https://doi.org/10.20965/jrm.2018.p0921

Paper:

Improved Artificial Bee Colony Algorithm and its Application in Classification

Haiquan Wang*, Jianhua Wei, Shengjun Wen*, Hongnian Yu***, and Xiguang Zhang***

*Zhongyuan Petersburg Aviation College, Zhongyuan University of Technology
41 Zhongyuan Road, Zhengzhou 450007, China
E-mail: wanghq@zut.edu.cn
**School of Electric and Information Engineering, Zhongyuan University of Technology
41 Zhongyuan Road, Zhengzhou 450007, China
E-mail: 18638923313@163.com
***Faculty of Science and Technology, Bournemouth University
Fern Barrow, Poole, Dorset BH12 5BB, United Kingdom
E-mail: yuh@bournemouth.ac.uk

[Received May 8, 2018; accepted September 26, 2018]

For improving the classification accuracy of the classifier, a novel classification methodology based on artificial bee colony algorithm is proposed for optimal feature and SVM parameters selection. In order to balance the ability of exploration and exploitation of traditional ABC algorithm, improvements are introduced for the generation of initial solution set and onlooker bee stage. The proposed algorithm is applied to four datasets with different attribute characteristics from UCI and efficiency of the algorithm is proved from the results.

Keywords: artificial bee colony algorithm, classifier, SVM, UCI database, wrapper method

1. Introduction

Classification is a data mining function that assigns items in a collection to target categories or classes after obtaining classification model represented as classification rules or mathematical formula [1, 2].

For the classification of data which consists of several samples within different features, the first step is feature selection. Generally speaking, the corresponding researches could be categorized into filter method and wrapper method [3]. The filter method, including minimum-redundancy-maximum-relevance (MRMR) [4], double input symmetrical relevance (DISR) [5], joint mutual information (JMI) [6], conditional infomax feature extraction (CIFE) [7] applies a statistical measure to calculate scores of each feature, then ranks and selects to build model by the scores. For wrapper method, different feature combinations are evaluated according to the predictive classification accuracy on test data with the help of classifier [8]. But the premise is that an efficient global search technique should be adopted for optimal feature selection in such a large feature space. A vast variety of algorithms have been explored and used especially in the intelligent algorithms, work in [9–11] used different genetic algorithm to examine and evaluate simultaneously candidate features. Sequential backward/forward selection methods were used and compared with genetic algorithm (GA) in [12]. The swarm intelligent algorithms such as particle swarm optimization (PSO) [8], binary bat algorithm [13], fire fly algorithm [14], ant colony algorithm (ACO) [15], elitist quantum inspired differential evolution [16] are also applied and have proved to be effective.

Based on the selected feature subset, classifier should be obtained for further recognition. Different methods such as decision tree, naive Bayes [17], extreme learning machine algorithm (ELM) [11], and proposed optimum-path forest method (OPF) [13] has been explored. As a state-of-art classifier, support vector machines (SVM) were frequently used whose quality of generalization and ease of training are far beyond the capacities of the traditional methods [18]. What should be noted is the effectiveness of SVM depends on the selected parameters such as penalty parameter C thus optimization problem about how to select the optimal parameters should be solved. PSO [19], ACO [20], gene selection [21], neural fuzzy inference system [22] were introduced for optimization of SVM with different kernel function.

For improving the performance of classification, limited researches considered the two optimization problems which are the minimum number of selected features and best SVM parameters simultaneously, such as novel gravitational search algorithm (GSA)-SVM method [23], PSO-CSVM model which combined the discrete PSO and the real-valued PSO [24], direct search and features ranking technology [25], ACO-based algorithm [26].

For this complex optimization problem which has large search space, in order to obtain near-global optimum solutions, novel artificial bee colony (ABC) algorithm which is proposed by D. Karaboga [27] is introduced in this paper. It employs fewer control parameters than other swarm intelligence algorithms, and possesses the ability to get out of a local minimum [28]. Meanwhile, considering the existing fact that ABC algorithm is better at exploration but poor at exploitation [29], corresponding improvements for ABC algorithm will be adopted.

This paper is organized as follows. Fundamentals about classification algorithm are introduced in Section 2. In Section 3, the optimization methodology with the proposed improved ABC algorithm (IABC)-SVM classification method is given. Subsequently in Section 4, experimental results for four University of California Irvine (UCI) datasets and discussions are presented. Finally, the conclusion is drawn in Section 5.

2. Classification Algorithm

Based on the pre-processed data including filtering and normalization, optimal features are selected from the original features for simplifying classifier models, and enhancing generalization by reducing over fitting [30], and then SVM based classifier could be applied to classification.

For a training set of instance-label pairs, the objective of SVM is to find hyper-plane or classifier based on the solution of following optimization problem [18]:

$$\begin{cases} \min \dfrac{1}{2}\omega^T\omega + C\sum_{i=1}^{l}\xi_i \\ \text{s.t. } y_i\left(\omega^T \cdot Z_i + b\right) \geq 1-\xi_i, \\ \xi_i \geq 0, \quad i=1,\ldots,l \end{cases} \quad \ldots\ldots \quad (1)$$

Here training vector x_i are mapped into a higher dimensional space by the function Φ as $Z_i = \Phi(x_i)$. $C > 0$ is the penalty parameter of error term.

Then $\omega = \sum_{i=1}^{l} \alpha_i y_i \Phi(x_i)$ and the nonlinear SVM classifier could be constructed as follows:

$$\begin{aligned} f(x) &= \operatorname{sign}(\omega^T\phi(x)+b) \\ &= \operatorname{sign}\left(\sum_{i=1}^{l}\alpha_i y_i K(x_i,x)+b\right) \quad \ldots\ldots \quad (2) \end{aligned}$$

Considering the fact that the Gaussian kernel can approximate most kernel functions if the parameter is chosen appropriately, this paper will focus on it and the form could be expressed as follows:

$$K(x_i,x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \quad \ldots\ldots \quad (3)$$

where γ is the kernel parameter [21].

Obviously, the two parameters of SVM classifier, penalty parameter C and the kernel parameter γ, should be optimized to obtain the optimal predictive accuracy.

3. Optimization Methodology for Classification

For data classification problem based on wrapper feature selection method and SVM classifier, two objectives which are minimum number of selected features and best SVM parameters should be satisfied, and improved artificial bee colony (IABC) algorithm will be introduced to solve the problem in this paper.

3.1. ABC Algorithm

As one of the most recent swarm intelligence approaches, ABC algorithm simulates the foraging behavior of honeybee swarm where the possible solution is represented by the position of food source in ABC algorithm, the quality of associated solution is equal to the nectar amount of food source [27]. The algorithm begins with randomly distributed initial population generation:

$$X_{i,j} = X_{min,j} + rand(0,1)(X_{max,j} - X_{min,j}) \quad \ldots \quad (4)$$

where i is the number of solutions as well as food source. $X_{min,j}$ and $X_{max,j}$ are the lower and upper limits of the j-th dimension of the i-th solution respectively. And then repeated search cycles will be executed to generate the population of solutions. During the cycles, the employed bee probabilistically produces a neighbor food source $X_{i,j}$ around $X_{a,j}$ as Eq. (5) and updates the solution based on its fitness value.

$$X_{i,j} = X_{a,j} + (X_{k,j} - X_{a,j})\varphi_{i,j} \quad (a \neq k) \quad \ldots \quad (5)$$

where $\varphi_{i,j}$ is an random number between -1 and 1. And then onlooker bee chooses a food source based on the probability calculated with roulette method and so on, and updates new solutions around it based on Eq. (5). If a solution does not improve for several iterations, the solution will be abandoned, the associated employed bee becomes scout bee and random search will be performed [31].

3.2. Improved ABC Algorithm

ABC algorithm is relatively simple and it has been proved to be able to produce good results at a low computational cost. But the algorithm still faces the problem that it is good at exploration, but poor at exploitation where the two indices contradict each other. In order to balance the two indices to obtain more accuracy solutions, two improvements will be proposed.

As mentioned above, the initial solution set is generated randomly as Eq. (6), the solutions are not uniformly distributed, or can't cover the search space of the optimization problem, the quality of optimal solution and the convergence of the algorithm will be affected. Thus the variable of $rand(0,1)$ in Eq. (4) could be replaced by a set of numbers which are created by extended integer tent maps method [32] as shown in Eq. (6).

When i is even,

$$r_{i+1} = \begin{cases} 2r_i + 1 & r_i \in [0, 2^{k-1}] \\ 2\left(2^k - 1 - r_i\right) & r_i \in (2^{k-1}, 2^k - 1] \end{cases} \quad . \quad (6)$$

when i is odd,

$$r_{i+1} = \begin{cases} 2r_i & r_i \in [0, 2^{k-1}] \\ 2\left(2^k - 1 - r_i\right) + 1 & r_i \in (2^{k-1}, 2^k - 1] \end{cases}$$

The range of generated sequential is $[0, 2^k + 1]$ and it could

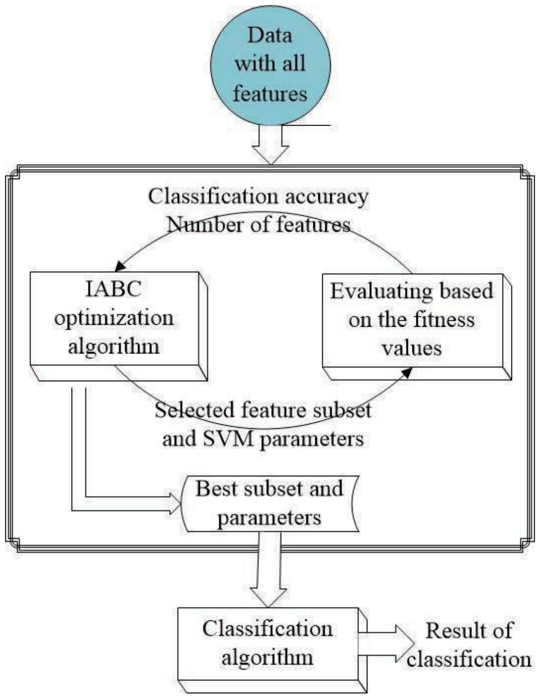

Fig. 1. Flowchart of classification based on IABC.

be converted to $[0,1]$ with $y = r/r_{max}$ where r_{max} is the generated maximum number.

Obviously the periodic points, which are chaotic and random distributed, are generated, the corresponding diversity of the initial population could be improved.

Meanwhile, considering the task of onlooker bee is to exploit local search region around the selected solution and obtain more precise solutions from limited solution space, $X_{Gbest,j}$ which is j-th dimension of the best solution so far is introduced to replace the term in Eq. (5):

$$X_{i,j} = X_{a,j} + (X_{Gbest,j} - X_{a,j})\varphi_{i,j} \quad (a \neq i) \quad . \quad . \quad (7)$$

where $X_{a,j}$ is a selected solution different from $X_{i,j}$. $\varphi_{i,j}$ is a random number between -1 and 1, which could be selected to adjust the evolution direction and the distance between $X_{a,j}$ and $X_{Gbest,j}$. With Eq. (7), the searching process could be concentrated around the optimal solution and the ability of exploitation could be improved.

3.3. IABC-SVM Classification

In order to obtain higher classification accuracy with less selected features, IABC will be introduced and combined with traditional SVM algorithm. The flowchart of the proposed algorithm is shown as **Fig. 1**, where the data set corresponding to the pre-selected features is used to train SVM classifier with the pre-selected parameters C and gamma firstly, then the solution will be evaluated and adjusted until obtaining the optimal performance of classification. Finally, with the selected feature subset and parameters, the classifier could be trained and the classification task will be completed.

4. Experiments and Results

In order to verify the effectiveness of the proposed algorithm, corresponding experiments and analyses will be done.

Four datasets with different attributes characteristics as shown in **Table 1** are selected from UCI datasets, and then scaled to $[-1,1]$. Subsequently, the training data and test data are selected randomly at the ratio of 3 : 1 and 10 fold cross-validation technique is used to evaluate predictive classification models.

The parameters of ABC and IABC used for optimization are set as follows: population size is 80, the number of optimized parameters $Dim = 2 + x$ where x represents the number of feature of different dataset, maximum iterations is set to be 200, and the limit for the introduction of scout bee is 20. The upper and lower bounds of the optimized SVM parameters are set to be 20 and 0, the bounds of selected features are $[0,1]$ where the optimized value which is no less than 0.5 represents selected and the one whose value is smaller than 0.5 isn't selected. The evaluation criterion of optimization problem could be described as Eq. (8):

$$fit = \omega \cdot C_{ac} + (1-\omega) \cdot \frac{1}{N_{sf}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where N_{sf} is the number of selected features and C_{ac} represents the cross-validation results with the selected features and SVM parameters, and the value of weight used to balance the proportion between the two indices is set to be 0.95.

For the four datasets, **Figs. 2–5** show the comparison between the results of fitness values during optimization process with ABC and IABC respectively. Obviously, ABC and improved ABC algorithm could fulfill the task of optimal SVM parameters selection and feature subset selection for four datasets. And the improved ABC algorithm possesses higher classification accuracy and faster convergence speed than traditional ABC algorithm for the same dataset, where the optimal solution could be obtained in the 61st iteration which is 124 iterations ahead of traditional ABC algorithm for Abalone dataset, and the fitness values is increased by 0.98%, 3.4%, 0.23%, and 0.68% respectively for the four datasets.

The related characteristic of selected feature/SVM parameters/classification accuracy are shown in **Tables 2** and **3**. The feature number could be reduced by 27.3% of the original dataset at least. Based on the fact that fitness values are combined cross-validation values with feature number as shown in Eq. (8), maximum fitness values which is the optimization objective doesn't mean possessing more feature number. So from **Tables 2** and **3**, for WDBC/Abalone datasets, there are more features selected based on IABC algorithm compared with ABC. According to **Table 4**, the proposed classification method with IABC possess higher classification accuracy and no more feature numbers than the four filter methods for all the datasets.

Table 1. UCI dataset.

	Heart	WDBC	Red wine quality	Abalone
Number of instances	303	569	1599	4177
Number of attributes	14	32	11	8
Number of classes	2	2	11	28
Attribute characteristics	Categorical, Integer, Real	Real	Real	Categorical, Integer, Real

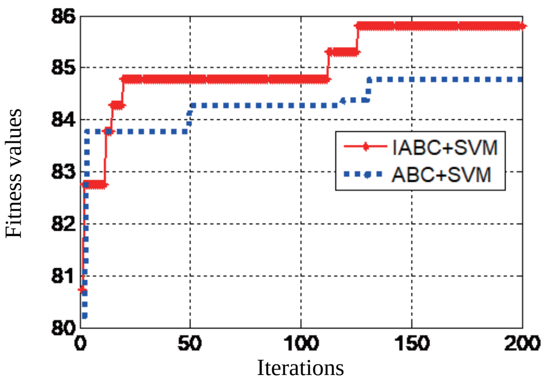

Fig. 2. Heart dataset.

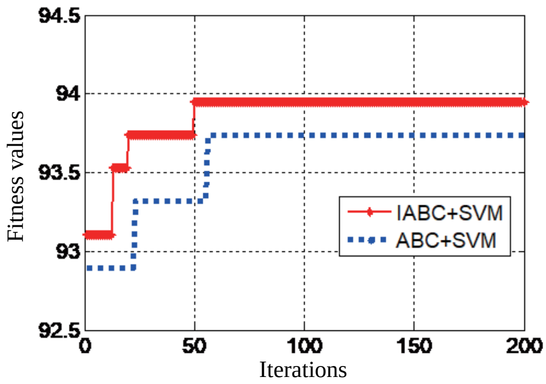

Fig. 3. WDBC dataset.

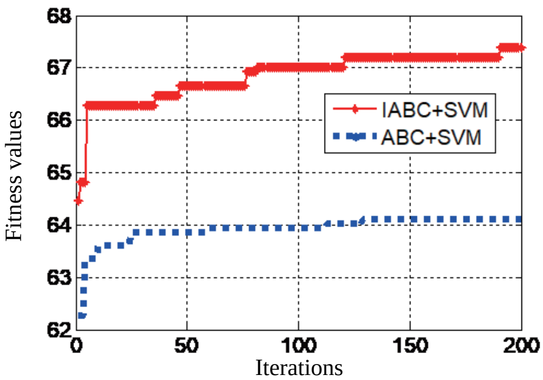

Fig. 4. Red wine quality dataset.

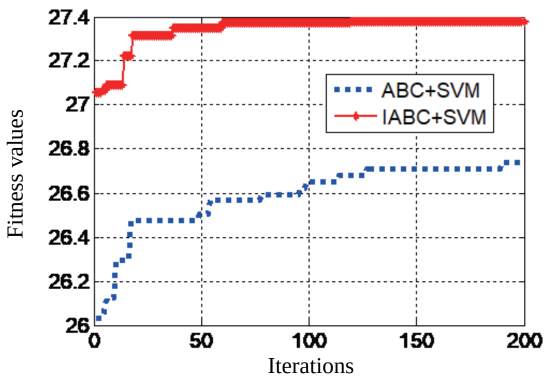

Fig. 5. Abalone dataset.

Table 2. Classification results with ABC-SVM.

Methods	Heart	WDBC	Red wine quality	Abalone
Number of selected feature	6	11	8	4
SVM parameters	1.47/6.35	12.55/0.36	1.09/18.99	16.02/7.21
Classification accuracy	83.52% (76/91)	97.32% (145/149)	57.75% (298/516)	29.55% (302/1022)

Table 3. Classification results with IABC-SVM.

Methods	Heart	WDBC	Red wine quality	Abalone
Number of selected feature	6	13	8	5
SVM parameters	3.24/9.29	11.44/0.19	3.67/14.56	4.76/8.09
Classification accuracy	85.71% (79/91)	98.66% (147/149)	59.11% (306/516)	31.01% (317/1022)
Improve	2.19%	1.34%	1.36%	1.46%

Table 4. Classification results with different filter method.

Methods	Heart	WDBC	Red wine quality	Abalone
mRMR	81.32% (74/91)	97.32% (145/149)	56.01% (289/516)	27.98% (286/1022)
JMI	80.22% (73/91)	97.32% (145/149)	55.43% (286/516)	27.98% (286/1022)
DISR	81.32% (74/91)	96.64% (144/149)	52.91% (273/516)	27.98% (286/1022)
CIFE	75.82% (69/91)	95.30% (142/149)	50.78% (262/516)	26.81% (274/1022)
Remarks	7 features selected from 13 features	13 features selected from 30 features	8 features selected from 11 features	5 features selected from 8 features

5. Conclusion

In this paper, ABC algorithm was applied to the selection of optimal feature subset and SVM parameters, and several improvements were proposed to balance the ability of exploration and exploitation of ABC algorithm. The experiment results demonstrated that the proposed algorithm could achieve higher accuracy compared with the traditional ABC-SVM method as well as other filter methods. But the problems that it will cost more time to perform the task of optimization should be solved for future work. Also multi-objective ABC algorithm could be applied to obtain better optimization results.

Acknowledgements

The authors acknowledge the support of Science & Technology program of Henan province (172102210588), and Program of Educational committee of Henan province (18A120005).

References:

[1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smith, "From Data Mining to Knowledge Discovery: An Overview, Advances in Knowledge Discovery and Data Mining," MIT Press, pp. 1-34, 1996.

[2] F. Otero, A. Freitas, and C. Johnson, "A new sequential covering strategy for inducing classification rules with ant colony algorithms," IEEE Trans. on Evolutionary Computation, Vol.17, No.1, pp. 64-76, 2013.

[3] H. Liu and H. Motoda, "Feature selection for knowledge discovery and data mining," Kluwer Academic, 1998.

[4] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.27, No.8, pp. 1226-1238, 2005.

[5] P. E. Meyer and G. Bontempi, "On the use of variable complementarity for feature selection in cancer classification," Proc. of European workshop on applications of evolutionary computing: Evo Workshops, pp. 91-102, 2006.

[6] H. H. Yang and J. Moody, "Feature selection based on joint mutual information," Proc. of Int. ICSC Symp. on Advances in Intelligent Data Analysis, pp. 1-8, 1999.

[7] D. Lin and X. Tang, "Conditional Infomax Learning: An Integrated Framework for Feature Extraction and Fusion," Proc. 9th European Conf. on Computer Vision, pp. 68-82, 2006.

[8] C. T. Tran, M. Zhang, P. Andreae, and B. Xue, "A Wrapper Feature Selection Approach to Classification with Missing Data," Lecture Notes in Computer Science, Vol.9597, pp. 685-700, 2016.

[9] A Wang, N. An, G. Chen, J. Yang, and L. Li, "Incremental wrapper based gene selection with markov blanket," IEEE Int. Conf. on Bioinformatics and Biomedicine, pp. 74-79, 2014.

[10] O. Soufan, D. Kleftogiannis, P. Kalnis, and V. B. Bajic, "DWFS: A Wrapper Feature Selection Tool Based on a Parallel Genetic Algorithm," PLoS One, Vol.10, No.2, 2015.

[11] D. Chyzhyk, A. Savio, and M. Grana, "Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI," Neurocomputing, Vol.128, pp. 73-80, 2014.

[12] R. Panthonga and A. Srivihokb, "Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm," Procedia Computer Science, Vol.72, pp. 162-169, 2015.

[13] D. Rodrigues et al., "A wrapper approach for feature selection based on Bat Algorithm and Optimum-Path Forest, Expert Systems with Applications," Vol.41, Issue 5, pp. 2250-2258, 2014.

[14] H. Banati and M. Bajaj, "Fire Fly Based Feature Selection Approach," Int. J. of Computer Science Issues, Vol.8, No.2, pp. 473-480, 2011.

[15] S. Kashef and H. Nezamabadi-pour, "An advanced ACO algorithm for feature subset selection," Neurocomputing, Vol.147, pp. 271-279, 2015.

[16] V. Srikrishna, R. Ghosh, V. Ravi, and K. Deb, "Elitist Quantum-Inspired Differential Evolution Based Wrapper for Feature Subset Selection," Lecture Notes in Computer Science, Vol.9426, pp. 113-124, 2015.

[17] R. Panthonga and A. Srivihokb, "Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm," Procedia Computer Science, Vol.72, pp. 162-169, 2015.

[18] P.-N. Tan and M. Steinbach, "Introduction to Data Mining," Addison-Wesley Companion Book, 2006.

[19] G. Gu, R. Hu, and Y. Li, "Study on Identification of Damage to Wind Turbine Blade Based on Support Vector Machine and Particle Swarm Optimization," J. Robot. Mechatron., Vol.27, No.3, pp. 244-250, 2015.

[20] H. Zhu and X. Li, "Research on a New Method based on Improved ACO Algorithm and SVM Model for Data Classification," Int. J. of Database Theory and Application, Vol.9, No.1, pp. 217-226, 2016.

[21] H. Gou, T. Shi, L. Yan, and J. Xiao, "Gait and Posture Analysis Method Based on Genetic Algorithm and Support Vector Machines with Acceleration Data," J. Robot. Mechatron., Vol.28, pp. 418-424, 2016.

[22] H. Ceylan and H. K. Ozturk, "Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach," Energy Convers Management, Vol.45, pp. 2525-2537, 2004.

[23] S. Sarafrazi and H. Nezamabadi-pour, "Facing the classification of binary problems with a GSA-SVM hybrid system," Mathematical and Computer Modelling, Vol.57, Issues 1-2, pp. 270-278, 2013.

[24] C.-L. Huang and J.-F. Dun, "A distributed PSO-CSVM hybrid system with features selection and parameter optimization," Applied Soft Computing, Vol.8, Issue 4, pp. 1381-1391, 2008.

[25] L. Zhou, K. K. Lai, and J. Yen, "Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimization," Int. J. of Systems Science, Vol.45, Issue 3, pp. 241-253, 2014.

[26] F. Samadzadegan, H. Hasania, and T. Schenk, "Simultaneous Feature Selection and SVM Parameter Determination in Classification of Hyperspectral Imagery Using Ant Colony Optimization," Canadian J. of Remote Sensing, Vol.38, Issue 2, pp. 139-156, 2012.

[27] D. Karaboga and B. Akay, "A survey: algorithms simulating bee swarm intelligence," Artificial Intelligence Review, Vol.31, No.1, pp. 68-85, 2009.

[28] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," J. of Global optimization, Vol.39, pp. 459-471, 2007.

[29] P. Dass, S. S. Jadon, H. Sharma, J. C. Bansal, and K. E. Nygard, "Hybridisation of classical unidimensional search with ABC to improve exploitation capability," Int. J. of Artificial Intelligence and Soft Computing, Vol.5, No.2, pp. 151-164, 2015.

[30] H. Liu and H. Motoda, "Feature selection for knowledge discovery and data mining," Kluwer Academic, 1998.

[31] Q.-K. Pan, "An effective co-evolutionary artificial bee colony algorithm for steelmaking-continuous casting scheduling," European J. of Operational Research, Vol.250, Issue 3, pp. 702-714, 2016.

[32] J. Liu, "Extended integer tent maps and dynamic hash function," J. of Communications, Vol.31, No.5, pp. 51-59, 2010.

Name:
Haiquan Wang

Affiliation:
Zhongyuan Petersburg Aviation College,
Zhongyuan University of Technology

Address:
41 Zhongyuan Road, Zhengzhou 450007, China

Brief Biographical History:
2009 Recevied Ph.D. from Northwestern Polytechnical University
2009-2013 Lecturer, Zhongyuan University of Technology
2013- Associate Professor, Zhongyuan University of Technology

Main Works:
- Intelligent computing and optimization
- Nonlinear systems
- Robust control

Name:
Hongnian Yu

Affiliation:
Faculty of Science and Technology,
Bournemouth University

Address:
Fern Barrow, Poole, Dorset BH12 5BB, United Kingdom

Brief Biographical History:
1994 Recevied Ph.D. from London University
1999-2002 Lecture, Exeter University
2002-2004 Senior Lecture, Bradford University
2004- Professor, Bournemouth University

Main Works:
- Control of non-linear systems
- Intelligent algorithm

Name:
Jianhua Wei

Affiliation:
School of Electric and Information Engineering,
Zhongyuan University of Technology

Address:
41 Zhongyuan Road, Zhengzhou 450007, China

Brief Biographical History:
2016 Received B.E. degree in Electrical Engineering and its Automation from Zhengzhou Institute of Aeronautical Industry Management
2017- Master Course Student, School of Electric and Information Engineering, Zhongyuan University of Technology

Main Works:
- Intelligent computing
- Path optimization
- Nonlinear systems
- Robust control

Name:
Xiguang Zhang

Affiliation:
Zhongyuan Petersburg Aviation College,
Zhongyuan University of Technology

Address:
41 Zhongyuan Road, Zhengzhou 450007, China

Brief Biographical History:
2006 Recevied M.Sc. from Beijing Institute of Technology
2009-2015 Lecturer, Zhongyuan University of Technology
2015- Associate Professor, Zhongyuan University of Technology

Main Works:
- Intelligent computing and optimization

Name:
Shengjun Wen

Affiliation:
Zhongyuan Petersburg Aviation College,
Zhongyuan University of Technology

Address:
41 Zhongyuan Road, Zhengzhou 450007, China

Brief Biographical History:
2011 Recevied Ph.D. from Tokyo University of Agriculture and Technology
2011-2013 Lecturer, Zhongyuan University of Technology
2013- Associate Professor, Zhongyuan University of Technology

Main Works:
- Nonlinear systems