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

# Security Check of Pattern Classification System

**B.Prasad**
HOD,
Department CSE,
Sri Chaitanya Technical Cmapus,
Ibrahimpatnam.

**B.Prasad**
Associate Professor,
Department CSE,
Sri Chaitanya Technical Cmapus,
Ibrahimpatnam.

**Sneha James**
P.G Scholar,
Department CSE,
Sri Chaitanya Technical Cmapus,
Ibrahimpatnam.

## 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. 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. We propose 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, performance evaluation, security evaluation.

## I.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. 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. This often gives rise to an arms race between the adversary and the classifier designer. Well known examples of attacks against pattern classifiers are:

submitting a fake biometric trait to a biometric authentication system (spoofing attack) [1], [2]; 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. Adversarial machine learning is a research field that lies at the intersection of machine learning and computer security. It aims to enable the safe adoption of machine learning techniques in adversarial settings like spam filtering, malware detection and biometric recognition. Examples include: attacks in spam filtering, where spam messages are obfuscated through misspelling of bad words or insertion of good words; attacks in computer security, e.g., to obfuscate malware code within network packets or mislead signature detection; attacks in biometric recognition, where fake biometric traits may be exploited to impersonate a legitimate user (biometric spoofing) or to compromise users' template galleries that are adaptively updated over time.[16] To understand the security properties of learning algorithms in adversarial settings, one should address the following main issues:

identifying potential vulnerabilities of machine learning algorithms during learning and classification; devising appropriate attacks that correspond to the identified threats and evaluating their impact on the targeted system; Proposing countermeasures to improve the security of machine learning algorithms against the considered attacks. 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. A few key aspects of a human physiological or behavioural trait that make for a strong biometric for recognition are universality, distinctiveness, permanence, and Collectability. 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

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

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.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 are 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.

These ensure that the trait is available from all people, is adequately variable among all people, does not change significantly over time, and is reasonably able to be measured. The problem with any human trait that meets these criteria is in the performance, acceptability, and circumvention of the biometric feature. Performance is an issue resulting mainly from the combination of lack of variability in the biometric trait, noise in the sensor data due to environmental factors, and robustness of the matching algorithm. Acceptability indicates how willing the client pool will be to use the biometric identifier regularly. Circumvention is the possibility of a non-client (impostor) getting past the system using deceptive methods. The key to creating a secure multimodal biometric system is in how the information from the different modalities is fused to make a final decision. There are two different categories of fusion schemes for multiple classifiers; rule based and supervised based. Supervised methods, on the other hand, require training but can often provide better results than the rule based methods. For example, a fusion strategy using a support vector machine (SVM) was able to out-perform a fusion algorithm using the sum rule.

Introducing a quality measure into a fusion algorithm is one method that has been used to boost performance in multibiometric systems. If for instance, a more secure biometric of high quality gives a low match score and a less secure biometric gives a high match score, then there is a high likelihood of a spoof attack.

It is commonly understood that one of the strengths of a multimodal system is in its ability to accommodate for noisy sensor data in an individual modality. In contrast, a more secure algorithm, in order to address the issue of a spoof attack on a partial subset of the biometric modalities, must require adequate performance in all modalities. This type of algorithm would invariably negate, to some extent, the contribution of a multimodal system to performance in the presence of noisy sensor data. A multimodal system improves the performance aspect but increases the security only slightly since it is still vulnerable to partial spoof attacks. Enhanced fusion methods which utilize approaches to improve security will again suffer decreased performance when presented with noisy Data.The support vector machine (SVM) is a exercise procedure for knowledge organization and reversion rubrics after statistics, for instance the SVM can be recycled to study polynomial, circular foundation purpose (RBF) then multi-layer perception (MLP) classifiers SVMs remained chief optional by Vapnik in the 1960s for organization to develop a part of penetrate in Investigate on owed to growths in the methods plus philosophy joined with postponements to reversion and Thickness approximation.

SVMs ascended after arithmetical knowledge philosophy the goal existence to resolve separate the problematic of attention deprived of resolving additional problematic as a middle stage. SVMs are founded on the physical threat minimisation code, carefully connected to regular inaction philosophy. This belief joins volume switch to stop over-fitting and therefore is ain complete response to the bias-variance trade-off quandary. Binary key rudiments in the application of SVM are the methods of precise software design and seed purposes. The limits are originated by resolving a quadratic software design problematic with direct parity and disparity restraints; slightly than by resolving a non-convex, unimpeded optimisation problem. The suppleness of seed purposes lets the SVM to exploration an extensive diversity of theory places.The geometrical clarification of support vector classification (SVC) is that the procedure pursuits for the best unravelling superficial, i.e. the hyper plane that is, in an intelligence, intermediate after the binary courses.

## III. PREVIOUS WORK ON SECURITY EVALUATION:

Previous work in adversarial learning can be categorized according to the two main steps of the proactive arms race described in the previous section.

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

The first research direction focuses on identifying potential vulnerabilities of learning algorithms and assessing the impact of the corresponding attacks on the targeted classifier; e.g[4, 5, 18, 36, 40, 41, 42, 46]. The second explores the development of proper countermeasures and learning algorithms robust to known attacks.e.g [26, 41, 57].Although some prior work does address aspects of the empirical evaluation of classifier security, which is often implicitly defined as the performance degradation incurred under a (simulated) attack, to our knowledge a systematic treatment of this process under a unifying perspective was only first described in our recent work [12]. Previously, security evaluation is generally conducted within a specific application domain such as spam filtering and network intrusion detection (e.g., [26, 31, 41, 47, 66]), in which a different application-dependent criteria is separately defined for each endeavour. Security evaluation is then implicitly undertaken by defining an attack and assessing its impact on the given classifier. For instance, in [31], the authors showed how camouflage network packets can mimic legitimate traffic to evade detection; and, similarly, in [26, 41, 47, 66], the content of spam emails was manipulated for evasion.

Although such analyses provide indispensable insights into specific problems, their results are difficult to generalize to other domains and provide little guidance for evaluating classifier security in a different application. Thus, in a new application domain, security evaluation often must begin anew and it is difficult to directly compare with prior studies. This shortcoming highlights the need for a more general set of security guidelines and a more systematic definition of classifier security evaluation that we began to address in [12].Apart from application-specific work, several theoretical models of adversarial learning has been proposed [4, 17, 26, 36, 40, 42, 46, 53]. These models frame the secure learning problem and provide a foundation for a proper security evaluation scheme. In particular, we build upon elements of the models of [4, 5, 36, 38, 40, 42], which were used in defining our framework for security evaluation [12]. Below we summarize these foundations.

## IV.SPAM FILTERING OVERVIEW:

Over the past few years, spam filtering software has gained popularity due to its relative accuracy and ease of deployment. With its roots in text classification research, spam filtering software seeks to answer the question "Whether the message x is spam or not?".

The means by which this question is addressed varies upon the type of classification algorithm in place. While the categorization method differs between statistical filters, their basic functionality is similar. The basic model is often known as the bag of words (multinomial) or multivariate model. Essentially, a document is distilled into a set of features such as words, phrases, meta-data, etc.This set of features can then be represented as a vector whose components are Boolean (multivariate) or real values (multinomial). One should note that with this model the ordering of features is ignored. Classification algorithm uses the feature vector as a basis upon which the document is judged. The usage of the feature vector varies between classification methods. As the name implies, rule based methods classify documents based on whether or not they meet a particular set of criteria. Machine learning algorithms are primarily driven by the statistics (e.g. word frequency) that can be derived from the feature vectors. One of the widely used methods, Bayesian classification, attempts to calculate the probability that a message is spam based upon previous feature frequencies in spam and legitimate e-mail.

## V.SPAM FILTERING:

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. This kind of classifier has been considered by several authors [6], [12], [13], and it is included in several real spam filters. In this example, we focus on model selection. We assume that the designer wants to choose between a support vector machine (SVM) with a linear kernel, and a logistic regression (LR) linear classifier. He also wants to choose a feature subset, among all the words occurring in training emails. A set D of legitimate and spam emails is available for this purpose. We assume that the designer wants to evaluate not only classifier accuracy in the absence of attacks, as in the classical design scenario, but also its security against the well-known bad word obfuscation (BWO) and good word insertion (GWI) attacks. They consist of modifying spam emails by inserting "good words" that are likely to appear in legitimate emails, and by obfuscating "bad words" that are typically present in spam [6]. The attack scenario can be modelled as follows. Attack scenario: Goal. The adversary aims at maximizing the percentage of spam emails misclassified as legitimate, which is an indiscriminate integrity violation.

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

Knowledge: As in [6], [10], the adversary is assumed to have perfect knowledge of the classifier, i.e., (k.ii) the feature set, (k.iii) the kind of decision function, and (k.iv) its parameters (the weight assigned to each feature, and the decision threshold). Assumptions on the knowledge of (k.i) the training data and (k.v) feedback from the classifier are not relevant in this case, as they do not provide any additional information.Capability: We assume that the adversary: (c.i) is only able to influence testing data (exploratory attack); (c.ii) cannot modify the class priors; (c.iii) can manipulate each malicious sample, but no legitimate ones; (c.iv) can manipulate any feature value (i.e., she can insert or obfuscate any word), but up to a maximum number n max of features in each spam email [6], [10].

## VI.PATTERN RECOGNITION:

Pattern recognition is a branch of machine learning that focuses on the recognition of patterns and regularities in data, although it is in some cases considered to be nearly synonymous with machine learning. Pattern recognition systems are in many cases trained from labelled "training" data (supervised learning), but when no labelled data are available other algorithms can be used to discover previously unknown patterns (unsupervised learning).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. In statistics, discriminate analysis was introduced for this same purpose in 1936.

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 labelling, 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.

Pattern recognition is studied in many fields, including psychology, psychiatry, ethnology, cognitive science, traffic flow and computer science.Pattern recognition is generally categorized according to the type of learning procedure used to generate the output value. Supervised learning assumes that a set of training data (the training set) has been provided, consisting of a set of instances that have been properly labelled by hand with the correct output. A learning procedure then generates a model that attempts to meet two sometimes conflicting objectives: Perform as well as possible on the training data, and generalize as well as possible to new data (usually, this means being as simple as possible, for some technical definition of "simple", in accordance with Occam's Razor, discussed below). Unsupervised learning, on the other hand, assumes training data that has not been hand-labelled, and attempts to find inherent patterns in the data that can then be used to determine the correct output value for new data instances.[2] A combination of the two that has recently been explored is semi-supervised learning,

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

which uses a combination of labelled and unlabeled data (typically a small set of labelled data combined with a large amount of unlabeled data). Note that in cases of unsupervised learning, there may be no training data at all to speak of; in other words, the data to be labelled is the training data.

## VII.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 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 differentclassifiers, 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 [12], [10]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 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.Attar, A., Rad, R.M., Atani, R.E.: A survey of image spamming and filtering techniques. Artif. Intell. Rev. 40(1), 71{105 (2013)

2.Barreno, M., Nelson, B., Sears, R., Joseph, A.D., Tygar, J.D.: Can machine learning be secure? In: Proc. of the 2006 ACM Symp. on Information, Computer and Comm. Sec. pp. 16{25. ACM, NY, USA (2006)

3.Biggio, B., Akhtar, Z., Fumera, G., Marcialis, G.L., Roli, F.: Security evaluation of biometric authentication systems under real spoofing attacks. IET Biometrics 1(1), 11{24 (2012)

4.Biggio, B., Didaci, L., Fumera, G., Roli, F.: Poisoning attacks to compromise face templates. In: 6th IAPR Int'l Conf. on Biometrics. pp. 1{7. (2013)

5.Biggio, B., Fumera, G., Pillai, I., Roli, F.: A survey and experimental evaluation of image spam filtering techniques. Pattern Rec. Letters 32(10), 1436 {1446 (2011)

6.Biggio, B., Fumera, G., Roli, F.: Security evaluation of pattern classiers under attack. IEEE Trans. on Knowledge and Data Engineering 99(PrePrints), 1 (2013)

7.D. Lowd and C. Meek, "Good Word Attacks on Statistical Spam Filters," Proc. Second Conf. Email and Anti-Spam, 2005

8.P. Fogla, M. Sharif, R. Perdisci, O. Kolesnikov, and W. Lee,"Polymorphic Blending Attacks," Proc. 15th Conf. USENIX Security Symp., 2006

9.P. Johnson, B. Tan, and S. Schuckers, "Multimodal Fusion Vulnerability to Non-Zero Effort (Spoof) Imposters," Proc. IEEE Int'l Workshop Information Forensics and Security, pp. 1-5, 2010

10.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