

A Context-aware Member Clustering Algorithm based on ant colony and Genetic optimization for P2P Mobile Social Network

Huaihu CAO

School of Information

Central University Finance and Economics
Beijing, China
caohhu@163.com

Yanmei ZHANG

School of Information

Central University Finance and Economics
Beijing, China
jlzym0309@sina.com

Abstract—It is one of important topics in social network research how to form member clustering according to potential social relations by automatic context-aware the user's behavior feature. This paper presents a context-aware mobile P2P social network framework, member clustering model and algorithm. The user's location information, environmental characteristics etc. are introduced to the clustering algorithm, which intelligently cluster to potential P2P social network. The experimental results show that the proposed approach and the algorithm have a higher response speed, load balance and adaptive ability.

Keywords- mobile social network; context-aware; clustering model; P2P

I. INTRODUCTION

Social Networks is originated from making friend online, the users organize and maintain the current social relationship and find new social relations by the Internet, which shows himself or herself and makes the social activities as a social person on the network [1, 2]. With the popularization of portable mobile devices such as smart phones, tablet, and laptop etc. and the development of mobile communication technology [3], the mobile wireless network that has highly distributed autonomous and topology dynamic characteristics is becoming the foundation of Social Network, the traditional Social Network relying on Web and providing centralized service is difficult to adapt to the current dynamic heterogeneous distributed mobile network environment, also can't satisfy the new demand such as independent found and shared data etc. by P2P. At the same time, people are no longer satisfied with simple dating activities and hope to be able to actively and intelligently to find potential social relationship according to the user's environment, personal interests, behavior characteristics and scene information, and then get more and more valuable entertainment, communication and cooperation opportunities [4]. Therefore, research how to actively structure P2P social network using the context-aware information, which have great significance for the further development of the Social Network.

In the real social interaction process, the user can collect the environmental information from mutual contact and informal communication by the face-to-face interaction. But in a social network environment, does not provide

environmental perception information, or is very limited, so lead to environment perception level in social network is generally low, and lost a lot of opportunities for interaction, cooperation and participation. At present, mobile devices will have more and more sensing function, including video, voice, GPS, gravity induction, acceleration induction, six axial gyro, infrared induction, so as to obtain the context-aware of node is no longer difficult and expensive. Therefore, this article will introduce more perception information, such as GPS, track, physical environment, interest preference, etc., to discuss Context-aware Member Clustering Algorithm for P2P Mobile Social Network.

II. THE HIERARCHICAL OVERLAY TOPOLOGY STRUCTURE FOR MOBILE P2P SOCIAL NETWORK

Because there are dynamic and heterogeneous characteristic on the ability and behavior of nodes for mobile network [5], we propose a hierarchical overlay topology structure for mobile P2P social network, it is illustrated in Figure 1. In this structure with layers, considering dynamic and heterogeneous nature of nodes, they were given to different responsibility according difference of node. Firstly, these nodes were clustered, then, M nodes were elected as main super node and back-up super node in every one clustering, which were given to more responsibility, for example, the function managing and defending mobile P2P social network. The different route algorithms were adopted in the inner of clustering, and the super node will accomplish the similar function with core gateway. Thereby the flexibility of the mobile P2P social network will be improved.

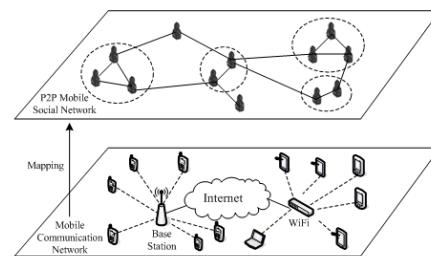


Figure 1. The hierarchical overlay topology structure

All kinds of mobile equipment access to the mobile social network in different ways, such as cell phone users can access to Internet through the 3G base station, Pad,

Tablet, notebook computer may be more is access to Internet through the WiFi, they formed a underlayer mixed communication network, and at the same time, should also pay attention to the mobile devices with bluetooth wireless technology can also short distances share data each other. The mobile device through the operation license client software at the bottom of mixed communication Network based on the scene perception, independent clustering to form a potential virtual P2P Overlay Social Network, on the basis of Social networks, the further active social found service can be realize, referring users to meet the requirements of the Social relations.

Because member clustering algorithm is very important for realizing mobile social network, therefore, we will of node will model the clustering problems in the next chapter, and put forward a feasible optimization algorithm.

III. MOBILE P2P SOCIAL NETWORK MEMBER CLUSTERING MODEL

According to above section mentioned, mobile P2P social network first automatically cluster to a potential overlay social networks according to the user context-aware information. If using graph theory to represent the problem of Mobile Node Clustering, then a hybrid network can be expressed as undirected graph $G = (V, E, W_G)$, the vertex set V denotes Mobile Node set, the side set E denotes the link set of mobile nodes, $W_G(u, v)$ ($u, v \in V$) denotes Mobile Node between u and v clustering vector distance (degree of similarity), Mobile Node clustering problem is dividing mobile nodes into different subgraph $g_1(v_1, e_1, w_1)$, $g_2(v_2, e_2, w_2)$, ..., $g_M(v_M, e_M, w_M)$ in the vector distance space (M denotes gathered number, v_i , e_i , w_i denote subgraph nodes, links between the nodes, and the vector distance set, $i = 1, \dots, M$), and satisfy the following conditions:

- a) $v_1 \cap v_2 \cap \dots \cap v_K = \emptyset$;
 - b) $v_1 \cup v_2 \cup \dots \cup v_K = V$;
 - c) $w_i(u, v) < W_G(u, v')$,
- $(u \in g_i, v \in g_i, v' \notin g_i), i=1, \dots, M.$ (1)

Clustering method is based on the similarity calculation of each mobile node, and it only depends on the characteristic matrix of nodes that the nodes were organized into many clusters. Clustering method can be described as: given N nodes $X = \{x_1, x_2, \dots, x_N\}$ with L dimension space, clustering N nodes into M subset of node C_1, C_2, \dots, C_M (clustering center are respectively y_1, y_2, \dots, y_M), made nodes in a cluster are similar as far as possible each other, usually $C_i \cap C_j = \emptyset$ ($i \neq j$), and $\bigcup_{i=1}^M C_i = X$.

The optimization model of node clustering problem can be reduced to the following minimization problem:

$$\text{MinD}(X, Y)$$

$$= \text{Min}(\sum_{i=1}^N \sum_{j=1}^M (\alpha \sqrt{\sum_{l=1}^L (x_{il} - y_{jl})^2} + \beta \frac{x_i^T y_j}{\|x_i\| \|y_j\|} + \gamma \frac{x_i^T y_j}{\|x_i\|^2 + \|y_j\|^2 - x_i^T y_j})^2) \quad (2)$$

Where $X_i = \{x_{i1}, x_{i2}, \dots, x_{iL}\}$ represents characteristic vector of node x_i (with L components), N is the total number of nodes, $Y_i = \{y_{j1}, y_{j2}, \dots, y_{jL}\}$ is the feature vector of clustering center node j (with L components), M is the total number of clustering center node, $\sqrt{\sum_{l=1}^L (x_{il} - y_{jl})^2}$ denotes environment similarity degree

between the node x_i and node y_j , $\frac{x_i^T y_j}{\|x_i\| \|y_j\|}$ denotes trajectory similarity degree, $\frac{x_i^T y_j}{\|x_i\|^2 + \|y_j\|^2 - x_i^T y_j}$ denotes history preference similarity degree. α, β, γ denotes the weight factor, which can be adjusted according to the actual need. Because of formula (2) is a non convex optimization problem, and there is no effective way to find the global optimal solution, the next section we will try to find a feasible algorithm.

IV. A CONTEXT-AWARE MEMBER CLUSTERING ALGORITHM BASED ON ANT COLONY AND GENETIC OPTIMIZATION

A. Flowchart of algorithm

Genetic algorithm has the ability of doing a global searching quickly and stochastically, But it can't make use of enough system output information. It has to do a large redundancy repeat for the result when solving to certain scope [9]. So the efficiency to solve precision result is reduced. Ant algorithm converges on the optimization process through information pheromone accumulation and renewal. It has the ability of parallel processing and global searching. The speed at which the ant algorithm gives the solution is slow, because there is little information pheromone on the path early. Hence, we present a Context-aware Member Clustering Algorithm based on Ant Colony and Genetic Optimization (ACGO), which integrates genetic algorithm and ant colony algorithm. First, it adopts genetic algorithm to distribute information pheromone. Second, it makes use of the ant colony to give the precision of the solution. Finally, it develops enough advantage of the two algorithms. Figure2. shows the complete ACGO algorithm flowchart.

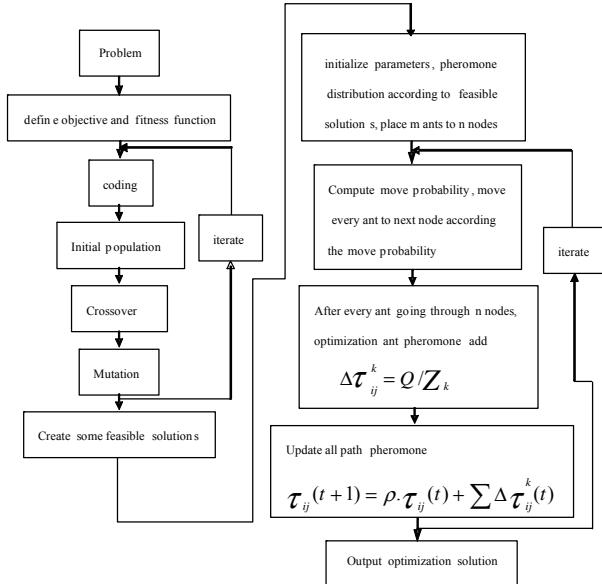


Figure 2. Flowchart of ACGO algorithm

B. The pseudocode of ACGO algorithm

```

1. Let  $\text{Ini}(V_i) = \text{average}\{w_i, q\}$ ;
2. while not all nodes clustered do
    {
        3. compute the  $\text{MinD}(X, Y)$  of all nodes;
        4. call Genetic Algorithm that minimize  $\text{MinD}(X, Y)$ 
        5. mark  $X_i$  clustered, adjust the ready node set;
    }
6. endwhile with not all nodes clustered
7. Initialize subgradient step and penalties  $\lambda_i$ ,  $i \in V$ 
8. Initialize trails  $[\tau_{ij}]$  and solutions  $X_k = \emptyset$ ,  $k = 1 \dots n$ 
9. repeat
10.   repeat
11.     for ants  $k=1$  to  $m$ 
12.       do Choose the next arc  $(ij) \in E$  to append to  $X_k$ 
13.        $X_k = X_k \cup X_j$ 
14.     until (all nodes have their cluster)
15.   Update trails
16. until (Termination condition)
    
```

The choice of the next arc to append, at step 12, is obviously made in probability. The formula (3) used at step 12 and at step 15 are those proposed in [6], that is, each k -th ant moves from its current state l to a feasible state ϕ in probability, according to the formula:

$$p_{l\phi}^k = \frac{\alpha \cdot \tau_{l\phi} + (1-\alpha) \cdot \eta_{l\phi}}{\sum_{(lv) \in tabuk} (\alpha \cdot \tau_{lv} + (1-\alpha) \cdot \eta_{lv})} \quad (3)$$

Whereas trails are updated by means of formula (4):

$$\tau_{l\phi}(t) = \tau_{l\phi}(t-1) + \tau_{l\phi}(0) \cdot (1 - \frac{Z_{curr} - LB}{\bar{Z} - LB}) \quad (4)$$

Where z_{curr} is the cost of each current solution, \bar{Z} is the average of the last computed solutions and LB is a lower bound to the optimal problem solution cost. Parameters to tune are m , number of ants, α , relative importance of visibility w.r.t. trail (multiplicative coefficient), and the termination condition, in our case maximum CPU time allowed. The cost average to use in the trail update formula at step 8 was made over the costs of all solutions computed at each iteration of loop 3-7. For the ants, parameters were $\alpha = 0.5$ and $num.ants = n/10$.

V. THE RESULTS OF SIMULATION EXPERIMENT

ACGO algorithm proposed by this paper is a kind of distributed parallel algorithm, run on the mobile device, because energy and computing ability of mobile devices is limited generally, so the costs and clustering speed and message number of communication of the algorithm is very important for the system performance. Therefore, we conducted large-scale simulation experiments to test its performance. The simulation experiment mainly was realized in the network topology generator BRITE of Java source version and discrete event driven simulation package Simjava based on the Java language, regardless the transmission delay, congestion, packet loss and details of the actual process of network.

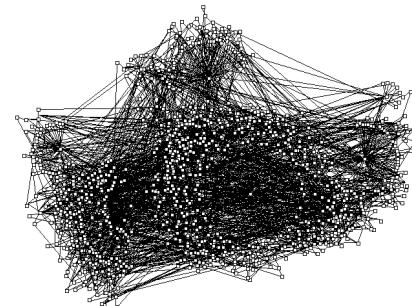


Figure 3. The clustering result of k-means algorithm

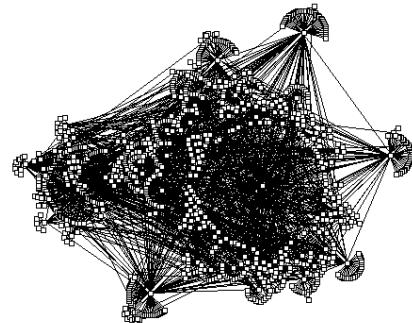


Figure 4. The clustering result of ACGO algorithm

The physical topology of nodes is generated by the BRITE (Boston University Representative Internet Topology Generator). BRITE is a general topology generator developed by Boston university, which realized the Waxman, BA, BA - 2, GLP topology generation algorithm and so on,

and produce network topology in accordance with the AS (autonomous domain) level and router level, with the top down and bottom up two methods to produce Transit - Stub hierarchical model. For as far as possible close to the reality of mobile hybrid network topology structure, physical network topology in this paper simulation experiment is generated according to GLP BRITE in router level, and random injecting location, temperature, light, trajectory and sensory information.

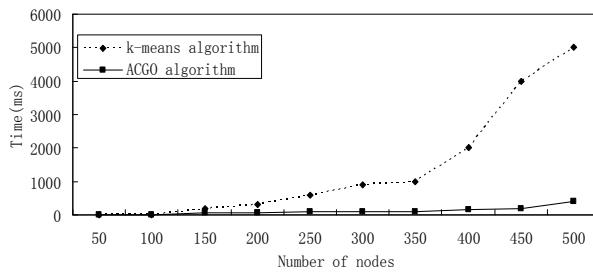


Figure 5. ACGO and k - means algorithm performance comparison

In the simulation system running the traditional k - means clustering algorithm and this paper proposed clustering ACGO algorithm, the results of clustering topology are shown in figure 8 and figure 9 respectively, the result of ACGO algorithm has appropriate scale and more close to the reality social network. The running time of algorithm is shown in Figure5. we can see that with the increase of the number of the nodes the cost time of ACGO algorithm growth gentle, has the good scalability, and k - means algorithm is present exponential growth trend. The total message number of ACGO algorithm is shown in Figure 6. Because ACGO algorithm is a kind of distributed parallel algorithm, and its number of interaction message presents a fast growing tendency, how to reduce the total amount of the news of interaction message will be the next step need to study problems.

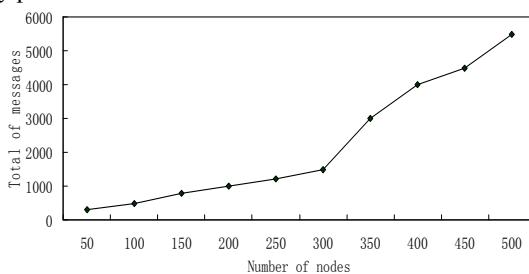


Figure 6. Chang of message number with the number of nodes

VI. CONCLUSION AND FUTURE WORK

This paper introduces the user's location information, environmental characteristics, such as trajectory into mobile social network, puts forward the scene perception mobile P2P social network topology structure, clustering model and algorithm, intelligently clustering to potential P2P social networks. And the algorithm was tested by experiments, and

the results show that the scheme and algorithm presented by this paper have high response speed, and at the same time, has certain load balance and adaptive ability.

But the experiment system is only a prototype system, the next step the practical application testing will be made, the communication between overlay P2P social network users is Flood protocol, this protocol is unstructured, the total amount of message is too much, if it is appropriate for the large scale application also need to be further tested, so as to improve the use efficiency of channel. In addition, privacy control and credit management is a important problem for mobile P2P social network, which also need to further research.

ACKNOWLEDGMENT

This work was partially supported by the National Natural Science Foundation of China (60970143), the Key Project of Chinese Ministry of Education (No109016), Supported by Program for Innovation Research in Central University of Finance and Economics, MOE (Ministry of Education in China) Project of Humanities and Social Sciences(Project No.11YJCZH006), 211 Project for Central University of Finance and Economics (the 3rd phase) and Discipline Construction Foundation of Central University of Finance and Economics.

REFERENCES

- [1] Cranshaw J, Toch E, Hong J, et al. Bridging the gap between physical location and online social networks. Proceedings of the 12th ACM international conference on Ubiquitous computing. New York, USA, 2010, pp.119-128.
- [2] Kimmerle J, Cress U. Group awareness and self presentation in computer-supported information ex- change. International Journal of Computer-Supported Collaborative Learning, 2008, 3(1), pp.85-97.
- [3] Tam Van Nguyen, Wontaek Lim, Huy Anh Nguyen et al. Context awareness framework based on contextual graph. Proceedings of the 5th International Conference on Wireless and Optical Communications Networks. Surabaya, East Java Indonesia, 2008, pp.1-5.
- [4] Buder J, Bodemer D. Supporting controversial CSCL discussions with augmented group awareness tools. International Journal of Computer-Supported Collaborative Lmning, 2008, 3(2), pp.123-139.
- [5] Schmidt A, Gellersen H-w. Visitor Awareness in the Web. Proceedings of the 10th International Conference on World Wide Web. Hong Kong, China, 2001, pp.745-753.
- [6] Acquisti A, GrossR. Imagined communities: Awareness, information sharing, and privacy on the Facebook. Privacy Enhancing Technologies. 2006, 3(1), pp.36-58.
- [7] D. Chopra, H. Schulzrinne, E. Marocco, E. Ivov. Peer-to-peer overlays for real-time communication: security issues and solutions, IEEE Journal on Communications Surveys & Tutorials, 2009, 11(3), pp.4-12.
- [8] Xin Yan Zhang, Qian Zhang, Zhengsheng Zhang, et al. A Construction of Locality-Aware Overlay Network: mOverlay and Its Performance]. IEEE Journal on Selected Areas in Communications, 2004, 22 (1), pp. 18-28.