Personalized POI Travel Recommendation from Social Media

J.Cibia¹, D. Vidya Bharathi²

¹M.E., ²Associate Professor, Computer Science and Engineering, Sona College of Technology, Salem.

Abstract—Travel based recommendation and journey planning are challenging tasks because of various interest preferences and trip restrictions such as limitation of time, source and destination points for each tourist. Large amount of data can be collected from the Internet and travel guides, but these resources normally recommend individual Point of Interest (POI) that is considered to be familiar, but they do not provide sufficient information to the interest preference of the users or hold to their trip constraints. In addition, the huge volume of information makes it a challenge for every tourist to pay attention to a potential set of POIs to make a visit in any unknown city. After the tourist discovers an acceptable set of POIs to go to, it’ll take abundant time and energy for him/her to make a brief outline of the suitable duration of the visit at every POI and the order in which to visit the POIs. To sort out these problems, an author topic collaborative filtering (ATCF) algorithm is suggested for personalized tours. This method suggests that the POIs are optimized to the users’ interest preferences and POI popularity. Hence, this method is elaborately explained here for tour recommendation problem based on similar user and similar city prediction, which considers user tags. It extends our method to provide personalized suggestions based on user geo co-ordinates points. In the first instance, the multiple users’ location histories are modeled using tree-based hierarchical graph (TBHG). Based on TBHG, HITS approach is developed in order to gather the interest level of a selected place and a user’s travel expertise (knowledge). Finally, HITS-based collaborative filtering technique is used to obtain GPS based personalized recommendation system. And for image based search similar images with the tag information are retrieved for the query image users have.

Keywords—City prediction, GPS trajectories, Place of interest Travel Recommendation, User preferences.

1. INTRODUCTION

In modern days, the rapid growth of cities has paved way for the development of a huge number of points of interest (POIs), e.g., stores, theatres, restaurants and residence that enliven and entertain the people, providing us with more choices of living experience than before. People routinely explore the city and neighbourhood in their daily life and decide where to go based on their personal interests and the various choices of POIs. At the same time, making an efficient and effective decision among the large number of POI choices becomes an annoying problem for the user. To facilitate the user’s exploration and decision making, POI recommendation has been introduced by location-based services.

However, such recommendation models are commonly based on the preference of most of the users on POIs, which ignore the user’s personal preference. When compared to the places visited that are best suited to the user’s interest, those places visited that are against the user’s interest could bring bitter exploration experience in a situation when the user travels to a new place.

Hence, the personalized POI recommendation is absolutely necessary in order to help the users filter out venues that are uninteresting according to their own taste, reduce the displeasing experience, and save their time in decision making. Collaborative filtering (CF) and Content-based (CB) are the most commonly used methods of personalized travel recommendation, CF, which is otherwise referred to as social filtering, filters information by using the recommendations of other people. The Recommender systems are being usually applied in many fields, like e-commerce, etc., to provide various products, important services and useful information to many customers. Collaborative filtering is considered to be the most successful approach. The contents it recommends to the current customers are mainly based on the past transactions and feedback of the various customers. However, it is not easy to distinguish the similar interests between customers because of the sparseness problem. The sparseness problem is caused by the insufficient number of transactions and feedback data, which limited the usage of the collaborative filtering method. Besides collaborative filtering, content-based filtering is yet another important class of recommenders systems. The content-based recommenders systems make travel recommendations by analysing the content of the textual information and finding the regularities in the content provided. The major difference between the collaborative filtering and the content-based recommenders systems is that the CF method only uses the user-item ratings data to maintain accuracy in predictions and effective recommendations. However, the content-based recommenders systems depend completely upon the features of the users and items for predictions.

Both the content-based recommenders systems and the CF systems have restrictions. While the CF systems do not explicitly incorporate feature information, the content-based systems do not incorporate the necessary and preferred information which is alike across the individuals.
When compared with other online social networks that have the user activities interacting with the virtual world, Location Based Social Networks (LBSNs) reflect the user’s geographical action in the real world, where the online world and the real world intersect; therefore, connecting the gap between the real world and the virtual world. The network provides challenges for researchers to investigate the users’ changing behavior for the point of interest recommendation in spatial, temporal, social and content aspects. The POI recommenders systems have a main role in LBSNs since they can not only meet the users’ personalized preferences for visiting new places, but also help LBSNs to enhance revenues by providing the users with intelligent location services, such as location-aware advertisements which is illustrated in fig 1.

The advanced GPS-enabled devices allow the people to record their location histories. They achieve this with GPS traces, which suggest the behaviour of the people and what all they prefer related to travel. We have performed two different types of travel recommendations by mining GPS traces of multiple users:

1. A generic one that recommends the user with the most interesting locations in a given spatial region.
2. A personalized recommendation that provides the user with suitable locations that matches his travel preferences.

To achieve the first recommendation, we develop multiple users’ location histories using a tree-based hierarchical graph (TBHG). We propose the HITS-based model to satisfy the user’s interest on visiting a particular location. In our personalized recommendation, we first study and understand the correlation between the locations.

Then we convert and incorporate this correlation into a collaborative filtering-based model. This predicts the user’s interests in an unvisited location based on his location histories and that of others.

Next the other type of recommendation is based on images (i.e.) the user can search similar images with the query image. This recommendation will retrieve all similar images with the associated information.

II. RELATED WORK

J. Bao, et.al., [2] learns the preferences of the users from her location history and models the preferred ideas with a weighted category hierarchy (WCH) and further approximately calculating the similarity between the two users’ preferences by calculating the similarity of WCHs between the two users. This method adds to user preference modeling and managing the data sparseness problem for location recommendations.

D. M. Blei, et.al, [3] described latent Dirichlet allocation, a versatile generative probabilistic model for collecting discrete information. LDA is established on an easy exchangeability assumption for the various words and topics in a document. It is so accomplished by a straightforward application of de Finetti’s illustration theorem. LDA is considered as a dimensionality reduction technique within the principle of LSI however with proper basic generative probabilistic semantics that’s logical for the kind of information that it models.

A. Cheng, et.al.[4] focuses on the customized recommendation framework to provide not solely a context-aware recommendation system however also a route planning application before the journey is initiated. The personalization is achieved by adopting specific user profiles with the automatically detected people attributes (e.g., gender, age and race) along with the trips undertaken.

M. Clements,et.al., [5] predicts similar locations based on the users’ geotags in a geographically remote location and view statistical enhancements over all users that visited largest cities and provides an example of efficient recommendation based on an artificial user profile and define a resemblance between the geotag distributions of two users based on a Gaussian kernel convolution. The geotags of most of the similar users are then combined to relocate the popular locations in the destined city personalized for this user.

H. Gao,et.al., [7] systematically studied the content information on LBSNs for POI recommendation and investigated various kinds of content data on LBSNs in terms of sentiment indications, user interests, and POI properties and model them below a unified POI recommendation framework.
As a result, the experiment demonstrated the importance of content data in explaining the user behavior and improvement of POI recommendation performance on LBSNs.

Zheng, et.al., [20] has proposed to deploy the activity correlation and location correlation of the user with regard to the location features to regularize the factorization of a location-activity matrix for location and activity recommendation.

All these approaches target issues associated with the overall plan and optimization of travel guides by relying on the available resources of knowledge. In contrast, this paper addresses a customized POI recommendation task. That is to mention those recommendations are generated by the individual preference of every user.

III. POI RECOMMENDATION USING ATCF

The rapidly growing of Social networks provides an outsized quantity of knowledge that allows the services based on point of interest. In this approach, a study of latest POI recommendation drawback to predict the users’ current cities is to be suggested. The challenge is tough to learn the user’s ordered information and provide personalized recommendation model. So implementation of Author Topic Modelling approach carries the personalized travel recommendation system that is the extended version of LDA technique. This system collects the knowledge of the author and therefore the cities. Through ATM, both the category and the user’s travel preferences are mined by modifying the latent model simultaneously. The ATM chiefly consists of two steps such as probabilistic generative model and Bayesian estimation model. Through ATM, we can determine the probabilities of every word to different topics. We also get author topic matrix for all the users. Then represent each POI as one point in a latent space. We assume that the Euclidean distance between the points of interests in the latent space reflects the transition probability. The larger the gap, the lower the strength of transitions. With all POIs embedded in a latent space, our model estimates the sensible transition probabilities of POIs. It is additionally attainable to assign meaningful possibilities to those unobserved transitions and it is shown in figure 2.

![Figure 2: ATCF framework](image)

Most of the people opt to micro blogging services like Twitter so as to assemble period news or opinion about famous people, popular things or events that interest much. Such services are used for social networking like being in touch with friends, relatives and colleagues. Micro blogging sites publish certain platforms to create and take content from sets of users with overlapping and various interests.

Example: A case study in Twitter
User 1: Charles
User 2: Robin

Charles follows Robin because of the latter’s posts about travel experience. But Robin also uses Twitter for sharing political views and cricket news with friends. Currently Charles has limited tools to filter cricket and political content from Charles. Twitter, however, assumes that Charles wants to read every post shared by other people. Similarly, Charles also has a limited set of options to identify new people to follow. He will be able to view the list of users in the social graph (e.g. those followed by Robin), Otherwise Charles can search by keyword and then browse the returned tweets’ posters. This remains very difficult to find people like Robin.
The textual information to predict user experience may be difficult if proper language is not used.

IV. PROPOSED SYSTEM

A. Search by Location

The latest GPS enabled devices allow the individual to ascertain their location histories with GPS records, which means human behaviour and preferences based on travel. In this paper, two sorts of travel recommendations are given by casting off multiple users’ GPS traces. The first kind recommends the user with prime fascinating locations and travel sequences in an exceedingly given geospatial region. The second may be a personalized recommendation that offers the user with locations matching her/his travel preference.

To model multiple user location history, tree-based hierarchical graph (TBHG) is employed. Tree based hierarchy is constructed by collecting multiple GPS logs and cluster them using “density based clustering” so that similar points will come under same cluster. The algorithm behind this technique is seen below

| Input: Multiple users GPS logs, r (radius), minpt (minimum number of point in cluster) |
| Output: Clusters |
| • Start with a point (GPS log), let it be the centre |
| • Extract neighbourhood of this point (i.e.) points with distance r. |
| • If sufficient neighbourhood, form a cluster or else consider as noise |

The distance between two points is calculated as

Distance (p, c) =sqrt ((px-cx)^2 + (py-cy)^2)

Where c is the centre, p is the point to be compared, px, py is the latitude and longitude of p, cx, cy be the latitude and longitude of centre point.

Based on tree based hierarchical graph, hypertext induced topic (HITS) search model is developed. This is a search-query dependent ranking algorithm for information retrieval and predicts various levels of location and knowledge about travel experiences. When the user enters a probe query, in the first instance, the HITS method lists out the relevant pages returned by a probe engine and then it brings out two sorts of rankings for the enlarged set of pages. Those rankings are called authority ranking and hub ranking.

In the expanded set, HITS assigns them an authority score (location that’s visited by most variety of user is given higher authority score) and a hub score (user who have visited most variety of places is given higher hub score).

There are two links that are discussed— in-links and out-links. An authority is a page with a number of in-links, and a hub is a page with a number of out-links. The main idea of HITS is that it will have good hub points to many good authorities, and a good authority is referred to a number of good hubs. Thus, both authorities and hubs have a mutual reinforcement relationship.

To be very precise, the authority score of a page is the sum of the hub scores of the pages pointed to it and its hub score is the integration of authority scores of the pages pointed to it. The authority and hub scores of every page can be calculated by using a power iteration method. According to the query topic, the main strength of HITS is ranking pages, which may provide more relevant authority and hub pages.

Finally implement CF approach to construct matrix based users, cities and GPS co-ordinate points. This is represented as shown in figure 3.

Figure 3: Mapping of user to cluster

This model contains two factors.

• The first one is the interest of a location which depends upon number of users visiting this location and the travel experiences of these users. The location interest is represented as

Location interest= \sum \ L_{ij}^k

And users experience is represented as hub score

H^k = h^k_i; 1 \leq i \leq L; 1 \leq j \leq C

Where c is the cluster and l is the level in tree graph.

• The second factor is that the users’ travel experiences and the location interests have a mutual reinforcement relationship.

134
B. Search by Image

Content based image retrieval (or) query by image content (QBIC) is the application of computer vision technique to the image retrieval problem from the large dataset. Content-based means analyzing the data of the image rather than the metadata such as keywords, tags or description associated with the image. Content refers to colors, texture or any other information that can be derived from the image. CBIR use query technique which involves an example image that it will then base its search upon. A pre-existing image can be used by the user to search. The result images should all share common elements with the provided example.

- The average of colour layout and edge histogram descriptor for all images is found and stored in the database.
- Calculate the distance of query image with that of images in the database using Euclidian distance

\[
H_1 = \text{Average histogram (image}_1) \\
H_2 = \text{Average histogram (image}_2) \\
\text{Distance} = \sqrt{\sum(H_1 - H_2)^2}
\]

Examining images based on the colors is one of the commonly used methods because it does not depend on image size or orientation. Thus it will retrieve similar image with the tag information associated with it.

C. Architecture

Figure shows the architecture diagram of our system, which is composed of the following two parts; search by location, search by image.

With the TBHG, we propose a hyper text induced topic search based inference model to estimate users’ travel experiences and location interests in a given region. Finally, the most experienced users X, top interesting locations M within the specified region can be suggested to the users.

Search by image: The user enters the search query in the system; from this image CEDD descriptor is extracted. This system uses image database in which indices contain the descriptor value for all images. Thus the images with most similar value are extracted with the tag information.

V. EXPERIMENTAL RESULTS

The experiment illustrates the framework of the evaluation, in which we respectively explore the effectiveness of the location and the travel sequence recommendation by performing the user study. In this study, .NET based application with Google API travel sequences is used which is based on ATCF and HITS approaches, including our methods and some baselines. As the subjects are familiar with this region, we are more likely to find out common ground truths shared by them. For location based recommendation, GPS logs of multiple users is taken as the input. It contains three fields. Each point is represented by an ID, an X coordinate (longitude), and a Y coordinate (latitude) with one point per line.

Table I shows four POIs with their latitude and longitude we mine in Madurai. The top ranked four POIs are famous POIs, such as Meenakshi Amman Temple, Vishal D Mall, Gandhi museum, Thirumalai Nayakar Mahal respectively. They could be mined by basic density based clustering.
Table I-
Famous poi in Madurai

<table>
<thead>
<tr>
<th>Latitude</th>
<th>Longitude</th>
<th>Poi names</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.90798</td>
<td>78.1021</td>
<td>Meenakshi Amman Temple</td>
</tr>
<tr>
<td>8.8045</td>
<td>79.2341</td>
<td>Gandhi Museum</td>
</tr>
<tr>
<td>9.9147</td>
<td>78.1242</td>
<td>Thirumalai Nayakar Mahal</td>
</tr>
<tr>
<td>9.60014</td>
<td>77.9506</td>
<td>Vishal D Mall</td>
</tr>
</tbody>
</table>

For image based search, a query image is taken as input and similar images with their tagged information are displayed as output. The most accurate images with least deviations are given high priority. The results are shown in figure 5.

![Image Search Result](image)

Figure 5: Results for image based search

And these results measured using response time is shown in figure 6.

![Response Time Graph](image)

Figure 6: Evaluation result

VI. CONCLUSION

In this paper, using the GPS trajectories generated by multiple users, we mined interesting locations and classical travel sequences within a given geospatial region. Such information can help us understand the correlation between the users and the locations, and enable travel recommendation as well as mobile tourist guidance. In this work, we consider an individual’s visit to a location as a link from the individual to the location, and weight these links in terms of the users’ travel experiences in various regions. A HITS-based model is proposed to infer the user’s travel experience and the interest of a location considering the following two aspects. One is the mutual reinforcement relationship between the location interest and the user travel experience. The other is that the user travel experience as well as the location interest is region-related. Later, we detected the classical travel sequences in a destined region using the location interests and the users’ travel experiences. Finally similar images with the tag information are retrieved for the query image users have.

REFERENCES


