Kamilė Stankevičiūtė

Graph neural networks for age prediction from neuroimaging data

Computer Science Tripos – Part II

Gonville & Caius College

2020

Declaration of originality

I, Kamilė Stankevičiūtė of Gonville & Caius College, being a candidate for Part II of the Computer Science Tripos, hereby declare that this dissertation and the work described in it are my own work, unaided except as may be specified below, and that the dissertation does not contain material that has already been used to any substantial extent for a comparable purpose.

Signed *Kamilė Stankevičiūtė*

Date 9th January 2020

Proforma

Candidate number:	?
Project title:	Graph neural networks for age prediction from neuroimaging data
Examination:	Computer Science Tripos – Part II 2020
Word count:	? [1]
Lines of code:	?
Project originator:	Mr Tiago Azevedo
Supervisors:	Mr Tiago Azevedo, Mr Alexander Campbell

Original aims of the project

To implement the population graph paradigm in modelling multi-modal structural and functional brain imaging data along with phenotype information of individual patients. Implement and apply different graph neural network architectures – Graph Convolutional Network (GCN) and Graph Attention Network (GAT) – for a patient age prediction task. Evaluate the robustness of graph neural networks to noisy and missing data, comparing their performance against each other and with other known benchmarks.

Work completed

Special difficulties

None.

[1] Counted using the LaTeX Utilities extension in Visual Studio Code.

Contents

Acknowledgements

Chapter 1

Introduction

The prevalence of age-associated disease and disability is increasing with the ageing population. Increasingly common and debilitating are the neurological and neurodegenerative disorders such as Alzheimer's disease, Parkinson's disease...

Brain maturity and ageing depends on many complex factors depending on both genetic and postnatal environmental experience, and is associated with cellular damage, changes in brain connectivity, structural gray matter and white matter changes, morphological and functional connectivity changes, frontal lobe thinning [2, 6, 3]. This is expressed in a cognitive decline, decreased memory capacity and a variety of increasingly common neurological and neurodegenerative disorders such as Alzheimer's and dementia. With the increasing global population lifespan, those disorders are increasingly disabling and therefore there is an interest in finding the biomarkers representing the brain health status and revealing the common risk factors, giving insight into the mechanisms of ageing, potentially delaying the onset of the disease, slowing it down and reducing its prevalence. With the increased availability of brain imaging data through the population-wide studies such as the UK Biobank, and increasing computing power enabling the analysis of rich and otherwise computationally intractable datasets (including the machine learning techniques), brain ageing is becoming a popular research direction.

One biomarker for estimating the overall health of the brain is the so-called *brain age*, which is a prediction of a person's *chronological* (actual) age based on the neuroimaging analysis of the patient, most commonly structural magnetic resonance imaging (MRI) data. Recent studies have linked the deviations between the brain age estimate and the chronological age to the occurrence of various neurological conditions such as Alzheimer's and dementia [5], presumably because the overestimate of the chronological age indicates the accelerating ageing and cellular damage accumulation of the brain.

Why population graphs are a nice approach (use of multimodal data; individual patient data *and* diagnoses of other patients; potential robustness to missing or noisy data of a patient as other patient can "make up" for it).

Why my work is novel (applying a new population graph framework to a different – and more challenging – problem).

1.1 Related work

1.1.1 Population graphs

Semisupervised population graphs in Parisot et al. for classifying healthy patients and patients having Mild Cognitive Impairment/Alzheimer's or autism spectrum disorder. These achieve state-of-the-art accuracies around 70% even for a binary classification task which indicates that brain conditions are generally a complex task. Due to the different nature of the task and different evaluation metrics, this performance cannot be directly compared to the predictive power of the brain age regression task.

1.1.2 Brain age prediction from neuroimaging data

Most of current machine learning-based methods for brain age gap prediction work on per-brain basis and do not consider pairwise similarities between patients, not taking into account the population as a whole.

Graphs are static so it makes sense that most practical approaches would not use this as their model, because it prevents using applying it for new patients

Kaufmann et al. [5] uses gradient boosting based techniques (XGBoost) [1] for brain age gap prediction from structural magnetic resonance imaging (MRI) data. The study uses a significantly larger dataset of 45,000 people and presents separate models for female and male brain age gap prediction, without considering any pairwise similarities between individuals.

Another framework for brain age gap prediction is based on Gaussian Processes regression, using raw T1-weighted MRI scans, segmenting them and using principal components analysis (PCA) for dimensionality reduction [2].

Chapter 2

Preparation

This section describes the main concepts related to the development of the population graph networks, the neuroimaging data they build on, and the neural network architectures that will be used to learn the patterns in those graphs.

2.1 Brain age estimation

2.1.1 Rationale behind the brain age

what is brain age how does it differ from the chronological age why chronological age can be used to predict the brain age

2.2 Data preprocessing

The United Kingdom Biobank (UK Biobank) [7]. This is a continuous, large, population-wide study of over 500,000 participants containing a wide range of phenotypic and genetic data that is used by the researchers to analyse the risk factors and development of various health conditions.

Of particular relevance to this dissertation are the UK Biobank participants with neuroimaging data records that have been denoised, motion-corrected and otherwise processed for further analysis, a total of 17,550 participants. The data has been initially preprocessed with the standard UK Biobank pipelines,[1] and further denoised, parcellated and kindly provided by Dr Richard Bethlehem of the Department of Psychiatry.[2] The details on this processing are described below.

[1] https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf

[2] Pipelines for preprocessing the UKB dataset at the Department of Psychiatry https://github.com/ucam-department-of-psychiatry/UKB

2.2.1 Structural data

Structural brain imaging data encapsulates brain features related to its structure, such as cortical thickness, white

Combined T1-weighted and T2-weighted FLAIR images which emphasise different aspects of the MRI scan and therefore might help to extract different structural features. For the dataset used in this dissertation, the combination of the two types of images to derive the structural features using the HCP Freesurfer pipeline.[3]

Euler indices

Euler indices are a quality control measure and corresponds to the number of times Freesurfer neuroimaging analysis software failed to connect two slices of the brain image. The higher the Euler index, the worse quality of the scan. These indices might be used to remove the subjects with low-quality scans to avoid them affecting the analysis [5]. Otherwise it can be used as a covariate in a machine learning model (as brain similarity metric or a node feature) to correct for any bias in prediction that might be related to scan quality.

Rafael Romero-Garica (rr480), Lisa Ronan (lr344) and Richard A.I. Bethlehem (rb643)

2.2.2 Functional data

Turns out functional data is not as effective as structural brain imaging data.

Description of correlation matrix computation.

2.2.3 Parcellation

A *parcellation* or *atlas* refers to the way the brain is split into meaningful regions for further analysis. Whether two voxels of a brain belong to the same parcel may depend on their proximity, empirical evidence of that the voxels are responsible for the same function and so forth, and can be used to compare the locations in two different brains. When the brain is imaged there is a choice whether to warp the image of the brain to the fixed atlas or whether to warp the atlas to match the variable brain images. The former makes it easier to process a dataset of many images and find the matching regions of two brains faster, but the latter remains more faithful to the unique structure of the individual patient's brain.

Both functional and structural datasets use one of the most common parcellations by Glasser et al. [4], which divides the brain into 360 cortical regions and 16 subcortical regions.

[3] https://www.ncbi.nlm.nih.gov/pubmed/23668970

2.2.4 Phenotype data

For the population graph construction, the neuroimaging features are associated with the individual subjects (nodes). The similarity metric is defined by phenotype data, which is all important but not directly neuroimaging-related data. Some examples of phenotype data that is related to the brain include the patient's sex, mental health, other potential health issues, full-time education, bipolar disorder status etc. Indeed the metric that is being predicted is correlated with the brain tissue age indicative of various neurological and neurodegenerative diseases.

2.3 Population graphs

2.3.1 Similarity metrics

2.3.2 Computational model

Train/validation/test split, cross-validation, patient selection and exclusion from results, stratification, graph representation (edge lists, node features, edge features,...)

2.4 Multilayer perceptrons

The multilayer perceptron maths, if appropriate.

2.5 Graph convolutional networks

2.6 Graph attention networks

2.7 Requirements analysis

Tasks to be implemented (according to proposal: work to be done, success criteria, possible extensions), their relative importance (priority) and difficulty. Provide the order in which the tasks should be carried out to show good planning skills and account for the changes in proposal where the preprocessing pipeline turned out to be more important than the neural network implementation.

2.8 Software engineering practice

Implementing a flexible preprocessing pipeline which could be customised in the future for a variety of machine learning tasks even outside graph neural networks (a package).

Modular structure encapsulating specific task and having well defined documentations of the others.

Description of software engineering techniques: planning out and executing the project based on requirements analysis, setting tasks, and smoothly meeting the success criteria.

Code reuse (of open source well tested libraries), follow documentation and follow the PEP-8 style guide (or whatever PyCharm encourages).

Incremental development.

Modular structure: e.g. data processing, graph construction, graph neural network modules, robustness evaluation framework. Figure out where validation and cross validation sections should be (while training, separately etc.)

Diagram of the pipelines and module interaction (like in google design docs)

2.9 Choice of tools

PyTorch, PyTorch geometric extension, graph spectral filters/convolutions, message passing, timeseries preprocessing into correlation matrices, IDEs, backup strategies

2.10 Starting point

- dataset, preprocessed by Dr Richard A.I. Bethlehem
- PyTorch, PyTorch geometric implementing GCN and GAT APIs and the graph API
- no previous experience with graph neural networks or the mathematics behind it
- no previous experience with PyTorch; limited experience with machine learning frameworks (basics of TensorFlow), no experience with neuroimaging data

Chapter 3

Implementation

Could be split into *preprocessing, parameter tuning, evaluation framework, software engineering techniques, repository overview*

3.0.1 Precomputation and preprocessing of connectivity matrices

3.0.2 Structural and functional data extraction

3.0.3 Graph construction pipeline

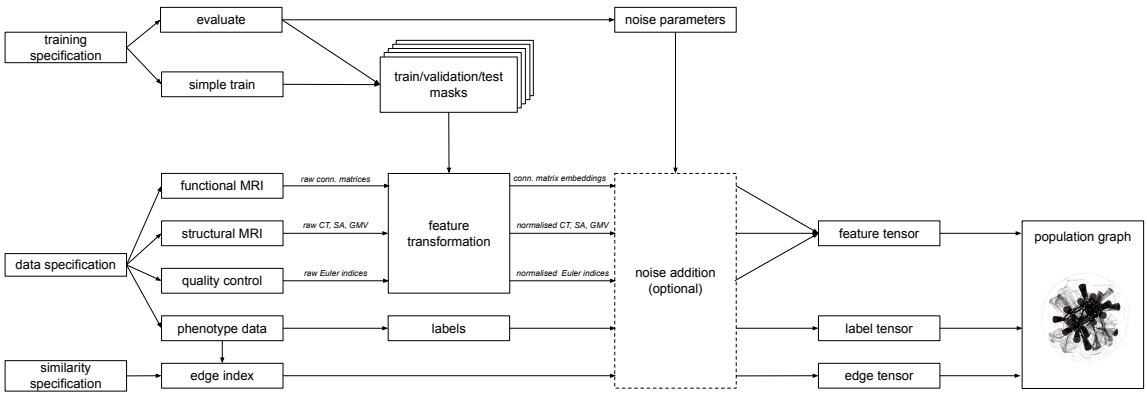

Figure 3.1: Graph pipeline.

3.0.4 Train, test, validation split

3.1 Non-graph baselines

I have played around with *xgboost*, ElasticNet and simple multilayer perceptrons to get a good idea of the minimum baselines my new architectures need to reach to be considered more effective.

3.2 Graph convolutional network

Describe the architecture in `BrainGCN`

3.3 Graph attention network

3.4 Robustness evaluation framework

3.5 Repository overview

Figure out how to split framework modules.

Could have a base BrainGNN class which can then be *extended* with BrainGAT, BrainGCN.

Chapter 4

Evaluation

4.1 Overview of the predictive power for the two graph neural network approaches

Discuss Pearson's r, coefficient of determination r^2 and other common performance metrics.

4.2 Hyperparameters

Effects of hyperparameter tuning

4.3 Unit testing

4.4 Other quantitative and qualitative results

Statistical significance of the results compared to the baseline (e.g. Pearson's r method in Python seems to return some p-value of it..?)

4.5 Robustness of the graph neural networks to noisy and missing data

Add noise to the graphs and measure the rate of drop in predictive power.

4.6 Comparison against existing benchmarks

Compare to the Kaufmann et al.'s *xgboost* approach [5] ($r \sim 0.93$); and the other package that was cited in the same paper.

Possibly compare to other non-graph (relatively baseline) (neural network) architectures, e.g. ElasticNet, MLP,...

4.7 Interpretation of the model behaviour

One of the advantages of the graphs that they *should* be interpretable I guess.

Chapter 5

Conclusion

5.1 Successes and failures

5.2 The project in hindsight (lessons learnt)

Lessons learnt: graph *representation* is more important than the framework used. Good representations of the dataset can help guide the learning algorithm in the right way as it gets the intuitions faster just because of the way the data was represented in the first place. It's not good if it captures bias, but in this case representation made the difference between the model not learning anything and the model getting great results.

Was a mistake analysing functional data first without analysing the other methods in detail. Functional imaging data by itself gave very high dimensionality which either could not be learnt by the network because of the low number of examples or other factors. Most of the literature uses just the structural data for age prediction, and indeed this turned out to be more effective. Also makes sense intuitively as structural features would be related to the signs of brain atrophy while it is not clear the pattern of how resting state brain activity would change with ageing brain.

Mention Niu et al. 2019 raising the issue that there is systematic bias in brain age gap prediction but not many studies use this knowledge to correct for it.

5.3 Possible continuations of the project

- Include DNA methylation data as it is widely used in other studies and is claimed to improve the predictive power of the model [2].

Bibliography

[1] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.

[2] James Cole, Stuart Ritchie, Mark Bastin, Maria Valdés Hernández, et al. Brain age predicts mortality. *Molecular Psychiatry*, 23(5):1385–1392, 2018.

[3] Katja Franke and Christian Gaser. Ten years of BrainAGE as a neuroimaging biomarker of brain aging: What insights have we gained? *Frontiers in neurology*, 10:789, 2019.

[4] Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171, 2016.

[5] Tobias Kaufmann, Dennis van der Meer, Nhat Trung Doan, Emanuel Schwarz, et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience*, 22(10):1617–1623, 2019.

[6] Xin Niu, Fengqing Zhang, John Kounios, and Hualou Liang. Improved prediction of brain age using multimodal neuroimaging data. *Human Brain Mapping*, 2019.

[7] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779, 2015.

Appendix A

Project Proposal