Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 6, June 2015, pg.266 – 270

SURVEY ARTICLE

SURVEY ON SECURITY EVALUATION OF PATTERN CLASSIFIER UNDER ATTACK

Avinash I Hatti

M.Tech Student. CSE Dept.

Bellary Institute of Technology &

Management, BELLARY

Email-avinashh.12@gmail.com

P.Phani Ram Prasad

Associate Prof., CSE Dept.

Bellary Institute of Technology &

Management, BELLARY

ABSTRACT- Pattern classification systems 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. In this paper, we address one of the main open issues: evaluating at design phase the security of pattern classifiers, namely, the performance degradation under potential attacks they may incur during operation. Reported results show that security evaluation can provide a more complete understanding of the classifier's behaviour in adversarial environments, and lead to better design choices.

Keywords: Pattern Classification, performance evaluation, security Evaluation, robustness evaluation.

1. Introduction

Pattern classification systems 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 address one of the main open issues: evaluating at design phase the security of pattern classifiers, namely the performance degradation under potential attacks they may incur during operation. 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 behaviour in adversarial environments, and lead to better design choices Adversarial scenarios can also occur in intelligent data analysis and information retrieval . 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, 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.

2. Literature Survey

1. Robustness of multi-model biometric verification systems under realistic spoofing attack

Recent works have shown that multi-model biometric systems are not robust against spoofing attacks. However, this conclusion has been obtained under the hypothesis of a “worst case” attack, where the attacker is able to replicate perfectly the genuine biometric traits. Aim of this paper is to analyse the robustness of some multi-modal verification systems, combining fingerprint and face biometrics, under realistic spoofing attacks, in order to investigate the validity of the results obtained under the worst-case attack assumption.

2. Adversarial Information retrieval: the manipulation of web content

In recent years several tools based on statistical methods and machine learning have been incorporated in security related tasks involving classification, such as intrusion detection systems (IDSs), fraud detection, spam filters, biometrics and multimedia forensics. Measuring the security performance of these classifiers is an essential part for facilitating decision making, determining the viability of the product, or for comparing multiple classifiers. There are however relevant considerations for security related problems that are sometimes ignored by traditional evaluation schemes. In this paper we identify two pervasive problems in security related applications. The first problem is the usually large class imbalance between normal events and attack events. This problem has been addressed by evaluating classifiers based on cost-sensitive metrics and with the introduction of Bayesian Receiver Operating Characteristic (B-ROC) curves. The second problem to consider is the fact that the classifier or learning rule will be deployed in an adversarial environment. This implies that good performance on average might not be a good performance measure, but rather we look for good performance under the worst type of adversarial attacks. In order to address this notion more precisely we provide a framework to model an adversary and define security notions based on evaluation metrics.

3. Features weighting for improved classifier robustness

There are often discrepancies between the learning sample and the evaluation environment, be it natural or adversarial. It is therefore desirable that classifiers are robust, i.e., not very sensitive to changes in data distribution. In this paper, we introduce a new methodology to measure the lower bound of classifier robustness under adversarial attack and show that simple averaged classifiers can improve classifier robustness significantly. In addition, we propose a new feature reweighting technique that rates the performance and robustness of standard classifiers at most twice the computational cost. We verify our claims in content based email spam classification experiments on some public and private datasets.

4. Multimodal fusion vulnerability to non-zero effort (spoof) imposter

In biometric systems, the threat of "spoofing", where an imposter will fake a biometric trait, has lead 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.

Conclusion

The current applicants which are using the Adversarial applications such as Biometric Authentication, Network Intrusion Detection, Spam Filtering in which data can be purposely manipulated by humans to undermine their operations. They don't know the ongoing things such as attacks on these data by Intruder. This can be resolved by evaluating the security of pattern so that the applicant gets benefited.

References:

[1] R. N. Rodrigues, L. L. Ling, and V. Govindaraju, "Robustness of multimodal biometric fusion methods against spoof attacks," J. Vis. Lang. Comput., vol. 20, no. 3, pp. 169–179, 2009.

[2] P. Johnson, B. Tan, and S. Schuckers, "Multimodal fusion vulnerability to non-zero effort (spoof) imposters," in IEEE Int'l Workshop on Inf. Forensics and Security, 2010, pp. 1–5.

[3] P. Fogla, M. Sharif, R. Perdisci, O. Kolesnikov, and W. Lee, “Polymorphic blending attacks,” in Proc. 15th Conf. on USENIX Security Symp. CA, USA: USENIX Association, 2006.

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

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

[6] D. Fetterly, “Adversarial information retrieval: The manipulation of web content,” ACM Computing Reviews, 2007.