AFRL-AFOSR-UK-TR-2022-0059

Using Deep Reinforcement Learning to Simulate Security Analyst

Pevny, Tomas
CZECH TECHNICAL UNIVERSITY IN PRAGUE
ZIKOVA 4
PRAGUE 6, , 166 36
CZ

06/03/2022
Final Technical Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
European Office of Aerospace Research and Development
Unit 4515 Box 14, APO AE 09421

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
20220603	Final	START DATE 20180901	END DATE 20211231

4. TITLE AND SUBTITLE
Using Deep Reinforcement Learning to Simulate Security Analyst

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
	FA9550-18-1-7008	61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Tomas Pevny

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CZECH TECHNICAL UNIVERSITY IN PRAGUE
ZIKOVA 4
PRAGUE 6 166 36
CZ

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
EOARD UNIT 4515 APO AE 09421-4515	AFRL/AFOSR IOE	AFRL-AFOSR-UK-TR-2022-0059

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The goal of this project was to overcome conventional limitations of deep learning approaches and advance the use of machine learning approaches to understand, identify, and analyze security incidents. In general, main-stream classifiers require a full description of the sample (in our scenario this means all possible information about the security incident) and perform the classification in one step, which is in a sharp contrast to the modus operandi of the analyst, whose investigation is composed of a sequence of actions and decisions.

During the performance of this grant, the PI (Tomáš Pevný) and fellow researchers (Viliam Lisý and Jaromír Janischand) leveraged recent reinforcement learning advancements to outperform classical heuristic solutions on a wide range of problems. They also demonstrated the flexibility of the proposed reinforcement learning approach allowing them to directly optimize their results with respect to either average budgeted resources or hard budget constraints per sample.

Notably, the team developed a Hierarchical Multiple Instance Learning (HMIL) framework that enabled them to address samples stored in hierarchical formats with a variable number of actions. They demonstrated their framework using it to solve seven different problems ranging from medical data to the investigation of security incidents. They further extended the approach to domains where one sample is a general graph with vertices of different types -- Vertices of different types can be described by different features and offer different actions to an agent. This yielded a very flexible and generalized framework that was used to solve three types of problems: BoxWorld, SysAdmin, and Sokoban.

In total, this project resulted in one workshop, one conference paper, one journal publication, and two draft conference submissions. All source code is freely available as indicated in corresponding publications. With more research and development, the PIs anticipate that the framework could be used in automated penetration testing which is an important area of interest to the DoD that costs millions each year.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	SAR	2

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
LOGAN MAILLOUX	314 235 6163

Progress report: Using deep reinforcement learning to simulate security analyst

Tomáš Pevný, Viliam Lisý and Jaromír Janisch

January 21, 2021

The goal of the project was to simulate by means of a machine learning the process of investigation of a security incident by an experienced network analyst. Main-stream classifiers require a full description of sample (in our scenario this means all possible information about the security incident) and perform the classification in one step, which is in a sharp contrast to a modus operandi of the analyst, whose investigation is composed of a sequence of actions and decisions, where actions correspond to either (i) deeper investigation of some part of the incident (sample) or to (ii) a decision about the incident if sufficient information about the incident has been collected. Moreover, a rationaly behaving analyst wants to investigate the incident as accurately and quickly as possible. This problem is known in the literature as *Classification with Costly Features* (CwCF). The cost of features can be monetary, where the analyst has to pay for querying third part intelligence providers, or it can reflect the time.

CwCF problem belongs to a class of problems known as *sequential decision making problem,* which are typically solved by Reinforcement learning (RL), as was also identified in 2011 in [1]. Despite this, subsequent works rely on heuristic solutions instead of the principled one offered by RL. We have demonstrated in [3] that RL is indeed a good approach. By implementing recent innovations in the field, RL scored better than heuristic solutions on a wide range of problems. We have further demonstrated the flexibility of the RL approach by allowing to directly optimize with respect to either average budget or hard budget per sample. This was summarized in a paper published in Machine Learning Journal [2].

All above works have assumed the sample to be described by a feature vector with a known and fixed dimension, where only some features are known. This simplifies the problem since the number of actions is known and fixed for all samples and feed-forward neural networks can be used. But it suffers two drawbacks: (i) most data describing security incidents have structure and they are stored in structured file formats like XML, JSON, ProtoBuffer, or MessagePack[1] which makes them difficult to represent as a vector such that above approaches can be used; (ii) the number of actions is dynamic and it can change between samples.

To tackle these problems, we have combined the HMIL framework from [6] with the hierarchical softmax. HMIL framework enables us to cope with samples stored in hierarchical formats and hierarchical softmax allows us to cope with a variable number of actions. The approach has been described in [4], where we have also demonstrated its generality by solving seven different problems ranging from medical data to the investigation of the security incident.[2] To our knowledge, our solution is unique due to its flexibility and generality. This work [4] is now being submitted to International Joint Conference on Artificial Intelligence 2021.

In the last step, we have extended our approach to domains where one sample is a general graph with vertices of different types. Vertices of different types can be described by different features (HMIL in general) and offer different actions to an agent. This yields a very flexible and general framework, which we have demonstrated in [5] (and will be submitted to International Conference on Machine Learning 2021) by solving three problems: BoxWorld, SysAdmin, and Sokoban. We can imagine this framework to be used in automated penetration testing after some more research work.

To summarize, our motivation of mimicking the work of a real security investigator lead to a very general solution to *Classification with Costly Features* applicable outside of the originally intended domain, which has been clearly demonstrated in our publications, and can providing substantial cost savings. For example in healthcare it allows to optimize the set of required diagnostic procedures for each. Since our solution relies on Reinforcement Learning, it can benefit from innovations there. In total, our work leads to one workshop, one conference [3], and one journal [2] publication with two more publications prepared to submission to IJCAI 2021 [4] and ICML 2021 [5]. All source codes are published as indicated in corresponding publications.

References

[1] Gabriel Dulac-Arnold et al. "Datum-wise classification: a sequential approach to sparsity". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2011, pp. 375–390.

[1] We call this type of samples HMIL standing for Hierarchical Multiple Instance Learning.

[2] We have simulated the scenario of our interest, investigation of the security incident, using data from a public service `threatcrowd.org/`.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited

[2] Jaromír Janisch, Tomáš Pevný, and Viliam Lisý. "Classification with costly features as a sequential decision-making problem". In: *Machine Learning* (2020), pp. 1–29.

[3] Jaromír Janisch, Tomáš Pevný, and Viliam Lisý. "Classification with costly features using deep reinforcement learning". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 3959–3966.

[4] Jaromír Janisch, Tomáš Pevný, and Viliam Lisý. "Cost-Efficient Hierarchical Knowledge Extraction with Deep Reinforcement Learning". In: *arXiv* (2019), arXiv–1911.

[5] Jaromír Janisch, Tomáš Pevný, and Viliam Lisý. "Symbolic Relational Deep Reinforcement Learning based on Graph Neural Networks". In: *arXiv preprint arXiv:2009.12462* (2020).

[6] Tomáš Pevný and Petr Somol. "Discriminative models for multi-instance problems with tree structure". In: *Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security*. 2016, pp. 83–91.

DISTRIBUTION A: Distribution approved for public release. Distribution unlimited