A Point of Interest Recommendation Engine for Suggesting Sightseeing Spots by User Reviews and Ratings

Suhanth P[1], Sathish Kumar G[2]
Mrs Priskilla Angel Rani J[3]

[1]Student, BE Computer Science Engineering, Anand Institute of Higher Technology, Tamil Nadu, India

[2]Student, BE Computer Science Engineering, Anand Institute of Higher Technology, Tamil Nadu, India

[3]Assistant Professor, ME Computer Science Engineering, Anand Institute of Higher Technology, Tamil Nadu, India

ABSTRACT

Tourism has become an important industry for most of the economies, especially for non-industrialized countries where it represents the main source of income. Recommendation systems are defined as the techniques used to predict the rating one individual will give to an item or social entity. These items can be places, books, movies, restaurants and things on which individuals have different preferences. These preferences are being predicted using two approaches first content-based approach which involves characteristics of an item and second collaborative filtering approaches which takes into account user's past behavior to make choices. Point of interest (POI) recommendation, which provides personalized recommendation of places to users. However, quite different from traditional interest oriented merchandise recommendation, POI recommendation is more complex due to the timing effects: we need to examine whether the POI fits a user's availability. With increasing adoption and presence of Online services, designing novel approaches for efficient and effective recommendation has become of paramount importance. In existing services discovery and recommendation approaches focus on keyword dominant Web service search engines, which possess many limitations such as poor recommendation performance and heavy dependence on correct and complex queries from users. Recent research efforts on Online service recommendation center on two prominent approaches: collaborative filtering and content based recommendation. Unfortunately, both approaches have some drawbacks, which restrict their applicability in Web service recommendation. In proposed system for recommendation we will be using Agglomerative Hierarchal Clustering and collaborative filtering for effective recommendation in this system.

Key Words – *Point of Interest, Keyword Dominant Web Services Search Engine, Collaborative Filtering***.**

1. INTRODUCTION

BIG data has emerged as a widely recognized trend, attracting attentions from government, industry and academia. Generally speaking, Big Data concerns large-volume, complex, growing data sets with multiple, autonomous sources. Big Data applications where data collection has grown tremendously and is beyond the ability of commonly used software tools to capture, manage, and process within a "tolerable elapsed time" is on the rise. The most fundamental challenge for the Big Data applications is to explore the large volumes of data and extract useful information or knowledge for future actions. Spurred by service computing and cloud computing, an increasing number of services are emerging on the Internet. As a result, service relevant data become too big to be effectively processed by traditional approaches. In view of this challenge, a Clustering-based Collaborative Filtering approach (Club CF) is proposed in this paper, which aims at recruiting similar services in the same clusters to recommend services collaboratively.

Recommender systems in location based social networks mainly take advantage of social and geographical influence in making personalized Points-of-interest (POI) recommendations. The social influence is obtained from social network friends or similar users based on matching visit history whilst the geographical influence is obtained from the geographical footprints users' leave when they check-in at different POIs. However, this approach may fall short when a user moves to a new region where they have little or no activity history. We propose a location aware POI recommendation system that models user preferences mainly based on; user reviews and categories of POIs. We evaluate our algorithm on the Yelp dataset and the experimental results show that our algorithm achieves a better accuracy.

Cluster-based recommendation is best thought of as a variant on user-based recommendation. Instead of recommending places to users, places are recommended to clusters of similar users. This entails a pre-processing phase, in which all users are partitioned into clusters. Recommendations are then produced for each cluster, such that the recommended items are most interesting to the largest number of users. The upside of this approach is that recommendation is fast at runtime because almost everything is pre-computed.

2. LITERATURE SURVEY

H. Yin, W. Wang, H. Wang, L. Chen, and X. Zhou [1] have presented an overview of the field of recommender systems and describes the current generation of recommendation methods that are usually classified into the following three main categories: content-based, collaborative, and hybrid recommendation approaches. These extensions include, among others, an improvement of understanding of users and items, incorporation of the contextual information into the recommendation process, support for multi criteria ratings, and a provision of more flexible and less intrusive types of recommendations.

In recent years, a variety of review-based recommender systems have been developed, with the goal of incorporating the valuable information in user-generated textual reviews in to the user modelling and recommending process [2]. They provided a comprehensive overview of how the review elements have been exploited to improve standard content-based recommending, collaborative filtering, and preference-based product ranking techniques. The review-based recommender system's ability to alleviate the well-known rating Sparsity and cold-start problems is emphasized. This survey classifies state-of-the-art studies into two principal branches: review-based user profile building and review-based product profile building.

The availability of user check-in data in large volume from the rapid growing location-based social networks (LBSNs) [3] enables a number of important location-aware services. Point-of-interest (POI) recommendation is one of such services, which is to recommend POIs that users have not visited before. For effective and efficient recommendation, they developed a preference propagation algorithm named Breadth-first Preference Propagation (BPP). The algorithm follows a relaxed breath-first search strategy and returns recommendation results within at most 6 propagation steps.

Making personalized and context-aware suggestions of venues to the users is very crucial in venue recommendation. These suggestions are often based on matching venues' features with users' preferences, which can be collected from previously visited locations. They have present a novel user modelling approach [4] which relies on a set of scoring functions for making personalized suggestions of venues based on venues content and reviews as well as user context. Our experiments, conducted on the dataset of the TREC Contextual Suggestion Track, proved that our methodology out performs state-of-the-art approaches by a significant margin.

With the increasing popularity of location-based services, such as tour guide and location-based social network, we now have accumulated many location data on the Web. In this paper [5], they show that, by using the location data based on GPS and users' comments at various locations, we can discover interesting locations and possible activities that can be performed there for recommendations. Our research is highlighted in the following location-related queries in our daily life.

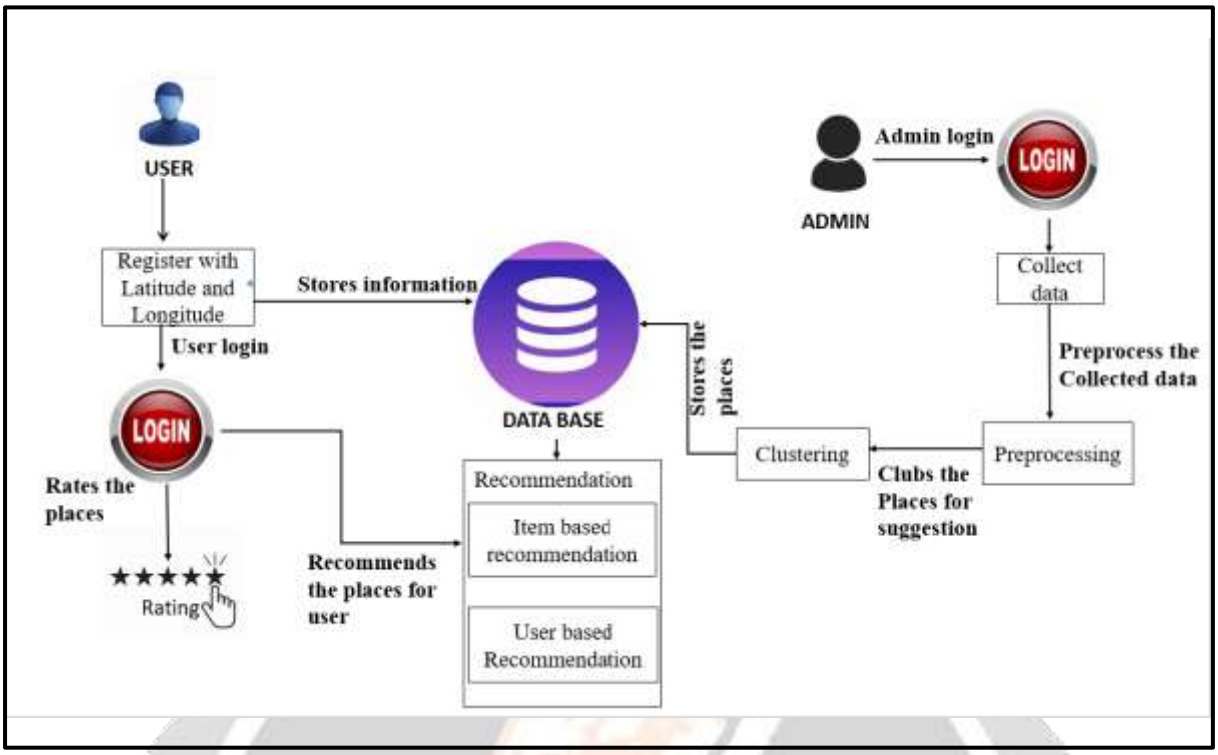

Figure 1: Core Architecture

3. SYSTEM DESIGN

The system architecture is described in Figure I. Initially, the new user has to register to access this application, Once the user is registered, the user id is automatically generated and user can enter the own password which has been registered in register module. In user module, User can able to rate the place which was visited by them and user can also view the recommended places according to the user tastes and recommendation also based on similar user who already had a same taste like them. In User side, two recommendation takes place. After that, in admin module, Admin had a default ID and password to login the admin page. In this page, admin can add the location list which was saved into the database. And then Preprocessing is takes place to clean the data. In order to recommend the accurate places to the user, Classification and clustering algorithm is used. In which algorithm all the features and properties are grouped together based on user taste. And it produces the Item based cluster and user-based cluster module. And finally, it recommends places with user interest and points on the map.

4. MODULES

The modules which we are developing using different strategies to give the accurate recommendation for the user.

- Gathering of data
- Add location Details
- Data Pre-testing
- Data clustering using collaborative filtering
- Recommendation Engine

4.1 GATHERING OF DATA

The dataset used for this study was prepared from the tourist places Repository. Here we having the three data sets are location, users and rating. The location data attributes include country name, city name, place name and the user data including the attributes are user id, age, gender and password. The rating dataset having the attributes are user id, place name and ratings. After collection of data, the dataset was prepared to apply the data mining techniques. Before application of prescribed model, data preprocessing was applied to measure the quality and suitability of data.

4.2 ADD LOCATION DETAILS

In this module admin can add new places with details and their location details. These details will add to the existing details. User can select the place details added in this module and they will rate the place based on their reviews. These details are used for cluster the data based on their ratings.

4.3 DATA PRE-TESTING

The training data, we are given a list of vectors (u; m; r; t), where u is a user ID, m is a places name; r is the rating u gave to m, and t is the date. After training, application output predictions for a list of user-movie pairs. Application measure error by using the root means squared error. After preprocessing application output the places name with the corresponding users and their ratings with separated files.

4.4 DATA CLUSTERING USING COLLABORATIVE FILTERING

Group based suggestion is best idea of as a variation on client-based proposal. Rather than prescribing things to clients, things are prescribed to groups of comparable clients. This involves a pre-handling stage, in which all clients are parceled into bunches Recommendations are then created for each group, with the end goal that the prescribed things are most fascinating to the biggest number of clients. The upside of this approach is that proposal is quick at runtime on the grounds that nearly everything is precompiled. One could contend that the proposals are less individual along these lines, since suggestions are processed for a gathering as opposed to a person. This approach might be more compelling at delivering proposals for new clients, who have little inclination information accessible. In this sense is important to consider that clusters are formed according the similarities and dissimilarities between the analyzed countries. Therefore, tourism professionals and researchers should consider that the identity and consciousness of the people.

4.5 RECOMMENDATION ENGINE

The hierarchical cluster analysis starts with the conglomerate data and considers each one of them as a cluster, so there are as many cases as groups. Subsequently, using an algorithm the SPSS sequentially combines the groups, reducing them until only one group remains. During this process the program calculates the distance between the data points to form differences or groups. The results are reflected in representing the groups and the distance between them. After Hierarchal Agglomerative approach filter we get a recommended data for individual user. Here, Map reduce algorithm is used to group the similar clustered users as a group. After Map-Reduce collect the highest rating of the particular users in the clustered data and that highest rating will send as a recommendation to the particular user.

5. MODULE EVALUATION

5.1 GATHWERING OF DATA

The user registers his details here like age password and location details. The figure 2 shows the registration details.

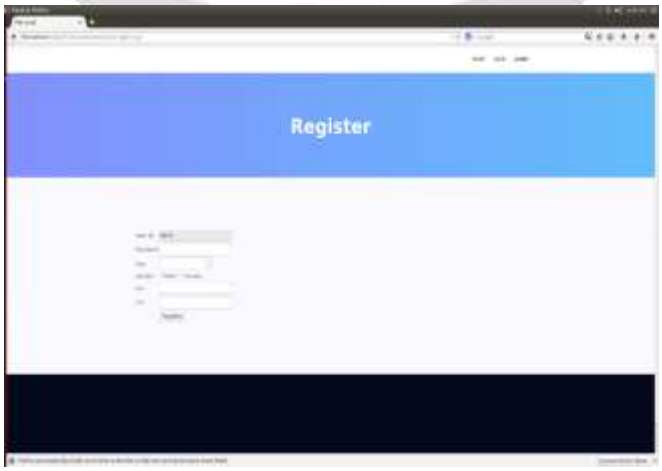

Figure 2: Gathering of data

5.2 ADD LOCATION DETAILS

The both users and admin can login where user can rate the places and admin can add new places. The figure 3 shows the login details.

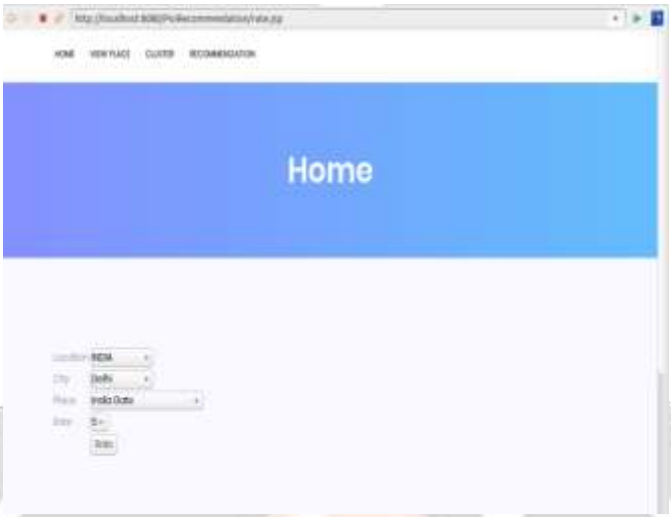

Figure 3: Add location details

5.3 DATA PRE-TESTING

The data is being pre-tested and removes the unwanted data by machine learning. The figure 4 shows the pre-tested details.

Figure 4: Data pre-testing.

5.4 DATA CLUSTERING

Based on the user rating the data are clustered and groups them by using collaborative filtering algorithm. The figure 5 shows the grouped data based on ratings.

Figure 5: Data clustering

5.5 RECOMMENDATION ENGINE

The places are recommended to the user based on his/her ratings by using Agglomerative Hierarchical Clustering. The figure 6 shows the recommended places to the users.

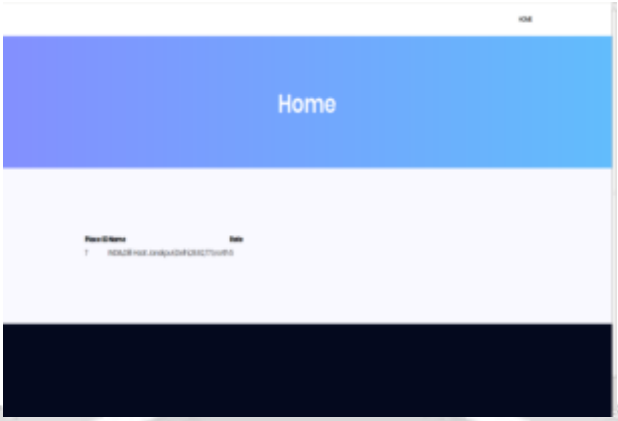

Figure 6: Recommendation engine

5.6 PERFORMANCE

The figure 7 shows the performance of the users and the system.

Figure 7: Performance

6. ALGORITHM USED

We are proposing two algorithms for the best recommendations for the users. These algorithms can give accurate and best solutions for the users.

- Collaborative filtering
- Agglomerative Hierarchical Clustering

6.1 Collaborative Filtering

Collaborative filtering is a technique used by recommender systems. Collaborative filtering has two senses, a narrow one and a more general one.

In the newer, narrower sense, collaborative filtering is a method of making automatic predictions about the interests of a user by collecting preferences or taste information from many users. The underlying assumption of the collaborative filtering approach is that if a person *A* has the same opinion as a person *B* on an issue, A is more likely to have B's opinion on a different issue than that of a randomly chosen person.

In the more general sense, collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. Applications of collaborative filtering typically involve very large data sets.

6.2 Agglomerative Hierarchical Clustering

In data-mining and statistics, hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. Agglomerative: This is a "bottom-up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

In general, the merges and splits are determined in a greedy manner. The results of hierarchical clustering are usually presented. The standard algorithm for hierarchical agglomerative clustering has a time complexity of {\display style {\math Cal {O}} (n^ {3})} and requires {\display style {\math Cal {O}} (n^ {2})} memory, which makes it too slow for even medium data sets. However, for some special cases, optimal efficient agglomerative methods {\display style {\math Cal {O}} (n^ {2})} are known: SLINK for single-linkage and CLINK for complete-linkage clustering. With a heap the runtime of the general case can be reduced to {\display style {\math Cal {O}} (n^ {2} \log n)}at the cost of further increasing the memory requirements. In many programming languages, the memory overheads of this approach are too large to make it practically usable. Except for the special case of single-linkage, none of the algorithms can be guaranteed to find the optimum solution

7. CONCLUSION

We present an Agglomerative Hierarchical clustering approach for big data applications relevant to service recommendation. Before applying CF technique, services are merged into some clusters via an AHC algorithm. Then the rating similarities between services within the same cluster are computed. As the number of services in a cluster is much less than that of in the whole system, Club CF costs less online computation time. Moreover, as the ratings of services in the same cluster are more relevant with each other than with the ones in other clusters. If we want, add the new places with name, city and country.

REFERENCES

[1] H. Yin, W. Wang, H. Wang, L. Chen and X. Zhou, "Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 11, pp. 2537-2551, 1 Nov. 2017.

[2] Chen, Li & Chen, Guanliang & Wang, Feng. (2015). *"Recommender systems based on user reviews: the state of the art. User Modeling and User-Adapted Interaction"*. 25. 10.1007/s11257015-9155-5.

[3] Yuan, Quan & Cong, Gao & Sun, Aixin. (2014). *"Graph-based Point-of-interest Recommendation with Geographical and Temporal Influences"*. CIKM 2014 - Proceedings of the 2014 ACM International Conference on Information and
Knowledge Management. 659-668.
10and groups them .1145/2661829.2661983.

[4] Aliannejadi, Mohammad & Mele, Ida & Crestani, Fabio. (2017). *"Personalized ranking for context-aware venue suggestion"*. 960-962. 10.1145/3019612.3019876.

[5] Zheng, Vincent & Zheng, Yu & Xie, Xing & Yang, Qiang. (2010). *"Collaborative location and activity recommendations with GPS history data"*.
Proceedings of the 19th International Conference on
World Wide Web, WWW '10. 1029-1038.
10.1145/1772690.1772795.