The techniques for computer security intrusion detection based on Preserving Embedding for Anomaly Detection

Chunxia Zhao, Wang linjing and Liao fan

Henan University of Chinese Medicine, Zhengzhou, China

Keywords: security intrusion detection, Preserving Embedding, Anomaly Detection

Abstract. Computer security has been attracting more and more attention ,since intrusion detection have become a significant threat in recent years. The techniques for intrusion detection are generally classified into two categories, which are anomaly detection and misuse detection respectively. In this paper, we mainly focus on anomaly detection on behavior of process which is in the form of system call traces. Each process trace is recorded by different system calls that can be naturally deemed as high dimensional data, as one operating system may have a great deal of different system calls. Thus, it is natural without any doubt to say that dimension reduction technique has the opportunity to make a better performance improvement by exploiting classifier in the low dimensional subspace.

Introduction

Unlike misuse detection attempting to model attacks as specific signatures, anomaly detection identifies activities that deviate from the normal behavior of the monitored system and thus has the potential to detect novel attack[1]. Thus, anomaly detection draws more preference than misuse detection from research community with the reason of its intelligent characteristics. Dimension reduction techniques, such as feature selection[2] and feature transformation[3], have been employed by anomaly detection for users, program or network behavior. However, wrapper filter and hybrid feature selection approaches are generally conducted in trivial way by exploiting a machine learning algorithm to evaluate the fineness of result of feature subset. As to feature transformation, Principal Component Analysis (PCA) [3] is recently proposed on traffic anomaly detection. Many challenges, however, are still faced by research community [4].

The representative nonlinear dimension reduction approaches include LLE, Laplacian Eigenmap[7], etc. However, they yield maps that are defined only on training data and the issue how to map test data to the low dimensional space remains difficult. Therefore, these nonlinear manifold dimension reduction algorithms can not be applied directly to classification problems. Recently, some manifold based linear dimensionality reduction methods, such as Neighborhood Preserving Embedding (SNPE), etc, have been proposed, which resolved the difficulty of how to implement the map on new test data. However, these methods may fail to achieve good performance when the data structure is nonlinear, because they are intrinsically linear dimensionality reduction technique. In this paper, we present a new approach based on SNPE, which is referred by Supervised Kernel Neighborhood Preserving Embedding (SKNPE), and introduce this dimension reduction technique into system call anomaly detection.

Our novel approach, SKNPE, makes three contributions as follows. First of all, SKNPE utilizes kernel trick to improve the ability of mapping nonlinear data structure by employing the neighborhood information in a supervised way. Second, SKNPE combined with KNN classifier, called SKNPE-KNN, shows promising capacity of noisy resistance, when training data is accompanied with noisy data. Third, our experiments show the superiority of SKNPE-KNN not

only in terms of high intrusion detection accuracy and low false positives but also in terms of superiority of running time over other techniques.

Related Work

This metric not only considered the frequency of all system call in one process, but also the binary based common degree the two processes have. The experiment had shown this new technique owns the same noisy resistant capacity as RSVM, and outperforms those previous works. Sharma continued to expand Liao's work, and incorporated the kernel trick into similarity measurement.The experiment argued that the best performance was proved against previous techniques using the same experiment dataset. However, 50 system calls as features were arbitrarily chosen in previous works to avoid the curse of high dimensionality. Whether 50 system calls as features are most representative or whether the number of chosen features are appropriate are still a little questionable. This paper utilize SKNPE in system call anomaly detection to find an optimal subspace that mostly preserves the neighborhood information, hoping that KNN has the capacity of performing best in lower dimensionality on both clean and noisy training data. Our approach, called SKNPE-KNN, is compared with the intrusion detection performance of Liao and Sharma's work in detection rate, false positive and computation time aspects.

Review of Supervised Neighborhood Preserving Embedding

Supervised Neighborhood Preserving Embedding (SNPE) is one of the most recent proposed linear dimensionality reduction techniques, which is to find an optimal embedding by representing each data point as a linear combination of the neighboring data points such that the neighborhood structure can be preserved in the dimensionality reduced space.

The essence of SNPE is briefly demonstrated as follow. Given a set of N points $X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$ in $\mathbb{R}^d$, construct neighborhood weights matrix W and find a transformation matrix A that maps these N points a set of points $Y = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N]$ in $\mathbb{R}^l$ ($l \ll d$), such that $\mathbf{y}_i$ represent $\mathbf{x}_i$, where $\mathbf{y}_i = A^T \mathbf{x}_i$ with $A = [\mathbf{a}_0, \mathbf{a}_1, ..., \mathbf{a}_{l-1}]$, and A is an $d \times l$ matrix.

Suppose that class labels are available and each point $\mathbf{x}_i$ can be reasonably reconstructed by linear combination of its neighbors. The reconstruction error needed to be minimized is defined as:

$$E(W) = \sum_i || \mathbf{x}_i - \sum_j w_{ij} \mathbf{x}_j ||^2 \qquad (1)$$

with constraint $\sum_j W_{ij} = 1$ ($j = 1, 2, ..., N$) and w_{ij} is equal to 0 if $\mathbf{x}_i$ and $\mathbf{x}_j$ come from different classes. Especially, points among the K nearest neighbors of $\mathbf{x}_i$ are utilized to compute w_{ij}, when $\mathbf{x}_i$ and $\mathbf{x}_j$ are from the same class. Others out of K nearest neighbors of $\mathbf{x}_i$ are still set to be 0, although they are from the same class. The neighborhood matrix W can be determined by solving the minimization of formula (1), which can be referred to LLE[6].

Since each data point can be represented as a linear combination of its neighbors, we reasonably assume that the image of each data point also can be represented as a linear combination of its neighbors with the same neighborhood matrix W. Suppose that the transformation is linear, that is, $\mathbf{y}^T = \mathbf{a}^T X$. Refined transformation A means to minimize the following cost function

$$\phi(\mathbf{y}) = \sum_i (\mathbf{y}_i - \sum_j w_{ij}\mathbf{y}_j)^2 \qquad (2)$$

with constraint $\mathbf{y}^T\mathbf{y} = 1$. This minimizing problem reduces to one of generalized eigenvalue problem

$$XMX^T\mathbf{a} = \lambda XX^T\mathbf{a} \qquad (3)$$

where $M = (I - W)^T(I - W)$ and I is an unit matrix..

Kernel Supervised Neighborhood Preserving Embedding

Although SNPE can preserve the local neighborhood structure well in low dimension space, it is also worthy to declare that SNPE is less efficient on mapping of nonlinear data structure. In this subsection, we utilize kernel trick to improve Supervised Neighborhood Preserving Embedding (SNPE). By employing kernel trick, SNPE can make itself more accessible to dealing with intrinsically nonlinear data, which is highly pervasive in real application.

As to establish the neighborhood matrix W, SKNPE shares the same way as SNPE. With respect to answer how to find the optimal transformation matrix, we need to make a reasonable deduction. To begin with, the data is mapped into an implicit feature space F, using a nonlinear transformation $\phi(x)$: $\mathbb{R}^d \to \mathbb{R}^m$, $d < m$. The data points are mapped as $\phi(X) = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), ..., \phi(\mathbf{x}_N)]$. The cost function we need to minimize in the feature space is described:

$$\phi(\mathbf{z}) = \sum_i (z_i - \sum_j w_{ij} z_j)^2 \qquad (4)$$

Suppose the transformation is linear $\mathbf{z}^T = \mathbf{a}^T\phi(X)$. We define $s_i = z_i - \sum_j W_{ij} z_j$, which can be written in vector form $\mathbf{s} = \mathbf{z} - W\mathbf{z} = (I - W)\mathbf{z}$. The cost function can be made a deduction as follow:

$$\begin{aligned} \phi(\mathbf{z}) &= \sum_i (z_i - \sum_j w_{ij} z_j)^2 \\ &= \sum_i (s_i)^2 \\ &= \mathbf{s}^T\mathbf{s} \\ &= \mathbf{z}^T(I - W)^T(I - W)\mathbf{z} \\ &= \mathbf{a}^T\phi(X)\,M\,\phi(X)^T\mathbf{a} \end{aligned} \qquad (5)$$

where $M = (I - W)^T(I - W)$. We also impose a constraint as follows: $\mathbf{z}^T\mathbf{z} = 1$. Since $\mathbf{a} \neq 0$, $\mathbf{a}$ must be in the span of $\phi(\mathbf{x}_i)$, and can be written as some linear combination of $\phi(\mathbf{x}_i)$. In other

words, there must exist some set of α_i such that $\mathbf{a} = \sum_i \alpha_i \phi(\mathbf{x}_i)$, which can be written in vector form $\mathbf{a} = \phi(X)\boldsymbol{\alpha}^T$, where $\boldsymbol{\alpha}$ is an $1 \times N$ vector. Thus, $\phi(Z)$ also can be reformulated as:

$$\phi(Z) = \boldsymbol{\alpha} KMK\boldsymbol{\alpha}, \tag{6}$$

where $K = \phi(X)^T \phi(X)$ is a $(N \times N)$ kernel matrix whose entries are $K(i, j) = (\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j))$. Finally, the constrained minimization problem reduces to finding

$$\arg\min_{\mathbf{z}^T\mathbf{z}=1} \boldsymbol{\alpha} KMK\boldsymbol{\alpha}. \tag{7}$$

The vector $\boldsymbol{\alpha}$ that minimizes the cost function is given by the minimum eigenvalue solution to the following generalized eigenvector problem:

$$KMK\boldsymbol{\alpha} = \lambda KK\boldsymbol{\alpha} \tag{8}$$

Thus, we choose eigenvectors corresponding to the bottom d nonzero eigenvalues to construct A_1, such that each data point $\mathbf{x}_i$ can be mapped to a d dimensional data point $\mathbf{z}_i$ by

$$\mathbf{z}_i = A^T \phi(\mathbf{x}) = A_1^T \theta,$$

where $\theta = [k(\mathbf{x}_1, \mathbf{x}), k(\mathbf{x}_2, \mathbf{x}), ..., k(\mathbf{x}_N, \mathbf{x})]^T$.

Experimental Setup and Result

In this section, we use system call database of Computer Immune System at NMU to prove the performance of KSNPE combined with KNN, called SKNPE-KNN, on both clean and noisy datasets. Prior to experiment, we need to organize the experiment benchmark and set several common parameters. As system call datasets provided by NMU contain system call sequence from a variety of programs and computer operation systems, we choose synthetic sendmail, synthetic lpr , and live lpr programs as the data source of our experiment. Note that system call traces from those three programs collected from SunOS 4.1.4 are recorded by 182 different system calls. We also extract 2398 unique normal process traces and 83 unique abnormal ones from these three programs. We then prepare the dataset for experiments illustrated in Table 1. It is also worthy of declaring that all the following experiments are conducted on WINDOW XP computer platform with Intel 2.4 GHz T8300 processor and 2G Main Memory. Note that without specific declaration, several common parameters needed for experiments are set as follows. The kernel type for SKNPE is polynomial kernel. K nearest neighborhood points that construct W for both SKNPE and SNPE are set to be 10. The number of nearest neighborhood used for KNN combined with SKNPE and SNPE are also set to be 10. The number of nearest neighborhood used for SBWRBF-KNN and KNN algorithms are, however, set to be 5.

Table 1 Experiments Data Preparation

	Clean Data	Noisy Data
Training	300 Normal Processes 48 Intrusive Processes	320 Normal Processes (20 mislabelled) 28 Intrusive Processes
Testing	600 Normal Processes, 35 Intrusive Processes	

In the following experiments, we compare SKNPE-KNN to previous works. SNPE-KNN (the combination of SNPE and KNN), Liao's research work referred as KNN, and Alok Sharma's successive work referred as SBWRBF-KNN, where smooth radial basis function is used, are selected as three competitors against SKNPE-KNN.

Conclusion

In this paper, we proposed a new dimension reduction approach, called Supervised Kernel Neighborhood Preserving Embedding (SKNPE), for system call intrusion detection. Experiments with system call database maintained by NMU show that SKNPE-KNN in lower dimension space possesses better detection rate, lower false positive and lower computation time on both clean and noisy training data. Especially, its lower computation time and capacity of resistant noisy training data make the proposed technique more suitable for practical system call intrusion detection utilization.

References

[1] H. Debar, M. Dacier and A. Wespi. "Towards a taxonomy of intrusion detection systems", *Computer Networks*, vol. 31, no.8, pp. 805-822, Apr. 1999.

[2] L. Song, A. Smola and A Gretton et al. "Feature Selection via Dependence Maximization", *Journal of Machine Learning Research*, pp. 1393-1434, 2012.

[3] L. Huang, X. Nguyen and M. Garofalakis et al. "In-network PCA and anomaly detection", *Advances in Neural Information Processing Systems 19*, pp. 617-624. MIT Press, Cambridge, MA, 2007.

[4] H. Ringberg , A. Soule and J. Rexford et al, "Sensitivity of PCA for traffic anomaly detection", *ACM SIGMETRICS Performance Evaluation Review*, Vol.35 No.1, June, 2007.

[5] L. Parsons, E. Haque and H. Liu. "Subspace clustering for high dimensional data: a review". *ACM SIGKDD Explorations*, vol. 6, no. 1, pp. 90–105, 2004.

[6] S. Roweis, and L. K. Saul. "Nonlinear dimensionality reduction by locally linear embedding", *Science*, vol.290, pp. 2323-2326, Dec. 2000.

[7] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering", *Advances in Neural Information Processing Systems 14*, Vancouver, British Columbia, Canada, 2001.