A Hybrid PSO-FSVM Model and Its Application to Imbalanced Classification of Mammograms

Hussein Samma[1], Chee Peng Lim[2,*], and Umi Kalthum Ngah[1]

[1] Imaging & Computational Intelligence Group (ICI), School of Electrical and Electronic Engineering, University of Science Malaysia, Malaysia

[2] Centre for Intelligent Systems Research, Deakin University, Australia

chee.lim@deakin.edu.au

Abstract. In this work, a hybrid model comprising Particle Swarm Optimization (PSO) and the Fuzzy Support Vector Machine (FSVM) for tackling imbalanced classification problems is proposed. A PSO algorithm, guided by the G-mean measure, is used to optimize the FSVM parameters in imbalanced classification problems. The hybrid PSO-FSVM model is evaluated using a mammogram mass classification problem. The experimental results are analyzed and compared with those from other methods. The outcomes positively demonstrate that the proposed PSO-FSVM model is able to achieve comparable, if not better, results for imbalanced data classification problems.

1 Introduction

Imbalanced classification problems occur when the instances in one class (i.e., the minority class) is very rare as compared with those in other classes (i.e., the majority classes). These problems exist in various fields, e.g. network intrusion detection [1] and breast cancer classification[2]. This study focuses on using a Fuzzy Support Vector Machine (FSVM) for undertaking imbalanced classification problems. While a number of computational intelligence models have been developed to undertake various problems in our previous studies [3-9], the models devised are not designed specifically for imbalanced learning and classification problems. As reported in the literature, Support Vector Machine (SVM)-based methods have been successfully applied to tackling imbalanced classification problems [10-20]. Generally these methods can be divided into two categories, i.e., integration of the SVM with data sampling techniques [11, 12, 14, 19] or modifications of the SVM [10, 13, 15-18, 20] to make it less sensitive to the class imbalanced problems. In the first category, sampling techniques (either over or under sampling) are first applied to the data level. Then, the SVM is used to process the sampled data.

Recently, Fuzzy Support Vector Machine (FSVM) [21] has been used for undertaking imbalanced learning problems [15]. Nonetheless, performances of an

[*] Corresponding author.

A. Selamat et al. (Eds.): ACIIDS 2013, Part I, LNAI 7802, pp. 275–284, 2013.

SVM-based model depends on the settings of its parameters, and inappropriate parameter selection leads to poor performances [22]. In this regards, Particle Swarm Optimization (PSO) has been used to optimize the SVM parameters [23, 24].

Motivated by the effectiveness of the FSVM in tackling imbalanced learning problems [15] and the usefulness of PSO in fine-tuning the SVM parameters [23, 24], this study proposes a hybrid PSO-FSVM model for undertaking imbalanced classification problems. The PSO algorithm is used to tune three FSVM parameters simultaneously, *viz.*, the penalty parameter (c) (which controls the trade-off between the model complexity and the training error rate); the kernel parameter (γ) of the kernel function (i.e., the Radial Basis Function); and the maximum fuzzy membership value (r). The applicability of the resulting PSO-FSVM model is demonstrated using a mammogram mass classification problem.

The organization of the paper is as follows. In Sections 2, and 3, the background pertaining to PSO and FSVM is given. Hybridization of PSO-FSVM is explained in Section 4. In Section 5, an experimental study to evaluate the effectiveness of PSO-FSVM using a mammogram mass classification problem is described. Concluding remarks and suggestions for further work are presented in Section 6.

2 Particle Swarm Optimization

PSO is a swarm-based optimization algorithm[25], which is inspired by the social behavior of organized colonies. A population in PSO is called a swarm, and it consists of particles which interact with each other to explore the search space. Each particle in the swarm is associated with two vectors, i.e., the velocity (V) and position (X) vectors, as follows:

$$X_i = [x_i^1, x_i^2, x_i^3, \ldots\ldots\ldots\ldots\ldots, x_i^D] \tag{1}$$

$$V_i = [v_i^1, v_i^2, v_i^3, \ldots\ldots\ldots\ldots\ldots, v_i^D] \tag{2}$$

where D represents the dimensions of the problem, i denotes the particle number in the swarm. Each particle is first initialized with random velocity and position vectors according to its solution range. During the search process, the velocity and the position vectors are updated, as follows:

$$V_i = w * V_i + c_1 * \text{rand}_1(\text{pBest} - X_i) + c_2 * \text{rand}_2(\text{gBest} - X_i) \tag{3}$$

$$X_i = X_i + V_i \tag{4}$$

where w is the inertia weight, c_1 and c_2 are the acceleration coefficients, $rand_1$ and $rand_2$ are two uniformly distributed random numbers in [0 1], *pBest* is the local best position achieved by the particle, and *gBest* is the global best position achieved by the swarm.

We adopt the procedure in [26] for setting the acceleration coefficients, i.e., the values of c_1 deceases from [2.5 to 0.5] and the values of c_2 increases from [0.5 to 2.5], respectively. As such, equations (5) and (6) are used:

$$c_1 = (0.5 - 2.5) * \frac{\text{current}_{\text{Iter}}}{\text{max}_{\text{Iter}}} + 2.5 \tag{5}$$

$$c_2 = 0.5 + \frac{\text{current}_{\text{Iter}}}{\text{max}_{\text{Iter}}} * (2.5 - .5) \tag{6}$$

where max_{Iter} is the maximum number of iterations, $\text{current}_{\text{Iter}}$ is the current iteration. The inertia weight parameter is a uniform random number $\epsilon[0\ 1]$, as suggested in [27].

3 Fuzzy Support Vector Machine

Originated from statistical learning theory, SVM aims to minimize structural risks by finding an optimal separating hyperplane which gives a low generalization error [21]. In SVM learning, minimization of the following function is considered

$$\min \frac{1}{2}\|w\|^2 + c \sum_{i=1}^{L} \xi_i \tag{7}$$

$$\text{subject to} \quad y_i (wx_i + b) \geq 1 \tag{8}$$

where ξ is the slack variable that measures the degree of misclassification, C is the penalty parameter that controls the trade-off between maximization of the margin and minimization of the classification error, b is the bias, and w is the weighted normal vector.

FSVM [21] is similar to SVM except that a fuzzy membership value, $m_{i.}$, is introduced, as follows

$$\min \frac{1}{2}\|w\|^2 + c \sum_{i=1}^{L} m_i \xi_i \tag{9}$$

$$\text{subject to} \quad y_i (wx_i + b) \geq 1 \tag{10}$$

The membership value, m_i, for each data point, x_i, in the training set is incorporated into the objective function (Eqn. 9). A smaller value of m_i makes the corresponding point x_i to be treated as less important, and vice versa. In this way, the cost of misclassification can be controlled by assigning lower membership values for less expensive classes and higher membership values for more expensive classes.

Another issue in FSVM is the determination of the membership value, $m_{i.}$ As stated in [21], the membership value can be computed as the distance between instance x_i and the class center. Let m_i+ to be the membership value for a more expensive class and m_i- is the membership value for a less expensive one, which are defined as follows:

$$m_i + = F(x_i) \tag{11}$$

$$m_i + = F(x_i) * \text{ratio} \tag{12}$$

where $ratio \in [\sigma \quad 1]$, and δ is a very small value selected by the user. As such, tuning the ratio value makes the FSVM classifier feasible in tackling imbalanced classification problems. .

4 A Hybrid PSO-FSVM Model

In this section, we proposed a hybrid PSO-FSVM model for undertaking imbalanced classification problems. The representation of the solution, fitness function, and procedure of the PSO-FSVM hybrid model are as follows.

(a) Solution representation

The FSVM with RBF kernels [15] , $k(x_i, x_j) = e^{-\gamma} \|x_i - x_j\|^2$, is used in our proposed PSO-FSVM hybrid model. As shown in Table 1, each particle in the swarm consists of three real-valued variables, i.e., the penalty parameter C, the RBF kernel function parameter γ, and the ratio parameter that controls the maximum fuzzy membership value for the majority class instances, r.

Table 1. Parameters of a particle

penalty parameter (C)	RBF kernel parameter (γ)	Ratio parameter(r)
$x_{i,1}$	$x_{i,2}$	$x_{i,3}$

(b) Procedure

The main steps of the proposed PSO-FSVM model are illustrated in Fig 1, as follows.

1- Data pre-processing: The training data samples are scaled to lie within [-1 1].
2- Training: The FSVM classifier is trained using the training set.
3- Fitness evaluation: A validation set is used with the G-mean measure as the fitness function, and it is calculated for each particle in the swarm.
4- Stopping criteria: Two stopping conditions for the PSO algorithm are applied, i.e., either the maximum number of the PSO operations is met, or a G-mean value of 100 is achieved during the execution process.
5- PSO updating: The position and velocity vectors of each particle are updated according to equations (3) and (4).

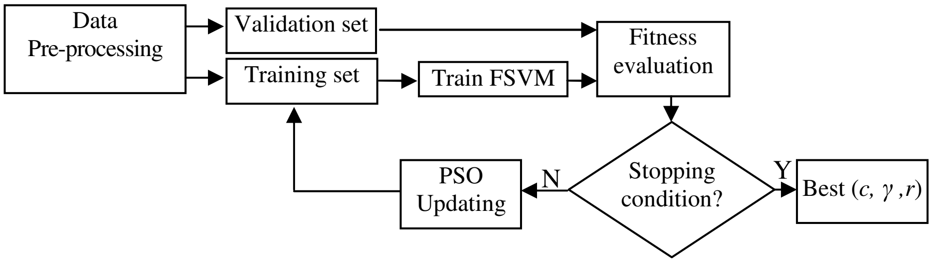

Fig. 1. PSO-FSVM model

(c) Fitness function

The G-mean measure is one of the assessment indicators that has been used to evaluate imbalanced data classification problems [15, 28, 29]. It is defined as follows.

$$\mathrm{G-mean} = \sqrt{\mathrm{Sensitivity} * \mathrm{Specificity}} \tag{13}$$

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{14}$$

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{15}$$

where TP is the true positive rate, FN is the false negative rate, TN is the true negative rate, and FP is the false positive rate.

5 An Experimental Study

The proposed PSO-FSVM model was evaluated using the MIAS mammogram data set [30]. A total of 92 abnormal Region of Interests (ROIs) (i.e., malignant and benign regions) and a total of 2090 normal ROIs were used in the experiment. As an example, a number of normal and abnormal ROI samples are shown in Fig 2. All ROI samples were re-scaled to 16x16 dimensions, and the intensity value was used to generate 256 feature vector lengths. The abnormal and normal ROIs were divided randomly into a training set and a testing set. For the training set, one third of the training samples were used for validation and the rest for training the FSVM classifier.

The PSO parameters were configured as in [26]. The size of the swarm and the maximum number of iterations were empirically chosen to be 20, and 1000, respectively. These values were found to be appropriate for this study by the trial-and-error method. The FSVM classifier with RBF kernels was used in the experiment. Both c and γ values were tuned within the logarithmic scale as recommended in [31], i.e., ln(c) =[-10,10], ln(γ) =[-10,10]. The last parameter, r, was tuned within the range r= [δ ,1],where δ is a very small value selected by the user (δ =10^(-6)).

Fig. 2. (a) abnormal ROIs , (b) normal ROIs

5.1 Results and Discussions

In this study, the Center Composite Design (CCD) technique [32] was employed to analyze the effect of the PSO parameters (i.e., size of the swarm, S, and the number of iterations, N) on the FSVM accuracy rates. The CCD technique has been used to investigate the interaction between PSO parameters [33, 34]. Its basic idea is to set the value of each parameter at different levels (i.e. Low, Medium, and High), as shown in Table 2. Then, the combination of the experimented parameters at a different level is generated to study the interaction between the parameters, as shown in Fig 3. In this experiment, a total of five experiments were carried out using different levels for both S and N settings.

Table 2. Parameters and levels of the CCD experiment

Parameter	Level		
	Low (-1)	Medium (0)	High (+1)
S	10	20	40
N	100	1000	2000

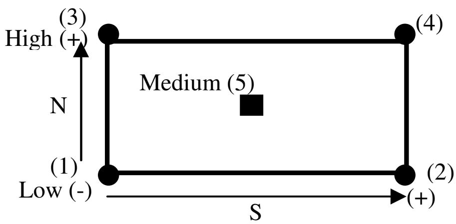

Fig. 3. A CCD experiment with two parameters and five points (i.e. one center and four corners)

Each experiment was repeated 10 times, and the average G-mean values were computed. The results are shown in Fig 4. As can be seen, the best G-mean value was produced by experiments 4 and 5. The parameter settings in experiment 5 (S=2, and N=1000) were used to compare PSO-FSVM with other methods in the next experiment.

A further experiment to compare the results of PSO-FSVM with those from PSO-SVM, FSVM, and SVM models, trained using the same data set, was conducted. In PSO-SVM, only two parameters were tuned by PSO, i.e. c and γ. In SVM and FSVM classifiers, three-fold cross validation was used to tune their parameters.

The experiment was repeated 10 times, and the mean and standard deviation of the sensitivity and specificity rates were computed. As shown in Table 3, the sensitivity rate and the G-mean value of PSO-FSVM are better than those from other SVM-based classifiers. However, its specificity rate is lower since the objective function used was maximization of the G-mean value, which was aimed to balance the true positive rate and the true negative rate.

Further analysis was conducted using the Receiver Operating Curve (ROC) as a graphical comparison method. As shown in Fig 5, PSO-FSVM produced a better ROC as compared with SVM-based classifiers. Moreover, the area under the ROC (AUC) was calculated, as in Table 3. The bootstrap method [35] was employed to assess the AUC values, with the significance level (α) set to 0.05 (i.e. 95% confidence level). As shown in Table 3, the *p-value* of the AUC measure is lower than α, which indicates that PSO-FSVM performed significantly better than the SVM-based FSVM classifiers (at the 95% confidence level).

Table 3. Comparison among different SVM classifiers

Model	Sensitivity (Std.dev)	Specificity (Std.dev)	G-mean (Std.dev)	AUC (Std.dev)	*p* value of bootstrap test
					AUC
MIAS dataset					
PSO-FSVM	**89.13** **(1.49e-014)**	90.21 (4.62e-002)	**89.67** **(2.29e-002)**	0.95 (1.31e-004)	-
PSO-SVM	49.35 (1.05)	**98.90** **(4.94e-002)**	69.86 (7.38e-001)	0.94 (4.59e-005)	0
SVM	51.09 (3.87)	98.68 (2.38e-001)	70.95 (2.63e+000)	0.94 (4.29e-003)	0
FSVM	51.30 (3.27)	98.74 (2.01e-001)	71.14 (2.21e+000)	0.94 (8.04e-003)	0.002

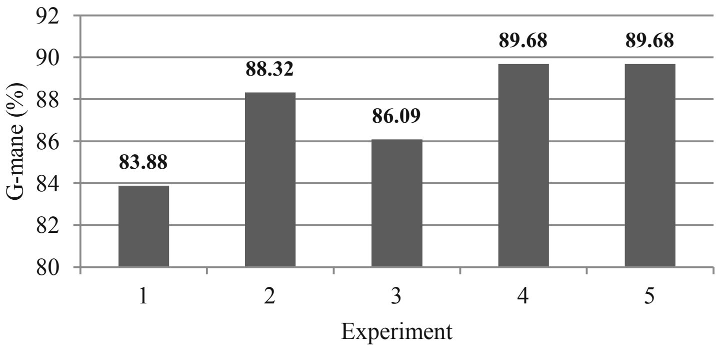

Fig. 4. Average G-mean values of PSO-FSVM

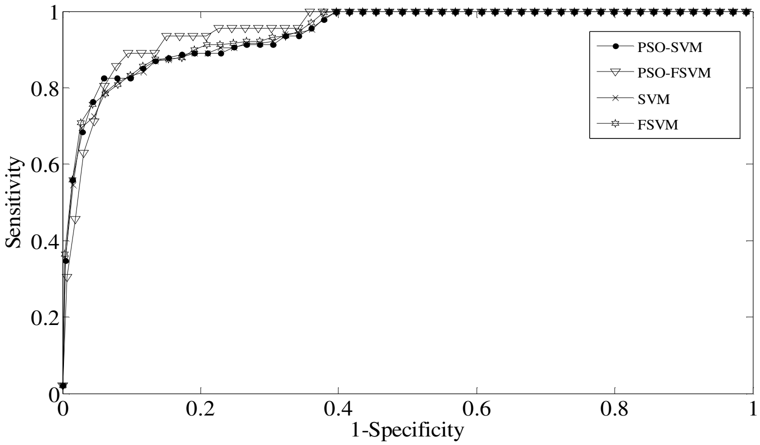

Fig. 5. Performance comparison with ROC with different PSO parameter settings

6 Summary

In this paper, a hybrid PSO-FSVM model has been proposed to undertake imbalanced classification problems. The effectiveness of PSO-FSVM has been evaluated using a mammogram mass classification problem. The results positively show that PSO-FSVM is able to produce good results, as compared to those from other methods.

While the performance of PSO-FSVM is encouraging, more experiments using imbalanced data sets are currently underway to further validate the usefulness of PSO-FSVM in undertaking imbalanced classification problems.

References

1. Vegard, E., Jonathan, V., Keith, P.: Enhancing network based intrusion detection for imbalanced data. Int. J. Know.-Based Intell. Eng. Syst. 12, 357–367 (2008)
2. Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., Tourassi, G.D.: Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. Neural Networks 21, 427–436 (2008)
3. Chen, K., Lim, C., Lai, W.: Application of a Neural Fuzzy System with Rule Extraction to Fault Detection and Diagnosis. Journal of Intelligent Manufacturing 16, 679–691 (2005)
4. Eric, W.M.L., Lee, Y.Y., Lim, C.P., Tang, C.Y.: Application of a noisy data classification technique to determine the occurrence of flashover in compartment fires. Adv. Eng. Inform. 20, 213–222 (2006)
5. Goh, W.Y., Lim, C.P., Peh, K.K., Subari, K.: Application of a Recurrent Neural Network to Prediction of Drug Dissolution Profiles. Neural Computing & Applications 10, 311–317 (2002)
6. Chee-Peng, L., Jenn-Hwai, L., Mei-Ming, K.: A Hybrid Neural Network System for Pattern Classification Tasks with Missing Features. IEEE Trans. Pattern Anal. Mach. Intell. 27, 648–653 (2005)

7. Hua Tan, K., Peng Lim, C., Platts, K., Shen Koay, H.: An intelligent decision support system for manufacturing technology investments. International Journal of Production Economics 104, 179–190 (2006)
8. Chee Peng, L., Harrison, R.F.: Online pattern classification with multiple neural network systems: an experimental study. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 33, 235–247 (2003)
9. Lim, C.P., Quek, S.S., Peh, K.K.: Prediction of drug release profiles using an intelligent learning system: an experimental study in transdermal iontophoresis. Journal of Pharmaceutical and Biomedical Analysis 31, 159–168 (2003)
10. Hsu, C.-C., Wang, K.-S., Chang, S.-H.: Bayesian decision theory for support vector machines: Imbalance measurement and feature optimization. Expert Systems with Applications 38, 4698–4704 (2011)
11. Liu, Y., An, A., Huang, X.: Boosting Prediction Accuracy on Imbalanced Datasets with SVM Ensembles. In: Ng, W.-K., Kitsuregawa, M., Li, J., Chang, K. (eds.) PAKDD 2006. LNCS (LNAI), vol. 3918, pp. 107–118. Springer, Heidelberg (2006)
12. Liu, Y., Yu, X., Huang, J.X., An, A.: Combining integrated sampling with SVM ensembles for learning from imbalanced datasets. Information Processing & Management 47, 617–631 (2011)
13. Veropoulos, K., Campbell, C., Cristianini, N.: Controlling the Sensitivity of Support Vector Machines. In: the International Joint Conference on AI, pp. 55–60 (1999)
14. Kang, P., Cho, S.: EUS SVMs: Ensemble of Under-Sampled SVMs for Data Imbalance Problems. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) ICONIP 2006. LNCS, vol. 4232, pp. 837–846. Springer, Heidelberg (2006)
15. Batuwita, R., Palade, V.: FSVM-CIL: Fuzzy Support Vector Machines for Class Imbalance Learning. IEEE Transactions on Fuzzy Systems 18, 558–571 (2010)
16. Wu, G., Chang, E.Y.: KBA: kernel boundary alignment considering imbalanced data distribution. IEEE Transactions on Knowledge and Data Engineering 17, 786–795 (2005)
17. Zhao, Z., Zhong, P., Zhao, Y.: Learning SVM with weighted maximum margin criterion for classification of imbalanced data. Mathematical and Computer Modelling 54, 1093–1099 (2011)
18. Hwang, J.P., Park, S., Kim, E.: A new weighted approach to imbalanced data classification problem via support vector machine with quadratic cost function. Expert Systems with Applications 38, 8580–8585 (2011)
19. Yuchun, T., Yan-Qing, Z., Nitesh, V.C., Sven, K.: SVMs modeling for highly imbalanced classification. Trans. Sys. Man Cyber. Part B 39, 281–288 (2009)
20. Imam, T., Ting, K.M., Kamruzzaman, J.: z-SVM: An SVM for Improved Classification of Imbalanced Data. In: Sattar, A., Kang, B.-H. (eds.) AI 2006. LNCS (LNAI), vol. 4304, pp. 264–273. Springer, Heidelberg (2006)
21. Chun-Fu, L., Sheng-De, W.: Fuzzy support vector machines. IEEE Transactions on Neural Networks 13, 464–471 (2002)
22. Keerthi, S.S., Lin, C.-J.: Asymptotic behaviors of support vector machines with Gaussian kernel. Neural Computation 15, 1667–1689 (2003)
23. Guo, X.C., Yang, J.H., Wu, C.G., Wang, C.Y., Liang, Y.C.: A novel LS-SVMs hyper-parameter selection based on particle swarm optimization. Neurocomputing 71, 3211–3215 (2008)
24. Lin, S.-W., Ying, K.-C., Chen, S.-C., Lee, Z.-J.: Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Systems with Applications 35, 1817–1824 (2008)

25. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: IEEE International Conference on Neural Networks, vol. 4, 1944, pp. 1942–1948 (1995)
26. Ratnaweera, A., Halgamuge, S.K., Watson, H.C.: Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. IEEE Transactions on Evolutionary Computation 8, 240–255 (2004)
27. Sheng, C., Xia, H., Harris, C.J.: Particle Swarm Optimization Aided Orthogonal Forward Regression for Unified Data Modeling. IEEE Transactions on Evolutionary Computation 14, 477–499 (2010)
28. Gao, M., Hong, X., Chen, S., Harris, C.J.: A combined SMOTE and PSO based RBF classifier for two-class imbalanced problems. Neurocomputing 74, 3456–3466 (2011)
29. Xia, H., Sheng, C., Harris, C.J.: A Kernel-Based Two-Class Classifier for Imbalanced Data Sets. IEEE Transactions on Neural Networks 18, 28–41 (2007)
30. The Mini-MIAS Database of Mammograms, `http://peipa.essex.ac.uk`
31. Li, S., Tan, M.: Tuning SVM parameters by using a hybrid CLPSO-BFGS algorithm. Neurocomputing 73, 2089–2096 (2010)
32. Montgomery, D.C.: Design and analysis of experiments. Wiley (1997)
33. Pan, Q.-K., Fatih Tasgetiren, M., Liang, Y.-C.: A discrete particle swarm optimization algorithm for the no-wait flowshop scheduling problem. Computers & Operations Research 35, 2807–2839 (2008)
34. Wang, C.-H., Lin, T.-W.: Improved particle swarm optimization to minimize periodic preventive maintenance cost for series-parallel systems. Expert Systems with Applications 38, 8963–8969
35. Efron, B.: Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics 7, 1–26 (1979)