Extensions and Applications of Graph-based Neural Networks

Context

Network-based analysis is a relatively recent and broad field of study. Indeed, researchers from many disciplines are now more than ever involved with the collection, modeling and analysis of network-indexed data. Undeniably, many phenomena arising from different situations can be represented using complex networks. For instance, in computer science the internet network; in biology, the protein-protein interaction (PPI) networks; in sociology, the social networks; and in medicine the disease propagation in human-interaction networks for infectious diseases or the genetic relationship networks for rare diseases. Graphs are mathematical representations of networks.

More recently, artificial intelligence approaches have become a very active and promising research field since there is now increasingly more data and scalable storage, memory and computing capacity to run efficient and effective algorithms on available data. More specifically, the branch of machine learning called deep learning revolutionized the tools commonly used in mainly two fields, computer vision and natural language processing (NLP). Deep learning techniques distinguish from typical statistics tools on the type of input data they treat, non-structured data, such as image, text or signals.

Unlike images and text, there is an increasing number of applications where data are represented as graphs with complex relationships and interdependency between objects. The complexity of graph data has imposed significant challenges on existing deep learning algorithms since graph data require robust approaches tailored to capture their structure. Graph Neural Network (GNN) is the branch of machine learning which concerns on building neural networks for graph data. Despite the potential applications of GNN, only recent studies exist about applying deep learning methods to graphs [1,2]. Many practical situations justify the need of using GNN. In computer vision, recognizing semantic relationships between objects facilitates the understanding of the meaning behind a visual scene; identifying human actions using Human joints which naturally form a graph. In addition, images often contain multiple objects, understanding the relationships (i.e., visual relationships) among them might be useful. In Natural language processing, text classification needs to take into account the inter-relations of documents or words to infer document labels and create a knowledge graph.

Challenges

The purpose of this thesis is to extend the approaches commonly used in deep learning to graph data. Then, consolidated approaches coming from graph mining can be applied or extended such as graph clustering or community detection and classification [3,4, 5,6]. Furthermore, the study of hybrid neural networks, where different type of input data, images, texts and graph are combined will be also considered. For instance, in medicine for disease diagnosis, image, as well as human-human interactions are together needed to understand the development of a disease. A special focus on medical applications will be given. The successful

candidate will perform experimentations in dataset repositories for COVID-19 diagnosis [7,8,9,10].

Required skills

• The candidate must hold an Engineering degree / MSc degree with an expertise with Machine Learning, Deep Learning and graph theory.

• Good programming skills in Python, Pytorch or Tensorflow.

• Fluent in English (C1 preferred/ C2).

Bibliography

[1] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," arXiv preprint arXiv:1901.00596, 2019.

[2] Zhang, S.; Tong, H.; Xu, J.; Maciejewski, R. Graph convolutional networks: A comprehensive review. Comput. Soc. Netw. 2019, 6, 11.

[3] Romain Campigotto, Patricia Conde Céspedes, and Jean-Loup Guillaume. A generalized and adaptive method for community detection. arXiv preprint arXiv:1406.2518, 2014.

[4] Conde-Céspedes, P.; Ngonmang, B.; Viennet, E. An efficient method for mining the Maximal α-quasi-clique- community of a given node in Complex Networks. Social Network Analysis and Mining 2018, 8. doi:10.1007/s13278-018-0497-y.

[5] Conde-Céspedes, P.; Marcotorchino, J.F.; Viennet, E. Comparison of linear modularization criteria using the relational formalism, an approach to easily identify resolution limit. In Advances in Knowledge Discovery and Management; Springer: Cham, Switzerland, 2017; pp. 101–120.

[6] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining. ACM, 2016.

[7] EU Open Data Portal: "COVID-19 Coronavirus data" https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data/resource/55e8f966-d5c8-438e-85bc-c7a5a26f4863

[8] World Health Organisation (WHO☺ Global research on coronavirus disease (COVID-19) https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov

[9] The Humanitarian Data Exchange: Novel Coronavirus (COVID-19) Cases Data https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases

[10] Our World in Data https://ourworldindata.org/coronavirus