A Critical Review on Location Based Hybrid Filtering Recommender Systems

K. G. Saranya[1], Aditya Sharma[2], Dharma Dhurai V[3], Harish J[4]

[1]Senior Assistant Professor, Department of Computer Science and Engineering, PSG College of Technology, Coimbatore

[2, 3, 4]Undergraduate Student, Department of Computer Science and Engineering, PSG College of Technology, Coimbatore

E-mail: [1]kgs.cse@psgtech.ac.in, [2]19z303@psgtech.ac.in, [2]19z312@psgtech.ac.in, [3]19z318@psgtech.ac.in

Abstract

A recommender system is basically a type of information filtering system that suggests/recommends items based on the factors that constitute what the user is most interested in. The recommendations are typically provided in relation to different decision-making processes. Tourism is a social phenomenon where people deliberately travel in search of recreation, well-being, cultural exploration or get themselves softened up. But, the amount of information available online keeps expanding at exponential rates and thus the users have expressed their feeling of frustration at how challenging it is to find the appropriate information. This problem is called information overload. This is where the recommendation system comes into play which helps in solving the information overload problem. The hybrid system addresses the disadvantages where location-based recommendation systems are used individually, of which the most notable is the cold start issue. Furthermore, in order to improve the accuracy of the prediction to recommend items, these systems search for the ideal fusion of different approaches. Thus, the hybrid recommendation method solves the challenges like 'cold start problem', inability to capture changes in user behavior, sparsity and selecting correct choices for users. This paper explores the hybrid recommendation systems and other filtering techniques used in various fields, their challenges, how they can also be used for tourism recommender systems based on the longitudes and latitudes.

Keywords: Information Overload, Cold Start, Content-Based, Collaborative Filtering, Weighted Hybrid, Switching Hybrid, Demographic Filtering.

1. Introduction

In order to recommend products that are likely to be of interest to users, recommender systems attempt to forecast their preferences. They are some of the most effective machine learning methods that internet merchants use to increase sales. A recommendation system can be advantageous for almost any organization. Thus, the applicable areas of recommendation systems are e-commerce, retail, banking, tourism and many more. E-commerce businesses are ideally placed to produce accurate suggestions because they have millions of clients and data on their online behavior. The most useful data is that related to shopping because it provides the clearest insight into a customer's intentions. Companies generating correct suggestions are led by retailers who have access to vast amounts of purchasing information. There are various types of filtering techniques discussed in this paper like the Content based, Memory based and Model based collaborative filtering techniques, Demographic and Knowledge Based filtering techniques. One of the most recent applications of recommendation systems is Netflix. The application uses their original recommendation technique called Netflix recommendation system which is used to recommend movies to the users according to their interests.

2. Literature Review

In the literature, there are several studies about the tourism recommender systems, hybrid recommender systems and how they were implemented for location-based recommendation. "Pravinkumar Swamy" et al. [1] have proposed a "Tourist Place Recommendation System" in which he used Collaborative Filtering and Content Based Filtering algorithms to recommend tourist places to users. At first, the datasets were extracted using Web Scraping technique. The extracted data were then cleaned and pre-processed. Two items are considered similar when they receive almost identical ratings from a certain user. This is the case in item based collaborative filtering. Then, a weighted mean of ratings on the items that are more comparable to the item that the user wants to rate was calculated and used as rating. This process was divided into steps of three. First, the depiction of user information, where the travel history and reviews/ratings of the user are checked. Secondly, the similarities between the tourists can be calculated on the basis of their past visiting history data and the CF algorithm proposed. Third step is to create a list of recommended places. After the CF, content-based filtering techniques are used which emphasize on the user's preferences and the properties of the item for recommendation. Here, they use the cosine similarity which gives the result in the range between 0 and 1. "Prof. P. A. Manjare et al. [2] have proposed a "Recommendation System

Based on Tourist Attraction" which finds a recommendation list of city wise tourist places and hotels using data mining techniques, analyses logged in users profile and filters the profiles using Collaborative Filtering. The system also city wise re-ranks the tourist places depending upon their ratings and reviews. Data mining techniques are used which are of different types. The first one is the User's Profile wise Collaborating Filtering in which a database of preferences for items by users is created. Different attributes are considered to describe the various degrees of detail of a user's profile. During Registration, the user will specify about their likings, his travel/search history wise collaborative filtering. When the user searches his tourist spot, the system tracks his click through data. Users will rate the tourist places. The second type is Collaborative Filtering Recommendation which calculates the accuracy of a collaborative filtering method depending on the number of items which can be associated with certain users. This application does not provide more accurate results by providing recommendations based on climate or time of the day. Does not tackle issues like location positioning etc. "Hela Masri" et al. [3] have proposed "A Personalized Hybrid Tourism Recommender System" that utilizes the three most popular recommendation techniques - The collaborative filtering (CF), the demographic filtering (DF) and the content-based filtering (CB). Various machine learning algorithms have been executed to implement and combine these recommendation techniques. The techniques are the decision tree for the DF and the K-nearest neighbors (K-NN) for both Collaborative and Content-based filtering. Firstly, data set was crawled from e-tourism website TripAdvisor.com using a web crawler called Web Harvy. The information presented in HTML format is converted to the structured data. Now, to determine the rating of the user on one particular product, the algorithm of nearest neighbor was used which computationally calculates the distance between the current item whose rating is needed and all the items that were previously rated by the user. The distance measure used was Euclid's Algorithm. This was the content-based approach. The demographic approach tries to make a user belong in a particular class based on his personal details like age, occupation gender and region to get his rating on one specific spot, so they commissioned the nodes as representative of the demographic details and the leaves as representative of the ratings. The ID3 decision tree algorithm was chosen because of its speed and the available discrete features. User based Collaborative method works by predicting the active user's ratings on the items that are not yet rated by the user. The key idea is to use the particular user and his friends' interests. Tanimoto was chosen to be the measure of similarity because it was the most accurate one for the dataset. Then, a weighted and switching hybrid technique was utilized as it alters between different recommender techniques' results to benefit from each type at different situations and

to get the most accurate rating result. This system aims to provide a systematic tour plan given the recommendations in the future. "Akbar Etebarian " et al. [4] have proposed "A hybrid recommender system based-on link prediction for movie baskets analysis" which uses a method on the basis of link prediction so that it meets the limitations of other techniques. The proposed solution in this paper contains four phases. The first phase is the Content Based Recommendation System. This phase makes sure that all users are clustered using Density-based spatial clustering of applications with noise algorithms (DBScan). After the clustering of the existing users, the new users are classified using the Deep Neural Network algorithm (DNN). The second phase consists of the Collaborative Recommender System (CRS) which uses the Hybrid Similarity Criterion. This criterion calculates the similarities between the new user and the existing users who are in the selected cluster according to a threshold. The criteria consist of gender, age and occupation. Phase three utilizes an improved Friendlink algorithm so as to calculate the similarity between the users. The final phase combines the collaborative recommender system's result and the improved Friendlink algorithm. This method faces the limitation of longer processing and execution time than existing solutions. "Riteshwari Ganjare" et al. [5] have proposed a Hybrid Recommendation System for Tourism Based Social Network, and AI, which is the implementation and combination of different types of recommender systems utilized in the tourism field. The main aim is to design a framework that clarifies the working of the Hybrid recommendation systems. The Online Social Networking (OSN) system module was developed in the first module for user registration and login. The process of user data collection involves collecting the ratings of the items based on its usage and the appreciation that it was provided, and the Demographic attributes of the user, such as occupation, age, gender, socio-professional category, location, personal status. Implementation of recommendations for new users were executed with the help of this.

The cold start problem of users was solved by this solution. For the content-based technique, the past visited places of the user are in the form of keyword vectors that are generated following an indexing phase. The social module is the collaborative phase consisting of the rating data of the consulted items by the other users. Finally, the user is recommended with the items that are considered relevant for him by the system based on his context. "Saman Forouzandeh" et al.[6] have proposed a "A Hybrid Method for Recommendation Systems based on Tourism with an Evolutionary Algorithm and Topsis Model" in which a new approach for the recommendation systems has been done in the tourism field. The approach identifies the combination of the Artificial Bee Colony technique and Fuzzy technique. The Techniques for the Order of Preference by Similarity to Ideal Solution (TOPSIS) is a multi-criteria solution

that is based on the decision-making model. It has been implemented to achieve the objective of optimization. The data was extracted from a multi question-answer on the Facebook site. The TOPSIS model defining a positive perfect solution was the first step and it was in the form consisting of a matrix with 4 columns. After that, the ABC program searches the destinations for recommending the best tourist attraction to the users. The limitation of this solution was the distribution of the questions as it was implemented by asking the questions through Facebook using several different accounts for collecting data from users. "Xixi Li" et al. [7] have proposed a "A Hybrid Recommendation Method Based on Feature for Offline Book Personalization" in which LDA model is used for the prediction of customers' choices on the topics of the books and the word2vec model is used for the calculation of the customers' choices on book types. For forecasting the book ratings, two factors have been taken into consideration. These are the similarities between the customers and the correlation between the books and buyers

3. State-of-the-Art Techniques

Collaborative Filtering

It is a technique that is widely used and can filter out the item that a user may like. It does not recommend the items using the properties of items rather than it is based on the reactions/likings of similar users. It works by means of searching a huge organization of people and locating a smaller set of users with tastes similar to a selected consumer. Then the items are preferred by the items liked by the organization of the user and then combine them to create a ranked list of pointers. The organizations are made on the similarity of the users and there are many ways to decide similarity between the users. The primary advantage of collaborative filtering is that it gives a broad spectrum of options to expose the user that other similar users show likings towards. The major disadvantages are that Collaborative filtering is unable to identify synonyms, as it totally relies on the historical data of the user and the item, initially new users don't have enough historical data (cold start). Another limitation is as the number of users will increase and the quantity of information expands the system will begin to go through a lower in performance virtually due to sheer increase in data extent.

There are two major categories of collaborative filtering:

- Memory based - It uses user's rating historical data in order to compute similarities.

- Model based - It used MLA to predict the user's rating of unrated items.

Collaborative filtering is used by many popular companies like Amazon, Netflix, etc. Amazon uses item based collaborative filtering. They use it because the average rating received by an item remains almost resolute as compared to the average ratings given by a user to different items.

Content-based filtering

It is a technique that can make recommendations to the user based on the similarities in the items as well as taking in regard the user's information. It tracks the user activity to make suggestions. The motive is to group the items with some specific keywords. The advantages of using content-based filtering techniques are that they are tailored to a user, they don't require information about other users. A large number of ratings made by people for an item is not required. The major disadvantages of using content-based filtering is that it gives suggestions purely on the basis of users current interest and doesn't allow the user to explore other potential areas. Also, the model needs to be hand engineered because it requires a vast domain knowledge of the items because of feature representation. Some popular platforms that use this type of techniques are Google, Wikipedia, etc. As in the case of google when you search for any group of words you get the results displayed which consist of those keywords
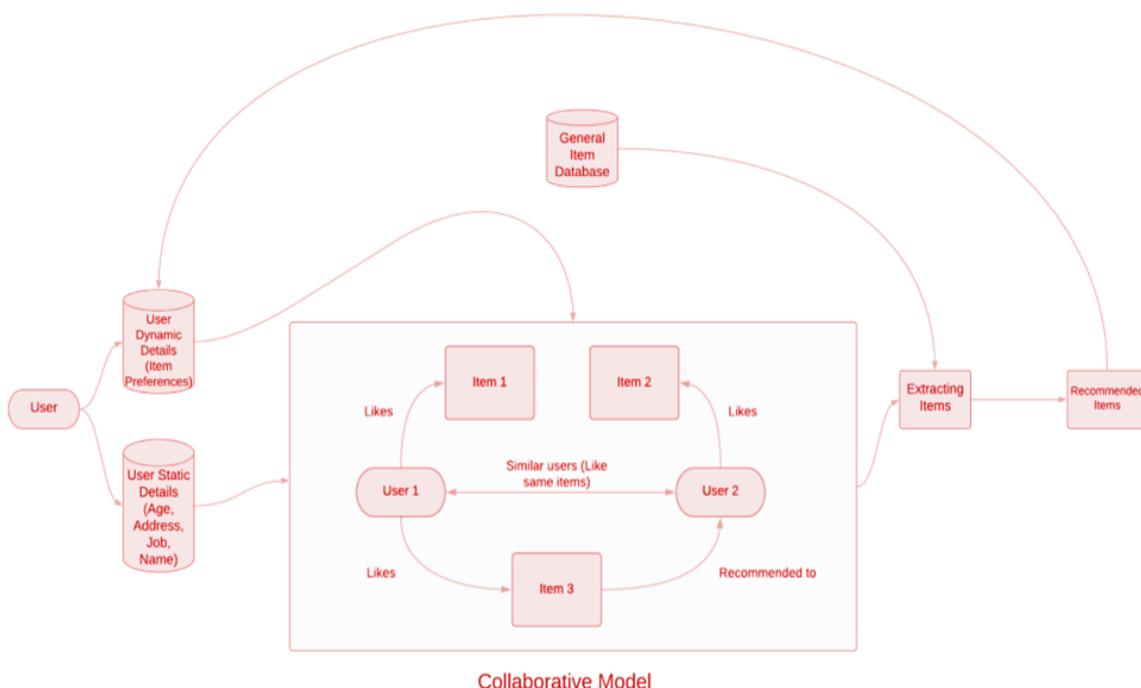
Figure 1. Generic Collaborative Model
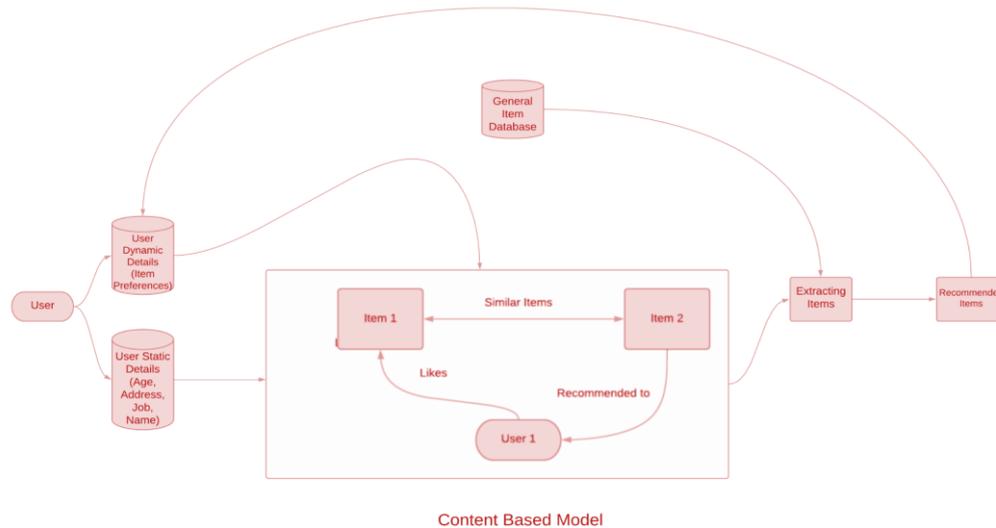
Figure 2. Generic Content Based Model

Demographic Filtering

It is a system that recommends items based on the user's demographic attributes. It doesn't require ratings. The users are classified into demographic classes on the basis of their personal attributes. The main advantage of using this filtering technique is that it is domain independent because item features are not required. They are easier to implement and don't require user ratings. Its disadvantage is that in order for recommendation it requires collection of demographic information which in turn gives rise to privacy issues. Many tourist recommendation systems use demographic filtering by utilizing the demographic data and then make recommendations based on demographic classes.

Hybrid Filtering

Hybrid filtering systems are designed by combining two or more types of filtering techniques to build a more robust model. By doing this we can eliminate the disadvantages of one type with the advantage of the other type. In paper [1] the author has created a model by combining both demographic and collaborative filtering techniques to create a movie recommendation system. In this way the recommendation quality was significantly improved. They were able to recommend by utilizing both user's demographic data as well as based on the user ratings on the movies.

Metrics

Evaluation metrics used to evaluate the recommendation system.

The common metrics used to evaluate the recommendation system are as follows: -

Predictive Accuracy Metrics: This metric measures how close the recommendation system is able to measure the estimated ratings with genuine user ratings. Techniques that can be used for it are mean absolute error, mean squared error, root mean squared error, etc.

Classification Accuracy Metrics: It is used to measure how accurately the recommendation system is able to make a successful decision. They have to check whether the item identified is relevant or irreverent. Based on this, the model calculates the number of right and wrong classifications.

Rank Accuracy Metrics: In this metric, the recommendation system is assessed by taking in evaluation of the recommendation's system ability to predict the items based on the right sequence of user's preference. It is also called rank correlation measurement.

4. Discussion

The study on the previous papers shows that the model building for the tourist recommendation system is expensive, time consuming and the model suffers from the cold start problem. The model used for tourism recommendation in [1] uses the traditional techniques and thus suffers the limitation of these algorithms like user and item cold start problems and it does not provide broader suggestions but the user interface developed is interactive, responsive and user adaptable. Issues of Distributed systems were considered while developing it. The proposed solution of [2] utilizes user preferences and broader suggestions and perspective to the users to recommend nearby hotels and attractions. This Application is trying to utilize the climate of a location that is the geographical conditions and the time of day for recommendation in upcoming days. It also lacks in tackling issues like location positioning technologies, Query processing, social media upload. Since [3] uses the hybrid model, accurate predictions are given as it solves the limitations of individually processed systems.

Similarly, [4] uses hybrid recommendation and solves the cold start problems in online movie systems and generates correct movie recommendations to new users with efficient accuracy but this method faces the limitation of longer processing and execution time than existing solutions.

[5] further contributes to the recommendation field by using a model that identifies the requirements of the tourists and indicates the resources best suited for them even if their

behavior changes in the future. It works in a dynamic way. The solution provided by [7] solves two key problems by considering the customers' preferences from multiple dimensions and by combining it with the offline book transaction. The limitations of the solution are it takes a qualitative step to consider the weights to calculate weighted similarity between the customers and this needs to be improved further and it uses the mean average to set the weights to adjust the predictions whose rationality needs to be further improved.

5. Conclusion

Current state and Future state: In the earlier stages of the recommendation systems, it was a necessary requirement to find the similarities between people and the products. But as the requirements grew, progress was made with the recommender systems. To calculate similarity of latent attributes, new methods have been proposed. One of the techniques is matrix factorization. The correlation between item and customer is represented in the form of a matrix by combining all of the attributes for an item and a customer in a way which indicates relationships that were not visible before. The common application of the matrix factorization is that it is used in the Movie recommendation system to determine movie genres even when the genres were not entered explicitly. This method can analyze the movie attributes like cast of the movie, movie title, movie crew, duration, and many others. By analyzing them, a new "genre" was given as output by the system. The notable point is that the solution does not give a new name for the "new genre". Similarly, for the tourism recommender system, the latitudes and longitudes are used to recommend places for the users. These attributes act as the coordinates of a place and the recommendation uses the distance measure to calculate similarities between places. Different distance measures result in different recommendations. Apart from them, other attributes like geographical conditions, amount are also considered by some recommender systems.

Recommendation systems are a very powerful weapon in the hands of a company. Therefore, even if they suffer from the mentioned limitations now, the future developments are definitely going to increase their business value even more

References

[1] Swamy, Pravinkumar, Sandeep Tiwari, and Kunal Pawar. "Tourist Place Recommendation System"

[2] Manjare, P. A., M. P. Vninawe, M. M. Dabhire, M. R. Bonde, and M. D. Charhate. "Recommendation System Based on Tourist Attraction." Int. Res. J. Eng. Technol. 3 (2016): 877-881

[3] Kbaier, Mohamed Elyes Ben Haj, Hela Masri, and Saoussen Krichen. "A personalized hybrid tourism recommender system." In 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), pp. 244-250. IEEE, 2017.

[4] Vahidi Farashah, Mohammadsadegh, Akbar Etebarian, Reza Azmi, and Reza Ebrahimzadeh Dastjerdi. "A hybrid recommender system based-on link prediction for movie baskets analysis." Journal of Big Data 8 (2021): 1-24.

[5] Ganjare, Riteshwari, Riya Sahu, Pratidnya Kharate, Vaishnavi Lohakare, and Gomati Sharnagat. "Hybrid Recommendation System for Tourism based Social Network, and AI."

[6] Forouzandeh, Saman, Mehrdad Rostami, and Kamal Berahmand. "A hybrid method for recommendation systems based on tourism with an evolutionary algorithm and topsis model." Fuzzy Information and Engineering 14, no. 1 (2022): 26-50.

[7] Li, Xixi, Jiahao Xing, Haihui Wang, Lingfang Zheng, Suling Jia, and Qiang Wang. "A hybrid recommendation method based on feature for offline book personalization." arXiv preprint arXiv:1804.11335 (2018).

[8] Yadav, Sambhav, and Sushama Nagpal. "An improved collaborative filtering based recommender system using bat algorithm." Procedia computer science 132 (2018): 1795-1803.

[9] Son, Le Hoang. "HU-FCF: a hybrid user-based fuzzy collaborative filtering method in recommender systems." Expert Systems with Applications: An International Journal 41, no. 15 (2014): 6861-6870.

[10] Ha, Taehyun, and Sangwon Lee. "Item-network-based collaborative filtering: A personalized recommendation method based on a user's item network." Information Processing & Management 53, no. 5 (2017): 1171-1184.

[11] Najafabadi, Maryam Khanian, Azlinah Hj Mohamed, and Mohd Naz'ri Mahrin. "A survey on data mining techniques in recommender systems." Soft Computing 23 (2019): 627-654.

[12] Silveira, Thiago, Min Zhang, Xiao Lin, Yiqun Liu, and Shaoping Ma. "How good your recommender system is? A survey on evaluations in recommendation." International Journal of Machine Learning and Cybernetics 10 (2019): 813-831.