Predicting Biological Age and Sex using Spatial Temporal Graph CNN for functional MRI data

Soham Gadgil
sgadgil@stanford.edu

Sunwoo Kang
swkang73@stanford.edu

Erick Fidel Siavichay-Velasco
esiavich@stanford.edu

Abstract

Functional Magnetic Resonance Imaging (fMRI) is a technique to measure brain activity by quantifying blood oxygen level (BOLD signal) (1). There have been different studies showing that human functional organization undergoes changes related to age and exhibits sexual dimorphism (2; 3; 4). We aim at leveraging deep learning techniques to develop a model that can be used to accurately predict the biological age and sex based on fMRI data. We used the state of art Spatial-Temporal Graph Convolutional Networks (ST-GCN), which moves beyond the limitations of previous methods by automatically learning both the spatial and temporal patterns from data. As the first attempt to apply spatial-temporal convolution on fMRI data, we overcame the challenge of limited sample size and are able to get to get an accuracy of 78% for sex and 63% for age (binary classification for young v/s old). These results manifest ST-GCN as the novel model to predict biological sex and age from fMRI data.

1 Introduction

Functional MRI (fMRI) is a state of art standard of care for monitoring brain activity in clinical settings. The goal of this project is to investigate whether functional MRI activities could be used to predict biological sex and age. Such analysis can enhance the understanding of normal neuromaturation, characterize developmental disruption caused by neurological disorders, and explain differences in cognitive performance among men and women. The input to our network was the resting state functional MRI, which contains the blood oxygen level (BOLD) of the different brain regions. Using Spatial Temporal Graph Convolutional Network (ST-GCN), we produced the output of biological sex and age. The final output of our model was binary sex (male / female) and binary age (young / old adult). [1]

2 Related work

Our work uses resting state fMRI from Human Connectome Project respository (5). There has been some previous work on done on this dataset. Zhang 2018 applied a predefined functional template and partial least squares regression modeling of the brain region for the purposes of sex classification (6). Quereshi et. al 2019 has achieved 98.09% accuracy for schizophrenia discrimination on resting state fMRI (7). The previous works on this dataset have relied on hand engineered features of the brain regions for sex classification. To our knowledge, our work would be the first effort to classify biological age as well as sex from the HCP dataset using a deep neural network.

[1] Our code is available at https://github.com/ericksiavichay/cs230-final-project

CS230: Deep Learning, Autumn 2019, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

The ST-GCN architecture we leveraged in our project was first developed by Yan, Xiong, and Lin 2018 for the purpose of human action recognition based on human skeleton models from video frames (8). The strength in their approach is that they account for both spatial and temporal features during the network training. The network has been used on dynamic action recognition in UFC sport video data (9), or hand gesture recognition (10). Our work includes applying this state of art network architecture to a different domain. We implemented their network architecture and altered the graph initialization through directly feeding in the adjacency matrix.

3 Dataset and Features

We used functional MRI scans from the public Human Connectome Project (5). Our data was pre-processed by the HCP functional pipeline available at the Connectcome DB. Our samples came from 1108 individuals, consisting of 603 female and 505 male subjects. The subjects came from the age range of 25 to 37. Each of our pre-processed data file consisted of oxygen level measurement from different brain regions for the time course of 1200 timesteps from a single subject. We initially used fine grain measurement from 360 different brain regions for the project, but during the network training, our network performance increased with coarser grain resolution. Thus, we used downgraded measurement from 22 different brain region for the network training (11).

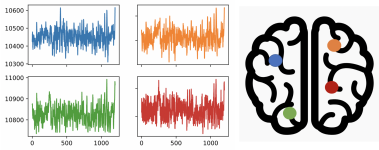

Figure 1: Sample average blood oxygen level (BOLD) resting-state fMRI sessions from a four arbitrarily chosen representative nodes

We performed data augmentation in which the fMRI data was fragmented by timesteps. The original file for each individual contained measurement of 22 brain regions over the course of 1200 timesteps; we split each file into 12 so that each input data contains fMRI measurement of 22 regions for the duration of 100 timesteps. That increased our sample size from 1106 to 13,082. We used 0.7/0.3 subject level split to create our training / testing set of size 9148 / 3933, respectively.

4 Methods

We used a novel model, called *Spatial-Temporal Graph Convolutional Network (ST-GCN)* (8) for our approach to predict sex and age. This involves the use of a graph network as the input to the GCN. In the case of f-MRI data, the 22 brain regions act as the nodes in the graph and the correlations between the regions act as the edges. The convolutional operation on graphs consists of the input feature map residing on a spatial graph. This means that the input feature map has a vector on each node of the graph.

For modeling the temporal component of the graph, we connected the same brain regions across consecutive timesteps. The number of timesteps to look ahead is defined by the temporal kernel size parameter, set to be 5 for our purposes. Figure 2 shows a visualization of the temporal sequence. The output feature map f_{out} for a given input feature map f_{in} is defined below (12)

$$f_{out} = D^{-1/2}(A+I)D^{-1/2}f_{in}W \tag{1}$$

where A is the adjacency matrix, I is the identity matrix, W is the weight matrix, and D is a diagonal matrix defined as $D^{ii} = \sum_j (A^{ij} + I^{ij})$. The input feature map can be represented as a tensor of dimensions (C, V, T), where C is the number of channels (1 in our case), V is the number of nodes in the graph (22 in our case), and T is the number of timesteps in the sample (100 in our case after data augmentation). The graph convolution is implemented by performing a 2D convolution followed by multiplying the resulting tensor with the normalized adjacency matrix $D^{-1/2}(A+I)D^{-1/2}$.

For sex classification, the loss function used was the binary cross entropy/log loss, as shown in equation 2.

$$\mathcal{L}(y_i, \hat{y_i}) = -\frac{1}{m}\sum_{i=1}^{m} y_i log(\hat{y_i}) + (1 - y_i) log(1 - \hat{y_i}) \tag{2}$$

where m is the number of training samples, y_i is the true label of the ith sample, and $\hat{y}_i$ is the predicted label of the ith sample.

For age classification, we first tried using a regression model with MSELoss as shown in Equation 3. We also tried converting the regression model into a classification problem by segregating the ages into two sets, old and young, based on a threshold of age 28. For this part, the loss function used was the log loss shown in Equation 2.

$$\mathcal{L}(y_i, \hat{y_i}) = \frac{1}{m}\sum_{i=1}^{m} (y_i - \hat{y_i})^2 \tag{3}$$

With this given setup, our input functional MRI was used to obtain the node values and edge weights to initialize the graph structure. The node values corresponded to the blood flow of the region, and the edge weights were the correlation values between two distinct regions in the brain.

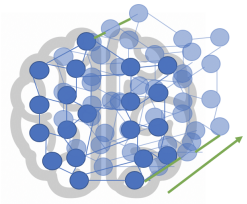

Figure 2: The spatial temporal graph of a fMRI sequence used in this work where the proposed ST-GCN operate on. Blue dots denote the brain region. The inter-frame edges connect the same regions between consecutive timelines.

5 Experiments/Results/Discussion

5.1 Network Architecture

The ST-GCN model was similar to the one used in (8). The same model was used for both sex and binary age classification. A batch normalization layer is used at the start to normalize the data. The model consists of 9 spatial temporal graph convolutional layers. The first three layers have 64 output channels, the next three layers have 128 output channels, and the last three layers have 256 output channels. The temporal kernel size used is 5. Each ST-GCN unit uses the Resnet mechanism (13) to prevent the problem of vanishing gradients by skipping some layers. The stride is set to 1 for all temporal convolutional layers except for pooling layers 4 and 7, where it is set to 2. Global average pooling is performed on the output of the last layer and it is fed into a SoftMax classifier with two outputs - Male, Female for sex classification and Old, Young for binary age classification. For predicting age with regression, the model is kept the same except there is no SoftMax classifier at the end and there is only a single output representing the predicted age. We used the $p2.xlarge$ aws instance for age and $p3.2xlarge$ aws instance for sex classification.

The models were trained using batch gradient descent with a batch size of 64. This size was chosen partly because it performed the best based on experimenting with different batch sizes and partly because the batch size was limited by the compute resources of the aws instances used for training. The learning rate was initialized to 0.1 for sex and binary age classification. For the full age regression task, the learning rate of 0.01 was used for 200 iterations in order to fine grain the weight training. Another parameter to tune was the window size, which selected a random time chunk from the input data to feed to the network. This was chosen to be 25 for our model. The PyTorch deep learning framework (14) for used for implementing this network.

5.2 Experiments and Results

5.2.1 Sex Classification with functional MRI data

Figure 3 shows the loss and accuracy curves obtained on the training and testing sets for sex classification. Both of the loss curves showed similar trends and the loss decreased as more epochs pass, as expected. However, the loss curves seemed erratic, and the loss jumped to higher values for some epochs. These spikes occurred around the same number of epochs for both the training and testing loss. This could be because of batch gradient descent being used which resulted in an erratic loss curve since weight updates were done more frequently. The training loss converged to a value lower than the testing loss, which showed that there might be some overfitting.

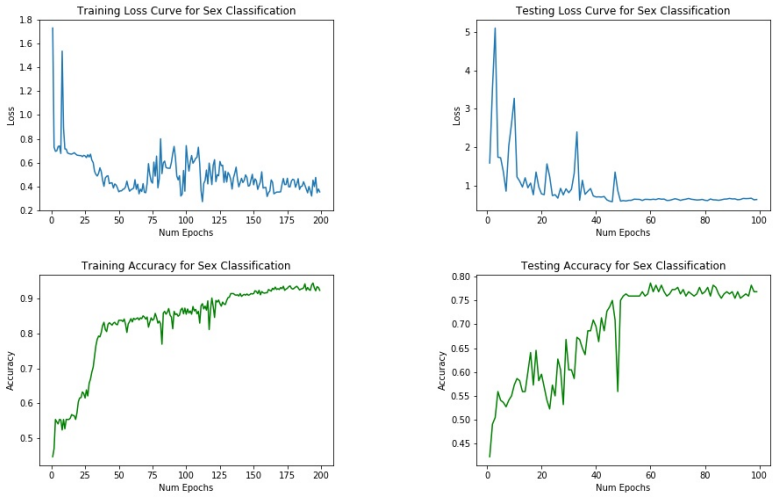

Figure 3: Plots showing the training and testing loss and accuracy curves for sex classification

The accuracy curves, as expected, showed similar trends as well, and the accuracy increased as more epochs passed. The converged training accuracy was higher than the converged testing accuracy, which again suggested overfitting. However, the testing accuracy was much higher than 50%, which showed that there are indeed functional differences in the way male and female brains function. The network was not guessing randomly and was able to learn these differences to separate male and female brains.

Figure 5 shows the confusion matrix obtained. Male was the positive label and Female was the negative label . Based on these labels, the Precision was 0.7563, the Recall was 0.833, and the F1 score was 0.793. The model predicted male and female brains with similar accuracy. The precision was lower than the recall, which showed that the model had more false positives indicating that it predicted the sex as male more often.

5.2.2 Age Classification with functional MRI data

Our initial effort was to treat age as an regression task, since biological age is continuous. To perform the regression, we modified our network to compute a single output value. We used Mean Square Error (MSE) loss and R^2 score as the measure of accuracy. But such a regression setting demonstrated aberrant behavior in which the training loss and accuracy converged with the iteration while the testing loss increased and accuracy fell (Figure 4). According to the documentation (15), negative R^2 score indicates worse regression performance than the mean average model. Implementing learning rate decay or changing the window size did not change this trend in training and testing. We speculate that these phenomena were due to the tight age range of the dataset being indifferentiable. The next step was to loosen the prediction criteria by transforming the fine-grained regression to a coarser binary classification task.

We discretized the age into a binary age classification problem where we tried to predict whether the sample was younger than the threshold age of 28. The threshold was set to most evenly bin the data

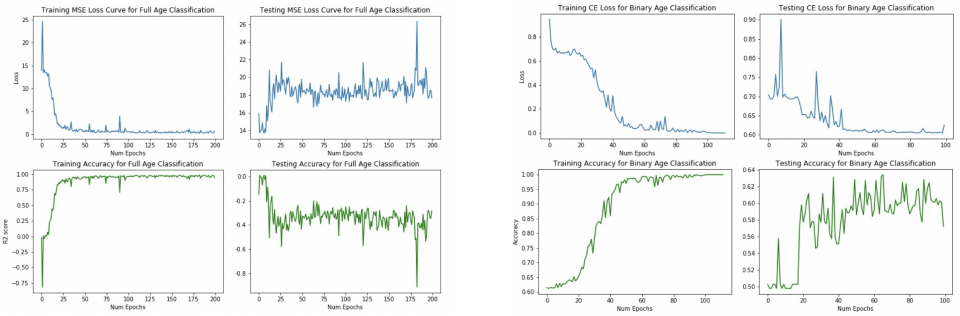

Figure 4: **Left**: Plots for training and testing accuracy for full age with MSE Loss. **Right**: Plots for training and testing accuracy for binary age with CE loss

into the two classes. This led to the expected trend of decrease in loss and increase in accuracy during both training and testing phase (Figure 4). The convergence in loss occurred at iteration of 60 in both training and testing dataset. The training loss converged below 0.1 whereas the testing loss converged at 0.6, which is one of the common marks for overfitting.

The sign of overfitting was also observed during the accuracy analysis (Figure 4). The maximum training accuracy was 1 which was markedly greater than the maximum testing accuracy of 0.63, suggesting an overfit in the data. But the prediction value greater than 0.5 indicates that the network learned the age dependent functional differences in the fMRI dataset.

Since the task is reduced to binary classification, we were able to further evaluate our network through the confusion matrix (Figure 5). Our network had a precision of 0.6457, a Recall of 0.8692, and the F1 score of 0.7410 with Young as the positive label. The recall was higher than the precision, which showed that the model is likely to result in the predict label of Young than Old. The true positives (113) were greater than the false positives (62) and the true negatives (27) were greater than the false negatives (17), which represented the success of the model to distinguish the biological age given fMRI.

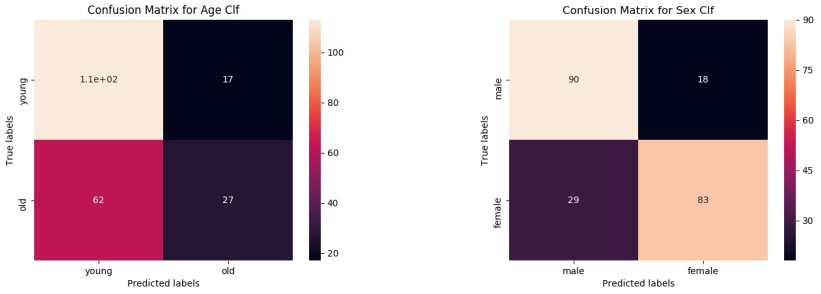

Figure 5: Confusion Matrix for Age and Sex Classification

6 Conclusion/Future Work

In this project, we performed both classification and regression tasks to predict biological age and sex given the resting state functional MRI dataset. We used a novel model architecture ST-GCN which had not been previously deployed in this field. The maximum testing accuracy for sex was 78%, and for binary age was 63%. This indicates that functional MRI is more informative to predict sex than age. A higher-than-random accuracy also suggests that there are indeed functional differences in the brain for different ages and genders. For both age and sex prediction, we encountered significant differences between train and test accuracy. This indicates the susceptibility of our model to overfitting. Given more time, we would like to reduce overfitting through dropout and weight regularization. We would also like to see whether subject level prediction instead of sample level prediction (because of data augmentation) increases the accuracy of the prediction.

7 Contributions

Sunwoo mainly contributed to the project by exploring the HCP subject levels, generating adjacency matrix, and running the age classification. During the age classification, she modified the network architecture to implement MSE Loss with a single output result. She took part in writing the project proposal, milestone, and final writeup. She was also in charge of employing AWS instances. Soham mainly worked on creating the testing and the training sets from the original dataset, running the sex classification, and tuning the hyperparameters of the network to improve performance. He also helped with the proposal, milestone, and the final writeup. Erick contributed to initial data analysis of HCP dataset. He took part in project proposal, milestone, and contributed to the final poster. He also maintains the master codebase, including data results and python notebooks, on Github. We would also like to thank our mentor, Qingyu Zhao, for his mentoring and support. Qingyu provided us with the original dataset and the coarser dataset, and also helped us in understanding the functional aspects of the ST-GCN model. He also provided us with an analysis of sex and age classification on structural MRI data, which helped us in identifying similarities with the functional MRI data. Throughout the project, he provided feedback through several in-person meetings.

References

[1] G. H. Glover, "Overview of functional magnetic resonance imaging," *Neurosurgery clinics of North America*, vol. 22, no. 2, pp. 133–vii, Apr 2011, 21435566[pmid]. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/21435566

[2] H. M. Aycheh, J.-K. Seong, J.-H. Shin, D. L. Na, B. Kang, S. W. Seo, and K.-A. Sohn, "Biological brain age prediction using cortical thickness data: A large scale cohort study," *Frontiers in Aging Neuroscience*, vol. 10, p. 252, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnagi.2018.00252

[3] J. Zhai and K. Li, "Predicting brain age based on spatial and temporal features of human brain functional networks," *Frontiers in human neuroscience*, vol. 13, pp. 62–62, Feb 2019, 30863296[pmid]. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/30863296

[4] S. Weis, K. R. Patil, F. Hoffstaedter, A. Nostro, B. T. T. Yeo, and S. B. Eickhoff, "Sex Classification by Resting State Brain Connectivity," *Cerebral Cortex*, 06 2019, bhz129. [Online]. Available: https://doi.org/10.1093/cercor/bhz129

[5] D. V. Essen, K. Ugurbil, E. Auerbach, D. Barch, T. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. Curtiss, S. D. Penna, D. Feinberg, M. Glasser, N. Harel, A. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S. Petersen, F. Prior, B. Schlaggar, S. Smith, A. Snyder, J. Xu, and E. Yacoub, "The human connectome project: A data acquisition perspective," *NeuroImage*, vol. 62, no. 4, pp. 2222 – 2231, 2012, connectivity. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811912001954

[6] C. Zhang, C. C. Dougherty, S. A. Baum, T. White, and A. M. Michael, "Functional connectivity predicts gender: Evidence for gender differences in resting brain connectivity," Apr 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/29322586

[7] M. N. I. Qureshi, J. Oh, and B. Lee, "3d-cnn based discrimination of schizophrenia using resting-state fmri," *Artificial Intelligence in Medicine*, vol. 98, pp. 10 – 17, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0933365719301393

[8] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," 2018.

[9] W. Zheng, P. Jing, and Q. Xu, "Action recognition based on spatial temporal graph convolutional networks," in *Proceedings of the 3rd International Conference on Computer Science and Application Engineering*, ser. CSAE 2019. New York, NY, USA: ACM, 2019, pp. 118:1–118:5. [Online]. Available: http://doi.acm.org/10.1145/3331453.3361651

[10] Y. Li, Z. He, X. Ye, Z. He, and K. Han, "Spatial temporal graph convolutional networks for skeleton-based dynamic hand gesture recognition," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, 2019.

[11] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, p. 171, 2016.

[12] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[14] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, pp. 8024–8035.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.