Sensors & Transducers, Vol. 159, Issue 11, November 2013, pp. 226-229

Sensors & Transducers

http://www.sensorsportal.com

Support Vector Machine Based Intrusion Detection Method Combined with Nonlinear Dimensionality Reduction Algorithm

Xiaoping Li
Jiangxi Vocational College of Finance and Economics, Jiujiang 332000, China

Received: 3 September 2013 /Accepted: 25 October 2013 /Published: 30 November 2013

Abstract: Network security is one of the most important issues in the field of computer science. The network intrusion may bring disaster to the network users. It is therefore critical to monitor the network intrusion to prevent the computers from attacking. The intrusion pattern identification is the key point in the intrusion detection. The use of the support vector machine (SVM) can provide intelligent intrusion detection even using a small amount of training sample data. However, the intrusion detection efficiency is still influenced by the input features of the ANN. This is because the original feature space always contains a certain number of redundant data. To solve this problem, a new network intrusion detection method based on nonlinear dimensionality reduction and least square support vector machines (LS-SVM) is proposed in this work. The Isometric Mapping (Isomap) was employed to reduce the dimensionality of the original intrusion feature vector. Then the LS-SVM detection model with proper input features was applied to the intrusion pattern recognition. The efficiency of the proposed method was evaluated with the real intrusion data. The analysis results show that the proposed approach has good intrusion detection rate, and is superior to the traditional LSSVM method with a 5.8 % increase of the detection precision. *Copyright © 2013 IFSA.*

Keywords: Intrusion detection, Nonlinear dimensionality reduction, Support vector machine, Isometric mapping.

1. Introduction

Intrusion detection is critical for the normal running of the internet. Terrible intrusion attacks may cripple the local network system. Hence, it is very urgent to recognize the intrusion in time to prevent broken-downs. Up to date, many advanced machine learning algorithms have been proposed to detect the network intrusion, such as evolution algorithm, artificial neural network (ANN), and support vector machine (SVM) and so on [1]. Among them, the SVM, which has the ability to find the decision function from low training set sizes, has been widely used as a learning algorithm in a wide variety of applications. The concept of the kernel trick allows SVM to perform regression and prediction even for nonlinear cases. However, SVM detection performance is affected by its input feature vector. In generally, the intrusion feature data often contains some redundancy, which may decrease the SVM performance [2]. Although the principal component analysis (PCA) [3] and its derivative algorithms have been proved to be a useful tool for feature reduction and extraction to improve the network attack detection accuracy, their main limitation lies in their ability is to capture the nonlinear properties of the original data [4-6]. The nonlinear dimensionality reduction (NDR) methods, including Isomap [4], locally linear embedding (LLE) [5] and Laplacian eigenmap [6], etc., are hence proposed to deal with the underlying nonlinear behavior of the data. Unlike the linear eigenvector-based feature extraction

algorithms, the Isomap algorithm extends the MDS method by introducing the conception of geodesic distances to preserve the geodesic of the underlying nonlinear manifold in a low-dimensional space [7]. The advantage of Isomap, compared with LLE and Laplacian eigenmap, is that it uses global geometric invariants, which are relatively less sensitive to measurement noise [8]. Thus, the Isomap was adopted to mine intrinsic geometry structure in feature vector of the intrusion data in this work. A new integration of Isomap and LS-SVM is proposed for the intrusion detection. The innovation of this new method is to achieve nonlinear feature reduction using Isomap and recognize intrusion patterns by LS-SVM.

By implementing the intrusion detection experiments, the analysis results show that the feature reduction is very essential in the intrusion detection because the original feature space have many redundant features to influence the intrusion identification. Eliminate these redundancies can enhance the intrusion detection.

This paper is organized as follows. In Section 2, the proposed integration method for network intrusion detection based on Isomap and LS-SVM is described. The application of the proposed method is presented for network intrusion detection in Section 3. The effectiveness of the proposed method is valued by analyzing the practice data. Conclusions are drawn in Section 4.

2. The Proposed Intrusion Detection Approach Based on Isomap and LSSVM

Though we can calculate characteristic information from the intrusion data, the original feature space is inevitably contaminated by redundancy. Isomap preserves the geodesic of the underlying nonlinear manifold in a low-dimensional space [7]. This advantage is essential to maintain the nonlinear properties of the input data and thus benefits the intrusion pattern identification.

Isometric Mapping (Isomap). Given a nonlinear high-dimensional dataset $F=[f_1 \quad f_2 \quad \cdots \quad f_n] \in R^p$, where n is the total sample number and p the dimensionality of each sample, the objective of Isomap is to reconstruct a nonlinear mapping to project F into a reduced manifold space $F_r=[f_{r1} \quad f_{r2} \quad \cdots \quad f_{rn}] \in R^q$ ($q<<p$).

For data lying on the manifold, the Isomap aims to find geodesic between two data points along the surface of the manifold, rather than the straight-line Euclidean distance. There are three steps to find this "true distance":

1) determine neighbors using the Euclidean distance $dE(i, j)$ between pairs of points i, j in the input space H,

2) estimate the geodesic distances $dG(i, j)$ between all pairs of points on the manifold by computing their shortest path distances and

3) employ MDS to the matrix of graph distances $\boldsymbol{M}=\{dG(i, j)\}$ to construct an embedding of H in a low-dimensional space.

The theory behind the Isomap can refer to [9]. In this paper, the input space $F_{m\times n}$ (m denotes the feature number, n denotes the samples of data) is projected into a low-dimensional space $\boldsymbol{F}_{p\times n}$ (p=3) by the use of Isomap.

The Least Square Support Vector Machines (LS-SVM). Since the least squares support vector machine (LS-SVM) is an improved algorithm based on SVM which adopts equality constraint to replace the inequality constraints for standard support vector machine, put the SVM quadratic programming problem into linear equations and realize the simplified algorithm. Due to the high accuracy of LS-SVM in classification and regression problems and non-sensitive of the sample dimension, the system adopted the LS-SVM for network intrusion detection. LS-SVM theory of detailed derivation and demonstration please refer to the reference [10], here only give the LS-SVM regression model:

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

where α is the Lagrange multiplier, b is the bias constants and RBF kernel adopted for kernel function

$$K(x_i,x)=\exp[-\frac{|x-x_i|^2}{2\sigma^2}] \tag{2}$$

The Proposed Detection Model. In this paper the Isomap and LS-SVM are proposed for the network intrusion detection. The workflow is given as follows:

Step 1: Pre-treat the original network intrusion data to standardized data format.

Step 2: Reduce the feature space by Isomap to extract distinguished features.

Step 3: Train the LS-SVM using the extracted features, and determine the network intrusion detection result according to the LS-SVM model output.

Step 4: Test the performance of the LS-SVM detection model, and provide the test result as the base for a valid network intrusion management decision. A flow chart of the proposed network intrusion detection method is illustrated in Fig. 1.

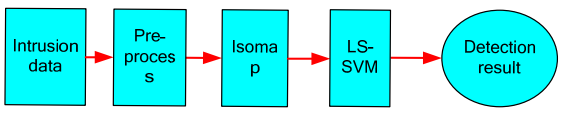

Fig. 1. The network intrusion detection system based on Isomap and LSSVM.

3. Experimental Analysis

In order to validate the performance of the proposed algorithm, the intrusion experiments were carried out in real intrusion data in the present work under four conditions, including normal, DoS, Probe and U2R. The recorded intrusion data describes 41 main attribute of the test network connection, including duration, service type, the bytes issued from source to destination, the bytes from destination to source, etc. In the intrusion detection, 6000 normal samples and 6000 intrusion samples were investigated.

In experiments, Isomap was adopted to reduce the 41 dimension of the original data to 3 dimensions. The performance of the nonlinear dimensionality reduction using Isomap is shown in Fig. 2. To highlight the efficiency of the Isomap, the nonlinear dimensionality reduction was compared with LLE and Laplacian eigenmap, shown in Fig. 3 - Fig. 4.

It can be seen from Fig. 2 - Fig. 4 that most of the information contained in the original data can be presented by the Isomap algorithm, and the intrusion data can be clustered correctly into 4 classes.

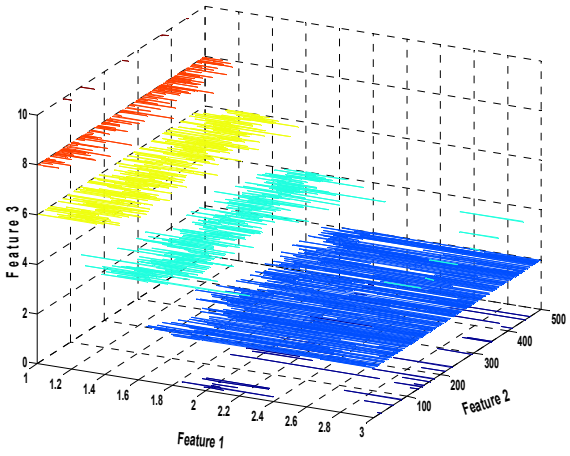

Fig. 2. The feature dimensionality reduction using Isomap.

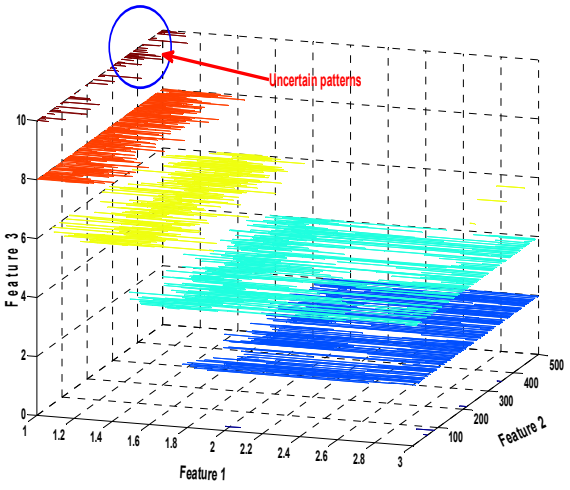

Fig. 3. The feature dimensionality reduction using LLE.

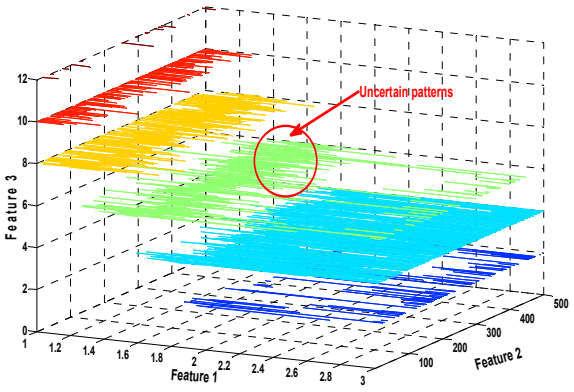

Fig. 4. The feature dimensionality reduction using Laplacian eigenmap.

However, the LLE and Laplacian eigenmap produce another uncertain intrusion pattern. Hence, the performance of the Isomap reduction is superior to the other two methods.

After the Isomap processing, the input of the LSSVM is better than before, and the intrusion detection abilities could be enhanced greatly. This is the advantages of the proposed intrusion detection method against the existing SVM based approaches. The intrusion detection performance is shown in Table 1.

Table 1. The intrusion detection performance of the LSSVM model.

Feature reduction algorithm	Intrusion detection error
None	5.38 %
LLE	1.09 %
Laplacian eigenmap	1.13 %
Isomap	0.92 %

The intrusion detection performance of the Isomap-LSSVM was compared with no feature reduction, LLE reduction and Laplacian eigenmap reduction in Table 1. It can be seen from Table 1 that by the nonlinear dimensionality reduction, the distinct features are obtained and thus the intrusion detection error is dominated significantly. Moreover, the Isomap reduction yields the best intrusion rate, 0.92 %. Hence, it can be seen that the proposed method can detect intrusion effectively.

4. Conclusions

Intelligent method has been widely used in intrusion detection, especially for the SVM based methods. However, reasonable input feature vector of the SVM model plays a critical role in the detection performance. This paper proposed a new intrusion detection method based on Isomap and LSSVM. The innovation is that the new method uses the Isomap based nonlinear dimension reduction to eliminate

useless information. The real practice data was applied to the validation of the proposed approach. The analysis results verify the effectiveness of this newly proposed method.

Acknowledgements

This work was supported in part by the Humanity and Social Science Youth foundation of Ministry of Education of China (13YJC870013).

References

[1]. X. Zhao, R. Jing and M. Gu, Adaptive intrusion detection algorithm based on rough sets, *J T Singhua Univ (Sci & Tech)*, 48, 2008, p. 1165–1168.

[2]. Z. Li and X. Yan, Independent component analysis and manifold learning with applications to fault diagnosis of VSC-HVDC systems, *Hsi-An Chiao Tung Ta Hsueh*, 45, 2011, p. 44–48.

[3]. A. Widodo and B. Yang, Wavelet support vector machine for induction machine fault diagnosis based on transient current signal, *Expert Sys. Appl.*, 35, 2008, p. 307–316.

[4]. J. B. Tenenbaum, V. Silva and J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science*, 290, 2000, p. 2319–2323.

[5]. S. Roweis and L. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science*, 290, 2000, p. 2323–2326.

[6]. M. Belkin and P. Niyogi, Laplacian Eigenmaps for Dimensionality Reduction and Data Representation, *Neural Comput.*, 15, 2003, p. 1373–1396.

[7]. H. Zha and Z. Zhang, Continuum Isomap for manifold learnings, *Comput. Stat. Data Anal.,* 52, 2007, p. 184–200.

[8]. G. Rosman, M. Bronstein, A. Bronstein and R. Kimmel, Nonlinear Dimensionality Reduction by Topologically Constrained Isometric Embedding, *Int. J. Comput. Vis.*, 89, 2010, p. 56–68.

[9]. Z. Cheng, A. Zhu and D. Chen, Fault diagnosis of chemical process using isometric feature mapping and linear discriminant analysis, *CIESC Journal,* 61, 2009, p. 122–126.

[10]. Jiang J., Ma H., Ren D., *et al*., A survey of intrusion detection research on network security, *Journal of Software*, 11, 2000, p. 1460–1466.
