2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 978-1-7281-8595-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCNT51525.2021.9579907

Travel Recommendation System Using Content and Collaborative Filtering - A Hybrid Approach

Vyshnavi Garipelly
Department of Computer Science and Engineering
SRM University, Amaravati
Email: garipelly_vyshnavi@srmap.edu.in

Padma Teja Adusumalli
Department of Computer Science and Engineering
SRM University, Amaravati
Email: padma_teja@srmap.edu.in

Priyanka Singh
Department of Computer Science and Engineering
SRM University, Amaravati
Email:priyanka.s@srmap.edu.in

***Abstract*—Tourism plays a major role in a country's economy. But there is still a lack of a platform that provides personalized information regarding tourist attractions. If there exists a system that can provide personalized accurate information to tourists about local attractions, food, and shopping it will be a huge benefit for tourists. In this paper, we are proposing a hybrid approach of recommended systems to recommend tourist attractions for users. This recommendation process involves a combination of both content and collaborative filtering approach. This Hybrid approach avoids the disadvantages in both the methods and provides users with accurate information. To calculate the similarity between items the cosine similarity method is adopted. We have applied a model-based collaborative filtering approach called SVD for better results. The weighted hybridization approach is used to combine the results of both methods. The data of tourist attractions and users have been collected for implementation. This approach has given better results compared to CB and CF filtering methods separately.**

Keywords: Recommendation, Hybrid recommend-er content filtering, Collaborative filtering, Tourism.

1. Introduction

Nowadays, many tourists depend on online services to plan a holiday trip. Tourists might find the information on many websites, newspapers, and blogs. Due to information overload on the internet finding travel destinations is a very challenging and time taking task. Users ended up taking some trip which is not based on preferences because most of the websites or agencies provide information without considering the user interests. This makes users fail to visit some of his/her favorite places. The tourism sector in India is the most important in the country's economy and is growing rapidly.

The World Travel and Tourism Council calculated that tourism generated 16.91 lakh crore(US$240 billion) or 9.2% of India's GDP in 2018 and supported 42.673 million jobs, 8.1% of its total employment.

We are proposing a system where users get recommendations of tourist attractions based on user interests. We are applying a Hybrid recommendation approach in machine learning to solve the problem of non-personalized recommendation. This method which consists of both content and collaborative filtering considers user preferences and also similarities between items. The weighted approach of hybridization combines the individual scores of both methods to predict the score. This technique overcomes the problem of cold start in collaborative filtering and the limited capability of content-based filtering in considering preferences. This makes recommendations more easy and accurate.

1.1. Related Work

A lot of research work has been done in this field of tourism to provide accurate recommendations to tourists. All the methods proposed so far in this domain have their own disadvantages. Most of these recommendations are based on user reviews. The main advantage of this method will be more personalization in the result list of recommendations. But as we know there are more people who don't rate products even if they like the items in this case the system fails to recommend the best items to users. Users might miss visiting the best places. Other research work involves using collaborative filtering methods to recommend tourist attractions to users. This helps to recommend places to users based on the user's past ratings or similar user's past preferences. In this method, more preference is given to user preferences. The main disadvantage of this method is that when a new item is added into the system that item cannot be recommended due to lack of rating. More items and few ratings will be an issue in this technique.

There is an existing recommendation system for Tourist spots using sentimental analysis based on CNN-LSTM. This is beneficial because the results are accurate because it considers time, season and weather. But it fails to give a Personalised list of recommendations as the System did not have user preferences.

One of the research works involved building personalized tourist recommendation systems has used the combination of more than 2-3 filtering methods that are Hybridization techniques. It gives a more accurate prediction than the other methods used separately. It provides more functions to evaluate the exact information for an active tourist. It can offer top-n-list recommendation services, refine the tourist places in a personalized way, and provide low-budget tour planning by genetic algorithm. but in this method, there is a challenge in the combination of map applications on the mobile app. The efficiency and effectiveness of the recommendation mechanism still need to be refined.

Few tourist recommendation systems using social media analysis to provide a more personalized list of places. It is a model with 58% prediction accuracy by considering social platforms like Twitter tweets. If In case the user is not available on social media platform So, this method is not efficient in recommending places. There is another method that is based on geo-tagged web photos, This recommendation gives more preference to users. But here Users had to provide the photos. If in case the user is not interested to take a photo then this recommendation system may fail and It also involves a lot of manual work.

Some researchers proposed a method of using smartphones and IOT environments. It provides an optimal route and considering behavior data to recommend user interests is an advantage. It is efficient in recommending travel routes. But designing travel sequences pre-processing is complex.

There exists many proposals in this field.some of them considered only one recommendation method.Some of them are based on google maps location which is not accurate and comfortable because it is not taking user interests into account.Some of them are complex to implement.Our method is simple and considered all user needs and interests before recommendation.

2. Methodology

Our paper aims to develop a recommend-er system that suggests users the most interesting tourist spots based on their interest. It is done based on the user profile, item profile, and previous ratings given to tourist places. We are proposing a hybrid recommend-er system technique that generates accurate recommendation lists by integrating content-based and collaborative filtering scores. This technique overcomes the drawbacks of each technique and uses its advantages for better recommendations.

Fig 1, explains the architecture of the proposed system. The architecture consists of input data, recommend-er techniques, and output. The system takes item data and user data as input data. The architecture follows splitting data and then training the data with recommend-er. For evaluation purposes, we use test data. As a final step, we apply a hybrid method to test data. Individual output values of both recommend-er are combined to generate a final ranked list of recommendations for the user as an output. The final recommendation list will be sorted based on scores of hybrid recommend-er.

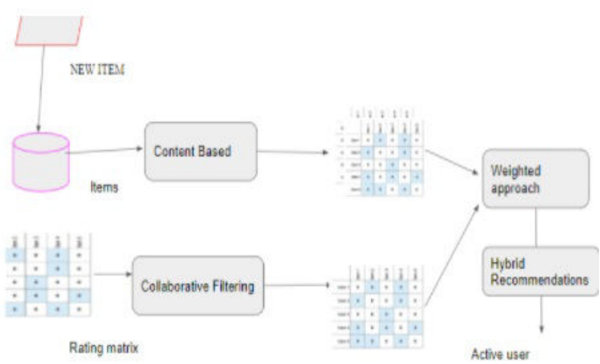

Figure 1. The architecture of the proposed system.

2.1. Dataset

To build recommend-er systems we need different data sets. We collected data from websites using a web scraping method. One of the data sets contains user data(user ID, item ID, ratings) that is user-given ratings for items and another dataset contains tourist attractions information. These IDs are used to extract metadata We split our dataset into training and test sets. a cross-validation approach is used and 20% of data is kept for the purpose of test data. To train recommend-ers and hybrid we used only training data. Data pre-processing and cleaning is done to use it for recommend er systems.

2.2. Collaborative filtering

This method recommends items to users which have never been rated by that user before. The recommendation will be done based on user interests and neighborhood similar user's interests. When we use collaborative filtering there will be diversity in the recommendation. A more effective recommendation process is because we consider users instead of items. Instead of memory-based collaborative filtering, we used the famous latent factor model Single Value Decomposition(SVD). It reduces the number of features of the dataset by reducing the space dimension. It handles the sparsity of matrices very well compared to memory-based ones and is helpful to achieve better results. The cold start problem is the disadvantage of collaborative filtering. When a new item is added due to lack of rating the item cannot predict ratings for users.

2.3. Content filtering

Content-based systems recommend relevant items to the items the user liked or reacted to before. In this method when a user arrives the system will create a user profile. It has information about users and a list of items. Based on this relevant items will be extracted based on user profiles to item profiles. Here we used TF-IDF. This converts unstructured text into vector structure. As all items will be represented in the same Vector Space Model it is to compute the cosine similarity between items. The main drawback is in this recommendation system there will be no diversity in recommended items. This method is not good while working with complex data.

$$similarity(A,B) = \frac{A*B}{||A||*||B||} = i = \sum_{i=1}^{n} \frac{A_i * B_i}{\sqrt{\sum_{i=1}^{N} A_i^2 * B_i^2}}$$

Where:

$$A_i = DimensionalVector$$

$$B_i = DimensionalVector$$

$$A_i.B_i = DotproductofVector$$

$$Formulatocalculatecosinesimilarity$$

2.4. Hybrid Recommendation System

The methods that we discussed before are suffering from one or more disadvantages. The main disadvantages are the cold start problem and sparsity problem. Cold start arises when no rating is available to the item to recommend for the user. Sparsity arises when there are few ratings available for a large number of items. To avoid those shortcomings and to combine the best features of both methods we used a technique called Hybrid recommender. This combines two or more recommendation systems for better results and to produce output. The recommendation accuracy is higher in hybrid recommendation systems compared to collaborative and content-based systems. This is because there is a lack of domain dependencies in collaborative filtering and in content-based filtering there is a lack of user preferences. The combination of both contributes to better recommendations. The advantages of both systems make it a better method for the recommendation. This extends collaborative filtering with content about items and content-based filtering with preferences and rating data.

2.5. Final Step

After testing the hybrid on test data we can check the accuracy and quality of this model.

2.6. Algorithm

Step 1 : Import libraries and data go through data cleaning.

Step 2 : We use the data to train the recommenders and hybrid recommenders.

Step 3 : We train the recommenders. We call the methods with a list of items rated by the user for generating predictions. For content based we use similarity to predict ratings.

Step 4 : Hybrid recommenders use a weighted average approach to combine content and collaborative filtering to generate results. compare the results and based on hybrid recommender scores we provide top N recommendations.

Step 5 : Checking the accuracy of the model for recall to know how many predictions are true and close to actual data. Recall metric is used to check whether the interacted item is in the Top N list of recommendations.

3. Experimental Results

In this section, we will describe the implementation part of our hybrid tourism recommendation system and some results we got from it. We have used web scraping techniques to extract the data from websites for implementation. We chose the Indian state Rajasthan which is one of the famous tourist attractions in India. For the Collaborative filtering approach in order to find cosine similarity, we have to convert matrix values into vectors. Our dataset contains 5 categories of attractions where the user has to choose one by giving input.

TABLE 1. RATING DATASET

Index	User-Id	Item-Id	Rating	timestamp
0	1	1	NaN	881250949
1	1	2	3.0	891717742
2	1	3	1.0	878887116
3	1	4	NaN	880606923
4	1	5	1.0	886397596

For evaluation purposes, we chose to mean absolute error (MAE) and root mean Squared Error (RMSE) values. We considered a 5-fold cross-validation approach Where one part is considered as test data and the remaining as training data. MAE is calculated by using

$$MAE = \sum_{i=1}^{n} \frac{|y_i - x_i|}{n}$$

where:

$$MAE = MeanAbsoluteError$$

$$y_i = Prediction$$

$$x_i = TrueValue$$

$$n = TotalNumberOfDataPoints$$

The lower the MAE the more accurate the output result We also used RMSE for evaluation purposes as it is one of the most common evaluation metrics used for recommendation systems.

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(Predicted_i - Actual_i)^2}{N}}$$

where:

$$RMSE = RootMeanSquareError$$

$$N = Numberofnon - missingdatapoints$$

We computed MAE and RMSE values for collaborative filtering, content-based filtering, and hybrid filtering approaches. To implement a hybrid recommender the weighted approach is used where we have taken equal weights of each method. Mean MAE and RMSE Values for the CB approach are 1.2 and 1.0 respectively. MAE and RMSE values for content filtering are 1.4 and 1.32, by comparing these results with the hybrid part. It has achieved 0.98 and 0.89 scores. By evaluating this we can say the hybrid system provides us a lower value of RMSE and MAE.

However, we have also calculated Recall at 5 and 10 recommendations for each method to check which method gives the best result. The recall is to find the relevant items found in the top recommendation list.

TABLE 2. RECALL RESULTS

Type	Recall@5	Recall@10
CB	16%	52%
CF	52%	46%
Hybrid	47%	67%

It shows how hybrid has given better results compared to CB and CF. It has given a better performance than CB and CF

4. Conclusion

Recommendations according to tourist preferences are much needed in today's world. Although there exist few platforms which can recommend tourist's best route and budget-friendly trips. Most of them are not accurate enough to consider all user interests. Our Hybrid approach has got more accuracy compared to the individual recommend-er systems approach. Using a Hybrid approach in recommendation systems for suggesting tourist spots helps people to make their trip fun and easy.Our system has not achieved expected accuracy. Improving accuracy rate and adding more data from traveller website can be seen as future work.

References

[1] Logesh. Ravi and Subramaniyaswamy Vairavasundaram, *A Collaborative Location Based Travel Recommendation System through Enhanced Rating Prediction for the Group of Users.* Vol-2016,Article I1291358, 28 pages, 2016. https://doi.org/10.1155/2016/129135

[2] Rula A. Hamid, A.S. Albahri, Jwan K. Alwan, *How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management.*Volume 39,2021,100337, https://doi.org/10.1016/j.cosrev.2020.100337.

[3] P. Yochum, L. Chang, T. Gu and M. Zhu, *"Linked Open Data in Location-Based Recommendation System on Tourism Domain".* vol. 8, pp. 16409-16439, 2020, doi:10.1109/ACCESS.2020.2967120.

[4] Zahra Abbasi-Moud, Hamed Vahdat-Nejad,*Tourism recommendation system based on semantic clustering and sentiment analysis.* Volume 167,2021,114324,ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2020.114324.

[5] C.-S. Jeong, K.-H. Ryu, J.-Y. Lee, *"Deep Learning-based Tourism Recommendation System using Social Network Analysis," International Journal of Internet, Broadcasting and Communication* , vol. 12, no. 2, pp. 113–119, May 2020.

[6] Wang M (2020), *Applying Internet information technology combined with deep learning to tourism collaborative recommendation system.* PLoS ONE 15(12): e0240656. https://doi.org/10.1371/journal.pone.0240656

[7] Shini, A. Sreekumar, *"An extensive study on the evolution of context-aware personalized travel recommender systems."* Information Processing Management 57, no. 1 (2020): 102078.

[8] Chenzhong , Tianlong Gu, Yanpeng Sun, Liang Chang*"A Travel Route Recommendation System Based on Smart Phones and IoT Environment", Wireless Communications and Mobile Computing.* vol. 2019, Article ID 7038259, 16 pages, 2019. https://doi.org/10.1155/2019/7038259

[9] Zihao Pu, Hongyu and Duanyu Feng. 2020. *Improved Tourism Recommendation System. In Proceedings of the 2020 12th International Conference on Machine Learning and Computing (ICMLC 2020). Association for Computing Machinery* New York, NY, USA, 121–126. DOI:https://doi.org/10.1145/3383972.3384074

[10] Dhaware M., Vanwari P. (2020) *A Tourism and Travel Recommendation System Based on User-Location Vector.* vol 601, https://doi.org/10.1007/978-981-15-1420-3152

[11] Ni Wayan Priscila Yuni.*"Literature Review Recommendation System Using Hybrid Method by Utilizing Social Media as Marketing."* Computer Engineering and Applications Journal 10, no. 2 (2021): 105-113.

[12] W. Yin, Y. Sun and J. Zhao, *"Personalized Tourism Route Recommendation System Based on Dynamic Clustering of User Groups," 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, 2021, pp. 1148-1151, doi: 10.1109/IPEC51340.2021.9421158.

[13] K. Fararni, F. Nafis, B. Aghoutane and A. Sabri, *"Hybrid recommender system for tourism based on big data and AI: A conceptual framework," in Big Data Mining and Analytics.* vol. 4, no. 1, pp. 47-55, March 2021, doi: 10.26599/BDMA.2020.9020015.

[14] K. Li and C. Qu .*"Design and Implementation of Tourism Route Recommendation System Based on LBS," 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC),* 2021, pp. 2748-2751, doi: 10.1109/IAEAC50856.2021.9391036.