International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

A Personalized Location Recommendation System based on Probability and Proximity

Aravind J, Parvathi R

Abstract**:** ***Big data is an emerging field plays a valuable role in the extraction of information from raw data. It has found its applications in areas such as predictive analytics, healthcare analytics, financial analytics, and retail analytics and so on. The enormous growth of the Internet has become a source for availability of the huge volume of data online. It is difficult to find out the necessary information from huge data within a short period. The availability of enormous data craves the need of an information filtering system and this information filtering systems are capable of providing the required data to users. The rapid growth of big data lays the path to recommendation systems. A recommendation system is an information filtering tool which has more impacts in day to day life of everyone and also redefines our lives. Recommendation systems provide suggestions based on user preferences, requirements, and interests. The reviews and rating values given by people are used to answer similar interest queries with predictions and suggestions. Reviews and Feedback play a key role in the decision-making process. People share their experiences in the form of feedback, ratings, and reviews and so on. If a user wants to visit a location and if he does not have prior knowledge of it, then he may use reviews and feedback given by others who visited the location already. It is not possible for a user to go through huge volumes of reviews and sometimes it may mislead the user to take wrong decisions if he goes by the review given by a person with a contrasting taste. In such cases, Recommendation systems are needed, which helps users in the decision-making process. In most of the existing methods, they used Point of Interest (POI) of users to recommend the locations. The main objective is to develop a Personalized Location Recommendation System, which will recommend the locations to users using Probability and Proximity. Our model uses Probability and Proximity measures to recommend the locations.***

Index Terms: Big data, Information Filtering, Probability, Proximity.

I. INTRODUCTION

Recommendation systems emerged in the 1990s mainly focused on retrieval and filtering of information from the enormous amount of raw data. Recommendation systems achieve widespread success in E-commerce, Advertisements and Social Networking. Various knowledge based techniques were discovered and applied to the data to get the preferences, interests and requirements of people. A Recommendation System is an information filtering tool which works with massive amount of data, filters data and provides data according to the users requirements. It acts as an interface between the people and services they need. It will figure out the services which are suitable for people based on their interests, preferences, and requirements, and suggest those services in an efficient manner. Recommendation systems have their roots in information retrieval, cognitive science and forecasting theories. They find their applications in areas such as movies, music, news articles, research articles, products, locations, social tags and so on. The recommendation system helps to people in their decision-making process and makes it so easy and simple.

II. RELATED WORKS

Jiang et al. [1] proposed a travel recommendation model by using travelogues and community contributed photos. This model recommends a travel package to users, which include places to visit, cost, visiting time and visiting season of each location. It also suggests the best routes to visit the locations. It ranks the routes based on the user's point of interest (POIs). Top K routes are recommended for users to visit a location. Kunhui Lin et al. [2] proposed an adaptive location recommendation model for users of location based social networks. This model formulated an algorithm which uses user collaborative filtering, similarity between the users and naive Bayesian classification. It finds out the current location of each user and recommends the locations similar to the current location. Ling Xing et al. [3] designed a novel personalized location recommendation model for social network users. It uses logistic regression and collaborative filtering to provide recommendations to users. Logistic regression is used to train the weights of features of items and collaborative filtering is used to find the similarity between the items as well as users. Ajantha et al. [4] proposed a personalized tourism and travel recommendation system which collects user's information from social websites and travelogues. The user's information such as travel history, posts, gender, emotions, reviews, ratings are collected and the data is preprocessed initially. The model finds the similar users based on the data collected by using similarity measures. It uses K Means clustering to group the users based on their age and gender. It computes the user location vector based on users travel history. The popular places are extracted from travel blogs and the places are arranged based on their popularity. The system recommends the top N locations to users by using user-location vector. Zeng et al.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Aravind J*, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India.
Parvathi R, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India.

[5] proposed a point of interest based location recommendation system. It captures the point of interest of users and recommends the locations which are similar to POI. It finds the similar locations by using time slots. It divides a day into 12 time slots. Each time slot is of duration 2 hours. The check-in times of users in locations are collected and based on the check-in times, the system finds the similarity between 2 locations. Finally the system recommends the locations which are similar to POI of users. Jiangning Hea et al. [8] proposed a model which recommends the travel package to users by analyzing the user behavior. It classifies the users into two groups, namely individual travellers and group travellers. It recommends the locations based on Point of Interest (POI) to individual travellers and Social Relationships in case of group travellers. Kesorn et al. [6] proposed a novel personalized recommendation system for tourists based on check-in data. This model extracts the check-ins of people from Facebook. It analyzes the behavior of users and finds the relationship between the users based on behavior and check-in data. Factors like affinity score, edge weight and time decay are used to find out the relationship between the users. The RF techniques calculate the user relationship scores and provide recommendations based on scores. Ziqing Zhu et al. [9] proposed a location-time-sociality aware personalized tourist attraction recommendation in LBSN. It provides recommendations based on user preference, social relationship and location popularity. It clusters the locations by using Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm by using check-in records of users and it extracts the preferences of users. It recommends the locations by combining the location clusters and user preferences. Zhiwen Yu et al. [7] proposed a model which uses the Apriori algorithm to obtain the frequent travel pattern of the user. This travel pattern is used to understand the users interest and location preferences in order to recommend the locations. Xie et al. [10] proposed a model which generates the travel packages for users. It gets the estimated budget of users for travel and then forms a travel package which will be covered within the budget..

III. PROPOSED ALGORITHM

The architecture of proposed system is given in the Fig.1.

A. *User Information Collection Module*

The system collects the travel history of users. The travel history contains information such as name of the user, location travelled, latitude and longitude of location, reviews and ratings given by users. Since it is a real time data analytics application, the streaming of data is needed. To stream the data in real time, the system uses Apache Kafka messaging architecture. Apache Kafka follows publish-subscribe model. The Kafka Producer collects the travel information at one end and forwards them to Kafka Broker. On the other end, Kafka Consumer receives the travel history of users from Kafka Broker and represent them in bipartite graph form. One set of vertices contain users and another set of vertices contain locations. There is a relationship between user and location if a user visited the location already.

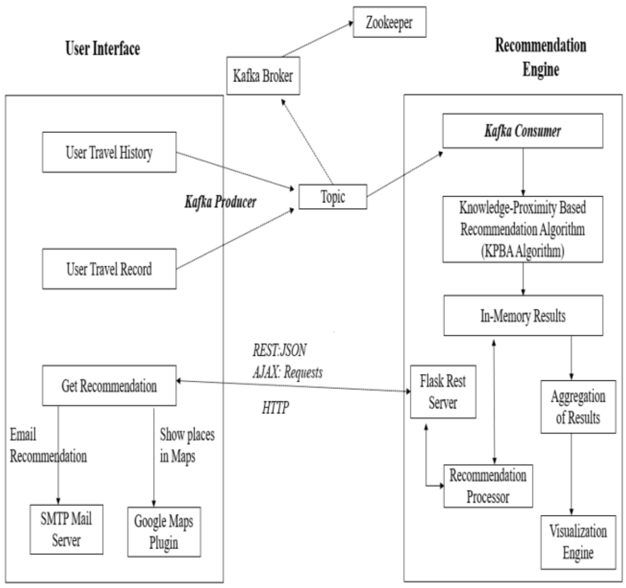

Fig.1. System Architecture

Data Source: Travel history of users (i) Represent the given data source in bipartite graph form (ii) Weight of edge between user X and location L is Weight of an edge= {1, if user X visited location L; {0, Otherwise

B. *Grouping of Locations and Users*

The system finds the number of people visited each location (i.e.) visited count of each location. It groups the locations based on visited count. The locations which have same visited count will be in one group. Users are then grouped based on the location groups. The users who visited locations in a particular location group will be in one group.

(i) Find the visited count (VC) of each location (ii) Group the locations based on visited count (i.e.) LG= {L1, L2, L3… Lk} L1, L2 in LG iff VC (L1) =VC (L2) (iii) Group the users based on locations group (i.e.) UG= {U1, U2, U3… Uk} U1, U2 in UG iff U1, U2 visited any of Locations in LG LG-Location Group, UG-User Group. L1, L2, L3...Lk-Locations. U1, U2, U3... Uk-Users.

C. *KPBA Algorithm*

The Kafka Consumer runs Knowledge Proximity Based Algorithm (KPBA) algorithm and recommends the locations to users. It provides recommendations based on Probability and Proximity measures. The flow of KPBA algorithm is given in the Fig.2.

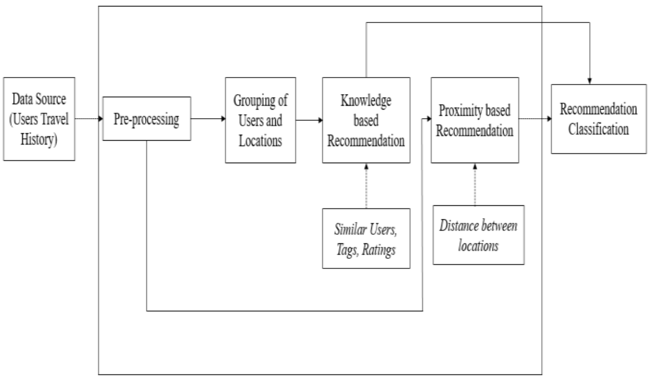

Fig.2. KPBA Algorithm

Recommendation Based on Probability

In each user group, User will gain knowledge about the locations in which he did not visit yet from users who visited, feedbacks and ratings given by users. Once he gets information about locations which did not visit yet, then the weight of edge between user u and location l will be updated from 0 to the value of knowledge gained. In each user group, find out the locations which are not visited by each user. Each user will get knowledge about locations which he did not visit already from other users who visited those locations within that user group. Knowledge value will be calculated by

$$\text{User X-Loc L} = \frac{\text{Number of users in the group visited Location L}}{\text{Total number of users in the group}} \quad (1)$$

User X-Loc L: User X gains knowledge about Location L
User X-Loc L knowledge value will be computed multiple times if user X present in multiple groups. In such cases, the Knowledge value will be maximum of knowledge values obtained in different user groups. People share their experiences on products or services or travels in the form of reviews or feedbacks, which enables other people for their decision making. Each user will gain knowledge about the locations which he did not visit yet from feedbacks which are given by users who visited those locations. Knowledge value will be calculated by

$$\text{User X-Loc L} = \frac{\text{Number of users given feedback about Location L}}{\text{Total number of users visited location L}} \quad (2)$$

Each user will gain knowledge about the locations which he did not visit yet from ratings which are given by users who visited those locations. Knowledge value will be calculated by

$$\text{User X-Loc L} = \frac{\text{Number of users given ratings about Location L}}{\text{Total number of users visited location L}} \quad (3)$$

The overall knowledge gained by users about locations is calculated by adding the knowledge value obtained from various sources. The value of knowledge gained by users about locations is calculated by

$$\begin{aligned}\text{User X-Loc L} = {} & 0.5(\text{Knowledge gained from similar users}) + \\ & 0.3(\text{Knowledge gained from feedbacks}) + \\ & 0.2(\text{Knowledge gained from ratings}) \quad (4)\end{aligned}$$

Location Recommendation based on Probability (i.e.) Knowledge value

(1) Find the mean of Knowledge gained by each user.
(2) Recommend the locations where Knowledge gained by user is more than or equal to mean.

Recommendation= {'Recommended Location',
if Knowledge gained greater than mean;
{'Knowledge not sufficient',
Otherwise;

Recommendation Based on Proximity

The system uses proximity measure to recommend the locations to users.

(1) Find the visited and non-visited locations of each user.
(2)Calculate the distance between the visited and non-visited locations.
(3) Recommend the locations which are nearer to visited locations of user.

D. *Recommendation*

The locations which are recommended based on both probability and proximity will be classified as "Highly Recommended Locations". If a user gives a request to our system to know the recommended locations, it forwards the request to the Flask REST server via JSON and AJAX. The Flask REST server gets the response from in-memory of Kafka Consumer and forwards the response to the system via HTTP. The system is integrated with Google maps and SMTP mail server. So the user gets the recommended locations via Email and also via Google maps.

IV. TOOLS AND TECHNOLOGIES

A. *Language: Python*

The system is developed using Python. Python is an object oriented and high level programming language. The coding in python is simple when compared to other languages and also it is platform independent. It is used for a variety of applications such as web applications, workflows, prototyping and so on. It is a more interactive programming language and it can be easily integrated with C, C++, and Java and so on.

B. *User Interface: Tkinter Python*

Tkinter is a package mainly used for creating Graphical User Interface (GUI) applications. It is a standard GUI library for python. It is used along with python. It consists of tools like widgets, labels, window, list box, menu button, entry, frame and so on.

C. *Dataset*

The system uses four square datasets to collect users travel history and also it collects the travel history of users in real time. Four square consists of user information such as id of users, locations they visited, latitude, longitude, reviews and ratings. It consists of check-in details of 3112 users and details of more than 2000 venues. The real time dataset consists of check-in details of 232 users and details of more than 50 venues.

D. *Apache Kafka*

Apache Kafka is a messaging system used for data streaming in real time applications in a distributed environment. It follows publish-subscribe model. It has a robust queue that is capable of handling huge volume of data and also enables us to transmit messages from one end to another end. It is suitable for both offline and online message transmission. Messages are replicated within the cluster to prevent data loss. It is built on the top of a Zookeeper, a synchronization mechanism. Zookeeper tells about the availability of Kafka Broker to Kafka Producer and Kafka Consumer. Apache Kafka is very fast, performs two million writes/sec. We can integrate Kafka with Apache Spark or Storm for real time data streaming applications.

V. RESULTS AND DISCUSSION

The recommendation system uses probability and proximity measures to recommend the locations. The system uses a hybrid filtering (i.e.) combining content based filtering and collaborative filtering to recommend the locations to users. The top 5 locations based on visited count is displayed in Fig.3. The Fig.4 and Fig.5 represents top 5 locations recommended to users based on probability and proximity respectively.

A. *Comparison of Results*

The proposed model uses Kafka technology and the execution of an algorithm is faster when compared to other algorithms. It can process huge amount of data within fraction of seconds. Our model is compared with Alternating Least Square (ALS), Simultaneous Matrix Completion Algorithm (SMCA), CD algorithm and the comparison results are given in the Fig.6. The models are compared in terms of execution and our model works better than the existing models.

Location	Visited Count (No of People visited)
Goa	600
Bangalore	500
Thailand	400
Mumbai	400
Chennai	400

Fig.3. Top 5 Locations based on Visited Count

Location	Count of People (Recommended)
Malaysia	1300
Madurai	1300
Kodaikanal	1300
Thanjavur	1100
Bangalore	1000

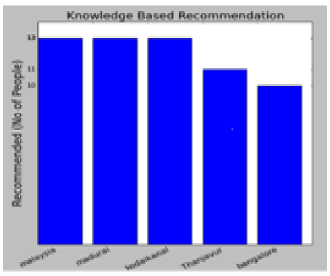

Fig.4. Top 5 Locations based on Probability

Location	Count of People (Recommended)
Kodaikanal	1300
Thanjavur	1300
Madurai	1200
Trichy	1100
Chennai	1000

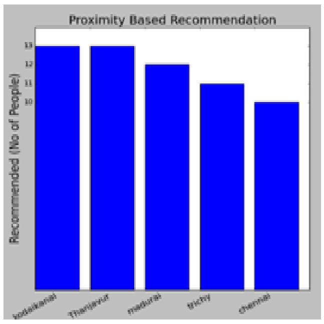

Fig.5. Top 5 Locations based on Proximity

Algorithm	Data Size (in %)	Time (in sec)	Data Size (in %)	Time (in sec)	Data Size (in %)	Time (in sec)	Data Size (in %)	Time (in sec)
SMCA	10	55	40	120	70	250	90	361
ALS	10	400	40	480	70	530	90	590
CD	10	60	40	85	70	140	90	165
KPBA	10	0.7	40	16	70	58	90	73

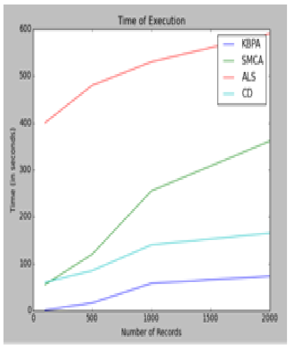

Fig.6. Performance Comparison of Various Algorithms

VI. CONCLUSION AND FUTURE WORK

In this work, We propose a Personalized location recommendation model which is used to recommend the locations to users which they did not visit yet already. This model provides a recommendation to users based on probability and proximity measures. This model is developed using single Kafka Producer and Consumer. Kafka Producer is the source of travel history of users and Kafka Consumer runs our algorithm and provides recommendation. This model will be enhanced in future in a way that multiple Kafka Consumers will be used so that it leads to parallel programming.

REFERENCES

1. Shuhui Jiang, Xueming Qian, Tao Mei, Yun Fu (2016), "*Personalized Travel Sequence Recommendation on Multi-Source Big Social Media*", IEEE Transactions on Big Data, VOL 2. NO.1
2. Kunhui Lin, Jingjin Wang, Zhongnan Zhang, Yating Chen, Zhentuan Xu (2015), "*Adaptive Location Recommendation Algorithm Based on Location-Based Social Networks*", The 10th International Conference on Computer Science and Education (ICCSE 2015)
3. Ling Xing, Qiang Ma, Song Chen (2017), "*A Novel Personalized Recommendation Model based on Location Computing*", Chinese Automation Congress, IEEE Conference 2017
4. Ajantha D, Jobi Vijay and Raji Sridhar (2017), "*A user-location vector based approach for personalized tourism and travel recommendation*", International Conference on Big Data Analytics and Computational Intelligence (ICBDAC) 2017
5. Jun Zeng, Yinghua Li, Feng Li , Junhao Wen , Sachio Hirokawa (2017), "*A Point-of-Interest Recommendation Method Using Location Similarity*", 6th IIAI International Congress on Advanced Applied Informatics 2017

6. Kesorn, K, Juraphanthong, W, and Salaiwarakul, A (2017), "*Personalized Attraction Recommendation System for Tourists through Check-In Data*", IEEE Access 2017
7. Zhiwen Yu, Huang Xu, Zhe Yang, and Bin Guo (2016), "*Personalized Travel Package with Multi Point of Interest Recommendation Based on Crowd sourced User Footprints*", IEEE Transactions On Human-machine Systems, VOL 46. NO.1
8. Jiangning Hea, Hongyan Liua, Hui Xiong (2016), "*SocoTraveler:Travel-package recommendations leveraging social influence of different relationship types*", Journal of Information and Management 2016
9. Ziqing Zhu, Jiuxin Cao, Chenghao Weng (2018), "*Location-Time-Sociality Aware Personalized Tourist Attraction Recommendation in LBSN*", IEEE 22nd International Conference on Computer Supported Cooperative Work in Design 2018
10. M. Xie, "*Breaking Out of the Box of Recommendations: From Items to Packages*", Proc. ACM RecSys 10, 2010
11. [11]Hsiu-Sen Chiang, Tien-Chi Huang (2015), "*User-adapted travel planning system for personalized schedule recommendation*", Journal of Information Fusion, Vol.21
12. K. W. T. Leung (2011), "*A Collaborative Location Recommendation Framework Based on Co-Clustering*", Proc. SIGIR '11
13. Bao J, Zheng Y, Wilkie D (2013), "*Survey on Recommendations in Location based Social Networks*", ACM Transaction on Intelligent Systems and Technology
14. Ference G, Ye M, Lee W C (2013), "*Location recommendation for out-of-town users in location-based social networks*", ACM international conference on Conference on information and knowledge management
15. X. Qian, H. Feng, G. Zhao, and T. Mei (2014), "*Personalized recommendation combining user interest and social circle*", IEEE Transactions on Knowledge Data Engineering
16. Q. Yuan, G. Cong, and A. Sun (2014), "*Graph-based point-of-interest recommendation with geographical and temporal influences*", International Conference on Information Knowledge Management
17. H.P. Hsieh and C.T. Li(2014), "*Mining and planning time-aware routes from check-in data*", ACM International Conference on Information and Knowledge Management
18. Y. Zheng and X. Xie(2011), "*Learning travel recommendations from user generated GPS traces*", ACM Transactions on Intelligent Systems Technology

AUTHORS PROFILE

Aravind J is pursuing M.Tech Computer Science and Engineering ar Vellore Institute of Technology, Chennai. He completed his B.Tech in Information Technology at Madras Institute of Technology, Anna University. His research interests include Data analytics, Recommendation Systems and Social Mining.

Parvathi R completed her doctoral degree from Anna University, Chennai, India, by contributing his ideas to the field of Spatial data mining. She has a teaching experience of over 20 years in the field of computer applications. Her research interests include Data Mining, Recommendation Systems and Social Network Analysis. She has authored articles in Big data analytics for renowned publications.