Computer Science Tripos – Part II – Project Proposal

A graph convolutional network for Alzheimer's disease classification

Kamilė Stankevičiūtė (`ks830`), Gonville & Caius College

September 26, 2019

Project Originator: Tiago Azevedo

Project Supervisor: Tiago Azevedo

Directors of Studies: Dr T. M. Jones, Prof P. Robinson, Dr G. Titmus

Project Overseers:

Introduction

TODO: *Description of/introduction to Alzheimer's disease—answering the question why is it important to study it and model it using neural networks.*

The use of neural networks provides the opportunity to capture the similarities between patients and trends which might help physicians to understand the mechanisms of the disease and in turn find more effective treatments.

One way of modelling the population with some patients having Alzheimer's is a graph, with vertices representing individuals and their features (including the diagnosis), and nodes corresponding to associations between individuals according to some heuristic or a formally defined similarity metric. Additionally, the graph structure is helpful in incorporate multiple modalities of data (e.g. imaging and non-imaging) that may help with discovery of additional patterns while mitigating the problem of incomplete patient information (e.g. not all patients have been imaged).

This project will explore such graph neural networks in the context of classification of patients into healthy controls and those having Alzheimer's in a semi-supervised manner.

Usability of the Parisot et al. paper is limited because the edges are set arbitrarily based on the specific set of features in the dataset and are not learnt by the model model as well as improve the performance through the use of GPUs.

Starting point

This project will be based on a state-of-the-art convolutional graph neural network (GCN) as described in a paper by Parisot et al. [1] In this paper, the GCN was used to classify Autism Spectrum Disorder and Alzheimer's patients and achieved the accuracies of 70.4% and 80.0% respectively.

The source code of this paper (written in TensorFlow) is publicly available at `github.com/parisots/population-gcn`. This will used as a basis for replication of the results on PyTorch and building additional improvements. PyTorch has been chosen for its libraries specialised for machine learning on structured graph data (particularly the `pytorch_geometric` package), which will make iteration and extensions to the model more flexible as well as improve its performance and simplify the APIs.

This project will use the ADNI dataset (same as in the paper) as the benchmark. The database is available at `adni.loni.usc.edu`.

TODO: *consider possible extension of the project to apply the improvements to other datasets, to demonstrate the flexibility in model transfer to the similar tasks on different dataset. This could be the ABIDE dataset for Autism Spectrum Disorder, PPMI (Parkinson's) dataset...*

Resources required

For the most part of this project I will be using my personal MacBook Pro (2019, quad-core 1.4 GHz Intel Core i5 processor, 8 GB LPDDR3 RAM) running macOS Catalina. Training the model will most likely require the use of GPUs provided by the Computational Biology Group (as confirmed by Prof Pietro Liò).

The following measures will be taken to store the work and reduce the likelihood of any loss of data:

- Saving the source code of the project on a GitHub repository;
- Storing the LaTeX source (as well as source code) on my machine, Google Drive and MCS.
- Regularly backing up the contents of my laptop on an external HDD.

The data required for training (ADNI database) is available on `adni.loni.usc.edu`, for which I requested and was granted access.

Work to be done

Success criteria

- Reimplement the model in PyTorch ()`pytorch_geometric`) package with the base accuracy of at least 80.0% as in [1]
- Compare the performance of this model with the

Possible extensions

Given the main success criteria have been achieved early and there is time left, some possible extensions include:

- ..
- ..
- ..

Timetable

Planned starting date is 18/10/2019.

1. **Michaelmas weeks 2–4** Learn to use X. Read book Y. Read papers Z.
2. **Michaelmas weeks 5–6** Do preliminary test of Q.
3. **Michaelmas weeks 7–8** Start implementation of main task A.
4. **Michaelmas vacation** Finish A and start main task B.
5. **Lent weeks 0–2** Write progress report. Generate corpus of test examples. Finish task B.
6. **Lent weeks 3–5** Run main experiments and achieve working project.
7. **Lent weeks 6–8** Second main deliverable here.
8. **Easter vacation:** Extensions and writing dissertation main chapters.
9. **Easter term 0–2:** Further evaluation and complete dissertation.
10. **Easter term 3:** Proof reading and then an early submission so as to concentrate on examination revision.

References

[1] Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrero, Ben Glocker, and Daniel Rueckert. Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer's disease. *Medical Image Analysis*, 48:117–130, August 2018.