MIXED VARIABLE ANT COLONY OPTIMIZATION TECHNIQUE FOR FEATURE SUBSET SELECTION AND MODEL SELECTION

Hiba Basim Alwan[1] and Ku Ruhana Ku-Mahamud[2]
[1, 2]*Universiti Utara Malaysia, Malaysia, hiba81basim@yahoo.com, ruhana@uum.edu.my*

ABSTRACT. This paper presents the integration of Mixed Variable Ant Colony Optimization and Support Vector Machine (SVM) to enhance the performance of SVM through simultaneously tuning its parameters and selecting a small number of features. The process of selecting a suitable feature subset and optimizing SVM parameters must occur simultaneously, because these processes affect each other which in turn will affect the SVM performance. Thus producing unacceptable classification accuracy. Five datasets from UCI were used to evaluate the proposed algorithm. Results showed that the proposed algorithm can enhance the classification accuracy with the small size of features subset.

Keywords: mixed variable ant colony optimization, support vector machine, features selection, model selection, pattern classification

INTRODUCTION

Pattern classification is an important area of machine learning and artificial intelligence. It attaches the input samples into one of a present number of groups through an approach. The approach is founded through learning the training data group (Wang et al., 2012). Certain pattern classification approaches allow the input data to include many features but in reality, only a few of them are relevant to classification. In certain circumstances, it is not suitable to choose a large number of features, as this may be difficult in calculating or it might be incomplete (Vieira, Sousa & Runkler, 2007).

The process of selecting features is called feature selection (FS). This process determines a subset of fields in the database and it minimizes the number of fields that appears during data classification (Huang, 2009). The main idea behind FS is to select a subset of input variables by deleting features that contain less or no information (Vieira, Sousa & Runkler, 2007). FS may be considered as an optimization problem which looks out for potential feature subsets which ultimately determines the optimal one (Abd-Alsabour, 2010).

Support Vector Machine (SVM) is an excellent classifier built on statistical learning approach, but it is not able to avoid the influence of the huge number of unrelated or redundant features on the classification results. Therefore, selecting a few numbers of suitable features would result in obtaining good classification accuracy (Liu & Zhang, 2009). The main concept of SVM is to obtain the Optimal Separating Hyperplane (OSH) between the positive and negative samples. This can be done through maximizing the margin between two parallel hyperplanes. After finding this plane, SVM can forecast the classification of unlabeled sample through asking on which side of the separating plan the sample lies (Vapnik & Vashist, 2009).Selecting the optimal feature subset and tuning SVM parameters to be used in SVM classifier are two problems in SVM classifier that influences the classification accuracy. These two problems must be solved because they affect each other (Huang, 2009).

The paper proposes an algorithm that is based on ACO_{MV} (Socha, 2008) to optimize SVM mixed variables. The work presented in this paper is an extension of Alwan & Ku-Mahamud (2012 and 2013) that can simultaneously optimize SVM parameters and feature subset. ACO_{MV} is known to have the ability to optimize discrete and continuous variables. The rest of the paper is organized as follows. Section 2 reviews previous studies on simultaneously optimizing SVM parameters and feature subset and Section 3 describes the proposed hybrid algorithm. Experimental results are discussed in Section 4 while concluding remarks and future works are presented in Section 5.

SIMULTANEOUS OPTIMIZATION OF FEATURE SUBSET SELECTION AND MODEL SELECTION FOR SUPPORT VECTOR MACHINE

Several studies suggested hybrid systems to enhance classification accuracy by using few and suitable feature subsets (Huang, 2009; Zhao et al., 2011; Sarafrazi & Nezamabadi-pour, 2013; Huang & Dun, 2008; and Lin et al., 2008. In these studies, feature subset and SVM parameters (C and γ RBF kernel) are simultaneously optimized. SVM is then used to measure the quality of the solution for all the hybrid systems. However, what differs is what the hybrid system is based on. GA is employed to select suitable feature simultaneously with optimize SVM parameter which were represented in the encoded chromosomes (Zhao et al., 2011). Discrete and continuous PSO values are mixed to simultaneously select suitable feature and optimize SVM parameter (Huang & Dun, 2008) while SA was used to simultaneously optimize model selection and feature subset selection (Lin et al., 2008). A hybrid system which is based on ACO and SVM was used in Huang (2009). The classical ACO has been employed to simultaneously select suitable feature and optimize SVM parameter. Two versions of Gravitational Search Algorithm (GSA) have been used in (Sarafrazi & Nezamabadi-pour, 2013): real value GSA (RGSA) to optimize the real value of SVM parameters and binary (discrete) value GSA (BGSA) to select a feature subset. These two algorithms have been applied to binary-class classification problem.

In conclusion, the above mentioned hybrid studies produced good results for classification accuracy with few numbers of selected features. Suggestions that were highlighted are as follows: Support Vector Regression (SVR) is to be applied for classification was suggested by Huang (2009) and Huang & Dun (2008), because SVR accuracy counts mainly on SVR parameters and selected feature subset. The use of continuous ACO to optimize the continuous value of SVM parameters was also suggested in Huang (2009). Other suggestions are to use other types of kernel function besides RBF and applying their works to solve on other real world problem.

PROPOSED HYBRID ALGORITHM

The proposed algorithm has employed the ACO_{MV} to optimize the feature subset selection and SVM classifier parameters. An ant’s solution is used to represent a combination of feature subset and 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 and pheromone table. 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 (Moustakidis & Theocharis, 2010), good performance in a major case (Zhang et al., 2009) and it only needs to use one parameter, which is the kernel parameter gamma (γ) (Huang & Wang, 2006). The overall process to hybridize ACO$_{MV}$ and SVM (ACO$_{MV}$-SVM) is as depicted in Figure 1.

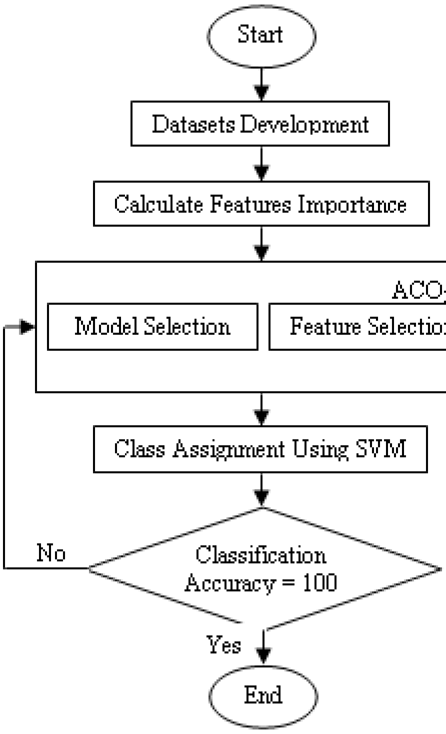

Figure 1. The proposed approach's flowchart

The main steps are (1) initializing solution archive, pheromone table and algorithm parameters, (2) solution construction for feature subset and determining the C and γ parameters (3) establishing an SVM classifier model and (4) updating solution archives and pheromone table.

F-score is used as a measurement to determine the importance of feature. This measurement is used to judge the favouritism capability of a feature. The high value of F-score indicates favourable feature. The calculation of the F - score is as follows (Huang, 2009):

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

where $i = 1, 2, \ldots, N_f$, $\mathbf{v}$ 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 ith feature with categorical value c, $c \in \{1, 2, \ldots, v\}$, $\bar{x}_{i,j}^{(c)}$ is the j^{th} training sample for the ith feature with categorical value c, $j \in \{1, 2, \ldots, N_i^{(c)}\}$, $\bar{x}_i$ is the ith feature, and $\bar{x}_i^{(c)}$ is the i^{th} feature with categorical value c.

In the initialization step, each ant established a solution path for parameter C, parameter γ, and feature subset. Three solution archives are needed to design the transition probabilities for the first feature in the feature subset, C and γ, and one pheromone table. The range for C and γ values will be sampled according to random parameter k which is the initial archive size of solutions archives for C and for γ, while the size of solution archive of features will be equal to the number of features. The weight vector, w, is then computed for each sample for C and γ as follow:

$$w_l = \frac{1}{qk\sqrt{2\pi}} e^{-\frac{(l-1)^2}{2q^2k^2}} \tag{2}$$

The features are computed using:

$$w_f = \frac{w_l}{u} + \frac{q}{\eta} \tag{3}$$

where u is the number of how many $feature_i$ is selected, $\eta$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 solution archives. Once this step is completed, the sampling procedure will be constructed for C and γ in two phases. Phase one involves choosing one of the weight vectors using a probability calculated as follows:

$$p_l = \frac{w_l}{\sum_{r=1}^{k} w_r} \tag{4}$$

The second phase involves sampling selecting w via a random number generator that is able to generate random numbers according to a parameterized normal distribution. This initialization will construct the transition probabilities. The probability transition that is used to move from γ to the first feature in features subset is as follow:

$$p_f = \frac{w_{f_i} * F - Score_{f_i}}{\sum_{i=1}^{n} w_{f_i} * F - Score_f} \tag{5}$$

In order to select other features that construct the features subset, the following probability transition is used:

$$Prob_{ij}^{k} = \begin{cases} \dfrac{(Prob_{ij})^{\alpha}(F - Score_j)^{\beta}}{\sum_{j \in I_i^k}(Prob_{ij})^{\alpha}(F - Score_j)^{\beta}} & if j \in I_i^k \\ 0 & otherwise \end{cases} \tag{6}$$

Like the solution archives and pheromone table, some important system parameters must be initialized as follows: the number of ants = 2, $q = 0.1$, number of runs = 10, $\alpha = 1$, $\beta = 2$, 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,γ and features subset. In order to construct the solution, three transition probabilities with various solutions archives and pheromone table are needed. These transition probabilities will be computed according to Eq. (4), Eq. (5), and Eq. (6).

Classifier model will be constructed in step four. Solution is generated by each ant and will be evaluated based on classification accuracy and feature weight obtained by SVM model utilizing k-fold Cross Validation (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 subgroups 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:

$$CV_{accuracy} = \frac{\sum_i test\_accuracy}{k}, i = 1, 2, \dots, k \tag{7}$$

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 compute as follows:

$$TestAccuracy = \frac{no.of correctlypredicteddata}{totaltestingdata} * 100\% \tag{8}$$

The benefits of using CV are (1) each of the test groups is independent and (2) the dependent outcomes can be enhanced (Huang, 2009).

The final step is related to updating the solution archives and pheromone table. The solution archives' modification will be done by appending the newly generated group solutions that gave the best classification accuracy to solution archive and then deleting the exact number of worst solutions. This will ensure the size of solution archive does not change. This procedure guaranteed that just good solutions are stored in the archive and it will efficiently influence the ants in the seek process. The pheromone table will be modified as follow:

$$T_{ij}(t+1) = pT_{ij} + \sum_{k=1}^{m} \Delta T_{ij}^{k}(t) \tag{9}$$

$$\Delta T_{ij}^{k} = \begin{cases} CVACC^{k} * Weight_i^k * Weight_j^k & if antkuseedge\ (i,j) \\ 0 & otherwise \end{cases} \tag{10}$$

EXPERIMENTAL RESULTS

Five datasets were used in evaluating the proposed ACO_{MV}-SVM algorithm. The datasets are Australian, Pima-Indian Diabetes, Heart, German, Splice datasets, available from University California, Irvin (UCI) Repository of Machine Learning Databases. The summary of these datasets are presented in Table 1.

Table 1. Summarization of UCI's Datasets Repository

Dataset	No. of Instances	No. of Features	No. of Classes	Type of Datasets
Australian	690	14	2	Categorical, Integer, Real
Pima-Indian Diabetes	760	8	2	Integer, Real
Heart	270	13	2	Categorical, Real
German	1000	24	2	Categorical, Integer
Splice	1000	60	2	Categorical

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.

$$\bar{x} = \frac{x - min_i}{max_i - min_i} \tag{11}$$

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 (Huang, 2009).

The performance of the proposed ACO_{MV}-SVM was compared with GA with feature chromosome (Zhao et al., 2011) and with previous work on ACO_R-SVM and was based on one run (Alwan & Ku-Mahamud, 2012).

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

Table 2 shows the average classification accuracy that produced in all the ten runs. The classification accuracy of the proposed ACO_{MV}-SVM algorithm was compared with GA-SVM and ACO_R-SVM results. The proposed approach classifies patterns with higher accuracy compared to GA-SVM and ACO_R-SVM for all five datasets. The average percentage increased in accuracy for all datasets is about 10.194 when compared with GA-SVM and 6.01 when compared with ACO_R-SVM. This is because the integration of ACO_{MV} with SVM, ACO_{MV} as an optimization approach improves SVM classification accuracy through optimizing its parameters which are the regularization parameter *C* and (γ) of RBF kernel function and selecting suitable features subset.

Table 2. Classification Accuracy (%)

Dataset	ACO_{MV}-SVM	GA-SVM	ACO_R-SVM
Australian	98.75	91.59	96.14
Pima-Indian Diabetes	96.00	83.84	87.79
Heart	98.70	95.56	89.99
German	98.00	86.10	94.00
Splice	99.00	90.53	96.22

Table 3 shows the average selected features subset size that was produced in all the ten runs. The average number of selected features of the proposed ACO_{MV}-SVM algorithm is the lowest when compared with GA-SVM and ACO_R-SVM. The biggest reduction in number of features is 76.42% for Splice dataset while the smallest number of feature reduction is 57.5% for Diabetes dataset.

Table 3. Average Selected Feature Subset Size

Dataset	No. of Features	ACO_{MV}-SVM	GA-SVM	ACO_R-SVM
Australian	14	4.350	5.2	3.2
Pima-Indian Diabetes	8	3.4	3.7	2.3
Heart	13	5.450	6.2	5.9
German	24	7.150	10.3	5.6
Splice	60	14.15	-	26.3

CONCLUSIONS

This study has investigated a hybrid ACO_{MV} and SVM technique to obtain optimal model parameters and features subset. Experimental results on five public UCI datasets have showed promising performance in terms of classification 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 data.

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

Abd-Alsabour N. (2010). Feature selection for classification using an ant system approach. In M. Hinchey, B. Kleinjohann, L. Kleinjohann, P. Lindsay, F. Ramming, J. Timmis & M. Wolf (Eds.), *Distributed, Parallel and Biologically Inspired Systems*, 233-241.

Alwan, H.B. & Ku-Mahamud, K.R. (2012). Optimizing support vector machine parameters using continuous ant colony optimization. *Proceedings of the 7th International Conference on Computing and Convergence Technology*, 164-169.

Alwan, H.B. & Ku-Mahamud, K.R. (2013). Solving support vector machine model selection problem using continuous ant colony optimization. *International Journal of Information Processing and Management*, *4*(2), 86-97.

Huang, C. & Dun, J. (2008). A Distributed PSO-SVM hybrid system with feature selection and parameter optimization. *Applied Soft Computing*, *8*(4), 1381-1391.

Huang, C. (2009). ACO-based hybrid classification system with feature subset selection and model parameters optimization. *Neurocomputing*, 73(1-3), 438-448.

Huang, C. & Wang, C. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications*, 31(2), 231-240.

Lin, S., Lee, Z.-J., Chen, S.-C. & Tseng, T.-Y. (2008). Parameter determination of support vector machine and feature selection using simulated annealing approach. *Applied Soft Computing*, *8*(4), 1505-1512.

Liu, W. & Zhang, D. (2009). Feature subset selection based on improved discrete particle swarm and support vector machine algorithm. *Proceedings of The International Conference on Information Engineering and Computer Science*, 1-4.

Moustakidis, S. & Theocharis, J. (2010). SVM-FuzCoC: A novel SVM-based feature selection method using a fuzzy complementary criterion. *Pattern Recognition*, *43*(11), 3712-3729.

Sarafrazi, S. & Nezamabadi-pour, H. (2013). Facing the classification of binary problems with a GSA-SVM hybrid system. *Mathematical and Computer Modeling*, *57*(1-2), 270-278.

Socha, K. (2008). Ant colony optimization for continuous and mixed-variables domain. (Doctoral dissertation, Universite' Libre de Bruxelles, 2008). Retrieved from iridia.ulb.ac.be/~mdorigo/HomePageDorigo/thesis/SochaPhD.pdf.

UCI Repository of Machine Learning Databases, Department of Information and Computer Science, University of California, Irvine, CA, <http://www.ics.uci.edu/mlearn/MLRepository>, 2012.

Vapnik, V. & Vashist, A. (2009). A new learning paradigm: learning using privileged information. *Neural Networks: The Official Journal of The International Neural Network Society*, *22*(5-6), 544–557.

Vieira, S., Sousa, J. & Runkler, T. (2007). Ant colony optimization applied to feature selection in fuzzy classifiers. In P. Melin, O. Castillo, L. Aguilar, J. Kacprzyk & W. Pedrycz (Eds.), *Foundations of Fuzzy Logic and Soft Computing*, 778-788.

Wang, X., Liu, X., Pedrycz, W., Zhu, X. & Hu, G. (2012). Mining axiomatic fuzzy association rules for classification problems. *European Journal of Operational Research*, *218*(1), 202-210.

Zhang, H., Xiang, M., Ma, C., Huang, Q., Li, W., Xie, W., Wei, Y. & Yang, S. (2009). Three-class classification models of LogS and LogP derived by using GA-CG-SVM approach. *Molecular Diversity*, *13*(2), 261-268.

Zhao M., Fu C., Ji, L., Tang K., & Zhou, M. (2011). Feature selection and parameter optimization for support vector machines: a new approach based on genetic algorithm with feature chromosomes, *Expert System with Applications*, *38(5)*, 5197-5204.