ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Novel Framework for Empirical Evaluation of Classifier Security

P.Divya Vani
M.Tech Student,
Department of CSE,
Vijaya Krishna Institute of Technology and Sciences,
Hyderabad.

B.Suman
Associate professor,
Department of CSE,
Vijaya Krishna Institute of Technology and Sciences,
Hyderabad.

Abstract:

Pattern classification is a branch of machine learning that focuses on recognition of patterns and regularities in data. In adversarial applications like biometric authentication, spam filtering, network intrusion detection the pattern classification systems are used. Pattern classification systems may exhibit vulnerabilities if adversarial scenario is not taken into account. Multimodal biometric systems are more robust to spoofing attacks, as they combine information coming from different biometric traits. In this paper, we assess the security of pattern classifiers that formalizes and generalizes the main ideas proposed in the literature and give examples of its use in three real applications. We put forward a framework for evaluation of pattern security, model of adversary for defining any attack scenario. 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.

Keywords:

Adversarial classification, adversarial scenario, pattern classification, security evaluation, machine learning.

Introduction:

The terms pattern recognition, machine learning, data mining and knowledge discovery in databases (KDD) are hard to separate, as they largely overlap in their scope. Machine learning is the common term for supervised learning methods and originates from artificial intelligence, whereas KDD and data mining have a larger focus on unsupervised methods and stronger connection to business use. Pattern recognition has its origins in engineering, and the term is popular in the context of computer vision: a leading computer vision conference is named Conference on Computer Vision and Pattern Recognition. In pattern recognition, there may be a higher interest to formalize, explain and visualize the pattern; whereas machine learning traditionally focuses on maximizing the recognition rates. Yet, all of these domains have evolved substantially from their roots in artificial intelligence, engineering and statistics; and have become increasingly similar by integrating developments and ideas from each other.In machine learning, pattern recognition is the assignment of a label to a given input value. An example of pattern recognition is classification, which attempts to assign each input value to one of a given set of classes (for example, determine whether a given email is "spam" or "non-spam"). However, pattern recognition is a more general problem that encompasses other types of output as well. Other examples are regression, which assigns a real-valued output to each input; sequence labeling, which assigns a class to each member of a sequence of values (for example, part of speech tagging, which assigns a part of speech to each word in an input sentence); and parsing, which assigns a parse tree to an input sentence, describing the syntactic structure of the sentence. Pattern recognition algorithms generally aim to provide a reasonable answer for all possible inputs and to perform "most likely" matching of the inputs, taking into account their statistical variation. This is opposed to pattern matching algorithms, which look for exact matches in the input with pre-existing patterns. A common example of a pattern-matching algorithm is regular expression matching, which looks for patterns of a given sort in textual data and is included in the search capabilities of many text editors and word processors. In contrast to pattern recognition, pattern matching is generally not considered a type of machine learning, although pattern-matching algorithms (especially with fairly general, carefully tailored patterns) can sometimes succeed in providing similar-quality output to the sort provided by pattern-recognition algorithms.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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 not possible using classical performance evaluation methods . (iii) developing novel design methods to guarantee classifier security in adversarial environments .

DISADVANTAGES OF EXISTING SYSTEM:

1.Poor analyzing the vulnerabilities of classification algorithms, and the corresponding attacks

2.A malicious webmaster may manipulate search engine rankings to artificially promote her1 website.

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 (Section 3). 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 encompasses and generalizes 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 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.

ADVANTAGES OF PROPOSED SYSTEM:

1.Prevents developing novel methods to assess classifier security against these attack.

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

Architecture:

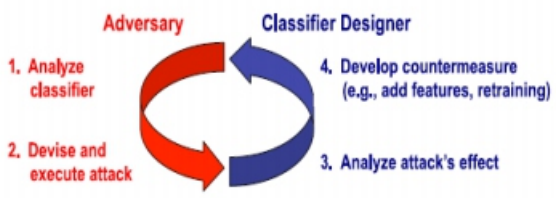
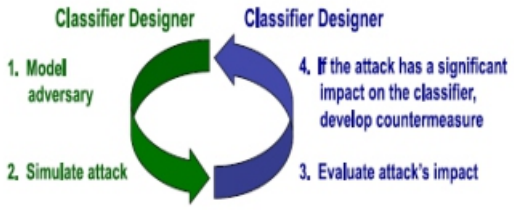


Fig.1. A conceptual representation in arm race in adversarial classification (a)The classical "reactive" arm race



(b)The "proactive" arm race.

FEATURES:

1.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.

2.Multimodal biometric systems for personal identity recognition have received great interest in the past few years.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

as 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'sface).

PROBLEM STATEMENT:

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.

MODULE DESCRIPTION:
Number of Modules:

After careful analysis the system has been identified to have the following modules:

1.Pattern classification Modules
2. Adversarial classification Modules
3.security Modules
4.Performance Modules

PATTERN CLASSIFICATION MODULES:

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 verifyifthishypothesisholds,beforedeployingthesystem,bysimulatingspoofingattacksagainsteach of the matchers.

ADVERSARIAL CLASSIFICATION MODULES :

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.

SECURITY MODULES:

Intrusion detection systems analyze network traffic to pre-vent and detect malicious activities like intrusion attempts,ROC curves of the considered multimodal biometric system under a simulated spoof attack against the fingerprint or the face matcher. port scans, and denial-of-service attacks.11 When suspected malicious traffic is detected, an alarm is raised by the IDSand 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 (e.g., Snort).12 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 (e.g., PAYL [49]),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, and should).

PERFORMANCE MODULES:

the performance is usually measured in terms of genuine acceptance rate (GAR) and false acceptance rate (FAR), respectively the fraction of genuine and impostor attempts that are accepted as genuine by the system. We use here the complete ROC curve, which shows the GAR as Under the above model selection setting (two classifiers, and four feature subsets) eight different classifier mode ls must be evaluated. Each model is trained on TR. SVMs are implemented with the Lib SV Ms Software The C parameter of their learning algorithm is chosen by maximizing theAUC10 percent through a 5-fold cross-validation on TR. An online gradient descent algorithm is used for LR.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Conclusion:

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 performanceevaluation 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. It is grounded on a formal model of the adversary, and on a model of data distribution that can represent all the attacks considered in previous work; provides a systematic method for the generation of training and testing sets that enables security evaluation and can accommodate application specific techniques for attack simulation.

References:

[1] Biggio, B., Fumera, G. ; Roli, F, Security Evaluation of Pattern Classifiers under Attack, Knowledge and Data Engineering, IEEE Transactions on Biometrics Compendium, IEEE (Volume:26 , Issue: 4).

[2] D. Lowd and C. Meek, "Good word attacks on statistical spam filters," in 2nd Conf. on Email and Anti-Spam, CA, USA, 2005.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. Wiley-Interscience Publication, 2000.

[4] A. Kolcz and C. H. Teo, "Feature weighting for improved classifier robustness," in 6th Conf. on Email and Anti-Spam, CA, USA, 2009.

[5] M. Barreno, B. Nelson, A. Joseph, and J. Tygar, "The security of machine learning," Machine Learning, vol. 81, pp. 121–148, 2010.

[6]. Biggio, B., Nelson, B., Laskov, P.: Poisoning attacks against support vector machines. In: Proceedings of the 29th International Conference on Machine Learning (2012).

[7] A. A. C´ardenas, J. S. Baras, and K. Seamon, "A framework for the evaluation of intrusion detection systems," in Proc. IEEE Symp. On Security and Privacy. DC, USA: IEEE CS, 2006, pp. 63–77.

[8]Fogla, P., Sharif, M., Perdisci, R., Kolesnikov, O., Lee, W.: Polymorphic blending attacks. In: Proceedings of the 15th Conference on USENIX Security Symposium (2006).

[9]. Cauwenberghs, G., Poggio, T.: Incremental and decremental support vector machine learning. In: T.K. Leen, T.G. Dietterich,V. Tresp (eds.) NIPS, pp. 409–415. MIT Press (2000).

[10] S. Rizzi, "What-if analysis," Enc. of Database Systems, pp. 3525– 3529, 2009.

[11] Huang, L., Joseph, A.D., Nelson, B., Rubinstein, B., Tygar, J.D.: Adversarial machine learning. In: Proceedings of the 4th ACMWorkshop on Artificial Intelligence and Security (AISec), pp. 43–57 (2011).

[12]Lowd, D., Meek, C.: Adversarial learning. In: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 641–647 (2005) .