An improved rule mining technology based on swarm intelligence computation

Zhaoyin Zhang *

School of Computer Science and Technology in Heilongjiang University 150010, China

Received 1 September 2014, www,cmnt,lv

Abstract

The traditional artificial fish swarm algorithm is easy to converge to local optimum. An improved artificial fish swarm algorithm is proposed which modifies position update formula of fish according to acceleration. Then rule mining algorithm based on improved artificial fish swarm is proposed, which includes rule coding, rule evaluation and determination of fitness function. UCI data set is used to test the performance of proposed algorithm. The experiment results show that the proposed algorithm has higher classification accuracy than particle swarm optimization and artificial fish swarm algorithm. It also has fast convergence speed compared with traditional artificial fish swarm algorithm.

Keywords: rule mining technology, artificial fish swarm, classification accuracy

1 Introduction

Classification of rules is an important task of data mining, which has become a hot spot of many scholars. Classification rules mining analyses the attributes of the known data object, and differentiate these attributes into different categories of rules. Then take advantage of these rules, the category division of the unknown data objects can be carried out. The classification rule mining algorithms mainly include the decision tree method, Bayes classification, intelligent optimization algorithm (genetic algorithm, particle swarm optimization, artificial fish algorithm, and ant colony algorithm), artificial neural network method, the rough set method, concept lattice, etc.

Data mining with an ant colony optimization algorithm was proposed by R. Parpinelli [1]. A new classification-rule pruning procedure for an ant colony algorithm was proposed by A.Chan [2]. A.A.Freitas also adopted ant colony algorithms for data classification [3]. Simultaneous ant colony optimization algorithms for learning linguistic fuzzy rules are proposed by M. Galea[4]. Extensions to the Ant-Miner classification rule discovery algorithm was proposed by K.M.Salama [5]. Multiple pheromone types and other extensions to the AntMiner classification rule discovery algorithm was also proposed by K.M.Salama [6]. A new version of the Ant-Miner algorithm discovering unordered rule sets was proposed by J. Smaldon [7]. Classification rule mining with an improved ant colony algorithm was proposed by Z. Wang [8]. Bees swarm optimization for web association rule mining was proposed by Y. Djenouri[9]. A hybrid bee's swarm optimization and tabu search algorithm for association rule tining was proposed by Youcef Djenouri [10]. Ant programming guided by grammar for building rule-based classifiers was proposed by J. Olmo [11]. Classification rule mining using ant programming guided by grammar with multiple pareto fronts was also proposed by J. Olmo [12]. A new sequential covering strategy for inducing classification rules with ant colony algorithms was proposed by F. Otero[13]. Ant colony algorithm is used for data classification by A. Freitas [14]. Classification rules method generated by multiple population PSO algorithm was proposed by Li-ping Yan,[15]. Classification rules discovery based on artificial fish algorithm was proposed by Jun-qing Chen [16]. Besides, intelligent computing is not only used in data mining, but also used in other areas [17-22]. All kinds of intelligent computing can provide reference for data mining.

The paper is organized as follows. In the next section, an improved artificial fish swarm algorithm is proposed. In Section 3, rule mining based on improved artificial fish swarm algorithm is proposed, including rule coding and determination of fitness function. In Section 4, in order to test classification accuracy of proposed algorithm, rule mining experiment is carried out based on UCI data set. In section5, some conclusions are given.

2 An improved artificial swarm algorithm

In the artificial fish swarm algorithm, n is the number of fish in the fish swarm, individual state of artificial fish is $X = (x_1, x_2, \cdots, x_n)$, and x_1, x_2,…, x_n represents optimized variables. Food concentration of artificial fish is measured by targeted function $Y = f(X)$. The

* Corresponding author Email:zhaoyingaga@126.com

distance between two fish is $d_{ij} = \left\| X_i - X_j \right\|$. $Visual$ Represents perception range of the fish, δ represents crowded degree, $Step$ and represents step length. try_number Represents try times. The principle of artificial fish swarm is as follows.

Step 1 Parameter n , $visual$, $step$, δ , and try_number are initialized and iteration times $n_c = 0$.

Step 2 Artificial fish swims randomly within their vision and when it finds food, the fish gets close to food.

Step 3 The foraging process is as follows.

(1) $m = 0$ and the current state of artificial fish is x_i .

(2) $x_j = random(n(x_i, visual))$, then a new state is generated and $m \leftarrow m+1$.

(3)If $f(x_i) < f(x_j)$, the state of artificial fish is modified according to $x_i \leftarrow x_i + random(step)(x_j - x_i) / \left\| x_j - x_i \right\|$.

If $m < try\_number$, return to (2).

(4)The fish swim randomly and go forward $x_i \leftarrow x_i + random(step)$.

Step 4 The swarm operator is as follows.

(1) The companion within its vision is $K = \{x_j \mid x_j - x_i \le visual\}(i, j = 1, 2, \cdots n)$

(2)The centre of K is calculated by $x_c = \sum x_j / n_f$ .The number of companion within its vision is n_f , $n_f \ge 1$.

(3)If $n_f / n < \delta$, $(0 < \delta < 1)$ and $f(x_i) < f(x_c)$, fish swim one step toward to x_c according to $x_{inext} = x_i + random(step) \cdot (x_c - x_i) / \left\| x_c - x_i \right\|$. Otherwise return to step 2.

Step 5 Follow operator.

(1) Look for the optimal companion $x_{\max}$ in its adjacent area. If $x_{\max}$ does not exist, go on to step 2.

(2)If $f(x_i) < f(x_{\max})$, and $\frac{n_f}{n} < \delta$, $(0 < \delta < 1)$, then the fish swim one step toward the position of $x_{\max}$ according to $x_i \leftarrow x_i + random(step)(x_{\max} - x_i) / \left\| x_{\max} - x_i \right\|$.

Step 6 If the algorithm does not meet the maximum iteration times, it turns to step 2, otherwise output the results. The improved artificial fish swarm algorithm is as follows.

$d_{ij}(t + \Delta t)$ Represents move direction of i under the impact of j in the next step. $X_i(t)$ And $X_j(t)$ represents the position of i and j . r_3 Represents radius of attracting area, r_1 represents radius of rejection area and r_2 represents proposed radius of middle area.

$$d_{ij}(t+\Delta t) = \begin{cases} (-)\dfrac{X_j(t) - X_i(t)}{\left|X_j(t) - X_i(t)\right|} & \left|X_j(t) - X_i(t)\right| \le r_1 \\ (+/-)\dfrac{X_j(t) - X_i(t)}{\left|X_j(t) - X_i(t)\right|} & r_1 < \left|X_j(t) - X_i(t)\right| \le r_2 \\ (+)\dfrac{X_j(t) - X_i(t)}{\left|X_j(t) - X_i(t)\right|} & r_2 < \left|X_j(t) - X_i(t)\right| \le r_3 \end{cases}$$

Current speed of $Fish_i$ is $Velocityy_i(t)$ and the speed of next step is determined by current speed and $Acceleration_i(t)$. The acceleration of rejection area is calculated as follows.

$Acceleration_{ir}(t) = 0$

For $j = 1, 2, \cdots, n$ and $j \ne i$ do

If

$\left|X_j(t) - X_i(t)\right| \in (0, r_1]$

$Acceleration_{ir}(t) = Acceleration_{ir}(t)$
$-rand() \cdot (X_j(t) - X_i(t))$

End For.

The acceleration of attraction area is calculated as follows.

$Acceleration_{ia}(t) = 0$

For $j = 1, 2, \cdots, n$ and $j \ne i$ do

If $\left|X_j(t) - X_i(t)\right| \in (r_2, r_3]$

$Acceleration_{ia}(t) = Acceleration_{ia}(t)$
$-rand() \cdot (X_j(t) - X_i(t))$

End for.

The acceleration of middle area is calculated by $Acceleration_{im}(t) =$

$$\sum_{j=1}^{N} rand() \cdot sign() \cdot (X_j(t) - X_i(t)) .$$

The acceleration of getting close to the best position is

$Acceleration_{ib}(t) = rand() \cdot (X_{best}(t) - X_i(t))$.

$X_{best}(t)$ Represents the best position at present. At last, acceleration can be calculated by

$$Acceleration_i(t) = \alpha_{ir} \cdot Acceleration_{ir}(t) + \alpha_{ia} \cdot Acceleration_{ia}(t) + \alpha_{im} \cdot Acceleration_{im}(t) + \alpha_{ib} \cdot Acceleration_{ib}(t).$$

α represents weight factor. Then speed and position can be calculated by

$$Velocity_i(t+\Delta t) = Velocity_i(t) + \Delta t \cdot Acceleration_i(t),$$

$$X_i(t+\Delta t) = X_i(t) + \Delta t \cdot Velocity_i(t+\Delta t).$$

Production cost, and production of different products has different unit production cost. Each distribution centre has the maximum capacity, and its stored products cannot exceed the maximum capacity.

3 Rule mining based on improved artificial fish swarm algorithm

IF…THEN is used to express the generated rules.

If Attribute1_min<< x_1 << $Attribute1_{\max}$ and

Attribute2_min<< x_2 << $Attribute2_{\max}$ and

…………..

AttributeN_min<< x_N << $AttributeN_{\max}$ and

Then class C

$Afish(x_1, x_2, \cdots, x_n, C)$ is used to express an artificial fish individual. The rule coding is show in table 1 and C represents class attribute. For feature attributes, real number coding method is adopted. For class attribute, binary code is used to represent each category.

The fitness function is

$$f(x) = w_1 \cdot S(R) + w_2 \cdot support + w_3 \cdot confidence.$$

$w_1 = 0.2$, $w_2 = 0.4$, $w_3 = 0.4$, $S(R) = M_c(R) - N_c(R)$, $support = \frac{tp}{tp+fp}$, $confidence = \frac{tn}{tn+fp}$.

$N_c(R)$ represents the number of attribute of one rule. $M_c(R)$ Represents the permitted maximum number of attribute of the front piece of the rule, tp represents the number of samples that meet the front piece of the rule and that are classified correctly. fp Represents the number of samples that meet the front piece of the rule and that are classified wrongly. fn Represents the number of samples that do not meet the front piece of the rule, but meet the conclusion of rule. tn represents the number of samples that not only do not match with the front piece of the rule, but also do not meet the conclusion of rule. The processes of classification rule extraction algorithm are as follows.

Step 1 Initialize artificial fish swarm.

Step 2 The data is divided into training data set and testing data set. Training data set accounts for 80% and testing set account for 20%.

Step 3 Artificial fish swarm algorithm is carried out. Then calculate fitness of the rule and the individual with the largest fitness value is put into bulletin board.

Step 4 If the bulletin board does not change in successive generations, turn to step 5. Otherwise turn to step 3.

Step 5 The generated rules are joined in the rule set. Repeat step one to step five, the rules set is obtained. If classification accuracy can meet the demands, the rule can be used to classify the data of unknown category.

TABLE 1 Rule coding

attribute	x_1	…	x_N
Upper bound	x_1 **_max**	…	x_N **_max**
Lower bound	x_1 **_min**	…	x_N **_min**

4 The experimental simulation and analysis

Iris data set and wine data set of UCI library are used for the experiment, and the data is divided into training data set and test data set, the proposed algorithm is used for mining classification rules corresponding the data set, and the mined rule set is use to classify the corresponding data set. The value of feature attribute is continuous variable and value of category attribute is discrete variable. The data set is shown in Table 2. The scale of swarm is 30, visual range is 2, the maximum iteration times are 50, step length is 0.5, the maximum try time is 30 and crowded factor is 0.8. The classification accuracy of proposed algorithm is compared with PSO algorithm [15] and traditional artificial fish swarm algorithm [16]. Table 3 is the classification accuracy of three different algorithms, which show that the proposed algorithm ha higher classification accuracy than the other two algorithms.

Convergence comparison of artificial fish swarm and proposed scheme on iris data set is shown in Figure 1. The red line represents the proposed algorithm, the blue line represents artificial fish swarm algorithm, the horizon axis represents iteration times and the vertical axis represents fitness value. It can be concluded that the proposed algorithm can converge to the optimal solution fast. Martin Gaddy function, Matyas function, Sumsquares function, Sphere function, Rastrgrin function, Grievank function and Ackly function are used to test the effectiveness of the proposed algorithm. Seven functions are shown in Table 4 and its corresponding parameters are shown in Table 5.

TABLE 2 UCI data set

data set	the number of samples	the number of features	the number of category
iris	150	4	**3**
wine	**178**	**13**	**3**

TABLE 3 Classification accuracy of three different algorithms

algorithm	classification accuracy		the number of rules	
	iris	**wine**	**iris**	**wine**
Proposed scheme	**96.2%**	**94.6%**	**3**	**3**
AFS	**93.10%**	**91.58%**	**3**	**3**
PSO	**93.51%**	**93.6%**	**5**	**4**

TABLE 4 Testing function

Martin Gaddy	$f(x)=(x_1-x_2)^2+((x_1+x_2-10)/3)^2$
Matyas	$f(x)=0.26(x_1^2+x_2^2)-0.48x_1x_2$
Sumsquares	$f(x)=\sum_{i=1}^{n} i\cdot x_i^2$
Sphere	$f(x)=\sum_{i=1}^{D} x_i^2$
Rastrgrin	$f(x)=\sum_{i=1}^{D}(x_i^2-10\cos(2\pi x_i))+10D$
Grievank	$f(x)=\frac{1}{4000}\sum_{i=1}^{D}x_i^2-\prod_{i=1}^{D}\cos\left(\frac{x_i}{\sqrt{i}}\right)+1$
Ackly	$f(x)=-20\exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2})$ $-\exp(\frac{1}{n}\sum_{i=1}^{n}\cos 2\pi x_i)+20+e$

TABLE 5 Parameters of testing function

Function	**dim**	**range**
Martin Gaddy	2	**[-20,20]**
Matyas	2	**[-10,10]**
Sumsquares	30	**[-10,10]**
Sphere	30	**[100,100]**
Rastrgrin	30	**[-5.12,5.12]**
Grievank	30	**[-600,600]**
Ackly	**30**	**[-30,30]**

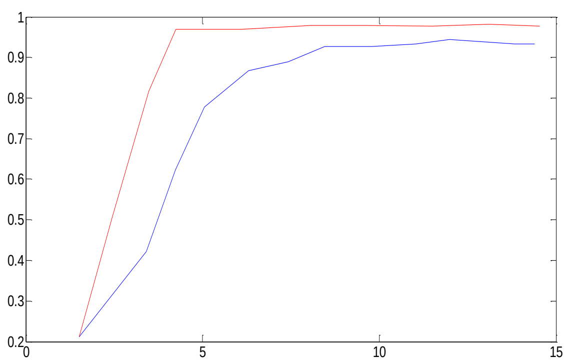

FIGURE 1 Convergence analysis on iris data set.

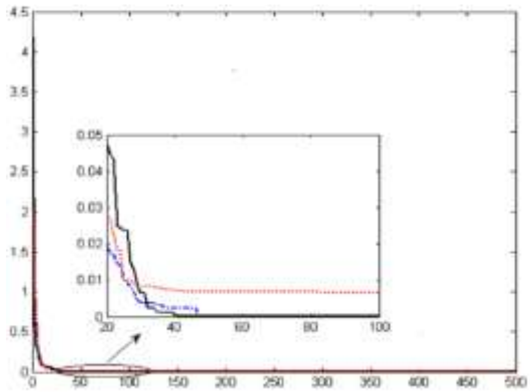

FIGURE 2 Martin Gaddy function

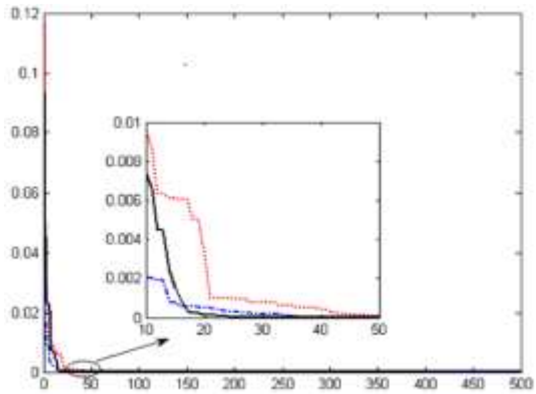

FIGURE 3 Matyas function

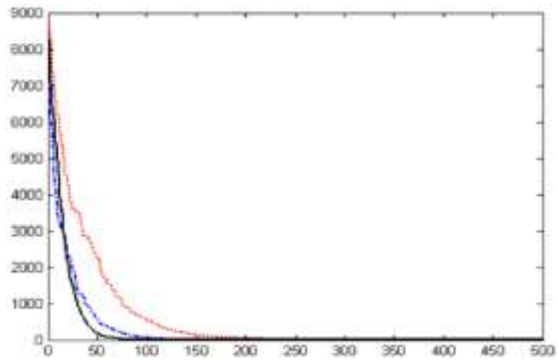

FIGURE 4 Sumsquares function

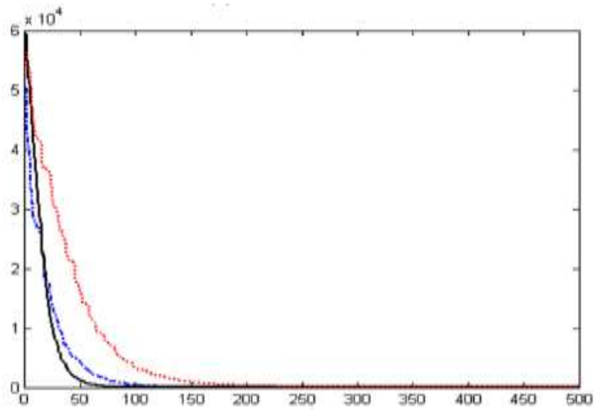

FIGURE 5 Sphere function

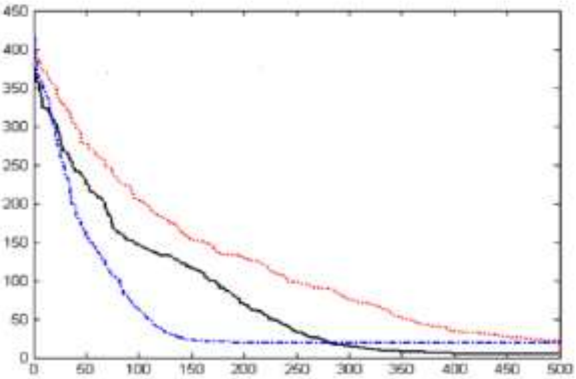

FIGURE 6 Rastrigin function

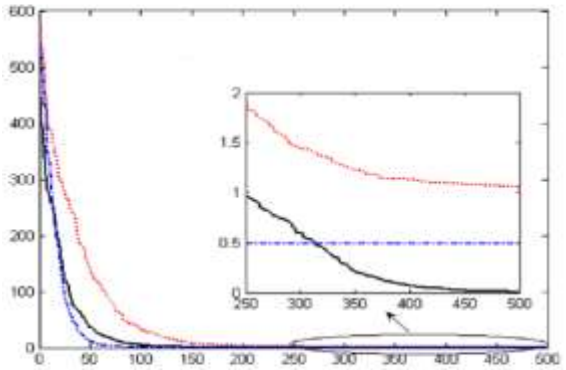

FIGURE 7 Griewank function

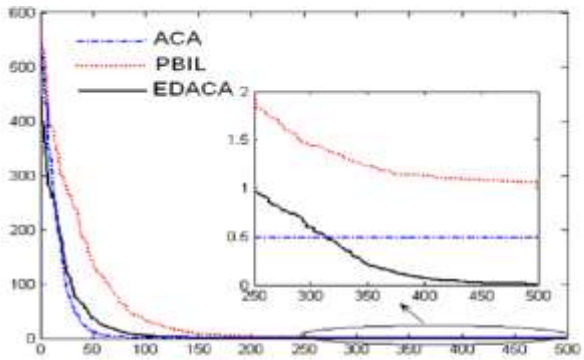

FIGURE 8 Ackly function

From Figure 2 to Figure 8, the blue line represents artificial fish algorithm, the blue line represents PSO algorithm and the black line represents the proposed algorithm. It can be seen that the proposed algorithm has faster convergence speed.

5 Conclusions

An improved artificial fish swarm algorithm is proposed. Then rule mining algorithm based on improved artificial fish swarm is proposed, which includes rule coding, rule evaluation and determination of fitness function. The experiment results show that the proposed algorithm has higher classification accuracy than particle swarm optimization and artificial fish swarm algorithm.

References

[1] Parpinelli R, Lopes H, and Freitas 2002 A Data mining with an ant colony optimization algorithm. *IEEE Transactions on Evolutionary Computation* **6**(4) 321-32

[2] Chan A, Freitas A 2006 A new classification-rule pruning procedure for an ant colony algorithm *Artificial Evolution*Lecture Notes in Computer Science* **3871** 25-36

[3] Freitas A A, Parpinelli R S, Lopes H S 2008 Ant colony algorithms for data classification *Encyclopedia of Information Science and Technology* Information Science Reference 154-9

[4] Galea M, Shen Q 2006 Simultaneous ant colony optimization algorithms for learning linguistic fuzzy rules *Swarm Intelligence in Data Mining* Springer-Verlag 75-99

[5] Salama K M, Abdelbar A M 2010 Extensions to the Ant-Miner classification rule discovery algorithm *The Seventh International Conference on Swarm Intelligence* 43-50

[6] Salama K M, Abdelbar A M, Freitas A A 2011 Multiple pheromone types and other extensions to the Ant-Miner classification rule discovery algorithm *The Eigth International Conference Swarm Intelligence* 149-82

[7] Smaldon J, Freitas A 2006 A new version of the Ant-Miner algorithm discovering unordered rule sets *Genetic and Evolutionary Computation Conference (GECCO)* 43-50

[8] Wang Z, Feng B 2004 Classification rule mining with an improved ant colony algorithm *Advances in Artificial Intelligence Lecture Notes in Computer Science* 357-67

[9] Djenouri Y, Drias H, Habbas Z and Mosteghanemi H 2012 Bees Swarm Optimization for Web Association Rule Mining *International Conferences on Web Intelligence and Intelligent Agent Technology* **3** 142-6

[10] Djenouri Y, Drias H and Chemchem A 2013 A Hybrid Bees Swarm Optimization and Tabu Search Algorithm for Association Rule Mining *World Congress on Nature and Biologically Inspired Computing* 120-5

[11] Olmo J, Romero J, and Ventura S 2011 Using Ant Programming Guided by Grammar for Building Rule-Based Classifiers *IEEE Transactions on Systems, Man, and Cybernetics* B: Cybernetics **41** 1585–99

[12] Olmo J, Romero J, and Ventura S 2012 Classification rule mining using ant programming guided by grammar with multiple pareto fronts *Soft Computing* **16** 2143-63

[13] Otero F, Freitas A, and Johnson C 2013 A New Sequential Covering Strategy for Inducing Classification Rules With Ant Colony Algorithms *IEEE Transactions on Evolutionary Computation* **17**(1) 64-76

[14] Freitas A, Parpinelli R, and Lopes H 2008 Ant colony algorithms for data classification *Encyclopedia of Information Science and Technology* **1** 154-9

[15] Yan Li-ping, Zeng Jian-chao 2007 Classification rules generated by multiple population PSO algorithm *Computer engineering and science* **29**(1) 94-6

[16] Chen Jun-qing, Zhu Wen-xing 2007 Classification rules discovery based on artificial fish algorithm *Journal of Fuzhou university* **35**(1) 25-30

[17] Fan Lilin, Meng Huili 2010 Application Research on Optimal Path Based on Genetic Algorithm *International Journal of Digital Content Technology and its Applications* **4**(8) 199-202

[18] Liu Jiansheng, Liu Jiansheng, Fu Hui, Liao Xinxing 2006 Combination Prediction for Short-term Traffic Flow Based on Artificial Neural Network *The Sixth World Congress on Intelligent Control and Automation* 8659-63

[19] Gu Yuesheng, Liu Yanpei, Xu Jiucheng 2011 Improved Hybrid Intelligent Method for Urban Road Traffic Flow Forecasting based on Chaos-PSO Optimization *International Journal of Advancements in Computing Technology* **3**(7) 282-90

[20] Zhang Yi, Zhang Hu, Liu Zheng, Li Huifang 2011 Crowding Self-Adaptive Genetic Algorithm and Its Application in Facility Layout

Problem", International Journal of Advancements in Computing Technology **4**(7) 203-12

[21] Leu Yungho, Chiu Tzu-I 2011 An Effective Stock Portfolio Trading Strategy using Genetic Algorithms and Weighted Fuzzy Time Series *International Journal of Digital Content Technology and its Applications* **6**(6) 333-44

[22] Yoon Gyung-Mok, Kim Jin, Kim Yong-Hyuk, Moon Byung-Ro 2012 Performance Improvement by Genetic Feature Selection and Adjusting Ratings' Mid-point Value in the Neural Network-Based Recommendation Models *Advances in Information Sciences and Service Sciences* **4**(11) 37-43

Author

Zhaoyin Zhang, 1955.10, Harbin County, Heilongjiang Province, P.R. China

Current position, grades: Associate professor of School of, Heilongjiang University, China.

University studies: received his B.Sc. in Electrical Engineering and Automation from Heilongjiang University of Heilongjiang in China. He received his M.Sc. from Heilongjiang University in China.

Scientific interest: Computer application；Artificial intelligence；Robot；Software engineering.

Publications: more than fifteen papers published in various journals.

Experience: He has teaching experience of thirty-two years, has completed three scientific research projects.