GREYC

Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen (UMR 6072)

24 août 2021

Graph decimation for Deep Graph Neural Networks

A PhD grant at the GREYC Laboratory (Caen, France)

keywords : Deep learning, Pooling, Graph Decimation, Graph Neural Networks.

Context Most of the objects of interest of our today's life are based on discrete objects with sequential (strings) or more complex (graph) relationships. We can evoke the relationships between people in social graphs, the bounds between atoms in a molecule or the topographic distance between speed sensors in traffic analysis, to name a few. The prediction of the properties of such objects falls in the scope of structural pattern recognition. For decades, this research field has been limited by costly (e.g. based on sub-graph isomorphism) or poorly efficient metrics usually combined with limited machine learning algorithms (mainly k-nearest neighbors algorithm). A first important breakthrough has been achieved by the introduction of structural kernels dedicated to strings or graph. In addition to provide efficient metrics on these discrete objects, the latter are gateway towards many machine learning methods. Hence, they reduce the gap between structural and statistical pattern recognition techniques. A second breakthrough in this field has been provided by the introduction of Graph Neural Networks (GNNs). As graph kernels, these networks provide a strong connection between graphs and machine leaning techniques. Moreover, as other deep learning techniques, GNNs avoid handcrafting the design of a similarity measure between graphs. GNNs are based on two operations, namely, Graph convolution and Graph decimation/pooling. However, these two operations still suffer from severe drawbacks. First, the expressive power of Graph convolution operations is limited in the spectral domain and usually corresponds to a low-pass filter. Secondly, the graph decimation operation is usually performed by existing graph clustering algorithms, while the equivalent operation in image neural networks corresponds to a sub-sampling, which offers guarantees in terms of decimation ratio and connectedness of the merged entities.

This PhD will focus on this last problem in close collaboration with other partners investigating the graph convolution framework.

Thesis topic In the following we distinguish two different concepts : Graph decimation, which consists in reducing the size of a graph by grouping connected sets of vertices, and Graph pooling, which consists in summing up a connected graph by a numerical value or vector.

The PhD will be roughly decomposed into three steps :

1. Graph decimation : The PhD student will first have to study Graph decimation techniques developed by our team in order to transpose them to a GPU implementation and the deep learning framework. These decimation scheme should insure :
   (a) A fixed decimation rate (ration between the sizes of two successive graphs),
   (b) A bounded (small) radius of the subgraphs grouped into one vertex by the decimation scheme.

2. Graph Spectral properties : The PhD student will have to study the literature related to decimation schemes preserving the spectral properties of Graphs. He will then have to propose new algorithms combining the findings of the previous step with these techniques in order to insure the preservation of spectral properties additionally to the fixed decimation ratio and the bounded sizes of clusters (subgraphs).
3. Learning decimation : This last step is certainly one of the most important. Existing techniques which learn a decimation scheme provide almost complete graphs hence discarding the graph structure. The PhD student will have to understand them and to improve the structural properties of the output based on previous findings.

Application. The main application field of the PhD will be devoted to drug design in close collaboration with an other laboratory specialized in this field. More precisely, the targeted applications are Alzheimer's disease and poly-pharmacology. Two complementary targets will be investigated for this application : 1) to improve the drug prediction activity on the datasets provided by our biological partner, 2) to group automatically atoms in pharmacophores(groups of atoms known to have a biological activity) through our decimation scheme. Let us note that our biological partner has already identified a set of handcrafted rules allowing to group atoms in pharmacophores. Our aim here will be to propose new atom groups through our learned decimation scheme.

Candidate Profile Curious and autonomous, the candidate should have a master degree or a diploma of Engineering in computer science or applied mathematics. A solid background in machine learning and/or deep learning will be appreciated. Additional skills in Graph theory or an interest in the subject would also be a plus. Due to short delays, an European Visa or an ability to get one quickly (e.g. thanks to your nationality) will be also appreciated.

Names of contacts

Luc Brun : luc.brun@unicaen.fr (PhD's supervisor)

Benoit Gaüzère : benoit.gauzere@insa-rouen.fr (co supervisor)

Information on the position

Location : Caen. This city is located in the north-west part of France (2h from Paris by train).

Salary : at least 1768 € without charges (corresponding to 1430€ charges included). This salary should be increased in 2022.

Limit date to postulate : October 2021

Start of the PhD : October/November 2021.