PhD student Andrea Vignali

Improving security of networked systems through an NLP-based Anomaly Detection approach

Tutor: G. Sperlì

Cycle: XXXVIII

co-Tutor: S.P. Romano

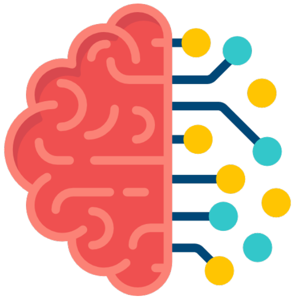
Year: First

My background

- MSc degree in Computer Engineering @ DIETI – Federico II
  - Thesis: “An active learning and similarity based augmentation approach for few-shot NER applications”
- Research group/laboratory: PICUSlab and ARCLab
- PhD start date: 01 November 2022
- Scholarship type: PNRR – DM 352
- Partner company:
  - AKKA Italia s.r.l. (former)
  - AKKODIS ITALY S.R.L.

Research field of interest

- Natural Language Processing (NLP)
  - Uses artificial intelligence to make natural language understandable and interpretable by a machine
  - Aims to analyze and elaborate text, speech and other linguistic data to extract or generate knowledge



- Anomaly Detection
  - Identifies anomalous and potentially harmful behaviors in a system or a network
  - Aims to prevent undesired events unleashed by those behaviors

Summary of study activities

- 8 ad hoc courses + 1 PhD school
  - Anomaly Detection:
    - On the challenges and impact of Artificial Intelligence in the Insurance domain
    - IoT Data Analysis
  - Networked systems:
    - CNTC (Complex networks and telecommunications 3rd edition: Towards 6G) – PhD school
  - NLP:
    - Artificial Intelligence and Natural Language Processing
  - Ethic:
    - Scienza moderna e disciplina giuridica dell'Intelligenza Artificiale
- 22 seminars
- conferences:
  - 21st IEEE Mediterranean Communication and Computer Networking Conference (MEDCOMNET2023)

Research activity: Overview (1/2)

- Problem*:*
  - *Anomalies suffer from data scarcity and are underrepresented in relation to the normal behavior of a system inside a dataset*
  - *In NLP, uncommon languages, specific domains and behaviors suffer from the same problems of data scarcity*
  - *NLP and Anomaly detection share similar problems which can be tackled in similar ways*
- *Use case (AKKA Italia s.r.l.):*
  - *Using NLP techniques to extract and represent the knowledge inferred from the committed changes in test suites and detect anomalies that can lead to failures*

Research activity: Overview (2/2)

- Methodologies*:*
  - *Data augmentation [P1; P2; P6] to cope with data scarcity*
  - *NLP analysis [P1; P2; P3; P6] for data representation and processing*
  - *Anomaly Detection [P4; P5] for models and techniques*
- *Use case (AKKA Italia s.r.l.):*
  - *Data representation: language embeddings extracted by transformers*
  - *Anomaly detection: clustering techniques and variational autoencoders*

Products

[P1]	***Learning how to augment data: an application to biomedical NER*** *– Vincenzo Moscato, Marco Postiglione, Guido Maria Secondulfo , Giancarlo Sperlì , Andrea Vignali – conference: 6th International Workshop on Knowledge Discovery from Healthcare Data (KDH-2023@IJCAI) – Published – 2023*
[P2]	***Data Augmentation via Context Similarity: an application to biomedical Named Entity Recognition*** *– Ilaria Bartolini, Vincenzo Moscato, Marco Postiglione, Giancarlo Sperlì, Andrea Vignali – journal: Information Systems – Published – 2023*
[P3]	***An NLP-Based Approach to Assessing a Company's Maturity Level in the Digital Era*** *– Simon Pietro Romano, Giancarlo Sperlì, Andrea Vignali – journal: Expert Systems With Applications– Submitted – 2023*

Products

[P4]	***CPS Security Unleashed: Anomaly Detection for Cyber-Physical Threats in Critical Infrastructures*** *– Roberto Canonico, Giovanni Esposito, Annalisa Navarro, Simon Pietro Romano, Giancarlo Sperlì, and Andrea Vignali – journal: IEEE Transaction on Dependable and Secure Computing – Submitted – 2023*
[P5]	***Network and Physical Data Fusion for Cyber-Physical Systems Protection*** *– Roberto Canonico, Giovanni Esposito, Annalisa Navarro, Simon Pietro Romano, Giancarlo Sperlì, and Andrea Vignali – journal: IEEE Transactions on Industrial Informatics – Submitted – 2023*
[P6]	***Active Learning based Data Augmentation for Named Entity Recognition*** *- Vincenzo Moscato, Marco Postiglione, Giancarlo Sperlì, and Andrea Vignali – journal: Transactions on Knowledge Discovery from Data – Submitted – 2023*

Tutorship

- Machine Learning and Big Data (U3422)
  - Mentorship and practical labs
- Machine Learning for Engineering (U4940)
  - Anomaly Detection and its basic methodologies
  - Autoencoders

Next year

- *Next year will be committed to*
  - *Broadening the knowledge of NLP and Anomaly Detection on different topics:*
    - *Social networks*
    - *Log analysis*
    - *Finance*
    - *Law*
  - *Information fusion of log analysis, network and physical data*

Thank you for your attention!