

Travel Recommendation using Review Based Sentiment Analysis

Shirish Suresh Arote^{#1}, Prof. Rahul L. Paikrao^{*2}

[#]Department of Computer Engineering, Savitribai Phule Pune University
¹shirisharote94@gmail.com, ²rahul.paikrao@avcoe.org

Abstract— Web-based social networking data offer tremendous sum of chances to take a shot at many different fields and one of them is travel suggestion. This paper depicts a modified travel proposal from the two experiences of a visitor and online networking contributed photographs and the differing sorts of metadata (e.g. geo-place, Seasons and date taken) connected with these photographs. The proposed approach is not in any respect just like the existing processes because it no longer simplest consider user's journey interest but additionally elements like time, season and value as properly. To predict customized POI arrangement, first of all, online networking is a totally famous way of communicating surveys and connecting with various individuals on the planet. At the point when taken in collection of critiques can offer a reflection of communal viewpoint toward places. Opinion assessment is utilized to separate between the terrible opinions and the great opinions, a good way to offer opinions about the POI and through which the tourist can determine whether or not to go there or now not. The System can be analyzed using metrics like precision and recall which provide better understanding of the results obtained.

Keywords— Recommendation, Filtering, Sentiment analysis, stemmer, travelogues.

I. INTRODUCTION

In today world travelling is a thing that is seen as a virtue of satisfaction or adventure or many other thing. Many people want to travel the world in a more precise or in more organized way to fulfil their dream to visit new culture or land to explore new things. Many of the travellers or tourists prefer to go on their own to explore without information but that can be exhausting and more dangerous if we don't have information that we should have while travelling a certain land but other travellers tend to get suggestions or recommendation from the people who have visited the place that is to be explored by tourist who haven't visited that place. Basically recommendation provides all the information that is needed about place like how the place is? , how far it is? , how are the people around the place? Etc. Thus, this information cannot be gained without recommendation.

Recommendation is used by many big multinational companies like Google, Facebook, flipchart, amazon etc. Some of them use recommendation for suggestion of goods and services that are supplied by the companies like E-commerce sites. Some of them use recommendation for suggesting friends or interests like social media platforms and some of them provide recommendation for searching something of interest on internet like search engines.

It is said that the amazon e-commerce sites has about 20 % of its earning from recommendation system so it is very important that recommendation should be used to provide essential information that has been required by the user, thus Recommendation becomes an effective as well as very important factor in area of travelling. Recommendations in any field can be provided by using many different approaches like content based recommendation, collaborative recommendation and also not so popular context driven approach for recommendations. We are talking in terms of travelling, so many of these approaches work on the history of the user that has visited the place before like collaborative approaches where the user's friends or related people's history is taken under consideration to provide recommendation for future tourists or travellers to use as guidance for travelling. And, on the other hand the content based recommendation approach which is a more comfortable approach in which the database itself is used and desired genre or type of spot is directly taken from the database to provide recommendation for travelling.

Before selecting any approach as a recommendation tactics we should know that every recommendation system has two basic problem that are cold start problem and sparsity problem. The cold start problem means that if a new user comes into the system then he cannot get recommendations directly if collaborative approach is used for recommendation. He or she has to wait until the new user himself or herself makes new friends and then further process can be considered. The sparsity problem is the problem related to the unavailability of the information as this problem is present in any of the above approaches used for recommendation which implies that even if the user has friends if the friends haven't travelled much or haven't done any activity in the system then the recommendation cannot be provided as there is not enough or no information available for the desired or required spot of user to recommend.

Thus the paper will focus on how to decrease the effect of sparsity and cold start problem and also on how to effectively provide recommendation to tourists and travellers around the world so that they can gain more information of the place and have an adventure to remember for their whole life.

II. LITERATURE REVIEW

The framework in [1] is the proposed system which automatically mines visitors tour interest from photograph collections furnished by means of users on social media which incorporates consumption capability, desired time and season this is essential to path making plans. The device recommends POIs to people on the idea of his personal interest not on interest of many people. The similarity among individual bundle and course package deal presents rating of well-known

Routes, and top ranked well-known routes are in addition optimized in line with recognize to a user's adventure information. The system affords Topical Package Model (TPM) technique to look at users and routes tour attributes. It decreases the difference between the explorers intrigue and course suggested by the model or framework.

A system in [2] by Y. Gao is a system which guides about landmarks used in the dense areas where there are many landmark and people are confused about which land mark to visit first. The system has it merits like Removal of noisy photos, Tag relevance calculation, Landmark ranking and demerits like two geographical adjacent places of interest are inclined to be converged at times. Dishonestly perceive an landmark and tag it as a place of interest.

Yan-Ying Chen, An-Jung Cheng defined a model for journey suggestion [3] which uses pictures to determine attributes about human beings like race, gender, age and also type of group i.e. Own family, pals, couple and additionally uses spatiotemporal analysis(time and location). It enables to analyse human activities and may be carried out in cellular surroundings. Because of strategies the technique has a boundless also, fundamental application area and it will jump at the chance to recognize among the couple and circle of relatives spots through thinking about trip routes taken by explorers. The frameworks central objectives are to utilization of the human properties and social setting found in those pictures to rely on the type of gathering demonstrated inside the photograph and furthermore to decide the association between individual's properties and more settings (e.g. Time, surely known spots) also, show the advantage of checking human travel ways. Enhance the customized travel suggestion, particularly in the locale wherein people have various options of the associative areas.

A system portrayed in [4] through Xueming Qi an influences utilization of social components to like client's own special leisure activity, relational preferring closeness, companion's effect on voyager and the other way around is utilized for suggestion of courses. It fathoms the inconvenience of new begin clients (the person that just began the utilization of the framework or character who has considerably less companions or less appraised items) and sparsity inconvenience (machine which has much less information roughly a city in this way, the suggestion isn't great) likewise because of these endowments the framework is used in overview of any historic point or area for tourism thought process and in areas were data is low and still the trekkers require suggestion. But since the obstructions are concerned if the client has more contacts the frameworks will be more productive and will tend toward precision in any other case the contraption will be less exact and components like clients POI, points of interest time, cost, separate, season should be taken underneath consideration in the meantime as suggesting a way.

S. Jiang created a collaborative filtering primarily based version [5] wherein issue of sparsity is tackled to a little degree. For instance, recommending a region in light according or by comparing to included areas of current user's companions and partners. The framework is utilized in locations where records are low and nonetheless the tourists need advice and additionally inside the towns wherein travel journals aren't accessible the framework can utilize geo-labelled pics. The overwhelming issues is that no utilization of travelogues which give more prominent contextual records roughly about POI than the geo-labelled pictures and model cannot be utilized as a part of city whose records is significantly less or the city in which people are journeying a great deal less (less pictures for suggestion).

Yu-Ting Wen proposed a system [6] wherein user can specify set of keywords and area to be queried, then checking information is used for course prediction. Combining a couple of measurements of routes and by means of the usage of course reconstruction approach various experience routes throughout the place that is decided and is displayed. This sort of mining improve coverage of input data. The recommender system can help to provide recommendation in any areas as it can provide region to define as well as keywords to specify different place of interests or landmarks or historic spots. The concepts and algorithms used make the approach more user friendly.

J. Li in [7] proposed a system which is Useful for people who want to consider many different factors while starting a travel. Its Advantages are like Inverted file structure Hierarchical structure Clustering used to find popular spots and also has some Disadvantages like Well-developed machine learning is needed. GPS should be great as it provides input of present location of user.

Y. Him implemented a system [8] which has many area of application i.e. Used by passengers to find path for travelling, Can be used by government to show people path of region which doesn't have roads using checkpoints, Application could be developed to help users plan trips. It has many advantages like exploration of region due to path collection, Path ranking, Indexing is used for paths. Disadvantages such as absence of Review System - As the users have no prior knowledge about the landmarks they are visiting and Sparsity problem.

III. SYSTEM OVERVIEW

A. Proposed model

The users tour attributes which include POI, expense and favoured time and season in which to visit are taken and this information given to the framework and a suggestion i.e. successive route of region of interests in advance with the people necessity, Representative labels are mined from travel journals and also the pictures from the online social networking. As there numerous insatiable calculations the most capable arrangement of standards of them on the thought of framework necessity could be travelling salesman problem, steps which could supply us the shortest or brief path among the areas and counselled by usage of the content material primarily based completely filtering algorithm.

- 1) The journals additionally accommodate suggestion about the money spent with the resource of the person or tourist on the place of pastimes inside the area.
- 2) Seasonal year is divided into a proper travelling season by using information from the photographs of social media like date and time metadata.
- 3) Critiques from customers visited earlier than supply the comments related to their journeying moments in terms of nice or awful level which can be elements to undergo in mind for the next site visitors.
- 4) An algorithm for stemming is used to rework the terms from its original form to its root form wherein the suffixes are added to its unique root form. The set of rules calculates every permutation and combination that can shape a phrase the use of suffixes. the assessment characteristic of the technique ought to additionally have stop word idea to do away with all of the vain phrases from critiques of users and additionally tokenization concept to facilitate the working of analysing the critiques as good or bad. Sentiment assessment is adequately benign in amusing media monitoring as it lets in us to an outline of the public towards the place.
- 5) Additionally using content material filtering approach alternatively of collaborative approach for overcoming the dilemma of a new start/sparsity trouble. the content filtering set of regulations right here does not examine the active traveller requirement with different tourist search as studied in collaborative filtering however the set of rules compares the input with the database immediately therefore putting off the trouble of getting buddies or maybe records approximately about other tourists likings.

B. Working

Dataset is stored on the server in tabular form taken from Trip Advisor site as it provides more information that can be used for recommendation like travel journals, longitude, latitude of place and review to min related to the place. First the review retrieved from the dataset are stemmed to its root form then they are sentimentally analysed to gain more information about the opinion of the public about the place and the it is shown in more understandable form so future visitors can assess the place, this all is done by using concepts of sentiment analysis and stemming. The second phase would be to take the users input according to his or her interest and comparing it with the database retrieved from dataset as it will lead to elimination of sparsity problem and thus providing us with POIs which satisfy the user's requirement along with metadata associated with it. Last but not the least, the output from the second phase which is filtered POIs according to users interest will be provided as an input to greedy algorithm to find out the shortest path between those POIs so that the user may visit these spots in lesser time.

C. System modules

By examining numerous previous papers and also through taking under interest the constraints and inadequacies of past methodologies six modules were created to overcome not everything but some of the challenges faced with preceding recommendation structures. The module may get and transmit information to different modules. The depiction of these modules is as per the following:

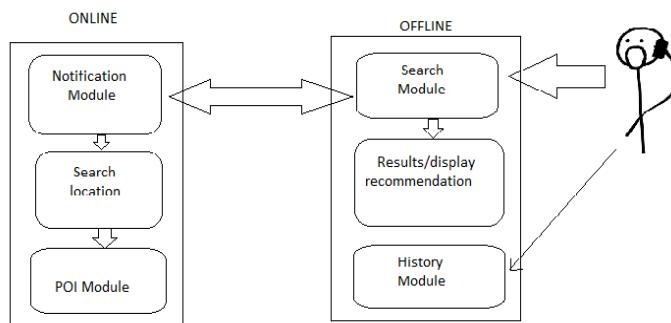


Fig. 1. System Module View

1) Online (administrator):**a) Seek module**

This module might be used by the administrator if a place which is sought by the individual isn't generally found after which by means of utilizing this module the administrator will transfer the sought POI by getting the notice.

b) Add/Remove module

This module may be dealt with by Admin to incorporate or to wipe out certain POI from the database.

c) Notification/alert module

At the point when a voyager looks to find a town and does now not discover it, a alert progressively given to the administrator for this motivation to include the coveted POI or town of the individual.

2) Offline (Client):**a) Seek module**

The individual enters a town's POI to which he or she wants to visit.

b) Module for previous search

To offer the client history about the before sought Places and furthermore spares preparing time when individual enters the same again.

c) Display end result

To display the recommendations according to the person desires.

C. Algorithms**1) Stemmer**

To convert the word with suffixes to its root form stemming algo is used. Initial step to check whether the word contain a vowel or consonant with proper word.

a) step1() : removal of plurals and different suffixes like -ed or -ing. e.g caring - care and deleted - delete

b) step2() : converts terminal y to i if there is another vowel in the stem.

c) step3() : maps double suffixes to single ones. So -ization (= -ize plus -ation) maps to -ize etc. note that the string before the suffix must give m() greater than 0 e.g industrialization - industrialize

d) step4() : includes with -ic-, -full, -ness etc. same working to step 3. Gyroscopic - gyroscope

e) step5() : removes -ant, -ence etc., in context (c)vcvc(v). e.g. dependence – depend

f) step6() : eliminates a final -e if m() greater than 1.

Stem the word set into the buffer of Stemmer through calling the add function. Returns genuine if the stemming process brought about a word not quite the same as the input information.

2) Content Filtering

Its basic approach include database as a whole search area instead of searching friends history in database, so the approach is to compare the query to the database of the spot of city and accordingly provide the result

And further review are used for recommendation

1. First positive and negative count of every spot that has been retrieved from content filtering approach is added
2. Then the number obtained by addition is used to obtain the percentage of the positive and negative review count for every spot
3. Final step is to take only those spot that satisfy the threshold value which decided previously(the spot which has more positive review than negative)

3) Travelling Salesman Algorithm

Initial the database provides us with the data about the distances between every locations and landmarks.

a) A graph is drawn from every location to every other location

b) From the initial point smallest distance among the every other point is calculated and put into the stack.

c) As there will be 2 point in the stack the initial one and the other where we reached so from where we reached we have to do the step 2 again till all the points recommended are not finished.

d) By doing so mark the visited point as 1 until all the point are visited

e) Finally when all the points are in the stack pop out them and join them and your shortest path will be available.

IV. SYSTEM ANALYSIS

As to illuminate the problem of cold start as the user doesn't need to have friends or any contact with other users of system as the users interest will be directly compared with database taken from Trip Advisor which are parameters like link reference(travelogues), longitude, latitude, location name and city itself as input for the algorithms. As the first algorithm implemented is for the purpose of sentiment analysis, the reviews are fetched from the link as the review words are tokenized to take the words separately for the input to porter stemming algorithm and then the remaining words in review are treated as stop-word and are eliminated.

For e.g. if the review is "This historical place is prideful and lovely to watch" then the tokens will be,
 (1, prideful)
 (2, lovely)

These tokens i.e. 1 and 2 for prideful and lovely resp. are provided to stemming algorithm to convert to its root form i.e. pride and love. The words are compared to the NLP dictionary to see the words are related to good or bad opinion and then added to the count to of good or bad reviews about the place depending on the results from NLP dictionary. As the porter stemming algorithm is used as it requires less software and hardware requirements and also less computation logic as compared to other techniques like lemmatization which requires database for morphological order as well as for vocabulary. Then the actual work of the system starts i.e. content filtering whose input will be users interest i.e. <city, POI> and then content based filtering will be carried on using the input and database and result will be the POI suggested by algorithm on the basis similarity between the users requirement and database for e.g. cosine similarity can be used. Then the output of content based filtering is used to give input to travelling salesman algorithm. Distances between all these spots which are suggested by content based algorithm are calculated previously using the dataset taken, so there is decrease in processing time required by the overall system. The output of the travelling salesman problem will be the final output of system i.e. sequence of spot recommended by the system according to the user's requirement. The travelling salesman algorithm is used because it provides an easier way to generate the adjacency matrix that includes distances between spot, which might be an overhead in different greedy algorithms.

The metrics used to analyse the overall result of the system are recall and precision, the division of pertinent spots among the recovered spots is precision (likewise positive predictive value), while the division of applicable detects that have been recovered over the aggregate sum of important spots is recall (AKA Sensitivity). Both accuracy and review are in this way in light of a comprehension and measure of pertinence.

$$\text{Precision} = \frac{\text{relevant locations} \cap \text{retrieved locations}}{\text{retrieved locations}}$$

$$\text{Recall} = \frac{\text{relevant locations} \cap \text{retrieved locations}}{\text{relevant locations}}$$

V. EXPERIMENTAL DETAILS

The dataset is Trip advisor from <https://www.tripadvisor.in/> website for the details of location. i.e. Location, review, poi, seasons and the attribute Location and poi is taken as a sensitive attribute in the dataset.

As mentioned earlier there are mainly two modules admin and user so the figure 2 show admin search module which basically searches dataset for retrieve data according to users requirement as the data retrieved is as follows longitude, latitude, link of the specific individual location, Name of location, POI under which it is categorized. So, For searching admin has to provide with keyword i.e. name of the city and POI by selecting from list of POI's. The Search will result into retrieval of all above data relevant to the given input and stored in the database.

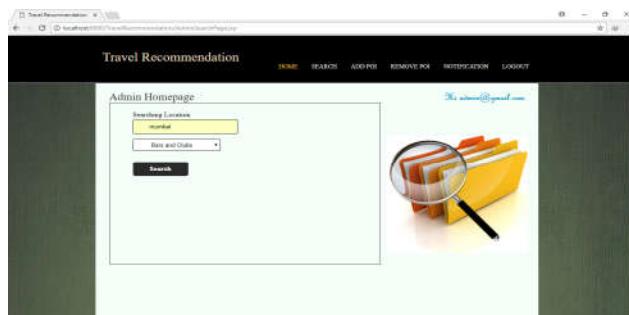


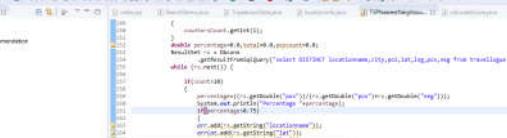
Fig. 2. Admin Search Module

When primary information has been retrieved then comes the time for retrieving reviews associated with the individual location. So, For that use of link that has been stored for individual location is used and then reviews on that page are retrieved and on that sentiment analysis is done and all that data is store in the database i.e new data stored will be reviews, season to visit, calculated score of reviews of that location in terms of positive and negative as shown in fig. 3

Then the calculated score plays a very important role in the analysis or evaluation phase of the system which is the primary objective of the system i.e. to provide more precise information for the desired requirement.

Fig. 3. Data Retrieved from Dataset

Fig.4. Below show how the working is done when the query or searching keyword (i.e city) along with the POI is given. The query is fired after the requirement of the user is given, the query is an SQL query based on the content filtering approach to find desired POI of that city and provide the results. Thereafter a threshold is defined to satisfy the criteria given in the algorithm to gain more precision.



The screenshot shows the IntelliJ IDEA interface with the following details:

- Project Explorer:** Shows a single package named "com".
- Code Editor:** Displays Java code for a class named "RecommendationSystem". The code includes imports for `java.util.List`, `java.util.ArrayList`, `java.util.HashMap`, `java.util.Map`, `java.util.Scanner`, and `java.util.Set`. It also imports `com.mongodb.MongoClient` and `com.mongodb.DB`.
- Code Snippets:** A snippet for `for (String s : set)` is visible.
- Toolbars:** Standard Java development tools like `File`, `Edit`, `Source`, `Refactor`, `Debug`, `Search`, `Project`, `Run`, `Window` are present.
- Status Bar:** Shows the file path as `C:\Users\hitesh\IdeaProjects\RecommendationSystem\src\main\java\com\RecommendationSystem.java` and the line number as `140`.

Fig.4. Content-Based Approach

After all the calculations are done then the TSP comes into play and then the output from content based approach is given as input to TSP and then the map shown in fig.5. below and is displayed as output of travelling salesman problem.

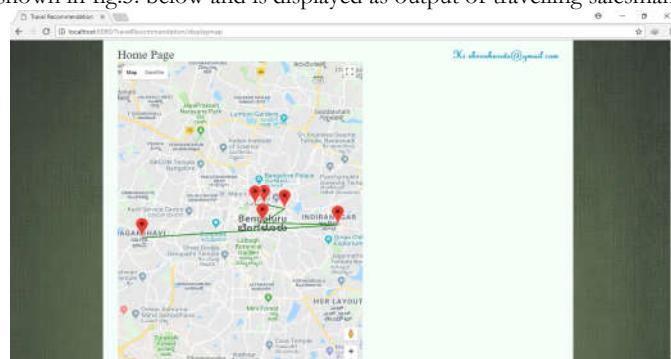


Fig.5. TSP

VI. RESULT ANALYSIS

In this section we compare our results with existing system. We would like to appreciate work of existing system authors. Our result shows that the performance of the proposed system is better than all other system that corresponds to this type of recommendation. As the reviews are the main criteria is to provide recommendation for user instead of cost and time so it provides more precision. After result analysis we will discuss limitations of our system.

In evaluation, Minimum requirement for execution of this application is JDK 1.7 and MySQL 5.5 for the implementation and run on 3.2 GHz Intel Core 2 Duo Processor machine with 4 GB RAM. The Microsoft Windows 7 Professional can be used as an operating system. So, four factors should be considered i.e. rationality, representation, diverseness, satisfaction. In evaluation, other than these four factors, the volunteers should also consider whether the routes meet the season preference, expenses and time and users interest.

In this work, users with different gender and generations should be used to conduct human analysis. Before evaluation, the travellers have to understand user's preference by taking under consideration travel records including photo from social media, travel journals and reviews about landmarks. For evaluation of performance the metrics used are the weighted average as well as average precision.

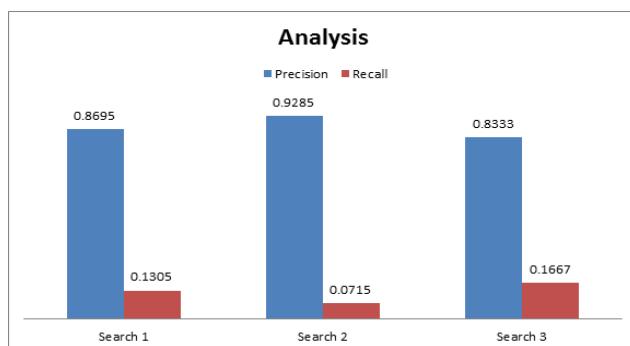


Fig. 6 Performance

The graph shown above shows the precision and recall of the system developed. Three searches were done for each city with its single POI and then its precision along with its average precision was calculated. So, three different cities and their different POI's were searched which shows precision and recall ratio is much better due all the factors i.e. reviews, season, distance considered. Thus the graph also shows how the criteria for better POI in cities with more good reviews changes the precision and recall ration.

The graph below shows that how the system is compared with all the other system which work with most of the same techniques and application but the difference is the factors used in the proposed system thus the average precision is more than all of the other system even if the criteria (review criteria) mentioned above is changed to any scenario.



Fig. 7 Comparison with other systems

The results table shown below basically shows that there are approaches like FAM i.e famous route planning without ranking and optimization, RFA i.e ranked famous route planning using users requirement without optimization, OPT i.e optimized route planning with ranking and optimizing which are compared with proposed system in terms of average precision. The average precision for proposed system is shown in the table.

| Search No. | FAM | RFA | POT | PROPOSED |
|------------|------|------|------|----------|
| Search 1 | 0.62 | 0.68 | 0.8 | 0.86 |
| Search 2 | 0.6 | 0.64 | 0.76 | 0.92 |
| Search 3 | 0.55 | 0.61 | 0.71 | 0.83 |

Table 1 - Performance result

A. Limitations

The data sparsity still remains the problem for all the recommendation system so we will try to solve that problem in future. The system is not been tested for all the dataset. More factors should be considered to get more precise recommendation and satisfy the user.

VII. CONCLUSIONS

The proposed version presents suggestion to clients on the possibility of their enthusiasm and by implies of consolidating their enthusiasm with different components like season, time, traveling hours which furthermore affect the proposal model. As the framework significant component that is assessment of the clients who went to place earlier than the current user gives the surveys about in their feeling like great or terrible level which can be a perspective to recall for the ensuing site tourist, with a view to accomplish better travel suggestion. The utilization of content material filtering instead of collaborative approach for beating the bind of cold begin inconvenience/sparsity to a degree. As there are various properties and elements utilized, this assists in additional use of the rich huge web-based social networking databases. The framework has numerous favors as it thinks about POI of each individual and takes under consideration numerous components to prompt a movement succession like the likeness between the sightseer's necessity and embraced accumulation might be a higher match than the previous methodologies.

In future work the system will be tested in different data sets as well as formats and the system will be also updated to improve its performance as well as efficiency and also in terms of security. The model will also try to reduce the effect of noisy data as well as the effect of sparsity further by using different technologies.

ACKNOWLEDGMENT

Not a single work can be viably completed without the assistance and inspiration from the scholarly individuals. I might want to broaden my much obliged gratitude to all associates who have been there at each progression for my assistance. I might likewise want to offer my thanks to my coordinator for his consistent help and inspiration that has supported me to think of this paper and furthermore exceptionally appreciative to family and companions who have given their whole hearted help dependably for the accomplishment of my paper.

REFERENCES

- [1] Shuhui Jiang, Xueming Qian, Tao Mei, Personalized Travel Sequence Recommendation on Multi-Source Big Social Media IEEE Transactions on Big Data, vol. x, no. x, September 2016.
- [2] Y. Gao, J. Tang, R. Hong, Q. Dai, T.-S. Chua, and R. Jain, W2go: A travel guidance system by automatic landmark ranking, in Proc. ACM Int. Conf. Multimedia, 2010.
- [3] A. Cheng, Y. Chen, Y. Huang, W. Hsu, and H. Liao, Personalized travel recommendation by mining people attributes from community-contributed photos, in Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011, pp. 8392.
- [4] X. Qian, H. Feng, G. Zhao, and T. Mei, Personalized recommendation combining user interest and social circle, IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 7, pp. 17631777, 2014.
- [5] S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, Author topic model based collaborative filtering for personalized poi recommendation, IEEE
- [6] Yu-Ting Wen, Jinyoung Yeo, Wen-Chih Peng, Seung-Won Hwang, Efficient Keyword-Aware Representative Travel Route Recommendation IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 8, august 2017.
- [7] J.Li, X. Qian, Y.Y.Tang, L.Yang and T.Mei," GPS estimation for places of interest from social users' uploaded photos" IEEE Trans. Multimedia, vol. 15, no. 8, pp. 2058–2071, Dec. 2013.
- [8] H. Yin, C. Wang, N. Yu, and L. Zhang, "Trip mining and recommendation from geo-tagged photos," in Proc. IEEE Int. Conf. Multimedia and Expo Workshops, Jul. 2012, pp. 540–545.