ATTENTION-BASED GRAPH CLASSIFICATION FOR PREDICTION IN IMMUNOTHERAPY RESPONSE

Anđela Todorović[1], Milan Bašić[2],

[1] Ecole Polytechnique, Institut Polytechnique de Paris
Email: andjelatodorovich@gmail.com

[2] Faculty of Sciences and Mathematics, University of Niš
Email: basic_milan@yahoo.com

Abstract

Traditional techniques of image processing, although powerful, usually lack understanding of spatial entity-level connectedness, thus justifying the widely investigated utilization of graphs for representing and manipulating imagery data. Although the later approach has shortcomings, mainly related to the time complexity of the algorithms, there have been proven problems in various areas of image processing, including digital pathology, where graph-representations are not only compelling to investigate, but also needed. One such example is prediction of an immunotherapy response, where it is commonly believed that in-depth understanding of the tumor tissue microenvironment and spatial relations play a crucial role for the successful prediction. In essence, immunotherapy response prediction is reduced to the problem of binary graph classification, with classes coresponding to either no response or complete response to immunotherapy. In this paper, we propose the usage of attention-based graph neural network architectures, more specifically, PiNet for learning inner tissue interactions and predicting immunotherapy response in patients diagnosed with melanoma, by obtaining a graph representation of whole-slide imagery data. PiNet a generalised differentiable a ttention-based pooling mechanism for utilising graph convolution operations for graph level classification. Furthermore, we compare the results obtained by PiNet to the state-of-the art Graph Convolutional Network (GCN) and propose solutions to the imbalance problem in the dataset. Finally, we present satisfactory classification r esults i n e xperiments conducted on the dataset of whole-slide images of melanoma tissue, with F_1 score obtaining a value of 0.45 compared to the same metric on GCN being equal to 0.1 in the under-represented dataset class.

Key words: graph representation learning, GCN, PiNet, digital pathology

1 Introduction

The prediction of the success level of immunotherapy, promising type of cancer treatment, has been proven to depend vastly on tracking and understanding interaction between different cell types, especially tumor-infiltrating lymphocytes in invasive margin. Therefore, in order to observe cell and graph level relations, a technique in digital pathology, known as WSI (whole-slide imaging) has been vastly used for collecting and representing the data. Whole-slide images, which refer to gigapixel images obtained by scanning tumor biopsies at a high resolution, are usually processed by tiling (dividing original images into smaller parts). The representations of the resultant tiles/patches, whether learned or hand-crafted, need to be aggregated in order to obtain slide-level representations or labels. To perform this task, we learn inter-entities spatial interaction by

representing individual tiles as nodes and constructing a high-level graph that quantifies the structural relationship between different tiles. After obtaining a graph representation, newly formed graphs can be prone to a variety of tasks (among which, node classification, link prediction, network similarity, community detection, graph classification etc.). In this paper, we tackle the problem of graph classification, i.e. predicting a label to each graph in a given set based on the certain statistics (i.e. graph features).

This paper is organized as follows: First, we will give some theoretical backgound describing the task of graph classification and graph neural network approach to this task. Ater this step, we give particular information about the dataset and its graph representation. Finally, we describe the conducted experiments with PiNet model and compare it's performance to the GCN.

2 Theoretical background

In this paper, we consider the problem of predicting a label $y \in \{0, 1\}$ for an unseen test graph, given a training set $\mathcal{G}$. A graph $G \in \mathcal{G}$ is defined by the order pair (A, X), where $A \in R^{N\times N}$ is the graph adjacency matrix, and $X \in R^{N\times d}$ is vertex features matrix. PiNet model consists of message passing convolution network $\psi : R^{N\times N} \times R^{N\times d} \rightarrow R^{N\times F'}$ with an arbitrary number of layers combined by a matrix product. Formally, the final output of the model $Z(G)$, where each coordinate is the predicted label, is given by the function

$$Z(G) = \sigma_s[g(\sigma_s(\psi_A(A, X)^T) \cdot \psi_X(A, X))W_D] \in R^2,$$

where σ_s denote softmax activation function, g is the function that reshapes a matrix into the vector by concatenating its rows, ψ_A and ψ_X are separate message passing networks for learning attention coefficients and vertex features, respectively and W_D is a weights matrix for a fully connected dense layer. PiNet model most commonly consists of a pair of double-stacked passing layers before product and dense layer, that is, $\psi_A(A, X) = \psi_X(A, X) = \sigma_r(\hat{A} \cdot \sigma_r(\hat{A}XW^{(0)})W^{(1)})$, where σ_r denotes the rectified linear unit function, $\hat{A} = (pI + (1-p)D)^{-\frac{1}{2}}(A + qI)(pI + (1-p)D)^{-\frac{1}{2}}$, D is the diagonal matrix of A and $0 \leq p, q \leq 1$. The use of two additional trainable parameters permits extra attention mechanism that enables the model to weight the importance of symmetric normalisation of the adjacency matrix, and the addition of self loops.

3 Dataset description and preprocessing

Our working dataset consists of 85 whole-slide images, out of which 66 from non-response class (0) and 19 images from a complete response class (1). Whole-slide images depict tumor tissue with different cells (T-cells, B-cells, macrophages etc.) and cell-level features (phenotypes), but in general divided into two categories: *Tumor* and *Stroma*.

Out of these whole-slide images, we build cell-level graphs connecting by modelling cells and tissue interactions as described by Yaner et al. [1]. After obtaining cell-level graphs, we group spatially neighboring cells into supercells or patches by considering square sliding windows of dimensions $l \times l$ for predetermined parameter l. Patches (supercells) are then connected into a grid graph (all horizontal and vertical neigbours are connected). Upon construction of the grid graph, we extract and assign node-level features such as phenotype count and count of edges connecting each type of cell. This allows us to perform a classification based on the feature matrix X, and adjacency matrix A, as described in the previous section.

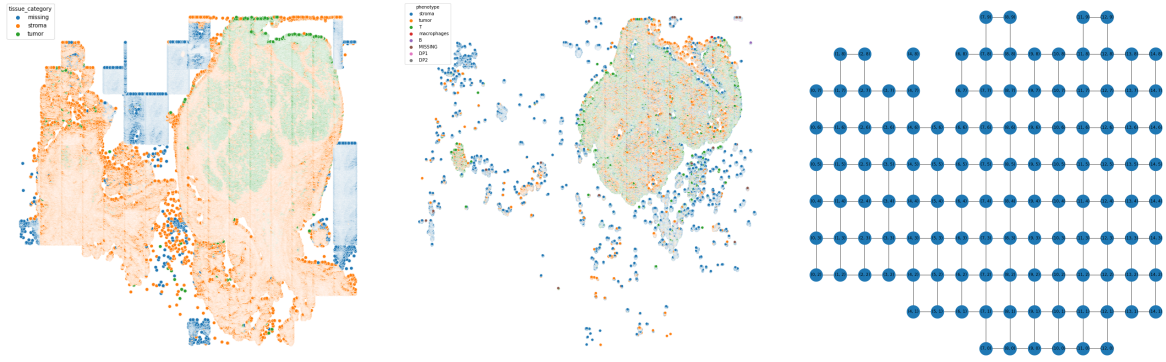

Table 1: Example slide of a whole slide image from a complete-response class colored to separate tumor tissue in green from stroma regions in orange (left); detailed scatterplot of the phenotypes in the stroma region (middle); grid graph (right)

4 Experimental setup

In order to ensure that the data is fairly distributed, we use a combination of stratified 10-folds and Monte Carlo sampling (since some splits could be repeated more than once), making sure that we maintain the same proportion of class distribution while choosing train and test sets each time. Considering the imbalance in classes, we introduce class penalty of 3 (proportional to the class frequency, and proven optimal for). We train both models for a maximum of 80 epochs (training iterations) using Adam as an optimizer with a learning rate of 0.01 and early stopping. As a loss function, we will use focal loss, which is a variant of the cross entropy loss that better handles class imbalance. Focal loss is defined as $FL(p) = -\alpha(1-p)^{\gamma}log(p)$. Parameters we choose for focal loss are $\alpha = \gamma = 0.5$

5 Results

The following figures depict accuracy of classification on train and test set for both PiNet and GCN models. Furthermore, we plot dice coefficient (F_1 score) as a metric of classification precision and recall in each of the classes.

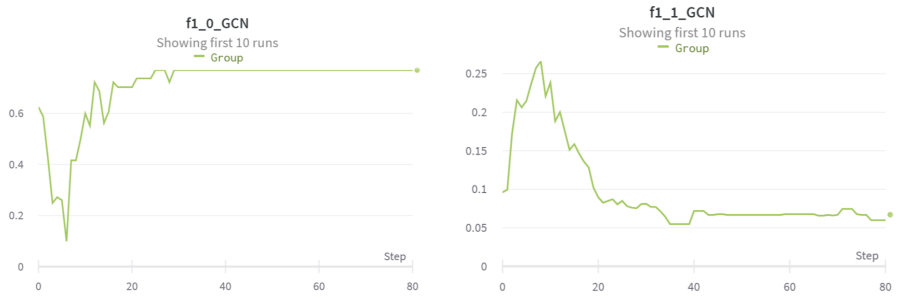

Figure 1: F_1 score on train and test set obtained by the GCN model with hyperparameters (leaning rate = 0.01, batch size = 10, in channels = 64, hidden channels = 10, dropout = 0.1).

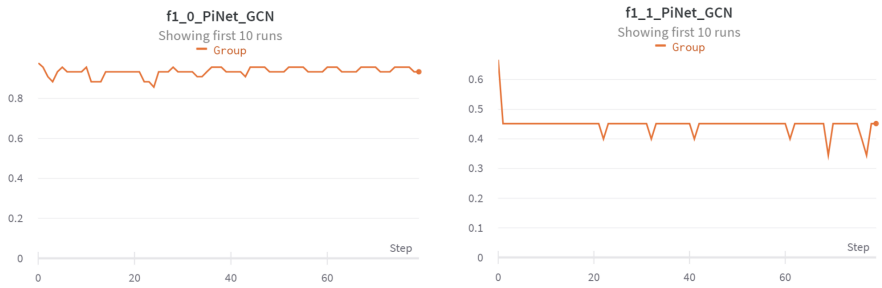

Figure 2: F_1 score (dice coefficient) obtained by the $PiNet$ model with hyperparameters (leaning rate = 0.01, batch size = 10, in channels = 100, hidden channels = 64, dropout = 0.5, $p = q = 0$).

$PiNet$ model achieves better accuracy on train set (0.889 compared to 0.7705) and test set (0.875 compared to 0.625). As mentioned earlier, class imbalance can cause the model to be biased towards one class during training, thus resulting in lower classification accuracy on the under represented class. $PiNet$ succeeds to overcome the problem of imbalance in classes by achieving decently high F_1 scores on both classes 0 and 1 (0.93 and 0.45, respectively), compared to the GCN model that suffers from abrupt drop in F_1 score on class 1 and converges around the value of 0.05.

6 Conclusions

In this paper, we determine to which extent novel attention-based graph classifiers can learn to separate the data subject to the immunotherapy response. Although satisfactory, results are more robust compared to the GCN, but still sensitive in convergence. In general, this can possibly be improved by running the model on larger number of epochs, using the techniques of handling dataset imbalance (such as undersampling), or by performing hyperparameter tunning. It is worthwhile to performing the pooling mechanism by symmetric feature-weighted mean operator for different values of p and q (also note that p and q may be different for each message passing layer).

References

[1] Yener, Bülent et. al, Cell-Graphs: Image-Driven Modeling of Structure-Function Relationship, Communications of the ACM, Volume 60, Issue 1,74–84, January 2017, https://doi.org/10.1145/2960404

[2] Peter Meltzer, Marcelo Daniel Gutierrez Mallea, Peter J. Bentley: A Permutation Invariant Graph Neural Network for Graph Classification, 2020, https://arxiv.org/abs/2008.04575

[3] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," International Conference on Learning Representations (ICLR), sep 2016, https://arxiv.org/abs/1609.02907