International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.8, No.8 (2015), pp.299-308
http://dx.doi.org/10.14257/ijsip.2015.8.8.31

Least Squares Fuzzy One-class Support Vector Machine for Imbalanced Data

Jingjing Zhang [1], Kuaini Wang [1], Wenxin Zhu [2] and Ping Zhong [1*]

[1] *College of Science, China Agricultural University, 100083 Beijing, China*
[2] *Department of Basic, Science Tianjin Agricultural University, 300384 Tianjin, China*
zhjingjing09@163.com, wangkuaini1219@sina.com, zhuwenxinyan@163.com
[*] *Corresponding author: zping@cau.edu.cn*

Abstract

Based on fuzzy one-class support vector machine (SVM) and least squares (LS) one-class SVM, we propose an LS fuzzy one-class SVM to deal with the class imbalanced problem. The LS fuzzy one-class SVM applies a fuzzy membership to each sample and attempts to solve the modified primal problem. Hence, we just need to solve a system of linear equations as opposed solving the quadratic programming problem (QPP) in fuzzy one-class SVM, which leads to an extremely simple and fast algorithm. Numerical experiments on several benchmark data sets demonstrate the feasibility and effectiveness of the proposed algorithm.

Keywords: *Class imbalanced learning, One-class SVM, Least squares, Fuzzy information*

1. Introduction

There are many class imbalanced problems in real-world applications such as information retrieval and filtering [1], the detection of oil spills [2] and the creditcard fraud detection [3]. The class imbalanced problem is characterized as one class is represented by a large number of samples while the other is represented by only a few samples. In traditional algorithms, the decision boundary is always skewed toward the minority class. Therefore, improving the accuracy of the minority class is a meaningful research. In the past few years, there have been many studies on reducing the bias of decision boundaries produced by the class imbalanced problem [4-6]. For the imbalanced learning, there are mainly two categories of methods to develop better performing classifiers with imbalanced data sets. One is sampling method, either randomly or intelligently, to obtain an altered class distribution [7-10]. The other is to modify the learning algorithms such as cost sensitive method [11-12], margin calibration [13] and kernel-based method [14]. On heavily-imbalanced data, applying one-class SVM to the class imbalanced problems is widely available [15-18].

The support vector machine (SVM) can effectively deal with the binary classification problem [19-21] by using both positive and negative samples. However, in one-class problems, usually, only one class is available, and others are difficult or expensive to obtain. The one-class SVM only uses positive samples in training process. It can generally be used to solve the following problems such as anomaly detection or novelty detection [22], class imbalanced problem [18] and image processing [23]. Schölkopf et al. [24] proposed a method of adapting the SVM conception for the one-class SVM. Its strategy is to map the samples into a high-dimensional feature space corresponding to a kernel and construct a maximal margin hyperplane to separate the training samples from

Copyright ⓒ 2015 SERSC

the origin. Instead of a hyperplane, a hypersphere is used to encircle the training samples, which is called support vector data description (SVDD) [25]. Many efforts have been made to reduce the time complexity or raise the classification accuracy. To decrease the training time, the author proposed the least squares version based on the standard one-class SVM [26]. To increase the accuracy, the authors assign each sample a membership value according to its importance in the data set. The benefit of this fuzzy information is raising the effectiveness of algorithm and reducing the effects of outliers or noise samples [27-28].

In this paper, we propose a least squares fuzzy one-class SVM (LS fuzzy one-class SVM). Firstly, different from the fuzzy one-class SVM, the proposed method uses a quadratic loss function in objective function and equalities in the constraints. The solution only requires solving a set of linear equations, instead of solving a complex QPP in the original method. Thus our method owns low time complexity. Secondly, different from the LS one-class SVM, our method embeds the fuzzy information into the corresponding optimization problem, which improves the effectiveness. Computational comparisons of the three algorithms on both generalization performance and training time have been made on several benchmark data sets. The experimental results show that the proposed method acquires the comparable accuracy with the fuzzy one-class SVM and the comparable fast learning speed with the LS one-class SVM.

This paper is organized as follows. Section 2 briefly dwells on the fuzzy one-class SVM and the LS one-class SVM. Section 3 gives our LS fuzzy one-class SVM. Experiments are presented in Section 4. Section 5 concludes this paper.

2. Background Review

2.1. Fuzzy One-class Support Vector Machine

Given a training set $T=\{x_1,x_2,...,x_l\},\ x_i\in R^n$. The fuzzy one-class SVM [27] searches for a crisp hyperplane that separates the image of the target class from the origin:

$$f(x)=w^T\varphi(x)+b \tag{1}$$

where $w\in R^n, b\in R,$ and $\varphi(\cdot)$ is to map a data set χ to high dimensional feature space H. Its prime problem can be expressed as

$$\begin{aligned}&\min_{w,b,\xi_i}\frac{1}{2}\|w\|^2+b+C\sum_{i=1}^{l}\mu_i\xi_i\\&s.t.\ \ w^T\varphi(x_i)+b\geq 0-\xi_i,\ \xi_i\geq 0,\ \text{for}\ i=1,...,l.\end{aligned} \tag{2}$$

where $0\leq\mu_i\leq 1$ is the fuzzy membership, and it is the attitude of the corresponding sample x_i toward the target class. The minimization of the regularization term $\|w\|^2$ is equivalent to the maximization of the margin between the target class from the origin. $\xi=(\xi_1,\xi_2,...,\xi_l)$ is called the slack variable. The parameter $C>0$ is to balance the model complexity and the empirical rick. Using the Karush-Kuhn-Tucker (KKT) conditions, we can formulate the dual problem as

 Copyright ⓒ 2015 SERSC

$$\max_{\alpha} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j K(x_i, x_j)$$
$$s.t.\quad 0 \le \alpha_i \le \mu_i C, \tag{3}$$
$$\sum_{i=1}^{l}\alpha_i = 1.$$

where $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ is a kernel function. Once the solution $\alpha = (\alpha_i, ..., \alpha_l)$ to the problem (3) has been obtained, the decision function $g(x)$ for fuzzy one-class SVM can be expressed as follows.

$$g(x) = \text{sgn}(f(x)) = \text{sgn}(\sum_{i=1}^{l}\alpha_i K(x_i, x) + b). \tag{4}$$

2.2. Least Squares One-class Support Vector Machine

Inspired by one-class SVM, Y. Choi [26] proposed least squares one-class SVM. The LS one-class SVM can be derived through solving the following optimization problem.

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + b + \frac{C}{2}\sum_{i=1}^{l}\xi_i^2$$
$$s.t.\quad w^T\varphi(x_i) + b + \xi_i = 0, \text{ for } i = 1, ..., l. \tag{5}$$

By introducing Lagrangian multipliers and constructing a Lagrangian function, one gets the following set of linear equations.

$$\begin{pmatrix} 0 & e^T \\ e & K + \frac{1}{C}I \end{pmatrix}\begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \tag{6}$$

where $e = (1,...,1)^T$, $\mathbf{0} = (0, ..., 0)^T$, I_n denotes $n \times n$ identity matrix. Once the solutions α and b can be obtained from (6), the hyperplane obtained from (6) can be written as follows.

$$f(x) = \sum_{i=1}^{l}\alpha_i K(x_i, x) + b. \tag{7}$$

3. Least Squares Fuzzy One-class Support Vector Machine

In this section, utilizing the fuzzy information and the LS technology, we propose the fuzzy LS one-class SVM. Its optimization problem can be described as

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + b + \frac{1}{2}C\sum_{i=1}^{l}\mu_i\xi_i^2$$
$$s.t.\quad w^T\varphi(x_i) + b + \xi_i = 0, \quad \text{for } i = 1, ..., l. \tag{8}$$

In (8), the first term $\frac{1}{2}\|w\|^2$ in the objective function is to control the model complexity. The variable ξ_i represents the loss caused by a training sample x_i with respect to the hyperplane, *i.e.* $\xi_i = -b - w^T\varphi(x_i)$. The third term minimizes the sum of the loss variables. Unlike the fuzzy one-class SVM, we reformulate our method by using the

Copyright ⓒ 2015 SERSC

quadratic loss function and the equality constraint. Unlike the LS one-class SVM, we add the fuzzy information to each sample.

In order to solve problem (8), we construct the Lagrangian function

$$L = \frac{1}{2}\|w\|^2 + b + \frac{1}{2}C\sum_{i=1}^{l}\mu_i\xi_i^2 - \sum_{i=1}^{l}\alpha_i(w^T\varphi(x_i) + b + \xi_i) \tag{9}$$

where the $\alpha = (\alpha_1, ..., \alpha_l)$ is the Lagrangian multiplier vector. Differentiating L with respect to w, b, ξ_i, α_i and setting the results to zero, we obtain

$$\partial L / \partial w = w - \sum_{i=1}^{l}\alpha_i\varphi(x_i) = 0 \Rightarrow w = \sum_{i=1}^{l}\alpha_i\varphi(x_i), \tag{10}$$

$$\partial L / \partial b = 1 - \sum_{i=1}^{l}\alpha_i = 0 \Rightarrow \sum_{i=1}^{l}\alpha_i = 1, \tag{11}$$

$$\partial L / \partial \xi_i = C\mu_i\xi_i - \alpha_i = 0 \Rightarrow \xi_i = \frac{\alpha_i}{C\mu_i}, \tag{12}$$

$$\partial L / \partial \alpha_i = w^T\varphi(x_i) + \xi_i + b = 0 \tag{13}$$

Eliminating w and ξ_i through substitution in (13) yields

$$\sum_{j=1}^{l}\alpha_j\varphi(x_j)^T\varphi(x_i) + b + \alpha_i / (C\mu_i) = 0. \tag{14}$$

Utilizing the (11) and (14), we can obtain α and b by the following equations.

$$\begin{pmatrix} 0 & e^T \\ e & K + S \end{pmatrix}\begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \tag{15}$$

where $e = (1, ..., 1)^T$, $\mathbf{0} = (0, ..., 0)^T$. K is the kernel matrix, S is a diagonal matrix given by $S = diag\{\frac{1}{C\mu_1}, \frac{1}{C\mu_2}, ..., \frac{1}{C\mu_l}\}$. μ_i is the fuzzy membership which assigns to each sample. Utilizing the solution of (15), we present the optimal hyperplane of the LS fuzzy one-class SVM.

$$f(x) = \sum_{i=1}^{l}\alpha_i K(x_i, x) + b. \tag{16}$$

4. Experiments

To check the validity of the proposed LS fuzzy one-class SVM in dealing with the class imbalanced problem, we compare it with the fuzzy one-class SVM and the LS one-class SVM on several UCI benchmark data sets [30]. All the algorithms are implemented on MATLAB 2012a under Windows 7 running on a PC with system configuration Intel(R) Core(TM) i3 processor with 4 GB RAM. We utilize "qp.m" for fuzzy one-class SVM, "inv.m" for LS fuzzy one-class SVM and LS one-class SVM. We consider the Gaussian kernel function $K(x, y) = \exp(-p\|x - y\|^2)$ for all data sets in the experiments. There exist two parameters (p, C) need to be selected in the three algorithms. To search the optimal parameters, we derive the (p, C) from the sets $\{2^i \mid i = -10, -9, ..., 10\} \times \{10^i \mid i = -4, -3, ..., 5\}$ on each training data set by cross-validation approach. Unlike the fuzzy one-class SVM, our method and LS one-class SVM [26] have no decision functions. However, the hyperplane $f(x)$ itself represents

 Copyright ⓒ 2015 SERSC

the optimal hyperplane in a least squares sense. As the optimal hyperplane is a distance function, we discriminate a test sample x as positive (i.e., it belongs to the target class) if it satisfies $\varepsilon^{-} \leq f(x) \leq \varepsilon^{+}$, where $\varepsilon^{+} = \max_k f(x_k)$, $\varepsilon^{-} = \min_k f(x_k)$ with x_k being a train sample.

Table 1. Description of Nine Data Sets

Data set	Minority	Majority	feature	Data size
Wine	59	119	13	178
Glass	29	185	9	214
Vehicle	168	428	18	596
Breast Cancer	239	444	10	683
German	300	700	24	1000
Diabetes	268	500	8	768
Segment	323	1879	18	2202
Abalone	391	3786	7	4177
Fourclass	305	550	2	855

Table 2. Confusion Matrix

	Predicted positive	Predicted negative
Positive class	True Positive (TP)	False Negative (FN)
Negative class	False Positive (FP)	True Negative (TN)

We evaluate the performance of fuzzy one-class SVM, LS one-class SVM and LS fuzzy one-class SVM on nine benchmark data sets, which consist of Wine, Glass, Vehicle, Breast Cancer, German, Diabetes, Segment, Abalone, and Fourclass. In these data sets, Breast Cancer, German, Diabetes and Fourclass are two-class imbalanced data sets. The minority class of each data set is regarded as the positive class, and the majority class of each data set is regarded as the negative class. The others are multi-class data sets. For Wine, Glass, Vehicle, Segment and Abalone, we use the 1st, 6th, 3th, 1st and 7th class as the positive class, respectively. Table 1 shows the characteristics of these data sets. For each data set, the 5 cross-validation technique is used in all algorithms to search for the optimal parameters. To test the performance of the algorithms on imbalanced data sets, accuracy can not exactly reflect the effectiveness of the algorithm. Here the confusion matrix in Table 2 is utilized to give a more comprehensive and objective assessment. In this paper, we take sensitivity, specificity, G-mean, F_1 and AUC as the performance measures as follows.

(1) $sensitivity = \dfrac{TP}{TP + FN}$,

(2) $specificity = \dfrac{TN}{TN + FP}$,

(3) $G - mean = \sqrt{sensitivity \times specificity}$,

(4) $F_1 = \frac{2 \times recall \times precision}{recall + precision}$, where $recall = \frac{TP}{TP+FN}$, $precision = \frac{TP}{TP+FP}$.

Sensitivity is defined as the accuracy on the positive class and the specificity is the accuracy on the negative class. $G-mean$ evaluates positive accuracy and negative accuracy. The value of $G-mean$ is high when both sensitivity and specificity are high. F_1 combines recall and precision to measure effectiveness of the algorithm. A higher value of F_1 implies that the algorithm has better performance. Specially, the area under the curve of receiver operating characteristics defined as AUC can be used to measure the performance of the algorithm in a quantitative manner. The larger AUC, the better the algorithm performs. $AUC = 0.5$ means random forecast, while $AUC = 1$ implies perfect forecasts.

Next we give the criterion to associate the fuzzy membership to each sample [27, 31]. Let the training samples are $(x_1, ..., x_l)$. Denote the mean of target class as x_{mean}, and the radius of the target class is defined as $r_{target} = \max_i |x_i - x_{mean}|, i = 1, ..., l.$ The fuzzy membership μ_i is defined as $\mu_i = 1 - \frac{|x_i - x_{mean}|}{(r_{target} + \delta)}$, where $\delta > 0$ is used to avoid $\mu_i = 0$.

Tabel 3. The Results of the Classifiers on Benchmark Data Sets; I, II and III Stand for Fuzzy One-class SVM, LS One-class SVM and LS Fuzzy One-Class SVM, Respectively

Data set	Alg	sen	spe	F_1	$G-mean$	AUC	Time(s)	(p,C)
Wine	I	0.4500	0.5534	0.2687	0.3063	0.6379	0.0263	$(2^{-3},1)$
	II	0.3846	0.7545	0.2648	0.5387	0.6288	0.0001	$(2^{-2},10^5)$
	III	0.4000	0.7479	0.2759	0.5470	0.6315	0.0001	$(2^{-8},10^3)$
Glass	I	0.9000	0.9892	0.8571	0.9435	0.9715	0.0085	$(2^{-1},10^{-2})$
	II	0.8000	1.0000	0.8889	0.8944	0.9845	0.0001	$(2^4,10^3)$
	III	0.9000	0.9946	0.9000	0.9461	0.9941	0.0001	$(2^{-1},1)$
Vehicle	I	0.4464	1.0000	0.6173	0.6682	0.8901	0.1072	$(2^0,10^4)$
	II	0.4632	1.0000	0.6310	0.6806	0.8071	0.0001	$(2^{-1},10^3)$
	III	0.4821	1.0000	0.6506	0.6944	0.8965	0.0004	$(2^{-4},10^5)$
Breast Cancer	I	0.9125	0.9595	0.8538	0.9357	0.9623	1.1317	$(2^{-10},10^5)$
	II	0.9750	0.8964	0.7647	0.9349	0.9565	0.0006	$(2^{-6},1)$
	III	0.9625	0.9347	0.8280	0.9485	0.9716	0.0007	$(2^{-7},10)$
German	I	0.4600	0.6657	0.2421	0.5534	0.5156	0.7287	$(2^{-3},10)$
	II	0.2300	0.8343	0.1925	0.4380	0.5121	0.0001	$(2^0,10^5)$
	III	0.3000	0.8443	0.2510	0.5033	0.5204	0.0001	$(2^0,10^4)$
Diabetes	I	0.5889	0.6657	0.4157	0.7760	0.6324	0.5932	$(1,10^4)$
	II	0.4433	0.8700	0.4021	0.6140	0.6005	0.0012	$(2^{-1},1)$
	III	0.5444	0.8230	0.4336	0.6706	0.6265	0.0014	$(2^{-3},10^2)$
Segment	I	0.6019	1.0000	0.7514	0.7758	0.8748	1.4916	$(1,10^5)$
	II	0.4815	1.0000	0.6500	0.6939	0.8591	0.0013	$(2^2,1)$
	III	0.6512	0.9234	0.5833	0.7754	0.8803	0.0018	

 Copyright ⓒ 2015 SERSC

								$(2,10^5)$
Abalone	I	0.7143	1.0000	0.8333	0.8452	0.8906	0.0960	$(2^{-3},1)$
	II	0.5238	0.9973	0.5868	0.7228	0.8785	0.0001	$(2^6,10)$
	III	0.7619	0.9963	0.7191	0.8713	0.8931	0.0003	$(2^{-4},1)$
Fourclass	I	0.9118	0.9855	0.9163	0.9479	0.9022	0.6334	$(2^2,10^5)$
	II	0.8725	1.0000	0.9313	0.9341	0.8916	0.0015	$(2^4,10^3)$
	III	0.9510	0.9691	0.8981	0.9600	0.9491	0.0019	$(2^4,10)$

Table 3 reports the comparison results of the three algorithms on nine data sets. It includes $sensitivity$ (sen), $specificity$ (spe), F_1, $G-mean$, the AUC, the training time and the optimal parameters. We observe that the sensitivity values of our method are better than those of LS one-class SVM on all data sets except Breast Cancer data set, which demonstrates that the fuzzy information can improve the performance of the positive class. In addition, the values of $G-mean$ and AUC in our method are better than those of LS one-class SVM, which demonstrates that the overall performance is better. For F_1 value, our method performs better than LS one-class SVM except the Segment and Fourclass data sets. For fuzzy one-class SVM and LS fuzzy one-class SVM, it can be seen that for sensitivity index, our algorithm has the better values on five of nine data sets, and achieves the same sensitivity value on Glass data set. For specificity index, our algorithm owns the comparable values on most of the data sets. For F_1 index, our algorithm has better values on five of nine data sets. The performance of $G-mean$ shows that our algorithm has better values on six data sets, and almost has equal values on Segment data set. As for AUC, our algorithm performs better on seven data sets except Wine and Diabetes data sets. The training CPU time shown in table 3 indicates that the LS fuzzy one-class SVM and the LS one-class SVM are almost hundreds of times faster than the fuzzy one-class SVM, since the algorithms solve a system of linear equations instead of the complex QPP in the fuzzy one-class SVM.

5. Conclusion

A least squares version of fuzzy one-class SVM is proposed to deal with imbalanced problem in this paper. The fuzzy membership to each sample is considered in LS fuzzy one-class SVM. Moreover, we solve just a system of linear equations and this makes our algorithm faster. Finally, experimental results on UCI benchmark data sets demonstrate that our algorithm is promising. However, the sparsity in our LS fuzzy one-class SVM disappears. Our further work should be concerned with the sparsity of this algorithm.

Acknowledgment

The work is supported by national Natural Science Foundation of China Grant No.11171346 and Chinese Universities Scientific Fund No.2013YJ010.

References

[1] D. Lewis and J. Catlett, "Training text classifiers by uncertainty sampling", Proceedings of the7th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, **(1994)** Dublin, Ireland
[2] M. Kubat and S. Matwin, "Learning when negative examples abound", Proceedings of the Ninth European Conference on Machine Learning, **(1997)**, pp.146-153.

[3] P. Chan and S. Stolfo, "Toward scalable learning with nonuniform class and cost distributions: a case study in credit card fraud detection", Proceedings of Knowledge Discovery and Data Mining, **(1998)**, pp. 164-168.
[4] F. Provost, "Machine learning from imbalanced data sets 101", Proceedings of the AAAI'2000Workshop on Imbalanced Data Sets, **(2000)**, pp. 64-68.
[5] N. V. Chawla, N. Japkowicz and A. Kotcz, "Editorial: special issue on learning from imbalanced data sets", ACM SIKDD Explorations Newsletters vol. 6, no. 1, (**2004)**, pp. 1-6.
[6] S. Visa and A. Ralescu, "Issues in mining imbalanced data sets—a review paper", Proceedings of the16th midwest artificial intelligenceand cognitive science conference, **(2005)**, pp. 67-73.
[7] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: one-sided selection", Proceedings of the 14th International Conference on Machine Learning, **(1997)**, pp**.** 179-186.
[8] T. Jo and N. Japkowicz, "Class imbalances versus small disjuncts", SIGKDD Explorations, vol. 6, no. 1, **(2004),** pp. 40-49.
[9] R. Barandela, R.M. Valdovinos, J.S. Sanchez and F.J. Ferri, "The imbalanced training sample problem: Under or over sampling?" In Joint IAPR International Workshops on Structural, Syntactic, and Statistical Pattern Recognition, Lecture Notes in Computer Science, vol. 3138, **(2004)**, pp. 806-814.
[10] N. Chawla, K. Bowyer, L.Hall and W.Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling technique", Journal of Artificial Intelligence Research, vol. 16, **(2002)**, pp. 321-357.
[11] D. Margineantu and T.G. Dietterich, "Bootstrap methods for the cost-sensitive evaluation of classifiers", Proceeding of International Conference on Machine Learning, **(2000)**, pp. 583-590.
[12] Y. Sun, M. Kamela, A. Wongb and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data", Pattern Recognition, vol. 40, no. 12, **(2007)**, pp. 3358-3378.
[13] Z. Zhao, P. Zhong and Y. Zhao, "Learning SVM with weighted maximum margin criterion for classification of imbalanced data", Mathematical and Computer Modelling, vol. 54, **(2011)**, pp. 1093-1099.
[14] L. Ding, J. Watada, and L.C. Chew, "A SVM-RBF method for solving imbalanced data problem", ICIC Express Letters, vol. 4, no. 6(B), **(2010)**, pp. 2419-2424.
[15] L.M. Manevitz and M. Yousef. "One-class SVMs for document classification", Journal of Machine Learning Research, vol. 2, **(2001)**, pp. 139-154.
[16] L. Zhuang and H. Dai, "Parameter optimization of kernel-based one-class classifier on imbalance learning", Journal of Computers, vol. 1, no. 7, **(2006)**, pp. 32-40.
[17] B. Raskutti and A. Kowalczyk, "Extreme rebalancing for SVMs: a case study", SIGKDD Explorations, vol. 6, no. 1, **(2004),** pp. 60-69.
[18] Z. Zhao, P. Zhong and Y. Zhao, "Reduced least squaers one-class SVM in empirical feature space for imbalnced data", ICIC Express Letters, vol. 5, no. 11, **(2010)**, pp. 4115-4121.
[19] C. Cortes, V.N. Vapnik, "Support vector networks", Machine Learning, vol. 20, (**1995)**, pp. 273-297.
[20] W. X. Zhu and P. Zhong, "A new one-class SVM based on hidden information", Knowledge-Based Systems, vol. 60, **(2014)**, pp. 35-43.
[21] K. N. Wang and P. Zhong, "Robust non-convex least squares loss function for regression with outliers", vol. 71, **(2014)**, pp. 290-302.
[22] S. R. Sain, H. L. Gray and W. A. Woodward, "Outlier detection from a mixture distribution when raining data are unlabeled", Bulletin of the Seismological Society of America, vol. 89, no. 1, **(1999)**, pp. 294-304.
[23] I. W. Tsang, J. T. Kwok and S. Li, "Learning the kernel in mahalanobis one-class support vector machines", International Joint Conference on Neural Networks, (**2006)**, pp. 1169- 1175.
[24] B. Schökopf, J. Platt, J. Shawe-Taylor, A. Smola and R. Williamson, "Estimating the support of a high-dimensional distribution", Neural Computation, vol. 13, no. 7, **(2001)**, pp. 1443-1471.
[25] D. M. J. Tax and R. P. W. Duin, "Support vector domain description", Pattern Recognition Letters, vol. 20, **(1999)**, pp. 1191-1199.
[26] Y. Choi, "Least squares one-class support vector machine", Pattern Recognition Letters, vol. 30, **(2009)**, pp. 1236-1240.
[27] P. Hao, "Fuzzy one-class support vector machines", Fuzzy Sets and Systems, vol. 159, **(2008)**, pp. 2317-2336.
[28] Y. Forghani, H. Sadoghi Yazdi, S. Effati, "An extension to fuzzy support vector data description", Pattern Anal Applic, vol. 15, **(2012)**, pp. 237-247.
[29] T. Wang, J. Chen, Y. Zhou and H. Snoussi, "Online Least Squares One-Class Support Vector Machines-Based Abnormal Visual Event Detection", Sensors, vol. 13, **(2013)**, pp. 17130-17155.
[30] C. L. Blake and C. J. Merz, "UCI Repository for Machine Learning Databases", Department of Information and Computer Sciences, University of California, Irvine, http://www.ics.uci.edu/~mlearn/MLRepository.html, **(1998)**
[31] C. Lin and S. Wang, "Fuzzy support vector machines", IEEE Trans. Neural Networks, vol. 13, no. 2, **(2002)**, pp. 464-471.

 Copyright ⓒ 2015 SERSC

Authors

Jingjing Zhang, she has received her B.S. degree in College of Mathematics and Information Science from Shaanxi Normal University, 2010, and received Master's degree in College of Science from China Agricultural University, China, in 2012. Currently, she is pursuing the Ph.D. degree at the College of Science of China Agricultural University. Her research interests include optimization methods, machine learning and data mining.

Kuaini Wang, she has received the B.S. degree in information and computation science from Jilin University, China, 2005, the M.S. degree in mathematics from China Agricultural University, 2009. She is currently pursuing the Ph.D. degree in operations research and management science at China Agricultural University. Her research interest includes machine learning and support vector machines.

Wenxin Zhu, he received the B.S. degree in mathematic from Sichuan Normal University in 2004. He received the MS degree in applied mathematics teacher of Tianjin Agricultural University. His current interests are in the area of pattern recognition, optimization method and machine learning. He has published several technical papers in these areas.

Ping Zhong, she has received the B.S degree and Master degree in mathematics in 1994 and 1997 both from Qufu Normal University, China, and her Ph. D degree in 2002 from China Agricultural University. Now she is a professor of China Agricultural University. She has published more than 40 papers cited by SCI and EI. Her research interest includes optimization theory, machine learning and support vector machines.

Copyright ⓒ 2015 SERSC