Proceedings of the 2017 International Conference on Advanced Mechatronic Systems, Xiamen, China, December 6-9, 2017

Human Activity Recognition Based on Improved Artificial Bee Colony Algorithm

Xuekai Sun
School of Electric and Information Engineering
Zhongyuan University of Technology
Zhengzhou,China
Email:sunxuekai1226@163.com

Haiquan Wang
Zhongyuan Petersburg Aviation College
Zhongyuan University of Technology
Zhengzhou,China
Email:wanghq@zut.edu.cn

Fanbing Zhu
School of Electric and Information Engineering
Zhongyuan University of Technology
Zhengzhou,China
Email:kangkanglaoda@163.com

Abstract—**Support vector machine is a new machine learning method, and its classification performance mainly depends on the selection of related parameters. An improved artificial colony algorithm is proposed to optimize the parameters of SVM and applied to human activity recognition. Compared with other optimization algorithms including basic artificial colony algorithm, genetic algorithm and particle swarm algorithm on standard datasets, the proposed algorithm can acquire higher classification precision. Compared with artificial colony algorithm based on all dimensional search, the improved algorithm costs less running time. The proposed method is used as the classifier of human activity and a high classification precision is acquired.**

Keywords*—*Improved artificial bee colony algorithm, Support vector machine, Parameter optimization, Activity recognition.

I. Introduction

China has the largest population in the world. The number of people aged 60 or older in 2014 was 21.2 million, accounting for 15.5 percent of the total population. This proportion will reach at 35 percent by 2050. The huge population of the elderly has given rise to a great demand for pension services. With the development of Internet technology,a new topic of family health care for elderly people based on the Internet matters has been paid attention to. The technology is based on the Internet matters, using wearable devices to collect data about the everyday life of the elderly, judge the current health state, and track the changes of the physical state, so as to detect adverse trends in time and take corresponding treatment measures quickly. Obviously, this technology can effectively solve the problem of the elderlys health care, and has very important theoretical and practical significance.

This paper use support vector machine to classify human activity data. Support vector machine (SVM) [1],[2] is a new method of machine learning, its basis is the statistical learning theory created by Vapnik [3],[4]. It can solve the practical problems of small sample which is nonlinear, and can also be widely used in 3D recognition [5], face image recognition [6], gene classification [7] and other fields. When the support vector machine is used for classification,model parameters have a great influence on classification accuracy.How to select model parameters to optimize the performance of the model is a hot issue in the field of support vector machine research.Current common-used model parameter optimization methods include genetic algorithm [8],[9], gradient descent method[10], simulated annealing algorithm[11],[12], particle swarm optimization algorithm[13] and ant colony optimization algorithm [14]. In the process of optimization, the above methods can get into local optimal solutions in different degrees,which can cause that the performance of the classification model cannot be maximized. Literature [15] proposed an multidimensional artificial bee colony (MABC), the algorithm is very good to keep the balance on the width and depth of mining search,and effectively avoid getting into a local optimal solution,but the method for selection of the full dimensional search strategy can greatly increase the cost of time.

According to the above analysis: Firstly, an improved artificial swarm algorithm (IABC) is proposed in this paper. Then, experiments on the standard data sets and the measured data were conducted, the results were analysised and show that the IABC has higher classification accuracy and better convergence, compared to the basicartificial colony algorithm [16], genetic algorithm and grid search algorithm; And the cost of the IABC is less in the case of non-loss classification accuracy, compared with MABC algorithm. At last, a conclusion was drawed: the experiments facts proved the feasibility and effectiveness of IABC algorithm.

II. The Analysis and Improvement of ABC

A. *Concept of Basic ABC*

Artificial bee colony algorithm (ABC) is a bionic algorithm proposed by Turkish scholar Karaboga in 2005 [17]. The algorithm mimicked the swarm's foraging behavior and divided the bees into three different types of work: gathers, observers, and scouts. Gathers and observers account for half of the population respectively, and a honey source corresponds to a bee. If a source has been abandoned due to the reaching of threshold value of the search number, then the scouts will randomly find a new source to replace the source.

It is assumed that there are food sources of N number, and each food source has D components. In the initial phase of ABC algorithm, the initial feasible solution is randomly generated by formula:

$$x_{id} = x_{idmin} + rand * (x_{idmax} - x_{idmin}) \qquad (1)$$

After initialization, the sources are searched by gathers according to formula (2). If the new source is more fit than the old one, the new will replace the old source.

$$x_{id} = x_{id} + rand(-1, 1) * (x_{id} - x_{jd}) \tag{2}$$

After the collecting of food source, the observers would choose the food source that should be followed according to the probability calculated by the following formula.

$$P_i = \frac{fit_i}{\sum_{k=1}^{N} fit_k} \tag{3}$$

Likewise,observer search for the food source that is followed through formula (2). Finally, update records of each food source are judged, if the number of updating of the source reaches the threshold, then the source is abandoned, and scouts generate a new source instead of abandoned according to the formula(1). ABC algorithm has completed one iteration at this time.

B. Improvement of ABC

In the basic ABC algorithm, gathers and observers randomly select one dimensional of a food source to search by formula (2). If there is a better solution on this dimension, but the next iteration will randomly select other one dimension, which can result in that a more optimal solution of this dimension does not be found, and then this source can be abandoned due to reaching the threshold. This will make the algorithm miss a lot of opportunities of finding the global optimal solution, which will increase the convergence time of the algorithm and also influence the accuracy of the final solution. For the ABC that adopts strategy of searching the full dimensional of each source, although to some extent, solve the above problems, but at the same time, it has searched some relative lower value dimension, which can lead to the increasing of the time cost of the algorithm.

Therefore,this paper proposes an improved multi-dimensional search strategy. The algorithm search each dimension of the sources for greedy choice when first iteration. If the try of the current dimension is successful, on the basis of retaining the current dimension correction, the algorithm save the dimension to an external document, and then go to the choice of next dimension. If the try is failing, the correction will not be retained and external documents also do not store the dimensions at the same time. Suppose that the food source for the current gathers or observers is $x_i = (x_{i1}, x_{i2}, , x_{ij}, , x_{iD})$, the bees search for new food source $x_i = (x_{i1}, x_{i2}, , x_{ij}, , x_{iD})$ near the food source. The specific steps are as follows:

First, the update flag bit is initialized as, $flag = 0$,and Dim is used to store the valuable dimensions in each food source, and Num represents the number of the valuable dimensions. In the first iteration, each dimension of the food source was performed as follows:

(1) Each dimension of the solution is searched according to the following formula: $x_{id} = x_{id} + rand(-1, 1) * (x_{id} - x_{jd})$,where $d = 1, 2, , D$,and $j = rand(N)||j \neq i$.

(2) Determine x_{id} whether to exceed the upper and lower limits of the dimension $[x_{idmin}, x_{idmax}]$,and adjust if it exceeds:if $x_{id} < x_{idmin}$,then let $x_{id} = x_{idmin}$;if $x_{id} > x_{idmax}$,then let $x_{id} = x_{idmax}$.

(3) To determine the fitness of the new solution, if the new solution is superior to the old one, the new solution is retained and the mark bit is updated: if $f(x_j) > f(x_i)$,then let $x_{id} = x_{jd}$,and make $Num = Num + 1$,$Dim(i, Num) = d$,$flag = 1$.

According to the above steps, after modified operation of each dimension, if flag is 1,then in all dimensions of correction,which means that there was successful correction of one dimension at least in the corrections of all dimension. If $flag$ is zero,the explore is failed.

After all of the above steps are completed, the valuable dimensions in each food source are stored in the Dim. In the next iteration, the above operations are performed on the stored dimensions in the Dim. After many iterations, the numbers of storage dimension will be reduced to 0.If $trial(i) < Limit$,the above operations will be performed anew for each dimension of the food sources.

III. The Selection of Support Vector Machine Parameter

When facing the nonlinear classification, support vector machine(SVM) transforms problem of the nonlinear classification into the linear classification problem of a high dimensional space through the high dimension space transform,and introducing Lagrangian function to solve extremum problem of quadratic function under the inequality constraints. Before using support vector machines for data classification, should select the kernel function type,penalty factor C and kernel function parameters. In this paper, the RBF radial basis function is used as the kernel function,as well as the width parameter g in RBF kernel function and the penalty factor C are optimization objective.

In order to illustrate the effect of penalty factor C and parameter g on model performance, table 1 and table 2 show the classification results of human activity data under different parameters. The penalty factor C plays a role in controlling the penalty of the wrong sub-sample, thus realizing the tradeoff between the proportion of the wrong subsample and the algorithm complexity. When C is lesser, the support vector machine has high tolerance of error, and the accuracy of the model will decrease, but its promotion degree is enhanced. When C is larger, the accuracy of the model is improved, but its complexity increases, so the promotion ability is also reduced. The selection of appropriate penalty factors is critical to the classification performance of support vector machines. Parameter g determines the complexity

TABLE I. Classification Results When C=2

g	Classified Accuracy/%
0.1	51.43
1	35.24
10	17.62
100	17.62

of data samples in high-dimensional feature space, which is called spatial dimension. This dimension defines the maximum VC dimension of the linear categorical surface that can be constructed in this space.The linear classification surface will be more complex with the higher the dimension, at the same

TABLE II. CLASSIFICATION RESULTS WHEN G=2

C	Classified Accuracy/%
0.1	17.62
1	26.67
10	27.14
100	27.14

time,the empirical risk will decrease and the confidence range will increase, and vice versa.As can be seen from table I and table II,penalty factor C and parameter g have significant influence on classification accuracy. Therefore, in order to obtain a SVM which has good performance,appropriate width parameters g should be chose firstly, and then look for the right punishment factor C in the decided feature space to makes the model fitting ability and generalization ability get the best combination.

IV. THE ANALYSIS OF SIMULATION RESULT

Using the IABC algorithm to optimize the penalty factor C and width parameter g of SVM,the relevant parameters and fitness functions are set as follows:

(1) Initialize the control parameters of the algorithm,including: population number N,cycle limit G and food source update threshold Limit.

(2) Set the fitness function in the algorithm.Because the purpose of using the algorithm to optimize the support vector machine is to obtain higher classification accuracy,so classification accuracy is used to determine the fitness function of the algorithm.

(3) Initialize the search scope of parameters. Known from the analysis of previous, penalty factor C and width parameter g will produce great influence on the performance of the SVM classification. So it can help get a better classification results to determine the search range of parameters in advance.

After the setting of the parameters and fitness, the model parameters of SVM are optimized by IABC algorithm. In order to verify the validity of IABC algorithm, the three sets of data sets of UCI standard data set are now used for training and test verification. Table 3 shows the names, dimensions and other information of data sets. Firstly, the IABC algorithm is compared vertically with ABC and MABC algorithms. And the IABC is compared parallelly with the grid search (GS) algorithm, genetic algorithm (GA) and particle swarm algorithm (PSO).

TABLE III. THE SPECIFICATION OF UCI TEST DATA SETS

Data sets	Data dimension	Training samples number	Test samples number
Wine	13	120	58
BreastTissue	9	80	26
Glass	9	140	74

A. The comparison before and after the improvement

The SVM model parameters of the three sets of samples were optimized respectively by IABC,MABC and ABC, and the test results were compared.The number of food sources of the algorithm is set as $N = 20$, and the threshold Limit of food source update is set as $Limit = 50$, and the number of iterations is set as $G = 500$, and the search range of SVM parameters is $[2e - 10, 2e + 10]$. Calculate the average results of the 10 experiments.Table IV shows the classification accuracy of each algorithm under different data sets.Table V is their running time. The data from table 4 and table 5 shows that the classification accuracy of SVM whose parameters are optimized by IABC and MABC is higher than the classification accuracy of ABC, and the running time of IABC algorithm is less by nearly one fifth than MABC. It shows that the IABC algorithm has a better convergence than ABC, as well as without the loss of classification accuracy, IABC reduces the running time of the algorithm compared with MABC.

TABLE IV. TEST RESULTS OF LONGITUDINAL COMPARISON OF EXPERIMENTS

Data sets	ABC/%	MABC/%	IABC/%
Wine	94.8	96.5	95.9
BreastTissue	67.7	73.1	72.5
Glass	64.2	66.8	67.6

TABLE V. THE RUNNING TIME OF LONGITUDINAL COMPARISON

Data sets	ABC	MABC	IABC
Wine	93.26	176.16	143.84
BreastTissue	82.89	151.02	120.27
Glass	180.57	332.35	265.44

B. The comparison with other algorithms

In order to further verify the performance of the improved algorithm, the IABC algorithm was compared with GS,GA and PSO algorithms, and the average results were carried out on 10 experiments. Table 6 shows the classification accuracy of four algorithms of different data set.

Can be seen from table 6, regardless of the amount of categories and dimension of the data sets, SVM model optimized by IABC has a higher classification accuracy than other algorithms. Therefore, the SVM model optimized by IABC combine the searching of local optimal solution and the global optimal solution, which shows a better convergence.

TABLE VI. TEST RESULTS OF HORIZONTAL CONTRAST TEST

Data sets	IABC/%	GS/%	GA/%	PSO/%
Wine	95.9	94.5	93.7	93.1
BreastTissue	72.5	69.2	63.5	65.3
Glass	67.6	63.5	62.7	63.1

V. INSTANCE ANALYSIS

A. Design of experimental scheme

This method is used to classify the measured data of human activities to verify the effectiveness of the improved algorithm.Participants must wear a smart watches equipped with acceleration,temperature,and height sensors. Data collected from the acceleration sensor, includes the acceleration of X,Y,and Z directions. The data from the sensors of temperature and height includes the experimenters surface temperatures and the distance between the experimenters and the ground.

The sampling frequency is set as 52Hz, and collect data of seven different human activities.Activities detailed information is shown in table VII.

TABLE VII. ACTIVITY DATA INFORMATION

Activity category label	specific description
1	Use computer to work or study
2	Stand, walk, and go up/down the stairs
3	Stand and still
4	Walk
5	go up/down the stairs
6	Talk to someone while walking
7	Talk to someone while standing

B. Comparison experiment of Activity data classification

The experiment was conducted using the MATLAB 2012b platform and using LIBSVM as the training and testing tool, which was carried out on the computer of CPU 3.3GHz and memory 2 GB.In this paper, we extracted 100 groups of samples each category, a total of 700 groups. Then 490 samples are randomly selected as training sets, and the remaining 210 samples were collected as test sets.

The results of parametric optimization by different algorithms are shown in Fig.1.The optimization process of GS, GA, PSO, ABC, MABC and IABC algorithms are respectively represented by a,b,c,d,e and f. The test sets are respectively put into SVM models obtained by different algorithms,and their classification accuracy is shown in table VIII. From figure 1 and table 8,the degree of fitting of IABC and MABC is better than other algorithms,and their convergence speed is quite fast,converge to the optimal value in the fifth and sixth generation respectively.But as mentioned above, under the condition that the number of iterations is 500,MABC algorithm optimization process took about 45.6 minutes, while the IABC algorithm optimization process took 33.3 minutes. The latter saves for nearly a third of the time than the former.The experimental results show that the SVM model whose parameters are optimized by IABC has higher classification accuracy.MABC has similar performance with IABC, but is inferior in running time.

TABLE VIII. DATA CLASSIFICATION RESULTS UNDER DIFFERENT ALGORITHMS

Algorithms	Classifying accuracy/%
GS	79.04
GA	82.86
PSO	81.13
ABC	83.29
MABC	86.57
IABC	85.93

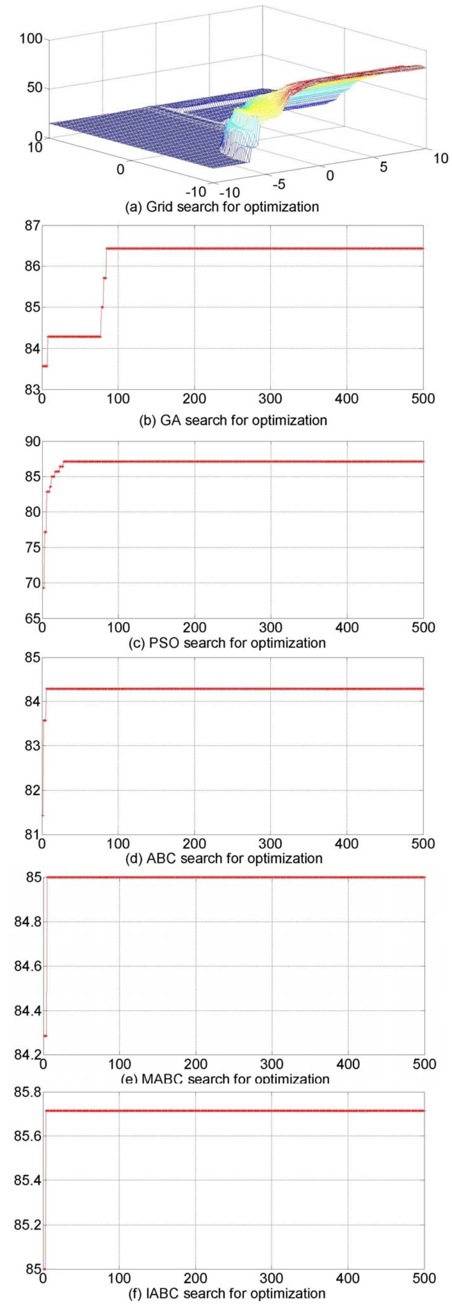

Fig. 1. The optimization curves of SVM parameters

VI. Conclusion

The model parameters of support vector machine (SVM) is one of the important factors affecting the classification accuracy. For most current problems, SVM training algorithm more likely have a larger computation complexity and a long running time, this paper proposes a new SVM parameters optimization method that bases on improved ABC. This method uses classification accuracy as the factor of fitness function, and uses the improved ABC to optimize the penalty factor C and function parameter g of RBF kernel function of support vector machine. Experiments on standard datasets show that the improved ABC has a better balance between the searching of width and mining of depth.Compared with other algorithms like basic artificial bee colony, genetic algorithm and particle swarm algorithm, IABC can achieve a higher classification accuracy. Compared with MABC algorithm, this method can effectively reduce searching time. This method is applied to the classification of the measured human activity data, and also obtains good classification accuracy, which proves that IABC algorithm has important practical application value.

REFERENCES

[1] Cristianini N,Taylor J.S. "*Support vector machine introduction*,"Microcomputer Development, Beijing: Electronic Industry Press,2004.

[2] S.f. Ding,B.j. Qi,H.y. Tan. "*Support vector machine theory and algorithm research review*,"Journal of electronic science and technology (natural science edition), Vol. 40,No. 1, pp. 2-10,2011.

[3] Vapnik V.N. "*The nature of statistical learning theory*,"New York:Springer-Verlag,1995.

[4] Vapnik V.N. "*Statistical learning theory*,"Beijing: Electronic Industry Press,2004.

[5] Pontil M,Verri A. "*Support vector machines for 3 D object recognition Pattern Analysis and Machine Intelligence*,"EEE Transactions, Vol. 20,No. 6, pp. 637-646,1998.

[6] S.q. Xie, F.m. Shen, X.n. Xiu. "*Face recognition method based on support vector machine*,"Computer Engineering, Vol. 35,No. 16, pp. 186-188,2009.

[7] Y.x. Li, X.g. Ruan. "*The genetic selection of tumor classification based on support vector machine*,"Computer research and development, Vol. 42,No. 10, pp. 1796-1801,2005.

[8] T. Meng. "*SVM parameter optimization based on adaptive genetic algorithm*,"Computer measurement and control, Vol. 24,No. 9, pp. 215-217,2016.

[9] N. Zhu,Z.g. Feng and Q. Wang. "*Parameter optimization of support vector machine classifier based on genetic algorithm of small habitat*,"Journal of nanjing university of technology (natural science edition), Vol. 33,No. 1, pp. 16-20,2009.

[10] Z.s. Zhang,L.j. Li and Z.j. He. "*The optimization research of fault classifier parameter based on support vector machine*,"Journal of xi 'an jiaotong university, Vol. 37,No. 11, pp. 1101-1104,2003.

[11] Y.x. Hu and H.t. Zhang. "*Based on simulated annealing algorithm - the classification of storage pests for support vector machines*,"Journal of agricultural machinery,Vol. 39,No. 9,pp. 108-111, 2008.

[12] F. Yan and S.y. Qin."*A kind of support vector machine overparameter optimization algorithm based on simulated annealing*,"Mission control,Vol. 26,No. 5,pp. 7-11,2008.

[13] Z.y. Lu,Y.y. Li and J. Lu. "*Particle swarm optimization algorithm optimizes RBF-SVM sandstorm prediction model parameters*,"Journal of tianjin university (natural science and engineering),Vol. 41,No. 4,pp. 413-418,2008.

[14] Z.y. Wu,H.q. Yuan. "*Application of ant colony support vector machine in internal combustion engine fault diagnosis*,"Vibration and shock, Vol. 28,No. 3,pp. 83-86,2009.

[15] S.q. Zhang,J.f. Teng and J.h. Gu. "*Artificial swarm algorithm based on multi-dimensional greedy search*,"Computer engineering, Vol. 40,No. 11,pp. 189-193,2014.

[16] L. Lu and T.y. Wang."*Optimization of support vector machine based on artificial swarm algorithm*",Journal of tianjin university (natural science and engineering),Vol. 44,No. 9,pp. 803-807,2011.

[17] Karaboga D."*An idea based on honey bee swarm for numerical optimization*",echnical Report-TR06 Engineering Faculty,Computer Engineering Department,Kayseri: Erciyes University,2005.