Integrating Sentiment Analysis and User Descriptors with Ratings in Sightseer Recommender System

Vaidehi Anil Chaudhari
Research Scholar
Department of Computer Science and Engineering
Government College of Engineering Aurangabad
Aurangabad,India
chaudharivaidehi77@gmail.com

Vivek Kshirsagar
Associate Professor
Department of Computer Science and Engineering
Government College of Engineering Aurangabad
Aurangabad,India
vkshirsagar@gmail.com

Meghana Nagori
Associate Professor
Department of Computer Science and Engineering
Government College of Engineering Aurangabad
Aurangabad,India
kshirsagarrmeghana@gmail.com

***Abstract*— Since last two decades there is a rapid growth in the globalized world for recommender systems. These provide users rich insights in diverse applications such as healthcare, e-commerce, education, and tourism etc. Hence there is a growing demand to accurately analyze the reviews posted by the users on different social media sites. Tourism industry's economy largely relies on analysis of the above mentioned data. Hence we have pursued the idea of building a recommender system which will provide users with valuable insights and help them in making the correct choice. In our approach we have experimented on the data collected for 150 locations from and near Pune city, in Maharashtra representing the country India. We first categorized the reviews into location specific details. The defined categories are "Temple", "Historical", "Hill Station" and "Educational". We then integrated the ratings provided by the previous users under each category with those of user's interests like Expense, total number of days, distance for trip and the user's interests. The combined approach will be capable of recommending a set of tours that most closely matches with the user's interests and thus enabling them to make the best choice.**

Keywords—Content based recommendation, Collaborative based recommenders, user profiles, user ratings, reviews, User attributes , Sentimental analysis.

I. INTRODUCTION

In our day to day life, whenever we need to search something we go for search engine but there is lots of irrelevant data so we need some systems which can provide us the relevant data. This gave rise to the evolution of recommendation systems. These systems make recommendations as per the user's interest. Recommendation System largely assists the decision making process in Ecommerce applications like Netflix which is used for making video recommendation, Amazon gives suggestions based on the users recent purchases and browsing habits, Linked-in collects similar tastes of the users whereas Pandora is used to recommend music, Tinder is popular date matching site, YouTube is popular for video recommendations. Touristic routes fall in the category of recommendation systems, which are actively researched and developed in the domain of Geographic Information Systems [1].Tour Recommendations are very complicated as each user has its own constraints be time, cost, weather, location etc. Hence this necessitates the need for a personalized recommendation system based on user's interest. As the RS field evolved, researchers studied the use of algorithms from machine learning (ML), an area of artificial intelligence (AI). Machine learning has been studied since the late 1950s, with the emergence of the field of AI.[2].Machine Learning (ML) uses computers to simulate human learning and allows computers to identify and acquire knowledge from the real world, and improve performance on some tasks based on this new knowledge[3].Today there are many Machine learning algorithm such as k-nearest neighbor, clustering, Bayesian, decision trees, matrix factorization etc. Learning is the process of knowledge acquisition. Humans naturally learn from experience because of their ability to reason. In contrast, computers do not learn by reasoning, but learn with algorithms. Today's machine learning algorithms are either supervised, unsupervised, semi- supervised, or based on reinforcement learning [3].

II. RECOMMENDATION SYSTEM

The first Recommendation System was introduced by Goldberg Nicols, and Oki & Terry in 1992[4].In this system collaborative filtering was used. After this many RS followed suit to overcome the limitations of existing systems. The function of a recommender system is to make predictions, suggestions, and opinions according to the user's configured data or any other necessary criteria.[5].A recommender system, is a system performing information filtering to convey information about items that can be movies, music, books, news, images, web pages etc. The aim of a recommender system is often to "help users learn about new products and desirable ones among myriad of choices".[6,7].Tourism is a leisure activity that involves complex decision making processes, for example, selecting attractions, activities, and services. Thus, Tour Recommendation System [TRS] has attracted the attention of many researchers from the fields of both academia and industry. Various TRS have been developed/deployed in and on many kinds of platforms (e.g. desktop, browser, mobile). TRSs recommend results to a user by estimating user interest, choosing Points of Interests (POIs), identifying services or routes, ranking them in sequence, or as a holistic trip plan [8].

Recommendation system can be categorized into different categories based on their design approach. The following section discusses the types of recommendation

systems. There are mainly three types of filtering used in the recommendation system which are Content based filtering, Collaborative filtering and hybrid filtering.

A. *Content Based Filtering:*

Content based filtering system, predicts the similar items of the user interest according to the previous liked items. Content-based filtering suggests activities, events, or services to a user by matching the user's interests with the information about these things. Travelocity.com [9], for example, deploys the content-based approach, in which users specify his or her requirements or interests to the system. The system will match the users' interests with the products available from the destinations. The filtering process applies the algorithms of either information retrieval or item attribute-based filtering [10].

Content-based recommendation systems do have some limitations. The user can only get recommendations based on his or her own preferences, without taking into account other users ratings and feedback [11].

B. *Collaborative Filtering:*

In Collaborative Filtering system, it predicts the taste of the user with respect to its previous ratings for an item as well as according to other users with similar taste. If two users have same rating for an item and if one of them likes a different item then it is automatically recommends the same item to the other user. It recommends considering the ratings of the user for an item. This approach is further divided as User-based approach and Item-based approach.
There are two types of collaborative filtering

1. *User based*: In this type of filtering, the recommendation system depends on similar user interests.
2. *Item based*: In this type of filtering, the recommendation system depends on similar items rated.

C. *Hybrid Approach:*

In this method, both Content based and Collaborative filtering is combined for better accuracy and performance.

III. SENTIMENTAL ANALYSIS:

Sentimental Analysis is also known as opinion mining. Sentiment is a natural ability of human being and predicting the nature of sentiment is known as sentimental analysis. Sentimental analysis is a computational study of attitudes, moods, emotions and sentiments etc. which are in the form of text. Suppose an example, the expression "I had a great day" is a simple positive sentence. The word great tells the attitude of the user from the whole sentence. The main aim of sentimental analysis is to take a decision for information which is present on the web in structured as well as unstructured data. Each user has his/her own different language in which they can express their views for a place.

In our system the sentimental analysis is done on the previously given reviews by any users for our selected place then the reviews are analyzed based on a dictionary of negative and positive words. The word count of the positive and negative words helps in determining the polarity of the reviews. These are integrated with ratings about the place under consideration.

IV. BACKGROUND

A. *Literature Survey*

Table 1. Existing Systems and Research

SR.NO	EXISTING SYSTEMS	DESCRIPTION	PUBLICATION	YEAR
1	GoTour	It is a mobile tourism application for android OS which was designed for Istanbul city.	Al-Rayes et al. [14]	2011
2	MyTourGuide	It was designed for travellers. It was used for recommending the places depending on the users current location and users preferences.	Husain et al. [15]	2012
3	iTravel	It is mobile tourism system in which collaborative filtering was applied which depends on the people who visit the same places can have similar tastes and needs.	WanShiou Yang, San-YihHwang [16]	2013
4	An Innovative Tour Recommendation System for Tourists in Japan.	This is a mobile prototype based on application which is based on the point of interests which takes four parameters from users like about their likes about weather, travel time, expenses which uses the heuristic greedy search.	Quang Thai LE, Devar PISHVA[17]	2015

5	A Content-Based Recommendation System using TrueSkill	In this system they used the content based approach however they overcome the problem of cold start problem which mainly occurs in the content based filtering which lead there system not to depend on the number of users and items.	Laura Cruz Quispe, Jos´e Eduardo Ochoa Luna.[18]	2015
6	Recommendation System for Big Data Applications Based on Set Similarity of User Preferences	This system works on a big data but rather than using data mining it focuses on the set of similarity join which provides the customized and personalized recommendation which is designed in hadoop with the MapReduce framework.	Arpan V Dev, Anuraj Mohan.[19]	2016
7	Personalized Attraction Recommendation System for Tourists Through Check-In Data	As now days the usage of social media like twitter, facebook where check ins are taken from user which can be used for recommending places therefore this system focuses on the social media domain like facebook check ins to recommend places.	K. Kesorn, W. Juraphanthong, and a. Salaiwarakul	2017
8	Hybrid Tourism Recommendation System Based on Functionality/Accessibility Levels	This system is based on user's point of interests where the user's physical and pysological functionality can help in recommending the tourist places.	Tarmo Robal , Yue Zhao, Christoph Lofi, Claudia Hauff.	2018

B. Publication of Recommendation System

In this section, previously study was done on publications of the papers on various domains. There are many of domains some of them are represented in graph as below given. From the above study, most of the research on recommendation system has been carried out in the movie domain due to its availability whereas other areas like books, web-pages, social analysis and tourism are less explored.

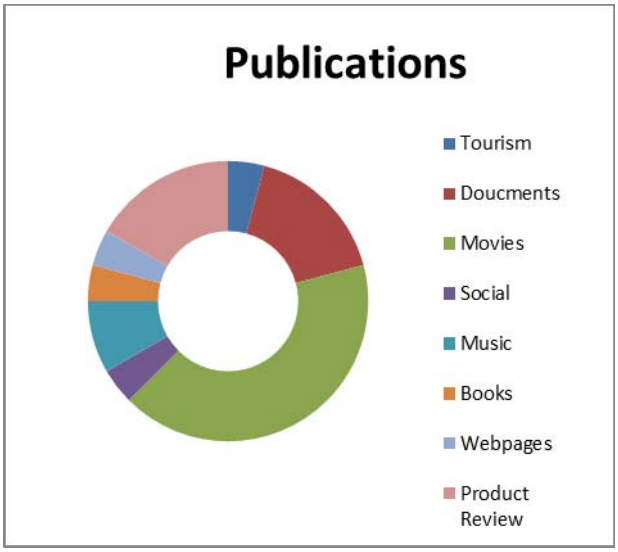

Figure 1 :Applications of Recommender systems in diverse domains

V. DATASET DETAILS:

Dataset can be collected in two ways as follows:

A. Explicit Data:

Explicit data can be the ranking and ratings for items. This can be selecting an item from all the items or it can also be choosing between two items.

B. Implicit Data:

Implicit data can be collecting the data from the history of user like viewing history of user, the purchased item, browsed item as well as social media. Our system collects the explicit data. Our system focuses on 150 places near from Pune city, and the data obtained for our system was collected from four websites named as Holidify.com ,Makemytrip.com,TripAdvisor.com and Trawell.com etc. sites websites provides the interactive database on the places near Pune to visit, along with the dataset of ratings of the users, etc. Our system interprets the star ratings provided by each user as for stars .Usually the 5 star ratings represents excellent which means user has extremely liked the place . While the 4 represents very good,3 represents average, 2 represents poor and 1 is represented as terrible which means user had not liked the place at all. The Makemytrip.com website provides user the most preferable packages and reviews. The TripAdvisor.com provides the best package combinations to the user's and the cost required for the defined locations. All these data has been collected by using the web scraping method which uses the extraction of useful data from different websites. We collected different parameters from different websites for more accuracy in our system.

Previous feedback for the place given by the travellers as well as it consists of the field for records of newly adding reviews by our system. The figure below illustrates a sample of words occurring in the reviews for historical place. Whenever the similar words are repeated then they appear in larger size. These are the reviews given by users in their own language.

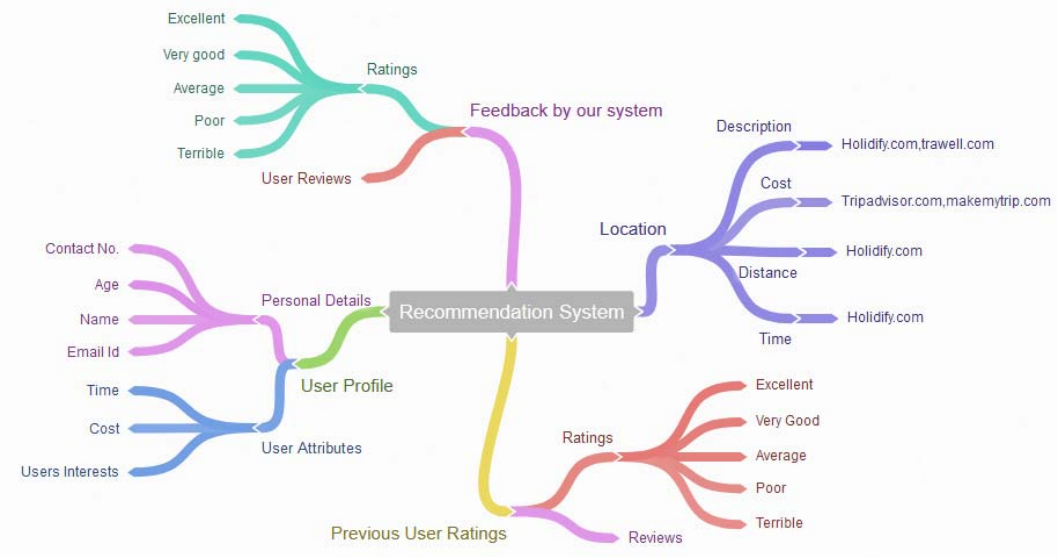

Figure 2: Visualization of Our System

Figure 3: Visualization of Our Dataset

.Figure 2 shows the visualization of our entire system.

VI. ARCHITECTURE OF PROPOSED SYSTEM:

This section describes the detail description about the system. Instead of concentrating on the recommendation system which relies on the ratings of the user for making predictions,our approach incorporates additional parameters defined as user's interest.To recommend a place first classification of reviews is carried out. The Users performing sacred rituals or visiting temples we classified under "Temple", for the fun lovers, adventure seekers and passionate about historical places we categorized as "Historical" and academia seekers we defined "Educational".

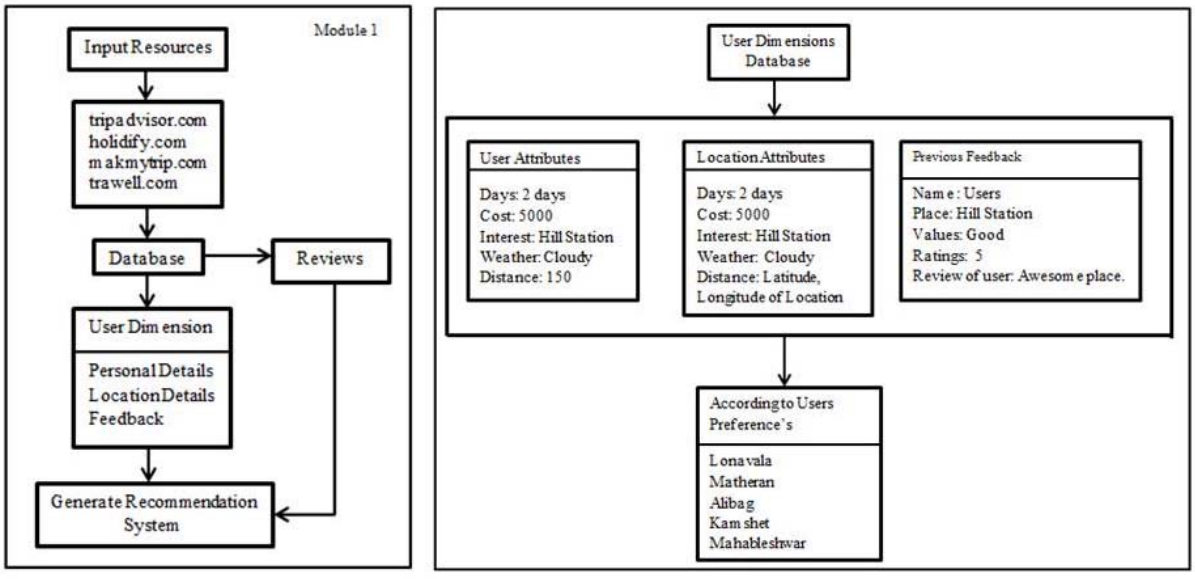

Figure 4:Architecture of Recommendation System

This allows us to work with a subset of datset and addresses the problem of dimensionality.On this reduced dataset sentiment anlaysis is conducted that classifies the tweets as positive and negative.Working on a much smaller dataset now obtained we proceed to recommend a set of tour's based on user's interest.Our TRS systems is thus able to resolve the problems of accurracy which is compromised in an recommnedation system based on ratings.Accurracy is improved by extracting and using the key descriptors from user's reviews as here they try to give actual experience of the place.

A. *Module 1 :* Responsible for creation of the database for the recommendation system from diverse sources.Our system uses content based filtering for recommending user for building the user profiles based on user perferences.

B. *Module 2 :* Feedback is most important aspect in any system,which assists other users in making choices.Our systems second module deals with the feedback of the user.The user experience is recorded. Feedback is also collected in form of ratings as well as reviews. Most of the time people are careless while rating but when it comes to a written review then we get proper description about their experience.

VII. ALGORITHM

The objective of our system is to maximize the number of visited places and minimize the time and budget of the user under user constraints. Figure 5 shows the algorithm we have used in our system for recommendation.

Pseudo Code

Input: Time, Cost, POI, Current location, File of Reviews.
Output: Final Destination.

Step1: Take input from user like user interests, Cost, time and user current location.
Step2: Search for the filtered place after Matching with the user interests with the Description of the place. And our system will rank the distance of places in descending order.
Step 3: If the POI and the user POI matches then it will add the places to the next Location to user for their trip.
Step 4: After filtering the places according to the parameters then the sentimental analysis is applied to the places filtered.
Step 5: After the analysis the polarity is calculated and the places reranked in a descending order and finally the top 5 places are been obtained.

Figure 5:Pseudo code for our System

VIII. RESULTS AND EXPERIMENTS

The table 2 consists of the details of sample reviews fo sentimental analysis,both ratings and reviews are combined to calculate the maximum recommendations by other users given to the place.In the table we have tried to describe the dataset for analys.

Table 2.Details of the Sample data for Sentimental Analysis

Sr no.	Category of Places	No. of Places	Ratings	Reviews:
1	Hill Station	24	116	We have good places to see in Lonavala like loin point, tiger point, lonavala lake, temples, museums . Because it is out of the way, this place has not really been touched by humans very much and all the pollution. The trek to the cave is also, amazing. I would HIGHLY recommend this as a day trek, just bring water.
2	Temple	56	2437	This is famous Ganpati Temple in Pune. This is situated on very narrow and single lane road. Avoid going by your personal vehicle. Rich ganpati temple one should compulsory need to see this temple if u 'll be coming to Pune. Nice place for everyone.
3	Historical	40	1268	Sinhagad is just 1 hr drive from pune railway station .one can also trek .it is a historical fort on top of the mountain .the locals dish out some nice snacks like kanda bhaji & Curd.they can also prepare Gavran chicken curry with bhakri. It is one of the largest forts in India and it has a great history and the architect of the fort is damn good.
4	Educational Institutes	30	165	One of the best nursing institutes in Pune ,here there are good facilities for nursing students as well as nursing staff .it's well maintained. Great college.

Table 3. Results generated considering Distance Parameter

CATEGORY	*RESULT*		
POIs	Short tours:One day	Long time Tours :Three days	Optimal Tours:Two days
Temple,Fort,Hill Sttaion,Educational Institution	Shaniwar Wada, Malhargad Fort, Dagdusheth Halwai Ganpati temple,Katraj Jain temple,Lonavala	Rajgad Fort, Lohgad Fort, Shivneri Fort, Korigad Fort, Pratapgarh Fort,	Sinhagad Fort, ISKCON, Fergusson College, ,Indira college,PICT College, Purandar Fort, Ghangad Fort,
TOP 5	70%	60%	50%
STOPs	3	2	5
Cluster Radius	50km	200km	150km

Table 4. Results generated considering Cost Parameter

CATEGORY	*RESULT*		
POIs	Short tours:One day	Long time Tours :Three days	Optimal Tours:Two days
Temple,Fort,Hill Sttaion,Educational Institution	Shaniwar Wada, Malhargad Fort, Dagdusheth Halwai Ganpati temple,Katraj Jain temple,Lonavala	Rajgad Fort, Lohgad Fort, Shivneri Fort, Korigad Fort, Pratapgarh Fort,	Sinhagad Fort, ISKCON, Fergusson College, ,Indira college,PICT College, Purandar Fort, Ghangad Fort,
TOP 5	80%	60%	50%
STOPs	3	2	5
Cluster Radius	50km	200km	150km

Table 3 & 4 describes about the results which are obtained after the implementation. The table 3 describes about the results obtained with the distance parameter while table 4 describes about the results which are obtained after considering the budget of the user.

IX. CONCLUSION

This paper illustrates the ongoing research in dynamically recommending a trip in India for the Pune city incorporating user's interests. The aim of the system is to combine the ratings and user reviews in the existing database for enhancing the accuracy of the Predictions. TRS thus falls in the category of content based recommendation. As a Future scope to our system we can continue to build a massive dataset by including all places that represents all the 29 states and the 7 union territories. This massive dataset can represent all the prominent places of our country and by employing graph partitioning algorithms we can process the same work efficiently in a distributed framework We can also incorporate in our database the information of the billions of users by the likes and dislikes from popular social media sites like Facebook, instagram, LinkedIn etc.

ACKNOWLEDGEMNENT

We are thankful to our entire faculty in the department for their support and encouraging words. We also thanks to our Principal Sir Dr. P.B.Mrunal for his Motivation and making timely availability of all the resources.

REFERENCES

[1]Yukiko Kawai, Jianwei Zhang, Hiroshi Kawasaki."Tour Recommendation System Based On Web Information And Gis".

[2]Martens, H. H. (1959). Two notes on machine "Learning". Information and Control, 2(4), 364-379.

[3]Ivens Portugal, Paulo Alencar, Donald Cowan.The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review.

[4]"Witten I. H and Frank I .data Mining, Morgan Kaufman Publishers San Francisco,2000".

[5]A. Hanze and S. Junmanee, "Travel recommendations in a mobile tourist information system",In: Proceedings of Information Systems and its Application ISTA, (2005), pp.86-99.

[6]Oh,H., Litterature review on advisor selection.

[7]Resnick, P., Varian, H. R., Recommender Systems, Communications of ACM, Vol. 40, No.3, 1997, pp. 56-58.

[8]Pree Thiengburanathum (PhD student)a, Shuang Cangb, Hongnian Yuc, "A Decision Tree based Recommendation System for Tourists".

[9]K.Kabassi,"Personalizing recommendations for tourists," Telemat. Inform., vol. 27, no. 1, pp. 51–66, 2010.

[10]A. Hanze and S. Junmanee, "Travel recommendations in a mobile tourist information system", In: Proceedings of Information Systems and its Application ISTA, (2005), pp. 86-99.

[11]Sebastia L, Garcia I, Onaindia E, Guzman C. e-Tourism: a tourist recommendation and planning application. In: ICTAI '08, vol. 2; 2008. p. 89–96.

[12]Husain W, Dih LY, Yen HF, Jothi N. Mytourguide.com: a framework of a location- based services for tourism industry. In: 2012 international conference on computer information science (ICCIS), vol. 1; 2012. p. 184–9.

[13]WanShiou Yang, San-Yih Hwang, "iTravel: A recommender system in mobile peer-to-peer environment", The Journal of Systems and Software 86, 12–20, 2013.

[14]Quang Thai LE*, Davar PISHVA. " An Innovative Tour Recommendation System for Tourists in Japan."

[15] Laura Cruz Quispe, Jos´e Eduardo Ochoa Luna "A Content-Based Recommendation System using TrueSkill."

[16]Arpan V Dev,Anuraj Mohan "Recommendation System For Big Data Applications Based On Set Similarity Of User Preferences".

[17]K. Kesorn,W. Juraphanthong, and A. Salaiwarakul "Personalized Attraction Recommendation System for Tourists Through Check-In Data". The PAAMS Collection – 15th International Conference, PAAMS 2017,Advances in Intelligent Systems and Computing 619.

[18] Tarmo Robal,Yue Zhao,Christoph Lofi, Claudia Hauff "Hybrid Tourism Recommendation System Based on Functionality/Accessibility Levels".In IEEE,vol 5, 2169– 3536, 2012.