IJCS 5,2

 $\mathbf{204}$

Received 3 March 2021 Revised 24 March 2021 Accepted 30 March 2021

Crowd evolution method based on intelligence level clustering

Lulu Ge

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China and North China University of Science and Technology, Tangshan Hebei, China

Zheming Yang

Institute of Computing Technology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China, and

Wen Ji

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Abstract

Purpose – The evolution of crowd intelligence is a mainly concerns issue in the field of crowd science. It is a kind of group behavior that is superior to the individual's ability to complete tasks through the cooperation of many agents. In this study, the evolution of crowd intelligence is studied through the clustering method and the particle swarm optimization (PSO) algorithm.

Design/methodology/approach – This study proposes a crowd evolution method based on intelligence level clustering. Based on clustering, this method uses the agents' intelligence level as the metric to cluster agents. Then, the agents evolve within the cluster on the basis of the PSO algorithm.

Findings – Two main simulation experiments are designed for the proposed method. First, agents are classified based on their intelligence level. Then, when evolving the agents, two different evolution centers are set. Besides, this paper uses different numbers of clusters to conduct experiments.

Practical implications – The experimental results show that the proposed method can effectively improve the crowd intelligence level and the cooperation ability between agents.

Originality/value – This paper proposes a crowd evolution method based on intelligence level clustering, which is based on the clustering method and the PSO algorithm to analyze the evolution.

Keywords Cluster analysis, Crowd intelligence, Crowd science, Crowd evolution, Intelligence measure

Paper type Research paper

1. Introduction

Evolution refers to the process in which organisms change over time in the natural world. In the classic work *The Origin of Species* (Darwin, 1902), Darwin proposed a theory to explain why life forms evolved and how the new structures are formed after the evolution of life.



International Journal of Crowd Science Vol. 5 No. 2, 2021 pp. 204-215 Emerald Publishing Limited 2398-7294 DOI 10.1108/IJCS.03-2021-0010

This work is supported by the National Key R&Program of China (2017YFB1400100), and the National Natural Science Foundation of China (62072440), and the Beijing Natural Science Foundation (4202072)

[©] Lulu Ge, Zheming Yang and Wen Ji. Published in *International Journal of Crowd Science*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http:// creativecommons.org/licences/by/4.0/legalcode

The theory is the natural selection that we are familiar with. It contains three basic elements, namely, variation, genetic and selection. This theory explains the emergence of new species, explicates the process of organic structure changing with time, also interprets why the components of these structures have obvious purposeful characteristics (Buss, 2015). In essence, evolution is a necessary process for organisms to survive better and organisms will be more competitive in the new environment. Many crowd intelligence phenomena such as the ant colony effect and bird flight, are universally evolving.

The phenomenon of crowd intelligence in human society, such as the process of enterprise management and the coordinated operation of the industrial chain (Yi-hong, 2005) in the economic field; various seminars and collective behavior processes (Andrusevich et al., 1988) in the social field and the national election in the political field; are all expected to achieve better or best results by gathering many individual intelligent. In the internet age, big data, artificial intelligence, the internet of things, Industry 4.0, cloud computing and other technologies have continuously expanded the depth and breadth of connection between people, enterprises, government and other institutions and intelligent things. Compared with the crowd intelligence phenomenon in human society, the crowd intelligence phenomenon in the network environment is large scale and closely connected. Many digital-selves are in the space of deep integration of physical space, consciousness space and information space, so the individuals in the crowd network are heterogeneous and may cause chaos and turbulence. Chai et al. (2017) put forward the concept of crowd science and engineering, aiming to study the principles, laws, methods, technologies and related engineering applications of the triple fusion system of information, physics and society under the large-scale internet online environment (for example, large-scale ecommerce platform, social network, etc.). Similar to the evolution in biology and human society, crowd evolution is one of the main issues in the scientific theory of crowd science. Crowd evolution can be defined as the intelligent individuals cooperate to complete a series of complex tasks so that the final effect is better than the individual completed.

Similar to crowd intelligence, there have been a series of studies on swarm intelligence, such as swarm intelligence (Kennedy, 2006), collective intelligence (Leimeister, 2010) and multi-agent system (Bellifemine *et al.*, 2000), their intelligence subjects are homogeneous and isomorphic. These disciplines use traditional evolutionary algorithms [e.g. particle swarm optimization (PSO)] to study the evolution of agents and have achieved good results. This inspired us to explore the potential relationship between the traditional evolutionary algorithm and the evolution method of crowd intelligence.

The intelligence measure is also an essential issue in the field of crowd science, and the intelligence level also affects the development of crowd intelligence. Therefore, the influence of the intelligence level can be considered when studying the issue of crowd evolution. Yang and Ji (2020) proposed an intelligence measure method for the hybrid property of heterogeneous agents. In addition, there have been relevant research studies on the evolution of crowd intelligence at present. Wang and Sun (2019) proposed an evolution simulation framework based on the ecological structure of the crowd network; Wang *et al.* (2019) proposed an adaptive information sharing method based on two-stage optimization; Wang and Sun (2020) proposed a novel simulation framework for crowd co-evolution. However, none of these studies considered using a universal intelligence measure method to analyze the evolution.

In this paper, we propose a crowd evolution method based on intelligence level clustering. First, the method uses the intelligence level of the digital selves as a measure to cluster digital selves. Clustering divides digital selves with similar attributes into one group. The traditional partition-based clustering methods use Euclidean distance or Manhattan distance as the measure standard, but digital-selves are heterogeneous in the crowd science, the digital-selves cannot be well-divided using these metrics. Due to the importance of the Crowd evolution method intelligence level, we cluster digital selves based on their intelligence level, the digital selves can be appropriately divided and the digital selves are no longer chaos. We use the intelligent measurement method proposed in Yang and Ji(2020), which is based on quality and time. Then inspired by the PSO, the method uses this algorithm to evolve the digitalselves after clustering, the interaction between digital-selves within the cluster will be more efficient. When the interaction within the cluster is completed that is to say, reaching a dynamic balance, the inter-cluster interaction will be carried out, finally, the crowd evolution is accomplished. We design simulation experiments and the results show that the proposed method can improve the crowd intelligence level, the population has evolved.

The contributions of this paper can be summarized as follows:

- We propose a crowd evolution method based on intelligence level clustering. The
 method uses the intelligence level of digital-selves as the clustering metric to cluster
 digital-selves so that the clustering method can be appropriately applied to the
 crowd science and the digital-selves will become orderly.
- The PSO algorithm is introduced to make the digital selves interact and collaborate to achieve crowd evolution. We update the position of the digital selves that is intelligence level and set up the evolution centers to let the digital selves evolve.
- We also design two main simulation experiments to verify the effectiveness of the proposed method. The results show that the proposed method can improve the crowd intelligence level.

The rest of the paper is organized as follows: Section 2 introduces the work on the measure and evolution of crowd intelligence. In Section 3, we describe the evolution method in detail. In Section 4, simulation experiments are designed to show the results of the evolution method and prove the effectiveness of the method. Finally, we present the conclusions and future work in Section 5.

2. Related work

Intelligence is an important factor in human development. The IQ test is a measurement method of human intelligence. Similarly, how to measure the intelligence level is an important issue in the development of crowd science. The intelligence measure methods can evaluate the intelligence of individuals and groups and offer support for evaluating the innovation potential of individuals and groups. In crowd science, intelligence level refers to the ability of the digital-selves to respond to environmental or a series of tasks, including the accuracy and time of response. However, due to the complexity and heterogeneity of digital selves in crowd networks, the intelligence level of digital selves is not easy to measure. To solve this problem, there have been some related works. These works have put forward measurement methods to evaluate crowd intelligence. Liu et al. (2018) proposed a universal measure method based on quality-time-complexity. This method takes into account three factors, namely, the quality of the environment, timeliness and test complexity and proved that the correlation among the three factors. Yang and Ji (2020) proposed a quality-time model of heterogeneous agents measure for crowd intelligence. For the hybrid property of heterogeneous agents, describing crowd intelligence as an aggregate of agent's multiple response-abilities to environment or external stimuli. In addition, response-ability is mainly measured by quality and time. These universal intelligent measure methods have laid a certain foundation for the development of crowd science.

In nature and human society, evolutionary phenomena exist universally. Human beings become what they are now through continuous evolution to adapt to the present natural

IICS

5,2

environment. Affected by the evolution of group behavior in nature, scholars have proposed many traditional evolutionary algorithms to enable information interaction and collaboration between biological individuals, such as genetic algorithm (GA), ant colony optimization (ACO) and PSO. GA (Genlin, 2004) is a method to search the optimal solution by simulating the natural evolution process, including selection, crossover and mutation operations, which has been widely used in machine learning. ACO (Dorigo *et al.*, 2006) is an evolutionary algorithm that uses the positive feedback learning mechanism of pheromone. PSO (Kennedy and Eberhart, 1995) is a cooperative evolutionary algorithm that simulates the foraging behavior of birds, the shared information is given by the global best particle. In machine learning, some studies (Chatterjee *et al.*, 2017) showed that PSO is also a potential neural network algorithm. Those evolution algorithms have played a good role in swarm intelligence, which makes us think about whether these methods can be applied to the crowd intelligence (Yu *et al.*, 2018).

There have been some studies on the evolution of crowd intelligence. Wang and Sun (2019) proposed an evolution simulation framework based on the ecological structure of the crowd network. The concept of ecological structure is put forward in crowd science for the first time, and this framework studied the change process of each intelligent subject in the ecological structure of the e-government system. To optimize and evaluate the performance of the crowd network, Wang et al. (2019) proposed an adaptive information-sharing method based on two-stage optimization, using two stages to analyze the information sharing mode in the crowd network. This method fully considered the factors affecting information sharing and combined local optimization with global optimization. Wang and Sun (2020) proposed a novel simulation framework for crowd co-evolution to study the evolution of the relationship between individuals in the crowd network. These works have studied the evolution of crowd intelligence, but they have not considered using a universal intelligence measure method to analyze the evolution. Therefore, we consider studying the crowd evolution based on intelligence level. We first clustering digital selves based on intelligence level, then use the algorithm of PSO to evolve digital selves because of the advantages of this method. In this case, the traditional evolution method can be applied to crowd intelligence.

3. Implementation of the evolution approach

Figure 1 shows the who process of our proposed method. The whole process consists of two parts, namely, the clustering process based on the intelligence measure and the evolution process of crowd intelligence, we will introduce them in detail.



Notes: The different shape and the different color of individuals represent different intelligence levels. The same shape represents the similar intelligence level, and the darker the color, the greater the intelligence level. First, the heterogeneous digital-selves are clustered through the intelligence level. Then, the evolution of the digital-selves within the cluster will be carried out, the color of the agent will change and the crowd intelligence level will be improved

Figure 1. The who process of our proposed method 3.1 Cluster digital selves based on the crowd intelligence measure

Clustering analysis (Everitt, 2018) is the process of dividing the collection of physical or abstract objects into several classes composed of similar objects. Clustering analysis does not need the predicted label, the essence is to explore the potential relationship within the data. It can be applied in many fields, in business, clustering can help operators analyze user preferences and provide corresponding services to users; in biology, clustering can classify genes to understand population structure; in the social field, clustering can divide people from different regions and fields for better communication.

K-medoids clustering (Park and Jun, 2009) is one of the classical and representative methods of cluster analysis, its cluster center is an object in the cluster. It uses Euclidean distance as a measure to determine, which category the object belongs to. Similar to it are k-means clustering, k-median clustering, their cluster center is the mean value of the cluster. These are clustering methods based on partition. K-medoids clustering is not sensitive to noise and abnormal data. The goal of the traditional K-medoids clustering algorithm is to minimize the error *E*:

$$E = \sum_{i=1}^{k} \sum_{o \in C_i} |o - \mu_i|$$
 (1)

where *o* represents the cluster object; $C_i(i = 1, 2, ..., k)$ represents the divided clusters; *k* represents the cluster number; μ_i is the cluster center of the cluster C_i .

In crowd science, the Euclidean distance cannot divide the digital-selves reasonably, and cannot reflect the intelligent level. This work proposes to use the intelligence level of the digital selves as the clustering measure to determine, which category the digital self belongs to. Intelligence is defined as the ability of the digital self to respond to the environment or external stimuli in crowd science, including response time, response quality. *I* represent the intelligent level of the digital self and the intelligent level set is expressed as; $I = \{I_1, I_2, I_3, \ldots, I_N\}$ *N* indicating the number of digital selves. In the quality-time model, the intelligent level is defined as:

$$I_j = \frac{Q_j}{T_j} (j = 1, 2, \dots, N)$$
 (2)

where Q_j represents the quality of the task completed by the *j*-th digital-self, and T_j represents the time when the *j*-th digital-self completes the task. Theoretically speaking, the range of values of the intelligence level, quality and time for completing tasks is $(0, +\infty)$, but in actual scenarios, the intelligence level of the digital-self is limited, the quality and the time are also. Considering this factor and to measure convenience, normalized the intelligence level, quality and time of the digital-self is necessary, as $I_j \in (0,1)$ and $Q_j \in (0,1)$. According to equation (2), it can be got $0 < \frac{Q_i}{T_j}$ (j = 1, 2, 3, ..., N) < 1 that is $0 < Q_j < T_j < 1$. This method uses the intelligence level of the final goal is to minimize the following errors:

$$E = \sum_{i=1}^{k} \sum_{o \in C_i} |I_o - I_{\mu_i}|$$
(3)

where $I_{\mu i}$ (i = 1, 2, 3, ..., k) denotes the intelligence level of the cluster center of the *j*-th cluster. After dividing the digital selves according to their own intelligence level, *k* clusters are obtained. The later evolution process is based on the *k* clusters.

208

IICS

5,2

<i>3.2 Crowd evolution</i> PSO performs well in traditional evolution methods. It needs to update the position and speed of particles, and the update formula is as follows:	Crowd evolution method
$v_j^{g+1} = v_j^g + c_1 r_1 \left(p_j - x_j^g \right) + c_2 r_2 \left(p_{gb} - x_j^g \right) $ (4)	
$x^{g+1} = x^g + v^{g+1} \tag{5}$	209

where v_j^{g} represents the speed of the *j*-th particle in the g-th iteration; x_j^{g} represents the position of the *j*-th particle in the g-th iteration; c_1, c_2 are the learning factor, usually non-negative constants; r_1 , r_2 are the random number between (0, 1); p_j represents the individual optimal value searched by the current particle; p_{gb} is the global optimal value, which represents the current optimal position of the whole population. Wang (2007) pointed out that the particle velocity term may cause the particles to deviate from the correct evolutionary direction, which will slow down the convergence rate in the later stage of evolution. Therefore, we use the idea of a simplified PSO algorithm that is only updates the position of particles without considering the speed, which can avoid or slow down the problem of slow convergence. In addition, we do not consider the individual optimal value and the global optimal value, set the evolution center instead. After the digital selves are divided, the simplified PSO algorithm is used to let digital selves interact with information and evolve. Use the following equation to update the intelligence level of the digital self:

$$x_{j}^{g+1} = x_{j}^{g} + c_{1} * r * \left(p_{ec} - x_{j}^{g} \right)$$
(6)

where r represents a random number between (0, 1). The first term on the right side of equation (6) represents the intelligence level of the digital-self in the g-th iteration

Algorithm 1 The Implementation Process of The Proposed Method.

- Input: the number of digital-selves N, the number of cluster k, parameters c₁, threshold α, maximum number of iterations G;
- 2: Initialize: Generate digital-selves o_j (j = 1, 2, ..., N), the cluster center μ_j ;
- 3: Use equation (3) to classify digital-selves;
- 4: while not converge $\|g < G do$
- 5: Use equation (6) to update the position of digital-selves (intelligence level);
- 6: endwhile
- 7: **Output:** Crowd intelligence level I_G , digital-selves after evolution.

and represents the influence of the previous intelligence level on the current intelligence level; the second term is the influence of the evolution center p_{ec} on the intelligent individuals within the cluster. By comparing and imitating with the evolution center, the information sharing and cooperation between the individuals are realized. Based on the above-mentioned two processes, the entire algorithm is shown in Algorithm 1.

In addition, the threshold α is set during the evolution process to control the degree of evolution. When the difference between the intelligence level of the digital self and the evolution center is greater than the threshold, the digital self is allowed to evolve and when it is less than the threshold, the digital self is evolved with the original intelligence level. When

the threshold is set to a large value, the number of evolving digital-selves will be less and the intelligence level will not be improved very high; when the threshold is set to a small value, the number of evolving digital-selves will be more and the intelligence level will increase.

4. Experiments and results

To verify the effectiveness of the proposed method, we design two simulation experiments to show the evolution process of the crowd intelligence, namely, the evolution center one is the digital self with the maximum intelligence level within the cluster, the other is the cluster center. In our experiments, the initial intelligent individuals are randomly generated in the domain, and the generation conditions meet the requirements described in Section 3.1 (as shown in Figure 2). We use a node to represent a digital self. Each digital self has its intelligence level. The position of the node in the coordinate system represents its intelligence level. The y-axis of the digital self represents the quality of completing the task, and the x-axis represents the time of completing the task.

In the experiments, the number of digital-selves is set to N = 200, the maximum number of iterations is M = 50. When the evolution center is the digital self with the maximum intelligence level within the cluster set threshold to $\alpha = 0.12$; when the evolution center is the cluster center set the threshold to $\alpha = 0.1$. Figures 3–5 show the evolution results of the proposed method, where the number of clusters k is 3–5, respectively. Other numbers can be set as needed, but it should not be too large or too small. If it is too large, it will reduce the intra-class spacing, but maybe damage the generalization of data and the interaction between digital-selves will still be very chaotic; if it is too small, the classification effect may be incomplete that is two digital-selves with great different intelligence levels are also divided into one category, which cannot achieve the expected effect. When k = 1 and k = 200, the final clustering effect is the same. Figures 3–5 show the results of the two processes of our method. Through Figures 3(a), 4(a) and 5(a), it can be seen that the effect of clustering digital selves according to their intelligence level, digital selves change from chaos to order. After that, each digital self

interacts with the evolution center in its own cluster to complete the evolution process.



Figure 2. The initial digitalselves, the number is 200

IICS

5,2



0.6

0.4 0.3 0.2 0.1

Figure 3. The two processes of the proposed method and the results of two experiments IJCS 5,2

212



with the maximum intelligence level within the cluster; (c) is the evolution result when the evolution center is the cluster center. The **Notes:** (a) Is the result of clustering the digital selves; (b) is the result of the evolution when the evolution center is an digital self number of clusters is 4

Figure 4. The two processes of the proposed method and the results of two experiments



IJCS 5,2 Figures 3(b), 4(b) and 5(b) show the results of the evolution when the evolution center is the digital self with the maximum intelligence level within the cluster. Figures 3(c), 4(c) and 5(c) show the results of the evolution when the evolution center is cluster center. The connection between digital-selves is closer. After the intra-cluster interaction is completed, the information interaction between the clusters is carried out, and finally, complete the evolution of the whole population. Figure 6(a) shows the changing trend of the overall intelligence level in the iteration process.

214

Figure 6. (a) shows the changing process of crowd intelligence level in the process of

(b) compares the crowd intelligence level before and after evolution when the evolution center is different and the number of clusters is different



In addition, Figure 6(b) compares the results before and after evolution. When the number of clusters is larger, the improvement of crowd intelligence level will be smaller. This is because the number of clusters is large, the maximum or mean intelligence level in each cluster will be smaller, so the final improvement of crowd intelligence level is smaller. However, regardless of the number of clusters, the final crowd intelligence level is steadily improving. At the end of the iteration, the process basically reached convergence and reached a dynamic balance state.

5. Conclusion and future work

In this paper, we propose a crowd evolution method based on the intelligence level clustering on the basis of the previous series of related work. We described the two processes of the proposed method in detail. When clustering the digital selves, the method uses the intelligence level of the digital selves as the metric; then adopts PSO algorithm to explore the crowd evolution. We design simulation experiments to prove the effectiveness of the proposed method. The results show that the method can improve the crowd intelligence level. In the future, we plan to solve the problem of unstable results due to the random selection of initial cluster centers. In addition, we will explore more effective evolution methods to improve group collaboration capabilities and the crowd intelligence level.

References

Andrusevich, V.V., Kveselava, A.D. and Kukuliev, G.Y. (1988), "Collective behavior processes in allocation of many-dimensional resources", *Automation and Remote Control*, Vol. 1988 No. 3.

Bellifemine, F., Poggi, A. and Rimassa, G. (2000), "Developing multi-agent systems with JADE", In International Workshop on Agent Theories, Architectures, and Languages, Springer, Berlin Heidelberg, pp. 89-103.

buss, D. (2013), Evolutionary Esychology. The iven Science of the Mina, Esychology Fless.	Crowd
Chai, Y., Miao, C., Sun, B., Zheng, Y. and Li, Q. (2017), "Crowd science and engineering: concept and research framework", <i>International Journal of Crowd Science</i> , Vol. 1 No. 1, pp. 2-8.	evolution method
Chatterjee, S., Sarkar, S., Hore, S., Dey, N., Ashour, A.S. and Balas, V.E. (2017), "Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings", <i>Neural Computing and Applications</i> , Vol. 28 No. 8, pp. 2005-2016.	
Darwin, C. (1902), "On the origin of species by the means of natural selection, or, the preservation of favoured races in the struggle for life", G. Richards.	215
Dorigo, M., Birattari, M. and Stützle, T. (2006), "Ant colony optimization", <i>IEEE Computational Intelligence Magazine</i> , Vol. 1 No. 4, pp. 28-39.	
Everitt, B.S. (2018), "Cluster analysis", <i>In Multivariate Analysis for the Behavioral Sciences</i> , CRC Press, pp. 341-363.	
Genlin, J. (2004), "Survey on genetic algorithm", Computer Applications and Software, Vol. 2 No. 1, pp. 69-73.	
Kennedy, J. (2006), "Swarm intelligence", In Handbook of Nature-Inspired and Innovative Computing, Springer, Boston, MA, pp. 187-219.	
Kennedy, J. and Eberhart, R. (1995), "Particle swarm optimization", Proceedings of ICNN'95- International Conference on Neural Networks, Perth, WA, Australia, Vol. 4, pp. 1942-1948.	
Leimeister, J.M. (2010), "Collective intelligence", <i>Business and Information Systems Engineering</i> , Vol. 2 No. 4, pp. 245-248.	
Liu, J., Pan, Z., Xu, J., Liang, B. and Ji, W. (2018), "Quality-time-complexity universal intelligence measurement", <i>International Journal of Crowd Science</i> , Vol. 2 No. 1, pp. 18-26.	
Park, H.S. and Jun, C.H. (2009), "A simple and fast algorithm for K-medoids clustering", <i>Expert Systems with Applications</i> , Vol. 36 No. 2, pp. 3336-3341.	
Wang, H.U. (2007), "A simpler and more effective particle swarm optimization algorithm", <i>Journal of Software</i> , Vol. 18 No. 4, pp. 861-868.	

 $\overline{}$

1

Wang, J. and Sun, H. (2019), "An evolution simulation framework for ecological structure of crowd networks", *International Journal of Crowd Science*, Vol. 4 No. 1, pp. 87-100.

- Wang, K. and Sun, H. (2020), "A novel simulation framework for crowd co- evolutions", *International Journal of Crowd Science*, Vol. 4 No. 3, pp. 245-254.
- Wang, X., Pan, Z., Li, Z., Ji, W. and Yang, F. (2019), "Adaptive information sharing approach for crowd networks based on two stage optimization", *International Journal of Crowd Science*, Vol. 3 No. 3, pp. 284-302.
- Yang, Z. and Ji, W. (2020), "A quality-time model of heterogeneous agents measure for crowd intelligence".
- Yi-Hong, Y.U. (2005), "Type of industrial chain and the benchmark of industrial chain efficiency", *China Industrial Economy*, Vol. 11, pp. 35-42.
- Yu, C., Chai, Y. and Liu, Y. (2018), "Literature review on collective intelligence: a crowd science perspective", *International Journal of Crowd Science*, Vol. 2 No. 1, pp. 64-73.

Corresponding author

Wen Ji can be contacted at: jiwen@ict.ac.cn