Analyzing Behavior of the Social Media Users Through Swarm Intelligence Perspective

Lin Bi, Yang Li ,Xiao-qiang Di
School of Computer Science and Technology
Changchun University of Science and Technology
Changchun, China
bilin7080@163.com

Ying Zhang
Information Center
Changchun Institute of Optics, Fine Mechanics and Physics,
Chinese Academy of Sciences
Changchun, China
starsugar_zy@163.com

***Abstract*—Swarm intelligence is defined as the properties of artificial systems. It is suited to depict people's daily behavior, such as social media users' behavior. Social media users have some characteristics. For one thing, users who have the same interest will focus on the same VIP (Very Important Person) users inside the industry. For another, the users concentrating on the same VIP user may focus on each other. It is formed by common interests as the cluster center. Based on the characteristics, we analyse the Sina Weibo as the research object. From the perspective of swarm intelligence, the multi-agent mechanism model has been established. In this paper, we implemented the user behavior model and studied the following aspects. Firstly, we clarify how the same interests can lead to the clustering of the users, and what the path of the clustering evolution would be. Secondly, we measure the influence of VIP users in the industry. In order to solve the above problems, we use the computational experiment method to simulate clustering process via the K-PSO algorithm (K-Particle Swarm Optimization). We select the RePast as the simulation experiment tool. The simulation results are highly consistent with the expected experimental results.**

Keywords—swarm intelligence; K-PSO algorithm; social media; behavior analyzing

I. INTRODUCTION

Social media is interactive, authentic, real-time and has the characteristic of transmission diversity. As one of the social networking applications, it gives users a great deal of participation and interactive space. It is a combination of communication, interaction and sharing, and makes each user become the recipient and the publisher through the social network platform. Social media has realized that the audience no longer just passively browsed the Internet content, but also involved in the manufacture, and it provides the corresponding services which are initiative, interactive, innovative and instructive. It has become the real core of the Internet. Every user can be independent for content creation, and he or she can exchange, share, and fully discuss topics of common interest. Social media is developing rapidly, and the ROI (Return On Investment) of social media will exist for a long time.

Social media users have such characteristics-For one thing, users who have the same interest, will focus on the same VIP users inside the industry. For another, the users concentrating on the same VIP users may focus on each other. It is formed by common interests as the cluster center. For the purpose of this kind of clustering, it has brought clear business promotion, and made the network economy more prosperous. The formation of the cluster is a dynamic process, from the attention towards each other to the production of active users. It also makes it easier to attract more attentions to VIP users. The clustering is an expression of swarm intelligence algorithm which was first presented in 1992 by Beni and Hackwood. Swarm intelligence algorithm can well express the self-organizing characteristics of complex systems, and the individual abstracts as a point in the model. The interests of the individual and collaborative goal become the fitness of individuals in the environment, in order to measure the abilities of the behavior of the individuals in the cluster and the model of individual learning, individual imitation, and the characteristics of evolution. In this paper, we use the K-PSO (K-Particle Swarm Optimization), such a species of intelligent algorithm, taking Sina Weibo as the research object, to describe and analyse the formation process of clustering in detail.

Swarm intelligence is defined as the properties of artificial systems. The design of artificial systems is aroused by the task effectiveness appearing in several natural groups. The main feature of the swarm intelligence refers to the collective behavior. It derives from the interactions among the individuals, such as the interactions between the individuals and the environment [1,2]. The swarming model determines the behavior of the system. Several swarming models have been presented in various fields. Specially, swarm intelligence is widely used in the fields of optimization [3,4], control and robotics [5–7], graphics [8] and computer networking [9,10].

It is crucial for understanding and modeling user behavior to build reliable and meaningful testing [11]. Nevertheless, we need to narrow the gap between techniques used in practice and suggested in theory [12].In the past few years, modeling and simulating user behavior has been studied[13].In recent years, the Internet has increasingly expanded from a technical level to commercial and social level. As a result, mass online social media have grown up. Based on these platforms, the interaction between users has been enhanced, such as emotional exchange, information diffusion, and commercial transaction [14].

II. APPLICATION OF K-PSO ALGORITHM IN THE SWARM INTELLIGENCE ISSUES

We proposed the clustering process of based on the K-PSO algorithm in this paper.

PSO (Particle Swarm Optimization) algorithm was invented by Eberhart and Kennedy. The basic PSO algorithm is developed by simulating the flock foraging behavior of random search algorithm based on group collaboration. It is thought to be a kind of swarm intelligence. It can be incorporated into the multi-agent system optimization.

PSO simulated birds feeding behavior. A flock of birds search food randomly, and all the birds don't know where the food is. The most simple and effective method is searching for the bird which is near the food in the surrounding area. Each solution of optimization problem in PSO is a bird in the search space. We call it particle. All particles have a fitness value. It is determined by the function of optimization of adaptive value. Each particle has a speed to decide the direction and distance next time. Then the particle to follow the current optimum particles can be searched in the search space. PSO initialization can realize a group of random particles (random solutions), and then find out the optimal solution, through the iteration of tracking two extreme values. The first value is the optimal solution of particles, and it is called individual extreme value *pBest*. Another extreme value is the entire optimal solution in the space, and this value is global extreme value *gBest*.

The agent in swarm intelligence algorithm and the particle in PSO algorithm are equal to the treatment in this paper. The reason is that the agent and particle have the same characteristics of autonomy, intelligence, and they can change their positions according to the environment. Through competition and cooperation between the agents, and combined with the iterative mechanism of PSO algorithm, the agent can update the position in the solution space.

We defined the position of the i_{th} agent in the space as $x_i=(x_{i1},x_{i2}\cdots x_{iN})$, where, $i=1,2,\cdots, n$. The speed of the agent is denoted as $v_i=(v_{i1},v_{i2}\cdots v_{iN})$, it is also vector of N-dimensional. For the i_{th} agent, the searched optimal position is defined as $p_i=(p_{i1},p_{i2}\cdots p_{iN})$. For the entire particle swarm, the searched optimal position is defined as $p_g=(p_{g1},p_{g2}\cdots p_{gN})$ .The agent update formula can be seen as follows [15]:

$$v_{id}(t+1)=\omega v_{id}(t)+c_1r_1(p_{id}-x_{id}(t))+c_2r_2(p_{gd}-x_{id}(t)) \quad (1)$$

$$x_{id}(t+1)=x_{id}(t)+v_{id}(t+1) \quad (2)$$

where, $i = 1, 2,\cdots$; C_1, C_2 represent the nonnegative constants respectively; r_1, r_2 denote the random number respectively, obeying uniform distribution in the region [0, 1]; $x_{id}(t)$ is the position of the i_{th} agent; p_{id} is the searched optimal location of the i_{th} agent. p_{gd} is the optimal position of the entire particle swarm; v_{id} is the current velocity of the i_{th} agent, $v_{id}\in[-V_{max}, V_{max}]$.

In this paper, in order to achieve clustering the industry users in the solution space, we use the K-PSO algorithm to optimize the process of clustering based on adaptive function. This algorithm is an unsupervised machine learning method to solve the clustering optimization.

We use the similarity difference to take user subject as many particles in the solution space so as to make all particles form K-clustering cluster, and select a cluster center from each clustering cluster. Then we use PSO algorithm to achieve clustering optimization. Through many testing results, it is indicated that the method is of higher efficiency and practicability, compared to the traditional K-means algorithm.

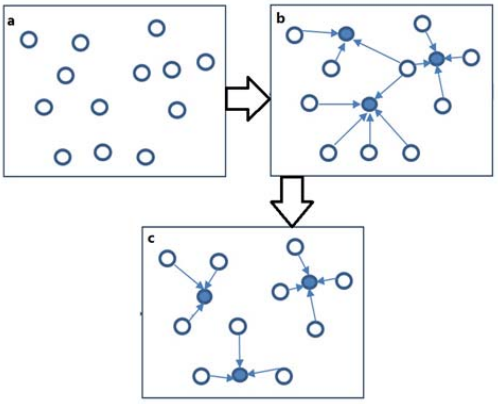

Fig.1 Schematic diagram of the K-PSO algorithm.(a. Disorder state of particles; b. K-clusterings; c. Searching for optimization dynamically by the PSO algorithm)

The update of velocities value of the particles depends on the change of the position. When the particles find the optimal value, they will be close to particle who has the optimal value at the fastest speed. In the K-PSO algorithm, it has not only one clustering center, which means that the particles in accordance with their own characteristics must be close to the clustering center. Otherwise, the velocity value of particle remains the same.

III. MODELING METHOD OF THE USER BEHAVIOR

A. Description of the Experimental Model

The definition of social media users is the subject which has the active behavior. Through the microscopic effects, it can export macro-effect by the top-down model. Between the subjects, the interaction between the subject and the environment and cooperation is the basic cause of system evolution. In order to achieve the operational purpose of or repeatable operation, it must carry on the abstract to the real problem, to establish a concept model. Taking the experiment of Sina Weibo as the research object, the Weibo users are independent and intelligent individuals. Subject in its own interest is continually to interact with other individuals, in order to access to information exchange and more attention. With the effect of information sharing, users can take a common interest hobby into a interest clustering. This effect is the evolution of the individual self interest and gives full play to the cluster result.

B. Modeling Method

Each user has the learning and adaptive ability, so we define them as individual agent. Every agent aims to seek more and more attention, so they are constantly focusing on other agents. Of course, their focus is their interest industry. Different users can be divided into VIP users, active users and common users. Common users expect to become active users, and the active users and VIP users expect to have more fans.

Assumptions of this experiment are as follows:

1. We define the agent number as n, and as for different times each agent can be regarded as a result of the interaction $x_i(t)$.

2. The influence of the growth rate under certain environmental resource constraint is defined as P_i.

3. $q_{i,j}$ is the cooperation effect of user j to the user i that can interact with each other. The influence of growth on the scale is as shown in the following formula:

$$\frac{dx_i(t)}{dt} = p_i(t)x_i(t)\left(1 - x_i(t) + \sum q_{ij}x_i(t)\right) \quad (3)$$

4. For user subject module, we define a two-tuple m=$\{A_i, A_j\}$, which can be seen as the most basic relationship. A_i and A_j in the two-tuple represent the different agents. When the agent A_i focuses on the agent A_j, the two-tuple m can represent the relationship. And each agent will focus on many other agents, of course, most of them have a common interest, or colleagues, friends and so on. Thus, in the midst of the whole space, it formed a reticular structure. In order to calculate and express it conveniently, each agent can be expressed as multi-dimensional particles based on the two-tuple.

The agent can be expressed as multi-dimensional vector:

$$A = \{m_{i1}, m_{i2}, \ldots\ldots, m_{in}\} \quad (4)$$

For the algorithm, it is important to choose suitable adaptation function, and it can be expressed as the following function:

$$F(x_i(t)) = p_i(t)x_i(t)\sum q_{ij}\frac{x_i(t)}{K} \quad (5)$$

where, F is fitness value function; T is on behalf of the cycle, K is a constant.

The relationship between two agents could be expressed by the function related to the distance, and the distance measured by common users, so $q_{i,j}$ can be expressed as the following formula:

$$q_{ij} = h \bullet d_{ij} \bullet e_{ij} \quad (6)$$

where, h represents the coefficient of impact factor, d_{ij} represents the distance between the two agents, e_{ij} represents that if the two agents focus on each other.

In the process of modeling, F is fitness value function to evaluate the correlation degree. The larger the F is, the better the clustering effect is.

We set optimal value of *pBest* as below:

$$pBest = \{d_{p_1}, d_{p_2}, \ldots, d_{p_n}\} \quad (7)$$

Every agent's global optimal value in the space can be defined as below:

$$gBest = \{d_1, d_2, \ldots, d_m\} \quad (8)$$

For each two-tuple, we should set a speed value v_{id} in this model, and represent it as the growth speed of fans. Value of vector v_{id} can be set as formula 1.

As a result, the v_{id} in he formula 1 of each agent can be expressed as below:

$$Vm_i = \{Vm_{i1}, Vm_{i2}, \ldots, Vm_{in}\} \quad (9)$$

IV. EXPERIMENTAL

A. Experimental Procedure

To build a multi-agent model, we use the simulation tool RePast (Recursive Porous Agent Simulation) to realize the user behavior model. It builds a model from bottom to up, and all kinds of agent model based on the Java language environment. We can make individual interaction events by RePast. In the experiment, agents number is defined as 1000, and the experimental cycle number is 400.

Through the analysis of the user behavior modeling method, we defined the social media users as the agents. In the improved strategies of particle swarm optimization algorithm, every agent is seen as a particle. The clustering process will be elaborated here.

Implementation steps:

Ⅰ. The size of the agent swarm can be regarded as a constant M;

The change of random initialization displacement for each particle can happen as $V_1, V_2, \cdots, V_s$, which can comprise $V(t)$.

Ⅱ. As for the k agent swarm, individual particles number for each particle swarm can be expressed as $D = M/k$.

Ⅲ. The optimization objective function is set as F;

Each module is distributed into the nearest clustering center C_i (I = 1, 2,..., k) as function F;

Ⅳ. According to the current position and velocity, we can calculate the searched optimal location $pBest_i$ of each agent so far. According to the position of each agent in the particle swarm C_i, we can calculate the optimal agent location $gBest_j$ so far in the C_i.

Ⅴ. Agents position: when agents move constantly in the process of clustering, if fitness value function F is less than an experience value, we can assume that the agents are already in the right clustering, which means that the agents can no longer move again in the module. At this time, we can set $V_i = 0$. Else, the agents will continue to look for the clustering center.

Ⅵ. We update variables v_i and p_i of the rest agents in every swarm particle;

Ⅶ. Repeat process (Ⅲ)-(Ⅵ) until it reaches the maximum iteration.

Ⅷ. After each process of the iteration, the agents can adjust $x_i(t)$ in the space, the new location will be determined by *pBest* and *gBest*.

Ⅸ. The experimental output is as follows:

We will get k classification results in the space, each user classification has common interests, that is, the users who are close to each other in the same classification. It is defined as the k agent clustering.

B. Experimental Results

1. In order to show that adding as friends to each other is contributed to the individual influence in the social media, Fig.2(a) reflects that focusing on the other users randomly can contribute to the clustering center formation. Fig. 2(b) reflects that the common interest as the dominant concern contributes to the influence on each other. So Fig.2 shows that the domination of common interests can expand the influence of the subjects.

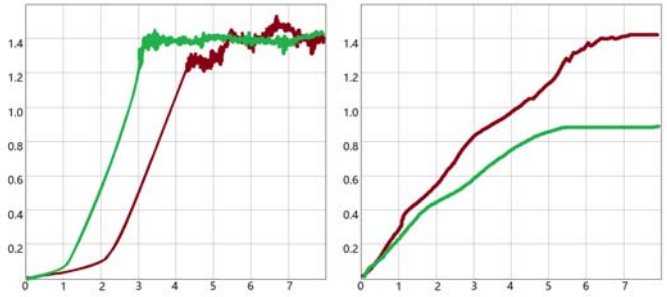

Fig.2 Schematic diagram of the influence (a. Focusing on the other user randomly; b. Common interest as the dominant concern contribution to influence)

2. In the experiments, the subject's influence can be randomly distributed, and the subject constantly interacts with the other users according to interests. In Fig.3, circle represents the common user, and square represents the active user, and triangle represents the VIP users. It can be seen from the Fig.3, in the evolution process, there are a large number of active users and several VIP users in origin, and evenly distribution is relative. There is no obvious stratification to be seen. With the experiment, the influence of each user can be differentiated, and numerous common users keep small influence. VIP users keep becoming the center of the cluster. On the one hand, expanding influence in the industry will get more attention. On the other hand, keeping active communication in the industry is also a necessary way to expand the influence.

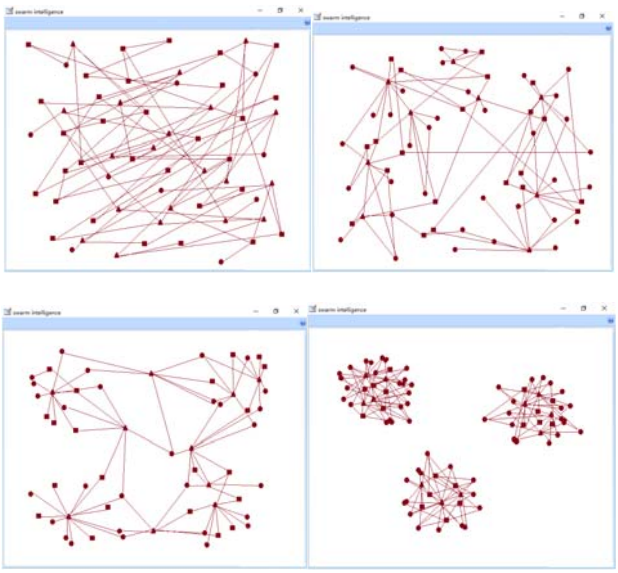

Fig.3 Schematic diagram of clustering

V. CONCLUSION

Social media is interactive, authentic, real-time and has the characteristic of transmission diversity. As one of the social networking applications, it gives users a great deal of participation and interactive space. Every user can be independent for content creation, and he or she can exchange and share, and fully discuss topics of common interest.

Social media users have some characteristics that users who have the same interest will focus on the same VIP (Very Important Person) users inside the industry. In addition, the users concentrating on the same VIP user may focus on each other. It is formed by common interests as the cluster center. For the purpose of this kind of clustering, it has brought clear business promotion, and made the network economy more prosperous. Based on the characteristics of social media, we analysed Sina Weibo as the research object. From the perspective of swarm intelligence, the multi-agent mechanism model has been established. In this paper, we implemented the user behavior model and studied the following aspects: Firstly, how can the same interests lead to clustering of the users, and what is the path of the clustering evolution. Secondly, it judges the influence of VIP users in the industry. In order to solve the above problems, we used the method of computational experiment, and simulated the clustering process through the K-PSO (K-Particle Swarm Optimization) algorithm. We selected the RePast as the simulation experiment tool. The simulation results are highly consistent with the expected experimental results.

ACKNOWLEDGMENT

This paper is supported by Science and Technology Research Project of the Education Department of Jilin Province for the 13th Five Year Plan of China（No.2016-379）.

REFERENCES

[1] J.F.Kennedy, R.C.Eberhart, “Swarm Intelligence, Morgan Kaufmann, London”, 2001.

[2] E.Bonabeau, M.Dorigo, G.Theraulaz,“Swarm Intelligence: From Natural to Artificial Systems”, Oxford: Oxford University Press, Inc., NewYork, NY,USA,1999.

[3] M.Dorigo,C.Blum, “Antcolony optimization theory: asurvey”, Theor. Comput. Sci., 2005,vol.344, pp.243–278.

[4] J.Kennedy, R.Eberhart, “Particle swarm optimization”, in: Proceedings of the IEEE International Conference on Neural Networks, 1995,vol.4, pp.1942–1948.

[5] R. Olfati-Saber, R.M.Murray, “Consensus problems in networks of agents with switching topology and time-delays”, IEEE Trans. Autom. Control, 2004,vol.49, pp.1520–1533.

[6] V.Gazi, K.M.Passino, “Stability analysis of social for aging swarms”, IEEE Trans. Syst. Man Cybern. PartB: Cybern.,2004,vol.34, pp.539–557.

[7] E.Şahin,“Swarm robotics: From sources of inspiration to domains of application, in: Swarm Robotics”, Springer, Heidelberg, 2005, pp.10–20.

[8] C.W.Reynolds, “Flocks, herdsand schools: adistributed behavioral model”, in: ACMSIGGRAPH Computer Graphics, ACM, NewYork, 1987, vol.21, pp.25–34.

[9] F.Dressler, O.B.Akan, “Asurvey on bio-inspired networking”, Comput. Netw. 2010,vol.54, pp.881–900.

[10] M.Saleem, G.A.DiCaro, M.Farooq, “Swarm intelligence based routing protocol for wireless sensor networks: survey and future directions”, Inf. Sci., 2011,vol.181, pp.4597–4624.

[11] D.G. Feitelson, “Work load modeling for computer systems performance evaluation”, 2015

[12] U.Schwiegelshohn,“How to design a job scheduling algorithm”.In JSSPP, 2014.

[13] S.Schlagkamp, “Influence of dynamic think times on parallel job scheduler performances in generative simulations”,InJSSPP, Hyderabad, India, 2015.

[14] L.David, P. Alex, A.Lada, Sinan, A, “Life in the network: the coming age of computational social science”,Science, 2009, vol.323, pp.721-723.

[15] A.G. Li, Z.Tan, F.M.Bao,“Particle swarm optimization algorithm”, Computer Engineering and Applications, 2002,vol.38, pp.1-3.