Intelligence level analysis for crowd networks based on business entropy

Zhouxia Li
Beijing Jiaotong University, Beijing, China
Zhiwen Pan
Institute of Computing Technology Chinese Academy of Sciences, Beijing, China
Xiaoni Wang
Beijing Jiaotong University, Beijing, China
Wen Ji
Department of Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology Chinese Academy of Sciences, Beijing, China, and
Feng Yang
Beijing Jiaotong University, Beijing, China

Received 28 May 2019
Accepted 1 September 2019

Abstract

Purpose – Intelligence level of a crowd network is defined as the expected reward of the network when completing the latest tasks (e.g. last N tasks). The purpose of this paper is to improve the intelligence level of a crowd network by optimizing the profession distribution of the crowd network.

Design/methodology/approach – Based on the concept of information entropy, this paper introduces the concept of business entropy and puts forward several factors affecting business entropy to analyze the relationship between the intelligence level and the profession distribution of the crowd network. This paper introduced Profession Distribution Deviation and Subject Interaction Pattern as the two factors which affect business entropy. By quantifying and combining the two factors, a Multi-Factor Business Entropy Quantitative (MFBEQ) model is proposed to calculate the business entropy of a crowd network. Finally, the differential evolution model and k-means clustering are applied to crowd intelligence network, and the species distribution of intelligent subjects is found, so as to achieve quantitative analysis of business entropy.

Findings – By establishing the MFBEQ model, this paper found that when the profession distribution of a crowd network is deviate less to the expected distribution, the intelligence level of a crowd network will be higher. Moreover, when subjects within the crowd network interact with each other more actively, the intelligence level of a crowd network becomes higher.

This work is supported by the National Key R&D Program of China (2017YFB1400100) and the National Basic Research Program of China under Grant No. 61671049, 61301002.

International Journal of Crowd Science
Vol. 3 No. 3, 2019
pp. 249-266
Emerald Publishing Limited
2398-7294
DOI 10.1108/IJCS-05-2019-0014

Originality/value – This paper aims to build the MFBEQ model according to factors that are related to business entropy and then uses the model to evaluate the intelligence level of a number of crowd networks.

Keywords Differential evolution, Business entropy, Crowd intelligence, Network profession distribution, Subject interaction patterns

Paper type Research paper

1. Introduction

Crowd networks where intelligent subjects are collaborated, interacted for a variety kinds of tasks, are widely existed in both human society and nature. In human society, crowd networks are pervasive in the field such as test automation (Bird and Munoz, 1983), customer preference analysis (Xiao *et al.*, 2016.), supply chain, collaborative decision-making (Chen *et al.*, 2017; Decker and Wattenhofer, 2013), currency network and so on. The more sufficiently multiple subjects work together, the easier our mission will be. Therefore, it is of great significance to study the collaborative effect of multiple intelligent subjects for optimizing this crowd network. In this paper, we mainly discuss some factors that affect the level of crowd network. Then we can update the network through these factors to make the network more powerful.

As a crowd network might be engaged in a variety kinds of tasks and these tasks are assigned to subjects within the network (Neubauer and Fink, 2009), the distribution of professional category of the crowd subjects have a great effect on crowd network's intelligence level. Imitating the concept of information entropy (Paninski, 2003), this paper will introduce business entropy to analyze crowd network's intelligence level so that the correlation between professional category of the crowd subjects and crowd network's intelligence level can be described. To be specific, according to the equation of business entropy, it can be found that when the category of crowd subject is 1, the value of business entropy is 0, which exactly corresponds to a single intelligent subject. However when the professional category of crowd subject obeys uniform distribution, the value is the maximum. In addition to the professional category, factors such as the intelligence level of crowd subjects (Nelson and O'Connel, 1978) and the amount of information shared among subjects (He *et al.*, 2006) also affect the crowd network's intelligence level. Hence, when describing the correlation between professional category of the crowd subjects and crowd network's intelligence level, we will also take other factors into consideration.

Section 2 introduces the basic concept of crowd intelligence network and relevant work. Section 3 introduces the concept of business entropy and its influencing factors. Section 4 establishes a multi-factor business entropy quantization (MFBEQ) model based on the influencing factors. In Section 5, the influencing factors and models presented in Section 4 are used to analyze 3 specific networks in detail. In Section 6, k-means clustering based on differential evolution is proposed to cluster intelligent subjects in a crowd intelligence network, so as to obtain the business entropy of a certain crowd intelligence network. Finally, it is concluded in Section 7.

2. The related work and background

2.1 Crowd network

The research object of crowd intelligence science and engineering is crowd intelligence network. The crowd intelligence network is a chaos field which connects the physical world information world consciousness world through the internet.

A particular crowd intelligence network can accomplish many different tasks. Liu *et al.* (2018b) pointed out that the intelligent subject within the crowd network can take collaborative actions to accomplish the task and gives feedbacks to the environment. Upon receiving the accomplishment feedback, the environment can quantified the quality of task accomplishment through a numerical value named reward and send back the crowd network. The illustration of such operational mechanism is shown in Figure 1.

In the Figure 1, first, the environment constantly assigns different tasks to the crowd network. Second, as the task can be refined into a number of sub-tasks, each intelligent subject in the crowd network accomplishes the corresponding sub-tasks through collaborating with other subject and relying on their own professional skill. Finally, the quality of these tasks can be quantified according to the rewards given by the environment. Intelligence level of a crowd network is defined as the expected reward of the network when completing the latest tasks (e.g. last N tasks). As the objective of our approach is to improve the intelligence level of a crowd network, this objective can be interpreted as increasing the expected rewards of the target crowd network for completing tasks in the long run.

According to the different functions of crowd intelligence network system, the basic problems of crowd intelligence science can be divided into crowd decision and crowd cooperation. Crowd decision problem refers to one or more collective results produced by the interaction and influence of many intelligent individuals; crowd cooperation problem refers to the group behavior that is superior to individual problem solving ability generated by many intelligent individuals through such mechanisms as competition and cooperation, differentiation and integration, feedback and selection, with clear goals, tasks and resource constraints.

2.2 Related work

The interaction pattern between intelligent subjects is one of the factors which can describe the distribution of professional categories of intelligent subjects. In Jennings (1995), it is proposed that the degree of mutual influence between intelligent subjects has a certain impact on the collaborative pattern of the whole network. The second factor is network distribution, that is, the number distribution of professional categories of intelligent subjects, which will affect network stability (Grefen *et al.*, 2000). The more stable the network is, the more static the business entropy will be, so the equilibrium (dominance) will

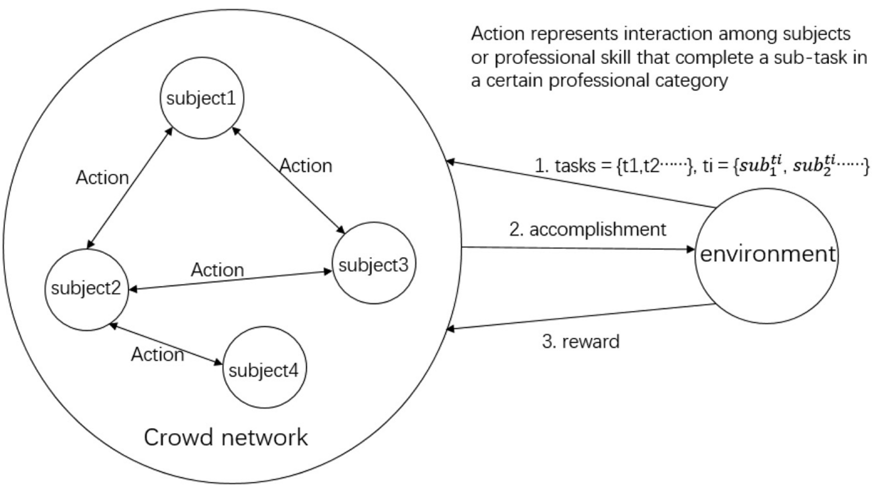

Figure 1.
The operational mechanism of crowd network

not decrease (increase). In Herrera-Viedma *et al.* (2005), authors pointed out that different intelligent subjects have different mentality, leading to decision-making preferences. For different preferences, the variety of intelligent subjects in the crowd network is also different. Therefore, the last factor is task types engaged by the crowd network.

The interaction among subjects is important for the efficiency of the whole network. These interactions can make subjects work collaboratively so that tasks can be fulfilled in a better fashion (Jennings, 1995). In this way, the intelligence level of the whole crowd network is increased. In Chapman *et al.* (2003), authors also proved that the positive collaborative effect is beneficial to the intelligence level of crowd networks such as logistics service supply chain, ant colony behavior and bit coin network. The network scale of crowd network can be evaluated by its stability. Grefen *et al.* (2000) proved that the more stable the network is, its business entropy will not change too much, so that the evolution speed of the network is relatively small. In Sinha (2010), Herrera-Viedma *et al.* (2005), crowd collaboration, crowd evolution and crowd decision-making problem types are explained in detail. According to these properties of crowd network, authors point out that task types also have impact on the operational efficiency of a crowd network (Sinha, 2010; Herrera-Viedma *et al.*, 2005). In Ren *et al.* (2013), it explicitly defines the measurement method of node importance, in which the clustering coefficient and degree of nodes are taken as the measurement of node importance.

The logistics service supply chain is a functional network structure mode formed by logistics service providers, logistics service integrator, manufacturers and retail enterprises (Persson and Virum, 2001). To a certain extent, this network model can describe the role played by each crowd subject within the crowd network. As to the network analysis of ant colony behavior, Colorni *et al.* (1991) proposed an evolutionary ant colony algorithm. The proposed ant colony algorithm can describe the collaboration mechanism of ants, as well as the intelligence level of the ant crowd network.

In Yuan and Wang (2016), the working mechanisms of the block chain are proposed. Based on these previous works, we carry out a comprehensive analysis of the block chain network. This paper will make a detailed case analysis of bit-coin network based on block chain network.

Through performing qualitative analysis on the business entropy we quantify the factors which may affect the business entropy according to the density and aggregation coefficient of the network. These quantitative factors we obtained can be used as characteristics to describe the business entropy information of the network. Next, the k-means clustering algorithm based on differential evolution proposed by Omran *et al.* (2005) is imitated to cluster a specific crowd intelligence network. After finding the category of a crowd network, the value of the equilibrium degree and the dominance degree of the business entropy can be obtained. Based on the degree of business entropy, the relationship between the business entropy, the professional category of the intelligent subject and the crowd network can be described quantitatively.

3. Business entropy and its influencing factors

3.1 The concept of business entropy

The definition of entropy in thermodynamics represents one of the states of the material parameters, whose physical meaning is a degree of chaos system. In this paper, we introduce the business entropy and professional category of intelligent subject to quantify the intelligence level of crowd network.

The professional category of intelligent subject refers to the intelligence level of an intelligent subject when completing a task in a crowd network (Lee and Kramer, 1992). In other word, we evaluate its intelligence level status by comparing its intelligence level with the similar products. Suppose the system has N intelligent subjects with M professional categories, the business entropy is defined as:

$$P_d = -\sum_{i=1}^{i=K} \frac{M_i}{N} * \ln\left(\frac{M_i}{N}\right), \tag{1}$$

where M_i is the number of intelligent subjects in the i-type professional types, which is considered as the relationship between crowd network's level and business entropy. Let the individual intelligence level of intelligent subject be I, I ∈ [0,1], and the crowd intelligence level formed by the cooperation of N intelligent subjects should be the sum of the intelligence levels of each subject. The crowd network's level and business entropy are abbreviated as CI and P_d, respectively.

According to equation (1) it can be seen that when the number of intelligent subjects with each professional category are the same, the business entropy reaches the maximum, corresponding to crowd network. On the contrary, when there is only one intelligent subject, the minimum value of business entropy is 0. Hence, the more evenly the professional categories of intelligent subjects are distributed in the crowd network, the higher the efficiency of intelligent cooperation in completing a task will be. Therefore, assuming that other conditions remain not changed, the intelligence level of crowd network is positively correlated with the professional category of the intelligent subject.

3.2 Factors affecting business entropy

The factors that affect the business entropy of a crowd network are listed in below:

3.2.1 Profession distribution deviation. Network profession distribution (NPD) is the distribution of profession category of intelligent subjects within the crowd network. As the profession categories of intelligent subjects are the major parameters of both network profession distribution and business entropy, the network profession distribution and business entropy are inherently correlated with each other. Such correlation can be described as the deviation between the current NPD and the optimized NPD. Suppose the total number of intelligent subjects in a crowd network is N, the number of profession category within the crowd network is K. Typically, for a crowd network whose objective is to be capable of fulfilling all kinds of tasks (instead of fulfilling specific kinds of tasks), its optimized NPD should be uniform distribution. In such case, the optimal NPD is expressed as 1/K of each professional category.

When the optimized NPD is uniformly distributed, a scatter diagram describing the correlation between deviation and business entropy is shown in Figure 2. When N = 15 and K = 3 in the network, the scatter diagram obtained is shown in Figure 2. It can be seen from the figure that the deviation and business entropy are linearly negatively correlated.

However, not all networks' NPD are uniformly distributed. For example, in ant colony networks, the number of worker ants is much larger than that of queens. This paper introduces the concept of expectation distribution (optimized NPD): A certain network can complete different kinds of tasks, and the optimized NPD for each particular task can be different. Hence, in this paper, we defined the optimized NPD of a crowd network as the expectation value of optimized NPD for the most recent T tasks that are given to the target network. The optimized NPD (OPD) is a one-dimension array and every item within the array represents the optimized OPD of a specific professional category.

3.2.2 Subject interaction pattern. A crowd network consists of not only subjects but also connections among these subjects which carries the interaction among subjects. This paper introduces subject interaction pattern to describe these edges in a crowd network.

Multi-subject coordination refers to the reasonable arrangement of resources among multiple subjects to maximize the overall intelligence level of the crowd network (Jennings, 1995). Therefore, the interaction among intelligent subjects actually reflects how the limited resources obtained by the crowd network are arranged. For example, in the logistics supply chain, the coordination of various elements within the system enables the resources to be fully used, system operating efficiency to be improved and the overall cost to be saved. This is the positively collaborative effect, and vice versa is the negatively collaborative effect and even no collaborative effect.

In a crowd network, subjects can be added in or removed from the network to improve the network efficiency, which also cause the increase of business entropy (He *et al.*, 2007). Moreover, the subjects with higher importance (e.g. higher intelligence level) are likely to be more active than others; hence, the subjects with higher importance may generated more interactions. Based on the above ideas, this paper describes the importance of intelligent subject by introducing the interactive pattern into the original business entropy equation, so that a more accurate business entropy quantization model is established.

4. Multi-factor business entropy quantitative model

In this section, the two factors including Subject Interaction Pattern and Profession Distribution Deviation are introduced into the business entropy model as parameters to make it more generalized and accurate.

4.1 Subject interaction pattern

Inspired by the work in Ren *et al.* (2013), the importance of a certain category is measured by the interaction among subjects. To describe the SIP of subjects within each profession category accurately, both the interactions between intelligent subjects of the same category and the interactions between intelligent subjects of different categories are considered. Consequently, importance of a professional category i can be measured as:

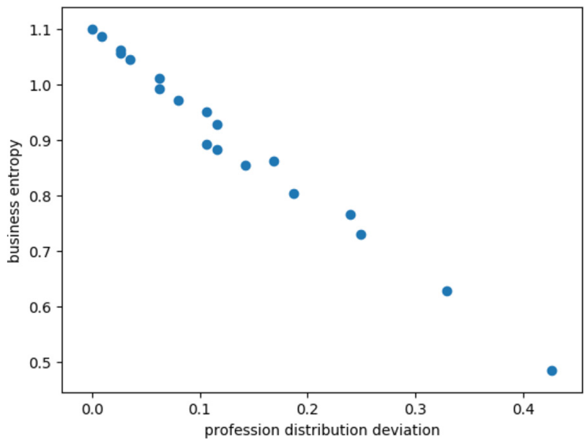

Figure 2.
Relationship between profession distribution deviation and business entropy by taking the expected distribution as uniform distribution

$$I_i = q * \frac{interaction\ with\ other\ classes}{total\ interaction\ among\ subjects\ in\ class\ i} + (1-q) * \frac{interaction\ within\ class\ i}{total\ interaction\ among\ subjects\ in\ class\ i}, \quad (2)$$

where q is a weight number whose value is between 0 and 1.

Based on the SIP of subjects within each profession category, SIP of each specific subject can be calculated by determining if the subject is constantly active (He *et al.*, 2007). Similarly, the active level of a subject can be quantified based on both the interactions between intelligent subjects of the same category and the interactions between intelligent subjects of different categories are considered. By referring to the definition of I_i, active level of SIP_{ij} can be quantified based on both the number of interactions among the target intelligent subject and intelligent subjects of other profession category and number of interactions among the target subject and intelligent subjects of the same category. Thus, the SIP of intelligent subject j whose profession category is i can be defined as:

$$SIP_{ij} = \sum_{i' \neq i} I_i' * \frac{interaction_{i'j}}{total\ interaction\ of\ subject\ j} + I_i * \frac{interaction_{ij}}{total\ interaction\ of\ subject\ j}, \quad (3)$$

where $interation_{i'j}$ represents the interaction between subject j and all subjects in the category i', $interation_{ij}$ represents the interaction between subject j and all the other subjects in the category i, and total interaction means the sum of $interation_{i'j}$ and $interation_{ij}$.

4.2 Profession distribution deviation

Suppose the number of intelligent subjects in a crowd intelligence network is K, the total number of intelligent subjects is N, and the number of intelligent subjects in profession category i is Mi, i =1, 2, 3, 4. . .. . . K. The deviations is defined as:

$$PDD = \sum_{i=1}^{i=K} \left(\frac{M_i}{N} - OPD_i \right)^2. \quad (4)$$

OPD_i represents the corresponding optimal NPD value of professional category i in the network.

4.3 The proposed multi-factor business entropy quantization model

By integrating the two factors together based on the concept of information entropy, our proposed MFBEQ model can be established as:

$$P_d = (-PDD) - \sum_{i=1}^{i=K} \sum_{j=1}^{j=Mi} SIP_{ij} * \ln(SIP_{ij}), \quad (5)$$

where SIP_{ij} represents Subject Interaction Pattern of the subject j in the professional category i. PDD represents the Profession Distribution Deviation of the entire crowd network. P_d is the business entropy of the target crowd network.

To conclude, the implementation of the MFBEQ model can be formalized as follows:

- *Step 1:* determine the optimized NPD by referring to the data or its definition in Section 3.2, and the hyper-parameter q involved in the model.
- *Step 2:* calculate the importance of all professional categories according to equation (2).
- *Step 3:* calculate the SIP of all the subjects according to equation (3).
- *Step 4:* calculate the PDD of the target network according to equation (4).
- *Step 5:* calculate the business entropy of the target network according to equation (5).

5. Case studies of analyzing crowd intelligence based on business entropy

According to the types of basic problems solved by crowd intelligence science and the functions of crowd intelligence network system, the basic problems of crowd intelligence science can be divided into crowd decision and crowd cooperation.

Crowd intelligence science and engineering is an interdisciplinary direction of natural science and social science, related to many research fields such as sociology, psychology, management, economics and so on. All the three networks can be regarded as three typical application scenarios of crowd intelligence. To be specific, the ant colony network belongs to the crowd intelligence network in nature, and the supply chain and block chain belong to the crowd intelligence network in human sociology. Besides, the cooperating patterns of these three networks are typical. The supply chain and ant colony network have very clear division of labor, so it is easy to distinguish the types of their intelligent subjects. The distinction of intelligent subjects in block chain is not obvious, since the subjects within the network cooperate or compete with others through doing the same labor. This paper analyzes the crowd intelligence based on specific crowd intelligence networks in this section.

When it comes to the model substitution, this section uses network topology among subjects as interaction. So SIP value can be achieved by using topology easily. And the detailed computing processes are shown in the following sections.

5.1 Collaborative operation of logistics service supply chain

Logistics service supply chain is a functional network structure mode connected by functional logistics service providers, logistics service integrators, manufacturing and retail enterprises (Persson and Virum, 2001).

Figure 3 is the schematic diagram of the logistics service supply chain network. H represents the distributor, W represents the temporary storage of the finished product and D is the distribution center. The figure illustrates the relationship among suppliers, manufacturers and distributors. These connections between logistics supply chain can describe flow of production, distribution and consumption. As logistics supply chain involves the coordinated operation of different industries and distributors, it is obvious that good coordination mechanism should be established between these companies. The mechanism should be carried by frequent information transmission and interaction, otherwise it will lead to cost waste and profit reduction. At present, the scale of logistics service supply chain can be said to be unprecedented which indicates that the scale of logistics service supply chain is constantly expanding, which may include more intelligent and powerful subjects.

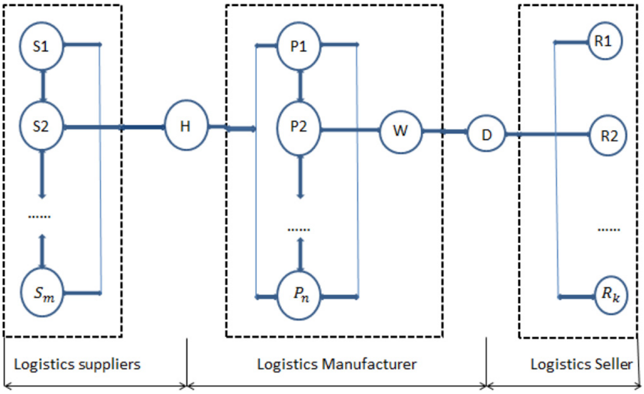

Figure 3.
Collaborative circulation process of logistics service supply chain

The interaction among subjects is of great significance to this supply chain, and the positively collaborative effect can reduce costs and increase benefits. From the overall large network, it is mainly suppliers, manufacturers and sellers.

The supplier timely informs the manufacturer of the supply of parts, and the manufacturer needs to inform the seller of the production situation. The seller feedbacks the sales situation and the demand situation to the manufacturer, and the producer also feedbacks the corresponding demand of original parts to the supplier. Only through this constant interaction and communication can the entire network operate effectively.

In this paper, a simple supply chain network is established as shown in the figure below. MFBEQ model is used for analysis, as shown in Figure 4:

The optimized NPD of this supply chain network: $OPD_{seller} = 0.6$, $OPD_{producer} = 0.2$ and $OPD_{supplier} = 0.2$; balance factor q: 2/3. In the example, this paper uses network topology as interaction among subjects and the computing process is as following:

The original network model in Figure 4(a) can be calculated as:

According to equation (2) and determined parameters, it can be obtained that $I_{seller} = I_{produer} = I_{supplier} = 2/3$.

According to equation (3), all subjects' activity level are calculated as:

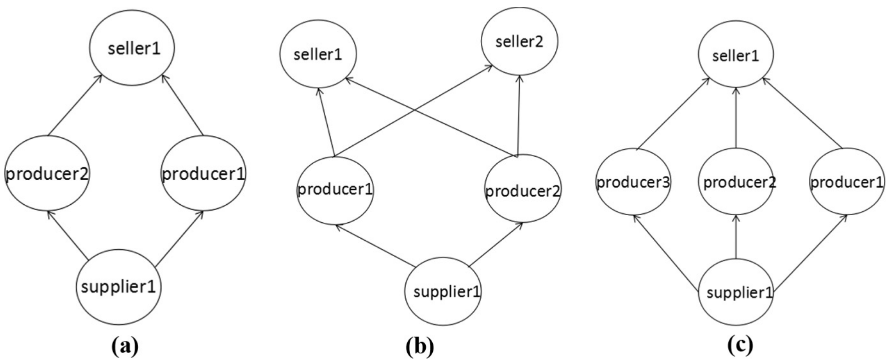

Figure 4.
Three supply chain diagrams

$$SIP_{seller1} = 1 * I_{producer} = 2/3, \ SIP_{producer1} = 1/2 * I_{seller} + 1/2 * I_{supplier} = 2/3,$$

$$SIP_{producer2} = 2/3, \ SIP_{supplier1} = 1 * I_{producer} = 2/3.$$

By substituting these values into equation (4), the value of *PDD* can be calculated as 0.1275.

Finally based on equation (5), the business entropy P_d can be calculated as0.9537.

Similarly, we implement the model on the crowd network in Figure 4(b):

$$I_{seller} = I_{produer} = I_{supplier} = 2/3,$$

$$SIP_{seller1} = 2/3, \ SIP_{seller2} = 2/3, \ SIP_{producer1} = 2/3 * I_{seller} + 1/3 * I_{supplier1} = 2/3,$$

$$SIP_{producer1} = SIP_{producer2} = SIP_{supplier1} = 2/3.$$

$$PDD = 0.08, \ P_d = 1.2716$$

Finally, we implement the model on the crowd network in Figure 4(c):

$$I_{seller} = I_{produer} = I_{supplier} = 2/3,$$

$$SIP_{seller1} = 1 * I_{producer} = 2/3, \ SIP_{producer1} = 1/2 * I_{seller} + 1/2 * I_{supplier} = 2/3,$$

$$SIP_{producer1} = SIP_{producer2} = SIP_{producer3} = 2/3, \ SIP_{supplier1} = 1 * I_{produer} = 2/3.$$

$$PDD = 0.32, \ P_d = 1.0316.$$

Because there are no interaction within the category, according to the simple calculation above, $I_{seller} = I_{produer} = I_{supplier} = q$. The three network differences are mainly in the model of *PDD*. A new seller is added to the original network [Figure 4(b)] and NPD of the network in the figure is closest to optimized NPD, so the business entropy value of this network is the maximum among the three networks.

5.2 Behavior analysis of ant society

In nature, there are a lot of animals can do thing that human beings are unable to perform, such as the bee nest. The bees build the nest structure particularly not only beautiful but also strong. Besides the bees' nest, the process of ants foraging can also confirm this. Single ants may not have any intelligence, but some complex tasks, such as foraging behavior, migration and cleaning the nest, can be completed through collaboration. The network of these animals is a prerequisite for their ability to complete the task. Ants release pheromones to realize information transmission and carry out their interactions (Colorni *et al.*, 1991). So there's a direct interaction among ants, and that interaction helps them distinguish different kinds of tasks.

Figure 5 is a figure describing ants foraging, this paper uses "#" to replace the ants. In the process of foraging, when there are no obstacles, the ants look for food as shown in Figure 5(a); when some ants are separated by obstacles, as shown in Figure 5(b); over time, more ants choose the path closer to the food, as shown in Figure 5(c). In the end, all the ants are on the shortest path from the food, as shown in Figure 5(d). They send out the pheromones for "communication", so more ants will choose shorter path to feed. Therefore, the degree of information interaction has a great impact on the network.

The queen's main job is to lay eggs, reproduce and manage the family members. The male's main job is to mate with the queen. Workers build and expand nests, gather food and feed larvae and queens. Soldiers can crush hard food and become weapons of war when defending colonies. For a network of ants, changes in the size of the network are closely related to reproduction and mortality. For the completion of a specific task, the allocation of task are assigned to ants based on based on their professional categories.

In the network shown in the Figure 6(a), there are four worker ants, one queen ant and two males. After removing a worker ant, Figure 6(a) is changed as Figure 6(b).

By setting q as 2/3, and the optimized NPD of this ant network: OPD_{worker} = 7/10, OPD_{queen} = 1/10, OPD_{male} = 2/10. As the example in ant network, the network topology is used to represent interaction among subjects (SIP) are regarded as interaction. Through implementing the MFBEQ model:

$$I_{worker} = 5/9,\ I_{queen} = 2/3,\ I_{male} = 2/3.$$

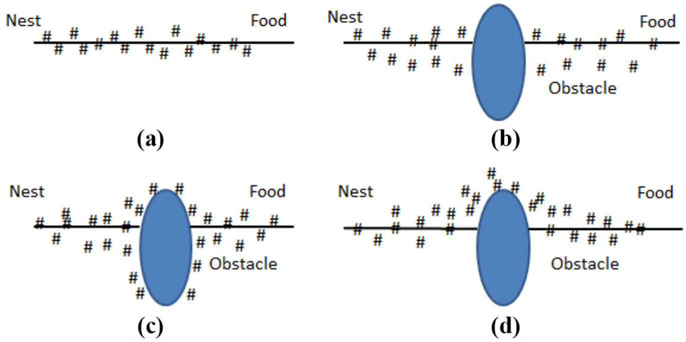

Figure 5. Foraging diagram of ants

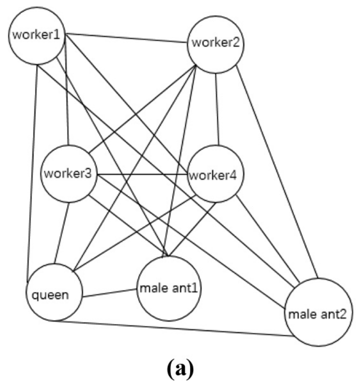

Figure 6. Two ant colony diagrams

Calculate the activity level of each intelligent subject according to equation (3):

$$SIP_{worker1} = 1/6 * I_{queen} + 1/3 * I_{male} + 1/2 * I_{worker} = 11/18, SIP_{worker1} = SIP_{worker2} = SIP_{worker3} = SIP_{worker4}$$

$$SIP_{queen} = 4/6 * I_{worker} + 2/6 * I_{queen} = 16/27.$$

$$SIP_{male1} = 4/5 * I_{worker} + 1/5 * \ = 26/45, SIP_{male1} = SIP_{male2}$$

Calculate the PDD according to equation (4):

$$\text{PDD} = 0.0257,$$

$$\text{So } Pd = 1.8051.$$

If removing a male ant, as is shown in Figure 6(b), a new crowd network can be established where $I_{worker} = 11/21$, $I_{queen} = 2/3$, $I_{male} = 2/3$. The MFBEQ model can be implemented as:
Calculate the activity level of each intelligent subject according to equation (3):

$$SIP_{worker1} = 3/5 * I_{worker} + 1/5 * I_{queen} + 1/5 * I_{male} = 61/105, \ SIP_{worker1} = SIP_{worker2} = SIP_{worker3} = SIP_{worker4}$$

$$SIP_{queen} = 4/5 * I_{worker} + 1/5 * I_{male} = 58/105, \ SIP_{male1} = 4/5 * I_{worker} + 1/5 * I_{queen} = 58/105,$$

Calculate the PDD according to equation (4):

$$PDD = 0.0067$$

$$\text{So } P_d = 1.911.$$

The result shows that: after removing a worker ant, PDD decreases and the second item value in the model increases, and the business entropy of the target network is increased.

5.3 Block chain

A typical example of block chain is the use of bit-coin, which is increasingly dominant as a new form of electronic currency payment transaction. In Figure 7, there is always an "authoritative" node S in the centralized network, while in the P2P network, the status of each role is equal. The bit-coin network consists of personal computers, servers, mining machines designed to mine bit-coin, mobile terminals and other subjects. Through broadcast mechanism to for information trading, information interaction in a block chain network is obviously a very full. Based on this mechanism, the system can decide whether there will be a transaction record into the block, whether it includes the new node into the bit-coin network. Therefore, broadcast mechanism is of great significance to make a decision for each subject.

As more people use bit-coin, the scale of this network can be said to be constantly expanding. The more nodes the whole network contains, the stronger the robustness of this network will be. Specifically, unless more than half of the nodes have problems, the network will operate safely and effectively (Yuan and Wang, 2016).

As block chain network is a decentralized network, the status of each node is equal. Therefore, to operate without third-party institutions like banks, it is necessary to work out some mechanisms that everyone must follow. Proof-of-Work consensus mechanism establishes mutual trust among nodes and incentive mechanism ensures that nodes provide computing power to maintain the whole network. Under these mechanisms, subjects compete and cooperate with each other to ensure the network runs efficiently.

As shown in Figure 8(a), M1, M2. . . . . . M6 refers to the six miners who dig the ore, and A and B are the two parties of the transaction. One miner records the transaction through bookkeeping competition among the miners, and the other five miners confirm the transaction histories between A and B, and then the transaction records are recorded in the block chain.

Assuming that M2 gains the bookkeeping right through competition, the activity level of M2 increases greatly in the network and becomes the right figure of Figure 8(b). PDD of the two figures does not change because of the same NPD, so only the subject interactive patterns need to be quantified. There are two profession categories, A, B and six miners. $q = 2/3$ is taken here. As for Figure 8(a):

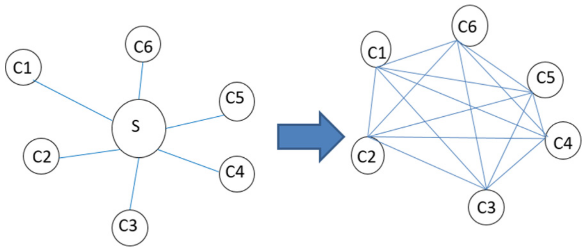

Figure 7.
Centralized network to P2P network

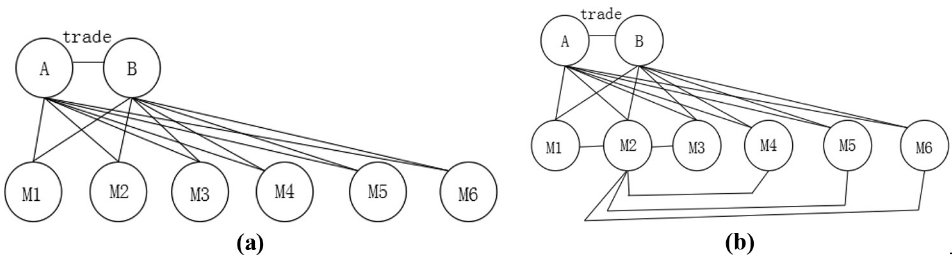

Figure 8.
Two block chain network diagrams

$$I_{AB} = 13/21,\ I_{\text{m}iners} = 2/3$$

Calculate the activity level of each intelligent subject according to equation (3):

$$SIP_A = 1/7 * I_{AB} + 6/7 * I_{miners} = 97/147,\ SIP_A = SIP_B.$$

$$SIP_{M1} = 1 * I_{AB} = 13/21,\ SIP_{M1} = SIP_{M2} = SIP_{M3} = SIP_{M4} = SIP_{M5} = SIP_{M6}$$

So the value of the second item in the model is 2.0556.

As for Figure 8(b):

$$I_{AB} = 13/21,\ I_{miners} = 29/51$$

Calculate the activity level of each intelligent subject according to equation (3):

$$SIP_A = 1/7 * I_{AB} + 6/7 * I_{miners} = 97/147,\ SIP_A = SIP_B$$

$$SIP_{M1} = 2/3 * I_{AB} + 1/3 * I_{miners} = 466/1071,\ SIP_{M1} = SIP_{M3} = SIP_{M4} = SIP_{M5} = SIP_{M6}$$

$$SIP_{M2} = 2/7 * I_{AB} + 5/7 * I_{miners} = 1457/2499.$$

So the value of the second item in the model is 2.673.

The increased links between M2 and the remaining five miners meant that M2's activity level was greatly increased in the network, which makes business entropy of the whole network become higher. Hence, in this use case, only the SIPs of subjects affect the intelligence level of the target network.

6. Differential evolution model and k-means clustering were used to distinguish intelligent subjects

In the previous sections, the paper discussed the related factors influencing the business entropy. The 3 network examples in the Section 5 can be classified by their division of labor. However, not all networks can find their professional categories only through division of labor. Therefore, in Section 6, k-means clustering algorithm with improved differential evolution is proposed.

Differential evolution (DE) was proposed by Storn and Price in 1995, which is an evolutionary algorithm based on population difference and random search. DE, characterized by internal information sharing and preserving the individual optimal solution, is a computational algorithm based on evolution of bionic intelligence. Differential evolution algorithm, as a new and efficiently heuristic parallel search technique, has been known for its fast convergence, few control parameters, simple training setting and robust optimization results (Storn and Price, 1997).

Let's say the initial population is PG $= \{X_{i,G}\ |\text{i} = 1,\ 2,\ 3\ldots\ \ldots N_p\}$, where $X_{i,G} = \left\{x^j_{i,G} | \text{j} = 1,\ 2,\ 3 \ldots\ldots \text{D}\right\}$ is the i-th individual in the population of generation G.

Each individual is described by D attributes, which affect business entropy analyzed in part 1:

- Mutation operation: Randomly select an individual as the parent basis vector and another two different individuals as the parent difference vectors from the population. Then produce a mutation individual based on the following equation:

$$V_{i,G} = X_{i,G} + F * \left(X_{b,G} - X_{c,G}\right). \tag{6}$$

F is the scaling factor which can be determined based on experience.

- Cross operation: equation (8) is used to perform crossover operation between the i-th individual $X_{i,G}$ in the population and its corresponding mutation individual $V_{i,G}$ from equation (6). And then get a new individual $U_{i,G}$:

$$U_{i,G} = \left(u_{i,G}^{1}, u_{i,G}^{2} \ldots\ldots u_{i,G}^{D}\right), \tag{7}$$

$$u_{i,G}^{j} = \begin{cases} v_{i,G}^{j} & if\ r_j \leq CR \\ x_{i,G}^{j} & otherwise \end{cases}. \tag{8}$$

r_j is a random number generated in each iteration and CR is a preset constant.

- Select operation: Substitute the original population member $X_{i,G}$ and the new member $U_{i,G}$ into the following equation. Individuals with better fitness are used to perform the next generation:

$$X_{i,G+1} = \begin{cases} Ui,G & if\ f\left(U_{i,G}\right) \leq f\left(X_{i,G}\right) \\ X_{i,G} & otherwise \end{cases}. \tag{9}$$

Due to the clustering operation, the fitness function is defined as follows: f(X_i) =J_c, where:

$$J_C = \sum_{j=1}^{k} \sum_{X_i \in Z_j} d\left(X_i, Z_j\right), \tag{10}$$

where Z_j is the j-th clustering center, and d(X_i, Z_j) is the Euclidean space distance from the sample to the corresponding clustering center.

Laplace mutation operator is adopted to select the scaling factor F in the paper. Laplace mutation operator can not only maintain the diversity of the population but also improve the convergence rate of the algorithm (Laszlo and Mukherjee, 2007). Laplace's distribution is written as follows:

$$f(x|\mu,\theta) = \frac{1}{2\mu} exp\left(\frac{-\left|x-\theta\right|}{\mu}\right), -\infty \leq \mathrm{x} \leq +\infty. \tag{11}$$

To prevent the phenomenon of premature and evolutionary stagnation which may occur in the late stage of differential evolution algorithm, the algorithm uses the following equation (12) for precocious judgment (Omran *et al.*, 2005):

$$p_i = \frac{f_i - f_{best}}{f_{worst} - f_{best}}. \tag{12}$$

The operational flow of algorithm is shown below in detail:

- *Step 1:* set the number of individuals N and the maximum number of iterations G_{max}.
- *Step 2:* randomly select samples as the clustering center, and calculate the fitness value of the current position, which is the population initialization.
- *Step 3:* for each individual $X_{i,G}$, generating the mutation operator F randomly according to Laplace distribution (11).
- *Step 4:* perform mutation operation according to equation (6) and crossover operation according to equation (8) respectively, generate test vector $U_{i,G}$ and perform selection operation according to equation (9).
- *Step 5:* reclassify the classification of samples via the nearest neighbor rule according to the clustering center coding of individuals.
- *Step 6:* recalculate the new clustering center to replace the original value.
- *Step 7:* check whether the individual falls into the local optimum from equation (12). If so, conduct a chaotic search on the individual variable scale to facilitate jumping out of the local optimum and go to step 3.
- *Step 8:* if the termination condition is not realized, go to step 3, and the value of G increases by 1; Otherwise, output the best individual value X_{best} and the best fitness value $f(X)_{best}$, and the algorithm ends.

7. Summary

This paper combines qualitative analysis of the factors affecting business entropy and quantitative expression of business entropy to discuss the relationship between professional categories of intelligent subjects in crowd intelligence network and the intelligence level of crowd network. There are many factors that affect the level of crowd intelligence network. Two important factors are listed in this paper, namely, Profession Distribution Deviation and Subject Interaction Pattern. PDD is quantified based on the deviation between the current NPD and optimized NPD. SIP is quantified according to the network topology. By combining these two factors, the MFBEQ model is proposed. By utilizing the model, three typical crowd intelligence networks (including logistics service supply chain, ant division and collaboration network and bit-coin network) are analyzed in detail.

Finally, the improved k-means clustering algorithm is introduced to cluster the intelligent subjects in the network. This algorithm is used to cluster the crowd network so that the entropy of the network can reach optimum. The simple entropy is not very useful for the evaluation of crowd intelligence network. The most important is to make a comprehensive evaluation of a certain network by combining those important factors that affect crowd network and business entropy.

Our future works are follows:

- the calculation of PDD can be performed based on cross entropy;
- the way that SIP is introduced into the MFBEQ model can be improved; and
- Evaluating the performance of our proposed model on real-world crowd network will be performed in the future.

References

Bird, D. and Munoz, C. (1983), "Automatic generation of random self-checking test cases", *IBM Systems Journal*, Vol. 22 No. 3, pp. 229-245.

Chapman, R.L., Soosay, C. and Kandampully, J. (2003), "Innovation in logistics services and the new business model: a conceptual framework", *International Journal of Physical Distribution and Logistics Management*, Vol. 33 No. 7, pp. 630-650.

Chen, G.R., Wang, X.F. and Li, X. (2015), *Introduction to Complex Networks Models, Structures and Dynamics*, Higher Education Press, Beijing, pp. 11-13.

Colorni, A., Dorigo, M. and Maniezzo, V. (1991), *Distributed Optimization by Ant-Colonies, European Conference on Artificial Life Pans*, Elsevier, Paris, pp. 134-142.

Decker, C. and Wattenhofer, R. (2013), "Information propagation in the bit coin network", *13th International Conference on Peer-to-Peer Computing (P2P)*, *IEEE*, pp. 1-10.

Grefen, P., Aberer, K., Hoffner, Y. and Ludwig, H. (2000), "Cross flow: cross – organizational workflow management in dynamic virtual enterprises", *International Journal of Computer Systems Science and Engineering*, Vol. 15 No. 5, pp. 227-290.

He, N., Li, D., Gan, W. and Zhu, X. (2007), "Computer science", *A Review of the Discovery of Important Nodes in Complex Networks*, Vol. 34 No. 12, pp. 1-5.

He, X., Zhu, Y., Hu, K. and Niu, B. (2006), "Information entropy and interaction optimization model based on swarm intelligence", in Jiao, L., Wang, L., Gao, X., Liu, J., Wu, F. (Ed.), *Lecture Notes in Computer Science on Advances in Natural Computation: ICNC 2006*, *Springer, Berlin, Heidelberg*, Vol. 4222.

Jennings, N.R. (1995), "Controlling cooperative problem solving in industrial multi-subject systems using joint intentions", *Artificial Intelligence*, Vol. 75 No. 2, pp. 195-240.

Laszlo, M. and Mukherjee, S. (2007), "A genetic algorithm that exchanges neighboring centers for k-means clustering", *Pattern Recognition Letters*, Vol. 28 No. 16, pp. 2359-2366.

Lee, J. and Kramer, B.M. (1992), "Analysis of machine degradation using neural networks based pattern discrimination n-model", *Journal of Manufacturing Systems*, Vol. 12 No. 5, pp. 379-387.

Liu, J., Pan, Z., Xu, J., Liang, B., Chen, Y. and Ji, W. (2018b), "Quality-time-complexity universal intelligence measurement", *International Journal of Crowd Science*, Vol. 2 No. 2, pp. 99-107.

Nelson, H. and O'Connel, A. (1978), "Estimation of premorbid intelligence levels using the new adult reading test", *Cortex*, Vol. 14 No. 2, pp. 234-244.

Neubauer, A. and Fink, A. (2009), "Intelligence and neural efficiency", *Neuroscience and Biobehavioral Reviews*, Vol. 33 No. 7, pp. 1004-1023.

Omran, M.G.H., Engelbrecht, A.P. and Salman, A. (2005), "Dynamic clustering using particle swarm optimization with application in unsupervised image classification", *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 9 No. 11, pp. 199-204.

Paninski, L. (2003), "Estimation of entropy and mutual information", *Neural Computation*, Vol. 15 No. 6, pp. 1191-1253.

Persson, G. and Virum, H. (2001), "Growth strategies for logistics service providers, a case study", *The International Journal of Logistics Management*, Vol. 12 No. 1, pp. 53-64.

Ren, Z., Shao, F., Liu, J. and Guo, Q. (2013), "BingHong Wang: research on importance measurement method of network nodes based on degree and clustering coefficient", *Acta Physica Sinica*, Vol. 62 No. 12, pp. 1-5.

Storn, R. and Price, K. (1997), "Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces", *Journal of Global Optimization*, Vol. 11 No. 4, pp. 341-359.

Xiao, S., Wei, C.-P. and Dong, M. (2016), "Crowd intelligence: analyzing online product reviews for preference measurement", *Information and Management*, Vol. 53 No. 2, pp. 169-182.

Yuan, Y. and Wang, F.Y. (2016), "Block chain: the state of the art and future trends", *Acta Automatica Sinica*, Vol. 42 No. 4, pp. 481-494.

Further reading

Chen, X., Li, H. and Ke, M. (2012), "Study on complex network stability based on the structure parameters", *IEEE/ACIS 11th International Conference on Computer and Information Science*, pp. 521-526.

Chen, J., Hu, Q. and Song, J.-S. (2017), "Effect of partial cross ownership on supply chain intelligence level", *European Journal of Operational Research*, Vol. 258 No. 2, pp. 525-536.

Chen, M., Mao, S. and Liu, Y. (2014), "Big data: a survey", *Mobile Networks and Applications*, Vol. 19 No. 2, pp. 71-209.

Dorigo, M., Maniezzo, V. and Colorni, A. (1996), "The ant system: optimization by a colony of cooperating subjects", *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, Vol. 26 No. 1, pp. 29-41.

Fakhrzad, M.B., Pourfereidouni, H. and Pourfereidouni, M. (2018), "Dual-channel supply chain synchronization with deterministic and stochastic demand under cost-sharing contract", *International Journal of Supply and Operations Management*, Vol. 5 No. 1, pp. 28-41.

Geng, H.Q., Gu, S.Z. and Guo, D.M. (2004), "Analyses on evolution of household energy consumption structure based on information entropy", *Journal of Natural Resources*, Vol. 19 No. 2, pp. 257-262.

Liu, F., Xiao, B., Jin, H. and Zhang, Q. (2018a), "A fusion method for node importance measurement in complex networks", *5th International Conference on Systems and Informatics*, pp. 682-686.

Shi, H., Ritter, G.X. and Wilson, J.N. (1994), "Parallel image processing with image algebra on SIMD mesh-connected computers", *Advances in Imaging and Electron Physics*, Vol. 90, pp. 353-431.

Corresponding author

Feng Yang can be contacted at: fengyang@bjtu.edu.cn