PhD in Information Technology and Electrical Engineering

Università degli Studi di Napoli Federico II

PhD Student: Rosario Catelli

XXXIII Cycle

Training and Research Activities Report – Third Year

Tutor: Prof. Valentina Casola (DIETI)

Co-Tutor: Dr. Massimo Esposito (ICAR CNR) from 15th October 2019

1. Information

I obtained my master's degree in Electronic Engineering in September 2017 at the University of Naples Federico II. My Ph.D. grant was funded until October 14, 2019 by Hitachi Rail STS (formerly Ansaldo STS), under the title "Cybersecurity tools and techniques to protect telecommunication and railway signalling systems", where I joined the CSAC (Cyber Security Assurance & Control) Department and its activities concerning (1) Policy, Guidelines and Compliance, (2) Standards and Metrics, (3) Security Incident Management, (4) Verification and Validation and (5) Internal Audits.

In October 2019 I won a research grant to conduct research activities under the PON OK-INSAID (Operational Knowledge from Insights and Analytics on Industrial Data) research program at the Institute for High Performance Computing and Networking (ICAR-CNR) where I started working on October 15, 2019.

2. Study and Training Activities

During this third year I followed these courses, external courses, and modules:

- ---

And these seminars:

- *Computational Biology: Large scale data analysis to understand the molecular bases of human diseases* (0.2 CFU) held by Prof. Michele Ceccarelli (Organizer Prof. Michele Ceccarelli).
- *How to Get Published with IEEE* (0.4 CFU) (Organizer Dr. Alessandra Scippa).
- *Large Scale Training of Deep Neural Networks* (0.5 CFU) held by Giuseppe Fiameni (Organizer Prof. Carlo Sansone).
- *SAS Analytics* (0.4 CFU) held by Dr. Cinzia Gianfiori, Dr. Costabile Santis, Dr. Daniele Goretti (Organizers Prof. Antonio Picariello and Prof. Vincenzo Moscato).
- *Science, Reality and Credibility. Talk with Nobel Laureate Saul Perlmutter.* (0.3 CFU) held by Nobel Laureate Saul Perlmutter (Organizer Prof. Mario Di Bernardo).
- *Patent Searching Best Practices with IEEE Xplore* (0.2 CFU) (Organizer Dr. Alessandra Scippa).
- *Network Systems, Kuramoto Oscillators, and Synchronous Power Flow* (0.3 CFU) held by Prof. Francesco Bullo (Organizer Dr. Marco Coraggio).

3. Research activity

I was involved until October 14, 2019 in the management and operational activities related to cybersecurity governance carried out by Hitachi Rail STS, which mainly focus on the implementation of an Information Security Management System (ISMS) for business systems and tools in accordance with ISO 27001 and GDPR (General Data Protection Regulation) with particular attention to privacy aspects.

The massive presence of activities carried out manually by humans has led me to move towards Natural Language Processing (NLP) techniques in order to process the large amount of information at stake.

In ICAR CNR, I started exploiting NLP techniques for textual analysis based both on classic approaches and deep learning. I experimented several systems for classification, in detail for Named Entity Recognition (NER) tasks with benchmark datasets, comparing their capabilities to represent words and related context. Finding an entity is a useful first step to answer questions or to link the text to information in structured knowledge sources, but many applications will also need to use specific types of entities to be defined as privacy-relevant information (date of birth, gender, etc.): finally, efforts were directed toward the use of deep learning techniques to preserve privacy. These techniques were applied to the field of clinical de-identification in which, through appropriately constructed Named Entity Recognition systems, it was possible

to achieve state-of-the-art performance. For this research, I had the pleasure of collaborating with Prof. Hamido Fujita and his laboratory at Iwate Prefectural University, Iwate, Japan.

Student: Rosario Catelli rosario.catelli@unina.it **Tutor: Valentina Casola** valentina.casola@unina.it **Cycle XXXIII**

	Credits year 1								Credits year 2								Credits year 3										
		1	2	3	4	5	6			1	2	3	4	5	6			1	2	3	4	5	6	7			
	Estimated	bimonthly	bimonthly	bimonthly	bimonthly	bimonthly	bimonthly	Summary	Estimated	bimonthly	bimonthly	bimonthly	bimonthly	bimonthly	bimonthly	Summary	Estimated	bimonthly	bimonthly	bimonthly	bimonthly	bimonthly	bimonthly	bimonthly	Summary	Total	Check
Modules	20	0	0	6	3	3	7.2	19.2	11	5	6	2	3	0	0	16	0	0	0	0	0	0	0	0	0	35.2	30-70
Seminars	5	0	0	0.6	5.2	0	0	5.8	5	0.6	0	0.2	0	1.2	0	2	2.2	0	0.6	0.9	0	0	0.8	0	2.3	10.1	10-30
Research	35	10	10	3.4	1.8	7	2.8	35	44	4.4	4	7.8	7	8.8	10	42	57.8	10	9.4	9.1	10	10	9.2	10	68	145	80-140
	60	10	10	10	10	10	10	60	60	10	10	10	10	10	10	60	60	10	10	10	10	10	10	10	70	190	180

4. Products

I published one conference paper:

- Valentina Casola and Rosario Catelli (Nov. 2020). “Semantic Management of Enterprise Information Systems through Ontologies”. In: Computer Science & Information Technology (CS & IT). AIRCC Publishing Corporation. doi: 10.5121/csit.2020.101403. url: https://doi.org/10.5121%2Fcsit.2020.101403

I published four journal papers:

- Rosario Catelli, Francesco Gargiulo, Valentina Casola, Giuseppe De Pietro, Hamido Fujita, and Massimo Esposito (2021). “A Novel COVID-19 Data Set and an Effective Deep Learning Approach for the De-Identification of Italian Medical Records”. In: IEEE Access 9, pp. 19097–19110. doi: 10.1109/access.2021.3054479. url: https://doi.org/10.1109/access.2021.3054479
- Marco Pota, Mirko Ventura, Rosario Catelli, and Massimo Esposito (Dec. 2020). “An Effective BERT-Based Pipeline for Twitter Sentiment Analysis: A Case Study in Italian”. In: Sensors 21.1, p. 133. doi: 10.3390/s21010133. url: https://doi.org/10.3390/s21010133
- Rosario Catelli, Valentina Casola, Giuseppe De Pietro, Hamido Fujita, and Massimo Esposito (Feb. 2021). “Combining contextualized word representation and sub-document level analysis through Bi-LSTM+CRF architecture for clinical de-identification”. In: Knowledge-Based Systems 213, p. 106649. doi: 10.1016/j.knosys.2020.106649. url: https://doi.org/10.1016/j.knosys.2020.106649
- Rosario Catelli, Francesco Gargiulo, Valentina Casola, Giuseppe De Pietro, Hamido Fujita, and Massimo Esposito (Dec. 2020). “Crosslingual named entity recognition for clinical de-identification applied to a COVID-19 Italian data set”. In: Applied Soft Computing 97, p. 106779. doi: 10.1016/j.asoc.2020.106779. url: https://doi.org/10.1016%2Fj.asoc.2020.106779

5. Conferences and seminars

I participated to the AIRCC Natural Language Processing 2020 Conference (NLP 2020) in November 2020, where I presented the paper “Semantic Management of Enterprise Information Systems through Ontologies” listed above.

6. Activities abroad

As reported above I worked remotely within the laboratory 'Intelligent Software Systems (Fujita Laboratory)', Iwate Prefectural University, Iwate, Japan, directed by Prof. Hamido Fujita, to study deep learning techniques to preserve privacy from August 1, 2020 to November 30, 2020.

7. Tutorship

-