PhD in Information Technology and Electrical Engineering

Università degli Studi di Napoli Federico II

PhD Student: Antonio Guerriero

XXXIV Cycle

Training and Research Activities Report – Second Year

Tutor: Stefano Russo – co-Tutor: Roberto Pietrantuono

1. Information

PhD candidate: Antonio Guerriero (mat. DR993614)
Date of birth: 06/08/1992
Master Science title: Master's degree in Computer Engineering (cum laude) on 25/05/2018, Università di Napoli Federico II
Doctoral Cycle: XXXIV
Fellowship type: UNINA
Tutor: Prof. Stefano Russo
Co-tutor: Prof. Roberto Pietrantuono
Year: Second
Publication:
- R. Pietrantuono, S.Russo, A. Guerriero, Run-time Reliability Estimation of Microservice Architectures, Proc. of the 2018 IEEE International Symposium on Software Reliability Engineering (ISSRE), Memphis, TN, USA, Oct. 15-18, IEEE, 2018. Winner of "Best Research Paper Award"
- R. Pietrantuono, S. Russo and A. Guerriero. "Testing microservice architectures for operational reliability". Software Testing, Verification and Reliability, 30(2) (2020).
- A. Guerriero, R. Mirandola, R. Pietrantuono and S. Russo, "A Hybrid Framework for Web Services Reliability and Performance Assessment," 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Berlin, Germany, 2019, pp. 185- 192.
- A. Bertolino, G. De Angelis, A. Guerriero, B. Miranda, R. Pietrantuono, S. Russo. "DevOpRET: Continuous reliability testing in DevOps". Journal of Software: Evolution and Process (2020). DOI: 10.1002/smr.2298
- A. Bertolino, A. Guerriero, B. Miranda, R. Pietrantuono, S. Russo. "Learning-to-Rank vs Ranking-to-Learn: Strategies for Regression Testing in Continuous Integration", 2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE), Seoul, Korea, July 2020.
- A. Guerriero. "Reliability evaluation of ML systems, the oracle problem". ISSRE 2020 Doctoral Symposium. In: Proc. of the 2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), pp. 127-130, IEEE.

2. Study and Training activities

In my second year I attended several Modules, Seminars, a PhD school, and an advanced course.

Lecture/Activity	Type	Hours	Credits	Dates	Organizer	Certificate
Complexity Management in the Design of Cyber-Physical Systems	Seminar	1	0.2	08/11/2019	Chinese University of Hong Kong	Yes
Intelligenza Artificiale ed Etica: La ricerca in IA alla prova delle sfide etiche	Ad hoc module	8	1.6	06/12/2019	Università di Napoli Federico II	Yes
5G PhD School	Doctoral School	23.5	3.0	02-04/12/2019	CNIT	Yes
Virtualization technologies and their applications	Ad hoc module	35	4.0	06/04/2020 – 15/05/2020	Università di Napoli Federico II	Yes
Adversarial attacks on image classifiers	Seminar	2	0.4	11/06/2020	University of Catania	Yes
3rd Advanced Course on Data Science & Machine Learning	Advanced Course	38	8.0	13-17/07/2020	Giuseppe Nicosia	Yes
IEEE Webinar How to Publish Open Access with IEEE	Seminar	1	0.2	23/08/2020	IEEE	Yes

Year 1	Year 2								Year 3	
		1	2	3	4	5	6			
Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	Total
25.9	**5**	4.6	0.0	0.0	4.0	8.0	0.0	**16.6**	**0**	**42.5**
12.8	**0**	0.2	0.0	0.0	0.4	0.0	0.2	**0.8**	**0**	**13.6**
37.0	**45**	4.0	9.0	9.0	6.0	7.0	9.0	**44.0**	**50**	**81.0**
75.7	**50**	8.8	9.0	9.0	10.4	15.0	9.2	**61.4**	**50**	**137.1**

3. Research activity

During the second year, I developed the research activities that started last year. In the following sections, I describe the new contribution related to each research activity.

3.1 Reliability and Performance Assessment of Microservices Architectures

With the research group of Barbara Russo, associate professor at Libera Università di Bolzano, we are working on a unified framework for performance and reliability assessment of microservice applications. Inspired by our previous works [1,2], the idea is to guide load testing by leveraging a specific operational profile definition. Data extracted during the testing activity provides a means to measure the reliability and performance of the microservice architecture. We aim to exploit coarse- and fine-grained metrics to provide insights to operators at the level of each microservice and considering the whole system.

3.2 Operational Testing of Deep Neural Networks

Thanks to the deep study on testing Machine Learning systems, I with my tutors, submitted a paper to ICSE 2021 about the operational testing of Machine Learning systems. In particular, the assessment of Deep Neural Networks (DNN) reliability is challenging due to the manual labeling of the operational data, whose expected output is unknown. The objective is to provide a technique able to select the most significant tests to be manually labeled to evaluate the reliability of the DNN under test, and maximizing the number of misclassifications detected.

Li *et al.* [3] proposed a sampling strategy based on an auxiliary variable extracted from the output of the last hidden layer of the DNN under test. Similarly, we propose an adaptive sampling technique based on different auxiliary variables: confidence, surprise adequacy metrics [4] (LSA and DSA), and a combination of confidence and DSA.

The results show that our approach can provide reliability estimates close to the real values as the technique proposed by Li *et al.*. Moreover, our approach is more effective in selecting operational data related to misclassifications of the DNN under test.

3.3 Reliability Evaluation of Machine Learning systems, the oracle problem

The growing adoption of machine learning (ML) also in safety-critical contexts makes reliability evaluation of ML systems a crucial task. Although testing represents one of the most used practices to evaluate the reliability of "traditional" systems, just a few techniques can be used to evaluate ML-systems' reliability due to the oracle problem. For this reason, I built a test oracle surrogate (TOS) able to automatically classify tests' outcome to obtain feedback about tests whose expected output is unknown. For this purpose, various sources of knowledge are considered to evaluate the outcome of each test. The aim is to exploit this test oracle surrogate to apply classical testing strategies to perform reliability assessment of ML systems.

A Machine Learning system (or ML system) takes in input a feature vector (f_1, …, f_n), exploited by an ML-based component to compute a response (r). At the same time, some other components of the system, not based on ML, compute additional output (a_1, …, a_n). The response and the additional output are combined to compute the final output (o).

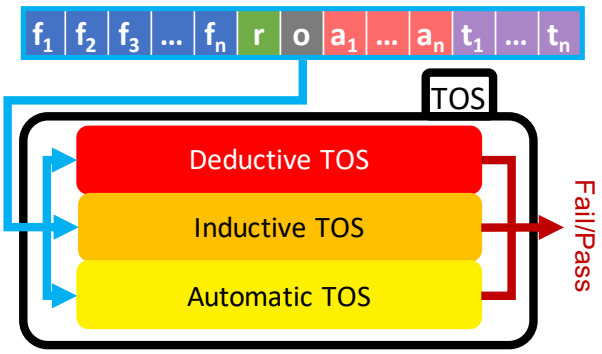

Figure 1: Test Oracle Surrogate

TOS architecture is depicted in Figure 1. In particular, I considered three levels: the *Deductive TOS* exploits domain-related constraints, and testing assumptions to classify tests; the *Inductive TOS* extracts a set of rules directly from the training set of the system under test; the *Automatic TOS* exploits the specific ML algorithm to detect patterns about how the ML system under test produces incorrect responses.

The preliminary results show that the reliability estimates obtained by exploiting the TOS are close to the real value, and they always represent an underestimate of true reliability, which represents a desired property in reliability evaluation. Further details about formulation and experimentation are reported in [5].

At the very end of the second year, I am working on a strong verticalization of the Test Oracle Surrogate in the *image classification* domain. In particular, I am working on a finer tuning and a deeper experimentation.

4. Products

During the second year I co-authored different papers and submissions along with my research group as well as with external researchers.

[P1]	R. Pietrantuono, S. Russo and A. Guerriero. "Testing microservice architectures for operational reliability". Software Testing, Verification and Reliability, 30(2) (2020).
[P2]	A. Guerriero, R. Mirandola, R. Pietrantuono and S. Russo, "A Hybrid Framework for Web Services Reliability and Performance Assessment," 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Berlin, Germany, 2019, pp. 185192.

[P3]	A. Bertolino, G. De Angelis, A. Guerriero, B. Miranda, R. Pietrantuono, S. Russo. "DevOpRET: Continuous reliability testing in DevOps". Journal of Software: Evolution and Process (2020). DOI: 10.1002/smr.2298
[P4]	A. Bertolino, A. Guerriero, B. Miranda, R. Pietrantuono, S. Russo. "Learning-to-Rank vs Ranking-to-Learn: Strategies for Regression Testing in Continuous Integration", 2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE), Seoul, Korea, July 2020.
[P5]	A. Guerriero. "Reliability evaluation of ML systems, the oracle problem". ISSRE 2020 Doctoral Symposium. In: Proc. of the 2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), pp. 127-130, IEEE.
[P6]	Stefano Russo, Roberto Pietrantuono, Antonio Guerriero. "Operation is the hardest teacher: estimating DNN accuracy looking for mispredictions". Submitted to the **43nd International Conference on Software Engineering** (ICSE 2021), Madrid, Spain, May 2021. Status: under review.

5. Conferences and Seminars

I participated to the following conferences and workshops:

Conference name	Place	Dates	Number of papers
2020 The 42nd International Conference on Software Engineering (ICSE)	Seoul, South Korea	July 6-11, 2020	293
2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Doctoral Symposium	Coimbra, Portugal	October 12, 2020	6
31st IEEE International Symposium on Software Reliability Engineering (ISSRE)	Coimbra, Portugal	October 13-15, 2020	38

As the author, I **presented** the following paper:

A. Guerriero. "Reliability evaluation of ML systems, the oracle problem". ISSRE 2020 Doctoral Symposium. In: Proc. of the 2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), pp. 127-130, IEEE. (see [P5])

6. Activity abroad

From the 2nd of November 2019 to the 13th of November, I was in Hong Kong at the Chinese University of Hong Kong (CUHK) as a visiting scholar, supervised by prof. Michael R. Lyu. I had to interrupt my period abroad due to many public disorders in Hong Kong, and also to CUHK.

For further details: https://en.wikipedia.org/wiki/Siege_of_the_Chinese_University_of_Hong_Kong

7. Tutorship

In this second year, I have been a teaching assistant for the course of Distributed Systems, a.a. 2019/2020, and I am a teaching assistant for the course of Software Engineering a.a. 2020/2021.

8. References

[1] Roberto Pietrantuono, Stefano Russo, and Antonio Guerriero. “Testing microservice architectures for operational reliability”. Software Testing, Verification and Reliability, 30(2):e1725, 2020. e1725 stvr.1725.

[2] Alberto Avritzer, Daniel Menasché, Vilc Rufino, Barbara Russo, Andrea Janes, Vincenzo Ferme, André van Hoorn, and Henning Schulz. “PPtam: Production and performance testing-based application monitoring”. In Companion of the 2019 ACM/SPEC International Conference on Performance Engineering, ICPE ’19, page 39–40, New York, NY, USA, 2019. Association for Computing Machinery.

[3] Zenan Li, Xiaoxing Ma, Chang Xu, Chun Cao, Jingwei Xu, and Jian Lu. “Boosting Operational DNN Testing Efficiency through Conditioning”. In Proc. of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ESEC/FSE, pages 499–509. ACM, 2019.

[4] Jinhan Kim, Robert Feldt, and Shin Yoo. “Guiding Deep Learning System Testing Using Surprise Adequacy”. In Proc. of the 41st Int. Conference on Software Engineering, ICSE, pages 1039–1049. IEEE, 2019.

[5] A. Guerriero. “Reliability evaluation of ML systems, the oracle problem”. ISSRE 2020 Doctoral Symposium. In: Proc. of the 2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), pp. 127-130, IEEE.