IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 7, 2019, accepted May 6, 2019, date of publication May 16, 2019, date of current version October 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917310

A Feature Selection Based Serial SVM Ensemble Classifier

JIANJUN CAO[1], (Member, IEEE), GUOJUN LV[2], (Member, IEEE), CHEN CHANG[2], (Member, IEEE), AND HONGMEI LI[2], (Fellow, IEEE)

[1]The 63rd Research Institute, National University of Defense Technology, Nanjing 210007, China
[2]PLA Army Engineering University, Nanjing 210007, China

Corresponding author: Guojun Lv (11983286838@outlook.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61371196, and in part by the China Postdoctoral Science Foundation under Grant 20090461425 and Grant 201003797.

ABSTRACT A serial support vector machine (SVM) ensemble classifier based on feature selection is designed to improve the efficiency and accuracy of unbalanced data classification so as to fully excavate the potential of features. The performance measurement indicators of the traditional SVM model "one-versus-others" are defined. With the goal of reaching higher classification precision and smaller feature set scale, the serial ensemble classifier is built using "one-versus-others" binary SVM classifier as a base classifier, and ant colony optimization (ACO) is used to determine the recognizable category and input feature subset of each level. Several experiments on vibration signals of real states of the engine are performed to contrast the classifying effects of the proposed approach, the traditional "one-versus-one" and "one-versus-others" parallel SVM classifiers, and the results demonstrate the superiority of the proposed approach.

INDEX TERMS Support vector machine, one-versus-others, unbalanced data classification, feature selection, serial SVM ensemble classifier.

I. INTRODUCTION

Multi-class imbalanced data classification is one of the crucial issues in the field of data classification. Many mature classification methods come up to a disappointed performance for imbalanced data which one or some classes have more samples in comparison to others. Generally, people tend to care more about minority class such as cancer diagnosis, in which we will pay a heavier price if we misjudge the minority class.

Approaches to deal with imbalanced data recognition can be summarized to data level approach, algorithm level approach and evaluation metrics.

Data level approach focuses on resizing the training datasets to balance all kinds of classes, the main techniques of resizing are over-sampling and under-sampling techniques. Chawla et al. proposed a method of over-sampling the minority class named SMOTE. Compared with the traditional method of random sampling, SMOTE creates synthetic minority class instances between the minority class instances and the surrounding instances, which solves the problem of overfitting effectively [1]. Later, Vorraboot proposed the data filtering technology to under-sample the samples of majority class in which hybrid algorithm is adapted [2].

Algorithm level approach focuses on carrying out modification on existing algorithm to strengthen the ability of learning from minority class, such as cost sensitive learning, one class learning and active learning. Cost sensitive learning attempts to increase the learning ability of classifiers by assigning larger misclassified cost for minority class samples. López et al. proposed Chi-FRBCS-BigDataCS algorithm to deal with large-scale imbalanced data using a fuzzy rule and cost sensitive learning techniques. This method uses the MapReduce framework to distribute the computational operations of the fuzzy model while using cost sensitive learning technique to deal with the imbalanced datasets [3]. Cao et al. proposed an optimized cost sensitive SVM for imbalanced data learning which optimizes the penalty term and the training parameters to improve the classification effect of SVM algorithm. This method optimizes the factors (ratio misclassification cost, feature set and intrinsic parameters of classifier) simultaneously for improving the performance of cost-sensitive SVM, while it uses the measure directly to train the classifier and discover the optimal parameter, ratio cost

The associate editor coordinating the review of this manuscript and approving it for publication was Zhao Zhang.

and feature subset based on different evaluation functions like G-mean or AUC to solve the problem [4]. However, the most popular one-class classifier used for imbalanced data classification is One-Class SVM (OCSVM). Tian et al. employed an algorithm to detect noises of minority samples and utilized the oversampled minority samples and a partial of majority samples to train SVM classifier [5]. What's more, Krawczyk et al. proposed a weighted OCSVM to model the minority class [6]. Later, Krawczyk and Woźniak introduced an ensemble classifier, in which OCSVM is used [7]. Besides of some methods based on SVM mentioned above, some other SVM models have been proposed to solve the unbalanced data classification due to its structural risk minimization principle [8]–[13]. Kim proposed a weighted k-means SVM for cancer prediction to circumvent the problem of imbalance in the data [8]. Richhariya et al. proposed a robust fuzzy least squares Twin Support Vector Machine(TWSVM) for class imbalance learning using 2-norm of the slack variables which makes the optimization problem strongly convex [9]. Cao et al. combined a re-sampling technique, which utilized over-sampling and under-sampling to balance the training data, with TWSVM to deal with imbalanced data classification [10]. Wang et al. modified SVM by presenting a new kernel function, which was extended and transformed by using the Chi square test and the weighting coefficients, in order to play a better classification performance for the minority class [11].

Since the most of traditional metrics are designed for addressing binary-class problems and they may ignore the minority class or treat them as noise, new metrics need to be proposed for multi-class imbalanced data classification. Gao et al. proposed a new class of neurofuzzy construction algorithms aiming at minimizing the leave-one-out mean square error which uses the modified AUC and F-value as performance metrics [14]. Krawczyk and Woźniak proposed a multiple classifier system, in which double-fault diversity measure is considered to prune those base classifiers that output similar hypotheses [7]. In fact, most of algorithms have abandoned using accuracy as performance metric, while ROC, G-mean and F-value are most popular used to cope with the class imbalanced problems [15].

In recent years, several ensemble algorithms combined with different approaches have been proposed to solve imbalanced data classification. Min proposed a method to integrate instance selection and bagging ensemble to improve the performance of the model, firstly using genetic algorithms to select the optimal instance subsets for base classifier learning which diversify the base classifiers in the ensemble model; then the majority voting scheme was used as a combining method; the results show a great improvement than some traditional bagging methods [16]. Last et al. proposed a feature-based serial combined classifier [17]. This combined classifier is a sequential chain of classifiers $C_1, C_2, \ldots, C_n$, which have an increasing number of features, whenever classifier C_{i-1} ($2 \leq i \leq n$) rejects the record, the next classifier in the chain, C_i, is called. Differently, we propose a serial SVM multi-class classifier for imbalanced data to improve the classification accuracy. On the basis of feature selection, "one-versus-others" SVM classifier is used as base classifier, with the goal of having higher classification precision and smaller feature set, the serial SVM ensemble classifier is built according to the cascade referring to the priorities of the assigned category. In this method, the number of classes is reduced continuously, at the same time, the number of classifiers among ensemble methods is controlled effectively, avoiding the situation of classifier redundancy [18], [19].

In this paper, section II defines several classification performance measurement indicators, section III builds the model of serial SVM ensemble classifier, section IV uses ACO to solve this model, section V verifies the performance of this model, and finally conclusions are given in section VI.

II. CLASSIFICATION PERFORMANCE MEASUREMENT

A. CLASSIFICATION PERFORMANCE MEASUREMENT INDICATORS

Equipment states are taken as an example to define several indicators to measure classification performance, and these indicators also apply to ordinary classes which are not equipment states.

Suppose $N_{correct}$ being the number of correctly classified samples, N_{total} being the total number of samples in testing, the classification precision P can be written as:

$$P = \frac{N_{correct}}{N_{total}} \times 100\% \tag{1}$$

Suppose N_{n2f} being the number of normal state samples diagnosed to be fault states and N_{normal} being the total number of normal state samples, the false alarm rate R_{fa} is:

$$R_{fa} = \frac{N_{n2f}}{N_{normal}} \times 100\% \tag{2}$$

Suppose N_{f2n} being the number of fault states samples diagnosed to be normal state, and N_{fault} being the total number of fault states samples, the missing alarm R_{ma} can be written as:

$$R_{ma} = \frac{N_{f2n}}{N_{fault}} \times 100\% \tag{3}$$

Suppose N_{f2f} being the number of a fault state diagnosed sample to be other fault states, the fault recognized rate R_{fr} is:

$$R_{fr} = \frac{N_{f2f}}{N_{fault}} \times 100\% \tag{4}$$

B. CLASSIFICATION PERFORMANCE MEASUREMENT INDICATORS OF "ONE-VERSUS-OTHERS" CLASSIFIER

The performance measurement indicators of "one-vs-others" SVM classifier are shown as follows:

The confusion matrix of classification result is:

$$p = [p_{ii'}], \quad i = 1, 2, \ldots, M; \ i' = 1, 2, \ldots, M \tag{5}$$

where M is the number of categories needs to be classified.

Suppose $N_{ii'}$ being the number of samples with the i-th category classified to the i'-th and N_i being the total number of the i-th category, $p_{ii'}$ in (5) is:

$$p_{ii'} = \frac{N_{ii'}}{N_i} \times 100\% \tag{6}$$

Suppose p_{k1} being the classification precision of the k-th classifier and p_{i2} ($1 \leq i \leq M$ and $i \neq k$) being the classification precision of the i-th classifier, the classification precision P shown is:

$$P = P_k p_{k1} + \sum_{i=1, i\neq k}^{M} P_i p_{i2} \tag{7}$$

where P_i ($1 \leq i \leq M$) is the prior probability of the i-th classifier, for a certain testing samples, P_i is:

$$P_i = \frac{N_i}{\sum_{j=1}^{M} N_j} \tag{8}$$

If there were a normal state in samples, the first category is defined as normal state, and the false alarm rate R_{fa} is:

$$R_{fa} = 1 - p_{11} \tag{9}$$

The missing alarm rate R_{ma} is:

$$R_{ma} = \sum_{i=2}^{M} P_{mai} p_{i1} \tag{10}$$

where P_{mai} is the fault states correctly diagnosed (for example, $i = a$ is fault state a and classified as fault state a) and it can be calculated as follow:

$$P_{mai} = \frac{N_i}{\sum_{j=2}^{M} N_j}, \quad i = 2, 3, \ldots, M \tag{11}$$

It should be noticed that no missing alarm rate and false alarm rate occur without any normal states in samples, and the classification precision can be calculated by (7).

The fault recognized as another fault R_{fr} is:

$$R_{fr} = 1 - P_k p_{k1} - \sum_{i=1, i\neq i'}^{M} P_i p_{i2} \tag{12}$$

III. SERIAL SVM CLASSIFIER MODEL

As to the classification problem with M ($M > 2$) categories, ($M - 1$) binary classifiers are designed, and the priorities of these categories are determined, then cascading them referring to their priorities respectively. And the priority of each category is not sure. For different states or experiment data, it may have different priorities, and it can be decided by experimenting. Their priorities from high to low are denoted as the 1-st category, the 2-nd category, ..., and at last the M-th category. Then the former s ($s = 1, 2, \ldots, M - 1$) levels find out the corresponding category s-th together, and the ($M - 1$) levels classifiers are designed using the rest two categories of ($M - 1$)-th and M-th which are not recognized by the former ($M - 2$) levels.

The design of s-th ($s = 1, 2, \ldots, M - 1$) level based on feature selection is given as follows:

The ($M - s + 1$) categories which cannot be recognized by the former $s - 1$ levels are denoted as $\{V_{i1}, V_{i2}, \ldots, V_{i(M-s+1)}\}$ and the many-to-one binary classifier is denoted by $\Lambda_{(M-s+1)k}$ ($k = 1, 2, \ldots, M - s + 1$). One category is i_k-th and the other is the rest of ($M - s + 1$) categories except the i_k-th. The classification precision is denoted as $P_{(M-s+1)k}$, the total number of feature subset is denoted as q, and the objective functions of serial SVM multi-class classifier are:

$$1)\ \max \max_{k=1}^{M-s+1} \{P_{(M-s+1)k}\} \tag{13}$$

$$2)\ \min q \tag{14}$$

The priorities of these two functions are from high to low, that means if and only if (13) has the max value, the min q will be considered. From (13), we can see that during each level the class with the highest classification precision will be classified. If each level reaches the maximum classification precision and minimum size of feature subset, the serial SVM classifier model will also reach the maximum classification precision and minimum size of feature subset. So the design purpose of s-th level is to maximize the classification precision and minimize feature subset. In the selection of input feature subsets and the construction of classifiers, the samples are divided into two parts, one of which is used to train classifier and the other to calculate the value of (14) to judge the selected subsets.

For the classification of M categories, SVM binary classifier is used as base classifier, the framework of serial multi-class classifier based on feature selection is showing in Figure 1. When samples are input to the serial multi-class classifier, the first category can be recognized by SVM1, the second recognized by SVM1 and SVM2. By parity of reasoning, the (M-2)-th category recognized by the former ($M - 2$) SVM classifiers, and the categories ($M - 1$)-th and M-th can be recognized by all the ($M - 1$) SVM classifiers. Mainly different from traditional cascade classifiers' adding predictions of the level to the next level, the cascade model that we proposed classifies the class with highest classification precision after each level, and the number of classes is reduced continuously.

The categories' separability of themselves won't affect classifiers of each level simultaneously because they are cascaded, which makes the model more convenient to dig features' classification potential.

IV. ANT COLONY OPTIMIZATION (ACO) REALIZATION

A. INTRODUCTION OF ACO

In multi-objective optimization, there is no global optimal solution for all objective functions, but a non-inferior solution set called Pareto optimal solution set could be found. The purpose of multi-objective optimization is to find a set of

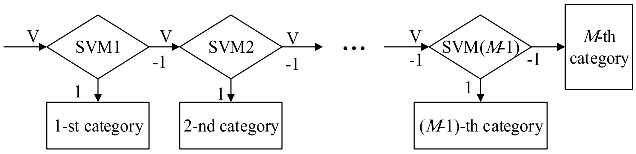


FIGURE 1. Framework of serial SVM multi-class classifier with M categories.

optimal solutions close to the Pareto optimal solution, and the decision maker can find the satisfactory solution set according to demands.

A kind of second best subsets solutions based on meta-heuristics algorithms for global search are developed and applied widely, such as simulated annealing, genetic algorithms, artificial immune algorithm, particle swarm algorithm, bat algorithm, firefly algorithm and ACO [20].

ACO is a classical kind of meta-heuristics algorithm, it is developed rapidly and employed widely for its strong ability to find better solution, better robustness, information positive feedback, parallel distributed computation and being easy to combine with other meta-heuristics algorithms. Especially in solving complex multi-objective combination optimization problems, ACO shows its superiority. Liu et al. proposed a new multi-objective ACO to solve high-dimensional objective subset selection problems which shows superior performance than other multi-objective methods like genetic algorithm and decomposition [21], Dilip et al. utilized multi-objective ACO to realize multi-criteria website optimization where a comparison between multi-objective ACO and multi-objective genetic algorithm has been done to show great performance of multi-objective ACO [22]. In this paper, ACO is used to solve each level's objective functions.

The steps of ACO are as follows:

Firstly, directed graph must be constructed for specific question, and it is the presentation of practical problem.

Secondly, the ants move according to the path transfer probability $h_{ij}(t)$ introduced from [23]:

$$h_{ij}(t) = \begin{cases} \dfrac{\tau_{ij}^{\alpha}(t-1)\eta_i^{\beta}}{\sum\limits_{e_{bj}\notin tabu_m} \tau_{bj}^{\alpha}(t-1)\eta_b^{\beta}}, & e_{ij} \notin tabu_m \\ 0, & otherwise \end{cases} \tag{15}$$

where α, β represent the importance of pheromone and heuristic factor respectively. e_{ij} is the edge in row i and column j, τ_{ij} represents how much pheromone in edge e_{ij}, η_i is the heuristic factor and represents the expectation of the i-th element, t is the moment, $tabu_m$ $(m = 1, 2, \ldots, M)$ is *tabu*, and it stores the edges which m-th ant traveled.

Later, each ant will be killed after finishing its travel, and the iteration will be finished if the colony is died. Then a solution which satisfies the constraint condition will be gotten.

Thirdly, the pheromone needs to be updated after each iteration, using equivalent route pheromone updating strategy proposed in [23]:

$$\tau_{ij}(t) = \begin{cases} (1-\rho)\tau_{ij}(t-1) + \dfrac{\Phi'(tabu^t)}{Q}, & e_{ij} \in \Gamma(tabu^t) \\ (1-\rho)\tau_{ij}(t-1), & otherwise \end{cases} \tag{16}$$

where $tabu^t$ is the path solution needed updating, $\Gamma\left(tabu^t\right)$ is the equivalent path, $\frac{\Phi'(tabu^t)}{Q}$ is the update equation of pheromone, $\Phi'\left(tabu^t\right)$ is the value of objective function where the pheromone of path solution needs to be strengthened, Q is constant.

The iteration will be continued until the solution cannot change or reach to the maximum number of iteration which has been set before.

The flow diagram of ACO is shown in Figure 2.

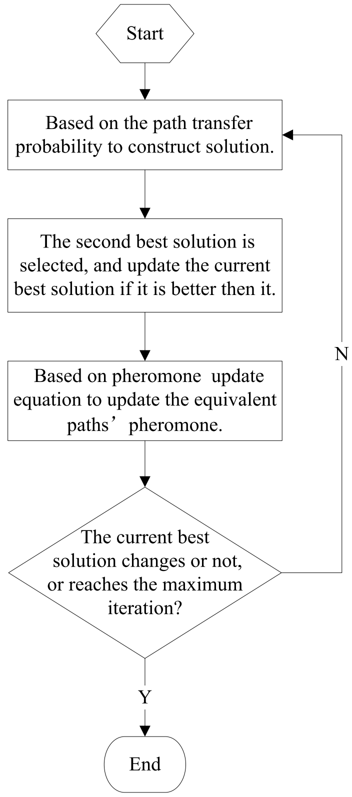


FIGURE 2. Flow diagram of ACO.

B. B.REALIZATION OF ACO

ACO is used to solve the serial multi-class classifier, the serial classifier optimal model in section III is analyzed as follows:

1) The value of q varying from 5 to 10 leads to better efficiency [24]. Generally, the binary problem is easier to be separated than the multi-class ones and may need fewer features to reach a relatively high classification precision since each level is a binary classifier. Here q varies from 1 to 10.

2) Fixing k, for classifier $\Lambda_{(M-s+1)k}$, searching q from small to large and comparing the objective values of

(13) to get the optimization solutions $P_{(M-s+1)kmax}$ and $q_{(M-s+1)kmin}$.

3) For classifier $\Lambda_{(M-s+1)k}$ $(k = 1, 2, \ldots, (M - s + 1))$ of the s-th level, initialize the global optimal solutions with $P_{current} = 0$, $q_{current} = 0$, and design classifier for each k respectively. If $k = 1$, according to 2), the global optimal solutions are $P_{current} = P_{(M-s+1)1max}$ and $q_{current} = q_{(M-s+1)1min}$. If $k = 2$, the optimal solutions are $P_{(M-s+1)2max}$ and $q_{(M-s+1)2min}$, and compare them with $P_{current}$ and $q_{current}$. Implement $P_{current} = P_{(M-s+1)2max}$ and $q_{current} = q_{(M-s+1)2min}$ when $P_{(M-s+1)2max} > P_{current}$ or $P_{(M-s+1)2max} = P_{current}$ and $q_{(M-s+1)2min} < q_{current}$. Do nothing under other conditions. Repeat this process until $k = (M - s + 1)$.

4) In the end, optimal feature subset, training results and state category of current priority are included in the optimal solutions.

According to the analysis above, the pseudocode of the s-th level classifier is shown in Algorithm 1.

Algorithm 1 Algorithm to Solve the s-th level's Classifier

```
Input:
  All collected features
Begin
  Initialization
   for ite = 1: max_ite
     for q =1:10
       Take function (13) as the objective function
       According to [23], use ACO to search
       Update optimal solutions according to 2)
     endfor
     Update optimal solutions according to 3)
     if (P_b =1, q_b =1)
       Break
     endif
   endfor
   Output optimal solutions[[period]]
End
Output:
   Optimal feature subset
   Training results
   Current priority state category
```

V. EXPERIMENTS ANALYSIS

A. DATA PREPARATION

The vibration signals of a cylinder head from a certain model of diesel engines are used as test data. The vibration signals are sampled by the sampling system shown in Figure 3.

The six right cylinder heads of the diesel engine (12150L, V12) are taken as test point. The cylinder head vibration signals, 512 points per working cycle, are collected with the rotation speed n =1000 r/m in idle case. The data of the three datasets for test are collected under five working states respectively of 1) normal, 2) misfire, 3) large clearance of exhaust valve, 4) small clearance of exhaust valve and 5) low pressure of injector.

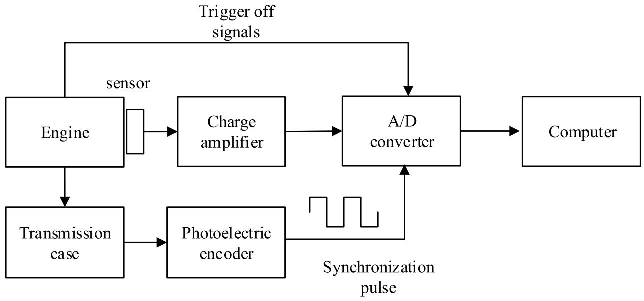


FIGURE 3. Synchronous sampling system for engine vibration signal.

TABLE 1. The number of data groups of each state in three datasets.

Datasets	1)	2)	3)	4)	5)
1	60	60	60	60	60
2	60	15	60	60	30
3	60	30	15	15	30

The feature parameter system of the cylinder head vibration signals is constructed by combining the lifting wavelet packet transform [25] as follows:

1) Extract 12 statistic parameters of original time domain signals. These parameters are mean value, mean amplitude, square root amplitude, standard deviation, effective value, peak-to-peak value, shape factor, impulse factor, crest factor, skewness, kurtosis and clearance factor respectively.

2) Extracted signals are resolved into 2 levels of lifting wavelet packet, which make the ignition explosion segment excitation response focus on the node (2, 1) (the second frequency band), the seating section in response to incentives of intake valve and exhaust valve focuses on nodes (2, 2) and (2, 3). The 12 feature parameters of the 4 nodes in 1) and standardized relative energy are extracted [26].

3) The time domain signals in the corresponding frequency bands can be obtained through single branch reconstruction for each node. And the statistic feature parameters in 1) are extracted from each of the 4 single branch reconstruction signals.

The 112 feature parameters are numbered in the following order: 12 statistic feature parameters, 4×12 feature parameters from the second level of lifting wavelet packet, 4×12 feature parameters from single branch reconstruction signal, and 4 standardized relative energies. The feature set is $Set = \{t|t = 1, 2, \ldots, 112\}$.

The features in $Set = \{t|t = 1, 2, \ldots, 112\}$ are extracted for each group of collected data. If there are M state categories, V_{ij} is denoted as the j-th feature vector sample of the i-th state category, the first category defined as normal state, N_i is the number of i-th state category, then the row of feature

TABLE 2. Results of ensemble classifiers design of three datasets.

Datasets	Classifier of each level	Samples' categories	q	set^q	P
1	SVM1	{2)},{1), 3), 4), 5)}	1	{76}	100
	SVM2	{3)}, {1), 4) ,5)}	2	{27,99}	100
	SVM3	{5)}, {1), 4) }	8	{19,36,51,72,79,80,86,112}	97.92
	SVM4	{1)},{4)}	3	{14,74,75}	96.88
2	SVM1	{2)},{1), 3), 4), 5)}	1	{86}	100
	SVM2	{3)}, {1), 4) ,5)}	2	{3,35}	100
	SVM3	{5)}, {1), 4) }	7	{17,40,47,61,64,74,77}	97.50
	SVM4	{4)},{1)}	9	{57,59,62,68,75,77,79,110,111}	91.67
3	SVM1	{2)},{1), 3), 4), 5)}	1	{87}	100
	SVM2	{3)}, {1), 4) ,5)}	1	{26}	100
	SVM3	{4)}, {1), 5) }	7	{61,74,75,76,89,95,98}	98.81
	SVM4	{1)},{5)}	2	{14,74}	100

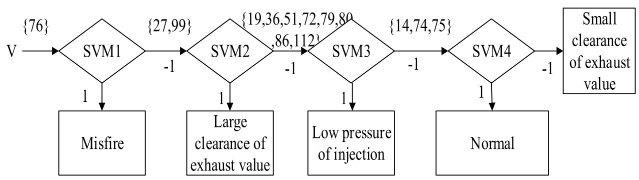


FIGURE 4. Framework of serial SVM classifier.

matrix V is

$$V_{ij} = \left(v_{ij1}, v_{ij2}, \ldots, v_{ijT}\right), \quad i = 2, 3, \ldots, M; \quad j = 1, 2, \ldots, N_i \qquad (17)$$

In this paper, $T = 112$, $M = 5$, and Ni is shown in Table 1.

From Table I, we can see that dataset 1 is a balanced dataset and other two are unbalanced.

B. REALIZATION OF SERIAL SVM CLASSIFIER

In order to determine the parameters of the binary SVM classifier of each level, we adopt the method of grid search and find that setting $\sigma = 15$ and $C = 80$ for the binary SVM classifier of each level will get better results. The data in the previous section are used to test, the former fifth of each class samples used to train the classifier, and the rest used to calculate values of objective functions. ACO is employed to solve the three datasets. The optimal feature subset, classification precision etc. are shown in Table II.

From Table II, take dataset 1 for example, the priorities of state categories from high to low are 2) misfire, 3) large clearance of exhaust valve, 5) low pressure of injector, 1) normal, and 4) small clearance of exhaust valve. The classifier is designed based on this result with the framework shown in Figure 4.

For dataset 1, inputting all test samples to the model in Figure 4, the whole classification precision is 96.67% and the optimal feature set subset which has 14 features by assembling each level's subset is written by $\text{subset}^{14} = \{14, 19, 27, 36, 51, 72, 74, 75, 76, 79, 80, 86, 99, 112\}$, the confusion matrix of classification results of the serial SVM multi-class classifier are shown in Table III.

TABLE 3. Confusion matrix of classification results for dataset 1.

$p_{ii'}$ (%)	1	2	3	4	5
1	91.67	0	0	6.25	2.08
2	0	100	0	0	0
3	0	0	100	0	0
4	6.25	0	0	93.75	0
5	2.08	0	0	0	97.92

TABLE 4. Confusion matrix of classification results for dataset 2.

$p_{ii'}$ (%)	1	2	3	4	5
1	93.75	0	0	4.17	2.08
2	0	100	0	0	0
3	0	0	100	0	0
4	8.33	0	0	91.67	0
5	8.33	0	0	0	91.67

TABLE 5. Confusion matrix of classification results for dataset 3.

$p_{ii'}$ (%)	1	2	3	4	5
1	97.92	0	0	0	2.08
2	0	100	0	0	0
3	0	0	100	0	0
4	0	0	0	100	0
5	0	0	0	0	100

From Table III, it can be calculated that $R_{fa} = 8.33\%$, $R_{ma} = 2.08\%$ and $R_{fr} = 0\%$. And classification precision is quite high of each state category as no value is below 80%, it has no fault recognized rate, normal diagnosed to be fault is caused by low pressure of injector and small clearance of

TABLE 6. Testing results of classifier's performance indicators.

Dataset	Classifier	N	Q	T_m	P(%)	R_{fa}(%)	R_{ma}(%)	R_{fr}(%)
1	Classifier1	10	6	0.0305	89.00	15.00	7.50	1.50
	Classifier2	5	5	0.0196	89.00	17.50	5.00	2.50
	Classifier3	100	112	0.0256	89.88	14.96	7.96	0.96
	Classifier4	4	14	0.0132	**96.67**	8.33	2.08	0
2	Classifier1	10	7	0.0311	88.00	20.00	5.45	3.64
	Classifier2	5	7	0.0203	87.33	25.00	4.55	3.64
	Classifier3	100	112	0.0197	89.68	16.74	7.56	0.41
	Classifier4	4	18	0.0167	**95.00**	6.25	4.55	0
3	Classifier1	10	4	0.0203	87.50	12.50	10.94	1.56
	Classifier2	5	4	0.0157	84.62	0	23.40	1.56
	Classifier3	100	112	0.0199	83.78	8.49	20.44	0.72
	Classifier4	4	10	0.0104	**99.17**	2.08	0	0

exhaust valve classifying to normal; missing alarm is caused by small clearance of exhaust value and low pressure of injector recognized as normal state.

The results for dataset 2 and dataset 3 are also shown in Table II, the confusion matrix of classification results are shown intuitively in the Table IV and Table V.

From Table IV and Table V, for dataset 2, it can be calculated that $R_{fa} = 6.25\%$, $R_{ma} = 4.55\%$ and $R_{fr} = 0\%$, and for dataset 3, it can be calculated that $R_{fa} = 2.08\%$, $R_{ma} = 0\%$ and $R_{fr} = 0\%$. Generally speaking, we can see that this method also gets good results in unbalanced datasets, and for some datasets with a reasonable sample size, it may have a higher classification precision than balanced datasets such as dataset 3. Detailed test results are given in Table VI in the next section.

C. EXPERIMENTS

Compare the classifying effects of "one-versus-one" SVM classifier, "one-versus-others" SVM classifier, the ensemble classifier proposed in [16] and the proposed classifier in this paper as follows. These four classifiers are labeled as Classifier 1, Classifier 2, Classifier 3 and Classifier 4 respectively to represent them more simply.

The experiment is conducted on a PC with Intel(R) Xeon(R) CPU E5-2609 2.5GHz, 64.00G RAM. The performance measurement of classifier contains the number of binary classifiers N (scale of classifier), total number of input features Q, mean recognize time T_m for each sample, classification precision P, false alarm rate R_{fa}, missing alarm rate R_{ma} and fault reorganization rate R_{fr}. The test results are shown in Table VI.

Impose $\sigma = 5$ and $C = 30$ for **Classifier 1** and **Classifier 2** while $\sigma = 15$ and $C = 80$ for **Classifier 4**. All of their feature subsets are obtained through ACO searching [24].

From Table VI, **Classifier 3** has the most binary classifiers and **Classifier 4** has the fewest. The feature's number for different datasets are different except **Classifier 3**. Though **Classifier 4** has the highest feature number compared to **Classifier 1** and **Classifier 2**, it has the least mean reorganization time and the highest classification precision. It presents that both in dealing with balanced dataset and unbalanced dataset, **Classifier 4** owns a better performance in terms of efficiency and precision, it may not all have lower missing alarm rate and fault reorganization rate and zero false alarm rate compared to the other classifiers, but it is not hard to find that values of these rates remain at a low level and change little. Because of lacking feature selection, **Classifier 3** focuses on sample selection and cannot get a good result when facing unbalanced data. In comparison, **Classifier 4** is really worthy for online fault diagnosis which focuses more on timeliness and classification precision.

VI. CONCLUSION

This paper proposes a serial SVM multi-class classifier based on feature selection. With the goal of improving classification precision and reducing feature subset size, the serial multi-class classifier is constructed with "one-versus-others" SVM classifier as the base classifier and solved using ACO. Signals of real equipment are used to verify its performance and conclusions are drawn as follows:

Proposed serial multi-class classifier requires fewer base classifiers as "one-versus-others" SVM is used as base classifier, the reorganization time is less through controlling of the input feature number with the goal of improving or stabilizing classification precise. Furthermore, the total classification precision can be improved highly for its distinguished classes. At last, the performances of proposed classifier are verified to be excellent by experimental results. Based on all these merits, the serial multi-class classifier is competent for recognizing problems such as online fault diagnosis that requiring high classification precision and less time.

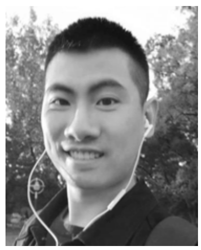
REFERENCES

[1] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[2] P. Vorraboot, S. Rasmequan, K. Chinnasarn, and C. Lursinsap, "Improving classification rate constrained to imbalanced data between overlapped and non-overlapped regions by hybrid algorithms," *Neurocomputing*, vol. 152, pp. 429–443, Mar. 2015.
[3] V. López, S. D. Río, J. M. Benítez, and F. Herrera, "Cost-sensitive linguistic fuzzy rule based classification systems under the mapreduce framework for imbalanced big data," *Fuzzy Sets Syst.*, vol. 258, pp. 5–38, Jan. 2015.
[4] P. Cao, D. Zhao, and O. Zaiane, "An optimized cost-sensitive SVM for imbalanced data learning," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2013, pp. 280–292.
[5] J. Tian, H. Gu, and W. Liu, "Imbalanced classification using support vector machine ensemble," *Neural Comput. Appl.*, vol. 20, no. 2, pp. 203–209, 2011.
[6] B. Krawczyk, M. Woźniak, and F. Herrera, "Weighted one-class classification for different types of minority class examples in imbalanced data," in *Proc. IEEE Symp. Comput. Intell. Data Mining*, Dec. 2014, pp. 337–344.
[7] B. Krawczyk and M. Woźniak, "Hypertension type classification using hierarchical ensemble of one-class classifiers for imbalanced data," in *ICT Innovations 2014* (Advances in Intelligent Systems and Computing), vol. 311. 2015, pp. 341–349.
[8] S. Kim, "Weighted K-means support vector machine for cancer prediction," *SpringerPlus*, vol. 5, no. 1, p. 1162, 2016.
[9] B. Richhariya and M. Tanveer, "A robust fuzzy least squares twin support vector machine for class imbalance learning," *Appl. Soft Comput.*, vol. 71, pp. 418–432, Oct. 2018.
[10] L. Cao and H. Shen, "Combining re-sampling with twin support vector machine for imbalanced data classification," in *Proc. 17th Int. Conf. Parallel Distrib. Comput., Appl. Technol.*, Dec. 2016, pp. 325–329.
[11] F. Wang *et al.*, "Imbalanced data classification algorithm with support vector machine kernel extensions," in *Evolutionary Intelligence*. 2017, pp. 1–7.
[12] D. Zhang, L. Jiao, X. Bai, S. Wang, and B. Hou, "A robust semi-supervised SVM via ensemble learning," *Appl. Soft Comput.*, vol. 65, pp. 632–643, Apr. 2018.
[13] W. Wang, X. Du, and N. Wang, "Building a cloud IDS using an efficient feature selection method and SVM," *IEEE Access*, vol. 7, pp. 1345–1354, 2019.
[14] M. Gao, X. Hong, and C. J. Harris, "Construction of neurofuzzy models for imbalanced data classification," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1472–1488, Dec. 2014.
[15] M. Last, H. Bunke, and A. Kandel, "A feature-based serial approach to classifier combination," *Pattern Anal. Appl.*, vol. 5, no. 4, pp. 385–398, 2002.
[16] S. H. Min, "Integrating instance selection and bagging ensemble using a genetic algorithm," *Int. J. Appl. Eng. Res.*, vol. 11, no. 7, pp. 5060–5066, 2016.
[17] L. Martí, N. Sanchez-Pi, J. M. M. López, and A. C. B. Garcia, "On the combination of support vector machines and segmentation algorithms for anomaly detection: A petroleum industry comparative study," *J. Appl. Log.*, vol. 24, pp. 71–84, Nov. 2017.
[18] G. Yao, H. Zeng, F. Chao, C. Su, C.-M. Lin, and C. Zhou, "Integration of classifier diversity measures for feature selection-based classifier ensemble reduction," *Soft Comput.*, vol. 20, no. 8, pp. 2995–3005, 2016.
[19] R. Diao, F. Chao, T. Peng, N. Snooke, and Q. Shen, "Feature selection inspired classifier ensemble reduction," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1259–1268, Aug. 2014.
[20] Y. Zhao, S. Xin, and L. Liu, *Emerging Meta Heuristic Optimization Method*. Beijing, China: Science Press, 2013.
[21] Y. Liu *et al.*, "An ant colony optimization algorithm for multiobjective subset selection problem," to be published.
[22] K. Dilip and T. V. V. Kumar, "Multi-criteria website optimization using multi-objective ACO," in *Proc. 4th Int. Conf. Rel., Technol. Optim. (INFOCOM)*, Noida, India, Sep. 2016, pp. 1–6.
[23] J.-J. Cao, P.-L. Zhang, Y.-X. Wang, G.-Q. Ren, and J.-P. Fu, "Graph-based ant system for subset problems," *J. Syst. Simul.*, vol. 20, no. 22, pp. 6146–6150, 2008.
[24] B.-S. Yang, D.-S. Lim, and J.-L. An, "Vibration diagnostic system of rotating machinery using artificial neural network and wavelet transform," in *Proc. 13th Int. Congr. Condition Monit. Diagnostic Eng. Manage.*, Houston, TX, USA, 2000, pp. 923–932.
[25] J.-J. Cao, P.-L. Zhang, Y.-T. Zhang, and G.-Q. Ren, "Feature extraction of an engine cylinder head vibration signalbased on lifting wavelet package transformation," *J. Vibrat. Shock*, vol. 27, no. 2, pp. 34–37, 2008.
[26] J.-J. Cao *et al.*, "Lifting wavelet package decomposing algorithm under the best bases and its applications in de-nosing for vibration signal," *J. Vibrat. Shock*, vol. 27, pp. 114–116, 2008.

JIANJUN CAO (M'90) received the M.S. degree from PLA Army Engineering University, Nanjing, China, in 2016, where he is currently pursuing the Ph.D. degree.

He is currently an Associate Research Fellow and a Master's Supervisor, mainly engaged in data quality control, data governance, data intelligence analysis, and application. He has presided over seven provincial key projects, has been authorized with four provincial and ministerial awards, four authorized invention patents, and four software copyrights, and has published more than 80 academic papers up to now. His research interests include data mining, machine learning and multi-modal data analysis. He received the National Excellent Doctoral Dissertation Nomination Award. He was a recipient of the China Postdoctoral Science Foundation Special Fund.



GUOJUN LV (M'95) received the bachelor's degree from the Nanjing University of Posts and Telecommunications, Nanjing, in 2017. He is currently pursuing the master's degree with PLA Army Engineering University, Nanjing. His research interests include machine learning, cross-modal retrieval, and data quality.

CHEN CHANG (M'94) received the bachelor's degree from the Nanjing University of Posts and Telecommunications, Nanjing, in 2017. He is currently pursuing the master's degree with PLA Army Engineering University, Nanjing. His research interests include data mining, data quality, and truth discovery.

HONGMEI LI (F'90) received the Ph.D. degree from PLA Army Engineering University, Nanjing, in 2018. Her research interests include collaborative filtering recommendation, data analysis and mining, and natural language processing.