Anomaly Detection in Time Series Data Using Support Vector Machines

Anomaly Detection in Time Series Data Using Support Vector Machines

Umaporn Yokkampon[1], Sakmongkon Chumkamon[1], Abbe Mowshowitz[2] , Eiji Hayashi[1]
[1]*Department of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan*
[2]*Department of Computer Science, The City College of New York, 160 Convent Avenue, New York, NY 10031, USA*
E-mail: may@mmcs.mse.kyutech.ac.jp, m-san@mmcs.mse.kyutech.ac.jp, amowshowitz@ccny.cuny.edu, haya@mse.kyutech.ac.jp
www.kyutech.ac.jp

Abstract

Analysis of large data sets is increasingly important in business and scientific research. One of the challenges in such analysis stems from uncertainty in data, which can produce anomalous results. In this paper, we propose a method of anomaly detection in time series data using a Support Vector Machine. Three different kernels of the Support Vector Machine are analyzed to predict anomalies in the UCR public data set. Comparison of the three kernels shows that the defined parameter values of the RBF kernel are critical for improving the validity and accuracy in anomaly detection. Our results show that the RBF kernel of the Support Vector Machine can be used to advantage in detecting anomalies.

Keywords: Anomaly detection, Support Vector Machine, Data mining, Factory automation.

1. Introduction

Research on anomaly detection is of great interest in machine learning and data mining. Detecting anomalies or finding outliers involves identifying abnormal or inconsistent patterns in a dataset. Abnormal data often results from unauthorized activity. Credit card fraud offers a well known example. Transactions with a stolen or fake credit card can produce suspicious data. A fake card can be made by copying information from an authorized card and using it to create a new unauthorized one. Data such as personal identifying information may be obtained through phishing or from employees who work in credit card companies [1]. Another source of abnormal data may derive from unauthorized intrusions in networks. Abnormal traffic or user actions are common signs of intrusions, which may occasion breaches of sensitive or confidential data. Intrusions may also cause sensor networks to generate erroneous data. When a sensor malfunctions, it is unable to capture data correctly and thus may produce anomalies. Abnormal changes in data sources may also result in anomalies [2]. Anomaly detection typically uses data mining and machine learning techniques to detect abnormal activities in systems. Over the past decade, many anomaly detection techniques have been developed, including Support Vector Machines (SVM), a supervised machine learning algorithm that can be used for classification or to solve regression problems. In practice, the SVM algorithm is applied with the kernel that transforms an input data space into the required form. Kernel function and kernel parameters affect the performance of SVM. The selection quality of SVM parameters and kernel functions has an effect on learning and generation performance. Appropriate kernel function and associated parameters should be selected to obtain optimal

classification performance. When an appropriate kernel function and parameters are selected, the prediction error of SVM can be minimized.

This paper reports on application of the support vector machine method to eight real world time series data sets to detect anomalies using three different kernels for analysis and prediction. In addition, SVM kernels are compared for effectiveness based on AUC, Precision, Recall, and F1-Score criteria.

2. Related Work

Anomaly detection is widely used in many fields, and various methods have been proposed over the years.

In 2003, Ma et al. [3] used one class SVMs for prediction which require a set of vectors as input instead of a time series. They convert the time series into a phase-space using a time-delay embedding process, i.e., by creating overlapping subsequences from a given long sequence. These vectors are projected into an orthogonal subspace that acts as a high pass filter to filter out the low frequency components and allow only high frequency ones (anomalies).

In 2005, Kim and Cha [4] tested the effectiveness of SVM in detecting masquerade activities. Their experiments showed that their model could detect this type of attack with an accuracy of 80.1%., thus empirically demonstrating that SVM offers an effective method for masquerade detection.

In 2008, Sugumaran et al. [5] developed fault diagnostics of roller bearings using a neighborhood score multiclass SVM in EDM machining. The roller bearing is one of the most widely used elements in a rotary machine. RBF is used as a kernel function. This research used a kernel based neighborhood score multiclass SVM for classification and a decision tree for addressing the future selection process. The study of a multiclass SVM showed its effectiveness in diagnosing the fault conditions of the bearing.

In 2012, Caydas and Ekici [6] applied SVM to develop prediction models for surface roughness in the turning process of AISI 304 austenitic stainless steel. The relevant parameters are cutting speed, feed rate and depth of cut using RBF as a kernel function. Three different SVM models were developed, namely, LS-SVM, spider SVM and SVM-KM. Spider SVM offered the best prediction model for surface roughness.

In 2015, Yu et al. [7] proposed a prediction model of bus arrival time based on SVM and a forgetting factor. The actual time of bus arrival at each time point is predicted by taking account of the time, weather and certain historical data as input vectors. A k-Nearest Neighbor Model for Multiple-Time-Step Prediction was introduced to predict short-term traffic conditions. Moreover, the Grubbs' test method was applied to remove outliers from the input data.

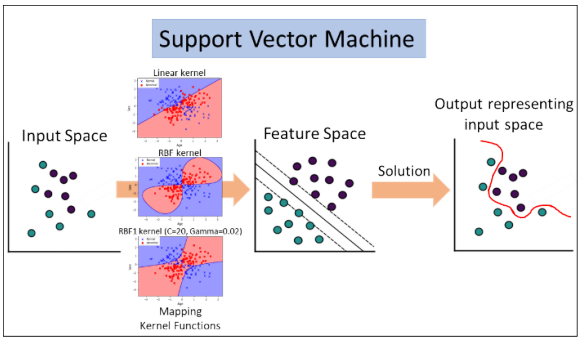

Fig. 1. Structure of anomaly detection in time series data used SVM. We used eight time series data sets processed by SVM, and three different kernels based on AUC, Precision, Recall and F1-Score criteria.

3. Support Vector Machine Algorithm

The Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for classification as well as regression problems. The SVM algorithm's goal is to create the best line or decision boundary that can decompose an n-dimensional space into sets supporting categorization of new data points. Hyperplanes define the boundaries in this space.

For a given dataset x with a number i of training data, SVM finds the maximum margin hyperplane separating different classes of data [8]:

$$x = (x_i, y_i),\ x_i \in \mathbb{R}^p,\ y_i \in \{-1,1\},\ \forall\ i = 1,2,...,N \quad (1)$$

where x_i is the p-dimensional input vector and y_i is the output value (1 or −1). A decision vector separating two classes is given by:

$$w^T \cdot x + b = 0 \quad (2)$$

where w^T is the optimal weighting vector and b is the bias. For linearly separable training data, margins are defined as:

$$w^T \cdot x + b = 1 \text{ and } w^T \cdot x + b = -1 \quad (3)$$

The distance between the margins is given by $2/\|w^T\|$. Hence, the objective function is to minimize $\|w^T\|$. In practice, it is not easy to linearly decompose the training dataset. Let C be the regularization parameter that defines the separation of two classes and the error when using a training dataset. The hyperplane is determined by minimizing:

$$C\sum_{i=1}^{N}\varepsilon_i + \frac{1}{2}\|w\|^2 \quad (4)$$

with constraints $t_i y(x_i) \geq 1-\varepsilon_i,\ i=1,...,N$ where t_i is the target value and ε_i is the set of slack variables.

Instead of employing a minimization model (4), the problem may be formulated using Lagrangian dual multipliers α as:

$$\max \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i, x_j) \quad (5)$$

subject to:

$$0 \leq \alpha_i \leq C \ \forall\ i=1,2,...,n \text{ and } \sum_{i=1}^{n}\alpha_i y_i = 0 \quad (6)$$

Using the "kernel trick" reduces the complexity of the optimization problem that now only depends on the input space instead of the feature space:

The objective function for SVM with nonlinear kernel has the form:

$$\max \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j k(x_i, x_j) \quad (7)$$

4. Kernels

Kernel methods form a popular class of machine learning techniques for various tasks. An important feature offered by kernel methods is the ability to model complex data through the use of the kernel trick (Schölkopf & Smola, 2002).

In this paper, two types of kernel functions are chosen and evaluated, namely the linear and radial basis function (RBF). The mathematical formula for the said functions are as follows:

4.1. *Linear kernel*

$$k(x_i, x_j) = (x_i, x_j) \quad (8)$$

4.2. *Radial Basis Function (RBF) kernel*

$$k(x_i, x_j) = \exp(-\| x_i - x_j \|^2 / (2\sigma^2)) \quad (9)$$

The two kernels have their own advantages and limitations. The linear kernel offers ease of performance as well as an ability to deal with small and linearly separable samples. The RBF kernel, on the other hand, is known as a good mapping function since it can be used for all kinds of samples, small or large with both high and low dimensions [9]. The SVM performance for each kernel will be evaluated in this study to determine the optimal kernel.

5. Experiments

This section introduces the data sets and the evaluation metric employed. We have compared the three kernels and evaluated their effectiveness for anomaly detection in Support Vector Machines.

5.1. *Data sets*

Time series data obtained from UCR public data set [10] were used to evaluate effectiveness. The details of the datasets are given in Table I. All datasets are presented in time series form, and every data point is manually labeled. For all datasets, we designated the minority class as an anomaly class. Twenty percent of the data was used for testing.

Table 1. Summary of the datasets

Datasets	Length	Number of instances	Anomaly Ratio
ItalyPowerDemand	24	1096	0.49
Wafer	152	7164	0.11
SonyAIBORobotSurface2	65	980	0.38
ECGFiveDays	136	884	0.50
TwoLeadECG	82	1162	0.50
MoteStrain	84	1272	0.46
Herring	512	128	0.40
Strawberry	235	983	0.36

5.2. *Performance Evaluation*

The accuracy of an anomaly detection method is evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC), Precision (Pre), Recall (Rec), and F1-Score, defined as follows:

$$\text{Pre} = \frac{TP}{TP+FP} \tag{10}$$

$$\text{Rec} = \frac{TP}{TP+FN} \tag{11}$$

$$\text{F1} = 2 \times \frac{\text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}} \tag{12}$$

where TP is the correctly detected anomaly, FP is the falsely detected anomaly, TN is the correctly assigned normal, and FN is the falsely assigned normal.

6. Results and Discussion

The efficiency of the following three SVM kernels are compared:

1. Linear Kernel
2. RBF Kernel (Default parameters value)
3. RBF1 Kernel (We define the parameters $C = 20$, $\gamma = 0.02$)

We performed experiments on accuracy of analysis and prediction of anomalies for eight time series data sets using the three different SVM kernels. Accuracy of analysis and prediction can be measured by the AUC as shown in the ROC in Fig. 2 - 9. The blue line is the Linear

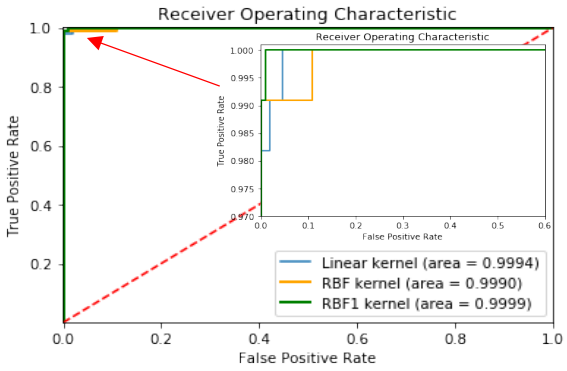

Fig. 2. The kernel performance comparison of Linear, RBF and RBF1 for testing ItalyPowerDemand dataset using ROC.

Kernel of SVM, the orange line is the RBF kernel, and the green line is the RBF1 kernel.

Fig. 2. shows that the RBF1 kernel is slightly more efficient than the Linear and RBF kernels for the ItalyPowerDemand data set. However, all three kernels yield almost 100 percent accuracy.

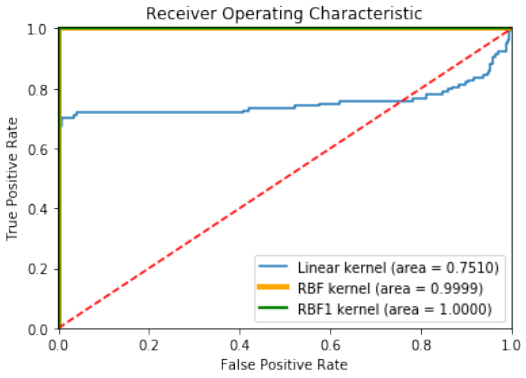

Fig. 3. The kernel performance comparison of Linear, RBF and RBF1 for testing Wafer dataset using ROC.

Fig. 3. shows that the RBF and RBF1 kernels are more efficient than the Linear kernel, and that the RBF

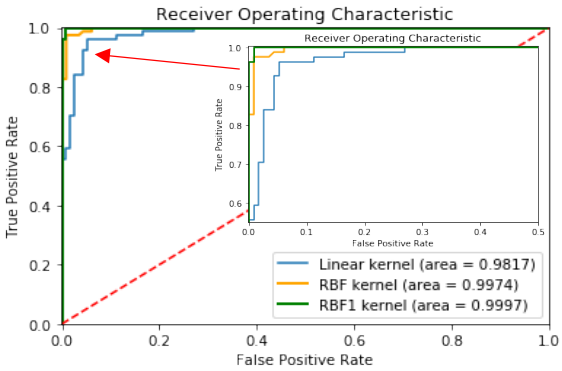

Fig. 4. The kernel performance comparison of Linear, RBF and RBF1 for testing SonyAIBORobotSurface2 dataset using ROC.

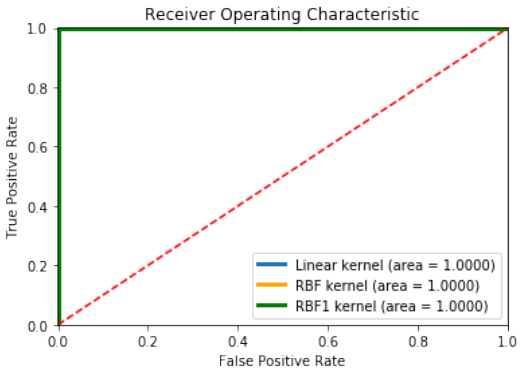

Fig. 5. The kernel performance comparison of Linear, RBF and RBF1 for testing ECGFiveDays dataset using ROC.

kernel is almost 100 percent accurate. In particular, the RBF1 kernel gives a ROC value perfectly for the Wafer data set.

Fig. 4. reveals that RBF1 is slightly more efficient than the linear and RBF kernels for the SonyAIBORobotSurface2 data set. In particular, the RBF and RBF1 kernels provide almost 100 percent accuracy.

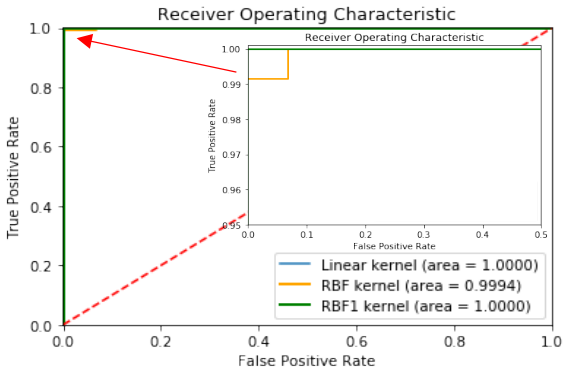

Fig. 6. The kernel performance comparison of Linear, RBF and RBF1 for testing TwoLeadECG dataset using ROC.

Fig. 5. shows that all three kernels give perfect ROC values for the ECGFiveDays data set.

Fig. 6. shows that the RBF kernel is almost 100 percent. Linear and RBF1 kernels give nearly perfect ROC values for the TwoLeadECG data set.

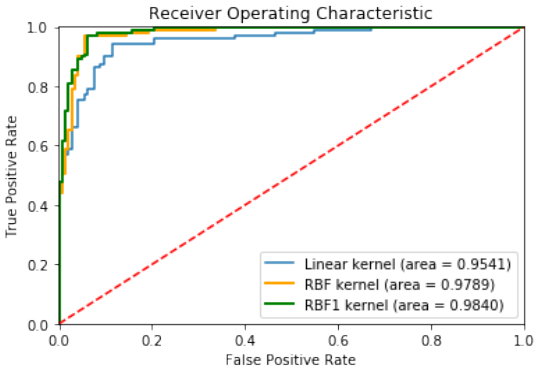

Fig. 7. The kernel performance comparison of Linear, RBF and RBF1 for testing MoteStrain dataset using ROC.

Fig. 7. shows that the RBF1 kernel to be slightly more efficient than the linear and RBF kernels for the MoteStrain data set.

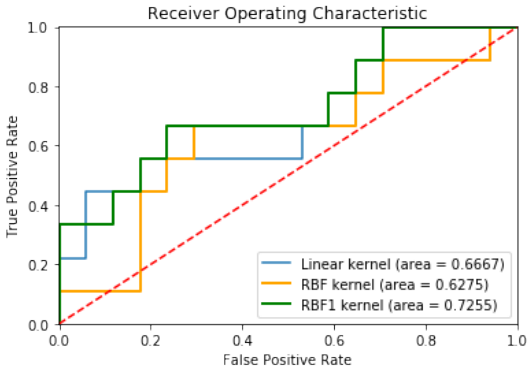

Fig. 8. The kernel performance comparison of Linear, RBF and RBF1 for testing Herring dataset using ROC.

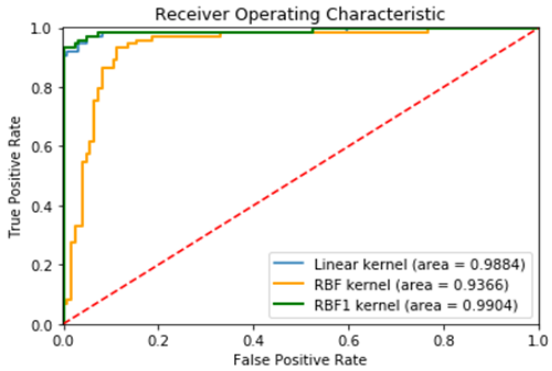

Fig. 9. The kernel performance comparison of Linear, RBF and RBF1 for testing Strawberry dataset using ROC.

Fig. 8. reveals that the Linear kernel gives slightly more accurate ROC values than does the RBF kernel, but the RBF1 kernel is the most accurate for the Herring data set.

Finally, Fig. 9. shows that the Linear kernel is more accurate than the RBF kernel. In particular, the RBF1 kernel is almost 100 percent for the Strawberry data set.

The anomaly detection results and comparisons are summarized in Table II. The results show that SVM with RBF1 kernel gives the highest accuracy and F1-Score on all aspects and data sets, except for the Herring data set, for which the highest of F1-Score is given by the RBF kernel. All three kernels gave perfect results for AUC,

Precision, Recall and F1-Score on the ECGFiveDays data set.

This shows that the RBF kernel with parameter values $C = 20$, $\gamma = 0.02$ exhibits good performance in anomaly detection for time series data.

Table 2. Summary of the kernel performance comparison of Linear, RBF, and RBF1.

Datasets	Kernel : Linear				Kernel : RBF (Default)				Kernel : RBF ($C = 20$, $\gamma = 0.02$)			
	AUC	Precision	Recall	F1-Score	AUC	Precision	Recall	F1-Score	AUC	Precision	Recall	F1-Score
ItalyPowerDemand	0.9994	0.9818	0.9818	0.9818	0.9990	0.9910	1.0000	**0.9955**	**0.9999**	0.9910	1.0000	**0.9955**
Wafer	0.7510	0.9687	0.9946	0.9815	0.9999	1.0000	0.9969	0.9985	**1.0000**	1.0000	0.9985	**0.9992**
SonyAIBORobotSurface2	0.9819	0.9646	0.9478	0.9561	0.9974	0.9910	0.9565	0.9735	**0.9997**	1.0000	0.9826	**0.9912**
ECGFiveDays	1.0000	1.0000	1.0000	1.0000	**1.0000**	1.0000	1.0000	1.0000	**1.0000**	1.0000	1.0000	**1.0000**
TwoLeadECG	1.0000	0.9915	1.0000	0.9957	0.9994	0.9915	1.0000	0.9957	**1.0000**	1.0000	1.0000	**1.0000**
MoteStrain	0.9540	0.9552	0.8767	0.9143	0.9789	0.9783	0.9247	0.9507	**0.9840**	0.9716	0.9384	**0.9547**
Herring	0.6667	0.7647	0.7647	0.7647	0.6275	0.6538	1.0000	**0.7907**	**0.7255**	0.7692	0.5882	0.6667
Strawberry	0.9884	0.9908	0.8710	0.9270	0.9366	0.6327	1.0000	0.7750	**0.9904**	0.9911	0.8952	**0.9407**

7. Conclusion

In this paper, we presented an analysis of anomaly detection in time series data using a Support Vector Machine with three different kernels, namely, Linear, RBF and RBF1. We evaluated the accuracy of anomaly detection methods based on AUC, Precision, Recall and F1-Score criteria. The evaluation results show that the kernel with defined parameters can improve accuracy on all aspects and data sets. This application of the Support Vector Machine method, with the RBF kernel, can be efficient for detecting anomalies in time series data. The results for data set ECGFiveDays show 100 percent accuracy with all three kernels, and the results for the TwoLeadECG show almost 100 percent with all three kernels. Moreover, the results indicate a high degree of accuracy for the three kernels on all the data sets, perhaps because our data was trained in supervised conditions.

In the future, we intend to implement the variational autoencoder method for detecting and predicting anomalies in time series and spectrum data in order to compare it with the autoencoder method.

Acknowledgements

This research was partially supported by a Japanese Government Project in collaboration with the Kyushu Institute of Technology, Yaskawa Electric Corporation, Kitakyushu Foundation for the Advancement of Industry, Science and Technology, and the Hayashi Laboratory.

References

1. E. Hormozi, M. Kazem Akbari, H. Hormozi, Accuracy evaluation of a credit card fraud detection system on Hadoop MapReduce, In *The 5th conference on information and knowledge technology*, 2013, pp. 35-39.
2. Varun Chandola, Arindam Banerjee, and Vipin Kumar, Anomaly detection: A survey, *ACM computing surveys (CSUR)*. **41**(3), 2009.
3. J. Ma and S. Perkins, Time-series novelty detection using one-class support vector machines, In *Proceedings of the International Joint Conference on Neural Networks, 2003*, vol. 3, pp. 1741-1745.
4. H.-S. Kim and S.-D. Cha, Empirical evaluation of SVM-based masquerade detection using UNIX commands, *Computers & Security*, **24**(2), 2005, pp. 160-168.
5. V. Sugumaran, G. R. Sabareesh, K. I. Ramachandran, Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine, *Expert Systems with Applications*, **34**(4), 2008, pp. 3090-3098.
6. U. Caydas and S. Ekici, Support vector machines models for surface roughness prediction in CNC turning of AISI 304 austenitic stainless steel, *Journal of intelligent Manufacturing*, **23**(3), 2012, pp. 639-650.
7. B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, k-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition, *Journal of Transportation Engineering*, **142**(6), 2016.
8. P. Batta, M. Singh, Z. Li, Q. Ding, and L. Trajkovic, Evaluation of Support Vector Machine Kernels for detecting network anomalies, In *2018 IEEE International Symposium on Circuits and Systems*, 2018, pp. 1-4.
9. W. Cao, L. Li and X. LV, Kernel function characteristic analysis based on support vector machine in face recognition, In *2007 International Conference on Machine Learning and Cybernetics*, 2007, vol. 5, pp. 2869-2873.
10. H. A. Dau, E. Keogh, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, Yanping, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, "The ucr

time series classification archive," November 2019, https://www.cs.ucr.edu/~eamonn/time series data 2018/.