Network Anomaly Detection using Fuzzy Gaussian Mixture Models

Dat Tran, Wanli Ma, and Dharmendra Sharma

Faculty of Information Sciences and Engineering,
University of Canberra, ACT 2601, Australia
{Dat.Tran, Wanli.Ma, Dharmendra.Sharma}@canberra.edu.au

Abstract. Fuzzy Gaussian mixture modeling method is proposed in this paper for network anomaly detection. A mixture of Gaussian distributions was used to represent the network data in multi-dimensional feature space. Gaussian parameters were estimated using fuzzy c-means estimation. The method was tested with the KDD Cup data set. Experimental results have shown that the proposed method is more effective than the vector quantization method.

Keywords: Fuzzy Gaussian mixture model, network anomaly detection.

1 Introduction

Network intrusion detection systems are automated systems that detect intrusions in computer network systems. An anomaly behavior detecting-based intrusion detection system builds normal traffic model and uses this model to detect abnormal traffic patterns and intrusion attempts. The goal of this anomaly detection system is to determine whether an unknown network data item belongs to normal or to an intrusive pattern [1]-[7].

Current network intrusion detection methods provide low detection rates because of the multi-dimensional data problem. For example, a simple variant of single-linkage clustering was applied to learn network traffic patterns on unlabelled noisy data [8]. The KDD CUP 1999 dataset [9] was used and this approach achieved from 40% to 55% detection rate and from 1.3% to 2.3% false positive rate.

We propose to use fuzzy Gaussian mixture modeling method to train the normal network model. Fuzzy Gaussian mixture model (FGMM) is an effective model capable of achieving high recognition accuracy for pattern recognition [10]. A number of prototypes are generated from the training network feature vectors by representing the feature space as a mixture of Gaussian distributions. Each prototype consists of a set of model parameters including mean vector, covariance matrix and mixture weight. Parameters are trained in an unsupervised learning method based on fuzzy estimation. Experimental results show that the proposed FGMM method provides a better detection result than the vector quantization method.

The rest of the paper is as follows. Section 2 presents the FGMM method. Section 3 describes network data and presents experimental results. Finally, we conclude the paper in Section 4.

2 Fuzzy Gaussian Mixture Models

Let $X = \{x_1, x_2, ..., x_T\}$ be a set of T feature vectors from the voice data of a person. Let $U = \{u_{it}\}$ be a fuzzy C-partition of X, each u_{it} represents the degree of vector x_t belonging to the ith mixture and is called the fuzzy membership function. For $1 \leq i \leq C$ and $1 \leq t \leq T$, we have

$$0 \leq u_{it} \leq 1 \qquad \sum_{i=1}^{C} u_{it} = 1 \qquad \text{and} \qquad 0 < \sum_{t=1}^{T} u_{it} < T \tag{1}$$

where C is the number of mixtures, $m > 1$ is a weighting exponent on each fuzzy membership u_{it} and is called the degree of fuzziness. Let λ denote a cell phase model consisting of a set of model parameters $\lambda = \{w_i, \mu_i, \Sigma_i\}$, where w_i, μ_i and Σ_i, i = 1,…, C, are mixture weights, mean vectors and covariance matrices. The fuzzy objective function was proposed as follows [1]

$$J_m(U, \lambda) = \sum_{i=1}^{C} \sum_{t=1}^{T} u_{it}^m d_{it}^2 \tag{2}$$

We generalize the fuzzy objective function through the use of fuzzy mean vector, fuzzy covariance matrix and fuzzy mixture weight. To obtain these, since the density of the data in cluster i is proportional to the joint mixture density function $P(x_t, i \mid \lambda)$, we can define the dissimilarity denoted by the distance in (2) as follows

$$d_{it}^2 = -\log P(x_t, i \mid \overline{\lambda}) = -\log[\overline{w}_i N(x_t, \overline{\mu}_i, \overline{\Sigma}_i)] \tag{3}$$

where

$$N(x_t, \mu_i, \Sigma_i) = \frac{\exp\left\{-\frac{1}{2}(x_t - \mu_i)' \Sigma_i^{-1} (x_t - \mu_i)\right\}}{(2\pi)^{d/2} \mid \Sigma_i \mid^{1/2}} \tag{4}$$

From (3) and (4), we have

$$d_{it}^2 = -\log \overline{w}_i + \tfrac{1}{2}\log(2\pi)^d \mid \overline{\Sigma}_i \mid + \tfrac{1}{2}(x_t - \overline{\mu}_i)' \overline{\Sigma}_i^{-1} (x_t - \overline{\mu}_i) \tag{5}$$

Substituting (3) into (2) gives

$$J_m(U, \mu, \Sigma, w; X) = -\sum_{i=1}^{C} \sum_{t=1}^{T} u_{it}^m \log \overline{w}_i \; - \sum_{i=1}^{C} \sum_{t=1}^{T} u_{it}^m \log N(x_t, \overline{\mu}_i, \overline{\Sigma}_i) \tag{6}$$

Minimizing J_m is performed by minimizing each term on the right hand side of (6).

To minimize the first term, note that

$$\sum_{i=1}^{C} \overline{w}_i = 1 \tag{7}$$

and after using the Lagrange multiplier method, we have

$$\overline{w}_i = \sum_{t=1}^{T} u_{it}^m \Big/ \sum_{i=1}^{C} \sum_{t=1}^{T} u_{it}^m \tag{8}$$

The expression of $\overline{w}_i$ in (8) is defined as the fuzzy mixture weight. Minimizing the second term on the right-hand side of (6) is obtained by using (4) and (5) and setting derivatives with respect to μ_i and Σ_i to zero for every i =1,…, C

$$\sum_{t=1}^{T} u_{it}^m \overline{\Sigma}_i^{-1} (x_t - \overline{\mu}_i) = 0 \tag{9}$$

$$\sum_{t=1}^{T} u_{it}^m [\overline{\Sigma}_i - (x_t - \overline{\mu}_i)(x_t - \overline{\mu}_i)'] = 0 \tag{10}$$

From (9) and (10) we have

$$\overline{\mu}_i = \sum_{t=1}^{T} u_{it}^m x_t \Big/ \sum_{t=1}^{T} u_{it}^m \tag{11}$$

$$\overline{\Sigma}_i = \frac{\sum_{t=1}^{T} u_{it}^m (x_t - \mu_i)(x_t - \mu_i)'}{\sum_{t=1}^{T} u_{it}^m} \tag{12}$$

where u_{it} is computed using (2) since it is derived from minimizing J_m with $\{u_{it}\}$ as variables. We obtain

$$u_{it} = \left[\sum_{k=1}^{C} (d_{it} / d_{kt})^{\frac{2}{m-1}} \right]^{-1} \tag{13}$$

The training and detection procedures of this FGMM algorithm are summarized as follows.

Training:

1. Given $X = \{x_1, x_2, ..., x_T\}$ as the *normal* data set of T network feature vectors
2. Train a FGMM model as follows
   a. Generate u_{it} at random satisfying (1)

b. Generate $\lambda = \{w_i, \mu_i, \Sigma_i\}$ at random satisfying (7)
c. Calculate $J_m(U, \lambda)$ using (2)
d. Update the *normal* model $\overline{\lambda} = \{\overline{w}_i, \overline{\mu}_i, \overline{\Sigma}_i\}$ using (8), (11) and (12)
e. Update $\overline{u}_{it}$ using (13)
f. Calculate $J_m(\overline{U}, \overline{\lambda})$
g. Stop if the difference between the fuzzy objective function $J_m(U, \lambda)$ and its update $J_m(\overline{U}, \overline{\lambda})$ is below a chosen threshold, otherwise go to step *d*.

Anomaly Detection:

1. Let λ be the *normal* model that has been trained. Given an unknown network feature vector *x*
2. Calculate the probability $P(x \mid \lambda)$ as follows

$$P(x \mid \lambda) = \sum_{i=1}^{C} P(x, i \mid \lambda) \tag{14}$$

where

$$P(x, i \mid \lambda) = w_i N(x, \mu_i, \Sigma_i) \tag{15}$$

w_i, μ_i and Σ_i are the mixture weight, mean vector and covariance matrix in the Gaussian mixture *i* of the *normal* model. The Gaussian $N(x, \mu_i, \Sigma_i)$ is calculated using (4)

3. Set a threshold value θ
4. If $P(x \mid \lambda) > \theta$ then *x* is normal else *x* is intrusive

It can be seen that when the threshold value increases, the anomaly detection rate and the false alarm rate also increase. If the false alarm rate is fixed, we can determine the corresponding values for the threshold value and the anomaly detection rate.

4 Experimental Results

We consider a sample dataset which is the KDD CUP 1999 dataset [9]. This dataset was based on MIT Lincoln Lab intrusion detection dataset, also known as DARPA dataset. The data was produced for "The Third International Knowledge Discovery and Data Mining Tools Competition", which was held in conjunction with the Fifth International Conference on Knowledge Discovery and Data Mining. The raw network traffic records have already been converted into vector format. Each feature vector consists of 41 features. The meanings of these features can be found in [9].

The proposed method for network intrusion detection was evaluated using the KDD CUP 1999 data set for training and the *Corrected* data set for testing. The testing data set contains 60593 feature vectors for the *normal* network pattern, and 306, 58001, 354, 1633 and 164091 feature vectors for the five attacks *ipsweep*, *neptune*, *portsweep*, *satan*, and *smurf*, respectively.

Table 1 presents false alarm rates in percentage for the FGMM method compared with the vector quantization using *K*-means clustering method. The threshold was set to a value such that anomaly detection rates are equal to 100%. The FGMM method achieved lower false alarm rates in different model sizes (number of Gaussians in a model).

Table 1. False alarm rates (in %) for vector quantization (VQ) and fuzzy Gaussian mixture model (FGMM). Anomaly detection is 100% for all.

Modeling method	False Alarm (%)	Number of clusters/Gaussians
VQ	18.8	2
FGMM	12.1	2
VQ	18.8	4
FGMM	11.9	4
VQ	18.5	8
FGMM	11.9	8
VQ	17.0	16
FGMM	11.5	16
VQ	17.0	32
FGMM	9.1	32
VQ	17.0	64
FGMM	9.0	64

4 Conclusion

We have presented fuzzy Gaussian mixture modeling method and applied it to building the normal network model for anomaly detection. We have used the KDD CUP 1999 dataset as the sample data for the study. Experimental results have shown that the proposed method is more effective than the vector quantization method.

References

1. Snort. Snort web site, http://www.snort.org.
2. Cisco.http://www.cisco.com/en/US/products/sw/secursw/ps2113/products/whitepaper 09186a008010e5c8.shtml.
3. V. Paxson, “Bro: A system for detecting network intruders in real-time”, in Proceedings of the 7th USENIX Security Symposium, 1998, Texas, USA, pp. 3-7.
4. Y. Yasami, M. Farahmand, V. Zargari, “An ARP-based Anomaly Detection Algorithm Using Hidden Markov Model in Enterprise Networks”, Second International Conference on Systems and Networks Comm, 2007, pp. 69 - 75
5. P.K. Chan, M.V. Mahoney, and M.H. Arshad, “A Machine Learning Approach to Anomaly Detection”, Technical Report CS-2003-06, 2003.
6. E. Eskin, “Anomaly Detection over Noisy Data Using Learned Probability Distributions”, in the 17th International Conference on Machine Learning, Morgan Kaufmann, San Francisco, USA, 2000, pp. 255-262.

7. W. Lee and D. Xiang, “Information theoretic measures for anomaly detection”, in 2001 IEEE Synposium on Security and Privacy, pp. 130-143.
8. L. Portnoy, E. Eskin, and S. Stolfo, “Intrusion detection with unlabeled data using clustering”, in Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001), 2001, Philadelphia, USA, pp. 333-342.
9. ACM KDD CUP 1999 Data Set, available at http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.
10. Dat Tran, Tuan Pham and Xiaobo Zhou, “Cell Phase Identification Using Fuzzy Gaussian Mixture Models”, the 2005 International Symposium on Intelligent Signal Processing and Communications Systems, pp. 465-468, December 2005, Hong Kong
11. Stanifor, Hoagland and McAlerney, “Practical Automated Detection of Stealthy PortScans”, Journal of Computer Security, 2002, vol. 10, no. 1, pp. 105-136
12. M. V. Mahoney and P.K. Chan, “PHAD: Packet Header Anomaly Detection for Identifying Hostile Network Traffic”, Technical report, Florida Tech., CS-2001-4, 2001
13. H. Yang, F. Xie, and Y. Lu, “Clustering and Classification Based Anomaly Detection”, Lecture Notes in Computer Science, 2006, vol. 4223, pp. 1611-3349.
14. C. Taylor and J. Alves-Foss, “An Empirical Analysis of NATE: Network Analysis of Anomalous Traffic Events”, in 10th New Security Paradigms Workshop, 2002, Virginia Beach, Virginia, USA, pp. 18-26.
15. D. Tran and T. Pham, “Modeling Methods for Cell Phase Classification”, Book chapter in the book Advanced Computational Methods for Biocomputing and Bioimaging, Editors: T.D. Pham, H. Yan, D. I. Crane, Nova Science Publishers, New York, USA, ISBN: 1-60021-278-6, 2007, chapter 7, pp. 143-166.
16. D. Tran, W. Ma, D. Sharma and T. Nguyen, “Fuzzy Vector Quantization for Network Intrusion Detection”, IEEE International Conference on Granular Computing, Silicon Valley, 2-4 November 2007, USA.
17. D. Tran and W. Wagner, “Fuzzy entropy clustering”, in Proceedings of FUZZ-IEEE Conference, 2000, vol. 1, pp. 152-157.
18. S.J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P.K. Chan, “Cost-based Modeling and Evaluation for Data Mining With Application to Fraud and Intrusion Detection: Results from the JAM Project”, in Proceedings of 2000 DARPA Information Survivability Conference and Exposition, 2000, pp. 1130-1144.
19. R. Anderson and A. Khattak, “The use of Information Retrieval Techniques for Intrusion Detection”, in First International Workshop on Recent Advances in Intrusion Detection (RAID'98), 1998, Louvain-la-Neuve, Belgium.