International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-03, Issue-09, Dec 2017

Survey on Recommendation System using Data Mining and Clustering Techniques

[1]Dinesh Vijay Attarde, [2]Prof. Manmohan Singh

[1]PG Student, [2]Professor, [1,2]Department of Computer Science & Engineering, RKDF School of Engineering, Indore, India

Abstract—The purpose of planned systems is to advocate the suitable appropriate things to the user using data mining and clustering techniques. Throughout this paper we tend to clarify the recommendation system connected analysis so Introduces varied techniques and approaches utilized by the recommender system User-based approach, Item based approach, Hybrid recommendation approaches and connected analysis at intervals the recommender system. Recommender systems profit the user by creating him suggestions on things that he's doubtless to buy and therefore the business by increase of sales. During this paper we tend to conjointly planned a brand new technique that overcomes the data-sparsity drawback and improve the performance accuracy.

Keywords—Recommender System, Types of Recommendation System, Recommendation methods.

I. INTRODUCTION

In everyday life, people rely on recommendations from other people by spoken words, reference letters, news reports from news media, general surveys, travel guides etc. so recommendations plays an important role in finding the best items. A recommender system is the information filtering that applies data analysis techniques to the problem of helping customers find the products they would like to purchase by producing a predicted likeness score or a list of recommended products for a given customer. Recommender systems work from a specific type of information filtering system technique that attempts to recommend information items (movies, TV program/show/episode, music, books, news, images, web pages, scientific literature etc.) or social elements (e.g. people, events or groups) that are likely to be of interest to the user [4].

The recommender system also compare the user profiles and seek to predict the ratings. With the help of Recommender systems are filtering and sorting data can be easily done. Moreover the Recommender system use opinions about the community of users and to determine content of interest using certain rules extractions. Recommendation systems are classified into 3 approaches i.e. collaborative, content based or knowledge-based method to have a better recommendation. The recommender system also compare the user profiles and seek to predict the ratings. With the help of Recommender systems are filtering and sorting data can be easily done. Moreover the Recommender system use opinions about the community of users and to determine content of interest using certain rules extractions. Recommendation systems are classified into 3 approaches i.e. collaborative, content based or knowledge-based method to have a better recommendation.

1.1 TYPES OF RECOMMENDATION SYSTEM

A. Collabrative Based Recommendation Systems

Collaborative recommender systems are basic forms of recommendation engines. In this type of recommendation engine, filtering items from a large set of alternatives is done collaboratively by users' preferences. The basic assumption in a collaborative recommender system is that if two users shared the same interests as each other in the past, they will also have similar tastes in the future. If, for example, user A and user B have similar movie preferences, and user A recently watched Titanic, which user B has not yet seen, then the idea is to recommend this unseen new movie to user B. The movie recommendations on Netflix are one good example of this type of recommender system [2].

There are two types of collaborative filtering recommender systems.

1) **User-based collaborative filtering:** In user-based collaborative filtering, recommendations are generated by considering the preferences in the user's neighborhood. User-based collaborative filtering is done in two steps as Identify similar users based on similar user preferences and Recommend new items to an active user based on the rating given by similar users on the items not rated by the active user respectively.

2) **Item-based collaborative filtering:** In item-based collaborative filtering, the recommendations are generated using the neighborhood of items. Unlike user-based collaborative filtering, we first find similarities between items and then recommend non-rated items which are similar to the items the active user has rated in past. Item-based recommender systems are constructed in two steps as Calculate the item similarity based on the item preferences and

 DOI : 10.18231/2454-9150.2017.0062

Find the top similar items to the non-rated items by active user and recommend them respectively.

B. Content Based Recommendation Systems

As the name indicates, a content-based recommender system uses the content information of the items for building the recommendation model. A content recommender system typically contains a user-profile-generation step, item-profile-generation step- and model-building step to generate recommendations for an active user. The content-based recommender system recommends items to users by taking the content or features of items and user profiles. As an example, if you have searched for videos of Lionel Messi on YouTube, then the content-based recommender system will learn your preference and recommend other videos related to Lionel Messi and other videos related to football [2].

In simpler terms, the system recommends items similar to those that the user has liked in the past. The similarity of items is calculated based on the features associated with the other compared items and is matched with the user's historical preferences.

C. Hybrid Based Recommendation Systems

This type of recommendation engine is built by combining various recommender systems to build a more robust system. By combining various recommender systems, we can replace the disadvantages of one system with the advantages of another system and thus build a more robust system. For example, by combining collaborative filtering methods, where the model fails when new items don't have ratings, with content-based systems, where feature information about the items is available, new items can be recommended more accurately and efficiently [2].

For example, if you are a frequent reader of news on Google News, the underlying recommendation engine recommends news articles to you by combining popular news articles read by people similar to you and using your personal preferences, calculated using your previous click information. With this type of recommendation system, collaborative filtering recommendations are combined with content-based recommendations before pushing recommendations.

II. LITERATURE REVIEW

In [11], the author suggested a novel clustering technique built on Latent Class Regression model (LCRM), which is basically ready to consider both the general ratings and feature-level opinion values (as extracted from textual reviews) to perceive reviewers' inclination homogeneity. In the examination, they tried the proposed recommender algorithm with two true datasets. More notably, they compared it with different related methodologies, including the non-review based technique and not-LCRM based variations.

In [12] author proposed an approach which includes item-to-item collaborative filtering to discover meaningful interesting videos among the large scale of the videos and this methodology is executed in Qizmt which is a.NET MapReduce framework. Merits are 1. Provides better recommendation for same item using interests of similar users. Demerits are 1. Does not consider similar Interests, 2. Complex to implement.

HamidrezaKoohi and KouroshKiani [13] the fuzzy C-means approach has been proposed for user-based Collaborative Filtering and its performance against different clustering approaches has been assessed. The MovieLens dataset was used to compare different clustering algorithms. the evaluated in terms of recommendation accuracy, precision, and recall. The empirical results indicated that a combination of Center of Gravity de-fuzzified Fuzzy Clustering and Pearson correlation coefficient can yield better recommendation results, compared to other techniques.

A multi-level recommendation method was proposed by Nikolaos Polatidis and, Christos K. Georgiadis [14] with its main purpose to assist users in decision making by providing recommendations of better quality. The proposed method can be applied in different online domains that use collaborative recommender systems, thus improving the overall user experience. The efficiency of the proposed method is shown by providing an extensive experimental evaluation using five real datasets and with comparisons to alternatives.

Phorasim, Phongsavanh, and Lasheng Yu [15] develop a movie recommender system using collaborative filtering technique and Kmeans, to improve data sparsity and scalability. this paper presented an approach based on user clustering to produce a recommendation for the active user by the new approach. k-means clustering technique is used to categorize users based on their interests.

III. RECOMMENDATIONS METHODS

A. Weighted Method

Here in Weighted method scores of several recommendations are combined together and it help to produce the single recommendation. The example of weighted method is P-TANGO system that uses hybrid Recommendations. Here first of all equal weight is assigned to both content and collaborative recommenders but gradually adjust the weights as the prediction of ratings are confirmed. Pazzani's combination hybrid does not use numeric scores, but rather use the output of each recommender as a set of votes, which are then combined in a consensus scheme.

B. Switching Method

Here in Switching method system uses some criterion to switch between recommendation techniques. The Daily Learner system uses a content/collaborative hybrid in which a content- based recommendation method is applied first. If the content-based system cannot make a recommendation with sufficient confidence, then a collaborative recommendation is

 DOI : 10.18231/2454-9150.2017.0062

attempted. This switching hybrid does not completely avoid problem.

C. *Mixed Method*

When large recommendations take place the mixed method come into the action. Here in this method is used in Television System used. First of all content based method is used for textual description of tv-shows and use of collaborative method for finding the preferences of the user and Recommendations from the two techniques lead to suggest a final program. With the help of this mixed method new item -start up problem can be overcome: the content-based component can be relied on to recommend new shows on the basis of their descriptions even if they have not been rated by anyone. It does not get around the “new user” start-up problem, since both the content and collaborative methods need some data about user preferences to get off the ground, but if such a system is integrated into a digital television, it can track what shows are watched (and for how long) and build its profiles accordingly.

D. *Feature Combination*

In Feature combination the features from different recommendation data sources are used together into a single recommendation algorithm. Feature combination hybrid lets the system consider collaborative data without relying on it exclusively, so it reduces the sensitivity of the system to the number of users who have rated an item. Conversely, it lets the system have information about the inherent similarity of items that are otherwise opaque to a collaborative system.

E. *Cascade*

Here in Cascade method it involves the stage process in this technique, one recommendation technique is employed to produce a ranking of candidates and a second technique refines the recommendation from the candidate set. The restaurant recommender Entree, described below, is a cascaded knowledge-based and collaborative recommender. Like Entree, it uses its knowledge of restaurants to make recommendations based on the user’s stated interests. The recommendations are placed in buckets of equal preference, and the collaborative technique is employed to break ties, further ranking the suggestions in each bucket.

F. *Feature Augumentation*

To improve the performance of a core system Feature Augmentation is used. For example Libra System makes content-based recommendations of books based on data found in Amazon.com, using a naive-bayes text classifier and this help in finding the quality of books. In Feature Augmentation one technique is used to produce a rating of an item and that information is then incorporated into the processing of the next recommendation technique. So the difference between the Cascade and augmentation are as follows: in feature augmentation the feature used by second recommendation is the one which is the output of the first one where as in cascading second recommender does not use the output of first one but the results of the two recommenders are combined in a prioritized manner [16].

G. *Metalevel*

Here two recommendation techniques can be merged by using the model generated by one as the input for another. The difference between the meta-level and augmentation is that in augmentation output of first recommender is used as input for second one where as in meta-level the entire model will be consider as a input for the second one . The first meta-level hybrid was the web filtering system Fab.

IV. Challenges And Issues Of Recommendation System

A. *Cold Start Problem*

The term derives from cars. When it’s really cold, the engine has problems with starting up, but once it reaches its optimal operating temperature, it will run smoothly. With recommendation engines, the “cold start” simply means that the circumstances are not yet optimal for the engine to provide the best possible results. In ecommerce, there are two distinct categories of cold start: product cold start and user cold starts. News sites, auction sites, ecommerce stores and classified sites all experience the product cold start. The user or visitor cold start simply means that a recommendation engine meets a new visitor for the first time. Because there is no user history about her, the system doesn’t know the personal preferences of the user. Getting to know your visitors is crucial in creating a great user experience for them [3].

B. *Data Sparsity*

In practice, many commercial recommender systems are based on large datasets. As a result, the user-item matrix used for collaborative filtering could be extremely large and sparse, which brings about the challenges in the performances of the recommendation. One typical problem caused by the data sparsity is the cold start problem. As collaborative filtering methods recommend items based on users' past preferences, new users will need to rate sufficient number of items to enable the system to capture their preferences accurately and thus provides reliable recommendations.

Similarly, new items also have the same problem. When new items are added to system, they need to be rated by substantial number of users before they could be recommended to users who have similar tastes with the ones who rated them. The new item problem does not limit the content-based recommendation, because the recommendation of an item is based on its discrete set of descriptive qualities rather than its ratings.

C. *Scalability*

As the numbers of users and items grow, traditional CF algorithms will suffer serious scalability problems. For example, with tens of millions of customers and millions of items, a CF algorithm with the complexity of n is already too

 DOI : 10.18231/2454-9150.2017.0062

large. As well, many systems need to react immediately to online requirements and make recommendations for all users regardless of their purchases and ratings history, which demands a higher scalability of a CF system. Large web companies such as Twitter use clusters of machines to scale recommendations for their millions of users, with most computations happening in very large memory machines.

D. ***Gray Sheep***

Gray sheep refers to the users whose opinions do not consistently agree or disagree with any group of people and thus do not benefit from collaborative filtering. Black sheep are the opposite group whose idiosyncratic tastes make recommendations nearly impossible. Although this is a failure of the recommender system, non-electronic recommenders also have great problems in these cases, so black sheep is an acceptable failure.

V. PROPOSED SYSTEM DESIGN

The main purpose of proposed system is to recommend best suitable item to the end user. With respect to the second decision, in order to test the performance of our model, we shall measure its capability to predict a user's true ratings or preferences, i.e. system accuracy. Following [17], we propose to use the mean absolute error (MAE) which measures how close system predictions are to the user's rating for each movie by considering the average absolute deviation between a predicted rating and the user's true rating.

$$MAE = \frac{\sum_{i=1}^{n} abs(p_i - r_i)}{N}$$

with N being the number of cases in the test set, pi the vote predicted for a movie, and ri the true rating

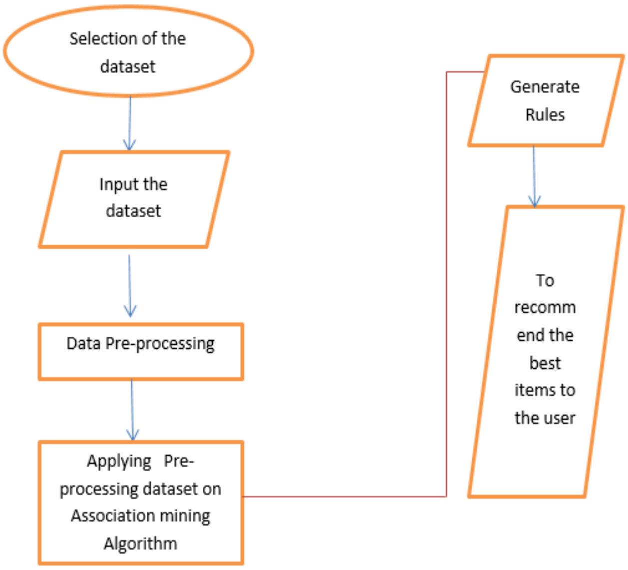

Fig. 1. System Flow

Figure 1 describes the system flow. In which we performs following steps

1. Selection of the dataset.
2. Preprocessing of the data.
3. Applying the association mining algorithm on different clustering groups.
4. Generation of the strong rules.
5. Applying priority among the best rules using top-n algorithms.
6. To recommend the best items to the user.

VI. CONCLUSION

This paper introduced the most popular data mining methods and techniques that can be implemented in the design of recommender systems. Proposed system which overcomes the data-sparsity problem and improve the performance accuracy of recommending the items to the end user. The paper mainly consist of approach where it help us to recommend the best suitable items to the user by applying association mining on clustering Moreover it also deals with various hybridization methods which is used to overcome the certain limitations.

REFERENCES

[1] Hong-Yi Chang, Shih-Chang Huang, Jia-Hao Wu, "Personalized music recommendation system based on electroencephalography feedback", © Springer Science+Business Media New York 2016

[2] Suresh Kumar Gorakala, "Building Recommendation Engines"

[3] http://www.yusp.com /blog/cold- start- problem-recommender- systems

[4] D. Fox Harrell, "Phantasmal Media: An Approach to Imagination, Computation, and Expression"

[5] Debashis Das, Laxman Sahoo, Sujoy Datta, "A Survey on Recommendation System", International Journal of Computer Applications (0975 – 8887), February 2017

[6] Sony V Hovale, Poonam G, "Survey Paper on Recommendation System using Data Mining Techniques", International Journal Of Engineering And Computer Science ISSN: 2319-7242, May 2016

[7] Srinivasa G, Archana M, Patil S S, "Survey Paper on Recommendation System using Data Mining Techniques", International Journal of Engineering and Technical Research (IJETR)ISSN: 2321-0869, Dec 2016

[8] Haruna K, Akmar Ismail M, Damiasih D, Sutopo J, Herawan T (2017) "A collaborative approach for research paper recommender system". PLoS ONE 12(10): e0184516. https://doi.org/10.1371/journal.pone.0184516 .

[9] Supriya Rajebhosale, Prof. Shilpa Gite, "Personalized Travel Recommendation Based On Hybrid Approach", International Journal for Research in Engineering Application & Management (IJREAM), Vol-02, Issue-04, July 2016, pp 01-06.

 DOI : 10.18231/2454-9150.2017.0062

[10] Mehrbakhsh Nilashi " An Overview of Data Mining Techniques in Recommender Systems", Journal of Soft Computing and Decision Support Systems (JSCDSS) E-ISSN: 2289-8603, Dec 2016.

[11] Xiang, Liang. "Recommendation system practice." BeiJing Youdian Publication House, BeiJing (2012): 6.

[12] Chen, Li, and Feng Wang. "Preference-based clustering reviews for augmenting e-commerce recommendation." Knowledge-Based Systems 50 (2013): 44-59.

[13] Jin, Yohan, et al. "MySpace video recommendation with map-reduce on qizmt." Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on. IEEE.

[14] Koohi, Hamidreza, and Kourosh Kiani. "User based Collaborative Filtering using fuzzy C-means." Measurement 91 (2016): 134-139.

[15] Polatidis, Nikolaos, and Christos K. Georgiadis. "A multi-level collaborative filtering method that improves recommendations." Expert Systems with Applications 48 (2016): 100-110.

[16] Phorasim, Phongsavanh, and Lasheng Yu. "Movies recommendation system using collaborative filtering and k-means." International Journal of Advanced Computer Research 7.29 (2017): 52.

[17] Jaimeel M. Shah1 ,and Lokesh Sahu1. "A Hybrid Based Recommendation System based on Clustering and Association."Binary Journal of Data Mining & Networking ISSN: 2229 -7170,5 (2015) 36-40.

[18] J. L. Herlocker, J.A. Konstan, L.G. Terveen, J.T. Riedl, Evaluating collaborative filtering recommender systems, ACM Transactions on Information Systems (2004) .

 DOI : 10.18231/2454-9150.2017.0062