Preimage Problem for Graph Data

Benoit Gaüzère and **Paul Honeine**

LITIS Lab, Rouen Normandy

Keywords

Machine Learning, Graph Neural Networks, Generative Adversarial Networks, Preimage Problem

Description

Graphs are a powerful and versatile data structure useful to encode many real-world data, such as networks, molecules and documents. However, their flexibility comes with some drawbacks including the complexity associated to elementary operations. For instance, deciding if two graphs are isomorphic (i.e., structurally equivalent) or computing a distance between two graphs are NP-Complete problems, and even hard to approximate.

Considering this, several strategies have been proposed to find some workarounds in order to be able to process graphs and use them in the Machine Learning pipeline. The simplest strategy is the explicit embedding of graphs to an Euclidean space [1], at the cost of losing information. To overcome this drawback, two major strategies have been recently investigated. The first one is improving the embedding through kernels defined on graphs [2,3]. The second and more recent strategy is the definition of Graph Neural Networks (GNNs) operating directly on graphs [4–8]. Using neural networks on graphs allows to learn a proper embedding of graphs given a problem to solve, and then alleviate the drawbacks of defining a priori an ad hoc embedding.

These embedding-based methods have been commonly investigated for supervised learning tasks, essentially classification and regression. However, their interpretability is one major drawback. Moreover, they were not efficient in many unsupervised learning tasks, such as estimating the data centroid in k-means or more generally generating a graph prototype (one graph representative of a set of graphs, e.g. a median graph). The main reason is the curse of the *preimage*, since one needs to reconstruct the solution in the graph-data space.

The preimage problem has already been addressed in various domains, mainly for kernel-based methods [9,10]. However, solving this problem for structured data remains an open problem and only very few attempts have been made on strings and some particular class of graphs [11].

The purpose of this PhD thesis is to alleviate the bottlenecks associated to the preimage problem on graphs, through the use of Generative Adversarial Network (GAN) [12,13]. GANs consists of two parts. First, the encoder aims to embed graphs to an Euclidean space of a predefined dimension. This can be implemented using existing GNNs and kernel-based methods. Second, the decoder part aims to reconstruct a graph given a vectorial representation. It may be considered as the inverse function of the encoder part. The purpose here is to define this decoder part to take Euclidean spaces as input and graphs as output, i.e., structured data. By investigating this approach, the PhD candidate will study particularly molecular generation.

Position Details

Location   The research will be conducted at LITIS Laboratory (Rouen, France) in Normandy. The LITIS (EA 4108) is affiliated to Normandie University, University of Rouen and INSA Rouen Normandie, and founding member of the CNRS Research Federation NormaSTIC.

Supervisors
Benoit Gaüzère, LITIS, INSA Rouen http://pagesperso.litislab.fr/~bgauzere
Paul Honeine, LITIS, University of Rouen http://honeine.fr/

Start date   September or October 2020 (or earlier)

Duration   36 months

Application

Required skills

- Master in Applied Mathematics, Computer Science, Data Science, or equivalent

- Experience in Python programming

- Skills in graph theory, neural networks or graph-based methods constitute an advantage

Contact   phd-gan_graphs@litislab.fr

Required documents

- Up-to-date CV

- A cover letter (research experience and interests)

- Recommendation letter or references

References

[1] Kaspar Riesen and Horst Bunke. Classification and Clustering of Vector Space Embedded Graphs. pages 49–70, 2011.

[2] Benoit Gaüzère, Luc Brun, and Didier Villemin. Two new graphs kernels in chemoinformatics. *Pattern Recognition Letters*, 33(15), 2012.

[3] Linlin Jia, Benoit Gaüzère, and Paul Honeine. Graph Kernels Based on Linear Patterns: Theoretical and Experimental Comparisons. *HAL preprint (hal-02053946)*, March 2019.

[4] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

[5] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 30(8):595–608, 2016.

[6] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

[7] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. A comprehensive survey on graph neural networks. *arXiv preprint arXiv:1901.00596*, 2019.

[8] Muhammet Balcilar, Guillaume Renton, Pierre Héroux, Benoît Gaüzère, Sébastien Adam, and Paul Honeine. Bridging the Gap Between Spectral and Spatial Domains in Graph Neural Networks. *HAL preprint (hal-02515637)*, March 2020.

[9] James Tin-Yau Kwok and Ivor Wai-hung Tsang. The Pre-Image Problem in Kernel Methods. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 15(6):1517–25, 2004.

[10] Paul Honeine and Cédric Richard. Preimage problem in kernel-based machine learning. *IEEE Signal Processing Magazine*, 28(2):77–88, 2011.

[11] Gökhan H. Bakır, Alexander Zien, and Koji Tsuda. Learning to find graph pre-images. In Carl Edward Rasmussen, Heinrich H. Bülthoff, Bernhard Schölkopf, and Martin A. Giese, editors, *Pattern Recognition*, pages 253–261, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[13] Nicola De Cao and Thomas Kipf. MolGAN: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*, 2018.