PhD position
Explicability of GNN decisions, Application to chemoinformatics

Abstract

The GREYC (computer science) and CERMN (pharmacology) research laboratories have been working for many years in a joint group to develop innovative computer methods for processing data relating to drugs and, in particular, to try to predict and understand how a new molecule acts from knowledge extracted from a set of molecules.

Since molecules can be modelled by graphs, it is possible to train a GNN (Graph Neural Network) to classify or predict the action of molecules, but currently no explanation of the network's decision-making is given. The explicability of networks is fundamental in predicting the therapeutic properties of molecules. Indeed, it allows

1. to validate predictions before committing efforts and resources to the synthesis of a molecular series
2. to gain insight into the key physico-chemical properties that a molecule must possess in order to have a targeted biological action.

The aim of this thesis is to propose solutions to explain the decisions made by a neural network operating on graphs with a view to applying it to data on chemical molecules. No knowledge of therapeutic chemistry is required.

Detailed subject : `https://lamotte.users.greyc.fr/files/theseCODAG-1.pdf`

Funding : €2,100 gross, see https://www.enseignementsup-recherche.gouv.fr/fr/le-financement-doctoral-46472

Location : University of Caen Normandy, GREYC laboratory, CODAG team

Contact : `jean-luc.lamotte@at@unicaen.fr`

Application deadline : 13 May 2024

Detailed subject : `https://lamotte.users.greyc.fr/files/theseCODAG-1.pdf`

Application : CV, transcript of Bachelor, MSc (or equivalent level) marks, covering letter explaining how your application matches the subject, and contact details for one or more referees.

Candidate profile : the candidate must be enrolled in the final year of a Master's degree or an engineering degree, or hold such a degree, in a field related to computer science or applied mathematics, and have solid programming skills. Experience in data science, deep learning, etc. would be a plus. The candidate must be able to write scientific reports and communicate research results at conferences in English.

1 Scientific context of the thesis

Deep learning has revolutionised many fields such as image analysis and automatic language processing. However, the inability of these systems to justify their decisions very quickly emerged as a major limitation. Methods for measuring the salience of each input have therefore been defined. These methods can be divided into two main groups: so-called "white box" methods, which assume that the network is known and are therefore based on all the information it contains: topology, learned weights and gradient calculated by backpropagation. Conversely, "black box" methods consider the network as an unknown function and measure the importance of each element by randomly masking parts of the input signal. The importance of each element is then defined as the average of the system responses over all the masks in which the element is not masked.

Chemoinformatics, the field of chemical information processing, has been strongly impacted by the arrival of deep learning and convolutional networks (CNNs). A Graph Neural Network (GNN) is the equivalent of a CNN operating on graphs. Convolution is replaced by graph convolution and pooling operations can also be defined. One notable difference between CNNs and GNNs is the pooling. In CNNs, pooling operates on fixed topologies and uses generally unlearned reduction functions such as max pooling or average pooling.

In GNNs, pooling learns both vertex groupings of arbitrary topology and reduction functions on these groupings.

The therapeutic action of a molecule depends on its ability to bind to the targeted proteins. The molecule's ability to do this is determined by its most likely conformations and by the presence of various functional groups, also known as pharmacophoric characteristics, which will enable the molecule to bind. There are two types of molecule representation for this type of prediction: the skeletal graph (vertices and edges represent the atoms and their covalent bonds respectively) and the pharmacopharmacological characteristics graph, where each vertex codes for a functional group of the molecule and the distance (topological or topographic) between two characteristics is coded by an edge. In this case, the graph is a complete graph with far fewer vertices than in the skeletal representation [11, 5].

The explicability of networks is fundamental in predicting the therapeutic properties of molecules. It allows :

1. to validate predictions before committing efforts and resources to the synthesis of a molecular series

2. to gain insight into the key physico-chemical properties that a molecule must possess in order to have a targeted biological action.

The methods explaining the action of networks in the context of graphs are an extension of the methods designed for CNNs. In particular, we can apply the black-box [17, 21, 13, 6]/white-box [14, 2, 4] distinction but also subdivide the approaches into local methods (such as the methods previously cited), which provide an explanation of the prediction of each input and the global [20, 10, 18] methods which attempt to put forward a global explanation of the key motifs for the prediction.

Note that the two approaches are not orthogonal. For example [1] agglomerates local explanations on a dataset to provide a global explanation. Other decomposition criteria are also possible. For example, approaches can be broken down into factual methods, for example [2], which look for patterns (nodes, edges, sub-graphs) that best explain the prediction, and counter-factual models [9, 3] which, on the contrary, look for minimal modifications to the input data that change the prediction. Among the factual methods, we can also distinguish methods that integrate pattern extraction into the prediction. These methods are therefore self-interpretable [12, 19] and methods that combine the prediction model with another method to generate explanations (e.g. [2]). The interested reader can refer to the following surveys for more information [8, 7]. It should be noted, however, that white-box methods have focused on GNN-type networks without pooling. To our knowledge, the use of pooling for explicability has not yet been explored.

2 Subject and progress of the thesis

The work will begin with a study and a comparison of methods for explaining GNNs on standard datasets and also on our protein/ligand interaction datasets on both complete graphs and skeletal graphs. In this respect, several criteria such as accuracy, area under the curve, fidelity, parsimony, have been defined.

On the basis of these comparisons, we plan to propose our own methods and, in particular, to explore the use of the pooling step [16, 15] for explainability. Indeed, a learned pooling must group together vertices that are important for the property to be predicted, which constitutes a form of explainability that has been little explored until now. Our aim is to obtain for each molecule a measure of the importance of its functional groups and their distances. This will be followed by a search of the entire database,

to identify important functional groups located at similar distances.

These groups will define what we call pharmacophores. This study will be extended to skeletal graphs. This will enable us to take into account the whole molecule and to learn which functional groups are active for a targeted biological action.

The pooling hierarchy will be studied in particular.

3 Supervision team

This thesis will be co-directed by Jean-Luc Lamotte (CODAG team - GREYC) and Luc Brun (Image team - GREYC) and co-supervised by Pierre Héroux (Learning team - LITIS). The relevance of this supervisory team to the proposed subject is discussed below.

4 Applicant profile

The candidate must be enrolled in the final year of a Master's degree or an engineering degree, or hold such a degree, in a field related to computer science or applied mathematics, and have solid programming skills. Experience in data science, deep learning, etc. would be a plus. The candidate must be able to write scientific reports and communicate research results at conferences in English.

5 Thesis conditions

The thesis will start in autumn 2024 - early September or early October. The work will take place mainly at GREYC in the CODAG and IMAGE teams, with strong interaction with the LITIS laboratory in Rouen on the deep learning and explainability part and with the CERMN laboratory in Caen for the analysis of the results.

6 Thesis environment

GREYC has an ongoing collaboration with the CERMN laboratory on data mining applied to therapeutic molecules.

GREYC also collaborates with LITIS on the theme of deep learning on graphs. In recent years, this collaboration has taken concrete form through the AGAC project (2018-2019) supported by the Normandy Region. The joint and complementary work of the two laboratories on the theme of deep learning on graphs has also been recognised by the selection and funding by the ANR of the CoDeGNN project (2022-2025) and the very high-level publications that have resulted from it balcilar2021analyzing, balcilar:hal-03410699. It should be noted that this thesis will strengthen the chemoinformatics aspect of this joint research project.

Chemoinformatics also provides LITIS with an application framework to support the scientific issues that have arisen from its collaboration with the COBRA chemistry laboratory (UMR-6014). The methods developed as part of this project could also find applications, for example, in identifying the molecular characteristics that polymers (to be synthesised) need to have certain physico-chemical properties.

These collaborations have produced results of international significance. Deep learning on graphs is currently the subject of a great deal of international research.

This thesis will bring together the GREYC and LITIS teams working on GNNs by developing a new line of research on the explicability of GNN decisions. This GREYC/LITIS collaboration will strengthen NormaSTIC's cross-disciplinary Graph axis and, more generally, the collaborations promoted by the federation. For CERMN, the work will make it possible to acquire new expertise in Deep Learning technologies, in a way that is efficient and fully adapted to their objects of study.

People implied in the project :

- GREYC : Jean-Luc Lamotte (co-supervisor), Luc Brun (co-supervisor), Bertrand Cuissart
- LITIS : Pierre Héroux, Paul Honeine
- CERMN : Ronan Bureau, Alban Lepailleur

Funding: €2,100 gross, see https://www.enseignementsup-recherche.gouv.fr/fr/le-financement-doctoral-46472

Location: University of Caen Normandie, GREYC laboratory, CODAG team,

Contact : `jean-luc.lamotte@at@unicaen.fr`

Application deadline: 13th of may, 2024

7 Application form

It consists of a curriculum vitae, a transcript of marks from bachelor and MSc (or equivalent level), a covering letter explaining why your application is relevant to the subject, and the contact details of one or more referees.

References

[1] Steve Azzolin, Antonio Longa, Pietro Barbiero, Pietro Liò, and Andrea Passerini. Global explainability of gnns via logic combination of learned concepts. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.

[2] Federico Baldassarre and Hossein Azizpour. Explainability techniques for graph convolutional networks. *CoRR*, abs/1905.13686, 2019.

[3] Ziheng Chen, Fabrizio Silvestri, Jia Wang, Yongfeng Zhang, Zhenhua Huang, Hongshik Ahn, and Gabriele Tolomei. GREASE: generate factual and counterfactual explanations for gnn-based recommendations. *CoRR*, abs/2208.04222, 2022.

[4] Qizhang Feng, Ninghao Liu, Fan Yang, Ruixiang Tang, Mengnan Du, and Xia Hu. DEGREE: decomposition based explanation for graph neural networks. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

[5] Damien Geslin, Alban Lepailleur, Jean-Luc Manguin, Nhat-Vinh Vo, Jean Luc Lamotte, Bertrand Cuissart, and Ronan Bureau. Deciphering a pharmacophore network: A case study using BCR-ABL data. *J. Chem. Inf. Model.*, 62(3):678–691, 2022.

[6] Qiang Huang, Makoto Yamada, Yuan Tian, Dinesh Singh, and Yi Chang. Graphlime: Local interpretable model explanations for graph neural networks. *IEEE Trans. Knowl. Data Eng.*, 35(7):6968–6972, 2023.

[7] Jaykumar Kakkad, Jaspal Jannu, Kartik Sharma, Charu C. Aggarwal, and Sourav Medya. A survey on explainability of graph neural networks. *CoRR*, abs/2306.01958, 2023.

[8] Yiqiao Li, Jianlong Zhou, Sunny Verma, and Fang Chen. A survey of explainable graph neural networks: Taxonomy and evaluation metrics. *CoRR*, abs/2207.12599, 2022.

[9] Ana Lucic, Maartje A. ter Hoeve, Gabriele Tolomei, Maarten de Rijke, and Fabrizio Silvestri. Cf-gnnexplainer: Counterfactual explanations for graph neural networks. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *International Conference on Artificial Intelligence and Statistics, AISTATS 2022, 28-30 March 2022, Virtual Event*, volume 151 of *Proceedings of Machine Learning Research*, pages 4499–4511. PMLR, 2022.

[10] Lucie Charlotte Magister, Dmitry Kazhdan, Vikash Singh, and Pietro Liò. Gcexplainer: Human-in-the-loop concept-based explanations for graph neural networks. *CoRR*, abs/2107.11889, 2021.

[11] Jean-Philippe Métivier, Bertrand Cuissart, Ronan Bureau, and Alban Lepailleur. The pharmacophore network: A computational method for exploring structure–activity relationships from a large chemical data set. *Journal of Medicinal Chemistry*, 61(8):3551–3564, 04 2018.

[12] Siqi Miao, Mia Liu, and Pan Li. Interpretable and generalizable graph learning via stochastic attention mechanism. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 15524–15543. PMLR, 2022.

[13] Tamara A. Pereira, Erik Nascimento, Lucas E. Resck, Diego Mesquita, and Amauri A. Souza. Distill n' explain: explaining graph neural networks using simple surrogates. In Francisco J. R. Ruiz, Jennifer G. Dy, and Jan-Willem van de Meent, editors, *International Conference on Artificial Intelligence and Statistics, 25-27 April 2023, Palau de Congressos, Valencia, Spain*, volume 206 of *Proceedings of Machine Learning Research*, pages 6199–6214. PMLR, 2023.

[14] Phillip E. Pope, Soheil Kolouri, Mohammad Rostami, Charles E. Martin, and Heiko Hoffmann. Explainability methods for graph convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 10772–10781. Computer Vision Foundation / IEEE, 2019.

[15] Stevan Stanovic, Benoit Ga"uz'ere, and Luc Brun. Maximal Independent Vertex Set applied to Graph Pooling. In *Structural and Syntactic Pattern Recognition (SSPR)*, Montr'eal, Canada, August 2022.

[16] Stevan Stanovic, Benoit Ga"uz'ere, and Luc Brun. Maximal independent sets for pooling in graph neural networks. In Mario Vento, Pasquale Foggia, Donatello Conte, and Vincenzo Carletti, editors, *Graph-Based Representations in Pattern Recognition - 13th IAPR-TC-15 International Workshop, GbRPR 2023, Vietri sul Mare, Italy, September 6-8, 2023, Proceedings*, volume 14121 of *Lecture Notes in Computer Science*, pages 113–124. Springer, 2023.

[17] Minh N. Vu and My T. Thai. Pgm-explainer: Probabilistic graphical model explanations for graph neural networks. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

[18] Han Xuanyuan, Pietro Barbiero, Dobrik Georgiev, Lucie Charlotte Magister, and Pietro Liò. Global concept-based interpretability for graph neural networks via neuron analysis. In Brian Williams, Yiling Chen, and Jennifer Neville, editors, *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 10675–10683. AAAI Press, 2023.

[19] Junchi Yu, Tingyang Xu, Yu Rong, Yatao Bian, Junzhou Huang, and Ran He. Graph information bottleneck for subgraph recognition. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

[20] Hao Yuan, Jiliang Tang, Xia Hu, and Shuiwang Ji. XGNN: towards model-level explanations of graph neural networks. In Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash, editors, *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 430–438. ACM, 2020.

[21] Yue Zhang, David DeFazio, and Arti Ramesh. Relex: A model-agnostic relational model explainer. In Marion Fourcade, Benjamin Kuipers, Seth Lazar, and Deirdre K. Mulligan, editors, *AIES '21: AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, USA, May 19-21, 2021*, pages 1042–1049. ACM, 2021.