Aangeboden projecten: Details project

Prediction of spurious and false associations in biological networks (Dr. Georgios Tsatsaronis)

Master Artificial Intelligence

13 Aug 2020

1.1 Proposal

AI Project Proposal

1.2 Name

Prediction of spurious and false associations in biological networks

1.3 Description

Elsevier is working in close collaboration with the Rochester General Hospital in harnessing prior knowledge from medical literature to assemble and analyse biological networks for the purpose of identifying intervention targets in complex chronic illnesses. Although this prior knowledge of regulatory interactions is being extracted from Elsevier's sizable corpus of over 5 million journal papers, the complexity of biological signalling is such much remains to be discovered. Hypothesis generation informed on the basis of regulatory interactions that are currently reported is a critical component in the optimal planning of new experiments. The prediction of spurious or indirect associations (false positives), as well as the absence of functionally important direct interactions (false negatives) has been proposed based on adherence to known fundamental topological properties of biological networks [1], albeit in undirected networks. The aim of this project is to extend this work to the prediction of novel interactions based on directed regulatory interactions and their *kinetic* properties. Applying Alon [2] regulatory motifs as basic building blocks is one potential approach. Machine learning techniques would be applied to artificial networks with known defects to create predictive models for inferring the such novel interactions. The results of the project will be summarized in a paper that will be submitted in high-impact peer review journal or conference.

References:

[1] Guimerà R, Sales-Pardo M. Missing and spurious interactions and the reconstruction of complex networks.Proc Natl Acad Sci U S A. 2009 Dec 29;106(52):22073-8.

[2] Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007 Jun;8(6):450-61.

1.4 Work environment

The student will be working within the Data Science department of Research Content Operations. The team comprises primarily machine learning and natural language processing experts, and data scientists, and is based at the headquarters of Elsevier in Amsterdam (Sloterdijk). The environment boosts innovation, by daily interactions of the group's members, who constantly challenge existing systems and processes, and aim at embedding state of the art technologies and novel ideas in the Elsevier's products and platforms. Elsevier offers access to large computational environments, in which the student can develop new ideas in a large variety of programming paradigms.

1.5 Expectations

Focused, open-minded, eager to innovate and make a change to existing technologies and products by applying novel ideas. Some prior experience with Natural Language Processing or Machine Learning is also a very positive characteristic, though it is not required, as the student can be trained in these methods during the Thesis period.

1.6 Duration

MSc AI: 8 months

1.7 Registration for thesis fair(s)

1 October 2020 AI Thesis Fair

1.8 Research Tags

Machine Learning and Deep Learning : Geometric deep learning, Information Retrieval and Language Processing : Language processing, MISC : AI for Health and Medicine

1.9 Programmes

This project is exclusively for MSc AI students

1.10 Contact

Dr. Georgios Tsatsaronis (gb.tsatsaronis@gmail.com, Extern)

2.1 Imported project id

2.2 Copied project id