COMPUTER AND CONTROL ENGINEERING

ADA University - Graph network models for Data Science

Funded By	ADA University [P.iva/CF:1301261241]
Supervisor	APILETTI DANIELE - daniele.apiletti@polito.it
Contact	
Context of the research activity	Machine learning approaches extract information from data with generalized optimization methods. However, besides the knowledge brought by the data, extra a-priori knowledge of the modeled phenomena is often available. Hence an inductive bias can be introduced from domain knowledge and physical constraints, as proposed by the emerging field of Theory-Guided Data Science. Within this broad field, the candidate will explore solutions exploiting the relational structure among data, represented by means of Graph Network approaches. Politecnico di Torino will reserve this position to selected candidates from ADA University according to the Cooperation Agreement in the framework of IAU University.
	Research Objectives The research aims at defining new methodologies for semantics embedding, propose novel algorithms and data structures, explore applications, investigate limitations, and advance the solutions based on different emerging Theory-guided Data Science approaches. The final goal is to contribute to improving the machine learning model performance by reducing the learning space thanks to the exploitation of existing domain knowledge in addition to the (often limited) available training data, pushing towards more unsupervised and semantically richer models. To this aim, the main research objective is to exploit the Graph Network frameworks in deep-learning architectures by addressing the following issues: - Improving state-of-the-art strategies of organizing and extracting information from structured data. - Overcoming the Graph-Network model limitation in training very deep architectures, with a consequent loss in expressive power of the solutions. - Advancing the state-of-the-art solutions to dynamic graphs, which can change nodes and mutual connections over time. Dynamic Networks can successfully learn the behavior of evolving systems. - Experimentally evaluate the novel techniques in large-scale systems, such as supply chains, social networks, collaborative smart-working platforms, etc.

Objectives	Currently, for most graph-embedding algorithms, the scalability of the structure is difficult to handle since each node has a peculiar neighborhood organization. - Applying the proposed algorithms to natively graph-unstructured data, such as texts, images, audio, etc. - Developing techniques to design ensemble graph architectures to capture domain-knowledge relationships and physical constraints. Outline 1st year. The candidate will explore the state-of-the art techniques of dealing with both structured and unstructured data, to integrate domain-knowledge strategies in network model architectures. Applications to physics phenomena, images and text, taken from real-world networks such as social platforms and supply chains will be considered. 2nd year. The candidate will define innovative solutions to overcome the limitations described in the research objectives, by experimenting the proposed techniques on the identified real-world problems. The development and the experimental phase will be conducted on public, synthetic, and possibly real-world datasets. New challenges and limitations are expected to be identified in this phase. During the 3rd year, the candidate will extend the research by widening the experimental evaluation to more complex phenomena able to better leverage the domain-knowledge provided by the Graph Networks. The candidate will perform optimizations on the designed algorithms, establishing limitations of the developed solutions and possible improvements in new application fields. Target publications IEEE TKDE (Trans. on Knowledge and Data Engineering) ACM TKDD (Trans. on Knowledge Discovery in Data) ACM TOIS (Trans. on Information Systems) ACM TOIT (Trans. on Internet Technology) ACM TIST (Trans. on Intelligent Systems and Technology) IEEE TPAMI (Trans. on Pattern Analysis and Machine Intelligence) Information sciences (Elsevier) Expert systems with Applications (Elsevier) Engineering Applications of Artificial Intelligence (Elsevier) Journal of Big Data (Springer) ACM Transactions on Spatial Algorithms and Systems (TSAS) IEEE Transactions on Big Data (TBD) Big Data Research IEEE Transactions on Emerging Topics in Computing (TETC) Information sciences (Elsevier)

Skills and competencies for the development of the activity	• Knowledge of the basic computer science concepts. • Programming skills in Python • Undergraduate experience with data mining and machine learning techniques • Knowledge of English, both written and spoken. • The ability to present the results of the work, both written (scientific writing and slide presentations) and oral. • Entrepreneurship, autonomous working, and goal oriented. • Flexibility and curiosity for different activities, from programming to teaching to presenting to writing. • Capability of guiding undergraduate students for thesis projects.