ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

The Conditional Random Fields of Layered Approach using Intrusion Detection

[1]Shaik. Riaz, [2]K. V. S. S. Ramakrishna
[1]Dept. of Computer Science, St. Mary's Group of Institutions, Guntur, AP, India
[2]Dept. of Computer Science, GVR & S College of Engineering & Technology, Guntur, AP, India

Abstract
Intrusion detection faces a number of challenges; an intrusion detection system must reliably detect malicious activities in a network and must perform efficiently to cope with the large amount of network traffic. Two issues of accuracy and efficiency using conditional Random Fields and Layered Approach. We demonstrate that high attack detection accuracy can be achieved by using conditional Random Fields and high efficiency by implementing the Layered Approach. Experimental results on the benchmark KDD'99 intrusion data set show that our proposal system based on Layered conditional Random fields outperforms other well-known methods such as the decision trees and the naïve Bayes. the improvement in attack detection accuracy is very high, particularly, for the U2R attacks and the R2L attacks. Statistical Tests also demonstrate higher confidence in detection accuracy for our method. Finally, we show that our system is robust and is able to handle noisy data without compromising performance.

Keywords
Instrusion Detection, R2L Attacks, U2R Attacks, Random Fields, Layered Approach

I. Introduction

Intrusion detection is one of the high priority and challenging tasks for network administrators and security professionals. More sophisticated security tools mean that the attackers come up with newer and more advanced penetration methods to defeat the installed security systems. However an accurate system that cannot handle large amount of network traffic and is slow in decision making will not fulfill the purpose of an intrusion detection system. We desire a system that detects most of the attacks, gives very few false alarms, copes with large amount of data, and is fast enough to make real-time decisions.

Another approach for detecting intrusions is to consider both the normal and the known anomalous patterns for training a system and the performing classification on the test data. Such a system incorporates the advantages of both the signature-based and the anomaly-based system and is known as the Hybrid system. Hybrid systems can be very efficient, subject t the classification method used, and can also e used to label unseen or new instances as they assign one of the known classes to every test instance. This is possible because during training the system learns features from all the classes. A Conditional Random Field (CRF) is a statistical modeling method often applied inpattern recognition. More specifically it is a type of discriminative undirective undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for labeling or parsing of sequential data, such as natural language text or biological sequences and in computer vision . Specifically, CRFs find applications in shallow parsing , named entity recognition and gene finding, among other tasks, being an alternative to the related hidden Markov models. In computer vision, CRFs are often used for object recognition and image segmentation.

The structured support vector machine is a machine learning algorithm that generalizes the Support Vector Machine (SVM) classifier. Whereas the SVM classifier supports binary classification ,multiclass classification and regression, the structured SVM allows training of a classifier for general structured output labels.

As an example, a sample instance might be a natural language sentence, and the output label is an annotated parse tree. Training a classifier consists of showing pairs of correct sample and output label pairs. After training, the structured SVM model allows one to predict for new sample instances the corresponding output label; that is, given a natural language sentence, the classifier can produce the most likely parse tree.

For a set of ℓ training instances $(\boldsymbol{x}_n, y_n) \in \mathcal{X} \times \mathcal{Y}, n = 1, \ldots, \ell$ from a sample space $\mathcal{X}$ and label space $\mathcal{Y}$, the structured SVM minimizes the following regularized risk function.The function is convex in $\boldsymbol{w}$ because the maximum of a set of affine functions is convex. The function $\Delta : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$ measures a distance in label space and is an arbitrary function (not necessarily a metric) satisfying $\Delta(y, z) \geq 0$ and $\Delta(y, y) = 0 \forall y, z \in \mathcal{Y}$. The function $\Psi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ is a feature function, extracting some feature vector from a given sample and label. The design of this function depends very much on the application.

Because the regularized risk function above is non-differentiable, it is often reformulated in terms of a quradratic program by introducing one slack variables ξn for each sample, each representing the value of the maximum.

A. Higher-order CRFs and semi-Markov CRFs

CRFs can be extended into higher order models by making each Y_i dependent on a fixed number o of previous variables $Y_{i-o}, \ldots, Y_{i-1}$. Training and inference are only practical for small values of o (such as $o \leq 5$), since their computational cost increases exponentially with o. Large-margin models for structured prediction, such as the structured support vector Machine can be seen as an alternative training procedure to CRFs.

There exists another generalization of CRFs, the semi-Markov conditional random field (semi-CRF), which models variable-length segmentations of the label sequence Y [6]. This provides much of the power of higher-order CRFs to model long-range dependencies of the Y_i, at a reasonable computational cost.

B. Conditional Random Fields

Consider a scenario where a hidden process is generating observables. Assume that the structure of the hidden process is known. For example, in NER and POS tagging tasks, we make the assumption that a particular POS tag (or named entity tag) depends only on the current word and the immediately previous and the immediately next tags. This corresponds to an undirected graphical model in the shape of a linear chain. Another example is the classification of a set of hyperlinked documents. The label of a document can be assumed to be dependent upon the document itself and the labels of the documents that link into it or out of it. Two tasks arise in these scenarios: 1. Learning: Given a sample set of the observables $\{x_1, \ldots, x_N\}$ along with the values of

the hidden labels $\{y_1, \ldots, y_N\}$, learn the best possible potential functions such that some criteria is maximized.

II. Inference

Given a new observable x, find the most likely set of hidden labels y_ for x, i.e. compute
(exactly or approximately):
y_ = arg max
y
P(y|x) (6)
Here, the graphical model would have some nodes (say Y_i's) and edges corresponding to the labels and the dependencies between them and atleast one more node (say X) corresponding to the observable x, along with some edges of the kind (X, Y_i). The joint probability distribution can thus be written as $P(x, y_1, \ldots, y_M)$
$=1Z_{\{X\}}(x)Y_{C2C,C6=\{X\}\,C}(x, y)$ (7)
Learning this joint distribution is both intractable (because the {X}(.) function is hard to approximate without making naive assumptions) as well as useless (because x is already provided to us). Thus, it makes sense to learn the following conditional distribution:
$P(y_1, \ldots, y_M|x) = 1Z_x Y_{C2C,C6=\{X\}\,C}(x, y)$ (8)
Note that the normalizer is now observable-specific.
The undirected graph with the set of nodes {X}[Y and the relevant Markovian properties is called a Conditional Random Field (CRF). From now on, we will assume that C excludes the singleton clique {X}.
An Intrusion Detection System (IDS) is a device or software application that monitors network and/or system activities for malicious activities or policy violations and produces reports to a Management Station. Some systems may attempt to stop an intrusion attempt but this is neither required nor expected of a monitoring system. Intrusion Detection and Prevention Systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, and reporting attempts. In addition, organizations use IDPSes for other purposes, such as identifying problems with security policies, documenting existing threats, and deterring individuals from violating security policies IDPSes have become a necessary addition to the security infrastructure of nearly every organization.
IDPSes typically record information related to observed events, notify security administrators of important observed events, and produce reports. Many IDPSes can also respond to a detected threat by attempting to prevent it from succeeding. They use several response techniques, which involve the IDPS stopping the attack itself, changing the security environment (e.g., reconfiguring a firewall), or changing the attack's content.
Conditional Random Fields:
In what follows, X is a random variable over data sequences to be labeled, and Y is a random variable over
corresponding label sequences. All components Y_i of Y are assumed to range over a finite label alphabet Y. For example, X might range over natural language sentences and Y range over part-of-speech taggings of those sentences, with Y the set of possible part-of-speech tags. The random variables X and Y are jointly distributed, but in a discriminative framework we construct a conditional model p(Y|X) from paired observation and label sequences, and do not explicitly model the marginal p(X). Definition. Let G = (V,E) be a graph such that $Y = (Y_v)_{v2V}$, so that Y is indexed by the vertices of G. Then (X,Y) is a conditional random field in case, when conditioned on X, the random variables Yv obey the Markov property with respect to the graph: $p(Y_v|X,Y_w,w 6= v) = p(Y_v|X,Y_w,w \_ v)$, where w _ v means that w and v are neighbors in G.
Thus, a CRF is a random field globally conditioned on the observation X. Throughout the paper we tacitly assume that the graph G is fixed. In the simplest and most impor- 3Weighted determinization and minimization techniques shift transition weights while preserving overall path weight (Mohri, 2000); their connection to this discussion deserves further study.
tant example for modeling sequences, G is a simple chain or line: G = (V = {1, 2, . . .m},E = {(i, i + 1)}).
X may also have a natural graph structure; yet in general it is not necessary to assume that X and Y have the same graphical structure, or even that X has any graphical structure at all. However, in this paper we will be most concerned with sequences $X = (X_1,X_2, \ldots,X_n)$ and $Y = (Y_1,Y_2, \ldots,Y_n)$. If the graph G = (V,E) of Y is a tree (of which a chain is the simplest example), its cliques are the edges and vertices. Therefore, by the fundamental theorem of random fields where x is a data sequence, y a label sequence, and y|S is the set of components of y associated with the vertices in subgraph S.We assume that the features fk and gk are given and fixed. For example, a Boolean vertex feature gk might be true if the word X_i is upper case and the tag Y_i is "proper noun." The parameter estimation problem is to determine the parameters _ = (_1, _2, . . . ; $\mu_1, \mu^2, \ldots$) from training data $D = \{(x(i), y(i))\}N_i$ =1 with empirical distribution ep(x, y). We describe an iterative scaling algorithm that maximizes the log-likelihood objective function ep(x, y) log p_(y | x). As a particular case, we can construct an HMM-like CRF by defining one feature for each state pair (y_0, y), and one feature for each state-observation pair (y, x): $f_{y0,y}(\langle u, v\rangle, y|_{\langle u,v\rangle}, x) = \_(y_u, y_0)\ \_(y_v, y)\ g_{y,x}(v, y|_v, x) = \_(y_v, y)\ \_(x_v, x)$. The corresponding parameters $\_{y0,y}$ and $\mu_{y,x}$ play a similar role to the (logarithms of the) usual HMM parameters $p(y_0 | y)$ and p(x|y). Boltzmann chain models (Saul & Jordan, 1996; MacKay, 1996) have a similar form but use a single normalization constant to yield a joint distribution, whereas CRFs use the observation-dependent normalizationZ(x) for conditional istributions. Although it encompasses HMM-like models, the class of conditional random fields is much more expressive, because it allows arbitrary dependencies on the observation

III. Conclusions

The dual problem of Accuracy and Efficiency for building robust and efficient intrusion detection systems. CRFs are very effective in improving the attack detection rate and decreasing the FAR. Having a low FAR is very important for any intrusion detection system. Further selection and implementing the Layered Approach significantly reduce the time required to train and test the model. even though we used a relational data set for our experiments. The areas for future research include the use of our method for extracting features that can aid in the development of signatures for signature-based systems. The signature-based systems can be deployed at the periphery of a network to filter out attacks that are frequent and previously known, leaving the detection of new unknown attacks for anomaly and hybrid systems. Sequence analysis methods such as the CRFs when applied to relational data give us the opportunity to employ the Layered Approach and to implement pipelining of layers in multi core processors, which is likely to result in very high performance.

References

[1] A.McCallum,"Efficiently Inducing Features of Conditional Random Fields".

[2] J.P.Anderson,"Computer security Threat Monitoring and Surveillance".

[3] C.Sutton, A.McCallum,"An Introduction to Conditional Random fields for Relational Learning".

[4] V. R. Borkar, K. Deshmukh, S. Sarawagi,"Automatic text segmentation for extracting structured records", In Proc. ACM SIGMOD International Conf. on Management of Data, Santa Barabara, USA, 2001.

[5] A. Borthwick, J. Sterling, E. Agichtein, R. Grishman, "Exploiting diverse knowledge sources via maximum entropy in named entity recognition", In Sixth Workshop on Very Large Corpora New Brunswick, New Jersey. Association for Computational Linguistics., 1998.

[6] R. Bunescu, R. J. Mooney,"Relational markov networks for collective information extraction", In Proceedings of the ICML-2004 Workshop on Statistical Relational Learning (SRL- 2004), Banff, Canada, July 2004.

[7] M. E. Califf, R. J. Mooney,"Bottom-up relational learning of pattern matching rules for information extraction", Journal of Machine Learning Research, Vol. 4, pp. 177–210, 2003.

[8] W. W. Cohen, P. Ravikumar, S. E. Fienberg,"A comparison of string distance metrics for name-matching tasks", In Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web (IIWeb-03), 2003.

[9] W. W. Cohen, S. Sarawagi,"Exploiting dictionaries in named entity extraction: Combining semi-markov extraction processes and data integration methods", In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004.

[10] M. Collins,"Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms", In Empirical Methods in Natural Language Processing (EMNLP), 2002.

[11] X. Ge.,"Segmental Semi-Markov Models and Applications to Sequence Analysis", Ph.D thesis, University of California, Irvine, December 2002.

Mr. ShaikRiaz,Received his B.E in CSE from AnnaUniversity,Chennai,T.N.India and M.Tech in CSE from Acharya Nagarjuna University Guntur,A.P.India. He was a lecturer,Assistant Professor and currently working as an Associate professor with the Dept of CSE,St.Mary's group of institutions,Guntur.His research interests include InformationSecurity and DataMining.

Mr. K.V.S.S.RamaKrishna is an excellent teacher with vast experience,received his M.C.A from Achayra Nagarujuna University, Guntur,A.P,India and M.Tech in CSE from AcharyaNagarjunaUniversity,Guntur,A.P,India.He is working as an Assistant Professor with the Dept of CSE in GVR&S group of institutions ,Guntur.His Research interests include Information Security,Image processing and Theory of Computation.