Post-Doctoral position (12 months extensible to 24 months)

GRADIVA: GRAphs for Deep neural network InVestigAtion

Proposal description:
Graphs are nowadays a common mathematical formalism used in various domains where the notion of network is significant, such as Genetics, Sociology, Ecology and Neurosciences for instance. For the latter graph representation allows to describe brain connectivity both at a structural and a functional level (1). Moreover, graph neural network is an emerging topic in data mining where the graph modelling allows the use of mathematical tools from graph theory in combination with deep learning approach.
The objective of this Post-Doctoral position is to use mathematical properties of classical graph neural networks in order to explore unresolved specific weaknesses of deep neural networks (DNN). We will focus on *catastrophic forgetting* (or catastrophic interference) and *adversarial attack*. The former hampers the training phase of DNN when the trained model forgets a previously learned pattern when confronted with new examples to learn. The latter refers to the vulnerability of DNN to a subtle carefully designed change in how inputs are presented completely alters its output and leads to wrong conclusion. To study these major DNN drawbacks notably for medical application (2), we propose to represent the DNN as a graph and track learning and prediction under different conditions of training and attack (3). The final goals are to respond to several questions: are specific hidden neurons (or layers) vulnerable to forgetting or attack? Which solutions can be implemented (introduction of penalty during the training phase, specific architectures including feedback connections, …) to design DNN more resilient to forgetting and attack?

(1) Hanczar, B., Zehraoui, F., Issa, T. *et al.* (2020) Biological interpretation of deep neural network for phenotype prediction based on gene expression. *BMC Bioinformatics* **21,** 501.
(2) Finlayson, S.G., Bowers, J.D., Ito, J., Zittrain, J.L., Beam, A.L., and Kohane, I.S. (2019). Adversarial attacks on medical machine learning. *Science* 363**,** 1287-1289.
(3) Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K., & Samek, W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. PLoS ONE, 10.

Key words: Machine learning; Multidimensional data; Neural Network

Environment: We offer a stimulating research environment gathering experts in Neurosciences & Neuroimaging and experts in Advanced Statistical and Machine Learning methods. The post-doc position will be available in the context of the Grenoble 3AI project (chair neuromorphometrics @MIAI https://miai.univ-grenoble-alpes.fr/). The postdoctoral fellow will work in close collaboration with a PhD student working on Graphs as model for brain network studies and with a Cea-List team which has developed a bioinspired architecture which could offer interesting resilient properties.

Supervision / contact: GIN-team « Functional neuroimaging and brain perfusion»: Michel Dojat (michel.dojat@inserm.fr), Inria-team Statify, Sophie Achard (sophie.achard@inria.fr) and Leti Marina Reyboz (marina.reyboz@cea.fr). Location: Grenoble Neurosciences Institut: https://neurosciences.univ-grenoble-alpes.fr & Inria Montbonnot : https://www.inria.fr/en/teams/mistis
Starting date: Autumn 2021
How to apply: Send an email directly to the supervisors with your CV. Applications will be accepted up to the 31st of August. The final decision will be given by the beginning of October.