PhD in Information Technology and Electrical Engineering

Università degli Studi di Napoli Federico II

PhD Student: Vincenzo Schiano Di Cola

XXIX Cycle

Training and Research Activities Report – Third Year

Tutors: prof. Nicola Mazzocca, prof. Francesco Piccialli
co-Tutor (aboard): prof. Gwanggil Jeon

1. Information

This report is written by Vincenzo Schiano Di Cola. I received my master's degree in Mathematics from this same University. I belong to the XXIX Ph.D. Cycle of the ITEE at the Università di Napoli Federico II.
I received a regional fellowship, the P.O.R. entitled: *Data Scientist for Predictive Analytics*; according to the "*Dottorati di ricerca con Caratterizzazione Industriale*" - DGR n. 156 del 21/03/2017 - DD n. 155 del 17/05/2018 - within the *POR Campania FSE 2014/2020 - Obiettivo Specifico 14 - Azione 10.4.5***.**
The tutors who are supervising me are Prof. Nicola Mazzocca and Prof. Francesco Piccialli, and in this third year, I have been additionally supervised by Prof. Gwanggil Jeon, from Incheon National University.

2. Study and Training activities

Throughout this third year, I obtained only 0,7 ECTS credits through seminars. As recommended by the ITEE guidelines, my primary focus was on the aspects of knowledge customization through research.

The seminars are listed, with their title, type, and the number of credits:

- Advances in Machine Learning, Seminar; 0,3
- Designing a Socially Assistive Robot; Seminar; 0,2
- Emotions in Reinforcement Learning Agents, Seminar; 0,2

3. Research activity

The POR project involves three primary participants, each of whom contributes unique objectives and expertise to my research within the broader context of Data Science. The DIETI at the University of Naples "Federico II," the IT company DATABOOZ ITALIA S.r.l., and the Incheon National University in South Korea are the three POR partners.

This third-year research focused on expanding previous work on Data Science applications by investigating innovative solutions and acquiring new strategies for advanced research. My research questions were as follows:

1. Which tools are most effective at recognizing a transparent object in a photograph?
2. How to extract features from photographs taken in a variety of lighting conditions?
3. How should Monte Carlo methods be applied optimally to simulate multiple diffusive models?
4. How should a Dirac delta function be approximated optimally?
5. How can PINNs be used to physically model and predict visitors' paths through a museum?

Consistent with the nature of Data Science, which is a synthesis of mathematics, information technology, and domain knowledge, the majority of collaborations have been with the Department of Mathematics and Applications (DMA) "Renato Caccioppoli," as well as other departments at the University of Naples and other foreign universities.

The following section will detail the research conducted during this third year, with an emphasis on extending research collaboration. The research can be summarized as follows: image processing, geostatistics, and Physics Informed neural networks (PINNs).

The initial research focused on applications of predictive analytics driven by chemicals. I used machine learning and image analysis to forecast the concentration of a particular contaminant (mercury) in a water sample. I developed a Python tool for recognizing well plates in images using SWIFT. Then, using machine learning techniques (XGBoost, MLP, Random Forest) on the diffused light intensity values, I forecasted the chemical contents of each well. I also studied the SWIFT algorithm formally, in detail, and provided feedback on the results and the overall architecture, by writing a report.

The second research focused on geoscience-related data science problems. I began studying Geostatistics Machine Learning, then I used a diffusion equation with sources and sinks to predict the behavior of pollution in aquifers. Using Python's gspython and fipy tools, I addressed probability estimations for thousands of pseudorandom parametrized simulations.
Additionally, I observed that by changing the approximation's support length on a non-integer scale, Dirac delta approximations behaved unexpectedly differently in the 2D case than in the 1D case. Finally, I wrote a report summarizing my findings published in [3].

The third project required me to conduct an independent review of the literature on Physics Informed Neural Networks (PINN) and Physics Constrained Neural Networks (PCNN). I reviewed over 150 papers on PINNs and related topics in physics-informed and physics-based neural networks, as well as machine learning. I discussed neural network design, partial differential equations, data processing, and the various applications and software available. Additionally, I included a Python implementation of the Schrodinger equation. Finally, I discussed the limitations and difficulties associated with PINNs, as well as future research directions.

Finally, at the end of the third year, the final month was spent writing the PhD thesis, collecting the results from the first two years and the results from the last year using biosensors.

Products

This third year of research lead to four publications:

1. “A virtual assistant in cultural heritage scenarios” on *Concurrency and Computation: Practice and Experience*. First published: 16 May 2019. DOI: 10.1002/cpe.5331
2. “The Role of Artificial Intelligence in Fighting the COVID-19 Pandemic” on *Information Systems Frontiers*. Published: 26 April 2021. DOI: 10.1007/s10796-021-10131-x
3. “Remarks on the numerical approximation of Dirac delta functions” on *Results in Applied Mathematics*. Available online: November 2021. DOI: 10.1016/j.rinam.2021.100200
4. “Scientific Machine Learning through Physics-Informed Neural Networks: Where we are and What's next” on *arXiv preprint*. Submitted online: 14 Jan 2022. DOI: 2201.05624

4. Conferences and Seminars

I presented my research at one conference and attended two other prestigious conferences.

I presented my work on Dirac delta functions and Monte Carlo Simulations, "Multi-dimensional Dirac delta Approximations for Uncertainty Quantification in Singular Source term Flows" in September at *SIMAI 2020+21*, in the mini-symposium "New Trends and Applications in Approximation Theory".

Additionally, I attended two conferences: Empirical Methods in Natural Language Processing (EMNLP 2021 - http://2021.emnlp.org/), organized by the Association for Computational Linguistics ACL, and the Thirty-fifth Neural Information Processing Systems Conference (NeurIPS 2021 - http://nips.cc/).

5. Activity abroad

Due to the COVID-19 health emergency, the research was conducted remotely (online) from Naples, Italy. All the research activities were carried out under the guidance of Prof. Gwanggil Jeon, Incheon National University (South Corea), in the framework of the research area accordingly to the PON scholarship. The remote supervision lasted an overall duration of one year, divided into three phases: from January 1st, 2021 to May 30th, 2021, then from June 21st, 2021 to August 27th, 2021, and finally from September 20th, 2021 to December 21st, 2021.

6. Tutorship

There has been no mentoring, as my POR fellowship is on a tight schedule and each hour must be reported and accounted for to the Campania Region.

	Credits year 1								Credits year 2								Credits year 3									
	Estimated	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	Summary	Estimated	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	Summary	Estimated	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	Summary	Total	Check
Modules	29	4	1,2	3	12	6	0	**26,2**	21	0	0	0	9	11	0	**20**	0							**0**	**46,2**	**30-70**
Seminars	7	0	2,5	0	1	4,8	0,4	**8,7**	5	0	0	0,4	4,9	0	0,3	**5,6**	0							**0**	**14,3**	**10-30**
Research	24	3	5	3	4	3	7,1	**25,1**	34	10	10	5,3	1,1	1	7	**34,4**	60							**0**	**59,5**	**80-140**
	60	7	8,7	6	17	13,8	7,5	**60**	60	10	10	5,7	15	12	7,3	**60**	60	0	0	0	0	0	0	**0**	**120**	**180**