Springer

KSME

Journal of Mechanical Science and Technology 26 (2) (2012) 601~608
www.springerlink.com/content/1738-494x
DOI 10.1007/s12206-011-1216-z

Fault diagnosis of rotary kiln using SVM and binary ACO †

Ouahab Kadri*, Leila Hayet Mouss and Mohamed Djamel Mouss
Laboratory of Automation & Production Engineering, University of Batna, 1 Rue Chahid Bokhlof, Batna, Algeria

(Manuscript Received April 21, 2011; Revised September 7, 2011; Accepted October 11, 2011)

Abstract

This paper proposes a novel hybrid algorithm for fault diagnosis of rotary kiln based on a binary ant colony (BACO) and support vector machine (SVM). The algorithm can find a subset selection which is attained through the elimination of the features that produce noise or are strictly correlated with other already selected features. The BACO algorithm can improve classification accuracy with an appropriate feature subset and optimal parameters of SVM. The proposed algorithm is easily implemented and because of use of a simple filter in that, its computational complexity is very low. The performance of the proposed algorithm is evaluated through two real Rotary Cement kiln datasets. The results show that our algorithm outperforms existing algorithms.

Keywords: Binary ant colony algorithm; Fault diagnosis; Feature selection; Support vector machine

---

1. Introduction

The principal function of the condition monitoring is to check the operating condition of the system. Our work falls under the condition monitoring and diagnosis of industrial system which is an important field of engineering study (in our case is a Rotary Cement kiln, see Fig. 1).

The diagnosis is made up of two parts which are detection and the diagnosis. The phase of detection makes it possible to determine the state of the system as being normal or abnormal. The phase of diagnosis consists in identifying the failing components and to find the causes starting from a whole of symptoms observed. In substance, diagnosis is considered as a classification problem [1-3].

An industrial system is described by a vector of numeric or nominal features. Some of these features may be irrelevant or redundant. Avoiding irrelevant or redundant features is important because they may have a negative effect on the accuracy of the classifier [1, 2]. In addition, by using fewer features we may reduce the cost of acquiring the data and improve the comprehensibility of the classification model (Fig. 2).

Feature extraction and subset selection are some frequently used techniques in data pre-processing. Feature extraction is a process that extracts a set of new features from the original features through some functional mapping [4].

Subset selection is different from feature extraction in that

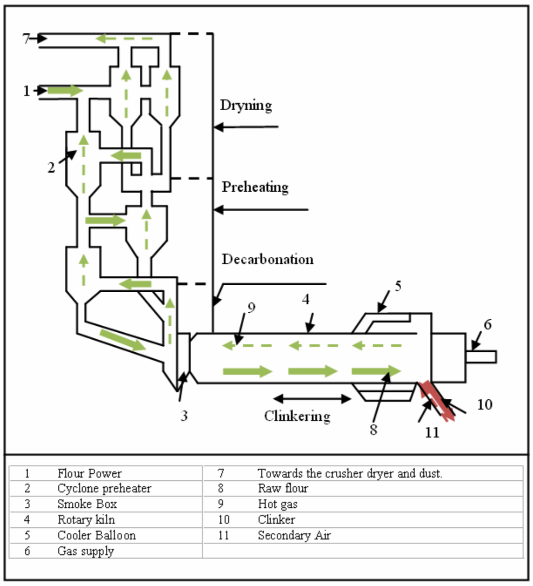


Fig. 1. Rotary cement kiln.

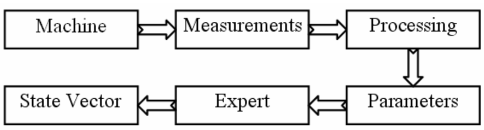


Fig. 2. Construction of the features subset.

no new features will be generated, but only a subset of original features is selected *and* feature space is reduced [5].

† This paper was recommended for publication in revised form by Editor Sung-Lim Ko

*Corresponding author. Tel.: +21333803396, Fax.: +21333803396
E-mail address: ouahabk@yahoo.fr

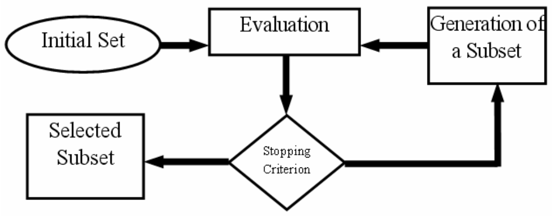


Fig. 3. Subset selection method.

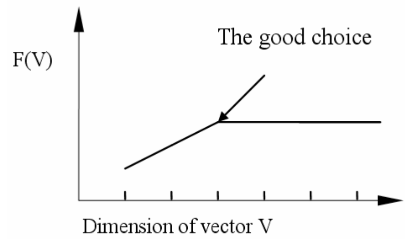


Fig. 4. The choice of the dimension of vector V.

The idea behind the selection approach is very simple and is shown in Fig. 3. Any method of selection of features consists of four essential points:

A starting subset, which represents the subset of features, initially is used by a search procedure. This set can be empty, or contains all the features or a random subset. The search procedure is the essential element of any method of selection. It turns over as result the features subset which answers the quality standard better. This criterion is returned by an evaluation function. This function determines the classification quality obtained by using a feature subset. A stopping criterion is used to finish the search procedure. This criterion depends to the evaluation function or with the configuration parameters which are defined by the user [6].

We present in this paper a hybrid approach based on ant colony optimization (ACO) and support vector machine (SVM) for feature selection problems using datasets from the field of industrial diagnosis.

This paper presents a novel approach for heuristic value calculation, which will reduce the set of available features. The rest of this paper is organized as follows. In section 2, different methods for feature selection problems are presented. An introduction on ACO applications in feature selection problems is discussed in Section 3. A brief introduction of SVM is presented in Section 4. In Sections 5 and 6, the proposed algorithm is discussed, followed by a discussion on the experimental setup, datasets used and the results.

2. Feature subset selection

Feature selection (FS) is included in discrete optimization problems. The whole search space for optimization contains all possible features subsets, meaning that its size is 2^n where n is the dimensionality (the features number). Usually FS algorithms involve heuristic or random search strategies in an attempt to avoid this prohibitive complexity. However, the optimality degree of the final feature subset is often reduced [7-9].

Two broad categories of optimal feature subset selection have been proposed based on whether feature selection is performed independently of the learning algorithm that constructs the classifier. They are the filter approach and the wrapper approach [10, 11]. The filter approach initially selects important features and then the classifier is used for classification while the wrapper uses the intended learning algorithm itself to evaluate the features usefulness [12]. The two famous algorithms of this category are sequential forward selection (SFS) and sequential backward selection (SBS) [6, 10]. In sequential forward selection, the features are sequentially added to an empty candidate set until the addition of further features does not decrease the criterion but in Sequential backward selection the features are sequentially removed from a full candidate set until the removal of further features increase the criterion. In our work, we use a hybrid wrapper/filter approach aiming to explore the qualities of both strategies and try to overcome some of their deficiencies [11].

The stopping criterion in Fig. 4 represents the dimension of the vector obtained by the algorithm where the quality standard does not evolve/move if we add another feature [13].

Where V represent the feature subset and F(V) is the evaluation function.

The first good use of ACO for feature selection seems to be reported in Ref. [13]. A. Al-Ani [13] proposes to use a hybrid evaluation measure that is able to estimate the overall performance of subsets as well as the local features importance. A classification algorithm is used to estimate the subsets performance. On the other hand, the local importance of a given feature is measured using the mutual information evaluation function. Vieira et al. [14] propose an algorithm for feature selection based on two cooperative ant colonies, which minimizes two objectives: the features number and the classification error. The first colony determines the features number (cardinality) and the second selects the features based on the cardinality given by the first colony. C.L. Huang [15] presents a hybrid ACO-based classifier model that combines ant colony optimization (ACO) and support vector machines (SVM). In his work, an ant's solution represents a combination of the feature subset and the classifier parameters, C and g, based on the radial basis function (RBF) kernel of the SVM classifier. The classification accuracy and the feature weights of the constructed SVM classifier are used to design the pheromone update strategy. Based on the pheromone table and measured relative feature importance, the transition probability is calculated to select a solution path for an ant. The major inconvenience with this method is the classifier parameters which are fixed during the program execution and they may have different value in each solution.

3. Ant colony optimization (ACO)

Ant colony optimization (ACO) is based on the cooperative behavior of real ant colonies, which are able to find the shortest path from their nest to a food source. ACO algorithms can be applied to any optimization problems that can be characterized as follows [16, 17]:

A finite set of components $C =\{ c_1, c_2, ..., c_N\}$ is given.

A finite L set of possible connections/transitions among the elements of C is defined over a subset C' of the Cartesian product $C \times C$, $L=\{C_iC_j\}|(c_i, c_j) \in C'\}$, $|L| \leq N2c'$.

For each $lCiCj \in L$ a connection cost function $JC_iC_j \equiv J(lC_iC_j, t)$, possibly parametrized by some time measure t, is defined.

A finite constraints set $\Omega \equiv \Omega(C, L, t)$ is assigned over the elements of C and L.

The states of the problem are defined in terms of sequences $s = (c_i, c_j, ... , c_k, ...)$ over the elements of C or of L. S' is a subset of S. The elements in S' define the problem's feasible states.

A neighbourhood structure is assigned as follows: the state s_2 is said to be a neighbor of s_1 if s_1 and s_2 are in S and the state s_2 can be reached from s_1 in one logical step, that is, if c_1 is the last component in the sequence determining the state s_1, it must exists $c_2 \in C$ such that $l_{C1C2} \in L$ and $s_2 \equiv \langle s_1, c_2 \rangle$.

A solution Ψ is an element of S' satisfying all the problem's requirements. A solution is said multi-dimensional if it is defined in terms of multiple distinct sequences over the C elements.

A cost $J_\Psi(L, t)$ is associated to each solution Ψ. $J_\Psi(L, t)$ is a function of all the costs J_{CiCj} of all the connections belonging to the solution.

It is worth mentioning that ACO makes probabilistic decision in terms of the artificial pheromone trails and the local heuristic information. This allows ACO to explore larger number of solutions than greedy heuristics. Another characteristic of the ACO algorithm is the pheromone trail evaporation, which is a process that leads to decreasing the pheromone trail intensity over time. Pheromone evaporation helps in avoiding rapid convergence of the algorithm towards a sub-optimal region [13, 16, 17].

4. Support vector machines

In our wrapper approach, we have used SVM as classifier. SVM is an attractive learning algorithm first introduced by Vapnik [18]. It has a competitive advantage compared to neural networks, and decision trees [19].

Given a data set $S =\{(x_1, y_1),...,(x_i, y_i),...,(x_m, y_m)\}$. Where $x_i \in R^N$ is features vector and $y_i \in \{-1, +1\}$ is a class label. The SVM goal is to find one of the forms

$$w\psi(x)+b=0 \quad \text{with} \quad y_i(w\psi(x_i)+b) \geq 1-\xi_i \tag{1}$$

that separates the S training dataset into two classes (positive and negative) (Fig. 5). In general, S cannot be partitioned by a linear hyperplane. However S can be transformed into higher dimensional feature space for making it linearly separable.

The mapping $\Psi(x)$ need not be computed explicitly; instead, an inner product Kernel of the form

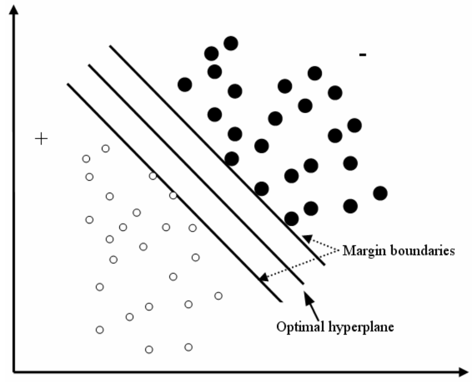


Fig. 5. Two-class SVM used in linear classification.

$$K(x_i,x_j)=\psi(x_i).\psi(x_j). \tag{2}$$

To solve the optimal hyperplane problem, we can construct a Lagrangian and transforming to the dual. Then, we can equivalently maximize

$$\sum_{i=1}^{m} a_i - \frac{1}{2}\sum_{i,j=1}^{m} a_i a_j y_i y_j K(x_i,x_j) \tag{3}$$

subject to

$$\sum_{i=1}^{m} a_i y_i = 0 \quad and \quad 0 \leq a_i \leq C. \tag{4}$$

For a test example z, we define the decision function as follow:

$$sign\left(\sum_{i=1}^{m} a_i y_i K(z_i,z)+b\right) \tag{5}$$

where

- α is the Lagrange multiplier.
- b is the bias term.
- C is the punishment parameter.
- w is the weight vector.

In the next section, we present our proposed SVM/Binary ACO algorithm [20], and explain how it is used for selecting an appropriate features subset.

5. Proposed approach

5.1 Description of the proposed approach

This research proposes a new implementation of Binary ACO algorithm [20] applied to feature selection, where the best features number is determined automatically. In this approach, each ant searches the same routine, and pheromone is left on each edge. As an intelligent body, each ant just chooses one edge of the two in the first step and in the second step the

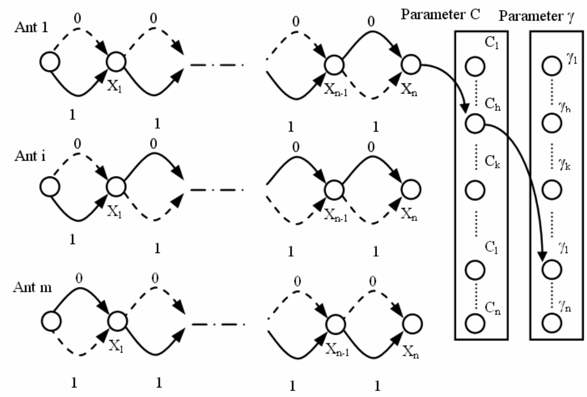


Fig. 6. The final net obtained by the BACO algorithm.

ant selects one value of parameters C and γ as shown in Fig 6. C and γ are the two parameters for the RBF kernel [3].

The intelligent behavior of ant is very simple, and the incidence matrix traversed by each ant needs only 2 × n + 2 steps, which to some extent solves the descriptive difficulty generated from long coding and the reduction of solution quality.

5.2 Probabilistic rule

Initially, the information quantity in each routine is randomly generated. During the movement, ant k shifts its direction according to the pheromone values concentration FP and the heuristic value FH. The heuristic value FH is computed using the Fisher discriminant criterion for feature selection [21, 22], which determines the importance of each feature, and it is described in more detail in Section 5.4. The probability that an ant k chooses the feature X_i is given by:

$$PS_{i1} = \frac{FP_{i1} + \frac{FP_{i0}}{Max(FH)} FH_i}{FP_{i1} + FP_{i0}}. \tag{6}$$

P_C represents the probability that an ant k chooses the parameter C. It is given by:

$$P_{C_{ki}} = \frac{FP_{C_{ki}}}{\sum_{i=1}^{n} \sum_{j=1}^{n} FP_{C_{iji}}} \tag{7}$$

and P_γ represents the probability that an ant *k* chooses the parameter γ. It is as follows:

$$P_{\gamma_{ki}} = \frac{FP_{\gamma_{ki}}}{\sum_{i=1}^{n} \sum_{j=1}^{n} FP_{\gamma_{iji}}}. \tag{8}$$

5.3 Updating rule

After all ants have completed their solutions, pheromone evaporation on all nodes is triggered, and then according to Eq. (7), pheromone concentration in the trails is updated.

$$FP \leftarrow (1-\rho) FP + \Delta FP \tag{9}$$

where $\rho \in]0, 1[$ is the pheromone evaporation and ΔFP is the pheromone deposited on the trails by the ant k that found the best solution for this tour:

$$\Delta FP = \frac{1}{1 + F(V) - F(V')} \tag{10}$$

where *F(V)* represents the best solution built since the beginning of the execution and *F(V')* represents the best solution built during the last tour.

F is the objective function of our optimization algorithm and V is the solution funded by the ant *k*.

The optimal subset is selected according to classifier performance and their length.

The results of this wrapper approach will be compared to a filter approach. The filter approach uses $F'(V)$ an evaluation function. F' is calculated using two concepts: the variance in each class and the variance between classes.

$$F'(V) = trace\left(\sum_W^{-1} . \sum_B\right) \tag{11}$$

where the variance matrix intra-class is calculated as follows:

$$\sum_W = \frac{1}{N} \sum_{C=1}^{M} \sum_{V=1}^{NR_C} (X_{CV} - m_C).(X_{CV} - m_C)^t \tag{12}$$

whereas the variance matrix inter-classes is calculated as follows:

$$\sum_B = \frac{1}{N} \sum_{C=1}^{M} (m_C - m).(m_C - m)^t \tag{13}$$

with:
- m : General gravity centre
- M : A classes number
- m_C : gravity Centre of the class number C
- X_{CV} : V th vector of the class number C
- NR_C : A vectors number of the class number C
- NR : the total number of vectors.

5.4 Heuristics

The heuristic value is computed using the Fisher discriminant criterion for feature selection [22]. Considering a classification problem with *M* possible classes, the Fisher discriminant criterion is described as follows:

$$FH(\alpha)=\sum_{c=1}^{M}\sum_{\substack{r=1\\ r\neq c}}^{M-1}\frac{m_c(\alpha)-m_r(\alpha)}{N_c\sigma_c^2(\alpha)-N_r\sigma_r^2(\alpha)} \quad (14)$$

where:

M represents the class's number;

$m_C(a)$ represent the gravity centre of the class number *C* by considering only the parameter α it is calculated as follows:

$$m_c(\alpha)=\frac{1}{N_c}\sum_{v=1}^{N_c}X_{cv}(\alpha) \quad (15)$$

with X_{cv} is the number *v* of the class number *C*. the value of *NR* equal to the vectors number of the class in question is the vector.

$\sigma_r^2(\alpha)$ is the *variance* of the component α of the vectors of the class number *C*.

$$\sigma_r^2(\alpha)=\frac{1}{N_c}\sum_{v=1}^{N_c}\left[X_{cv}(\alpha)-m_c(\alpha)\right]^2 \quad (16)$$

Algorithm 1 presents the description of the Binary ACO-SVM feature selection algorithm.

The time complexity of proposed algorithm is *O(Im)*, where *I* is the iterations number, and m the ants number. This can be seen from Fig. 6. In the worst case, each ant selects all the features. As the objective function is evaluated after all ants have completed their solutions, this will result in m evaluations. After *I* iterations, the objective function will be evaluated *Im* times.

Algorithm 1. Binary ACO/SVM feature selection algorithm.

1	Initiate the pheromone of the net;
2	Compute the $FH(\alpha)$ using (12);
3	Ants search using (6), (7) & (8);
4	Evaluate the solutions founded by the ant colony algorithm using SVM classifier, and reserve the optimal;
5	Upgrade pheromone of the net by optimal solution using (9);
6	Judge whether the stopping condition of the qualification is met, if qualified, ends; otherwise, Go to step 3;

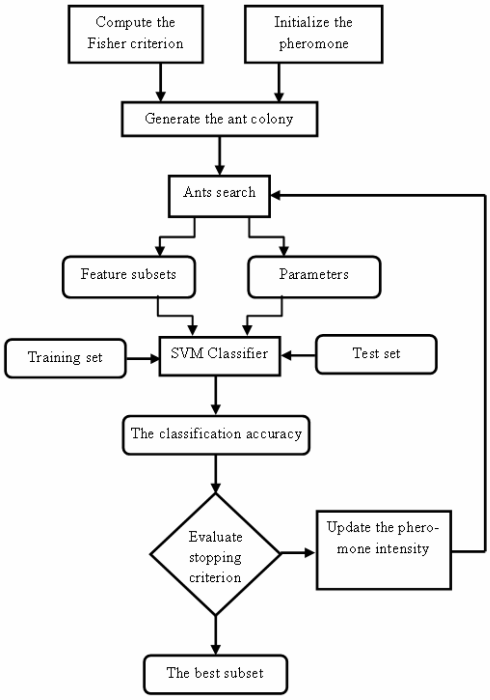


Fig. 7. SVM-Binary ACO feature selection algorithm.

6. Experimental results

6.1 Test data

The experimental results comparing the binary ACO algorithm with genetic algorithm are provided for two industrial datasets of Rotary Cement kiln (RCK1 and RCK2) [1]. RCK1 consists of 200 recordings which represent 4 classes. RCK2 consists of 500 recordings which represent 2 classes (Table 1). Cement rotary kiln is the most essential element of a cement factory whose outcome is cement clinker. A rotary kiln is a cylinder with a length of around 70 meters and a diameter of around 5 meters in a factory with a producing capacity about 1560 tons of clinker in a day. The kiln is rotated by a powerful electrical motor. The temperature in the hottest point in the kiln is up to 4500ºC. The typical clinker composition is CaO= 65 ± 3%, SiO_2= 21 ± 2%, Al_2O_3= 5 ± 1.5%, and FeO_3 = 3 ± 1% [23].

6.2 Parameters of the selection algorithm

Like any other algorithm, before passing to the selection phase. Some parameters should be fixed. This problem represents one of the disadvantages of the bio-mimetic methods. Since the parameters values are related to the number of individuals and the data distribution on the representation area. The following table presents the parameters values of our algorithm. These parameters are fixed after the execution of several simulations by using as entered a restricted whole of data. The search range for parameter *C* is [2^3, 2^{11}] and the search range for parameter γ is [2^{-12}, 2^2]. The search ranges are divided into many discrete search points with discretization interval lengths of 0.25 [15].

6.3 Heuristic factor FH

Heuristic factor FH is taken into account that by the ants which have a behavior related to the probability PS. The ants which have a random behavior are used to discover new search space. The following figures represent the heuristic values factor FH by using RCK1 and RCK2 datasets.

According to Fig. 8, we notice that the 41st feature has the greatest value of FH. Consequently it will be present in the final subset.

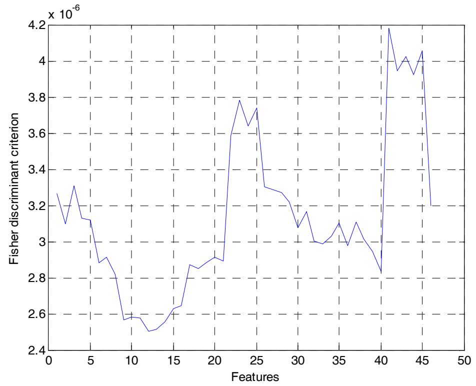


Fig. 8. The Fisher discriminant criterion for RCK1 dataset.

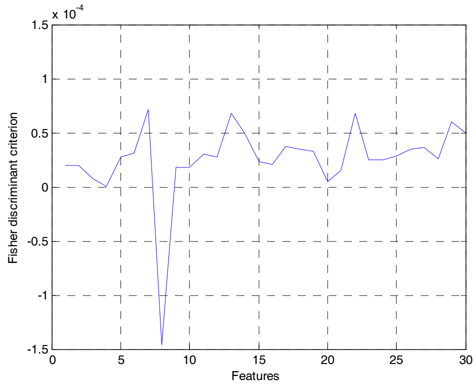


Fig. 9. The Fisher discriminant criterion for RCK2 dataset.

6.4 Results

We tested the performances of our algorithm by using the following classification error measure:

$$p = 1 - \left(\frac{1}{N} \times \sum_{i=1}^{Ct} \max M_i\right) \tag{17}$$

where p is the purity of a class, Ct is the class number, M is the confusion matrix and N is the total data number.

Table 3 shows the classification quality while using:

a) The best discriminating feature;

b) The best features subset generated;

c) All features.

The implementation platform was implemented in *Matlab* 7.9, which is a general mathematical development tool. The Bioinformatics Toolbox functions *svmclassify* and *svmtrain* were used as the SVM classifier. The empirical evaluation was performed using an Intel Pentium Dual Core T4400 2.2 GHz with 3 GB RAM.

Using the parameters presented in the Table 1, the following results were obtained by taking the best solution after 20 BACOs trials.

The Table 3 gives the best solutions obtained for each dataset (RCK1 & RCK2). For the two datasets, the FV of the best solution is indicated with the corresponding error Rate. We conducted a performance comparison between the proposed wrapper-based (ACO–SVM), the filter-based ACO and the filter-based GA.

Table 1. Detail of the RCK datasets used in experimentation.

Dataset	Class	Description	Training data (observations)	Testing data (observations)
RCK1 30 features	1	Disrupted operation	31	30
	2	Moving area	20	20
	3	Inferior product	21	21
	4	Energy loosed	29	28
RCK2 46 features	1	Normal operation	125	125
	2	Abnormal operation	125	125

Table 2. Parameters of binary ant colony.

Parameter	Value	Description
N_A	20	A number of agents
F_a	0.2	Random rate of behavior
ρ	0.3	Rate of evaporation
S_Min_FP	Min FH	Minimal threshold of pheromone
S_Max_FP	Max FH	Maximum threshold of pheromone

Table 3. Performances of classification by using the various entries.

Algorithm	Features	RCK1		RCK2	
		Error rate	F (V)	Error rate	F (V)
Hybrid wrapper/filter-based ACO-SVM	Generated subset	11 %	0.7717	10 %	0.5178
Filter-based ACO		13 %	0.4347	15 %	0.5010
Filter-based GA		11 %	0.6537	15 %	0.5010
All algorithms	One feature	75 %	0.0325	46 %	0.0061
	All features	7 %	0.7875	10 %	0.5218

Table 3 shows that we obtain an acceptable error rate with the subset generated by our algorithm. It is also noticed that the FV value reflects well the quality of classification. The Fig. 10 shows the FV value obtained by each agent during the last iteration using Rotary Cement kiln dataset.

The GA parameters used in this experimentation are: population size = 50, generations = 100, crossover probability = 0.8 and mutation probability = 0.05. For the Filter based ACO algorithm, we have used the parameters presented in the Table 1. In the two algorithms, the Fisher discriminant criterion is used as a filter.

The best pair of (C, γ) using Hybrid Wrapper/Filter based ACO-SVM algorithm for the two datasets is (2^3, 2^{-5}).

We notice that after the last iteration, more than 80% of agents find a good solution. This is due to the pheromone density which is updated at the end of each iteration.

Fig. 11 shows that we obtain the optimal solution after the 6th iteration which shows the effectiveness and the speed of

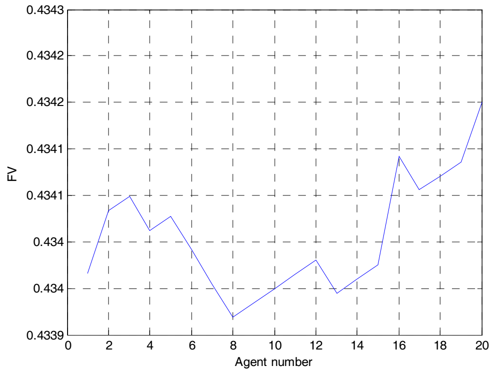


Fig. 10. The FV value obtained by each agent during the last iteration (RCK1).

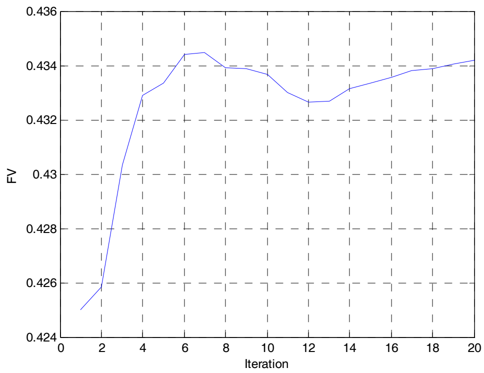


Fig. 11. The best solution obtained at the end of each iteration (RCK1).

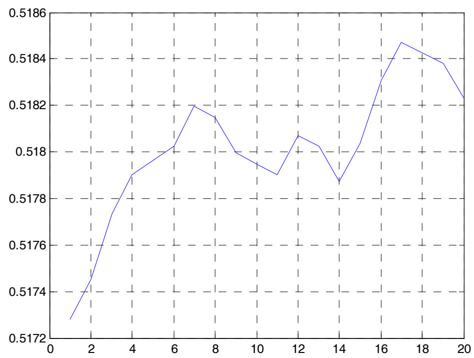


Fig. 12. The best solution obtained at the end of each iteration (RCK2).

our algorithm. The convergence time of the presented algorithm can be reduced using a lower number of ants. This number is related to the features number in the dataset.

Table 4 shows that our algorithm discards a bigger percentage of features for the RCK dataset case. However, the selected features are not always the same, once there are features that are weakly relevant and have a similar influence in the classifier.

The results given in Figs. 10-12 and Table 3 show that our approach (Wrapper/Filter-based ACO-SVM) is very precise. In other word, it gives the optimal solution compared to those obtained by other algorithms. In fact, the results obtained on

Table 4. Description of selected feature subset (RCK1).

P_1	COC	Cyclone outlet CO content A50	P_{11}	O_2C	Teneur O_2 sortie cyclone A50
P_2	A50T1	Cyclone gas outlet temperature A50	P_{12}	V31F1	Gas flow
P_3	A52T1	Cyclone gas outlet temperature A52	P_{13}	V01F1	Gas flow
P_4	A52P2	Cyclone pressure A52	P_{14}	W01X1	Oven time
P_5	A53T1	Temperature gas cyclone A53	P_{15}	W01S1	Speed oven
P_6	A53T2	Material temperature cyclone A53	P_{16}	TV	Kiln shell temperature
P_7	A53P1	Cyclone pressure A53	P_{17}	V07P1	Primary air pressure
P_8	A54T1	Temperature gas cyclone A54	P_{18}	U01T1	Clinker temperature
P_9	A54P2	Cyclone pressure A54t.	P_{19}	A54T2	Material temperature cyclone A54
P_{10}	COP	CO content smokebox	P_{20}	K01T1	Secondary air temperature

the Rotary Cement kiln datasets show that our approach converges to the global optimum in all of runs.

7. Conclusion

In this work, a new approach for selecting best discriminates features subset using Binary ACO algorithm is presented. The ACO is chosen for this study because it is the newest metaheuristic. The goal is to select the best subset that is sufficient to perform a good classification and obtain acceptable error rate. We have tested the proposed method on two datasets. The experimental results indicate that the proposed Binary ACO algorithm can be applied for a large number of features.

The classifier induced in the experiments is a SVM. This classifier is chosen because it does not suffer from the local minima problem, it has fewer learning parameters to select, and it produces stable and reproducible results, but our wrapper method can be used with any other supervised classifiers.

In the near future, the performance of the proposed algorithm will be compared with other features selection methods to improve that our algorithm achieving equal or better performance. Moreover, we will combine our algorithm with other intelligent classifiers, such as neural networks classifiers.

Acknowledgment

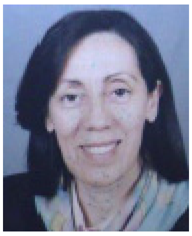
The research has been generously supported by Batna University. The authors would like to express their sincere appreciation for all support provided.

References

[1] R. Mahdaoui and L. H. Mouss, NEFDIAG, A new approach for industrial diagnosis by neuro-fuzzy systems: Application to manufacturing system, *Proc. of ERIMA*, 2 (2) (2008) 144-

151.
[2] Z. Zhang, W. Cheng and X. Zhou, Research on intelligent diagnosis of mechanical fault based on ant colony algorithm, *The Sixth International Symposium on Neural Networks, Springer* Berlin, Heidelberg, 56 (2009) 631-640.
[3] Q. Wu, Fault diagnosis model based on Gaussian support vector classifier machine, *Expert Systems with Applications*, 37 (9) (2010) 6251-6256.
[4] N. Wyse, R. Dubes and A. K. Jain, A critical evaluation of intrinsic dimensionality algorithms, *Proc. of Pattern Recognition in Practice*, Morgan Kaufmann Publishers, Inc, (1980) 415-425.
[5] M. Dash and H. Liu, Feature selection methods for classifications, *Intelligent Data Analysis*, 3 (1) (1997) 131-156.
[6] A. L. Blum and P. Langley, Selection of relevant features and examples in machine learning artificial, *Artificial Intelligence*, 97 (1997) 245- 271.
[7] P. M. Narendra and K. Fukunaga, A branh and bound algorithm for feature subset selection, *IEEE Transactions on Computers*, 26 (1977) 917-922.
[8] J. H. Yang and V. Honavar, Feature subset selection using a genetic algorithm, *IEEE Intelligent Systems*, 13 (2) (1998) 44-49.
[9] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn and A. K. Jain, Dimensionality reduction using genetic algorithms, *IEEE Transactions on Evolutionary Computation*, 4 (2) (2000) 164-171.
[10] J. Kittler, Feature set search algorithms, *Proc. of Pattern Recognition and Signal,* C. H. Chen, editor, Sijhoff and Noordhoff, the Netherlands (1978).
[11] O. Boz, Feature subset selection by using sorted feature relevance, *International Conference on Machine Learning and Application*, Las Vegas City (2002) 147-153.
[12] I. Guyon and A. Elisseeff, An introduction to variable and feature selection, *Journal of Machine Learning Research*, 3 (2003) 1157-1182.
[13] A. Al-Ani, Feature subset selection using ant colony optimization, *International Journal of Computational Intelligence*, 2 (1) (2005) 53-58.
[14] S. M. Vieira, J. C. Sousa and T. A. Runkler, Two cooperative ant colonies for feature selection using fuzzy models, *Expert Systems with Applications*, 37 (4) (2010) 2714-2723.
[15] C. L. Huang, ACO-based hybrid classification system with feature subset selection and model parameters optimization, *Neurocomputing*, 73 (2009) 438-448.
[16] V. Maniezzo, M. Dorigo and A. Colorni, The ant system: Optimization by a colony of cooperating agents, *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 26 (1) (1996) 29-41.
[17] M. Dorigo and G. D. Caro, The ant colony optimization meta-heuristic, In *New Ideas in Optimization*, McGraw-Hill, London, UK (1999) 11-32.
[18] W. N. Vapnik, An overview of statistical learning theory, *IEEE Transactions of Neural Networks*, 10 (1999) 988-999.
[19] M. F. Pasha and R. Budtarto, Evolvable-NEURAL- based fuzzy inference system and its application for adaptive network anomaly detection, *Advances in machine learning and cybernetics*, 3930 (2006) 662-671.
[20] X. Weiqing, Y. Chenyang and W. Liuyi, Binary ant colony evolutionary algorithm, *International Journal of Information Technology*, 12 (3) (2006) 10-20.
[21] E. Youn, Support vector-based feature selection using Fisher's linear discriminant and Support Vector Machine, *Expert Systems with Applications*, 37 (9) (2010) 6148-6156.
[22] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern classification*, Second Ed. Wiley Interscience Publication (2001).
[23] S. O. Ogbeide, Developing an optimization model for CO2 reduction in cement production process*, Journal of Engineering Science and Technology Review,* 3 (1) (2010) 85-88.

Ouahab Kadri received his magister degree from the Department of Computer Science, University of Batna, Algeria, in 2004. He is currently an assistant professor at the University of Khenchela, Algeria. He is currently a Doctoral student in the Department of Industrial Engineering, University of Batna, Algeria. His current research interests include evolutionary computation, artificial intelligence, etc.



Leila Hayet Mouss was born in Batna, Algeria, in 1954. She received the B.Sc. degree in Electrical Engineering, in 1979, from the National Polytechnic School of Algiers, Algeria; the M.Sc. degree in Electrical and Computer Engineering, in 1982, from the ENSERB, France; and finally the Ph.D. degree also in Electrical and Computer Engineering, in 1985, Bordeaux University, France. After graduation, she joined the University of Batna, Algeria, where she is an Associate Professor of Electrical and Computer Engineering. Pr. Mouss is a member of New York Science Academy. She is the head of Automatic and Computer Integrated Manufacturing Laboratory. Pr. Mouss current research interests include industrial Diagnosis of production system using the artificial intelligence techniques in the LAP Lab (Laboratoire d'Automatique et Productique) at Batna, Algeria.