On Learning and Recognition of Secure Patterns

Invited Keynote

Battista Biggio
Università di Cagliari
Piazza d'Armi
09123, Cagliari, Italy
battista.biggio@diee.unica.it

ABSTRACT

Learning and recognition of secure patterns is a well-known problem in nature. Mimicry and camouflage are widely-spread techniques in the arms race between predators and preys. All of the information acquired by our senses is therefore not necessarily secure or reliable. In machine learning and pattern recognition systems, we have started investigating these issues only recently, with the goal of learning to discriminate between secure and hostile patterns. This phenomenon has been especially observed in the context of adversarial settings like biometric recognition, malware detection and spam filtering, in which data can be adversely manipulated by humans to undermine the outcomes of an automatic analysis. As current pattern recognition methods are not natively designed to deal with the intrinsic, adversarial nature of these problems, they exhibit specific vulnerabilities that an adversary may exploit either to mislead learning or to avoid detection. Identifying these vulnerabilities and analyzing the impact of the corresponding attacks on pattern classifiers is one of the main open issues in the novel research field of adversarial machine learning.

In the first part of this talk, I introduce a general framework that encompasses and unifies previous work in the field, allowing one to systematically evaluate classifier security against different, potential attacks. As an example of application of this framework, in the second part of the talk, I discuss evasion attacks, where malicious samples are manipulated at test time to avoid detection. I then show how carefully-designed poisoning attacks can mislead learning of support vector machines by manipulating a small fraction of their training data, and how to poison adaptive biometric verification systems to compromise the biometric templates (face images) of the enrolled clients. Finally, I briefly discuss our ongoing work on attacks against clustering algorithms, and sketch some possible future research directions.

CCS'14, November 3–7, 2014, Scottsdale, Arizona, USA.
ACM 978-1-4503-2957-6/14/11.
http://dx.doi.org/10.1145/2666652.2666653.

Categories and Subject Descriptors

I.5.0 [**Pattern recognition**]: General; G.3 [**Probability and Statistics**]: Statistical computing; I.5.1 [**Models**]: Statistical; D.4.6 [**Security and Protection**]: Invasive software (e.g., viruses, worms, Trojan horses); I.5.3 [**Clustering**]: Algorithms

Keywords

Secure Pattern Recognition; Adversarial Machine Learning; Evasion Attacks; Poisoning Attacks

Short Biography

Battista Biggio received the M. Sc. degree in Electronic Eng., with honors, and the Ph. D. in Electronic Eng. and Computer Science, respectively in 2006 and 2010, from the University of Cagliari, Italy. Since 2007 he has been working for the Dept. of Electrical and Electronic Eng. of the same University, where he holds now a postdoctoral position. In 2011, he visited the University of Tübingen, Germany, and worked on the security of machine learning algorithms to contamination of training data. His research interests currently include: secure machine learning and pattern recognition methods, multiple classifier systems, kernel methods, biometric authentication, spam filtering, and computer security. He serves as a reviewer for several international conferences and journals, including Pattern Recognition and Pattern Recognition Letters. Dr. Biggio is a member of the IEEE (Computer Society, and Systems, Man and Cybernetics Society), and of the Italian Group of Italian Researchers in Pattern Recognition (GIRPR), affiliated to the IAPR.

1. REFERENCES

[1] M. Barreno, B. Nelson, A. Joseph, and J. Tygar. The security of machine learning. *Machine Learning*, 81:121–148, 2010.

[2] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar. Can machine learning be secure? In *Proc. ACM Symp. Information, Computer and Comm. Sec.*, ASIACCS '06, pages 16–25, New York, NY, USA, 2006. ACM.

[3] B. Biggio, S. R. Bulò, I. Pillai, M. Mura, E. Z. Mequanint, M. Pelillo, and F. Roli. Poisoning complete-linkage hierarchical clustering. In P. Franti, G. Brown, M. Loog, F. Escolano, and M. Pelillo, editors, *Joint IAPR Int'l Workshop on Structural, Syntactic, and Statistical Patt. Recogn.*, volume 8621 of *LNCS*, pages 42–52, Joensuu, Finland, 2014. Springer Berlin Heidelberg.

[4] B. Biggio, I. Corona, D. Maiorca, B. Nelson, N. Šrndić, P. Laskov, G. Giacinto, and F. Roli. Evasion attacks against machine learning at test time. In H. Blockeel, K. Kersting, S. Nijssen, and F. Železný, editors, *European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Part III*, volume 8190 of *LNCS*, pages 387–402. Springer Berlin Heidelberg, 2013.

[5] B. Biggio, I. Corona, B. Nelson, B. Rubinstein, D. Maiorca, G. Fumera, G. Giacinto, and F. Roli. Security evaluation of support vector machines in adversarial environments. In Y. Ma and G. Guo, editors, *Support Vector Machines Applications*, pages 105–153. Springer International Publishing, 2014.

[6] B. Biggio, L. Didaci, G. Fumera, and F. Roli. Poisoning attacks to compromise face templates. In *6th IAPR Int'l Conf. on Biometrics (ICB 2013)*, pages 1–7, Madrid, Spain, 2013.

[7] B. Biggio, G. Fumera, and F. Roli. Multiple classifier systems for robust classifier design in adversarial environments. *Int'l J. Mach. Learn. and Cybernetics*, 1(1):27–41, 2010.

[8] B. Biggio, G. Fumera, and F. Roli. Security evaluation of pattern classifiers under attack. *IEEE Trans. on Knowl. and Data Eng.*, 26(4):984–996, April 2014.

[9] B. Biggio, G. Fumera, and F. Roli. Pattern recognition systems under attack: Design issues and research challenges. *Int'l J. Patt. Recogn. Artif. Intell.*, 2014, In press.

[10] B. Biggio, G. Fumera, F. Roli, and L. Didaci. Poisoning adaptive biometric systems. In G. Gimel'farb, E. Hancock, A. Imiya, A. Kuijper, M. Kudo, S. Omachi, T. Windeatt, and K. Yamada, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 7626 of *LNCS*, pages 417–425. Springer Berlin Heidelberg, 2012.

[11] B. Biggio, B. Nelson, and P. Laskov. Poisoning attacks against support vector machines. In J. Langford and J. Pineau, editors, *29th Int'l Conf. on Machine Learning*, pages 1807–1814. Omnipress, 2012.

[12] B. Biggio, I. Pillai, S. R. Bulò, D. Ariu, M. Pelillo, and F. Roli. Is data clustering in adversarial settings secure? In *Proc. 2013 ACM Workshop on Artificial Intell. and Security*, AISec '13, pages 87–98, New York, NY, USA, 2013. ACM.

[13] L. Huang, A. D. Joseph, B. Nelson, B. Rubinstein, and J. D. Tygar. Adversarial machine learning. In *4th ACM Workshop on Artif. Intell. and Sec.*, AISec '11, pages 43–57, Chicago, IL, USA, October 2011.

[14] M. Kloft and P. Laskov. Online anomaly detection under adversarial impact. In *Proc. 13th Int'l Conf. on Artif. Intell. and Statistics*, pages 405–412, 2010.

[15] M. Kloft and P. Laskov. Security analysis of online centroid anomaly detection. *Journal of Machine Learning Research*, 13:3647–3690, 2012.

[16] D. Maiorca, I. Corona, and G. Giacinto. Looking at the bag is not enough to find the bomb: an evasion of structural methods for malicious pdf files detection. In *Proc. 8th ACM SIGSAC Symp. on Information, Computer and Comm. Sec.*, ASIA CCS '13, pages 119–130, New York, NY, USA, 2013. ACM.

[17] F. Roli, B. Biggio, and G. Fumera. Pattern recognition systems under attack. In J. Ruiz-Shulcloper and G. S. di Baja, editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 8258 of *LNCS*, pages 1–8. Springer, 2013.

[18] B. I. Rubinstein, B. Nelson, L. Huang, A. D. Joseph, S.-h. Lau, S. Rao, N. Taft, and J. D. Tygar. Antidote: understanding and defending against poisoning of anomaly detectors. In *Proc. 9th ACM SIGCOMM Internet Measurement Conf.*, IMC '09, pages 1–14, New York, NY, USA, 2009. ACM.

[19] N. Šrndić and P. Laskov. Detection of malicious pdf files based on hierarchical document structure. In *Proc. 20th Annual Network & Distributed System Security Symposium*. The Internet Society, 2013.