTS

Director, producer, screenplay, writer, actors contribute to the success of the movie. The votes given to the movie depends on all these people who are responsible for making the movie, thus we recommend the best cast and crew for making the

particular genre of movie. the genre for which the recommendation needs to be given is taken from the user. The genres of movie, cast, crew and the average votes given to the movie on the scale of 10 are extracted from the dataset using python for giving recommendation.

The steps for giving recommendations is given below:

Step 1: The cast, crew and ratings information of all the movies which are made in that particular genre of movie is extracted.

Step 2: From this set the movies whose average rating is greater than 7 is extracted.

Step 3: Apriori algorithm is applied to the cast column in the data that is extracted.

Step 4: Apriori algorithm is applied to the crew column.

Step 5: After applying the algorithm the values are sorted.

Step 6: The cast and crew members for which the values are high is given as recommendation to the movie makers.

The apriori algorithm on cast and crew column gives the support values less than 0.003 since there are 1000s of people involved in making the movie. The results are displayed below:

Table- I: Crew Recommendations

Crew	support	confidence	lift
Haruko Sugimura, Chishu Ryu	0.002597	0.41666	106.9444
Toshirō Mifune, Daisuke Katô	0.002077	0.5	87.5
Ingrid Caven, Kurt Raab	0.002077	0.72727	280.0
Takashi Shimura,Kamatari Fujiwara	0.0023376	0.81818	175.0
Toshirō Mifune, Kamatari Fujiwara	0.0020779	0.72727	127.2727

Table- II: Cast Recommendations

Cast	support	confidence	Lift
Lyle R. Wheeler, Ben Nye	0.0033766	0.86667	151.667
Bob Weinstein, Harvey Weinstein	0.0033766	1.0	296.153
Wally Westmore, Edith Head	0.00311	0.48	132.0
Eric Fellner, Tim Bevan	0.003376	0.92857	275.0
TJoel Coen, Ethan Coen	0.003116	1.0	320.8333

VI. CONCLUSION

In this article apriori algorithm is applied to the cast and crew dataset of the movies by extracting the movies for which the rating is high. Since the dataset consists of 45000 movies and each movie has more than 5 cast and crew members the support value after applying the apriori algorithm is not more than 0.03 hence we can conclude that this algorithm is not efficient in finding the best cast and crew for making the movie of particular genre. The probability of the movie getting high votes is very less since there are several directors, producers, actors who make the movie and also whether the movie gets high rating or not also depends on individual user interest. Hence this algorithm cannot be used on cast and crew data to give the recommendations.

REFERENCES

1. A. Bhave, H. Kulkarni, V. Biramane and P. Kosamkar, "Role of different factors in predicting movie success," *2015 International Conference on Pervasive Computing (ICPC)*, Pune, 2015, pp. 1-4.

2. J. Ahmad, P. Duraisamy, A. Yousef and B. Buckles, "Movie success prediction using data mining," *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Delhi, 2017, pp. 1-4.

3. Kumar, Manoj & Yadav, D.K. & Singh, Ankur & Kr, Vijay. (2015). A Movie Recommender System: MOVREC. International Journal of Computer Applications. 124. 7-11. 10.5120/ijca2015904111.

4. C. Christakou and A. Stafylopatis, "A hybrid movie recommender system based on neural networks," *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*, Warsaw, 2005, pp. 500-505.

5. Katarya, Rahul & Verma, Om. (2016). An effective collaborative movie recommender system with cuckoo search. Egyptian Informatics Journal. 18. 10.1016/j.eij.2016.10.002.

6. Wang, Zan & Yu, Xue & Feng, Nan & Wang, Zhenhua. (2014). An Improved Collaborative Movie Recommendation System using Computational Intelligence. Journal of Visual Languages & Computing. 25. 10.1016/j.jvlc.2014.09.011.

7. A. D. Chang, J. Liao, P. Chang, C. Teng and M. Chen, "Application of artificial immune systems combines collaborative filtering in movie recommendation system," *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, Hsinchu, 2014, pp. 277-282.

8. N. Yi, C. Li, X. Feng and M. Shi, "Design and Implementation of Movie Recommender System Based on Graph Database," *2017 14th Web Information Systems and Applications Conference (WISA)*, Liuzhou, 2017, pp. 132-135.

9. Diao, Qiming & Qiu, Minghui & Wu, Chao-Yuan & J. Smola, Alexander & Jiang, Jing & Wang, Chong. (2014). Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 10.1145/2623330.2623758.

10. Kim, Mucheol and Sang Oh Park. "Group affinity based social trust model for an intelligent movie recommender system." *Multimedia Tools and Applications* 64 (2011): 505-516.

11. Azaria, Amos et al. "Movie recommender system for profit maximization." *RecSys*(2013).

12. Miller, Bradley & Konstan, Joseph & Riedl, John. (2004). PocketLens: Toward a personal recommender system.. ACM Trans. Inf. Syst.. 22. 437-476.

13. Katarya, Rahul & Verma, Om. (2016). A collaborative recommender system enhanced with particle swarm optimization technique. Multimedia Tools and Applications. 75. 10.1007/s11042-016-3481-4.

14. Kavitha, Mrs. M. and Ms. S. T. Tamil Selvi. "Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons." (2016).

AUTHORS PROFILE

Ms. Ranjitha B R, MSc computer Science Christ (Deemed to be university) department of computer science.