HydraGCN: Multi-modal data analysis framework for medical applications (focusing on Graph Convolutional Networks)

1. General Info

Project Title: HydraGCN: Multi-modal data analysis framework for medical applications (focusing on Graph Convolutional Networks)

Contact Person: Anees Kazi, Ahmad Ahmadi, Gerome Vivar, Hendrik Burwinkel

Contact Email: anees.kazi@tum.de, mangotee@gmail.com, gerome.vivar@gmail.com, hendrik.burwinkel@tum.de

Outcome: The result of the project will potentially be published in an A-level med-tech journal.

2. Project Abstract

In medical research, multi-modal datasets from large-scale population-based studies are an essential tool towards better diagnosis and treatment of disease. Multimodal data comprises imaging and non-imaging data and is available in many domains. In technical research, such datasets serve as enablers of Computer-Aided Diagnosis (CADx) with machine learning (ML). In this project, we will be focusing on four such multi-modal dataset HCP dataset and UK Biobank for the application of age and gender prediction.

In this project, we analyze multi-modal datasets with all its challenges such as imbalance, small size, missing values etc with the perspective of graph convolutional networks. We will work on a framework developed at CAMP called 'HydraGCN' which is designed with a multi-head input and output architecture, which allows for flexible modeling of multimodal input data and which enables multi-target learning through separate output heads and losses.

3. Background and Motivation

Graph Convolutional Networks (GCNs) are a subset of models in Geometric Deep Learning (GDL) and have made a remarkable impact on the usage of such data in various non-medical domains, e.g. in computer vision, natural language processing or chemistry. However, GCNs are still to be explored to full extent for medical applications. Recently, it was discovered that it is greatly beneficial to model an entire patient population as a graph, using disease-relevant patient characteristics taken from clinical health records (CHR). GCNs can natively integrate this graph to analyze multimodal data across neighborhoods of patients, similar to concepts in social network analysis. In this project, we will aim at brain graph learning using GCNs. Along with learning the graph, the task will be age and gender prediction.

In recent literature, it can be seen that GCNs provide a principled and dynamic way to integrate imaging and non-imaging modalities in the medical domain. So far, GCNs have been explored for a variety of applications such as disease prediction, segmentation, and matrix completion by using massive, multi-modal datasets. In this project, we want to explore how one GCN network can be applied to solve all the challenges in one framework, in terms of performance, and in terms of understanding the prediction mechanisms.

4. Technical Prerequisites

- Good background in statistics and deep learning
- Very Good skills in Python
- Good skills in PyTorch / Tensorflow(+Keras).

5. Benefits:

- Learn about multi-modal data analysis using graph deep learning.
- Interacting with professionals working with GCNs in industry.

- Possibility of writing a scientific paper.

6. Students' Tasks Description

Students' tasks would be the following:

Group 1:

- Understanding and implementing basic GCNs.
- Understanding the implementation of HydraGCN.
- Running experiments on dataset 1 and dataset 2
- Discuss the limitation of methods and prepare final report

Group 2:

- Understanding and implementing basic GCNs.
- Understanding the implementation of HydraGCN.
- Running experiments on dataset 3 and dataset 4.
- Discuss the limitation of methods and prepare final report

7. Work-packages and Time-plan:

	Description	#Students	From	To
WP 1	Understanding and implementing basic GCNs.	4	29.04	05.04

WP 2	Familiarize with tensorflow/ pytorch	4	06.04	12.05
WP 3	Understanding the implementation of HydraGCN.	4	13.05	26.05
WP 4	Group 1: Experiments on dataset 1 and dataset 2 Group 1: Experiments on dataset 3 and dataset 4	2/2	27.05	09.06
M1	Intermediate Presentation II	4	10.06	
WP 5	Running necessary ablation test	4	17.06	23.06
WP 7	Prepare final results and compare implemented possible comparative methods.	4	24.06	07.07
WP 8	Prepare Final Documentation	4	08.07	14.07
M2	Final Presentation	4	15.07.2021	