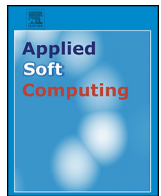




Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A feature selection method based on modified binary coded ant colony optimization algorithm

Youchuan Wan^a, Mingwei Wang^a, Zhiwei Ye^{b,*}, Xudong Lai^a^a School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, PR China^b School of Computer Science, Hubei University of Technology, Wuhan 430068, PR China

ARTICLE INFO

Article history:

Received 25 February 2016

Received in revised form 28 June 2016

Accepted 4 August 2016

Available online xxx

Keywords:

Feature selection

Genetic algorithm

Modified binary coded ant colony optimization

Visibility density model

Pheromone density model

ABSTRACT

Feature selection is a significant task for data mining and pattern recognition. It aims to select the optimal feature subset with the minimum redundancy and the maximum discriminating ability. In the paper, a feature selection approach based on a modified binary coded ant colony optimization algorithm (MBACO) combined with genetic algorithm (GA) is proposed. The method comprises two models, which are the visibility density model (VMBACO) and the pheromone density model (PMBACO). In VMBACO, the solution obtained by GA is used as visibility information; on the other hand, in PMBACO, the solution obtained by GA is used as initial pheromone information. In the method, each feature is treated as a binary bit and each bit has two orientations, one is for selecting the feature and another is for deselecting. The proposed method is also compared with that of GA, binary coded ant colony optimization (BACO), advanced BACO (ABACO), binary coded particle swarm optimization (BPSO), binary coded differential evolution (BDE) and a hybrid GA-ACO algorithm on some well-known UCI datasets; furthermore, it is also compared with some other existing techniques such as minimum Redundancy Maximum Relevance (mRMR), Relief algorithm for a comprehensive comparison. Experimental results display that the proposed method is robust, adaptive and exhibits the better performance than other methods involved in the paper.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Feature selection is a fundamental task for pattern recognition and data mining applications especially for high-dimensional datasets, which is the process of selecting a subset of relevant features for use in model construction [1]. Ideally, objects may be described more completely with more features and each feature should supplement an independent set of information. Unfortunately, features are correlated each other in common, which may lead to a large number of useless calculation with lower accuracy. As a result, many approaches have been proposed around the topic, most of those fall into two categories: unsupervised feature selection methods and supervised feature selection methods. For instance, Kim [2] proposed an unsupervised feature selection method with weighted principal components (PCs), and the weighted PC was obtained by the weighted sum of the important PCs of interest. Mitra [3] presented an unsupervised feature selection algorithm based on measuring similarities of each features.

Wang [4] came up with a new criterion called maximum projection and minimum redundancy, to address unsupervised learning scenarios, and experimental results demonstrated the superiority of the method. Gao [5] utilized the Relief algorithm to solve the problem of feature selection. Song [6] and Li [7] respectively proposed two novel unsupervised algorithms based on clustering analysis theory; experimental results showed that the proposed method not only produced the smaller subset of features, but also improved the performance of the classifier. In all, unsupervised methods do not apply any prior information; the performance might need to be further improved. As for supervised methods, Basu [8] proposed a new supervised feature selection approach by generating a score of each term based on their similarity with all the classes and then all the terms will be ranked accordingly. Frénay [9] presented a method for mutual information estimation, which was based on a probabilistic label noise model combined with a nearest neighbours-based entropy estimator. Liu [10] proposed a method of feature selection by using the AdaBoost algorithm and naive Bayes nearest-neighbor (NBNN) classifier for action recognition. Bermejo [11] presented a novel approach depending on the complexity of the wrapper search and filter ranking, whose main goal was to drastically reduce the number of wrapper evaluations while maintaining good classification accuracy. Park [12] presented an ensemble-based algorithm

* Corresponding author.

E-mail addresses: yichuan@whu.edu.cn (Y. Wan), wmwscola@sina.com (M. Wang), weizhiye121@126.com (Z. Ye), laixudong@whu.edu.cn (X. Lai).

with K-Nearest Neighbor classifier and random forests algorithm by using an iterative procedure. Verikas [13] and Li [14] utilized artificial neural network (ANN) respectively to select important features from the originally given plausible features; experimental results shown that the proposed method outperformed the other methods by achieving higher classification accuracy on all the problems tested. Liu [15] proposed a method to pick important features by using information criteria, which worked like the process of feature clustering in an agglomerative way. As the most widely used supervised learning method, support vector machine (SVM) is a powerful machine learning approach [16]. It has been widely used in many fields, such as peer-to-peer network traffic identification [17], real-time detection system [18], and breast cancer detection [19] and so on. Moreover, SVM has also been successfully utilized for feature selection. One way is to directly obtain the optimal feature subset by using SVM [20,21]; on the other hand, it is used as a classifier [22,23]. In brief, supervised methods have better performance than unsupervised methods, but the problem of feature selection has not been fully resolved.

In nature, feature selection is a typical combinational optimization problem. Its computing complexity is $O(2^N)$, where N is the number of features and it is a non-polynomial hard problem in essence. For non-polynomial hard problem, a lot of research has focused on how to obtain the satisfied solution by using some optimization algorithms in limited time. Some evolutionary algorithms and swarm intelligence algorithms (genetic algorithm (GA), ant colony optimization (ACO), particle swarm optimization (PSO), and differential evolution (DE) and so on) have been successfully applied continuous optimization and combinatorial optimization problems, such as function optimization [24], travelling salesman problem [25], job-shop scheduling [26], proportion integral derivative controller [27], UWB antenna [28]. Moreover, evolutionary algorithm and swarm intelligence have been employed to solve the feature selection problem. For example, Kannan [29] proposed a new intrusion detection model in which combining a GA based feature selection algorithm and an existing fuzzy SVM for effective classification as a solution. Chu [30] and Pedernana [31] proposed a hybrid feature selection strategy based on GA and SVM (GA-SVM), which formed to search for the best combination of features with higher classification accuracy. Abdi [32] developed a diagnosis model based on PSO and SVM to select the optimal feature subset from the original feature set and diagnose erythematous-squamous diseases. Unler [33] presented a hybrid filter wrapper feature subset selection algorithm based on maximum relevance minimum redundancy PSO (mr2PSO) for SVM classification. Bhattacharyya [34] proposed an efficient feature selection technique, realized by means of DE, which attempted to overcome the shortcomings of several state-of-the-art approaches in the topic.

Ant colony optimization (ACO) [35] is a population-based stochastic global search algorithm, which was initially proposed by Dorigo in 1991. Nowadays, ACO has been widely used in diverse fields, such as network routing [36], software project scheduling [37], fuzzy control [38], etc. As for feature selection, Chen [39] utilized ACO and rough set theory to find a minimal subset of the features. Chen [40] presented a feature selection algorithm based on ACO for image classification and recognition. Huang [41] presented a novel hybrid ACO-based classifier model that combines ACO and SVM to improve classification accuracy with a small and appropriate feature subset. However, the convergence speed of ACO is slow, and it is highly possible to plunge into the local optimum. Therefore, Lee [42] and Changdar [43] have proposed a hybrid ant colony optimization via genetic algorithm (GA-ACO), which simply utilized GA to generate feature subset, and adopted ACO to enhance the performance of GA by incorporating local search. On the other hand, in order to deal with the continuous optimization problem, a binary coded ant colony optimization algorithm (BACO)

[44] is proposed for the unconstrained function optimization problem. Jang [45] proposed a novel BACO algorithm based on ACO, and developed to solve the unit commitment problem of power systems. Ahmed [46] proposed a novel feature selection algorithm based on BACO to select the optimal feature subset. Kashef [47] presented a novel feature selection algorithm based on advanced BACO, which verified that the algorithm provided a suitable feature subset with good classification accuracy using a small feature set. Kadri [48] proposed a novel hybrid algorithm for fault diagnosis of rotary kiln based on BACO and SVM to find an optimization subset which was achieved through eliminating the noise feature. Although the simple BACO has superior optimization ability, it is not robust enough and the computing time complexity of ABACO is relatively high. Moreover, at the beginning of BACO, the heuristic information is very important, which influences its robustness and efficiency to a certain degree. GA could obtain a good solution without any heuristic information, which remedies the disadvantage of BACO. Thus, a modified binary coded ant colony optimization algorithm combined with genetic algorithm (MBACO) is considered here and employed for feature selection is investigated.

The rest of the paper is structured as below. Section 2 illustrates the modified binary coded ant colony optimization algorithm combined with genetic algorithm. The proposed feature selection method is detailed in Section 3. Section 4 displays the experimental results and discussion. Finally, the paper is concluded in Section 5.

2. The modified binary coded ant colony optimization algorithm

2.1. Overview of binary coded ant colony optimization algorithm

Binary coded ant colony optimization algorithm (BACO) [44] is a kind of swarm intelligence algorithm based on standard ACO, which has been used to solve continuous optimization problems. Supposed a continuous optimization problem is described as Eq. (1).

$$\begin{cases} \max f(x) \\ x = (x_1, \dots, x_k, \dots, x_n) \in R^n, x_k \in X_k \subset R \end{cases} \quad (1)$$

A solution $x^* \in R^n$ is called the global optimal solution in case that $f(x^*) = \max\{f(x) | x \in R^n\}$. The feasible regions of all variables in x are coded with a binary string to construct search space before BACO starts. Let a binary string $(b_{k,L-1}, b_{k,L-2}, \dots, b_{k,i}, \dots, b_{k,0})$, $b_{k,i} \in \{0, 1\}$ denotes a feasible region of x_k , L is the string length. The process of finding the optimal solution is to decide the best state for each bit, which could be viewed as a problem of searching the optimal path in a directed graph shown in Fig. 1. The vertices of the graph composed of “0” and “1” that are the state candidates of every bit. At a bit, each ant needs to choose an alternative state of two vertices. The space of incidence matrix traversed by every ant is $2 \times n$.

At each iteration, an ant travels all nodes to build a candidate solution. It departs from the lowest bit in each digraph and travels through the digraph along arcs. Thus, its trace will generate a binary string with L bits; by this means, ant colony will construct a group

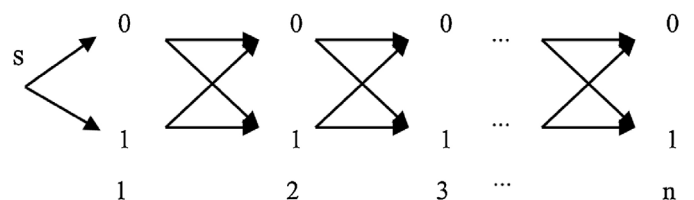


Fig. 1. Traverse road map in the binary coded ant colony algorithm.

of candidate solutions. The size of solutions is equal to ant colony population. The k th binary string can be mapped into the solution x by decoding it to a decimal number.

Initially, the concentration of pheromone is equal in both route of “0” and “1”. Let τ_{ij} be the pheromone on the bit “ i ” to state “ j ”, with $j \in \{0, 1\}$. As shown in Fig. 1, there are two arcs adjacent to next vertex for each bit. An ant decides its route on the basis of the pheromone distribution on both arcs. The ant “ k ” ($k = 1, 2, \dots, M$) decides the direction by concentration of pheromone in the different routes. The movement probability is defined as follows

$$p_{ij}^k(0) = \frac{[\tau_{ij}(0)]^\alpha \cdot [\eta_{ij}(0)]^\beta}{[\tau_{ij}(0)]^\alpha \cdot [\eta_{ij}(0)]^\beta + [\tau_{ij}(1)]^\alpha \cdot [\eta_{ij}(1)]^\beta} \quad (2)$$

$$p_{ij}^k(1) = 1 - p_{ij}^k(0) \quad (3)$$

where p_{ij}^k is the probability of movement from the bit “ i ” to the state “ j ” at time t . The parameter “ α ” ($\alpha > 0$) is the relative importance of the pheromone. The parameter “ β ” ($\beta > 0$) is the relative importance of the visibility. $\tau_{ij}(0)$ is the pheromone density of the side which the “ j ” is “0”, while $\tau_{ij}(1)$ is the pheromone density of the side which the “ j ” is “1”. $\eta_{ij}(0)$ is the visibility density of the side which the “ j ” is “0”, while $\eta_{ij}(1)$ is the visibility density of the side which the “ j ” is “1”. The ant decides the route by the concentration of pheromone on two sides. In addition, the pheromone density will evaporate as time goes. Let $\tau_{ij}(t+1)$ be the intensity at time $t+1$ (t is number of iterations), the trial information is updated as Eqs. (4) and (5).

$$\tau_{ij}(0)(t+1) = (1 - \rho)\tau_{ij}(0)(t) + \Delta\tau_{ij}^{best} \quad (4)$$

$$\tau_{ij}(1)(t+1) = (1 - \rho)\tau_{ij}(1)(t) + \Delta\tau_{ij}^{best} \quad (5)$$

where ρ is a coefficient representing evaporation rate and $\rho \in [0, 1]$, $\Delta\tau_{ij}^{best}$ is the incremental pheromone, which is defined as $\Delta\tau_{ij}^{best} = 1/f(s^{best})$, and the $f(s^{best})$ is equal to the best fitness value, if value “ j ” is selected for i -bit of the solution built by ant and zero otherwise. At the end of each iteration, the optimal solution x_t^* is acquired from the candidates of ant colony. BACO retains the record of optimal solution found so far, which is called the global best solution x^* . If x_t^* is superior to x^* , then x^* is replaced by x_t^* . As BACO seeks the optimum in parameter space and makes no use of heuristic information, the randomness of bit change may lead to the local optima.

2.2. Proposed binary coded ant colony optimization model

Genetic algorithm (GA) [49] is one of the most commonly used evolutionary computation methods, which follows the process of natural selection and evolution. It is widely applied to generate useful solutions to optimization and search problems especially when there is no heuristic information of the problem to be handled. It is noted that the initial heuristic information for BACO is randomly generated and BACO does not utilize information inspired by the handled problem itself, meanwhile; the heuristic information properly set could help BACO obtain the better result [44]. Fortunately, GA needs not any heuristic information in the process of evolution and could acquire satisfactory solution in most cases. However, it is vulnerable to premature convergence and has very low convergent speed at later evolution process. Hence, in order to take the advantages of the two algorithms, BACO may be combined with GA, here it is called MBACO. Further, BACO and GA utilize binary encoding, which make them easy to be cooperated. Two different models of MBACO are put forward in the paper called as the pheromone density model and the visibility density model respectively.

2.2.1. The pheromone density model

In order to quickly obtain the optimal solution, the influence of visibility density of BACO is ignored in this model; that is to say the parameters of α and β is set as $\alpha = 1$, $\beta = 0$. The solution of GA is used as initial pheromone density τ_{ij} for BACO, the model is relatively simple, that is, as an iteration of GA ended, a small proportion of the optimal individuals will be used in BACO, and the initial pheromone density is set according to the optimal binary strings generated by GA, details are as following. If certain bit of the strings is with value “0” in the binary strings of GA, then its pheromone density to “0” is set with a big value and its pheromone density to “1” is set with a small value, for example, $\tau_{ij}(0) = 0.8$, $\tau_{ij}(1) = 0.2$. Suppose that the binary string of GA is “01010001”, the pheromone density of the current individual can be respectively set from left to right as $\tau_i(0) = \{0.8, 0.2, 0.8, 0.2, 0.8, 0.8, 0.8, 0.2\}$, $\tau_i(1) = \{0.2, 0.8, 0.2, 0.8, 0.2, 0.2, 0.2, 0.8\}$; and then, generate the binary strings of BACO by the results of Eqs. (2) and (3). If the fitness value of current individual is better than that of GA, the individual will be used to replace the corresponding individual of GA and GA restarts to run, thus GA and BACO take turns to run until the termination condition is attained.

2.2.2. The visibility density model

The pheromone and visibility density are all utilized in the model. In order to emphasize the influence of visibility density, the parameters of α and β is set as $\alpha = 1$, $\beta = 3$. The initial pheromone density τ_{ij} is set as a small random number; the solution of GA is used as initial visibility density η_{ij} , and the pheromone and visibility density will simultaneously impact the result of BACO. As an iteration of GA ended, a small proportion of the optimal individuals will be used in BACO, and the initial visibility density is set by the optimal binary strings, details are as following. If certain bit of the strings is with value “0” in the binary strings of GA, then its visibility density to “0” is set with a big value and its visibility density to “1” is set with a small value, for example, $\eta_{ij}(0) = 0.8$, $\eta_{ij}(1) = 0.2$. Suppose that the binary string of GA is “01010001”, the visibility density of the current individual can be respectively set from left to right as $\eta_i(0) = \{0.8, 0.2, 0.8, 0.2, 0.8, 0.8, 0.8, 0.2\}$, $\eta_i(1) = \{0.2, 0.8, 0.2, 0.8, 0.2, 0.2, 0.2, 0.8\}$; and then, generate the binary strings of BACO by Eqs. (2) and (3). If the fitness value of current individual is better than that of GA, the individual will be used to substitute the corresponding individual of GA and GA rolls back, thus GA and BACO perform in turn until end condition is met. By the influence of pheromone and visibility density, each individual will have a multi-orientation search process comparing with pheromone density model.

3. The proposed method

In this section, the proposed feature selection method is developed to maximize the classification accuracy meanwhile minimize the number of redundant features by using MBACO. As one of the optimal classifiers, SVM is utilized to do classification here. Moreover, all features make up a fully connected graph where each bits represent features. As it is shown in Fig. 1, there are two orientations assigned to each feature, one for selecting the feature and the other for deselecting.

In the proposed method, the classification accuracy by using SVM is considered as the main factor of the fitness value.

$$Accuracy(i) = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

In Eq. (6), the meaning of parameters is as below: TP (True Positive): in case of test sample is positive and it is identified as positive, it is considered as a true positive; TN (True Negative): in case of test sample is negative and it is identified as negative, it is considered as a true negative. FP (False Positive), in case of test sample is negative

and it is identified as positive, it is considered as a false positive. F_N (False Negative) in case of test sample is positive and it is identified as negative, it is considered as a false negative.

3.1. Coding scheme of the proposed method

It is a key issue to make a suitable mapping between problem solution and MBACO. As for the problem of feature selection, each feature has two candidate states, it is selected or ignored, which is easy to be mapped with the binary code. Here, the coding length is equal to the number of the features. Each bit of MBACO is represented by “0” or “1”, “1” indicates this feature will be chosen for classification, and “0” indicates this feature will not be chosen. Supposed that the whole dataset has 8 features, the coding of MBACO is “01010001”. That is, the 2nd, 4th and 8th features will be chosen to make up a feature subset that is utilized to complete the task of classification by SVM, and other features will be deserted; the quality of the feature subset will be evaluated by the objective function. For example, {1,2,3,4,5,6,7,8} is a sample of the dataset, each row expresses a feature. According to the preceding discussion, the data {2,4,8} will be chosen to compose the feature subset, and import to the SVM classifier. The exchange of individual will generate different feature subsets. As the algorithm reach the termination condition, the optimal feature subset could be obtained by the coding of MBACO.

3.2. Objective function

For feature selection, the classification accuracy is the one significant goal. The ultimate goal of feature selection is to obtain the higher classification result by using the fewer features as possible. Thus, the objective function of feature selection is defined as below [50]:

$$F(i) = \frac{Accuracy(i)}{1 + \lambda \cdot n(i)} \quad (7)$$

where $F(i)$ is the fitness value of i th ant, $n(i)$ is the number of selected feature, $Accuracy(i)$ is the classification accuracy. λ is a weighting parameter, which is usually set as $\lambda = 0.01$.

3.3. Implementation of the proposed method

The proposed method is simple and easy to be implemented. The main process of the proposed MBACO for feature selection is as follows:

Algorithm 1. Feature selection optimized by MBACO

Input: The training dataset and testing dataset, and the iteration number of algorithm is $t=0$.

Output: The classification accuracy based on the optimal feature subsets.

```
1: Set GA parameters and generate initial populations;
2: while The algorithm does not reach the termination condition do
3:    $t = t + 1$ ;
4:   Calculate fitness value by using Eq. (7);
5:   Execute GA operation, such as selection, crossover, and mutation;
6:   Set pheromone (visibility) density according to GA and generate
   solution of BACO;
7:   if BACO does not reach 5 iterations then
8:     Ants search by Eqs. (2) and (3);
9:   end if
10:  Compute the fitness value of each solution generated by BACO and
   keeps the current optimal solution;
11:  if The solution of BACO is better than GA then
12:    Replace the current individual;
13:  else
14:    The individual of GA does not change;
15:  end if
```

```
16:   Update pheromone density by Eqs. (4 and 5);
17: end while
18: return The optimal solution, and the classification accuracy with the
   optimal feature subsets.
```

According to the operational process of evolutionary computation algorithm, the results of GA and BACO depend on parameters setting in some extent; fine tuning of the parameters could produce good results. Tables 1 and 2 show the parameters used in GA and BACO.

As well as some of the commonly used evolutionary algorithms or swarm intelligence based feature selection methods are also carried out in the paper. As it is illustrated in Section 2, MBACO is used in this paper. In order to make an intuitional comparison, all of the algorithms adopt the binary form. GA, BACO [44], binary coded PSO (BPSO) [51] and binary coded DE (BDE) [52] are all used with its standard mode. As simple GA-based approach may result in premature convergence, some strategies have been adopted to prevent premature convergence of GA. For instance, an efficient GA (EGA) which has a highly efficient uniform crossover [53] and an improved GA (IGA) by using adaptive crossover and mutation operators [54] have been put forward, which are also employed to handle with the problem of feature selection in the paper. In addition, because of its better optimization ability, the 4th heuristic information measurement of ABACO model which is proposed by Kashef [47] is also used to make a comparison. Finally, as to make a further comparison, a GA-ACO hybrid approach which is proposed by Changdar [43] is also used in this paper, and the parameters for all comparison algorithms are the same with simple GA and BACO. Tables 3 and 4 display the parameters of BPSO and BDE algorithm.

Table 1
Parameters used in GA.

Parameter	Explanation	Value
N	Number of genetic(s)	20
P_s	Selection ratio	0.9
P_c	Crossover ratio	0.8
P_m	Mutation ratio	0.1

Table 2
Parameters used in ACO.

Parameter	Explanation	Value
N	Number of ant(s)	20
α	Pheromone	1.0
β	Visibility	0.0,3.0
ρ	Evaporation of pheromone	[0,1]

Table 3
Parameters used in BPSO.

Parameter	Explanation	Value
N	Number of particle(s)	20
c_1, c_2	Acceleration constants	2.0
r_1, r_2	Random numbers	[0,1]

Table 4
Parameters used in BDE.

Parameter	Explanation	Value
N	Number of individual(s)	20
f_m	Mutation factor	0.6
C_R	Crossover rate	0.9

Table 5
General case of 10 public datasets.

Name	Sample size	Number of features	Number of classes
Wine	178	13	3
Credit	690	14	2
Forest	523	18	4
Image	2100	19	7
Cardiotocography	2126	21	3
Steel	1941	33	2
Dermatology	366	34	6
Biodegradation	1055	41	2
Spambase	2300	57	2
Handwritten	1797	64	9

4. Simulation results and discussion

The proposed algorithm is implemented by the language of Matlab 2014b on a personal computer with a 2.30 GHz CPU, 8.00G RAM under Windows 8 system.

In order to evaluate the performance of feature selection based on two models of MBACO (pheromone density model (PMBACO) and visibility density model (VMBACO) in Selection 2.2. 10 public datasets from UCI machine learning repository named “Wine”,

“Credit”, “Forest”, “Image”, “Cardiotocography”, “Steel”, “Dermatology”, “Biodegradation”, “Spambase” and “Handwritten” are used in this section. The general case of these 10 datasets has been shown in Table 5. For all datasets, we randomly choose half of data as training data, and last half are selected as testing data. The objective function is defined as Eq. (7). A higher fitness value of objective function indicates better optimization ability, and classification result can be calculated according to the fitness value.

To make an impartial comparison, all the algorithms will terminate after being executed 30 iterations; that is, all algorithms will stop when the iterations reach 30, and all the algorithms make 40 independent operations. For MBACO, BACO will run 5 iterations independently, as GA ends an iteration, and the computing complexity of the proposed MBACO is $O(N^2)$. On the other hand, some other feature selection techniques such as minimum Redundancy Maximum Relevance (mRMR) [4], Relief algorithm [5] and K-Nearest Neighbor algorithm (KNN) [12] are also used to make a comparison here. In order to make an overall comparison, the original SVM classification technique with all the features is also used in this section. Among them, the selected feature number of mRMR method is equal to the proposed method. In this section, we present some contrastive experimental results, including illustrative examples and performance evaluating tables, which clearly demonstrate

Table 6
Classification accuracy of different algorithms.

Dataset	Meas.	GA	EGA	IGA	BPSO	BDE	BACO	ABACO	GA-ACO	PMBACO	VMBACO
Wine	Acc (%)	96.2921	96.5169	96.6854	97.0292	97.9775	98.2270	98.4831	98.3472	98.7079	99.1011
	Std _a (%)	2.3342	2.0848	1.8382	1.5509	1.4853	1.2565	1.1685	1.2153	0.9883	1.0050
Credit	Acc (%)	83.5942	83.9087	84.1884	84.3768	84.5507	85.0928	85.2957	85.4101	85.5797	85.7826
	Std _a (%)	2.2216	1.9904	1.7659	1.5178	1.4638	1.3473	1.2467	1.3034	1.2094	1.1972
Forest	Acc (%)	67.2989	67.5414	68.1494	67.7866	68.6854	69.2286	69.9276	69.5870	70.8812	71.4176
	Std _a (%)	3.1282	2.8192	2.7279	2.7435	2.6372	2.5780	2.2371	2.4194	2.1832	2.0365
Image	Acc (%)	87.0952	87.3571	87.8095	87.5619	88.1190	88.8333	89.4619	89.0095	89.8238	90.2000
	Std _a (%)	2.6140	2.1186	1.7279	1.9439	1.5670	1.3382	1.1476	1.2580	1.0433	0.8712
Cardiotocography	Acc (%)	84.1364	84.6294	84.5484	84.3509	84.8119	85.2446	85.4184	85.5033	85.6419	86.2409
	Std _a (%)	2.3980	1.7898	1.8354	1.9125	1.7037	1.5284	1.4226	1.3950	1.3344	1.2220
Steel	Acc (%)	97.5660	97.8918	98.2784	98.1701	98.9350	99.6320	99.7732	99.6970	99.8454	100
	Std _a (%)	1.7711	1.6170	1.4171	1.5330	1.1189	0.8250	0.7053	0.7526	0.6916	0
Dermatology	Acc (%)	90.1366	91.4173	91.1804	90.7650	92.1967	93.0328	93.8361	93.8962	94.6448	95.1639
	Std _a (%)	6.0146	4.7573	4.8203	4.9504	4.6051	4.2759	3.5782	3.4313	3.1346	2.6551
Biodegradation	Acc (%)	72.6603	72.9208	73.5539	73.2486	74.3112	75.7211	76.0683	76.0000	76.8216	78.2258
	Std _a (%)	5.6850	5.4538	4.8013	4.9345	4.5573	4.0108	3.4301	3.4950	2.4972	1.8638
Spambase	Acc (%)	85.9000	86.2478	86.2739	85.0174	86.5391	87.3087	88.0652	87.7739	88.4783	89.4174
	Std _a (%)	2.7992	2.7077	2.6872	2.8693	2.5961	2.1028	1.8227	1.9756	1.6924	1.4428
Handwritten	Acc (%)	91.1704	91.7040	91.8408	91.4207	91.8563	92.1659	92.3719	92.4212	92.5891	92.9728
	Std _a (%)	2.0982	1.7295	1.6676	1.8121	1.6489	1.5462	1.4088	1.3830	1.3486	1.2515

Table 7
Fitness value of different algorithms.

Dataset	Meas.	GA	EGA	IGA	BPSO	BDE	BACO	ABACO	GA-ACO	PMBACO	VMBACO
Wine	Fiv	0.9141	0.9176	0.9191	0.9219	0.9303	0.9350	0.9361	0.9355	0.9375	0.9396
	Std _f	0.0207	0.0192	0.0177	0.0170	0.0141	0.0106	0.0079	0.0101	0.0098	0.0084
Credit	Fiv	0.8174	0.8205	0.8233	0.8267	0.8299	0.8367	0.8413	0.8430	0.8478	0.8525
	Std _f	0.0167	0.0178	0.0151	0.0195	0.0140	0.0128	0.0122	0.0114	0.0116	0.0119
Forest	Fiv	0.6242	0.6305	0.6449	0.6417	0.6522	0.6635	0.6688	0.6659	0.6704	0.6779
	Std _f	0.0325	0.0297	0.0271	0.0410	0.0282	0.0265	0.0248	0.0253	0.0229	0.0238
Image	Fiv	0.8027	0.8039	0.8109	0.8050	0.8215	0.8415	0.8457	0.8431	0.8480	0.8528
	Std _f	0.0226	0.0213	0.0201	0.0230	0.0197	0.0119	0.0112	0.0117	0.0095	0.0100
Cardiotocography	Fiv	0.7837	0.7913	0.7902	0.7893	0.7938	0.8126	0.8149	0.8158	0.8173	0.8213
	Std _f	0.0146	0.0132	0.0137	0.0134	0.0125	0.0104	0.0092	0.0103	0.0114	0.0076
Steel	Fiv	0.8278	0.8317	0.8413	0.8371	0.8495	0.8698	0.8889	0.8768	0.9037	0.9206
	Std _f	0.0223	0.0213	0.0186	0.0205	0.0140	0.0109	0.0094	0.0101	0.0104	0.0059
Dermatology	Fiv	0.7932	0.8006	0.7977	0.7945	0.8237	0.8391	0.8465	0.8494	0.8571	0.8610
	Std _f	0.0348	0.0291	0.0334	0.0379	0.0311	0.0256	0.0203	0.0198	0.0171	0.0136
Biodegradation	Fiv	0.6275	0.6315	0.6403	0.6367	0.6606	0.6785	0.6862	0.6854	0.7049	0.7147
	Std _f	0.0427	0.0360	0.0315	0.0326	0.0326	0.0305	0.0294	0.0279	0.0258	0.0247
Spambase	Fiv	0.6996	0.7087	0.7103	0.6929	0.7149	0.7223	0.7302	0.7268	0.7373	0.7539
	Std _f	0.0216	0.0171	0.0163	0.0178	0.0155	0.0112	0.0093	0.0104	0.0099	0.0082
Handwritten	Fiv	0.7131	0.7295	0.7317	0.7230	0.7397	0.7577	0.7646	0.7673	0.7745	0.7861
	Std _f	0.0175	0.0162	0.0157	0.0181	0.0150	0.0117	0.0102	0.0106	0.0125	0.0095

Table 8
Selected feature number of different algorithms.

Dataset	Meas.	GA	EGA	IGA	BPSO	BDE	BACO	ABACO	GA-ACO	PMBACO	VMBACO
Wine	Fn_a	6.35	6.25	6.10	6.00	5.80	5.55	5.25	5.40	5.30	5.15
	Fn_o	5	5	5	5	4	4	4	4	4	4
Credit	Fn_a	5.95	5.35	5.05	4.85	4.10	2.75	2.70	2.55	2.10	2.00
	Fn_o	5	4	4	3	3	2	2	2	2	2
Forest	Fn_a	5.70	5.60	5.35	5.45	5.25	4.80	4.60	4.70	4.40	4.25
	Fn_o	5	5	5	5	5	4	4	4	4	4
Image	Fn_a	8.20	7.95	7.40	7.60	7.15	6.60	5.95	6.35	5.35	5.00
	Fn_o	7	6	6	6	6	6	5	6	5	5
Cardiotocography	Fn_a	7.65	6.65	6.70	6.85	6.35	5.80	5.55	5.45	5.30	4.40
	Fn_o	6	6	6	6	6	5	5	5	5	4
Steel	Fn_a	17.30	16.80	15.70	16.05	14.30	13.10	12.25	12.80	10.50	9.40
	Fn_o	15	14	13	14	12	10	10	10	9	8
Dermatology	Fn_a	14.70	13.25	13.45	13.55	12.70	12.45	12.20	11.75	11.30	10.45
	Fn_o	14	13	13	13	12	12	12	11	11	10
Biodegradation	Fn_a	16.85	16.40	15.50	16.15	13.60	12.40	11.35	11.50	9.70	7.85
	Fn_o	16	16	15	16	13	11	10	10	8	7
Spambase	Fn_a	22.95	22.10	22.00	23.30	21.40	20.50	19.75	20.05	18.65	16.35
	Fn_o	20	20	20	21	19	18	18	18	18	16
Handwritten	Fn_a	27.55	25.05	24.95	25.35	24.85	21.85	21.05	20.85	19.80	17.85
	Fn_o	23	23	23	23	23	20	19	18	17	16

Table 9
CPU time of different algorithms.

Dataset	Meas.	GA	EGA	IGA	BPSO	BDE	BACO	ABACO	GA-ACO	PMBACO	VMBACO
Wine	Time	0.2918	0.3367	0.3103	0.4830	0.4496	0.4316	0.4989	0.3216	0.3705	0.3952
Credit	Time	2.0741	2.2474	2.1549	3.1990	2.7219	2.5766	3.2208	2.1954	2.2771	2.3830
Forest	Time	2.3638	2.9090	2.6342	3.8202	3.5134	3.4175	3.9719	2.7863	3.0922	3.2324
Image	Time	26.2871	30.9155	28.5272	37.2149	36.5631	35.1567	38.9459	29.6682	33.0334	34.5732
Cardiotocography	Time	15.5493	18.0098	16.6224	24.7104	22.9309	20.4749	25.2936	17.1499	18.1326	18.9623
Steel	Time	24.6279	28.5985	26.3358	34.4034	32.8970	31.7493	35.3008	27.5045	28.9580	29.5083
Dermatology	Time	1.5701	1.9823	1.6958	2.3887	2.2625	2.1646	2.5058	1.8121	1.8918	2.0874
Biodegradation	Time	9.2872	10.8854	9.7780	14.7017	13.7324	13.0465	15.3082	10.3761	11.1327	12.1003
Spambase	Time	78.6670	87.5357	81.3967	114.5644	106.8571	103.7968	124.7460	84.5962	89.8254	96.7572
Handwritten	Time	59.3283	62.8977	60.3020	81.2072	74.3889	69.2895	85.6119	61.5157	64.0222	67.5725

the merits of the proposed method. All the algorithms are evaluated using the same objective function. Our primary interest is the optimal feature subset, which is shown by the fitness value of objective function defined as Eq. (7), and the classification accuracy by using the selected features.

In experiments, classification is accomplished by using the selected features. Tables 6–8 show the classification accuracy, fitness value and selected feature number of feature selection handled by different algorithms, and Table 9 shows the

classification accuracy by some commonly used feature selection methods. The average fitness evolving curve of 6 algorithms is shown in Fig. 2, and the red line for MBACO is all for VMBACO model.

In Table 6, Acc and Std_a respectively indicate average and standard deviation of the classification accuracy by making 40 independent operations. In Table 7, Fiv and Std_f respectively denote average and standard deviation of the fitness value by making 40 independent operations. In Tables 8 and 9, Fn_o and Fn_a respectively

Table 10
Classification accuracy of commonly used methods.

Dataset	Meas.	SVM	MBACO	mRMR	Relief	KNN
Wine	Fn_o	–	4	4	4	4
	Acc_o (%)	97.7528	100	89.8876	84.2679	91.1002
Credit	Fn_o	–	2	2	5	5
	Acc_o (%)	87.2761	88.6957	72.1739	88.4058	88.6957
Forest	Fn_o	–	4	4	6	6
	Acc_o (%)	71.6475	76.6284	72.7969	67.0498	73.5632
Image	Fn_o	–	5	5	6	6
	Acc_o (%)	86.9524	90.7619	87.6190	85.2381	87.1429
Cardiotocography	Fn_o	–	5	5	5	9
	Acc_o (%)	86.0771	88.2408	87.1119	81.8438	87.8645
Steel	Fn_o	–	8	8	10	9
	Acc_o (%)	99.4845	100	98.9691	89.6907	97.9381
Dermatology	Fn_o	–	10	10	13	9
	Acc_o (%)	96.7213	97.2678	85.2459	85.2459	89.0710
Biodegradation	Fn_o	–	7	7	11	9
	Acc_o (%)	79.3169	81.4042	73.4345	70.5882	70.5882
Spambase	Fn_o	–	16	16	32	25
	Acc_o (%)	88.6957	90.9565	84.8696	87.6522	84.3478
Handwritten	Fn_o	–	16	16	22	19
	Acc_o (%)	94.6584	95.1002	91.0913	81.2918	85.1893

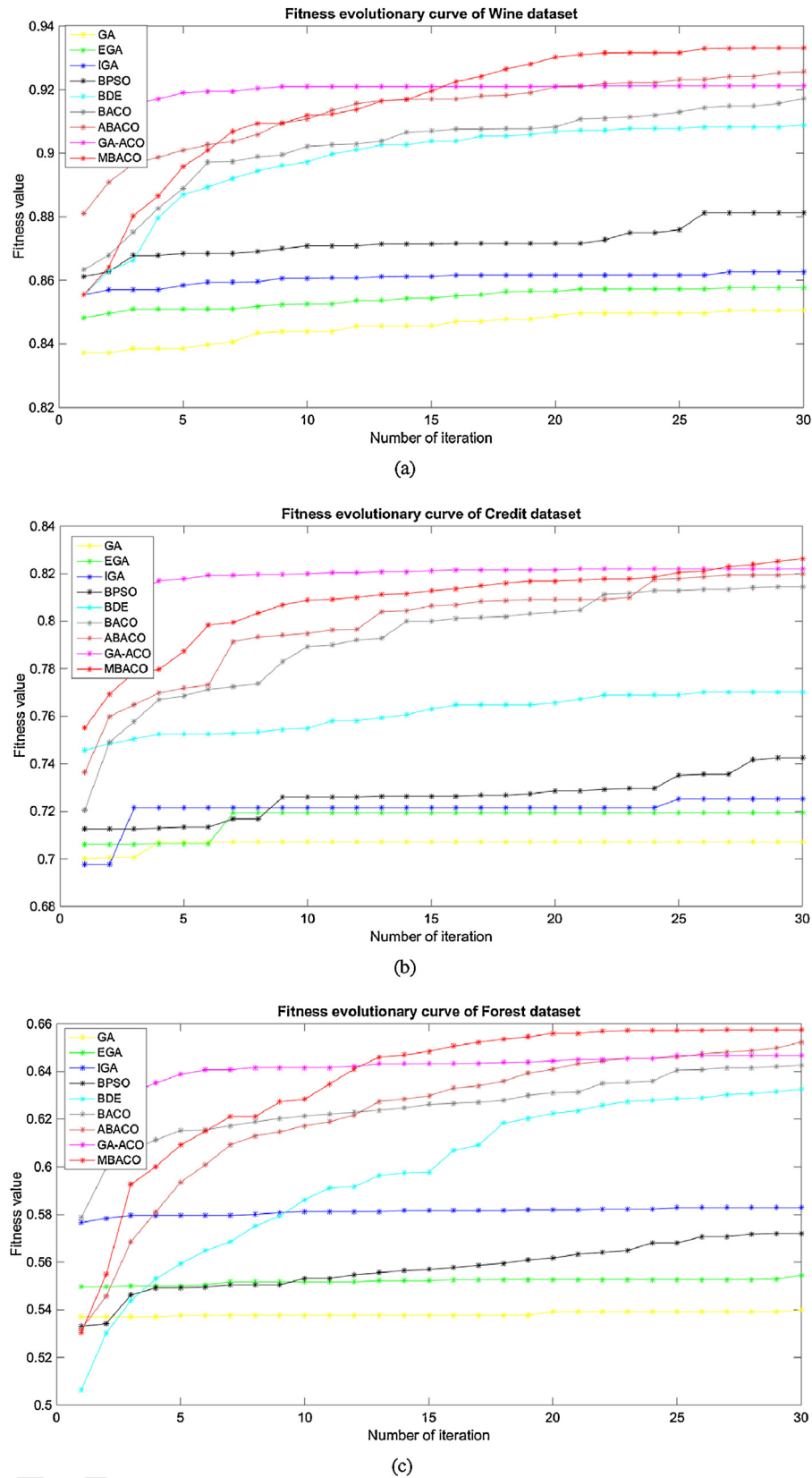


Fig. 2. Average fitness evolutionary curve: (a) "Wine" dataset, (b) "Credit" dataset, (c) "Forest" dataset, (d) "Image" dataset, (e) "Cardiotocography" dataset, (f) "Steel" dataset, (g) "Dermatology" dataset, (h) "Biodegradation" dataset, (i) "Spambase" dataset, (j) "Handwritten" dataset.

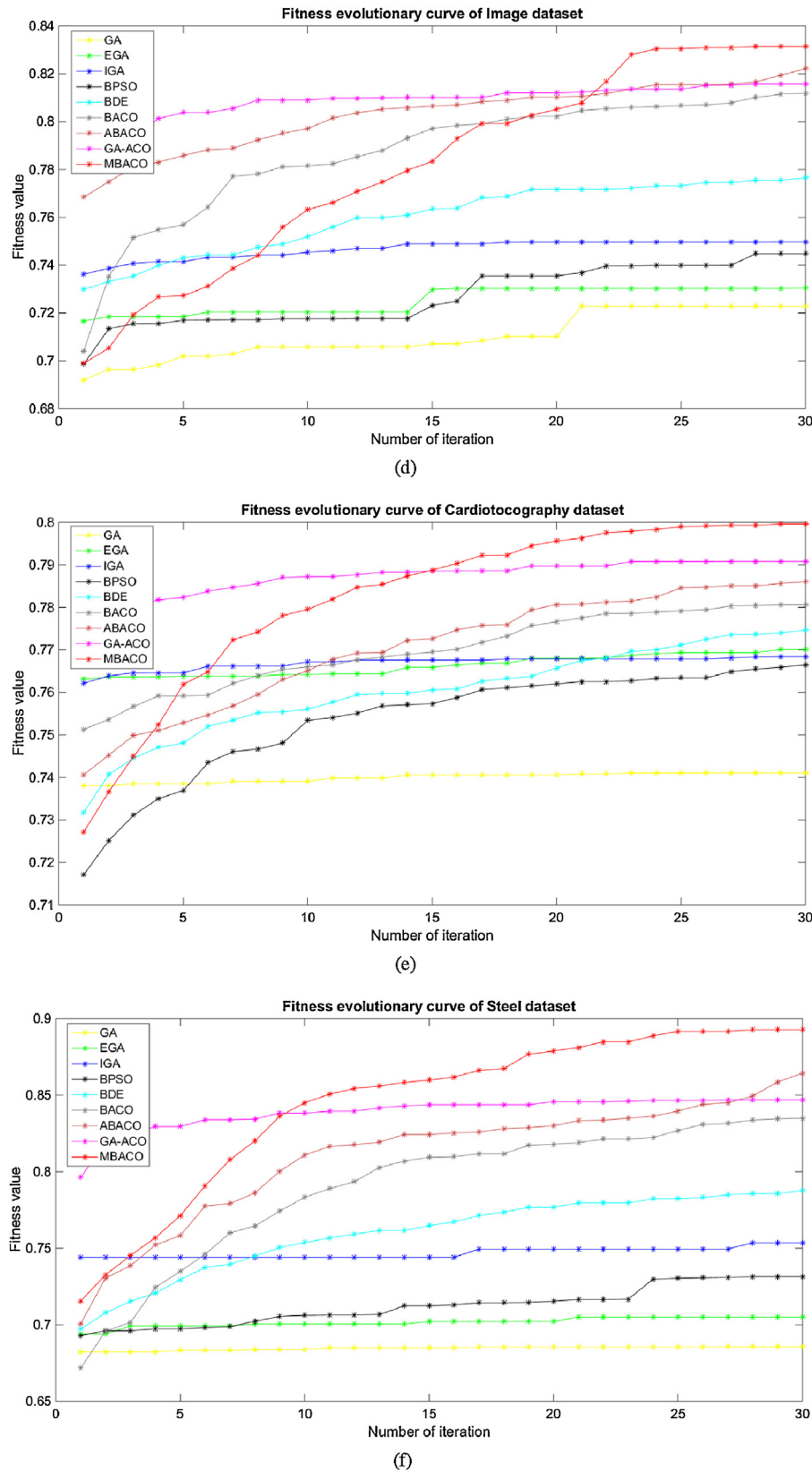


Fig. 2. (Continued)

indicate the optimal and average selected feature number in 40 independent operations. Time is the CPU time of each computing, its unit is second, Acc_o is the optimal classification accuracy in 40 independent operations.

According to the data in Table 6, it is observed that the average classification accuracy is close for 9 algorithms, the maximum difference value is only 5.5%; and for “Credit” and “Cardiotocography” datasets, the difference value is only 2.2%. Among these

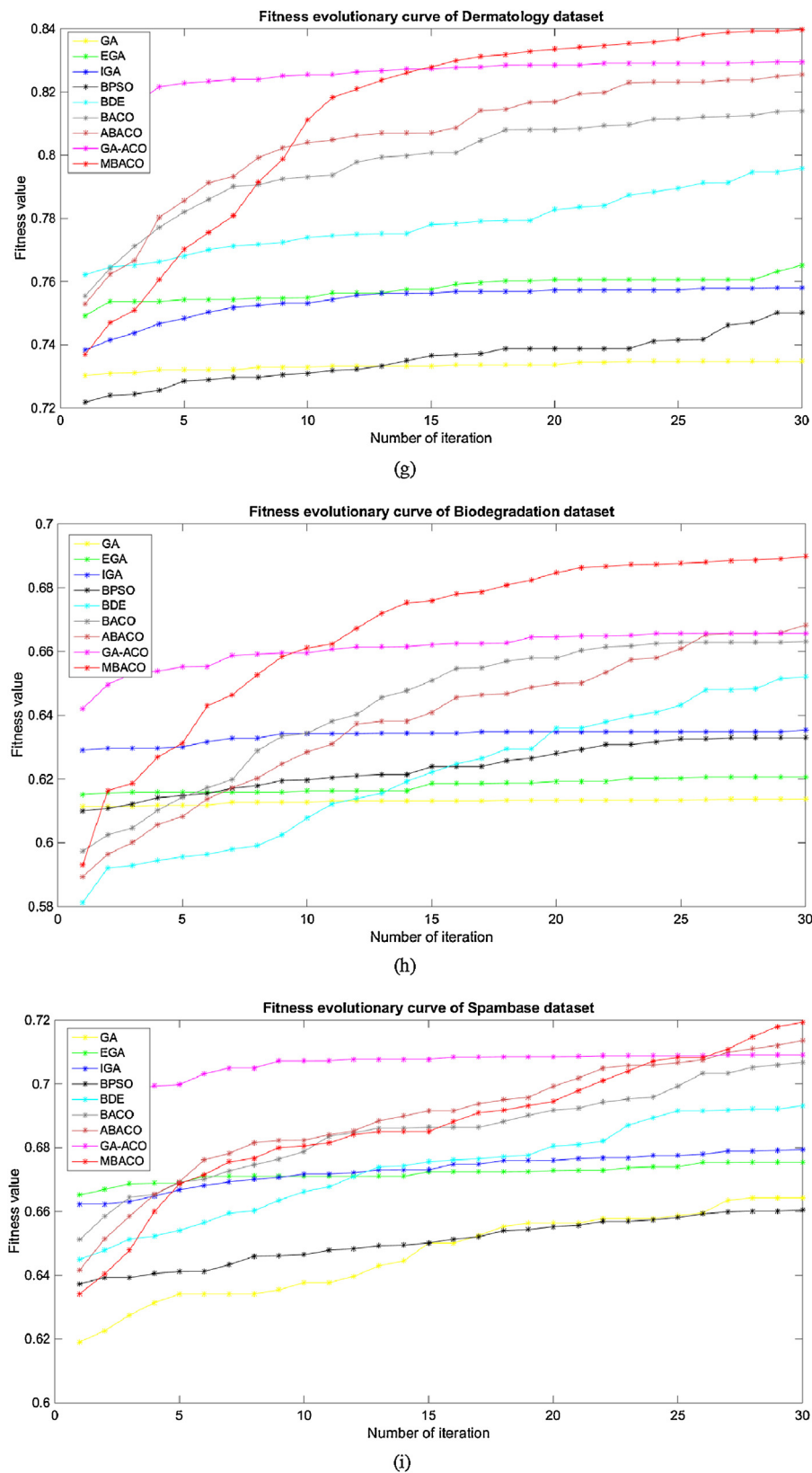


Fig. 2. (Continued)

methods, MBACO has the optimal classification accuracy for all algorithms; the average classification accuracy has reached 70% for all datasets, and the average classification accuracy has exceeded 90% for 5 datasets, especially for “Steel” dataset, it has reached

100%. In Table 7, the average fitness value of MBACO has reached 0.7 for 9 datasets, which illustrate that the optimization ability of MBACO has a distinct advantage comparing with the other 5 algorithms. Although for “Cardiotocography”, “Steel”, “Spambase” and

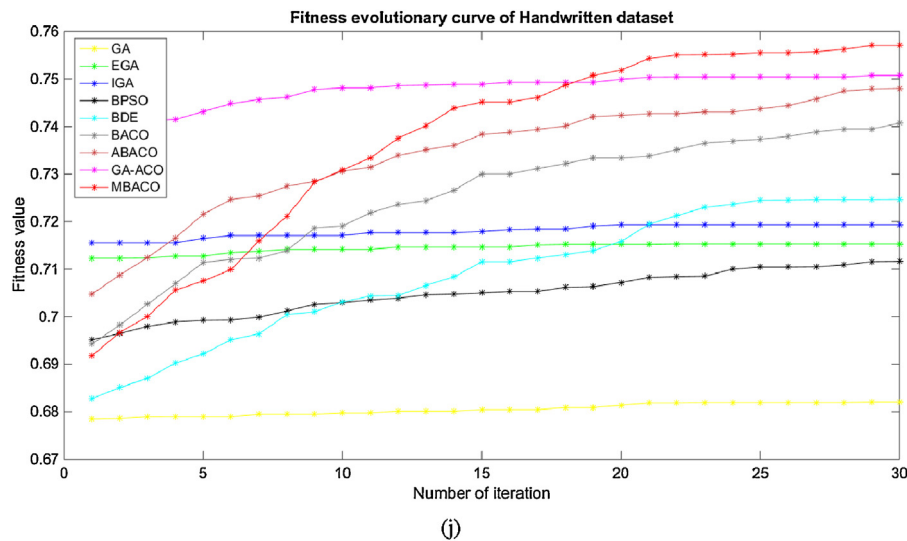


Fig. 2. (Continued).

“Handwritten” datasets, the standard deviation of fitness value by using PMBACO model is a few higher than that by using ABACO, the maximum difference is only 0.0023; and for “Wine” and “Credit” dataset, the standard deviation of fitness value by using PMBACO model and VMBACO model are all higher than that by using ABACO or GA-ACO algorithm, the difference is very small, which is less than 0.002. In Table 8, the selected feature number by using MBACO is all less than that by using other algorithms; especially for VMBACO model, the selected feature number is the optimal, which are all lower than 1/3 of the original feature number; especially for “Biodegradation” dataset, it only selects 7 features from the whole 41 features with rather good classification accuracy.

Furthermore, for computing efficiency in Table 9, BPSO, BDE, BACO and ABACO algorithm have a slow convergence speed, and the maximum difference of CPU time between them is more than 40s for “Spambase” dataset. It is worth noting that GA has the fastest convergence speed. However, the average fitness value of GA is obviously worse than MBACO; the maximum difference is over 0.08 for “Biodegradation” dataset. In Table 10, the selected feature number of the proposed method is superior for 9 datasets; although for “Dermatology” dataset, the selected feature number by using KNN algorithm is lower than that by using the proposed method, the classification accuracy has a distinctly gap, which has exceeded 8%. On the other hand, the process of feature selection by Relief algorithm could obtain higher classification accuracy comparing with the proposed method for “Spambase” dataset, its selected feature number is nearly double of the proposed method, and the computational efficiency is very low. On the whole, it may deduce that the proposed method has the better performance on the consideration of classification accuracy, computing efficiency and the number of selected features, which is a robust, reliable and efficient method.

5. Conclusion

In sum, a feature selection method optimized with MBACO algorithm is proposed. 10 datasets with different data dimensions from UCI machine learning repository are used to make an evaluation for the proposed method. Results are compared with some other classification techniques optimized by GA, EGA, IGA, BPSO, BACO, ABACO, BDE and GA-ACO. In general, it is observed that evolutionary algorithm and swarm intelligence algorithm can be well used to solve the problem of feature selection based on SVM. Among these algorithms, MBACO has a better performance, the average

fitness value is higher than GA, EGA, IGA, BPSO, BACO, ABACO, BDE and GA-ACO; that is MBACO is more appropriate to be employed to reduce the data dimension based on SVM than GA, EGA, IGA, BPSO, BACO, ABACO, BDE and GA-ACO. Moreover, in terms of CPU time, MBACO can quickly converge to the optimal solution; the number of selected features is lower than half of the total feature number, which is quite fast enough to meet real-time applications. On the other hand, in order to make a more comprehensive comparison, some commonly used feature selection techniques like minimum Redundancy Maximum Relevance (mRMR), Relief algorithm and K-Nearest Neighbor algorithm (KNN) are also used in this paper. It is revealed that the proposed method has a better performance; the classification accuracy is obviously higher than traditional feature selection approaches. As a whole, SVM has a stable performance for classification in most cases, and feature selection has a very significant effect for computing efficiency; especially for high dimension datasets, the efficiency is more significant. Further, the disadvantage of heavy computation efficiency could be conquered at the maximum degree when it is combined with MBACO. The proposed method is able to keep a good balance on the efficiency and classification accuracy, which makes it more suitable for some feature selection applications.

Acknowledgements

This work is funded by the 863 High Technology Program of China under Grant No. 2013AA122104, the National Science & Technology Pillar Program under Grant No. 2014BAL05B07, and the National Natural Science Foundation of China under Grant No. 41301371.

References

- [1] M. Dash, H. Liu, Feature selection for classification, *Intell. Data Anal.* 1 (1997) 131–156.
- [2] S.B. Kim, P. Rattakorn, Unsupervised feature selection using weighted principal components, *Expert Syst. Appl.* 38 (2011) 5704–5710.
- [3] P. Mitra, C. Murthy, S.K. Pal, Unsupervised feature selection using feature similarity, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (2002) 301–312.
- [4] S. Wang, W. Pedrycz, Q. Zhu, W. Zhu, Unsupervised feature selection via maximum projection and minimum redundancy, *Knowl. Based Syst.* 75 (2015) 19–29.
- [5] L. Gao, T. Li, L. Yao, F. Wen, Research and application of data mining feature selection based on relief algorithm, *J. Softw.* 9 (2014) 515–522.
- [6] Q. Song, J. Ni, G. Wang, A fast clustering-based feature subset selection algorithm for high-dimensional data, *IEEE Trans. Knowl. Data Eng.* 25 (2013) 1–14.

- [7] Z. Li, J. Liu, Y. Yang, X. Zhou, H. Lu, Clustering-guided sparse structural learning for unsupervised feature selection, *IEEE Trans. Knowl. Data Eng.* 26 (2014) 2138–2150.
- [8] T. Basu, C. Murthy, Effective text classification by a supervised feature selection approach, in: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW), 2012, pp. 918–925.
- [9] B. Frénay, G. Doquire, M. Verleysen, Estimating mutual information for feature selection in the presence of label noise, *Comput. Stat. Data Anal.* 71 (2014) 832–848.
- [10] L. Liu, L. Shao, P. Rockett, Human action recognition based on boosted feature selection and naive Bayes nearest-neighbor classification, *Signal Process.* 93 (2013) 1521–1530.
- [11] P. Bermejo, L. de la Ossa, J.A. Gámez, J.M. Puerta, Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking, *Knowl. Based Syst.* 25 (2012) 35–44.
- [12] C.H. Park, S.B. Kim, Sequential random k-nearest neighbor feature selection for high-dimensional data, *Expert Syst. Appl.* 42 (2015) 2336–2342.
- [13] A. Verikas, M. Bacauskiene, Feature selection with neural networks, *Pattern Recognit. Lett.* 23 (2002) 1323–1335.
- [14] R.-P. Li, M. Mukaidono, I.B. Turksen, A fuzzy neural network for pattern classification and feature selection, *Fuzzy Sets Syst.* 130 (2002) 101–108.
- [15] H. Liu, X. Wu, S. Zhang, A new supervised feature selection method for pattern classification, *Comput. Intell.* 30 (2014) 342–361.
- [16] L. Wang, *Support Vector Machines: Theory and Applications*, vol. 177, Springer Science & Business Media, 2005.
- [17] Z. Ye, M. Wang, C. Wang, H. Xu, P2P traffic identification using support vector machine and cuckoo search algorithm combined with particle swarm optimization algorithm, in: *Frontiers in Internet Technologies*, Springer, 2014, pp. 118–132.
- [18] N. Shibuya, B. Nukala, A. Rodriguez, J. Tsay, T. Nguyen, S. Zupancic, D. Lie, A real-time fall detection system using a wearable gait analysis sensor and a support vector machine (SVM) classifier, in: 2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU), 2015, pp. 66–67.
- [19] U.R. Acharya, E. Ng, J.-H. Tan, S.V. Sree, Thermography based breast cancer detection using texture features and support vector machine, *J. Med. Syst.* 36 (2012) 1503–1510.
- [20] M. Sheikhan, M. Bejani, D. Gharavian, Modular neural-SVM scheme for speech emotion recognition using ANOVA feature selection method, *Neural Comput. Appl.* 23 (2013) 215–227.
- [21] C. Shang, D. Barnes, Fuzzy-rough feature selection aided support vector machines for mars image classification, *Comput. Vis. Image Underst.* 117 (2013) 202–213.
- [22] J. Neumann, C. Schnörr, G. Steidl, Combined SVM-based feature selection and classification, *Mach. Learn.* 61 (2005) 129–150.
- [23] D. Liu, H. Qian, G. Dai, Z. Zhang, An iterative SVM approach to feature selection and classification in high-dimensional datasets, *Pattern Recogn.* 46 (2013) 2531–2537.
- [24] B.-Y. Qu, P.N. Suganthan, J.-J. Liang, Differential evolution with neighborhood mutation for multimodal optimization, *IEEE Trans. Evolut. Comput.* 16 (2012) 601–614.
- [25] M. Mahi, Ö.K. Baykan, H. Kodaz, A new hybrid method based on particle swarm optimization, ant colony optimization and 3-opt algorithms for traveling salesman problem, *Appl. Soft Comput.* 30 (2015) 484–490.
- [26] M.R. Singh, S. Mahapatra, A quantum behaved particle swarm optimization for flexible job shop scheduling, *Comput. Ind. Eng.* 93 (2016) 36–44.
- [27] A. Moharam, M.A. El-Hosseini, H.A. Ali, Design of optimal PID controller using hybrid differential evolution and particle swarm optimization with an aging leader and challengers, *Appl. Soft Comput.* 38 (2016) 727–737.
- [28] A.E. Teirab, J.A. Jervase, S.S. Mneina, Design of UWB monopole antenna using genetic algorithms, in: 2013 7th IEEE GCC Conference and Exhibition (GCC), 2013, pp. 89–92.
- [29] A. Kannan, G.Q. Maguire, A. Sharma, P. Schoo, Genetic algorithm based feature selection algorithm for effective intrusion detection in cloud networks, in: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW), 2012, pp. 416–423.
- [30] H. Chu, L. Ge, A. Ng, C. Rizos, Application of genetic algorithm and support vector machine in classification of multisource remote sensing data, *Int. J. Remote Sens. Appl.* 2 (2012) 1–11.
- [31] M. Pedergrana, P.R. Marpu, M.D. Mura, J.A. Benediktsson, L. Bruzzone, A novel technique for optimal feature selection in attribute profiles based on genetic algorithms, *IEEE Trans. Geosci. Remote Sens.* 51 (2013) 3514–3528.
- [32] M.J. Abdi, D. Giveki, Automatic detection of erythematous-squamous diseases using PSO-SVM based on association rules, *Eng. Appl. Artif. Intell.* 26 (2013) 603–608.
- [33] A. Unler, A. Murat, R.B. Chinnam, mr2pso: a maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification, *Inf. Sci.* 181 (2011) 4625–4641.
- [34] S. Bhattacharyya, A. Sengupta, T. Chakraborti, A. Konar, D. Tibarewala, Automatic feature selection of motor imagery EEG signals using differential evolution and learning automata, *Med. Biol. Eng. Comput.* 52 (2014) 131–139.
- [35] M. Dorigo, *Ant Colony Optimization: New Optimization Techniques in Engineering*, Springer-Verlag, Berlin Heidelberg, 1991, pp. 101–117.
- [36] B.C. Mohan, R. Baskaran, A survey: ant colony optimization based recent research and implementation on several engineering domain, *Expert Syst. Appl.* 39 (2012) 4618–4627.
- [37] W.-N. Chen, J. Zhang, Ant colony optimization for software project scheduling and staffing with an event-based scheduler, *IEEE Trans. Softw. Eng.* 39 (2013) 1–17.
- [38] O. Castillo, H. Neyoy, J. Soria, P. Melin, F. Valdez, A new approach for dynamic fuzzy logic parameter tuning in ant colony optimization and its application in fuzzy control of a mobile robot, *Appl. Soft Comput.* 28 (2015) 150–159.
- [39] Y. Chen, D. Miao, R. Wang, A rough set approach to feature selection based on ant colony optimization, *Pattern Recognit. Lett.* 31 (2010) 226–233.
- [40] B. Chen, L. Chen, Y. Chen, Efficient ant colony optimization for image feature selection, *Signal Process.* 93 (2013) 1566–1576.
- [41] C.-L. Huang, ACO-based hybrid classification system with feature subset selection and model parameters optimization, *Neurocomputing* 73 (2009) 438–448.
- [42] Z.-J. Lee, S.-F. Su, C.-C. Chuang, K.-H. Liu, Genetic algorithm with ant colony optimization (GA-ACO) for multiple sequence alignment, *Appl. Soft Comput.* 8 (2008) 55–78.
- [43] C. Changdar, R.K. Pal, G. Mahapatra, A genetic ant colony optimization based algorithm for solid multiple travelling salesmen problem in fuzzy rough environment, *Soft Comput.* (2016) 1–15.
- [44] M. Kong, P. Tian, A binary ant colony optimization for the unconstrained function optimization problem, in: *Computational Intelligence and Security*, Springer, 2005, pp. 682–687.
- [45] S.-H. Jang, J.-H. Roh, W. Kim, T. Sherpa, J.-H. Kim, J.-B. Park, A novel binary ant colony optimization: application to the unit commitment problem of power systems, *J. Electr. Eng. Technol.* 6 (2011) 174–181.
- [46] A.-A. Ahmed, Feature subset selection using ant colony optimization, *Int. J. Comput. Intell.* 2 (2005) 53–58.
- [47] S. Kashef, H. Nezamabadi-pour, An advanced ACO algorithm for feature subset selection, *Neurocomputing* 147 (2015) 271–279.
- [48] O. Kadri, L.H. Mouss, M.D. Mouss, Fault diagnosis of rotary kiln using SVM and binary ACO, *J. Mech. Sci. Technol.* 26 (2012) 601–608.
- [49] M. Srinivas, L.M. Patnaik, Genetic algorithms: a survey, *Computer* 37 (1994) 17–26.
- [50] Z. Ye, Z. Zheng, Y. Wan, A novel approach for feature selection based on ant colony optimization algorithm, *Geomat. Inf. Sci. Wuhan Univ.* 32 (2007) 1127–1130.
- [51] G.-C. Luh, C.-Y. Lin, Y.-S. Lin, A binary particle swarm optimization for continuum structural topology optimization, *Appl. Soft Comput.* 11 (2011) 2833–2844.
- [52] S. Das, P.N. Suganthan, Differential evolution: a survey of the state-of-the-art, *IEEE Trans. Evolut. Comput.* 15 (2011) 4–31.
- [53] X.-B. Hu, E. Di Paolo, An efficient genetic algorithm with uniform crossover for air traffic control, *Comput. Oper. Res.* 36 (2009) 245–259.
- [54] Z. Liao, X. Mao, P.M. Hannam, T. Zhao, Adaptation methodology of CBR for environmental emergency preparedness system based on an improved genetic algorithm, *Expert Syst. Appl.* 39 (2012) 7029–7040.