Nonlinear Dyn (2010) 61: 303–310
DOI 10.1007/s11071-009-9650-5

ORIGINAL PAPER

Anomaly detection combining one-class SVMs and particle swarm optimization algorithms

Jiang Tian · Hong Gu

Received: 14 November 2009 / Accepted: 20 December 2009 / Published online: 8 January 2010

Abstract Anomalies are patterns in data that do not conform to a well-defined notion of normal behavior. One-class Support Vector Machines calculate a hyperplane in the feature space to distinguish anomalies, but the false positive rate is always high and parameter selection is a key issue. So, we propose a novel one-class framework for detecting anomalies, which takes the advantages of both boundary movement strategy and the effectiveness of evaluation algorithm on parameters optimization. First, we search the parameters by using a particle swarm optimization algorithm. Each particle suggests a group of parameters, the area under receiver operating characteristic curve is chosen as the fitness of the object function. Second, we improve the original decision function with a boundary movement. After the threshold has been adjusted, the final detection function will bring about a high detection rate with a lower false positive rate. Experimental results on UCI data sets show that the proposed method can achieve better performance than other one class learning schemes.

Keywords Outlier detection · Particle swarm optimization · Support vector machine · Anomaly detection · One-class classification

J. Tian (✉) · H. Gu
School of Electronic and Information Engineering, Dalian University of Technology, Dalian, China
e-mail: tianjiang@gmail.com

1 Introduction

Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior. These nonconforming patterns are often referred to as anomalies, outliers, discordant observations, exceptions, aberrations, surprises, peculiarities, or contaminants in different application domains. With the development of information technology, vast quantities of data are captured and stored. The capacity, dimensions, and complexity of database have grown rapidly, but usually the real data set could not be used directly for data mining due to ignorance, human errors, rounding errors, transcription factor, instrument malfunction, and biases. Anomaly detection is used to find the objects that do not comply with the general behavior of the data and then lead to potentially useful information. Anomalies can be translated to significant or critical actionable information. Anomaly detection can be applied to many fields, such as credit card fraud detection, security systems, medical diagnosis, network intrusion detection, and information recovery [1, 2].

One-class classification based anomaly detection techniques assume that all training instances have only one-class label. Such techniques learn a discriminative boundary around the normal instances using a one-class classification algorithm. Any test instance that does not fall within the learned boundary is declared as anomalies. Support Vector Machines (SVMs) have been applied to anomaly detection in the one-class setting, e.g., one-class SVMs [3, 4]. A variant of the basic technique [5, 6] finds the smallest hypersphere in the kernel space, which contains all training instances, and then determines which side of that hypersphere a test instance lies. If a test instance lies outside the hypersphere, it is declared to be anomalous. Many SVM-based techniques have been proposed for anomaly detection in musical signal data [7], fault diagnosis in machinery [8], novelty detection in power generation plants [9], seizure analysis from intracranial electroencephalogram [10], and intrusion detection [11–13].

In practice, availability of labeled data for training and validation of models used by anomaly detection techniques is a major issue, there are only few labeled anomalies in the data sets. The unsupervised SVMs are promising in detecting new anomalies, however, they usually gain high detection rate but also high false positive rate, also. With regard to the applications sensitive to the false positives, keeping the resultant false positive rate (FPR) under a maximal user tolerance is usually a concern. In addition, proper parameters are important to the learning results and generalization ability of the classifier. Our research aims to achieve a true positive rate (TPR) with a low FPR, so we propose a novel method which utilizes tactics of parameter selection and boundary movement. A new one-class framework for anomaly detection was provided based on an One Class SVM (OCSVM). The parameters selection was considered as a compound optimization problem in this study. In our research, the area under the Receiver Operating Characteristics (ROC) curve [14] was chosen as evaluating metric of parameters selection. The area under the ROC curve (AUC) actually describes the probability that a randomly chosen positive example is correctly ranked with greater suspicion than a randomly chosen negative example [15] is. After the objective function of optimization problem had been set, an PSO algorithm was utilized to search the global ideal SVM parameters. Corresponding to the optimal AUC, the threshold of decision function was adjusted and the detection model would bring about the best combination of TPR and FPR was obtained. To the best of our knowledge, we are the first to carry out anomaly detection combining One-Class SVMs and PSO techniques, which aims to provide a robust and effective method to detection anomalies. The implemented software was evaluated on benchmark experiments, results showed its effectiveness.

2 One class support vector machine

Support Vector Machine is a new method to deal with the highly nonlinear classification and regression problems, which is based on statistic learning theory [16]. Scholkopf et al. [4] used the max margin theory of SVM to calculate the hyperplane which distinguished anomalies from others. The hyperplane has maximal margin to the origin and a pre-specified fraction of the training examples are laid beyond it. Training examples that lay under the hyperplane are detected as anomalies. OCSVM was proposed for estimating the support of a data distribution instead of the full density. As shown in Fig. 1, an OCSVM calculates a hyperplane in the feature space so that a prespecified fraction of the training example will lie beyond that hyperplane, while the hyperplane has maximal margin to the origin [17]. Given l samples, we solve the following quadratic problem [4]:

$$\begin{aligned} \min \quad & \frac{1}{2}\|w\|^2 + \frac{1}{\nu\ell}\sum \xi_i - \rho, \\ \text{s.t.} \quad & w \cdot \phi(x_i) \geq \rho - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, \ldots, l. \end{aligned} \tag{1}$$

Since nonzero slack variables ξ_i are penalized in the objective function, if w and ρ solve this problem, then the decision function will be positive for most examples x_i contained in the training set, while the regularization term $\|w\|$ will still be small. Here, $\nu \in (0, 1]$ is a parameter controlling this actual tradeoff. The dual

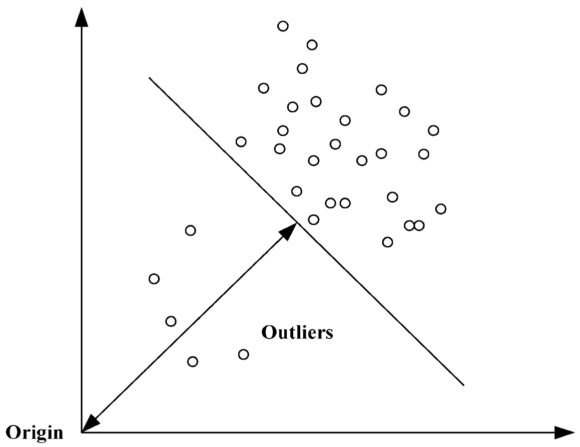


Fig. 1 One-Class SVM constructs an N-dimensional hyperplane that optimally separating the data into two categories

problem is

$$\begin{aligned} \min \quad & \frac{1}{2}\sum \alpha_i \alpha_j K(x_i, x_j), \\ \text{s.t.} \quad & 0 \le \alpha_i \le 1/(\nu\ell), \\ & \sum \alpha_i = 1, \end{aligned} \tag{2}$$

where α is the Lagrangian parameter, the point x_i with the corresponding $\alpha_i > 0$ is called a support vector (SV), K is the positive definite kernel function satisfying $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. After deriving the dual problem, the decision function can be shown to have a SV explanation:

$$f(x) = \operatorname{sgn}\left(\sum \alpha_i K(x_i, x) - \rho\right). \tag{3}$$

The offset ρ can be recovered by exploiting that for any α_i which is not at the upper or lower bound, the corresponding pattern x_i satisfies

$$\rho = \left(w \cdot \phi(x_i)\right) = \sum \alpha_j k(x_j, x_i). \tag{4}$$

When we apply OCSVM algorithm to anomaly detection applications, training examples whose decision outputs are greater than zero are considered as normal, while the others are detected as anomalies. From (3), the decision outcome is affected by the support vectors and the parameter ρ. Using a Gaussian RBF Kernel, which is commonly used for classification problems, the prediction function of OCSVM becomes

$$f(x) = \operatorname{sgn}\left(\sum \alpha_i \exp\left(-\frac{\| x_i - x \|^2}{2\gamma^2}\right) - \rho\right), \tag{5}$$

where γ is the kernel width parameter of a Gaussian RBF kernel.

From (1) and (5), the parameters of ν and kernel width γ are key issues for the final decision function. Proper parameters are important to the learning results and generalization ability. Although OCSVMs are promising techniques in detecting anomalies, they usually have high detection rate but high false positive rate, also. Our research aims to find a method to keep high detection rate while lowering the false positive rate, so we adopt two tactics for this problem. First, we choose PSO to select parameters. One particle represents two important parameters: ν and γ. The object of the fitness function is the AUC. Second, we apply a boundary movement strategy on standard OCSVM. The offset ρ depicts the location of the hyperplane, so a subjective movement on this variable can improve the accuracy of the classifier. The amount of movement is calculated following the parameter selection process, as it will lead to a higher performance. The class decision boundary becomes

$$f(x) = \operatorname{sgn}\left(\sum \alpha_i \exp\left(-\frac{\| x_i - x \|^2}{2\gamma^2}\right) - \rho + \Delta\rho\right). \tag{6}$$

The variable $\Delta\rho$ represents the threshold which corresponds to the optimized detection performance. In this study, the parameters of the detection model are optimized by PSO algorithm. The final decision function will achieve better performance.

3 Method

To achieve the best detection model, we will take the following steps. First, the training examples are trained using an OCSVM classifier to calculate the AUC. Second, PSO algorithm with perturbations will not halt updating the particles' velocities and positions until the stop criterion is satisfied. In this step, the AUC is used as the fitness function values for updating the particles, and the optimized parameters are achieved. Finally, after adjusting the offset to make a boundary movement of (6), we obtain the ideal decision function with high TPR and lower FPR.

3.1 One-class classifier and ROC analysis

OCSVM is an unsupervised learning algorithm, which utilizes the hyperplane in feature space to classify the training examples. In anomaly detection applications, the decision function is usually trained on unsupervised examples. Accuracy is not a sufficient metric, because the classifiers can obtain higher predictive accuracies without considering the anomaly class. This apparent good performance can be identified as meaningless. Also, the costs of misclassifications are commonly not equal. For example, high accuracy is occasionally unmeaningful such as fraud detection and medical diagnosis. So we choose ROC analysis to balance TPR and FPR.

The ROC curve has been introduced by the signal processing community for evaluating the capability of a human operator to distinguish informative radar signal from noise [18]. ROC curve is a two-dimensional measure of classification performance, and it can be understood as a plot of the probability of classifying correctly the positive examples versus the rate of incorrectly classifying true negative examples. In this sense, one can interpret this curve as a comparison of the classifier across the entire range of class distributions and error costs. The best possible classifier would yield a point in the upper left corner or coordinate (0, 1) of the ROC curve, representing all true positives are found and no false positives are appeared. In practical applications of anomaly detection, it was expected that the TPR is as high as possible while the FPR is as low as possible. The most frequently used performance measure extracted for the ROC curve is the value of the area under the curve (AUC) [19, 20], which is a reasonable performance statistic for classifier assuming no knowledge of the true ratio of misclassification costs.

Gaussian kernel is used in our algorithm, the selection of parameters (ν, γ) is indeed an optimal searching process. Each particle indicates a group of SVM parameters, and the AUC is used as the fitness function of PSO. To calculate the AUC, a tow-phase technique based on OCSVM was adopted. All unsupervised training examples are used to determine the training model, then the predict outputs on the supervised training examples are calculated to calculate AUC. This strategy is corresponding to real applications, where large number of unlabeled samples and small number of labeled samples are exist. This is similar to a semi-supervised recognition or detection task, which aims to define a boundary of normalcy [1, 21]. AUC is used to evaluate objective function of PSO, the following steps show how to calculate the AUC.

Step 1: initialize the OCSVM model, load the training examples including both the unsupervised and supervised patterns.

Step 2: determine the origin decision function $f(x)$ (3), the unsupervised patterns are trained using an OCSVM, which lead to a density estimation of the data.

Step 3: utilize $f(x)$ to predict the outputs of the supervised training examples. In this step, the predictions are stored for further operation. Because the ROC analysis technique needs generated scores from a classifier rather than just a class label [20], we select the distance between a point and the hyperplane as the predicting score.

Step 4: calculate the AUC of OCSVM. In this step, we use the algorithm described in [20], which calculate the AUC from the classification scores and original target labels.

3.2 Model selection

Kennedy and Eberhart proposed the PSO method in 1996 [22], which was derived from the simulation of the birds in finding foods. It is one kind of global search algorithm, which searches for the optimal value by sharing historical an d social information among the particles. This algorithm has many merits such as fast convergence rate, simple concept and so on. PSO techniques had been applied for solving model selection problems of SVM [23–25], where each group of SVM parameters was trained as a particle.

The fitness function of PSO is used to evaluate the proposed solution to a given problem. PSO regards each individual in the swarm as a particle, which is a candidate solution. This particle can fly in a D-dimensional search space at a certain velocity, and the velocity of a particle is regulated by the flight experiences of the particle itself and those of others. PSO is initialized with a group of random particles and then search for optimal fitness by updating generations. The particles iteratively evaluate the fitness of the candidate solutions and remember the location where they had their best success. During iterations, each particle is updated by the following two values, the best fitness it has achieved so far which is called *localbest* and the best fitness the whole swarm has achieved which is called *globalbest*. Each particle makes this information available to their neighbors. They are also able to see where their neighbors have had success. The particle updates its velocity and positions with following formulas:

$$\begin{aligned} v_{ik+1} = {} & c_0 v_{ik} + c_1 \tau_1 (localbest_{ik} - x_{ik}) \\ & + c_2 \tau_2 (globalbest_k - x_{ik}), \end{aligned} \tag{7}$$

$$x_{ik+1} = x_{ik} + v_{ik+1}, \tag{8}$$

where c_0 is an inertial weight which denotes the influence of the previous velocity of a particle upon its current velocity. c_1 and c_2 are acceleration constants, τ_1 and τ_2 are the random numbers between $[0, 1]$. v_{ik} is the velocity of i-th particle in iteration k, and x_{ik} is the position, which represent the parameters of (ν, γ). The optimized parameters of OCSVM are then used to train the detection model.

3.3 Anomaly detection algorithm

Combining OCSVM and ROC analysis, the proposed method includes a series of different procedures (Fig. 2). The detailed procedures are described as follows:

Step 1: system initialization. Set the particle numbers 25, inertial weight c_0, acceleration constants c_1 and c_2, iteration numbers 100.

Step 2: update the positions, begin loop operations.

Step 2.1: calculate the fitness of all particles. In this step, we need to calculate new velocity by (7) and new position by (8). Then map the positions of particle into SVM parameters (ν,γ), and update parameters of the OCSVM. We train the one-class model to calculate AUC, which is used as the fitness value $f(x_{ik+1})$ of PSO.

Step 2.2: update particle historical best position $localbest_{ik}$. Compare the current fitness of each particle with its own historical best position $localbest_{ik}$. If its own historical best position $localbest_{ik}$ is smaller, then it is replaced by the current fitness.

Step 2.3: update swarm historical best position $globalbest_k$. Compare the best current position of all particles with the historical best position of the whole swarm $globalbest_k$. If the historical best position $globalbest_k$ is smaller, replace it with the best current position of all particles.

Step 2.4: judge if the exit condition is satisfactory. If AUC is larger than the critical value, exit loop operations. This critical value is set high enough to gain optimized parameters. Otherwise, return to step 2.1.

Step 3: adjust the offset corresponding to the best fitness and best positions. Calculate the best AUC and the largest $TPR\sqrt{1-FPR}$. Find the critical output and adjust the offset of decision function.

Step 4: construct the detection model. Map the positions into parameters and construct the final detect model.

Step 5: make predictions on the test patterns.

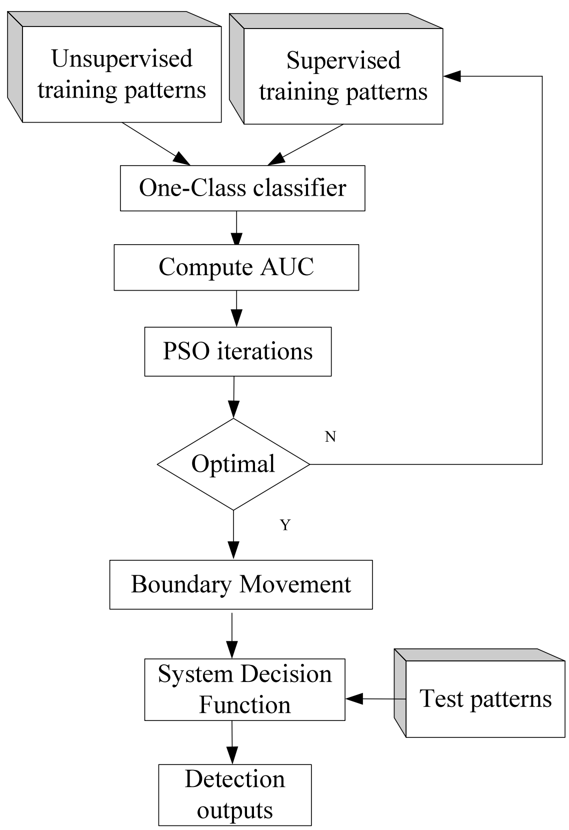


Fig. 2 The detailed procedure of proposed method. To construct and validate the one-class model, the training data set is separated into two parts

4 Experiments

4.1 Data sets

We validated our method on five different data sets from the UCI Repository for machine learning. These data sets belong to conventional classification problems with multiple classes. In order to make them suitable for anomaly detection applications, the data sets were converted to binary class problems. We selected one of the classes to depict the anomalies, while the remaining classes were gathered together to form the normal class. This is similar to the procedure used by Cao et al. [26] and Oliveira et al. [27]. Table 1 lists the data sets, along with some fitting statistics, such as the number of positive and negative examples, their ratios. Because the information about the UCI data sets

Table 1 Description of four data sets for anomaly detection experiments

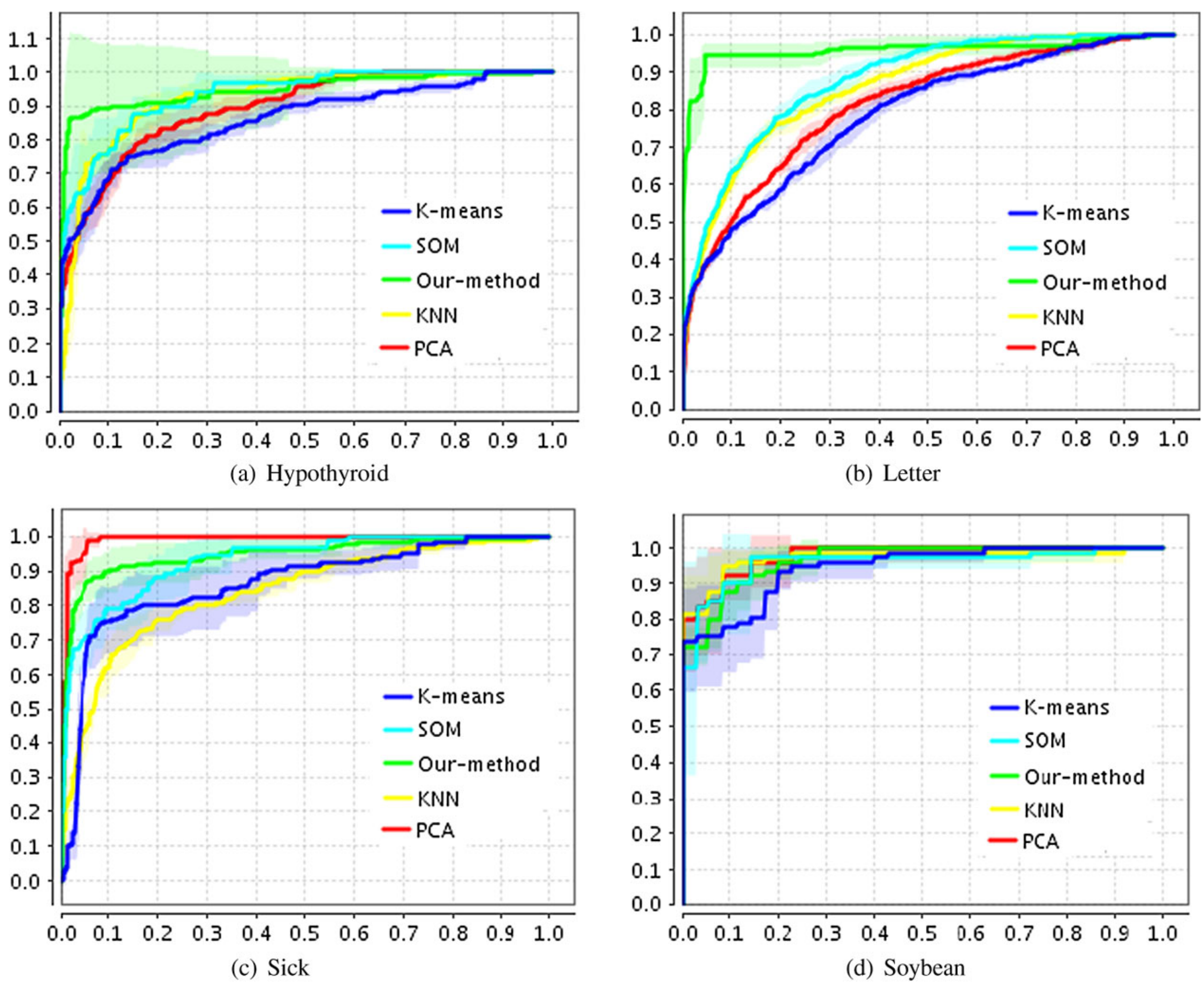
Data set	Training samples		Test samples	
	Normal	Anomalies	Normal	Anomalies
Hypothyroid	1820/480	60	1181	231
Letter	10236/2547	124	6483	610
Sick	1854/480	46	1210	185
Soybean	334/82	15	176	76

is public, we do not report it here. The names of the anomaly class are “Sick,” “Positive,” “H,” “Brown-spot.” In our method, the training data set is separated into two parts, one is for constructing a one-class model, the other is for validating the model.

4.2 Results and discussion

We compared our method with four other anomaly detection methods, K-NN (K-Nearest Neighbor) [28], K-Means [29], SOM (Self-Organizing Map) [30], and PCA (Principal Component Analysis) [31]. We used a 10-fold cross validation strategy to estimate the optimizing parameters of the other four one-class classifiers. These methods were compared using ROC curves and respective AUCs; all ROC curves were constructed using the test data sets. For the experiments, RapidMiner was used [32].

Figure 3 depicts the ROC curves obtained by our method and other techniques. For these data sets, the performance of our method is satisfactory and stable. For the Hypothyroid data set, PCA and SOM obtain similar results, their ROC curves show better performance than that of K-means, while KNN obtains a second-best ROC curve. For the Letter data set, KNN, SOM and our method showed better performance than PCA and K-means. For the Sick data set, PCA and our method show preferable results, while the other methods obtain similar results. For the Soybean data set, the ROC curves interlace with each other, so it



(a) Hypothyroid (b) Letter (c) Sick (d) Soybean

Fig. 3 ROC curves for different data sets. Five methods are compared: K-means, SOM, KNN, PCA, and our method, which are described using different *colored lines*

Table 2 Comparisons of area under ROC curves

AUC	KNN	K-means	SOM	PCA	Our-method
Hypothyroid	0.924	0.861	0.881	0.898	0.949
Letter	0.865	0.788	0.879	0.811	0.889
Sick	0.843	0.852	0.895	0.923	0.952
Soybean	0.889	0.909	0.852	0.869	0.925

is difficult to evaluate the performance only according to the observation. We have also calculated the AUCs for the ROC curves to compare the results (Table 2). Note there is a remarkable improvement in the detection performance. The average AUC of our method is 0.929, which is the highest among these five methods. The average AUCs of KNN, K-means, SOM and PCA are 0.88, 0.853, 0.877, 0.875 respectively.

For our method, it can be inferred that the anomalies can be effectively detected by the optimizing classifier, which is constructed using parameter selection and boundary movement strategies. The results suggest that our method improves the robustness of the overall decision. While the other methods show different performances on different data sets. The KNN technique works well on the Hypothyroid data set, while shows difficulty on the Sick data set. The other three techniques have the similar characteristic; they achieve better or worse results on different data sets. Results indicate that the proposed framework can find the optimized parameters of OCSVM, and the final decision function modified by a boundary movement can improve the performance. Moreover, our method is robust to detect anomalies in different data sets, as means it is more data independent.

It is important to mention that we have also validated our method on data sets using other classes as the anomaly class. The results were similar to those reported in this section. One limitation of our method is that we assume there exist some labeled anomalies in training data sets, these anomalies are used to calculate the AUCs, while in some applications, the labeled anomalies are expensive to acquire, or even nonentity. We aim to detect anomalies when the data sets contain many labeled normal patterns and a small number of labeled anomalies. Not withstanding its limitation, this study does suggest the proposed method can improve the performance of classical OCSVM. Results confirm that our methods can achieve a more robust detection performance than other data description techniques.

5 Conclusion

Anomaly detection is an important area of data mining. Based only on examples of a normal profile, one-class classification techniques are able to induce a classifier that is capable of detecting anomalies. In this study, a novel detection framework was proposed which combined the idea of unsupervised learning method and the supervised strategy. Instead of calculating the accuracy, ROC analysis was utilized to evaluate the detection performance. AUC was selected as the fitness of each particle, PSO algorithm was executed for global optimal parameters of SVM. Best combination of TPR with FPR was achieved after adjusting the offset of the detection function, a boundary movement approach could help to improve the performance. The effectiveness of our method for anomaly detection was demonstrated on four benchmark data sets, results showed satisfactory performance.

References

1. Hodge, V., Austin, J.: A survey of outlier detection methodologies. Artif. Intell. Rev. **22**(2), 85–126 (2004)
2. Filzmoser, P., Maronna, R., Werner, M.: Outlier identification in high dimensions. Comput. Stat. Data Anal. **52**(3), 1694–1711 (2008)
3. Scholkopf, B., Williamson, R.C., Smola, A.J., Shawe-Taylor, J., Platt, J.: Support vector method for novelty detection. Adv. Neural Inf. Process. Syst. **12**, 582–588 (2000)
4. Scholkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural Comput. **13**(7), 1443–1471 (2001)
5. Tax, D.M.J., Duin, R.P.W.: Support vector domain description. Pattern Recogn. Lett. **20**(11–13), 1191–1199 (1999)
6. Tax, D.M.J., Duin, R.P.W.: Support vector data description. Mach. Learn. **54**(1), 45–66 (2004)
7. Davy, M., Desobry, F., Gretton, A., Doncarli, C.: An online support vector machine for abnormal events detection. Signal Process. **86**(8), 2009–2025 (2006)
8. Zhang, Y., Liu, X.D., Xie, F.D., Li, K.Q.: Fault classifier of rotating machinery based on weighted support vector data description. Expert Syst. Appl. **36**(4), 7928–7932 (2009)
9. King, S.P., King, D.M., Astley, K., Tarassenko, L., Hayton, P., Utete, S.: The use of novelty detection techniques for monitoring high-integrity plant. In: Proceedings of the 2002 International Conference on Control Applications, Cancun, Mexico, vol. 1, pp. 221–226 (2002)
10. Gardner, A.B., Krieger, A.M., Vachtsevanos, G., Litt, B.: One-class novelty detection for seizure analysis from intracranial eeg. J. Mach. Learn. Res. **7**, 1025–1044 (2006)
11. Eskin, E., Arnold, A., Prerau, M., Portnoy, L., Stolfo, S.: A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data. In: Data Mining for Security Applications, vol. 19 (2002)

12. Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., Srivastava, J.: A comparative study of anomaly detection schemes in network intrusion detection. In: Proceedings of Third SIAM Conference on Data Mining, San Francisco, vol. 3 (2003)
13. Giacinto, G., Perdisci, R., Del Rio, M., Roli, F.: Intrusion detection in computer networks by a modular ensemble of one-class classifiers. Inf. Fusion **9**(1), 69–82 (2008)
14. Bradley, A.P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn. **30**(7), 1145–1159 (1997)
15. Hassan, M.R., Hossain, M.M., Bailey, J., Ramamohanarao, K.: Improving k-nearest neighbour classification with distance functions based on receiver operating characteristics. In: Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases—Part I, Antwerp, Belgium. Lecture Notes in Artificial Intelligence, vol. 5211, pp. 489–504. Springer, Heidelberg (2008)
16. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, Berlin (2000)
17. Muller, K.R., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B.: An introduction to kernel-based learning algorithms. IEEE Trans. Neural Netw. **12**(2), 181 (2001)
18. Egan, J.P.: Signal Detection Theory and ROC Analysis. Academic Press, New York (1975)
19. Fawcett, T.: ROC graphs: Notes and practical considerations for data mining researchers. Tech. Rep. (2004)
20. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)
21. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. In: ACM Computing Surveys (2009)
22. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Neural Networks, 1995. Proceedings. IEEE International Conference on, Piscataway, NJ, vol. 4 (1995)
23. Bazi, Y., Melgani, F.: Semisupervised pso-svm regression for biophysical parameter estimation. IEEE Trans. Geosci. Remote Sens. **45**(6), 1887–1895 (2007)
24. Melgani, F., Bazi, Y.: Classification of electrocardiogram signals with support vector machines and particle swarm optimization. IEEE Trans. Information Technol. Biomed. **12**(5), 667–677 (2008)
25. Peng, T., Zuo, W.L., He, F.L.: Svm based adaptive learning method for text classification from positive and unlabeled documents. Knowl. Inf. Syst. **16**(3), 281–301 (2008)
26. Cao, L.J., Lee, H.P., Chong, W.K.: Modified support vector novelty detector using training data with outliers. Pattern Recogn. Lett. **24**(14), 2479–2487 (2003)
27. Oliveira, A.L.I., Costa, F.R.G., Filho, C.O.S.: Novelty detection with constructive probabilistic neural networks. Neurocomputing **71**(4–6), 1046–1053 (2008)
28. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 3rd edn. Academic Press, San Diego (2006)
29. Bishop, C.: Neural Networks for Pattern Recognition. Oxford University Press, Oxford (1995)
30. Kohonen, T.: The self-organizing map. Neurocomputing **21**(1–3), 1–6 (1998)
31. Jolliffe, I.: Principal Component Analysis. Springer, New York (2002)
32. Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T.: Yale: Rapid prototyping for complex data mining tasks. In: International Conference on Knowledge Discovery and Data Mining, pp. 935–940. Association for Computing Machinery, Philadelphia (2006)