Security Evaluation of Pattern Classifiers under Attack

P. Gnana Pavan¹ M.Tech CSE K. Venkatesh-Asst.Professor ² V. Rajesh-Asst. professor³
Prasad V. Potluri Siddhartha Institute of Science and Technology, Kanuru, Vijayawada, A.P-520007, Autonomous, Accredited by NAAC, All UG Programs are Provisionally Accredited by NBA, Permanently Affiliated to JNTU Kakinada

Abstract: The systems which can be used for pattern classification are used in adversarial application, for example spam filtering, network intrusion detection system, biometric authentication. This adversarial scenario’s exploitation may sometimes affect their performance and limit their practical utility. In case of pattern classification conception and contrive methods to adversarial environment is a novel and relevant research direction, which has not yet pursued in a systematic way. To address one main open issue: evaluating at contrive phase the security of pattern classifiers (for example the performance degradation under potential attacks which incurs during the operation). To propose a framework for evaluation of classifier security and also this framework can be applied to different classifiers on one of the application from the spam filtering, biometric authentication and network intrusion detection.

Keywords: Machine learning system, Security evaluation, Adversarial classification, Arms Race, Spam Filtering.

Introduction

In Pattern classification systems machine learning algorithms are used to perform security-related applications like biometric authentication, network intrusion detection, and spam filtering, to distinguish between a “legitimate” and a “malicious” pattern class. The input data can be purposely manipulated by an adversary to make classifiers to produce false negative. This often gives rise to an arms race between the adversary and the classifier designer. Well known examples of attacks are: Spoofing attacks where one person or program purposely falsifying data and thereby gaining an illegitimate advantage[1][2],modifying network packets belonging to intrusive traffic manipulating contents of emails[3],modifying network packets belonging to intrusive traffic. Mainly three main open issues are identified: (i) analyzing the vulnerabilities of classification algorithms, and the corresponding attacks (ii) developing novel methods to assess classifier security against these attacks (iii) developing novel design methods to guarantee classifier security in adversarial environments.

Machine learning is used to prevent illegal or unsanctioned activity which are created from adversary. Machine learning is used in security related tasks involving classification, such as intrusion detection systems, spam filters, biometric authentic action. Measuring the security performance of these classifiers is an essential part for facilitating decision making. The security in Machine Learning Systems besides of spam filtering (spam-mails) and network intrusion detection systems that is NIDS. The Machine learning systems have been employed in different number of applications which contains Online Deputy Systems (ODS), Clump Supervising (cluster monitoring), and toxin detection same as virus detection and some dynamic operations applications. There are some algorithms with accurate performance in the case of adversarial condition like Secure Learning Algorithms [2]. Some Classifiers are utilized to generate some contrasts which promote security intention. For example, the intention of a toxin (virus) detection system is to diminish vulnerabilities.

The toxins (virus) give antecedent to contamination or by detecting the contamination. An adversary’s attempt to procure the data which are nothing but the domestic state of a Machine Learning System (MLS) to-(i) infuse the personal data which is encrypted in its domestic state otherwise (ii) originate the data which sanction the adversary to effectually onslaught the system

Literature Survey

Unsolicited commercial email is a significant problem for users and providers of email services. While statistical spam filters have proven useful, senders of spam are learning to bypass these filters by systematically modifying their email messages. In a good word attack, a spammer modifies a spam message by inserting or appending words indicative of legitimate email. We describe and evaluate the effectiveness of active and passive good word attacks against two types of statistical spam filters: naïve Bayes and maximum entropy filters[4]. Spoof attacks consist in submitting fake biometric traits to biometric systems, and this
is a major threat in security. Multi-modal biometric systems are commonly used in spoof attacks. Multimodal biometric systems for personal identity recognition is very useful from past few years. It has been shown that combining information coming from different biometric traits can overcome the limits and the weaknesses inherent in every individual biometric, resulting in a higher accuracy[1][2]. Intrusion detection systems analyze network traffic to prevent and detect malicious activities like intrusion attempts, port scans, and denial-of-service attacks. When suspected malicious traffic is detected, an alarm is raised by the IDS and subsequently handled by the system administrator. Two main kinds of IDSs exist: misuse detectors and anomaly-based ones. Data refers both to the data used by the learning algorithm during classifier design, coming from D(where D is data set), and to the data collected during operation to retrain the classifier through online learning algorithms. “Testing” data refers both to the data drawn from D to evaluate classifier performance during design, and to the data classified during operation. We propose an algorithm to sample training (TR) and testing (TS) sets of any desired size from the distributions. Training and Test sets have been obtained from distribution using a classical resampling technique like cross validation or bootstrapping. Security evaluation is carried out by averaging the performance of the trained and tested data.

“R.N. Rodrigues, L.L. Ling, and V. Govindaraju Proposed [1] that, we address the security of multimodal biometric systems when one of the modes is successfully spoofed. We propose two novel fusion schemes that can increase the security of multimodal biometric systems. The first is an extension of the likelihood ratio based fusion scheme and the other uses fuzzy logic. Besides the matching score and sample quality score, our proposed fusion schemes also take into account the intrinsic security of each biometric system being fused. Experimental results have shown that the proposed methods are more robust against spoof attacks when compared with traditional fusion methods [1]. “P. Johnson, B. Tan, and S. Shockers” Proposed [2] that biometric systems, the threat of “spoofing”, where an imposter will fake a biometric trait, have led to the increased use of multimodal biometric systems. It is assumed that an imposter must spoof all modalities in the system to be accepted. This paper looks at the cases where some but Not all modalities are spoofed. The contribution of this paper is to outline a method for assessment of multimodal systems and underlying fusion algorithms. The framework for this method is described and experiments are conducted on a multimodal database of face, iris, and fingerprint match scores[2].

Methodology

Construction of Training (TR) and Testing (TS):
Generation of training and test data sets from gathered data is an important task in developing a classifier with high generation ability. Reassembling techniques are used in statistical analysis, are used for model selection by estimating the classification performance of classifiers. Reassembling techniques are used for estimating statistics such as the mean and the median by randomly selecting data from the given data set, calculating statistics on that data and repeating above procedure many times. “Training”

Architecture

Fig.1. A conceptual representation in arm Race in adversarial classification

(a) The classical “reactive” arm race

b) The “proactive” arm race

In “reactive” arms race, the designer reacts to the attack by analyzing the attack’s effects and developing countermeasures. In “proactive” arms race, the designer tries to anticipate the adversary by
simulating potential attacks, evaluating their effects and developing countermeasures if necessary. We summarize the three main concepts in our framework for security evaluation: 1) Arms race and security by design: since it is not possible to predict how many and which kinds of attacks a classifier will incur during operation, classifier security should be proactively evaluated using a what-if analysis, by simulating potential attack scenarios. 2) Adversary modeling: effective simulation of attack scenarios requires a formal model of the adversary. 3) Data distribution under attack: the distribution of testing data may differ from that of training data, when the classifier is under attack.

Conclusion and Future Work

This paper presented an overview of work related to the security of pattern classification systems with the goal of imparting useful guidelines on how to improve their design and assess their security specific attacks. Also the paper focused on innovative security evaluation of pattern classifiers that deployed in adversarial environments. Main contribution is a framework for verifiable security evaluation that construes and establishes the notion from previous work, and can be utilized to different classifiers, learning algorithms, and classification tasks. In the future, clustering methods can be integrated with the existing technique in order to get better results. Further, this approach can be applied to the application which makes the classification problem highly non-stationary.

REFERENCES


Author-1:
P. Gnana Pavani is presently pursuing M.Tech in Dept of CSE
In PVPST, Vijayawada. She received her B.Tech in 2015 from Loyola Institute of Technology & Management. The research interests are Network Security, Cloud Computing and Operating Systems

Author-2:
K. Venkatesh is currently working as Assistant Professor in Dept of CSE, PVPST, Vijayawada, A.P. and his research activities are Networking, Network Security, Mobile Computing, Cloud Computing.

Author-3:
V. Rajesh is currently working as Assistant Professor in Dept. Of CSE, PVPST, Vijayawada, A.P. and his research activities are Networking, Network Security, Mobile Computing, Cloud Computing.