Location-Based Mobile Community Using Ants-Based Cluster Algorithm

Chetneti Srisa-an [1,*]

[1] College of Information Technology and Communication Rangsit University, Thailand
[1] chetneti@rsu.ac.th*;
* corresponding author

(Received April 24, 2021, Accepted April 24, 2021, Available online April 25, 2021)

Abstract

A location based service (LBS) is widely used on modern smartphone around the world asits built-in features. Each smartphone can access a google API or map. People can therefore share their location (latitude and longitude) among friends. Many LBS spots can easily form "location based mobile community (LBMC)." Since the nodes are mobile, the community group changes dynamically and is unstructured. Ant-based clustering algorithm is a special kind of optimization technique which is highly suitable for finding the adaptive clustering for volatile networks. This Paper Aims To form a location based mobile community (LBMC) by using Ant-based clustering algorithm. Due to the mobile type community, a vanishing community problem is also stated in this paper. Instead of redo a whole algorithm again, we modify an original algorithm by applying a pheromone concept to handle a change. Our algorithm is named as ABCA & VP which stands for Ant-Based Clustering Algorithm with Vanishing problem. More than 5,000 samples from their latitude and longitude coordinates in Thailand.From an experiment, K-means clustering work well in small data size and low number of clusters. In Small size of data between 50 and 1000, our algorithm runs battery when a number of clusters reach 15 clusters. In a big data size (between 1,000 and 5,000 samples), our algorithm outperforms K-means clustering when a number of clusters reach 20 clusters.

Keywords: Mobile location based service; WebMining; Ant-based clustering;

1. Introduction

Today, smartphones allow people to easily share their location on the internet. People can easily find out and find out where their friends are. This new capability can form a small virtual community. "Location Based Mobile Community (LBMC)" and "Vanishing Community" are two new terms introduced in this paper because of this phenomenal new case. If the location is in close proximity, it means that they are in the same cluster. Grouping algorithms such as K-mean or Ant-based clustering are good for this type of problem by aggregating a set of data into the same similar group.

Clustering is one of the fundamental examples in data mining. The most famous algorithm in the clustering field is k-means clustering because of its simplicity. It partitions the objects into clusters by minimizing the sum of the squared distances between the objects and the centroid of the clusters. The k-means clustering is simple but it rather static and complex. (ACO) algorithm [11]. Unlike an Ant colony algorithm, an ant-based clustering algorithm does not use a pheromone concept to guide a path to destination. With a dynamic environment, an ant-based clustering algorithm has been accepted as a better way for data clustering problems because it can easily adapt to real world applications such as document management, packaging, networking, etc [12].

Suppose that more than 50% of members in a cluster have disappeared, the status of the cluster becomes unstable. If a density of a cluster is not enough compared to dissimilarity, the rest of the members will belong to a new nearest cluster. This event is called the "vanishing community problem". By modifying the original algorithm, we adopt a pheromone concept to handle a vanishing community problem. We however adopt both concepts in our algorithm. We use an Ant-based clustering algorithm to form a cluster (community) and then a pheromone concept for a vanishing community problem.

The paper is organized as follows: section 2 presents a related work. Section 3 presents our proposed. Section 4 demonstrates experimental results.Section 5 presents a conclusion.

2. Related Works

A.E.F. Hegazy and Omar, W., A. Badr [1] proposed a new hybrid ant-based clustering technique with new modifications to enhance the operations of ants, picking up and dropping of the data items.

Berger, S., H. Lehmann & F. Lehner [2] state that location based service (LBS) have already been introduced to the tourism market. A location based service (LBS) can add significant value to local tourism products and even create new tourism activities.

E.D. Lumer and B. Faieta [3] proposed a new concept by introducing the notion of short-term memory within each agent. Each ant remembers a small number of locations where it has successfully dropped an item. And so, when picking a new item this memory is consulted in order to bias the direction in which the ant will move. Thus, the ant tends to move towards the location it last drops a similar item.

J.-L. Deneubourg and his team [4] introduced the first Ant-based clustering model. The ants would randomly pick n drop one data workspace. item is increased if a data item is surrounded by dissimilar data, or when there is no data in its neighborhood.

Rui Cheng, Zhuo Yang, Feng Xia [7] presented the design and a prototype implementation of the iZone system. A combination of wireless technology, J2ME, LBS and GIS are equipped into a system therefore it is able to locate users and send information based on their locations.

R. Vallikannu and A. George [8] presented an evaluation of an Autonomous Location based Energy Efficient routing Protocol with Ant Colony algorithm for MANETS. A Mobile Ad Hoc Network (MANET) is a network independent of pre-installed infrastructure. Ant Colony Optimization (ACO) is applied for finding the adaptive routing for this network.

Saroj Bala, S. I. Ahson and R. P. Agarwal [9] presented a pheromone based technique that did not need a short term memory and needed not to store the information of all the objects recently met during its walk. The proposed method needs very few parameters and less calculation.

Stefan Berger, Hans Lehmann and Franz Lehner [10] said that LBS are labeled as a killer application in their paper. LBS is useful for tourism because of their potential to add real value for the traveler.

3. Method

3.1. Location Retrieval

From Princeton Survey Research Associates International [5] between April 17 and May 19, 2013, it shows that people use their smartphones to navigate the world. More than 74% of adult smartphone owners say they use their phone to get directions or other information based on their current location. Some 12% of adult smartphone owners say they use a geosocial service to “check in” to certain locations or share their location with friends.

The location of the device can be retrieved by 1) Mobile Phone Service Provider Network-The current cell ID is used to locate the Base Transceiver Station (BTS) that the mobile phone is interacting with and the location of that BTS. 2) The Global Positioning System (GPS) uses a constellation of 24 satellites orbiting the earth. Since the smart phone has an in-built GPS receiver, BY default of the new smart phone, GPS is built-in features. GPS method is therefore chosen because of its easy acquisition and accuracy.

3.2. ISO Standard

A location based service (LBS) became an ISO standard in 2009. From ISO / TC 211Geographic Information / Geomantic, StandardsGuidehandled the LBS standards in the ISO 19132 (Location based services possible standard), 19133 (Location based services tracking and navigation) and 19134 (Multi-modal location based services for routing and navigation) documents. A manufacturer can add location based service (LBS) feature as a built-in service in all smart phones since 2009. On May 10, 2011, Google announced the Google Places API. The Google Places API is a service that returns data about Places. Place response specifies locations as latitude / longitude coordinates.

3.3. A distance between longitudes and latitudes

All recent smartphones can host google applications such as Google Map and Chrome. Beside, google API is a web service application; therefore, it can get longitudes and latitudes easily. From google API, longitudes and latitudes are retrieved. The haversine formula is an equation important in calculating distances between two points on a sphere from their longitudes and latitudes. This uses the 'haversine' formula to calculate the great-circle distance between two points. Formula:

$$d_{lon} = lon_2 - lon_1$$

$$d_{lat} = lat_2 - lat_1$$

$$a = \left(sin\left(\frac{d_{lat}}{2}\right)\right) + cos(lat2) * \left(\sin sin\left(\frac{d_{lon}}{2}\right)\right)$$

$$c = 2 * a \tan tan\, 2\left(\sqrt{a}, \sqrt{(1 - a)}\right)$$

$$dis = R * c\ (where\ R\ is\ the\ radius\ of\ the\ earth)$$

3.4. Ant-based Clustering Algorithm

An ant-based clustering algorithm is implemented to form a cluster. In this paper, each cluster stands for “location based mobile community (LBMC)”. According to this algorithm, a pick up and drop strategy is declared. We adopted the Lumer and Faieta [9] model as follows. The picking probability is defined by:

$$P_p(i) = \left(\frac{k_p}{k_p + f(i)}\right)^2 \qquad (2)$$

$$P_d(i) = 2f(i)\ \ if\ \ f(i) < k_d 1, \quad if\ \ f(i) \geq k_d \qquad (3)$$

$$f(i) = \frac{1}{d^2}\sum_j (1 - d(i,j)|a\ \ if\ \ f > 0)\ 0\ otherwise \qquad (4)$$

Where

- $f(i)\ is\ a\ local\ estimation\ of\ density\ of\ elements\ and\ their\ similarity\ to\ i$
- k_p and k_d is a constant
- $d(i,j)$ measures the dissimilarity between the pair of elements(i,j)
- $\propto$ is a constant that scale a dissimilarities

Assume that the elements (data) represent points in an ann-dimensional space so that d (i, j) is simply the Euclidian distance between I and j. The normalizing term d2 equals the total number of sites in the local area of interest, and introduces a density dependency in$f$ (i). Whenever a loaded ant decides to drop its element, it looks for the first empty site in its vicinity in which to do so. A time step finishes with the selected ant moving to one of its four adjacent nodes, each direction of motion being equally likely.

3.5. Handling a Vanishing Community Problem

Vanishing community is a new issue to address in this paper. Most clustering problems are statistics or relatively steady compared to other areas. In the mobile community, they come and go quite often. A lot of users tend to turn off a location based service when they finish their business. In this case some clusters are impacted by vanishing some members or even all members. The cluster would disappear when it happens. All clusters might need to reformed or merge with others.

Unlike the Ant colony algorithm (ACO), an ant-based clustering algorithm does not rely on a pheromone concept. It is based upon the brood sorting behavior of a real ant. Once all clusters are formed by an ant-based cluster algorithm, data is sorted and kept in a database. From this point, a center of each cluster is calculated. Eachcenter consists of a queen ant and becomes a nest of a cluster. This is only knowledge we learn from the first step. The problem is when a cluster is formed, there is no knowledge about any other clusters. Each cluster does not have any information about any other clusters or communities. When more than 50% of members disappear from its cluster, a queen ant dies. All other ants will find a new nest.

We therefore apply a pheromone concept to handle this issue. Ants are sent out in all 8 directions surrounding a center of a vanishing cluster, those ants who find a next cluster will drop a pheromone on the way back to nest (source). In case of a multiple path, the shortest path has the strongest intensity of pheromone. We can simply know the nearest cluster for each data (member) for each existing data. From those paths, we can form a new cluster easily by merging to a nearest existing cluster. In this case, the rest of ants will sense the strongest concentration of phenome, then they will pick up each member (data) and find a new cluster automatically.

3.6. An Ant-Based Clustering Algorithm with Vanishing problem

Our algorithm is named ABCA & VP which stands for Ant-Based Clustering Algorithm with Vanishing problem. ABCA & VP start with anant-based clustering algorithm presented by Lumer and Faieta [7]. It uses a pick-up or drop strategy to move and sort data [13]. Each ant will pick up data and try to drop it into a new cluster that has a similar characteristics (similar distance). The probability of dropping an item is increased if ants are surrounded with similar distance in the8 neighborhood data [7]. In the same token, each ant will pick up data when it feels dissimilarity among its neighborhood and try to drop into a new area. Eventually, all clusters are formed. Once all clusters are formed, data is sorted and kept in a database. From this point, a center of each cluster is calculated. This is a knowledge we learn from the first step. A queen ant is lay down at the center.

If more than 50% of members in a cluster disappear from a cluster then a queen ant die. The rest of embryos (data) will be adjusted into a new cluster. Instead of redo an ant-cluster algorithm again, we apply a knowledge that learn from the first step. Ants are sent out in all 8 directions surrounding a center of a vanishing cluster, those ants who find a nearest cluster will come back the nest first and deposita pheromone on the way back to nest (source). Therefore; the shortest path have the strongest intensity of pheromone. We can simply know the nearest cluster for each data (member) for each existing data. From those paths, we can form a new cluster easily by merge to a nearest existing cluster. In this case, the rest of ants will sense the strongest concentration of phenome, then they will pick up each embryos (data) and find a new cluster automatically.

An artificial ant starts a route finding process. The first ant that can establish the pheromone track back to its original source node. Other ants follow their pioneer by trailing the path to a destination. The concentration of pheromone becomes stronger when more ants follow. This step is called a path establishment. This is a natural way of wanting to establish a path. If multiple paths are found, the longer path the lower the concentration of pheromone. A pheromone evaporates when time passes due to a longer distance. The path maintenance is done by the evaporation process of pheromone.

In case of new members, our algorithm will scan every 15 minutes; therefore, each round will have all clusters and all data are reset. This way, we do not need to maintain a database. The worst case of a vanishing community is when some clusters vanish while all clusters are not yet formed completely. The only way to solve this problem is to stop a clustering process and redo all processes again from the beginning.

4. Experimental

This program was developed by MATLAB software, Version 7.1 with several samples of latitude and Longitude in Thailandcollected from Google API. A simulation runs from 50up to 5,000 samples of data. K-Mean is used to compare our algorithm. We separate our algorithm into two scenarios as follows: Ant-based clustering algorithm without vanishing problem (ABCA) and with vanishing problem (ABCA & VP). As shown, the runtime is small in K-Mean in a small number of clusters. With a small size of samples between 50and 1000, we found that K-means clustering ran better with a small size of sample as shown. Our algorithm runs when a number of clusters reach 15 clusters in figure 1.

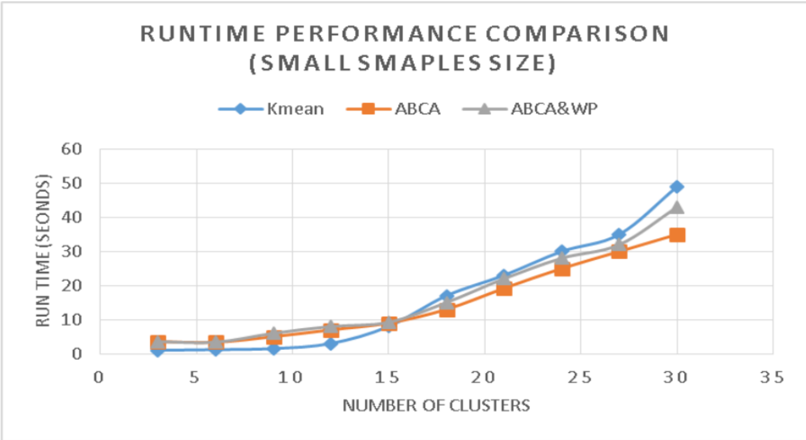

Fig. 1. Performance comparison (small sample size)

Figure 2 shows that our algorithm becomes superior when a number of clusters reach to 20clustersrun between 1,000 and 5,000 samples.

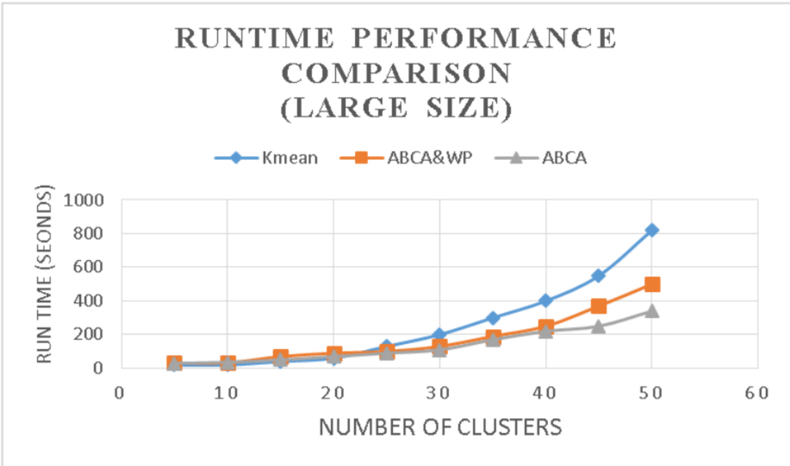

Fig. 2. Performance Comparison (Large sample size)

5. Conclusion

A "location based mobile community (LBMC)" is introduced in this paper. Ant-based clustering algorithm is a special kind of optimization technique which is highly suitable for finding the adaptive clustering for volatile networks. A vanishing community problem is also stated in this paper. Instead of redoing a whole ant-based clustering algorithm again, we modify the algorithm by applying a pheromone concept to handle a change. From 500 up to 10000 samples of the latitude and longitude coordinates in Campus, some clusters are dynamically formed and grouped. We found some interesting tourism areas that form a community. This information is a benefit to the tourism industry.

With small size of data between 50 and 1000, we found that K-means clustering ran better with a small size of sample when a number cluster is less than 15. With small data size, our algorithm runs batter when a number of clusters reach 15 clusters . In a big data size (between 1,000 and 5,000 samples), our algorithm becomes superior when a number of clusters reach 20 clusters.

References

[1] A.E.F. Hegazy and Omar, W., A. Badr, Hybrid Ant-Based Clustering Algorithm with Cluster Analysis Techniques. Journal of Computer Science, 9 (6): 780-793, 2013

[2] Berger, S., H. Lehmann and F. Lehner, Location-based Services in the Tourist Industry. Information Technology & Tourism, 5 (4), 243–256, 2003

[3] E.D. Lumer and B. Faieta, “Diversity and adaptation in populations of clustering ants”, Cambridge: MIT Press, In D. Cliff, P. Husbands, J.-A. Meyer, & S.W. Wilson (Eds.), From animals to animats: Proceedings of the Third International Conference on Simulation of Adaptive Behavior, 1994, pp. 501-508.

[4] DeneubourgJ.L., GrossS., FranksN., Sendova-FranksA., Detrain C. and ChretienL., The dynamics of collective sorting: Robot-like ants and ant-like robots, In Proceedings of the First International Conference on Simulation of Adaptive Behavior : From Animals to Animals, Cambridge, MA, MIT Press, 1991, pp. 356-363.

[5] Kathryn Zickuhr (September 12, 2013), Location-Based Services, http://www.pewinternet.org/2013/09/12/location-based-services/

[6] Messaoud Hameurlaine, Abdelouahab Moussaoui, and Hadda Cherroun, AntMeans: A New Hybrid Algorithm based on Ant Colonies for Complex Data Mining, International Journal of Computer Applications (0975 –8887) Volume 60 – No.17, December 2012

[7] Rui Cheng, Zhuo Yang, Feng Xia, iZone, A Location-Based Mobile Social Networking System, Third International Symposium on Parallel Architectures, Algorithms and Programming (PAAP), P 33 –38, 2010

[8] R. Vallikannu and A. George, Performance analysis of autonomous location based energy efficient ACO routing protocol with dissimilar MANET mobility models, ARPN Journal of Engineering and Applied Sciences, 10 (4), 2015

[9] Saroj Bala, S. I. Ahson and R. P. Agarwal, A Pheromone Based Model for Ant Based Clustering, (IJACSA) International Journal of Advanced Computer Science and Applications, 3 (11), 2012

[10] Stefan Berger, Hans Lehmann and Franz Lehner, Location Based Services In The Tourist Industry, Information Technology and Tourism, 5 (4), 243-56, 2010

[11] Henderi and T. Wahyuningsih, “Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer,” *IJIIS Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, 2021, doi: 10.47738/ijiis.v4i1.73.

[12] M. G. Pradana and H. T. Ha, “Maximizing Strategy Improvement in Mall Customer Segmentation using K-means Clustering,” *J. Appl. Data Sci.*, vol. 2, no. 1, pp. 19–25, 2021.

[13] N. Dzakiyullah, “Semi-Supervised Classification on Credit Card Fraud Detection using AutoEncoders,” *J. Appl. Data Sci.*, vol. 2, no. 1, pp. 1–7, 2021, doi: 10.47738/jads.v2i1.16.