Volume 6, Issue 11, November 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Find the Movement of the User Based on Latitude, Longitude and Clustering by Using K Meansalgorithm

Suman Bala
Student, Doaba Institute of Engineering and Technology, Kharar, PTU University, Punjab, India

Maninder Kaur
Assistant Professor, Doaba Institute of Engineering and Technology, Kharar, Punjab, India

Abstract: The increasing availability of GPS-enabled devices is changing the way people interact with the Web, and brings us a large amount of GPS trajectories representing people's location histories. In this paper based on Gps trajectories we have developed an algorithm which extract the data of most liked users from Gps trajectories and finally display on user smart phone with location and their number of user via android application.

Keywords—GPS, Spatial database

I. INTRODUCTION

Recent advancement in the telecommunication technologies and positioning devices like Global Positioning Devices (GPS) has generated lot of interest amongst researchers to mine and learn patterns of human behavior from spatio-temporal data. The easy availability of these GPS devices has enabled convenient logging of movement and travel histories of users. In recent past, several users have started recording their movements with GPS enabled devices and uploading their GPS traces on web. These traces can now be used to infer how people move around in the city and extract their context and habits. For instance, GPS data tracked on mobile devices contains rich information about human activities and preferences. Analysis of the collected traces is an important research problem which can find applications in recommending interesting places to the tourists, traffic planning, planning locations for advertisement hoardings, itinerary planning etc. This paper aims to mine popular places from spatio temporal data using semantics of the places. By popular places we mean the places which are frequently visited and Mining Popular Places in a Geo-Spatial Region also liked by a large number of people. Example of such places is historical monuments, temples, parks, restaurants, shopping malls etc. There has been lot of prior work on finding interesting places in a geo-spatial region. In most of these approaches a location which is visited by many people is considered as interesting place. We must consider the fact that there are places which are visited by several users frequently but these may not be popular places such as traffic signals. If some comments regarding what user did, what user saw and how he/she felt on some place are attached, then it may bring more significance. Therefore, we aim to use multiple users' GPS log along with location semantics to find interesting locations in a geospatial region. We are proposing an approach that takes into account the semantic features of the places (such as type of place, what user did, how user felt at this place etc) in order to find interesting places for tourists in a geo-spatial region. We propose to attach semantic tags to each place which is visited by several users over a period of time. Finally, the places are ranked based on their popularity (frequency of visit by multiple users) and the semantic tags attached..

II. SPATIAL ASSOCIATION RULE MINING

Spatial association rule mining extracts association rules whose support and confidence are not less than the user-defined min_sup and min_conf, respectively. The problem to be solved by data mining can be decomposed into two sub-problems: (1) how to locate all the frequent itemsets whose support is at or more than the min_sup, and (2) how to generate all the association rules whose confidence is at or more than the min_conf for the frequent itemsets. The solution to the second sub-problem is straightforward, so the focus here is to develop a new efficient algorithm to solve the first sub-problem.

Process of Association Rule

Suppose $I = \{i1, i2, \ldots, in\}$ is a set of items—that is, an itemset. An itemset is a collection of items: k itemset is an itemset containing k items. If $A \subseteq I$, $B \subseteq I$, and $A \cap B = \emptyset$, then the association rule is an implicative form like $A \Rightarrow B$. D is the set of transactional data relevant to the mining task. Each transaction T with an identifier TID is the itemset enabling $T \subseteq I$. A frequent itemset is one where the occurrence frequency is no smaller than the product of min_sup (the minimum support threshold) and the total number of transactions in the transaction set D. Another threshold is min_conf (the minimum confidence threshold). The relationship is mapped among a collection of objects or attributes. The composite operation of the two mappings is defined as a closed operation $\gamma(l) = \alpha(l) \circ \beta(l) = \beta(\alpha(l))$ for the itemset l, whose closed operation $\gamma(l)$ is the largest collection of all objects that contain l. Definition 8.1 For frequent itemset l, $\gamma(l)$

= l means that the set is a closed itemset. When the condition also satisfies min_supp, it is called a frequent closed itemset According to the definition, it is intuitive for an itemset l to generate the frequent closed itemset with the following method. First, perform the operation α to identify the set of all the objects that contain l. If the percentage containing such objects versus all the collections is no smaller than the degree of the support threshold, then it is evident that the collection is frequent; Second, perform the operation β and then calculate the joint itemset of all the objects in the collection. However, the efficiency of such an implementation is very low. In the process of association rule mining, after fully understanding the data, the user must define the objective, prepare the data, select the appropriate min_sup and min_conf, and understand these association rules. Thus, the mining process of association rules includes the following steps: (1) Prepare the dataset. (2) Set the min_sup and min_conf. (3) Find all the frequent itemsets whose support is no smaller than the min_sup with the mining algorithm. (4) Generate all the strong rules whose confidence is no smaller than the min_conf according to the frequent itemsets. (5) Adjust the support threshold and confidence threshold for regenerating the strong rules if the amount of generated rules is too many or too few. (6) Refine the association rules of interest by using professional knowledge to understanding the generated rules. Within these steps, the most complex and time-consuming job is step (3) of generating the frequent itemsets. Based on the frequent itemsets, step (4) to generate association rules is relatively simple, but how to avoid the generation of too many redundant and excessive rules also needs to be carefully considered. Other steps may be thought of as related and auxiliary.

III. LITERATURE SURVEY

A research study done by Taysir H. A. Soliman et al. [1] in their research As crime rates are increasing worldwide, crime mining requires more efficient algorithms that can handle current situations. Identifying crime hot spot areas via clustering spatio-temporal data is an emerging research area. In this paper, dynamic clustering algorithms for spatio-temporal crime data are proposed to detect hot crime spots in Kuwait. Kuwait governorates are taken as case study: the capital, Hawalli, Al-Ahmady, Al-Jahra, Al-Farawaniya, and Mubarak Al-kebeer. In addition, different crime types are considered: act of discharge and humiliation, adultery, aggravated assault, bribery, counter fitting, drugs, embezzlement, fight or resist employee on job, forging of official documents, weapon, robbery and attempted robbery, suicide and attempted suicide, and bank theft. Applying Random subspace classification to those clustered data, 98% accuracy and 99.4% ROC are obtained, having precision (98.7%), recall (98.4%), and F1 (98.28%).

A research study done by Xi C. Chen et al. [2] in their research Clustering has gained widespread use, especially for static data. However, the rapid growth of spatio-temporal data from numerous instruments, such as earth-orbiting satellites, has created a need for spatio-temporal clustering methods to extract and monitor dynamic clusters. Dynamic spatio- temporal clustering faces two major challenges: First, the clusters are dynamic and may change in size, shape, and statistical properties over time. Second, numerous spatio-temporal data are incom- plete, noisy, heterogeneous, and highly variable (over space and time). We propose a new spatio- temporal data mining paradigm, to autonomously identify dynamic spatio-temporal clusters in the presence of noise and missing data. Our proposed approach is more robust than traditional cluster- ing and image segmentation techniques in the case of dynamic patterns, non-stationary, heterogeneity, and missing data. We demonstrate our method's performance on a real-world application of moni- toring in-land water bodies on a global scale.

A research study done by Kamalpreet Kaur Jassar et al. [3] in their research Spatial data mining is a mining knowledge from large amounts of spatial data. Spatial data mining algorithms can be separated into four general categories: clustering and outlier detection, association and co-location method, trend detection and classification. All these methods have been compared according to various attributes. This paper introduces the fundamental concepts of widely known spatial data mining algorithms in a comparative way. It focuses on techniques and their unique features.

A research study done by Fre ́de ́ric Flouvat et al . [4] in their research Co-location mining is a classical problem in spatial pattern mining. Considering a set of boolean spatial features, the goal is to find subsets of features frequently located together. It has wide applications in environmental management, public safety, transporta- tion or tourism. These last years, many algorithms have been proposed to extract frequent co-locations. However, most solutions do a "data-centered knowledge discovery" instead of a "expert-centered knowledge discovery". Successfully providing useful and interpretable patterns to experts is still an open problem. In this setting, we propose a domain-driven co-location mining approach that combines constraint-based mining and cartographic visu- alization. Experts can push new domain constraints into the mining algorithm, resulting in more relevant patterns and more efficient extraction. Then, they can visualize solutions using a new concise and intuitive cartographic visualization of co-locations. Using this original visualization approach, they identify new interesting patterns, and use uninterest- ing ones to define new constraints and refine their analysis. These proposals have been integrated into a prototype based on PostGIS geographic information system. Experiments have been done using a real geological datasets studying soil erosion, and results have been validated by a domain expert.

A research study done by M.Vignesh et al. [5] in their research Spatial data mining, i.e., mining knowledge from large amounts of spatial data, is a highly demanding field because huge amounts of spatial data have been collected in various applications, ranging from remote sensing, to geographical information systems (GIS), computer cartography, environ-mental assessment and planning, etc. The collected data for exceeded human's ability to analyze. Recent studies on data mining have extended the scope of data mining from relational and transactional databases to spatial databases. This paper summarizes recent works on spatial data mining, from spatial data generalization, to spatial data clustering, mining spatial association rules, etc. It shows that spatial data mining is a promising field, with fruitful research results and many challenging issues.

A research study done by Sunita Tiwari et al. [6] in their research the increasing availability of Global Positioning System (GPS) enabled devices has given an opportunity for learning patterns of human behavior from the GPS traces. This paper describes how to extract popular and significant places (locations) by analyzing the GPS traces of multiple users. In contrast to the existing techniques, this approach takes into account the semantic aspects of the places in order to find interesting places in a geo-spatial region. GPS traces of multiple users are used for mining the places which are frequently visited by multiple users. However, the semantic meanings, such as 'historical monument', 'traffic signal', etc can further improve the ranking of popular places. The end result is the ranked list of popular places in a given geo-spatial region. This information can be useful for recommending interesting places to the tourists, planning locations for advertisement hoardings, traffic planning, etc.

A research study done by Ch. Mallikarjuna Rao et al. [7] in their research any geographic location undergoes changes over a period of time. These changes can be observed by naked eye, only if they are huge in number spread over a small area. However, when the changes are small and spread over a large area, it is very difficult to observe or extract the changes. Presently, there are few methods available for tackling these types of problems, such as GRID, DBSCAN etc. However, these existing mechanisms are not adequate for finding an accurate changes or observation which is essential with respect to most important geometrical changes such as deforestations and land grabbing etc.,. This paper proposes new mechanism to solve the above problem. In this proposed method, spatial image changes are compared over a period of time taken by the satellite. Partitioning the satellite image in to grids, employed in the proposed hybrid method, provides finer details of the image which are responsible for improving the precision of clustering compared to whole image manipulation, used in DBSCAN, at a time .The simplicity of DBSCAN explored while processing portioned grid portion.

IV. PROPOSED ALGORITHM

Decision tree
1. t = create Node()
2. label(t) = most common Regions list ID(D, Region array)
3. IF regions find then set flag=true
4.Then make region id is called an parent node and start traverse below
4. FOREACH a ∈ Array∗ DO
Da = {(x, c(x)) ∈ D : x|A∗ = a}
IF Da = ∅ THEN // find the respected region value in tree nodes
sum=sum+path traversing
ELSE
ENDIF
ENDDO
5. return false
K means clustering
Let X = {x1,x2,x3,……..,xn} be the set of Regions cluster center
For each i<x
1) Randomly select *'c'* cluster centers.
2) Calculate the distance between user and cluster centers.
3) Assign the user to the cluster whose distance from the cluster center is minimum of all the cluster centers..
5) Recalculate the distance between each user and new cluster centers if generated.
if i>x
return false

V. RESULTS AND DISCUSSION

1) Purity

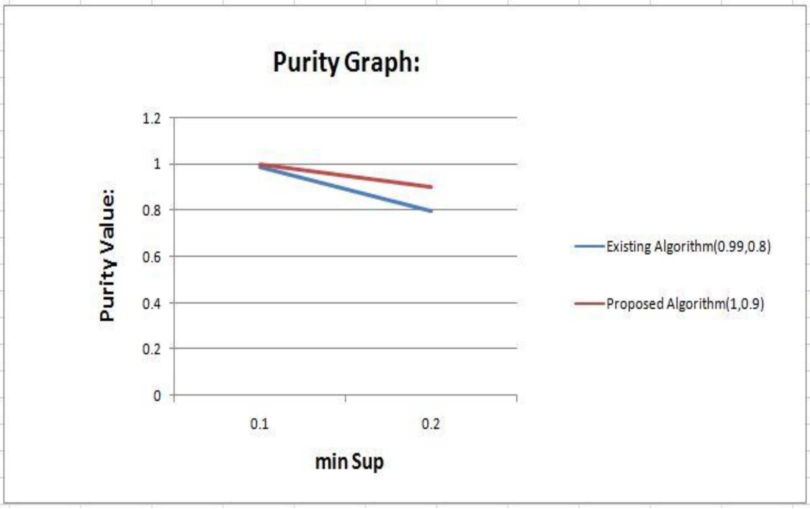
	min sup value=0.1	min sup value=0.2
Existing Algorithm Values	0.99	0.8
Proposed Algorithm Values	1	0.9

Formula purity

Purity is a function of the relative size of the largest class in the resulting communities.

$$\text{PURITY}(T,X) = Entropy(T) - Entropy(T,X)$$

$$G(\text{LAKE,CITCO}) = E(\text{LAKE}) - E(\text{LAKE,CITCO})$$
$$= 0.940 - 0.693 = 0.247$$



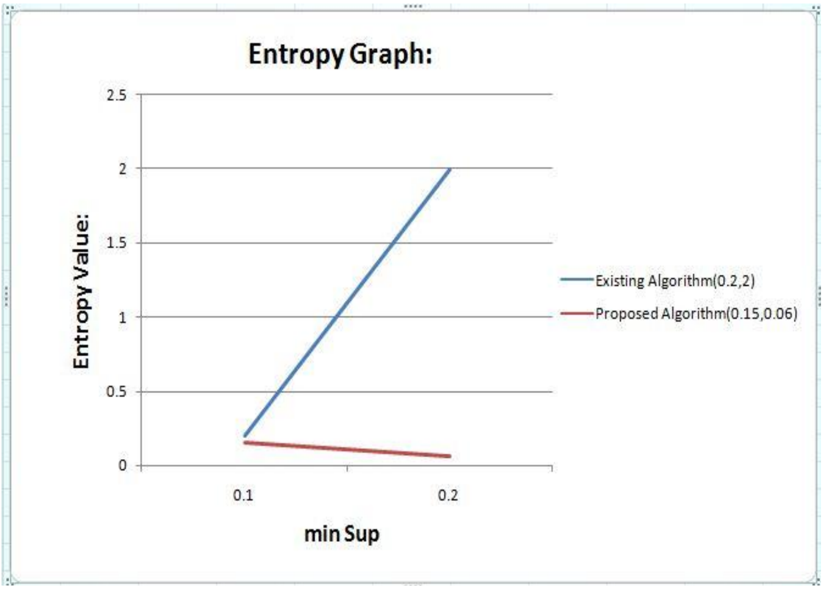
2) Entropy

	min sup value=0.1	min sup value=0.2
Existing Algorithm Values	0.2	0.22
Proposed Algorithm Values	0.15	0.06

Entropy formula

Entropy is a function of the distribution of classes in the resulting communities.

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$



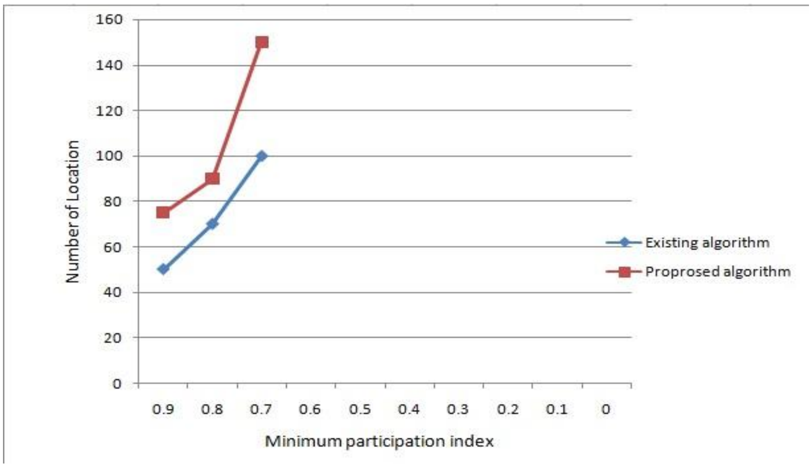
3) Minimum participation value

Minimum participation	No of location in proposed algorithm	No of location in Existing algorithm
0.9	75	50
0.8	90	70
0.7	155	100

Minimum Participation formula

No .of location with respect to minimum participation of user in particular location.

Participation Index=Total visit by users with respect to locations/Total users in database.



VI. CONCLUSION AND FUTURE SCOPE

Spatial Data Mining is the process of discovering interesting and previously unknown, but potentially useful patterns from large spatial data sets. Extracting interesting and useful patterns from spatial data sets is more difficult than extracting the corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial auto correlation. in our research we have developed an technique to find the most visited regions by users. in this research vai gps technique we are allowing the user to see current location on google map and store that location in database. Further we allow the users to share the location with other user using latitude and longitude sending vai google cloud messaging api. There is one limitation in this research if gps is not working then we cannot store the location locally on smart phone so that later it can sync on server database so in future this work will be consider.

REFERENCES

[1] Taysir H. A. Soliman 1,*, Khulood Al Ommar 2, Youssef B. Mahdy 3 " Developing spatio -Temporal Dynamic clustering Algorithm for identifying crime hot spot in kuwait".Information Systems Dept., Faculty of Computers and Information, Assiut University, Egypt

[2] Xi C. Chen, James H. Faghmous, Ankush Khandelwal, Vipin Kumar "Clustering Dynamic Spatio-Temporal Patterns in the Presence of Noise and Missing Data"Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015).

[3] Kamalpreet Kaur Jassar ,Kanwalvir Singh Dhindsa,Ph.D " Comparative Study of Spatial Data Mining Techniques"International Journal of Computer Applications (0975 – 8887) Volume 112 – No 14, February 2015.

[4] Fre´de´ric Flouvat · Jean -Franc¸ois N'guyen Van Soc · Elise Desmier · Nazha Selmaoui -Folcher " Domain-driven co-location mining Extraction, visualization and integration in a GIS " © Springer Science+Business Media New York 2014.

[5] M.Vignesh "Spatial Data Mining: Progress and Challenges " International Journal of Computer Science and Information Technology Research Vol. 2, Issue 3, pp: (1-16), Month: July-September 2014 .

[6] Sunita Tiwari1,3 and Saroj Kaushik2 " Mining Popular Places in a Geo-spatial Region Based on GPS Data Using Semantic Information "Springer-Verlag Berlin Heidelberg 2013 .

[7] Ch. Mallikarjuna Rao1 and Dr.Ananda Rao.Akepogu2 " A HYBRID APPROACH FOR ANALYSIS OF DYNAMIC CHANGES IN SPATIAL DATA " International Journal of Database Management Systems (IJDMS) Vol.6, No.1, February 2014 .