

Hybrid Pattern Classification Algorithm based on ACO-SVM

HIBA BASIM ALWAN, KU RUHANA KU-MAHAMUD

School of Computing, College of Art and Sciences

University Utara Malaysia

06010 Sintok, Kedah

MALAYSIA

hiba81basim@yahoo.com

ruhana@uum.edu.my

Abstract: - The performance of Support Vector Machine (SVM) classifier can be improved by simultaneously optimizing its parameters and features subset selection. The problem of selecting suitable features subset and optimizing SVM parameters must occur simultaneously, because these problems are essential and they affect each other which in turn will affect the SVM classification accuracy. Ant Colony Optimization (ACO) originally deals with discrete optimization problem. In applying ACO for optimizing SVM parameters which are continuous variables, there is a need to discretize the continuously value into discrete value. This discretize process would result in loss of some information and hence affect the classification accuracy and seeking time. This study proposed two algorithms, the first algorithm deals with optimizing only SVM parameter using Incremental Continuous Ant Colony Optimization without the need to discretize continuous value for support vector machine parameters, while the second algorithm can simultaneously optimize SVM parameters and features subset selection using Incremental Mixed Variable Ant Colony Optimization. Five benchmark datasets were used to evaluate the proposed algorithms. Results showed that the proposed algorithms can enhance the classification accuracy with small size of features subset.

Key-Words: - Support Vector Machine, Ant Colony Optimization, Feature Selection, and Pattern Classification.

1 Introduction

SVM is an excellent classifier built on statistical learning approach [1]. The main concept of SVM is to obtain the Optimal Separating Hyperplane (OSH) between positive and negative samples. This can be done through maximizing the margin between two parallel hyperplanes. Once the plane is found, SVM can forecast the classification of unlabeled sample through identifying the side where the sample lies [2]. Tuning SVM parameters and selecting optimal feature subset for SVM classifier are two problems that influence the classification accuracy. These problems affect each other because error produced from the feature subset selection phase will affect the values of the SVM parameters [3].

There is no regular methodology that accepts advance approximation of optimal values for SVM parameter. Currently, almost all SVM researches choose these variables experimentally via searching a bounded number of values and preserving those that supplies the lowest amount of errors. This approach requires a grid search through the area of variable values and requires identifying the range of executable solution and best sampling step. This is a tricky task because best sampling step changes from kernel to kernel and grid range may not be simple to identify without advance knowledge of the problem.

Furthermore, when hyper parameter exceeds two of the manual prototype chosen, it may become intractable [4]. Approaches such as trial and error, grid search, cross validation, generalization error estimation and gradient descent can be used to find optimal parameter values for SVM. Evolutionary approaches such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) may also be utilized.

Feature Selection (FS) is a process of determining a subset of fields in database and it minimizes the number of fields that appears during data classification [3], [5], and [6]. The main idea behind FS is to select a subset of input variables by deleting features that contain less or no information [7]. FS aims to decrease the dimension of the initial features group by determining the unauthentic features which would eventually supply the best performance under certain classification dataset [8] and [9], and to delete unrelated, unneeded, or noisy features while preserving the richness of the instructive ones [10] and [11]. FS may be considered as an optimization problem which looks out for potential feature subsets which ultimately determines the optimal one [10].

ACO algorithms have been applied to tune SVM parameters. These algorithms work through

repetitive creation procedures where each procedure directs a dependent heuristic by intelligently mixed various ideas for exploring and exploiting the seek space. The learning fashions are utilized to construct information to efficiently obtain near optimal solutions. ACO algorithms deal with discrete and continuous variables. ACO that deals with continuous variables is considered as a recent research field [13].

Continuous Ant Colony Optimization (ACO_R) uses probability density function instead of discrete probability distribution to determine the direction that an ant should follow. The main disadvantages of ACO_R are the high running time and the application gap with the-state-of-art continuous solvers. To overcome these disadvantages, the IACO_R Algorithm is introduced. The enhancements are as follow: the size of solution archive in IACO_R begins with an initial size and increases over time until it reaches a maximum size. The variable *Growth* monitors the percentage at which the archive grows. This algorithm differs from the original ACO_R in the case of choosing the solution that directs the creation of new solutions. The new mechanism counts on a variable $\rho \in [0, 1]$, which monitors the probability of utilizing just the best solution in the archive as a directing solution. In order to avoid stagnation, IACO_R involves an algorithm-level diversification mechanism. The mechanism includes in restarting the algorithm and initializing the new initial archive with the best-so-far solution. The restart condition is the number of succeeding iterations, *MaxStagIter*, with a relate solution enhance lower than a certain threshold [14].

IACO_{MV} algorithm that is adopted in this research is based on [14] and it has been used to simultaneously optimize SVM parameters and feature subset selection. Incremental mixed variable Ant Colony Optimization (IACO_{MV}) contains two main parts. The first part is related with optimizing SVM continuous parameter as explained in IACO_R and the second part deals with optimizing the discrete variable and related with selecting suitable feature subset.

The rest of the paper is organized as follows. Section 2 reviews previous studies on tuning SVM parameters as well as to simultaneously optimize SVM parameters and features subset. Section 3 describes the proposed hybrid algorithms. Experimental results are discussed in Section 4 while concluding remarks and future work is presented in Section 5.

2 Related Works with SVM

Bare Bones Differential Evolution (BBDE) to optimize SVM parameters was proposed by [15]. It deletes the monitoring variables of PSO and substitutes the static differential evolution monitor variables with dynamically alternating variables to generate a general parameter-free, self-adaptive, optimization algorithm. Four UCI datasets were used to test the performance of the proposed method and results were compared with grid algorithm. Results showed that BBDE-SVM and DE-SVM take shorter time to compute and produced higher classification accuracy. This implies that the parameters chosen by BBDE and DE are better than grid search. Three factors (accuracy, sensitivity and specificity) were utilized to generate fitness function to be maximised [16]. In their research, they proposed meta-classification prototype (SVMDE_{Meta}) and four types of kernel function were used. The kernel functions are linear, polynomial, radial basis function, and sigmoid. SVMDE_{Meta} was evaluated on three real-life datasets. The datasets are TargetMiner, and two UCI datasets. A method to optimize hyper-parameters for SVM was proposed by [17]. This method uses PSO and grid search in a dynamic environment via alternating between three stages: A framework which consists of three basic modules was developed. The first module controls the quality of the model selection tasks and the second module adapted grid search to provide optimum solutions. The third module utilized Dynamic Particle Swarm Optimization (DPSO) to obtain new solutions through re-optimization procedures. The authors examined their approach on fourteen datasets and compared their work with classical Grid Search (GS), 1st Grid Search (1st GS), Chained PSO (CPSO), and Dynamic Model Selection (DMS). Another recent study conducted by [18] try to find the optimal value for the kernel (σ) and SVM parameter (C) to enhance its performance on classified Chinese text provided by Sogou Labs. The author in their study utilized exponent-weighted ACO to optimize SVM parameters and the reason for utilizing exponent weight was to avoid the local optimization and slow convergence speed of ACO. A Kernel Condition Random Field (KCRFs) was utilized as a pre-processing step to segment words, extract features, and representation, then a combination of Principle Component Analysis (PCA) and noise elimination threshold was utilized to select feature subset. One-to-One strategy was utilized in this study to classify multi-class text and the result was compared with standard SVM, linear kernel function SVM (LKF-SVM), AdaBoost, and Naive Bayes. Results

obtained showed that the proposed method had optimal execution in terms of classification accuracy and efficiency. The authors suggested using mutation procedure in ACO to avoid the local convergence problem as well as to use other swarm optimization techniques.

There are four similar studies [3], [6], [12], and [19] that suggested using hybrid systems to enhance classification accuracy by using few and suitable feature subsets. In these studies, feature subset and SVM parameters (C and γ RBF kernel) variable are simultaneously optimized. SVM is then used to measure the quality of the solution for all the hybrid systems. However, the quality of the solution is dependent on the hybrid system being used. Studies by [3] and [12] proposed the use of a hybrid system which is based on GA and SVM. GA is used to select suitable feature simultaneously with optimized SVM parameter which were represented in the encoded chromosomes. The hybrid system which is used by [6] is based on ACO and SVM. Classical ACO was used to simultaneously select suitable feature and optimize SVM parameter. Bees algorithm is proposed in [19] to simultaneously choose the best combination of feature subset and SVM parameters values for the process of classifying faults in wood layer pieces. The general conclusions derived from these studies are acceptable and reliable with regards to classification accuracy and a few numbers of selected features. Other suggestions from related studies are as follows: Support Vector Regression (SVR) is suggested in the studies by [12] and 16, because SVR accuracy counts mainly on SVR parameters and selecting of feature subset. They also suggested on using other types of kernel function. [3] Suggested applying their work on other real world problem and finally [6] suggested using continuous ACO to optimize the continuous value of SVM parameters.

3 Proposed Hybrid Algorithms

The first proposed algorithm utilizes IACO_R to optimize only SVM classifier parameters. An ant's solution is used to represent a combination of the classifier parameters, C and γ , based on the Radial Basis Function (RBF) kernel of the SVM classifier. The classification accuracy of the built SVM classifier is utilized to direct the updating of solution archives. Based on the solution archive, the transition probability is computed to choose a solution path for an ant. In implementing the proposed scheme, this study utilizes the RBF kernel function for SVM classifier because of its capability

to manage high dimensional data, good performance in major cases, and it only needs to use one parameter, which is kernel parameter gamma (γ) [20-22]. The main steps are (1) selecting feature subset (2) initializing solution archive and algorithm parameters, (3) solution construction for C and γ , and (4) establishing SVM classifier model.

In features subset selection, F-score is used as a measurement to determine feature importance. This measurement is used to judge the favouritism capability of a feature. High value of F-score indicates favourable feature. The calculation of F-score is as follow [20-22]:

$$F - Score_i = \frac{\sum_{c=1}^v (\bar{x}_i^{(c)} - \bar{x}_i)^2}{\sum_{c=1}^v \left(\frac{1}{N_i^{(c)}} \sum_{j=1}^{N_i^{(c)}} (x_{ij}^{(c)} - \bar{x}_i^{(c)})^2 \right)} \quad (1)$$

where $i = 1, 2, \dots, N_f$, v is the number of categories of target variable, N_f is the number of features, $N_i^{(c)}$ is the number of samples of the i^{th} feature with categorical value c , $c \in \{1, 2, \dots, v\}$, $\bar{x}_{i,j}^{(c)}$ is the j^{th} training sample for the i^{th} feature with categorical value c , $j \in \{1, 2, \dots, N_i^{(c)}\}$, \bar{x}_i is the i^{th} feature, and $\bar{x}_i^{(c)}$ is the i^{th} feature with categorical value c .

After computing F-score for each feature in the dataset, average F-score will be computed and it will be considered as threshold for choosing a feature in the feature subset. Feature with F-score equal to or greater than the threshold will be chosen and put in the feature subset. This subset will be presented to SVM.

In the initialization step, each ant established a solution path for parameter C and parameter γ . Two solution archives are needed to design the transition probabilities for C and γ . The range for C and γ values will be sampled according to random parameter k which is the initial archive size of solutions archives. The weight vector, w is then computed for each sample for C and γ as follow:

$$w_i = \frac{1}{qk\sqrt{2\pi}} e^{-\frac{(1-i)^2}{2q^2k^2}} \quad (2)$$

where q is the algorithm's parameter to control diversification of search process. These values will be stored in solution archives. Once this step is completed, the sampling procedure will be constructed in two phases. Phase one involves choosing one of the weight vectors as following:

$$p_i = \frac{w_i}{\sum_{r=1}^k w_r} \quad (3)$$

The second phase involves sampling selecting w via a random number generator according to a parameterized normal distribution.

This initialization will construct the transition probabilities. Like the solution archives, some important system parameters must be initialized as follows: the number of ants = 2, $q = 0.1$, *initial*

archive size = 10, Growth = 5, maximum archive size = 15, MaxStagIter = 2, number of runs = 10, C range $\in [2^{-1}, 2^{12}]$ and $\gamma \in [2^{-12}, 2^2]$.

The third step is related to solution construction where each ant builds its own solution. This solution will be a combination of C and γ . In order to construct the solution, two transition probabilities with various solutions archives are needed. These transitions will be computed according to Eq. (2) and Eq. (3).

Classifier model will be constructed in step four. Solution is generated by each ant and will be evaluated based on classification accuracy obtained by SVM model utilizing k -fold CV with the training set. In k -fold CV, training data group is partitioned into k subgroups, and the holdout approach is repeated k times. One of the k sub-groups is utilized as the test set and the remaining $k-1$ subgroups are combined to construct the training group. The average errors along with all the k trails are calculated. CV accuracy is calculated as follows:

$$CV_{accuracy} = \frac{\sum_i test_accuracy}{k}, i = 1, 2, \dots, k \quad (4)$$

Test accuracy is used to evaluate the percentage of samples that are classified in the right way to determine k -folds and it will be computed as follows:

$$Test\ Accuracy = \frac{no.of\ correctly\ predicted\ data}{total\ testing\ data} * 100\% \quad (5)$$

The benefits of using CV are (1) each of the test groups is independent and (2) the dependent outcomes can be enhanced [20-22].

The second proposed algorithm has adopted $IACO_{MV}$ to optimize features subset selection and SVM classifier parameters. An ant's solution is used to represent a combination of features subset and the classifier parameters, C and γ , and not only SVM parameters as in first proposed algorithm stated above. $IACO_{MV}$ is like $IACO_R$ in part of optimizing SVM model, but is different during the selection of feature subset. $IACO_{MV}$ will not use filter technique to selecting feature subset as in $IACO_R$, but the ant will select suitable feature subset. The ant solution therefore will include three parts; the first two parts are for C and γ while the third part is for feature subset. $IACO_{MV}$ will start with initializing three solution archives and these solution archives are needed to design the transition probabilities for first feature in the features subset, C and γ , and one pheromone table. The process for the ant to select a suitable feature subset begins with computing the F-score for each feature according to Eq. (1) and then computes the weight vector w for each feature as follows:

$$w_f = \frac{w_i}{u} + \frac{q}{\eta} \quad (6)$$

where u is the number of $feature_i$ is selected, η is the number of none selected features, and q in both equations is the algorithm's parameter to control diversification of search process. These values will be stored in the solution archive for features. In order to select other features that construct the features subset, the following probability transition is used:

$$Prob_{ij}^k = \begin{cases} \frac{(Prob_{ij})^\alpha (F-Score_j)^\beta}{\sum_{j \in I_i^k} (Prob_{ij})^\alpha (F-Score_j)^\beta} & \text{if } j \in I_i^k \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Similarly $IACO_R$ parameters are used in $IACO_{MV}$ with the same values. Two additional parameters that appear in $IACO_{MV}$ are α which is set to 1 and β which is set to 2.

In order to construct the transition probability for the ant to construct feature subset, the solution archive for feature as well as the pheromone table is needed. The transition probability for feature will be computed according to Eq. (6).

The pheromone table will be modifying as follow:

$$T_{ij}(t+1) = pT_{ij} + \sum_{k=1}^m \Delta T_{ij}^k(t) \quad (8)$$

$$\Delta T_{ij}^k = \begin{cases} CVACC^k * Weight_i^k * Weight_j^k & \text{if ant } k \text{ use edge } (i, j) \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

4 Experimental Results

Five datasets were used in evaluating the proposed $IACO_R/IACO_{MV}$ -SVM algorithms. The datasets are Australian, Pima-Indian Diabetes, Heart, German, and Ionosphere datasets, available from UCI Repository of Machine Learning Databases [23]. The summary of these datasets are presented in Table 1.

Table 1 Summarization of UCI's Datasets

Dataset	No. of Instances	No. of Features	Type of Datasets
Australian	690	14	Categorical, Integer, Real
Diabetes	760	8	Integer, Real
Heart	270	13	Categorical, Real
German	1000	24	Categorical, Integer
Ionosphere	351	34	Integer, Real

All input variables were scaled during data pre-processing phase to avoid features with higher numerical ranges from dominating those in lower numerical ranges and also to reduce the computational effort. The following formula was used to linearly scale each feature to $[0, 1]$ range.

$$\bar{x} = \frac{x - \min_i}{\max_i - \min_i} \quad (10)$$

where x is the original value, \bar{x} is the scaled value, and \max_i and \min_i are the maximum and minimum values of $feature_i$, respectively [20-22].

Each dataset is randomly re-arranged and divided into ten approximately equal size subsets, one subset as testing set and the remaining as training sets and

repeated ten times. The performance of the proposed $IACO_R/IACO_{MV}$ -SVM was compared with GA with feature chromosome [12].

C programming language was used to implement the $IACO_R/IACO_{MV}$ -SVM. Experiments were performed on Intel(R) Core (TM) 2 Duo CPU T5750, running at 2.00 GHz with 4.00 GB RAM and 32-bit operating system.

Table 2 shows the optimal values for C and γ that gave the highest classification accuracy in all ten runs that have been produced by the proposed algorithms as well as p value which is the algorithm parameter generated randomly and used to monitor the probability of utilizing just best solution in the archive as a directing solution.

Table 2 Optimal Value for C , γ , and p

Dataset	IACO-SVM			IACO _{MV} -SVM		
	C	γ	p	C	γ	p
Australian	440.86	0.75	0.371	416.29	0.75	0.539
Diabetes	450.21	0.76	0.428	440.53	0.76	0.280
Heart	448.81	0.76	0.403	439.33	0.76	0.516
German	123.2	0.120	0.878	437.68	0.74	0.506
Ionosphere	448.33	0.75	0.459	447.5	0.73	0.979

Table 3 shows the average classification accuracy that was produced in all the ten runs. The classification accuracy of classify pattern of the proposed $IACO_R/IACO_{MV}$ -SVM algorithms is compared with GA-SVM [12] results. The proposed algorithms classify patterns with higher accuracy compared to the GA-SVM [12] in all five datasets. This is because the proposed algorithms handle directly the continuous value of SVM parameters without the need to discretize it. Also, the performance of the proposed algorithm $IACO_{MV}$ -SVM was better than the proposed algorithm $IACO_R$ -SVM. This is because the proposed algorithm $IACO_{MV}$ -SVM simultaneously optimizes feature subset selection and model selection for SVM.

Table 3 Classification Accuracy%

Dataset	IACO _R -SVM	GA-SVM without Feature Selection	IACO _{MV} -SVM	GA-SVM with Feature Selection
Australian	94.6±0.5	86.8±3.6	96.96±0.5	91.6±2.1
Diabetes	91.2±0.1	81.97±5.3	97.2±0.8	83.8±5.1
Heart	96.3±0.2	91.1±2.6	98.01±0.4	95.6±2.3
German	91.6±0.1	80.8±2.1	97.2±0.5	86.1±1.97
Ionosphere	98.9±0.5	98.6±2.02	99.99±0.02	99.4±1.2

Table 4 shows the average selected features subset size that was produced in all the ten runs. The proposed algorithms produced lower average number of selected features when compared with GA-SVM [12]. However, the proposed $IACO_{MV}$ -SVM algorithm produced lower feature subset size when compared with the proposed $IACO_R$ -SVM algorithm. This is because the proposed $IACO_{MV}$ -

SVM algorithm is based on wrapped feature technique that depends on the inductive learning approach which has the ability to integrate the classification accuracy and necessary features introduced to the SVM classifier from $IACO_{MV}$ algorithm. The biggest reduction in number of features for $IACO_{MV}$ -SVM was 87% for Australian dataset while the smallest number of feature reduction was 75% for Diabetes dataset. For $IACO_R$ -SVM, the biggest reduction in number of features was 76% for Australian dataset while the smallest number of feature reduction was 54% for Heart dataset.

Table 4 Average Selected Feature Subset Size

Dataset	IACO-SVM	IACO _{MV} -SVM	GA-SVM without Feature Selection	GA-SVM with Feature Selection
Australian	3.3±0.46	1.8±0.4	6.7±3.16	5.2±2.15
Diabetes	2.6±0.49	2±0	5.1±1.63	3.7±1.26
Heart	6±0.45	2±0	7±1.05	6.2±1.12
German	6.4±0.63	3.9±0.3	11.8±3.33	10.3±1.76
Ionosphere	10.6±1.74	6±0	15.4±3.32	13.9±3.45

5 Conclusions

This study has investigated the hybridization of $IACO_R/IACO_{MV}$ and SVM technique to obtain optimal model parameters as well as to simultaneously optimize SVM parameters and features subset. Experimental results on five public UCI datasets showed promising performance in terms of test accuracy and features subset size. Possible extensions can focus on the area where other kernel parameters besides RBF, application to other SVM variants and multiclass.

ACKNOWLEDGEMENTS

The authors wish to thank the Ministry of Higher Education Malaysia for funding this study under Fundamental Research Grant Scheme, S/O code 12377 and RIMC, Universiti Utara Malaysia, Kedah for the administration of this study.

References:

- [1] Liu, W., & Zhang, D., Feature subset selection based on improved discrete particle swarm and support vector machine algorithm. *Paper presented at the IEEE International Conference of the Information Engineering and Computer Science*, (2009), (pp.1-4).
- [2] Vapnik, V., & Vashist, A., A new learning paradigm: learning using privileged information. *Neural networks: the official journal of the International Neural Network Society*, Vol. 22, 2009, No.5-6, pp.544-557.

- [3] Huang, C., & Wang, C., A GA-based feature selection and parameters optimization for support vector machines. *Expert system with Applications*, Vol.31, 2006, No.2, pp.231-240.
- [4] Ayat, N., Cheriet, M., & Suen, C., Automatic model selection for the optimization of SVM kernels. *Pattern Recognition*, Vol.38, 2005, No.10, pp.1733-1745.
- [5] R. Sivagaminathan and S. Ramakrishnan, A Hybrid Approach for Feature Subset Selection Using Neural Networks and Ant Colony Optimization, *Expert Systems with Applications*, Vol. 33, No. 1, pp. 49-60, 2007.
- [6] Huang C., ACO-Based Hybrid Classification System with Feature Subset Selection and Model Parameters Optimization, *Neurocomputing*, Vol.73, No.1-3, 2009, pp.438-448.
- [7] Vieira S., Sousa J, & Runkler T., Ant Colony Optimization Applied to Feature Selection in Fuzzy Classifiers. *Paper Presented at IEEE International Conference on Fuzzy System 2008*, (pp. 778-788).
- [8] Kabir M., Shahjahan M., & Murase K., An Efficient Feature Selection Using Ant Colony Optimization Algorithm. In C. Leung & J. Chan (Eds.), *Neural Information Processing*, (pp. 242-252). Springer-Verlag: Berlin Heidelberg, 2009.
- [9] Kabir M., Shahjahan M., Murase K., A New Hybrid Ant Colony Optimization Algorithm for Feature Selection, *Expert Systems with Applications*, Vol. 39, No. 3, 2012, pp. 3747-3763.
- [10] Abd-Alsabour N., Feature Selection for Classification Using an Ant System Approach, *Distributed, Parallel and Biologically Inspired Systems, IFIP Advances in Information and Communication Technology*, Vol. 329, 2010, pp 233-241.
- [11] Sanz S., Cumplido M., Cruz F., & Calzon C., Feature Selection via Genetic Optimization. In J. Dorronsoro (Ed.), *Artificial Neural Networks*, (pp. 547-552). Springer-Verlag: Berlin Heidelberg, 2002.
- [12] Zhao M., Fu C., Ji L., Tang K., & Zhou M., Feature Selection and Parameter Optimization for Support Vector Machines: A New Approach Based on Genetic Algorithm with Feature Chromosomes, *Expert System with Applications*, Vol. 38, No. 5, 2011, pp. 5197-5204.
- [13] Socha, K. & Dorigo, M., Ant colony optimization for continuous domain. *European Journal of Operational Research*, Vol.185, 2008, No.3, pp.1155-1173.
- [14] Liao, T., Dorigo, M., & Stutzle, T. (2011). Improved ant colony optimization algorithms for continuous and mixed discrete-continuous optimization problems. Retrieved from <http://www.swarmbots.org/~mdorigo/HomePageDorigo/thesis/dea/LiaoMAS.pdf>.
- [15] Qiu, D., Li, Y., Zhang, X., & Gu, B., Support vector machine with parameter optimization by bare bones differential evolution. *Paper presented at the 7th IEEE International Conference of the Natural Computation ICNC*, (pp. 263-266), 2011.
- [16] B. Tapas, B. Sanghamitra, & M. Ujjwal, Differential evolution based optimization of SVM parameters for meta classifier design. *Paper presented at Proceeding of 2nd International Conference on Computer, Communication, Control and Information Technology (C3IT-2012)*.
- [17] Kapp, M. N., Sabourin, R., & Maupin, P., A dynamic model selection strategy for support vector machine classifiers. *Applied Soft Computing*, Vol.12, 2012, No.8, pp.2550-2565.
- [18] L. La & Q. Guo, Text categorization using SVM with exponent weighted ACO. *Paper presented at Proceeding of the 31st Chinese Control Conference, 2012*. Hefei, China.
- [19] Pham D., Muhamad Z., Mahmuddin M., Ghanbarzadeh A., Koc E., & Otri S. Using the Bees Algorithm to Optimise a Support Vector Machine for Wood Defect Classification. *Paper presented at Innovative Production Machines and Systems Virtual Conference 2007*.
- [20] Alwan, H.B & Ku-Mahamud, K.R., Incremental Continuous Ant Colony Optimization Technique for Support Vector Machine Model Selection Problem. *Paper Presented at Proceedings the 17th WSEAS International Conference on Applied Mathematics (AMATH '12) 2012*. Montreux, Switzerland.
- [21] Alwan, H.B. & Ku-Mahamud, K.R., Optimizing support vector machine parameters using continuous ant colony optimization. *Paper Presented at Proceedings the 7th ICCCT International Conference on Computing and Convergence Technology (ICCIT, ICEI and ICACT) 2012*. Seoul, South Korea.
- [22] Alwan, H.B. & Ku-Mahamud, K.R., Incremental Mixed Variable Ant Colony Optimization for Feature Subset Selection and Model Selection. *Paper Presented at Proceedings the 12th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED '13) 2012*. Cambridge, UK.
- [23] *UCI Repository of Machine Learning Databases, Department of Information and Computer Science, University of California, Irvine, CA, <http://www.ics.uci.edu/mllearn/MLRepository>, 2012.*