International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072

Hybrid Recommendation System For Movies

Rohan Nayak[1], Aniket Mirajkar[2], Jeetesh Rokade[3], Prof. Girish Wadhwa[4]

[1,2,3] B.E. Student, Department of Information Technology Engineering , Vidyalankar Institute of Technology, Maharashtra, India
[4]Assistant Professor, Department of Information Technology Engineering , Vidyalankar Institute of Technology, Maharashtra, India

---

Abstract - *Hybrid Recommendation System for Movies uses the combination of collaborative and content based filtering in the context of web-based recommender systems. In particular, we will link the well-known Movie Lens rating data with supplementary IMDB content information. The resulting network of user-item relations and associated content features will be converted into a unified mathematical model, which is applicable to our underlying neighbor-based prediction algorithm. By means of various experiments, we will demonstrate the influence of supplementary user as well as item features on the prediction accuracy of our proposed hybrid recommender. In order to decrease system runtime and to reveal latent user and item relations, we will factorize our hybrid model via singular value decomposition (SVD). Due to the enormous amount of information available online, the need for highly developed personalisation and filtering systems is growing permanently. Recommendation systems constitute a specific type of information filtering that attempt to present items according to the interests expressed by a user.*

Key Words- *Content-Based-Filtering, Collaborative Filtering, prediction, hybrid recommender, IMDB, personalisation.*

1. INTRODUCTION

Due to the enormous amount of information available online, the need for highly developed personalization and filtering systems is growing permanently. Recommender systems constitute a specific type of information filtering that attempt to present items according the interests expressed by a user[1]. Most web recommenders are employed for e-commerce applications or customer adapted websites, which assist users in decision making by providing personalized information[5]. Modern recommendation systems make use of two basic types of recommendation techniques, namely content-based filtering and collaborative filtering. Beside recommendation precision, computation efficiency is a key consideration in all fields of computer science . Usually a recommender needs to deal with millions of users and items, computing rating estimations in an instant or even in real time. Under the restrictions of memory and time consumption many prediction algorithms quickly reach their limit of possible manageable data volume. In order to handle large- scale datasets, further improvements on information representation and recommendation modeling need to be done.

2. AIM AND OBJECTIVES

1. The main aim would be to develop a hybrid recommender system which incorporates and enhances properties of existing recommendation systems along with a new approach in order to decrease system runtime and to reveal latent user and item relations with great accuracy.

2. Developing a popularity score which will help users judge the movie in a better way and success prediction for movies before release will provide better feedback to movie makers.

3. To find a general way to make recommendation methods more effective in a broader range of applications. Although our experiments merely focus on one specific dataset, we desire to develop a universal model that can be applied to any other problem domain.

3. PROBLEM STATEMENT AND SCOPE

Formally speaking we aim to develop a recommendation system that enhances the properties of existing system with a newer and a more efficient approach that reduces the system run time and determine item relations with a greater accuracy. The project scope encompasses a hybrid recommendation system which will make use of item-based and user-based filtering to provide personalized recommendations. The project will incorporate sentiment analysis based on movie reviews and will also incorporate a success predictor to estimate the success rate of upcoming movies based on various parameters. The most obvious ideas is to add features to suggest movies with common actors, directors or writers.

4. PROPOSED SYSTEM

The proposed system will be a website that will contain a database consisting many movies. New users will have to sign up using the user interface provided on the website. The users will be asked to provide feedback on certain movies and movie genre. Based on the feedback provided, the user will be segregated, and a set of recommendations will be provided. Real time analysis ensures that the system will adopt dynamically based on user behavior[5]. Registered users will be able to access various features such as viewing movie details, add movies to watchlist.

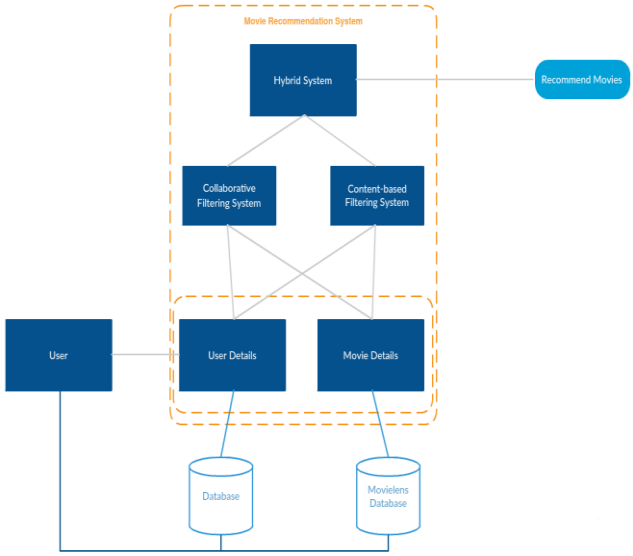

Fig -1: System Architecture

5. METHODOLOGY

We have planned on using the following approaches to solve the given problem: Datamining tool (WEKA),R-programming language . Hadoop, JavaScript, Apache Spark Data mining tools like WEKA will be used for performing database operations such as classification, clustering and outlier analysis. Hadoop, R- programming and Apache Spark will be used for developing the back-end and performing real time analysis. The front-end will be developed using JavaScript and HTML/CSS. For our intended movie recommendation system we decided to use the feature combination strategy [10], making use of a contributing recommender that provides supplementary features for the actual recommender. To be more specific, content features are utilized to strengthen collaborative recommendation. Instead of employing the retrieved content features to interpolate missing ratings [11], new features are attached to the sparse user-item matrix right away. The general idea behind this approach is that the expanded matrix gives more precise information about item-item similarities. Due to the fact that we are going to employ the item-based Nearest Neighborhood algorithm for our examination, these item similarities will have a significant influence on the final recommendation results.
Another crucial point for the success of our novel hybrid recommender system is an elaborated normalization method. This is due to the fact that rating information usually exhibit large user and item effects [8]. Some user tend to give higher ratings than others, and likewise some items receive higher ratings than others. Collaborative Filtering (CF) is the process of evaluating information using the opinion of other people [4]. Typically, predictions about user interests are made by collecting taste information from many other similar users. Thereby it is assumed that those individuals agreed in the past tend to agree again in the future. Often CF systems need to process huge amounts of information, including large-scale datasets such as in electronic commerce and web applications. Within the last decade CF has been improved continually and finally became one of the most prominent personalization techniques in the field of recommendation systems.

6. DATASET

We are considering the Movie Lens dataset which contains approximately 100k ratings distributed across 943 users and 1682 movies. Addtionally, Movie Lens also provides user information such as gender, age, occupation, etc. For retrieving information related to movies, our project makes use of IMDB which is available online. IMDB contains item features such as movie title, cast, release date, imdb url, etc.

7. ALGORITHMS

For our project, we focused on two main algorithms for recommendations:-

Collaborative filtering & Content-based filtering.

1.Collaborative Filtering:-

Collaborative Filtering techniques make recommendations for a user based on ratings and preferences data of many users[7]. The main underlying idea is that if two users have both liked certain common items, then the items that one user has liked that the other user. has not yet tried can be recommended to him. We see collaborative filtering techniques in action on various Internet platforms such as Amazon.com, Netflix, Facebook. We are recommended items based on the ratings and purchase data that these platforms collect from their user base. We also explored one algorithm for Collaborative Filtering known as the Nearest Neighbors Algorithm. This approach relies on the idea that users who have similar rating behaviors so far, share the same tastes and will likely exhibit similar rating behaviors going forward[2]. The algorithm first computes the similarity between users by using the row vector in the ratings matrix corresponding to a user as a representation for that user. The similarity is computed by using either cosine similarity or Pearson Correlation. [13]In order to predict the rating for a particular user for a given movie j, we find the top k similar users to this particular user and then take a weighted average of the ratings of the k similar users with the weights being the similarity values. Now we move on further to the Second algorithm known as Content Based Filtering Algorithm.

2.Content Based Filtering Algorithm

Content Based Filtering algorithm takes into account the likes and dislikes of the user and generates a User Profile. For generating a user profile, we take into account the item profiles(vector describing an item) and their corresponding user rating[6]. The user profile is the

weighted sum of the item profiles with weights being the ratings user rated. Once the user profile is generated, we calculate the similarity of the user profile with all the items in the dataset, which is calculated using cosine similarity between the user profile and item profile. Advantages of Content Based approach is that data of other users is not required and the recommender engine can recommend new items which are not rated currently [8].

8. DESIGN

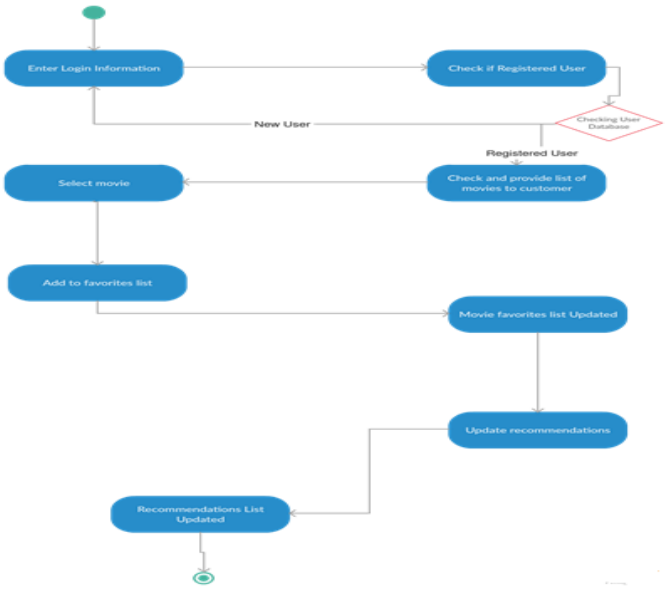

Fig -2: Activity Diagram

Fig -3: Use Case Diagram

The Above Diagram shows that the User has the following authorities:-

- He can create and edit his/her account
- He can view profiles
- He can search for various movies
- He can rate and edit movie rating as per his choice

The Admin has the following authorities:-

- He can edit account only if the user permits him to do so
- He can view profiles
- He can search for movies
- He can edit movie description

The Movie Maker has the following authorities

- He can edit movie description
- He can check for success prediction

9. FEASIBILITY STUDY

1. Technology Considerations

Movie recommendation systems available in the market are dependent on the dataset to contain large of clusters of similar users and items.[9] They also do not provide services such as effective remote access via cloud, customer interaction modules, etc. to be solved with the proposed system.

2. Product/Service Marketplace

The Movie recommendation system will impact client institutions in several ways. The following provides a high-level explanation of how the organization, tools, processes, and roles and responsibilities will be affected as a result of the movie recommendation system implementation:-

Tools: The existing requirement for on site management systems will be eliminated completely with the availability of a cloud based system.

Processes: With the Movie recommendation system comes more efficient and streamlined administrative and customer relations processes.

Hardware/Software: Clients will need to handle no extra software or hardware apart from a stable high speed Internet connection and a computer device.

3. Operational Feasibility

The project will be implemented in a way that it will allow the functioning of recommendations smoothly.[13] It will provide a user-friendly user interface in a modular fashion.

10. CONCLUSION

There are plenty of way to expand on the work done in this project. Firstly, the content based method can be expanded to include more criteria to help categorize the movies. The most obvious ideas is to add features to suggest movies with common actors, directors or writers. In addition, movies released within the same time period could also receive a boost in likelihood for recommendation. Similarly, the movies total gross could be used to identify a users taste in terms of whether he/she prefers large release blockbusters, or smaller indie films. However, the above ideas may lead to overfitting, given that a users taste can be highly varied, and we only have a guarantee that 20 movies (less than 0.2%) have been reviewed by the user.

REFERENCES

[1] Adomavicius and Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEETKDE: IEEE Transactions on Knowledge and Data Engineering, 17, 2005.

[2] R. M. Bell and Y. Koren. Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In ICDM '07: Proceedings of the 2007 Seventh IEEE International Conference on Data Mining, pages 43–52, Washington, DC, USA, 2007. IEEE Computer Society.

[3] R. D. Burke. Hybrid web recommender systems. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, The Adaptive Web, Methods and Strategies of Web Personalization, volume 4321 of Lecture Notes in Computer Science, pages 377–408. Springer, 2007.

[4] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 426–434, New York, NY, USA, 2008. ACM.

[5] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7(1), 2003.

[6] P. Melville, R. J. Mooney, and R. Nagarajan .Content boosted collaborative filtering for improved recommendations. In ACM SIGIR Workshop on Recommender Systems, pages 187–192, 2002.

[7] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, The Adaptive Web, volume 4321 of Lecture Notes in Computer Science, chapter 10,pages 325–341. Springer-Verlag, Berlin, Germany, May 2007.

[8] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl. Item-based collaborative filtering recommendation algorithms. In WWW '01: Proceedings of the 10th international conference on World Wide Web, pages 285–295, New York, NY, USA, 2001. ACM.

[9] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl. Application of dimensionality reduction in recommender system - a case study. In In ACM WebKDD Workshop, 2000.

[10] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen. Collaborative filtering recommender systems. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, The Adaptive Web, volume 4321 of Lecture Notes in Computer Science, chapter 9, pages 291–324. Springer-Verlag, Berlin, Germany, may 2007.

[11] D. H. Stern, R. Herbrich, and T. Graepel. Matchbox: Large scale online bayesian recommendations. In WWW '09: Proceedings of the 18th international conference on World wide web, pages 111–120, New York, NY, USA, 2009. ACM.

[12] A. Toescher, M. Jahrer, and R. Legenstein. Improved neighborhood-based algorithms for large-scale recommender systems. In KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Graz, Austria, 2008. ACM.

[13] J. Wang, A. P. de Vries, and M. J. T. Reinders. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pages 501–508