A Hybrid Feature Subset Selection Approach based on SVM and Binary ACO. Application to Industrial Diagnosis

O. Kadri, M. D. Mouss, L.H. Mouss, F. Merah

Abstract—This paper proposes a novel hybrid algorithm for feature selection based on a binary ant colony and SVM. The final subset selection is attained through the elimination of the features that produce noise or, are strictly correlated with other already selected features. Our algorithm can improve classification accuracy with a small and appropriate feature subset. 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 a real Rotary Cement kiln dataset. The results show that our algorithm outperforms existing algorithms.

Keywords—Binary Ant Colony algorithm, Support Vector Machine, feature selection, classification.

I. INTRODUCTION

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). In substance, condition monitoring is a classification problem [12]. The principal function of the condition monitoring is to check the operating condition of the system. It 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 [7, 10, 12].

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 [7,10]. 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).

O. Kadri is with Laboratory Automation & Production Engineering, University of Batna, 1, Rue Chahid Boukhlouf 05000 Batna, Algeria (phone: 21333803396; fax: 21333803396; e-mail: ouahabk@yahoo.fr).

M. D. Mouss is with Laboratory Automation & Production Engineering, University of Batna, 1, Rue Chahid Boukhlouf 05000 Batna, Algeria (e-mail: Djamelm@yahoo.fr)

L.H. Mouss is with Laboratory Automation & Production Engineering, University of Batna, 1, Rue Chahid Boukhlouf 05000 Batna, Algeria (e-mail: hayet_mouss@yahoo.fr).

F. Merah is with Department of Mathematics, University of Khenchela, Route de Batna, BP: 1252, El Houria, 40004 Khenchela, Algeria (phone: 21332331966; fax: 21332331963;e-mail: merahfateh@yahoo.fr).
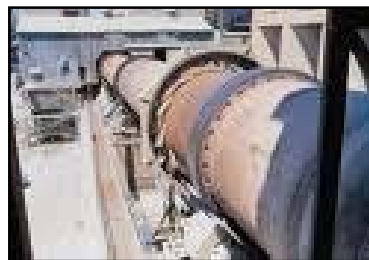
Fig. 1 Rotary Cement kiln

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 [15]. Subset selection is different from feature extraction in that no new features will be generated, but only a subset of original features is selected and feature space is reduced [14].
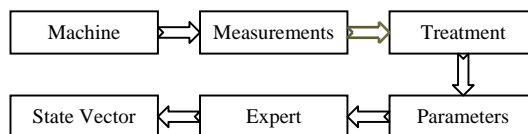

Fig. 2 Construction of the set of features

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:
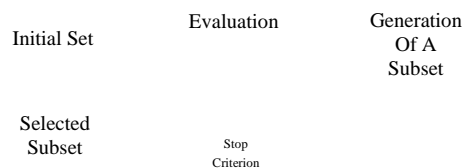

Fig. 3 Subset Selection Method

A starting subset, which represents the subset of features, initially is used by a procedure of research. This set can be empty, or contains all the features or a random subset. The procedure of research is the essential element of any method of selection. It turns over as result the subset of features which answer the quality standard better. This criterion is turned over by a function of evaluation. This function determines the

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:4, No:12, 2010

quality of classification obtained by using a subset of feature. A criterion of stop is used to finish the procedure of research. This criterion depends to the evaluation function or with the parameters of configuration which are defined by the user [1]. 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.

II. FEATURE SUBSET SELECTION

Feature selection is included in discrete optimization problems. The whole search space for optimization contains all possible subsets of features, meaning that its size is 2^n where n is the dimensionality (the number of features). Usually FS algorithms involve heuristic or random search strategies in an attempt to avoid this prohibitive complexity. However, the degree of optimality of the final feature subset is often reduced [2, 4, 8].

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 [3, 22]. 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 usefulness of features [13]. The two famous algorithms of this category are Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) [1, 3]. 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 [22]. The criterion of stop represents the dimension of the vector obtained by the algorithm where the quality standard does not evolve/move if we add another feature [9].
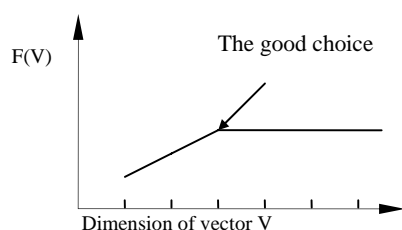

Fig. 4 The choice of the dimension of vector V

The first good use of ACO for feature selection seems to be reported in [9]. A. Al-Ani [9] proposes to use a hybrid evaluation measure that is able to estimate the overall performance of subsets as well as the local importance of features. A classification algorithm is used to estimate the performance of subsets. On the other hand, the local importance of a given feature is measured using the Mutual Information Evaluation Function. Susana [19] proposes an algorithm for feature selection based on two cooperative ant colonies, which minimizes two objectives: the number of features and the classification error. The first colony determines the number (cardinality) of features and the second selects the features based on the cardinality given by the first colony. C.L. Huang [20] 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 parameters of classifier are fixed during the execution of program and they may have different value in each solution.

III. 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 [5, 16]:

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

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

3. 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.

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

5. 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.

6. A neighborhood 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 $s_2 \equiv \langle s_1, c_2 \rangle$.

7. 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 elements of C.

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:4, No:12, 2010

8. 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 [5, 9, 16].

IV. SUPPORT VECTOR MACHINES

In our wrapper approach, we have used SVM as classifier. SVM is an attractive learning algorithm first introduced by Vapnik [23]. It has a competitive advantage Compared to neural networks, and decision trees.

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

$$w\varphi(x) + b = 0 \; with \; y_i(w\varphi(x_i) + b) \geq 1 - \xi_i \qquad (1)$$

that separating the S set of training data 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.
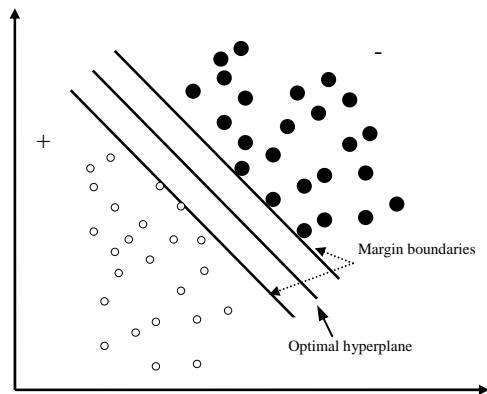

Fig. 5 Two-Class SVM used in linear classification

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

$$K(x_i, x_j) = \varphi(x_i).\varphi(x_j) \qquad (2)$$

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

$$\sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \qquad (3)$$

subject to

$$\sum_{i=1}^{m} a_i y_i = 0 \; and \; 0 \leq a_i \leq C \qquad (4)$$

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

$$sign(\sum_{i=1}^{m} \alpha_i y_i K(z_i, z) + b) \qquad (5)$$

where

- w is the weight vector.

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

In the next section, we present our proposed SVM/Binary ACO algorithm [6], and explain how it is used for selecting an appropriate subset of features.

V. PROPOSED APPROACH

A. Description of the proposed approach

This research proposes a new implementation of Binary ACO algorithm [6] applied to feature selection, where the best number of features 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 as shown in Fig 6. The intelligent behavior of ant is very simple, and the incidence matrix traversed by each ant needs only $2 \times n$'s space, which to some extent solves the descriptive difficulty generated from long coding and the reduction of solution quality.
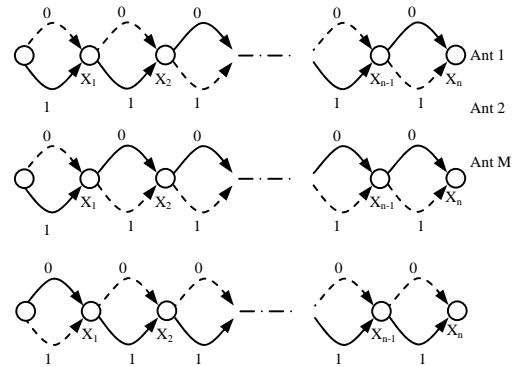

Fig. 6 The final net obtained by the binary ACO algorithm

B. Probabilistic rule

Initially, the quantity of information in each routine is randomly generated. During the movement, ant k shifts its direction according to the values of pheromone concentration FP and the heuristic value FH. The heuristic value FH is computed using the Fisher discriminant criterion for feature selection [11] [17], 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 Xi is given by:

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

C. Updating rule

After all ants have completed their solutions, pheromone evaporation on all nodes is triggered, and then according to (2), pheromone concentration in the trails is updated.

$$FP \leftarrow (1 - \rho)FP + \Delta FP \qquad (7)$$

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:

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:4, No:12, 2010

$$\Delta FP = \frac{1}{1 + F(V) - F(V^{'})} \qquad (8)$$

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.

$$F'(V) = trace\left(\sum{}_{W}^{-1} \cdot \sum{}_{B}\right) \qquad (9)$$

Where the matrix of variance intra-class is calculated as follows:

$$\sum{}_{W} = \frac{1}{N} \sum_{C=1}^{M} \sum_{V=1}^{N_C} (X_{CV} - m_C)(X_{CV} - m_C)^t \qquad (10)$$

Whereas the matrix of the variance inter-classes is calculated as follows:

$$\sum{}_{B} = \frac{1}{N} \sum_{C=1}^{M} (m_C - m)(m_C - m)^t \qquad (11)$$

With:
- m: General centre of gravity
- M: A number of classes
- m_C: Centre of gravity of the class number C
- X_{cv}: Vth vector of the class number C
- NR_C: A number of vectors of the class number C
- NR: Numbers total vectoriels

D. Heuristics

The heuristic value is computed using the Fisher discriminant criterion for feature selection [17]. Considering a classification problem with M possible classes, the Fisher discriminant criterion is described as follow:

$$FH(\alpha) = \sum_{c=1}^{M} \sum_{\substack{r=1 \\ r \neq c}}^{M-1} \frac{m_c(\alpha) - m_r(\alpha)}{N_c \sigma_c^2(\alpha) - N_r \sigma_r^2(\alpha)} \qquad (12)$$

Where:

M represents the number of classes;

$m_c(\alpha)$ Represent the centre of gravity of the class number C by considering only the parameter α it is calculated as follows:

$$m_c(\alpha) = \frac{1}{N_c} \sum_{v=1}^{N_c} X_{cv}(\alpha) \qquad (13)$$

With X_{cv} is the number v of the class number C. the value of NR equal to the number of vectors 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.

$$\sigma_r^2(\alpha) = \frac{1}{N_c} \sum_{v=1}^{N_c} [X_{cv}(\alpha) - m_c(\alpha)]^2 \qquad (14)$$
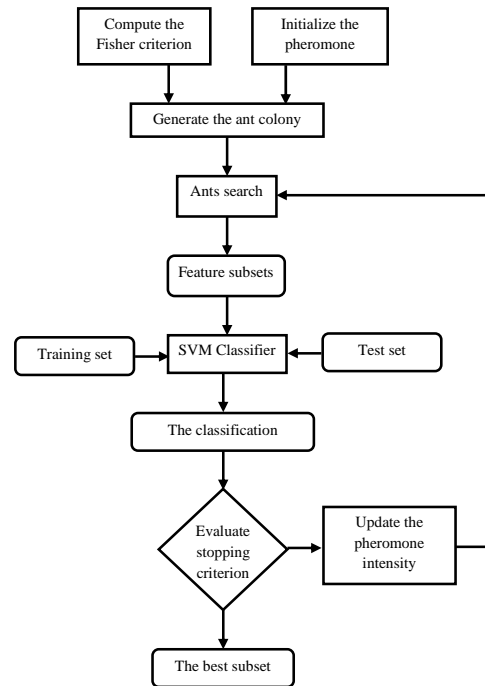

Fig. 7 SVM-Binary ACO feature selection algorithm

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

ALGORITHM I
BINARY ACO/SVM FEATURE SELECTION ALGORITHM

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

The time complexity of proposed algorithm is $O(Im)$, where I is the number of iterations, and m the number of ants. 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.

VI. EXPERIMENTAL RESULTS

A. Data of test

The experimental results comparing the binary ACO algorithm with genetic algorithm are provided for real-life dataset (Vehicle) [18] and industrial dataset (Rotary Cement kiln) [7]. The Vehicle base is collected in 1987 by Siebert. Vehicle consists of 846 recordings which represent 4 classes.

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:4, No:12, 2010

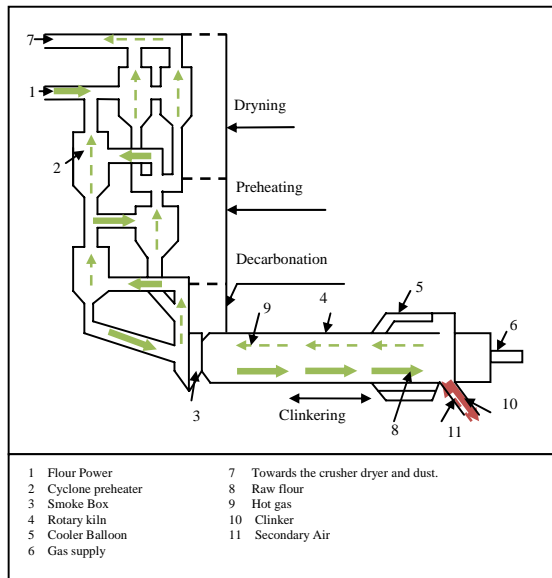RCK consists of 200 recordings which represent 4 classes.

Fig. 8 Flow Diagram of a Rotary Cement Kiln [7]

B. Parameters of the algorithm of selection

Like any other algorithm, before passing to the phase of selection. Some parameters should be fixed. This problem represents one of the disadvantages of the biomimetic methods. Since the values of the parameters are related to the number of individuals and the distribution of the data on the beach of representation. The following table presents the values of the parameters of our algorithm. These parameters are fixed after the execution of several simulations by using as entered a restricted whole of data.

TABLE I
PARAMETERS OF BINARY ANT COLONY

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

C. 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 spaces of research. The following figures represent the values of heuristic factor FH by using the dataset of the Rotary cement kiln and Vehicle dataset.

According to fig. 9, we notice that the 43rd feature has the greatest value of FH. Consequently it will be present in the final subset.
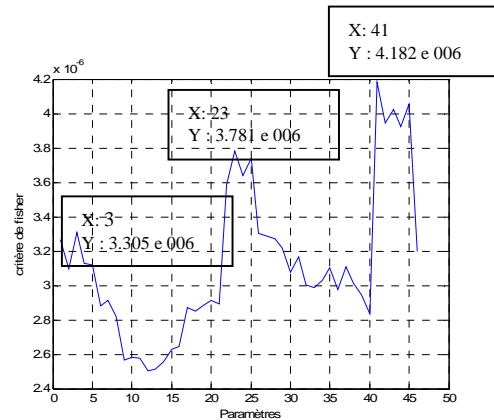

Fig. 9 The Fisher discriminant criterion for the dataset of the Rotary cement kiln
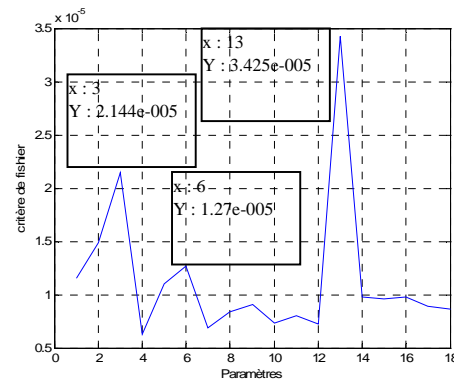

Fig. 10 The Fisher discriminant criterion for Vehicle dataset

D. Results

We tested the performances of our algorithm by using the evolutionary method of classification ECMC [21] and the following table shows the quality of classification while using:

a) The best discriminating feature;
b) The best subset of features 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 2.2GHz with 3GB 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.1 gives the best solutions obtained for each dataset (Rotary Cement kiln & Vehicle). For the two datasets, the FV of the best solution is indicated with the corresponding Rate of error. We conducted a performance comparison between the proposed wrapper-based ACO (ACO–SVM), the filter-based ACO and the filter-based GA.

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:4, No:12, 2010

TABLE II
PERFORMANCES OF CLASSIFICATION BY USING THE VARIOUS ENTRIES

Algorithm	Features	Vehicle		Rotary Cement kiln	
		Error rate	F (V)	Error rate	F (V)
Wrapper-based ACO/SVM	One feature	75 %	0.0325	46 %	0,0194
	Generated subset	11 %	0.4717	10 %	0.5078
	All features	07 %	0.7875	10 %	0,5082
Filter-based ACO	One feature	75 %	0.0325	46 %	0.0061
	Generated subset	13 %	0.6537	15 %	0.0210
	All features	07 %	0.7875	10 %	0.0218
Filter-based GA	One feature	75 %	0.0325	46 %	0.0061
	Generated subset	11 %	0.7717	15 %	0.0210
	All features	07 %	0.7875	10 %	0.0218

Table 2 shows that we obtain an acceptable rate of error with the subset generated by our algorithm. It is also noticed that the value of FV reflects well the quality of classification. The Fig. 9 shows the value of FV obtained by each agent during the last iteration using Rotary Cement kiln dataset.
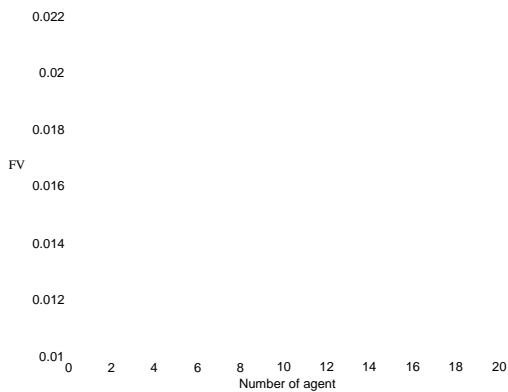

Fig. 11 The value of FV obtained by each agent during the last iteration

We notice that after the last iteration, more than 33% of agents find the optimal solution. This is due to the pheromone density which is updated at the end of each iteration.
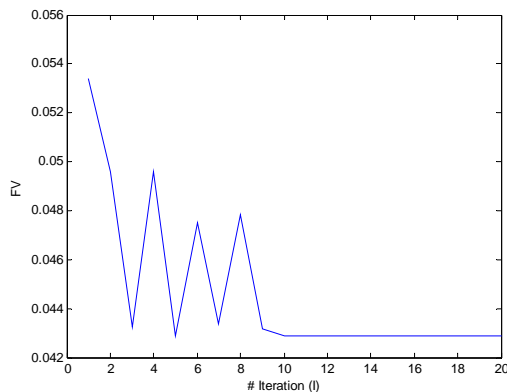

Fig. 12 The best solution obtained during each iteration

Fig. 12 shows that we obtain the optimal solution after the

5nd iteration which shows the effectiveness and the speed of our algorithm. The time of convergence of the presented algorithm can be reduced using a lower number of ants. This number is related to the number of features in the dataset.
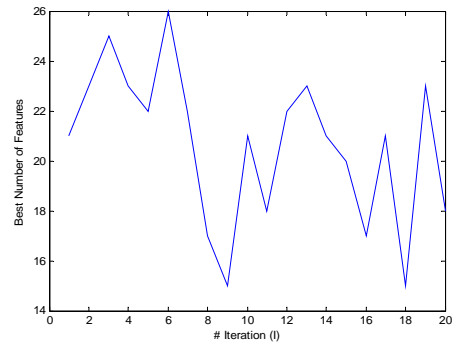

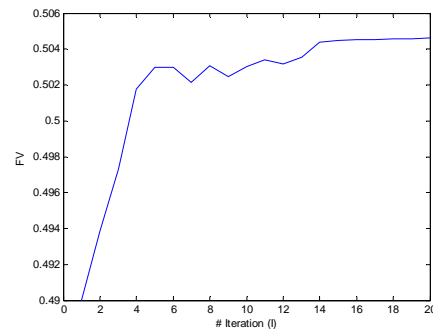Fig. 13 Best number of features for each iteration in Rotary Cement kiln dataset

Fig. 14 The best solution obtained during each iteration (Vehicle)

Fig. 14 shows that our algorithm discards a bigger percentage of features for the case of Vehicle dataset. 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 figures 11-14 and Tables 2 show that our approach (Wrapper-based ACO-SVM) is very precise. In other word, it gives the optimal solution in compare to those of obtained by the other algorithms. In fact, the results obtained on the Rotary Cement kiln dataset show that our approach converges to the global optimum in all of runs.

VII. 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 rate of error. We have tested the proposed method on two datasets. The experimental results indicate that the proposed Binary ACO algorithm can be applied for larger number of features.

The classifier induced in the experiments was a SVM. This classifier was chosen because it does not suffer from the local minima problem, it has fewer learning parameters to select,

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:4, No:12, 2010

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. And we will combine our algorithm with other intelligent classifiers, such as neural networks classifiers.

REFERENCES

[1] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning artificial," *Artificial Intelligence*, vol. 97, pp. 245–271, 1997.
[2] P.M. Narendra and K. Fukunaga, "A branh and bound algorithm for feature subset selection." *IEEE Transactions on Computers*, vol 26, pp. 917–922, 1977.
[3] J. Kittler, "Feature set search algorithms," In *C. H. Chen*, editor, Pattern Recognition and Signal Processing. Sijhoff and Noordhoff, the Netherlands, 1978.
[4] J. H. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intelligent Systems*, Vol. 13, no. 2, pp. 44-49, 1998.
[5] 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*, vol 26, no.1, pp. 29–41, 1996.
[6] X.Weiqing, Y.Chenyang and W.Liuyi, "Binary Ant Colony Evolutionary Algorithm," *International Journal of Information Technology*, Vol. 12, no. 3, 2006.
[7] R.Mahdaoui, L.H. Mouss, "NEFDIAG. a New Approach for Industrial diagnosis by Neuro-Fuzzy systems: Application to manufacturing System," *ERIMA*, vol.2, no. 2, pp. 144–151, 2008.
[8] 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*, vol 4, no. 2, pp 164–171, 2000.
[9] A. Al-Ani, "Feature Subset Selection Using Ant Colony Optimization," *International Journal of Computational Intelligence*, vol 2, no. 1, pp 53–58, 2005.
[10] 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, Vol 56, pp 631–640, 2009.
[11] E. Youn, "Support vector-based feature selection using Fisher's linear discriminant and Support Vector Machine," *Expert Systems with Applications*, vol, 37, no 9, pp. 6148–6156, 2010.
[12] Q. Wu, Fault diagnosis model based on Gaussian support vector classifier machine. Expert Systems with Applications, vol, 37, no 9, pp.6251-6256, 2010.
[13] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol 3, pp 1157–1182, 2003.
[14] M. Dash and H. Liu, "Feature selection methods for classifications," *Intelligent Data Analysis*, vol 3, no.1, pp. 131–156, 1997.
[15] N. Wyse, R. Dubes, and A.K. Jain, "A critical evaluation of intrinsic dimensionality algorithms," In *E.S. Gelsema and Kanal L.N*, editors, *Pattern Recognition in Practice*, Morgan Kaufmann Publishers, Inc, pp 415–425. 1980.
[16] M. Dorigo and G. Di Caro, "The ant colony optimization meta-heuristic," In *New Ideas in Optimization*, McGraw-Hill, London, UK, pp 11–32, 1999.
[17] R. O. Duda, P. E.Hart and D. G. Stork, in *Pattern classification*, 2nd ed. Wiley Interscience Publication, 2001.
[18] A. Asuncion, and D. J. Newman, *UCI machine learning repository*, 2007.
[19] S. M. Vieira, J.C. Sousa and T.A. Runkler, "Two cooperative ant colonies for feature selection using fuzzy models," *Expert Systems with Applications*, vol. 37, no.4, pp. 2714–2723, 2010.
[20] C. L. Huang, "ACO-based hybrid classification system with feature subset selection and model parameters optimization," *Neurocomputing*, vol 73, pp 438–448, 2009.
[21] M. F. Pasha, R.Budtarto, "Evolvable-NEURAL- Based Fuzzy Inference System and Its Application for Adaptive Network Anomaly Detection," *Advances in machine learning and cybernetics*, vol. 3930, pp. 662–671, 2006.
[22] O. Boz, "Feature subset selection by using sorted feature relevance," *International Conference on Machine Learning and Application*, Las Vegas City, , P147-153, 2002.
[23] W.N.Vapnik, "An overview of statistical learning theory," *IEEE Transactions of Neural Networks*, vol 10, pp. 988–999, 1999.