Logic, Machine Learning, and Security

V. S. Subrahmanian(✉)

Department of Computer Science,
Institute for Security, Technology, and Society,
Dartmouth College, Hanover, NH 03755, USA
vs@dartmouth.edu

Abstract. Logic stands at the very heart of computer science. In this talk, I will argue that logic is also an essential part of machine learning and that it has a fundamental role to play in both international security and counter-terrorism. I will first briefly describe the use of logic for high-level reasoning in counter-terrorism applications and then describe the BEEF system to explain the forecasts generated by virtually any machine learning classifier. Finally, I will describe one use of logic in deceiving cyber-adversaries who may have successfully compromised an enterprise network.

Keywords: Logic · Deception · Counter-terrorism · Machine learning · AI · Cybersecurity

1 Introduction

In this talk, I will describe the role of logic in 3 broad areas: the use of logic in counter-terrorism applications, the use of logic to explain the results generated by very diverse and potentially very complex machine learning classification algorithms, and the role of logic in deceiving malicious hackers who may have successfully entered an enterprise network.

2 Logic for Counter-Terrorism

Since approximately 2004, my research group (then at the University of Maryland College Park) and I have worked on the problem of predicting the behaviors of terrorist groups and reshaping their behavior when the forecasts of their behavior were not to our liking. The key aspect of our work was to develop forecasts that were: (i) accurate, and (ii) easily explainable to policy makers. As policy makers are drawn from diverse backgrounds ranging from lawyers to social scientists to business people, the explanations had to be both easy to grasp and compelling. We turned immediately to probabilistic logic programs [6] and temporal probabilistic logic programs [3]. We wrote the first ever paper on computational predictive models in counter-terrorism. The paper, about Hezbollah,

A. Cuzzocrea et al. (Eds.): FQAS 2019, LNAI 11529, pp. 3–6, 2019.
https://doi.org/10.1007/978-3-030-27629-4_1

presented probabilistic rules about Hezbollah's behavior [7] that were simple enough for journalists and Hezbollah to understand—so much so that Hezbollah even issued a comment to the Beirut Daily Star about the paper on Oct 22 2008[1]. The fact that Hezbollah could understand our paper gave us the confidence to believe that we were on the right track—and later, we were able to develop the first ever thorough study of terrorist group behavior by analyzing Lashkar-e-Taiba, the terrorist group that carried out the infamous 2008 Mumbai attacks [11]. This was quickly followed by a similar study of the Indian Mujahideen [10] using temporal probabilistic rules. Our group put out several live forecasts of the behaviors of these two groups which were mostly correct. We subsequently developed methods to reshape the behaviors of these groups and formulate policies against them. For instance, in citesimari2013parallel, we showed that a form of abduction could be used to generate policies that would reduce—with maximal possible probability—the different types of attacks that the group would carry out. Later, we showed how to combine temporal probabilistic rules and game theoretic reasoning to show that strategically disclosing the behavioral rules we had learned about the groups could help reshape the action of the group to help deter/influence them [8,9].

3 Logic for Explaining Forecasts Generated by Machine Learning Classifiers

More recently, the field of "explainable" machine learning has become very important. Machine learning classifiers such as support vector machines [2] and ensemble classifiers such as random forest [1] often generate highly accurate forecasts, but explaining them in plain English can be a major challenge. In the second part of my talk, I will describe a system called BEEF (Balanced English Explanation of Forecasts) developed by us [4]. Given any machine learning algorithm (in a black box) and given a forecast F made by that algorithm, BEEF introduces the concept of a balanced explanation. A balanced explanation consists of arguments both *for* and *against* the forecast. The need for balanced explanations was motivated by my prior work on counter-terrorism where I was repeatedly asked to provide explanations for both why the forecasts we made were correct as well as to explain why they may be incorrect. We show that the problem of generating balanced explanations has both a geometric and a logical interpretation. We built out a prototype system and ran experiments showing that BEEF provides intuitive explanations that were deemed more compelling by human subjects than other methods.

4 Logic for Deceiving Cyber-Adversaries

Today, most enterprises are aware that they need to be ready to be the target of cyber-attacks. When malicious hackers successfully enter a network, they often

[1] http://www.dailystar.com.lb/News/Lebanon-News/2008/Oct-22/54721-us-academics-design-software-to-predict-hizbullah-behavior.ashx.

move laterally in the network by scanning nodes in the network, understanding what kinds of vulnerabilities exist in the scanned nodes, and then move through the network by exploiting those vulnerabilities. As they move from node to node, they may carry out a host of malicious activities ranging from reconnaissance and surveillance to exfiltration of data or intellectual property, to planting malware and backdoors, or carrying out denial of service attacks. I will discuss one way to disrupt the hacker's ability to damage an enterprise even after the enterprise has been compromised. We introduce the idea of generating fake scan results [5] that lead a hacker away from the crown jewels of an enterprise and minimize the expected damage caused by the hacker.

5 Conclusion

This talk describes results generated by my research group along with several students, postdocs, and colleagues. Our work shows that logic is a rich and fertile mechanism for helping humans understand the behavior of both humans and programs and has demonstrated the potential to help secure us—both in the physical world and in cyberspace.

Acknowledgement. Different parts of this work were supported by ONR grants N000141612739, N00014-16-1-2918, N00014-18-1-2670 and N00014-16-1-2896.

References

1. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
2. Cortes, C., Vapnik, V.: Support vector machine. Mach. Learn. **20**, 273–297 (1995)
3. Dekhtyar, A., Dekhtyar, M.I., Subrahmanian, V.: Temporal probabilistic logic programs. In: ICLP, vol. 99, pp. 109–123 (1999)
4. Grover, S., Pulice, C., Simari, G.I., Subrahmanian, V.: BEEF: balanced English explanations of forecasts. IEEE Trans. Comput. Soc. Syst. **6**(2), 350–364 (2019)
5. Jajodia, S., et al.: A probabilistic logic of cyber deception. IEEE Inf. Forensics Secur. **12**(11), 2532–2544 (2017)
6. Khuller, S., Martinez, M.V., Nau, D., Sliva, A., Simari, G.I., Subrahmanian, V.S.: Computing most probable worlds of action probabilistic logic programs: scalable estimation for 10 30,000 worlds. Ann. Math. Artif. Intell. **51**(2–4), 295–331 (2007)
7. Mannes, A., Michael, M., Pate, A., Sliva, A., Subrahmanian, V.S., Wilkenfeld, J.: Stochastic opponent modeling agents: a case study with Hezbollah. In: Liu, H., Salerno, J.J., Young, M.J. (eds.) Social Computing, Behavioral Modeling, and Prediction, pp. 37–45. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-77672-9_6
8. Serra, E., Subrahmanian, V.: A survey of quantitative models of terror group behavior and an analysis of strategic disclosure of behavioral models. IEEE Trans. Comput. Soc. Syst. **1**(1), 66–88 (2014)
9. Simari, G.I., Dickerson, J.P., Sliva, A., Subrahmanian, V.: Parallel abductive query answering in probabilistic logic programs. ACM Trans. Comput. Logic (TOCL) **14**(2), 12 (2013)

10. Subrahmanian, V.S., Mannes, A., Roul, A., Raghavan, R.: Indian Mujahideen: Computational Analysis and Public Policy. TESECO. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-02818-7
11. Subrahmanian, V.S., Mannes, A., Sliva, A., Shakarian, J., Dickerson, J.P.: Computational Analysis of Terrorist Groups: Lashkar-e-Taiba. Springer, New York (2012). https://doi.org/10.1007/978-1-4614-4769-6