

Movie Recommendation System

Bennet Praba¹, Abhinav Kumar², Prakhar Gupta³

¹Asst. Prof., Dept. Computer Science and Engg., SRM Institute Of Science And Technology, Chennai, India

^{2,3}Student, Dept. Computer Science and Engg., SRM Institute Of Science And Technology, Chennai, India

Abstract—Recommendation systems have gained tremendous concern over the last recent years. As all know recommendation systems have provided a boon in the field of online shopping and other browsing portals. Recommendations system use machine learning techniques to predict what the user may like based on his history of transaction with a system full of items. At present there are many ways in which implementation a recommendation system can take place. The proposed system with the help of data collected brings out a list of possible movies a user can like.

The algorithm that works will determine the type of research that will happen. Hence, the designed system successfully implements the algorithm that will bring the type of results which will be very close to users likes. The system architecture is such that first the user will give ratings to the movies, then the recommended movies will come depending upon the type of ratings given the different type of movies. The given ratings will be stored in the data base.

Index Terms—Recommendations, efficient, database, algorithm

I. INTRODUCTION

On the online transactions & many other internet related things where many choices are rash, there is a huge amount of data present online which is very difficult to manage as a user. Hence there is a need for recommendation system in order filter out the unwanted information so as the user is only left with the information which is useful for him.

The recommendation system are based on many different selective criteria for example on the basis of content, on collaboration .The proposed system is of movie recommendation system in which the user will be able to have a list of all the movies of his choice with the help of ratings given by him.

The recommendation system will be proper and will be filtering out the useful information fragment from a large amount of dynamic info based on user preference , internal and behavior about the item, which all depends upon the ratings he has given. The data will be collected from the data base where all the user ratings will be stored and later on that information will be vital in carrying out the movie recommendations. For example suppose a user searches for a movie online the person will find a huge never ending list. Through this recommender system we tend to simplify the job of searcher so that finding movies of similar interest prove to be not so much of a headache.

The proposed system is better as the proposed system will take the information direct from the user itself.

By this way the chances of the suggested movie by the system has a greater probability of getting liked by the person. Clearly the choice of movie suggested will totally depend on the highest ratings given to the type of movie liked by the user. This proposed system has better chance to be accurate than any other earlier proposed system.

II. RELATED WORKS

A huge number of recommended systems are being proposed now a day's which work both on collaborative filtering and content based filtering. Over the past few years the recommender system have taken a great turn in the field of movie recommendations one of the best type of system is that a recommender system was proposed by using a text categorization techniques to learn from movie synopses. The recommending process was comprised of the following process, first the user was asked to divide a particular no of movies into different categories be it

- a) Terrible
- b) Average
- c) Above average
- d) Good
- e) Excellent

Another example of the latest recommender system proposed was by one of the world's best Dictionary site Wikipedia. Wikipedia was used to estimate the similarity between the movies and therefore was able to provide more accurate predictions for Netflix prize competition.

Various algorithms have been designed in order to design a better movie recommender system which will combine the tag based profiling techniques compared with additional content based recommender strategies. But the problem with these is that none of these proposed systems were accurate i.e. none of the systems demonstrated any significant improvement.

III. SYSTEM ARCHITECTURE

The system architecture consists of two types of information storing databases one which is viewing events and the other is purchasing events. The next part is training algorithm in which the coding is done in order to correlate all the previous user ratings and cross examining all the user ratings and the resultant list of the recommended movie is obtained. The list which is obtained is also termed as the indicators. Now the further fields of the movie recommendation architecture consists of User

viewing history.

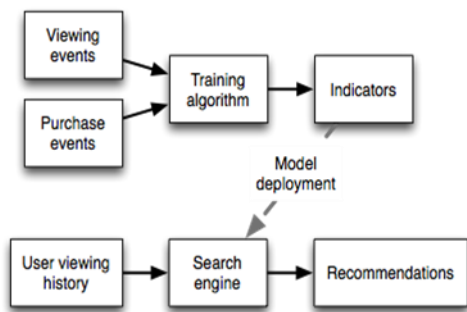


Fig. 1. System architecture

IV. PROPOSED SYSTEM

The proposed system consists of viewing events, purchase events, the training algorithm, and indicators.

To start with two different types of data have been stored which contain the different information about a wide varieties of movies. Depending on the type of information about the movie the categories are divided into viewing events and purchase events.

A. Viewing Events

The viewing events category consists of the information related to the movie about its genre and its name and year .these are the main topics in which the movie information is stored.

B. Purchasing Events

The purchasing events consists of the information about the movie .for example for every movie the user id has been assigned and the rating of all the movies have been collected online from various number of sites online. Hence this is one of the prominent reasons which proves how much the recommendation list will be accurate. Both of the above data is collected online and is been collected and compared from a lot of different sites. Collection of data online from sites like Wikipedia and ‘wiki how’ and one of the viewers most trusted sites like IMDB has been done.

The next phase of the program is training algorithm. With the help of this algorithm the movie recommendations will be done. The list of recommended movie list will come as the output based on the information stored. The following is the explanation of how the users ratings based filtration works. Suppose there are three users User A User B User C and there are 3 movies 1,2,3. All the 3 users have watched all the 3 movies. User A liked movie 1 and 2, User B liked 2 and 3. User C liked movie 1 so what are the chances that he will like 3 or 2. Based on the information the selection of the movie will take place. As from the given information both the users A and C like the movie 1 there is a greater chance that the second movie that the user c will like will be movie C. The above mentioned is one of the example will demonstrate the idea of the proposed system.

The third and the last working phase of our project is the indicator. The indicator is the output obtained i.e the list of

movies recommended based on the users choices and users ratings

V. FEATURE ENCODING

Feature encoding refers to the various features this system possess. The proposed system collects the data from online market from various sites and then compares to the list of ratings given by the user and hence comes to an output which is a list of recommended movies. The following are the feature which are possessed by the proposed system features. The data is being compared and collected on the fields such as titles, genres, directors and short abstracts can be easily grouped into these three types of features.

A. Movie Synopsis

The synopsis can be explained as the short description of the movie content. It is very similar to document modeling, words from all the synopses are extracted in the given data set and stemming is subsequently applied to each word. Words like ‘A’, ‘The’, ‘An’ are not compared while the selection for the recommended movies is done. A vocabulary is then constructed for forming a graph for each synopsis. Thus, each synopsis can be represented by a graph X.

B. Actor

The another aspect is the keeping a track of the actor list. As many people are fan of few actors and hence tend to watch all the movies of them. So keeping a record on the actor list can be important sometimes. The proposed system’s future enhancements examples is introducing a new field in ACTOR .The system will be able to analyze the users favorite list of actors and hence the recommendations will be done keeping in mind the actors list too.

C. User Comment

With such an advancement people are able to share their views on any platforms via simply “commenting ‘.This is the power of the User comment. Given a chance a popular movie may receive hundreds s of comments from different users. One possible way to analyze the comments on the movies is by grouping the comments together and then analyzing the comment section. This is a simple way, however, cannot capture the discriminative information of different user opinions. In this paper, we introduce a new method to project these user comments into a vector.

VI. BENEFICIARY OF THE SYSTEM

The following written are the benefits i.e. the advantages of the system,

- 1) The main aim and objective of the system which was designing a very accurate recommendation system has been complete
- 2) With the combination of both content based and collaborative filtering the system has designed which makes the system more and more accurate as well as

very precise in its research.

- 3) The collection of the data has been done from very profound sites hence the data is correct and accurate.
- 4) The algorithm designed has been such that the result obtained is optimum.

VII. EXPERIMENTS

TABLE I
MOVIES AND RATINGS

title	Correlation	num of ratings
Liar Liar (1997)	1.000000	485
Batman Forever (1995)	0.516968	114
Mask, The (1994)	0.484650	129
Down Periscope (1996)	0.472681	101
Con Air (1997)	0.469828	137

The above table has been created with the help of data which is collected online from all the different sites particularly about these movies

The table shows the co relation between them and no of ratings given to each movies. This is an example graph which shows the relation between the movies which are old. As it can be easily inferred from the graph that the movie selection year is in the period 1990-1999.

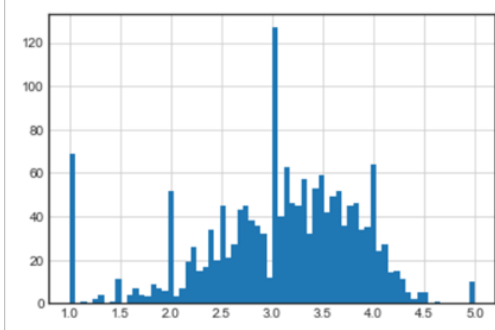


Fig. 2. Rating graph

The above graph has been statically drawn comparing the rating that have been given to the above listen movies which came out during the period 1990-1999. The graph clearly shows that most of the users have given 3 star ratings to the movies. Now these ratings will be further compared to the movies of same group and another set of data will be obtained, which will further help us to make our system more precise.

VIII. PROSPECTIVE AND CONCLUSION

The proposed system has the following future scope. The movie recommendation system can also be designed better once the user recommendations could be done from users search history also. ML (machine learning) will be used in order to get the results of the user based search history.

The another way in which the recommendations can be made easy is that designing the search engine itself. Hence the user will be able directly search the movie and compare the ratings and read about the genre, movie actor, actresses.

The search engine designed will directly show the output rather than showing the list of movie recommendations which we all know the process is very time consuming. With the help of ML the process of movie recommendations can be made very simple and easy but a whole lot information about the movies are required. Hence designing the information of all the movies present till date won't be possible. This has proven to be one of the prominent reasons why the recommendations systems doesn't still use machine learning as their domain.

REFERENCES

- [1] Gonçalves, Diogo; Costa, Miguel; Couto, Francisco M. (2016-09-15). "A Flexible Recommendation System for Cable TV". 3rd Workshop on Recommendation Systems for Television and online Video (RecSysTV), At Boston, MA, USA.
- [2] Montaner, Miquel, L, pez, Beatriz, de la Rosa, Josep Llu{\i}s (2002). "Developing trust in recommender agents". Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1. pp. 304–305.
- [3] Beel, Joeran, Langer, Stefan, Genzmehr, Marcel (September 2013). "Sponsored vs. Organic (Research Paper) Recommendations and the Impact of Labeling". In Trond Aalberg; Milena Dobрева; Christos Papatheodorou; Giannis Tsakonas; Charles Farrugia. Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), pp. 395–399.
- [4] Yong Ge; Hui Xiong; Alexander Tuzhilin; Keli Xiao; Marco Gruteser; Michael J. Pazzani (2010). An Energy-Efficient Mobile Recommender System (PDF). Proceedings of the 16th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York City, New York: ACM. pp. 899–908.
- [5] Bouneffouf, Djallel (2012), "Following the User's Interests in Mobile Context-Aware Recommender Systems: The Hybrid-e-greedy Algorithm", Proceedings of the 2012 26th International Conference on Advanced Information Networking and Applications Workshops (PDF), Lecture Notes in Computer Science, IEEE Computer Society, pp. 657–662.
- [6] Tuukka Ruotsalo; Krister Haav; Antony Stoyanov; Sylvain Roche; Elena Fani; Romina Deliai; Eetu Mäkelä; Tomi Kauppinen; Eero Hyvönen (2013). "SMARTMUSEUM: A Mobile Recommender System for the Web of Data". Web Semantics: Science, Services and Agents on the World Wide Web. Elsevier. 20: 657–662.
- [7] Bouneffouf, Djallel (2013), DRARS, A Dynamic Risk-Aware Recommender System (Ph.D.), Institut National des Télécommunications
- [8] Lohr, Steve. "A \$1 Million Research Bargain for Netflix, and Maybe a Model for Others". The New York Times.
- [9] R. Bell; Y. Koren; C. Volinsky (2007). "The BellKor solution to the Netflix Prize".
- [10] Bodoky, Thomas. "Mátrixfaktorizáció one million dollars". Index.
- [11] Lathia, N., Hailes, S., Capra, L., Amatriain, X.: Temporal diversity in recommender systems. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, pp. 210–217. ACM, New York
- [12] Turpin, Andrew H, Hersh, William (2001). "Why batch and user evaluations do not give the same results". Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 225–231.
- [13] Beel, Joeran; Genzmehr, Marcel; Langer, Stefan; Nürnberger, Andreas; Gipp, Bela (2013-01-01). "A Comparative Analysis of Offline and Online Evaluations and Discussion of Research Paper Recommender System Evaluation". Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation. RepSys '13. New York, NY, USA: ACM: 7–14.

- [14] K. Lakiotaki, N. F. Matsatsinis and A. Tsoukias, "Multicriteria User Modeling in Recommender Systems," in *IEEE Intelligent Systems*, vol. 26, no. 2, pp. 64-76, March-April 2011.
- [15] Adomavicius G., Manouselis N., Kwon Y. (2011) Multi-Criteria Recommender Systems. In: Ricci F., Rokach L., Shapira B., Kantor P. (eds) Recommender Systems Handbook. Springer, Boston, MA.