ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Implementation of Pattern Classifiers under Attack Using Security Evaluation

Rupali B. Navalkar, Prof. Rajeshri R. Shelke

M. E Second Yr (CSE) H.V.P.M's COET, Amravati, S.G.B Amt University, Maharashtra, India
Assistant Professor, ME (CSE), Ph.D (Pursuing) H.V.P.M's COET, Amravati, S.G.B Amt University, Maharashtra, India

ABSTRACT: Pattern classification is a branch of machine learning that focuses on recognition of patterns and regularities in data. This Pattern classification system are commonly used in adversarial applications, like biometric authentication, network intrusion detection, and spam filtering, in which data can be purposely manipulated by humans to undermine their operation. As this adversarial scenario is not taken into account by classical design methods, pattern classification systems may exhibit vulnerabilities, whose exploitation may severely affect their performance, and consequently limit their practical utility. Extending pattern classification theory and design methods to adversarial settings is thus a novel and very relevant research direction, which has not yet been pursued in a systematic way. In this paper, we propose a framework for empirical evaluation of classifier security that formalizes and generalizes the main ideas proposed in the literature, and give examples of its use in real applications. Reported results show that security evaluation can provide a more complete understanding of the classifier's behavior in adversarial environments, and lead to better design choices. This framework can be applied to different classifiers on one of the application from the spam filtering, biometric authentication and network intrusion detection. Considering Multimodal system, that the proposed methodology to rank score fusion rules is capable of providing correct ranking of score fusion rules under spoof attack. So in this we propose an algorithm for the generation of training and testing sets to be used for security evaluation.

KEYWORDS: Pattern classification, Adversarial classification, performance evaluation, security evaluation.

I. INTRODUCTION

Pattern classification systems based on machine learning algorithms are commonly used in security-related applications like biometric authentication, network intrusion detection, and spam filtering, to discriminate between a "legitimate" and a "malicious" pattern class. Contrary to traditional ones, these applications have an intrinsic adversarial nature since the input data can be purposely manipulated by an intelligent and adaptive adversary to undermine classifier operation. Well known examples of attacks against pattern classifiers are: submitting a fake biometric trait to a biometric authentication system (spoofing attack), modifying network packets belonging to intrusive traffic to evade intrusion detection systems (IDSs), manipulating the content of spam emails to get them past spam filters (e.g., by misspelling common spam words to avoid their detection). It is now acknowledged that, since pattern classification systems based on classical theory and design methods do not take into account adversarial settings, they exhibit vulnerabilities to several potential attacks, allowing adversaries to undermine their effectiveness. A systematic and unified treatment of this issue is thus needed to allow the trusted adoption of pattern classifiers in adversarial environments.

We propose a framework for empirical evaluation of classifier security that formalizes and generalizes the main ideas proposed in the literature, and give examples of its use in three real applications. Reported results show that security evaluation can provide a more complete understanding of the classifier's behavior in adversarial environments, and lead to better design choices. This framework can be applied to different classifiers on one of the application from the spam filtering, biometric authentication and network intrusion detection. Considering Multimodal system, we address the security of multimodal biometric systems when one of the modes is successfully spoofed. that the proposed methodology to rank score fusion rules is capable of providing correct ranking of score fusion rules under spoof attack.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Spoofing attacks where one person or program purposely falsifying data and there by gaining an illegitimate advantage. We discuss how the classical design cycle of pattern classifiers should be revised to take security into account. Finally, we summarize our contributions, the limitations of our framework, and some open issues.

II. RELATED WORK

Here we review previous &related work, Pattern classification systems based on machine learning algorithms are commonly used in security-related applications like biometric authentication, network intrusion detection, and spam filtering, to discriminate between a "legitimate" and a "malicious" pattern class[1][2]. Pattern classification systems based on classical theory and design methods do not take into account adversarial settings. They exhibit vulnerabilities to several potential attacks, allowing adversaries to undermine their effectiveness [3]. Biometric systems have been found to be useful tools for person identification and verification. A biometric characteristic is any physiological of behavioural trait of a person that can be used to distinguish that person from other people [2][6]. Spoof attacks consist in submitting fake biometric traits to biometric systems [2][4], and this is a major threat in security.

The presence of an intelligent and adaptive adversary makes the classification problem highly non-stationary[1], and makes it difficult to predict how many and which kinds of attacks a classifier will be subject to during operation, that is, how the data distribution will change. In particular, the testing data processed by the trained classifier can be affected by both exploratory and causative attacks, while the training data can only be affected by causative attacks, In both cases, during operation, testing data may follow a different distribution than that of training data, when the classifier is under attack. Therefore, security evaluation cannot be carried out according to the classical paradigm of performance evaluation [1][2][3].

Security problems often lead to a "reactive" arms race between the adversary and the classifier designer [2]. At each step, the adversary analyzes the classifier defences, and develops an attack strategy to overcome them [1]. Many authors implicitly performed security evaluation as a what-if analysis, based on empirical simulation methods; they mainly focused on a specific application, classifier and attack, their goal was either to point out a previously unknown vulnerability, or to evaluate security against a known attack.

III. EXISTING SYSTEM

Pattern classification systems based on classical theory and design methods do not take into account adversarial settings, they exhibit vulnerabilities to several potential attacks, allowing adversaries to undermine their effectiveness. A systematic and unified treatment of this issue is thus needed to allow the trusted adoption of pattern classifiers in adversarial environments, starting from the theoretical foundations up to novel design methods, extending the classical design cycle of. In particular, three main open issues can be identified: (i) Analyzing the vulnerabilities of classification algorithms, and the corresponding attacks. (ii) Developing novel methods to assess classifier security against these attacks, which is a not possible using classical performance evaluation method. (iii) Developing novel design methods to guarantee classifier security in adversarial environments.

IV. PROPOSED SYSTEM

In this work we address issues above by developing a framework for the empirical evaluation of classifier security at design phase that extends the model selection and performance evaluation steps of the classical design cycle .We summarize previous work, and point out three main ideas that emerge from it. We then formalize and generalize them in our framework. First, to pursue security in the context of an arms race it is not sufficient to react to observed attacks, but it is also necessary to proactively anticipate the adversary by predicting the most relevant, potential attacks through a what-if analysis; this allows one to develop suitable countermeasures before the attack actually occurs, according to the principle of security by design. Second, to provide practical guidelines for simulating realistic attack scenarios, we define a general model of the adversary, in terms of her goal, knowledge, and capability, which encompass and generalize models proposed in previous work. Third, since the presence of carefully targeted attacks may affect the distribution of training and testing data separately, we propose a model of the data distribution that can formally

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

characterize this behavior, and that allows us to take into account a large number of potential attacks; we also propose an algorithm for the generation of training and testing sets to be used for security evaluation, which can naturally accommodate application-specific and heuristic techniques for simulating attacks.

a) **Advantages of proposed system**

1. Proposed system prevents developing novel methods to assess classifier security against these attacks.

2.The presence of an intelligent and adaptive adversary makes the classification problem highly non-stationary.

b) **Objectives**

1. Our future work will be devoted to develop techniques for simulating attacks for different applications.

2. Prevents developing novel methods to assess classifier security against these attacks.

3. The presence of an intelligent and adaptive adversary makes the classification problem highly non- stationary.
4. We also propose an algorithm for the generation of training and testing sets to be used for security evaluation, which can naturally accommodate application-specific and heuristic techniques for simulating attacks**.**

V. MODULES

a) Attack Scenario and Model of the Adversary
b) Pattern Classification
c) Adversarial classification:
d) Security modules

a). Attack Scenario and Model of the Adversary

The definition of attack scenarios is ultimately an application-specific issue, it is possible to give general guidelines that can help the designer of a pattern recognition system. Here we propose to specify the attack scenario in terms of a conceptual model of the adversary that encompasses, unifies, and extends different ideas from previous work. Our model is based on the assumption that the adversary acts rationally to attain a given goal, according to her knowledge of the classifier, and her capability of manipulating data. This allows one to derive the corresponding optimal attack strategy.

b). Pattern Classification

Multimodal biometric systems for personal identity recognition have received great interest in the 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. Moreover, it is commonly believed that multimodal systems also improve security against Spoofing attacks, which consist of claiming a false identity and submitting at least one fake biometric trait to the system (e.g., a “gummy” fingerprint or a photograph of a user’s face). The reason is that, to evade multimodal system, one expects that the adversary should spoof all the corresponding biometric traits.

In this application example, we show how the designer of a multimodal system can verify if this hypothesis holds, before deploying the system, by simulating spoofing attacks against each of the matchers.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

c). **Adversarial Classification**

Assume that a classifier has to discriminate between legitimate and spam emails on the basis of their textual content, and that the bag-of-words feature representation has been chosen, with binary features denoting the occurrence of a given set of words.

d). **Security Modules**

Intrusion detection systems analyze network traffic to prevent and detect malicious activities like intrusion attempts, biometric system under a simulated spoof attack against the fingerprint or the face matcher. 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. Misuse detectors match the analyzed network traffic against a database of signatures of known malicious activities. The main drawback is that they are not able to detect never-before-seen malicious activities, or even variants of known ones. To overcome this issue, anomaly-based detectors have been proposed. They build a statistical model of the normal traffic using machine learning techniques, usually one-class classifiers, and raise an alarm when anomalous traffic is detected.

Their training set is constructed, and periodically updated to follow the changes of normal traffic, by collecting unsupervised network traffic during operation, assuming that it is normal (it can be filtered by a misuse detector).

Fig. 1: A conceptual representation of the arms race in adversarial classification.

The "proactive" arms race advocated in this paper. The designer tries to anticipate the adversary by simulating potential attacks, evaluating their effects, and developing countermeasures if necessary. Detailed guidelines require one to take into account application-specific constraints and adversary models**.**

VI. CONCLUSION AND FUTURE WORK

In this paper we focused on empirical security evaluation of pattern classifiers that have to be deployed in adversarial environments, and proposed how to revise the classical performance evaluation design step. In this paper the main contribution is a framework for empirical security evaluation that formalizes and generalizes ideas from previous work, and can be applied to different classifiers, learning algorithms and classification tasks An intrinsic limitation of our work is that security evaluation is carried out empirically, and it is thus data dependent; on the other hand, model-driven analyses require a full analytical model of the problem and of the adversary's behaviour, that may be very difficult to develop for real-world applications. Another intrinsic limitation is due to fact that our method is not

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

application-specific, and, therefore, provides only high-level guidelines for simulating attacks. Indeed, detailed guidelines require one to take into account application-specific constraints and adversary models..

REFERENCES

[1] Battista Biggio, Member, IEEE , Giorgio Fumera, Member, IEEE, and Fabio Roli, Fellow, IEEE ," Security Evaluation of Pattern Classifiers under Attack", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 4, APRIL 2014.
[2] S.P.Mohana Priya[1], S.Pothumani , "Identifying Security Evaluation of Pattern Classifiers Under attack", International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization)Vol. 4, Issue 3, March 2015.
[3] Kale Tai. , Prof. Bere S. S., "A Survey on: Security Evaluation of Pattern Classifiers under Attack", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 11, November 2015.
[4]Yadigar Imamverdiyev , Lala Karimova, Vugar Musayev , James Wayman, " TESTING BIOMETRIC SYSTEMS AGAINST SPOOFING ATTACKS" , The Second International Conference "Problems of Cybernetics and Informatics" September 10-12, 2008, Baku, Azerbaijan.
[5] Tanisha Aggarwal, Dr. ChanderKant Verma, "Spoofing Technique for Fingerprint Biometric system" ,IJSRD - International Journal for Scientific Research & Development| Vol. 2, Issue 03, 2014 .
[6] Arun Ross and Anil K. Jain,"MULTIMODAL BIOMETRICS: AN OVERVIEW" , Appeared in Proc. of 12th European Signal Processing Conference (EUSIPCO), (Vienna, Austria), pp. 1221-1224, September 2004.
[7] R.N. Rodrigues, L.L. Ling, and V. Govindaraju, "Robustness of Multimodal Biometric Fusion Methods against Spoof Attacks" ,J. Visual Languages and Computing, vol. 20, no. 3, pp. 169-179, 2009.
[8] D.B. Skillicorn,"Adversarial Knowledge Discovery," IEEE Intelligent Systems, vol. 24, no. 6, Nov./Dec. 2009.
[9] Shaik Zeeshan, Md. Amanatulla," Implementation of Security Evaluation of Pattern Classifiers under Attack", IJCSIET-- International Journal of Computer Science information and Engg., Technologies ISSN 2277-4408.
[10]. R. R. Shelke ,Dr. V. M. Thakare, Dr. R . V. Dharaskar, "Study of Data Mining Approach for Mobile Computing Environment",International Journal on Computer Science and Engineering (IJCSE) , ISSN : 0975-3397, Vol. 4, 12 Dec 2012 , pp.1920-1923.
[11] D. Fetterly, "Adversarial Information Retrieval: The Manipulation of Web Content," ACM Computing Rev., 2007.
[12] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification. Wiley-Interscience Publication, 2000.