Internship proposal (Master) on “Graph Neural Network for Phenotype prediction”

This internship is intended to be pursued by a PhD. We are thus looking for a candidate in his/her 2nd year of master studies, who is willing to apply with us to a PhD funding program.

Context

Large-scale biological datasets have been populated thanks to modern high-throughput technologies, the decreasing costs of data generation, unprecedented improvement in data processing and analysis and the increasing capacity to save and store these datasets. However, the challenge of well predicting an observable disease or trait of interest (phenotype) from these data is still partially unmet, while it is key to address important societal needs in medicine, agriculture or biotechnologies.

Some of the reasons that explain that the exploitation of existing biological data is still insufficient are, from one part, the very large dimensional of the datasets that poses computational and statistical issues, and, for the other part, the very complex relations existing between all the biological entities at different levels of the living organism and that impact the phenotype. An improved understanding of phenotypes should be made possible by the development of new predictive models for living beings.

In this proposal, we intend to use recently developed Graph Neural Network [Scarcelli, *et al.* 2009; Micheli 2009] approaches to predict phenotypes from gene expression and other omics data, incorporating relationship information. Such methods have been widely developed recently in a variety of applications involving relations between entities of interest https://rlgm.github.io/papers/ but has not been tested yet for such applications with only one recent exception [Chereda, *et al 2019*], as far as we can tell.

Objectives

The objective of the master internship will be to test the benefit of incorporating network information into expression-to-phenotype neural network and to compare the performance with that of other machine learning approaches or to neural networks blind of network information. This will require to:

- set up a test case dataset for benchmarking;
- test several architectures and/or implementations [Grattarola & Alippi 2020; Chen, *et al.* 2020];
- compare with other approaches.

The conclusions of the internship will be incorporating to specify a PhD project that intends to use such approaches to jointly predict the phenotype and learn the network structure.

Conditions

The intern will be hosted at MIAT laboratory from INRAE Toulouse (at Castanet Tolosan). The intern will benefit from the facilities of the INRAE center (working station, lab servers, cafeteria, …) and from a dynamic working environment. He/she will receive the legal French allowance for master internships (approximately 600€/month).

Required skills

Good programming skills (Master in computer science, data science, statistics, …). Knowledge in computational biology is not mandatory but will be a plus. Master students in bioinformatics with solid background in computer science and motivation for the topic are also encouraged to apply.

Contacts

- Nathalie Vialaneix (INRAE, MIAT): nathalie.vialaneix@inrae.fr
- Céline Brouard (INRAE, MIAT): celine.brouard@inrae.fr

To apply to this offer, please, send an email to both contacts, with a CV and an application letter briefly describing which skills you have to address this topic and why is the topic of interest for you.

References

[Chen, *et al.*, 2020] Chen, Jacob, and Mairal (2020) Convolutional kernel networks for graph-structured data. *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, Vienna, Austria, PMLR 108, 2020.

[Chereda, *et al.* 2019] Chereda, Bleckmann, Kramer, Leha, and Beissbarth (2019) Utilizing molecular network information via graph convolutional neural networks to predict metastatic event in breast cancer. *Studies in Health Technology and Informatics*, 267, 181-186.

[Grattarola & Alippi 2020] Grattarola and Alippi (2020) Graph neural networks in TensorFlow and Keras with Spektral. *Graph Representation Learning and Beyond – ICML 2020 Workshop*.

[Micheli 2009] Micheli (2009) Neural network for graphs: a contextual constructive approach. *IEEE Transactions on Neural Networks*, 20(3), 498-511.

[Scarcelli, *et al.* 2009] Scarcelli, Gori, Tsoi, Hagenbuchner, and Monfardini (2009) The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1), 61-80.