Brain Network Transformer

Xuan Kan[1] Wei Dai[2] Hejie Cui[1] Zilong Zhang[3] Ying Guo[1] Carl Yang[1]

[1]Emory University [2]Stanford University [3]University of International Business and Economics

arXiv Code

Summary

Human brains are commonly modeled as networks of Regions of Interest (ROIs) and their connections for the understanding of brain functions and mental disorders. Recently, Transformer-based models have been studied over different types of data, including graphs, shown to bring performance gains widely. In this work, we study Transformer-based models for brain network analysis. Driven by the unique properties of data, we model brain networks as graphs with nodes of fixed size and order, which allows us to (1) use connection profiles as node features to provide natural and low-cost positional information and (2) learn pair-wise connection strengths among ROIs with efficient attention weights across individuals that are predictive towards downstream analysis tasks. Moreover, we propose an Orthonormal Clustering Readout operation based on self-supervised soft clustering and orthonormal projection. This design accounts for the underlying functional modules that determine similar behaviors among groups of ROIs, leading to distinguishable cluster-aware node embeddings and informative graph embeddings. Finally, we re-standardize the evaluation pipeline on the only one publicly available large-scale brain network dataset of ABIDE, to enable meaningful comparison of different models. Experiment results show clear improvements of our proposed Brain Network Transformer on both the public ABIDE and our restricted ABCD datasets.

Challenges

Complete graph. The simplest and most frequently used methods to construct a brain network is via pairwise correlations between BOLD time courses from two ROIs. This impedes the designs like centrality, spatial, and edge encoding because each node in the brain network has the same degree and connects to every other node by one hop.
Positional embeddings. In brain networks, the connection profile, which is defined as each node's corresponding row in the brain network adjacency matrix, is recognized as the most effective node feature. This node feature naturally encodes both structural and positional information, making the aforementioned positional embedding design based on eigenvalues and eigenvectors redundant.
Scalability. The numbers of nodes in molecule graphs are less than 50. However, the node number for brain networks is generally around 100 to 400, while the edge number can be up to 160,000. Therefore, operations like the generation of all edge features in existing graph transformers can be time-consuming, if not infeasible.

Framework

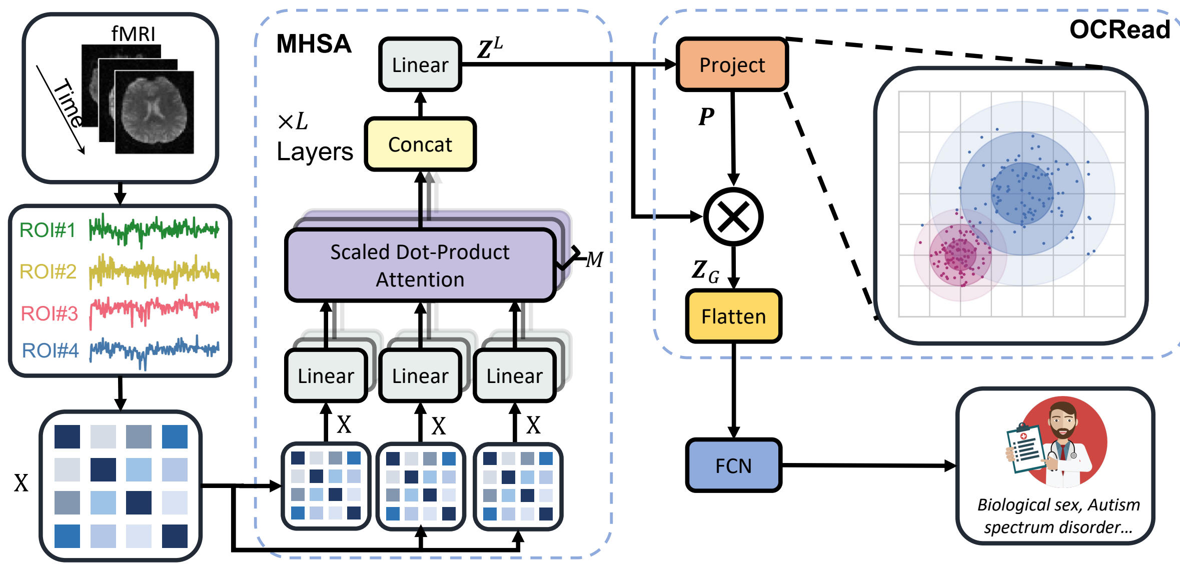
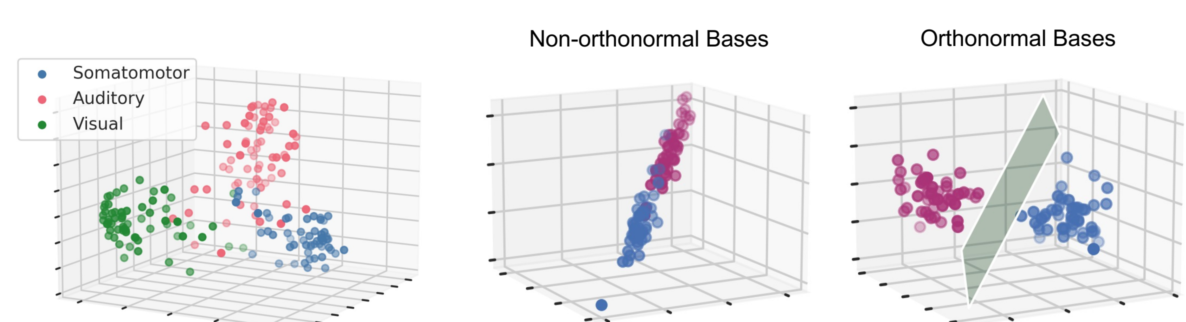


Figure: The overall framework of our proposed Brain Network Transformer.

Multi-Head Self-Attention (MHSA). Formally, we leverage a L-layer non-linear mapping module to generate more expressive node features $\boldsymbol{Z}^L = \mathrm{MHSA}(\boldsymbol{X}) \in \mathbb{R}^{V\times V}$. For each layer l, the output $\boldsymbol{Z}^l$ is obtained by

$$\boldsymbol{Z}^l = (\|_{m=1}^{M} \boldsymbol{h}^{l,m})\boldsymbol{W}_{\mathcal{O}}^{l}, \boldsymbol{h}^{l,m} = \mathrm{Softmax}\left(\frac{\boldsymbol{W}_{\mathcal{Q}}^{l,m}\boldsymbol{Z}^{l-1}(\boldsymbol{W}_{\mathcal{K}}^{l,m}\boldsymbol{Z}^{l-1})^\top}{\sqrt{d_{\mathcal{K}}^{l,m}}}\right)\boldsymbol{W}_{\mathcal{V}}^{l,m}\boldsymbol{Z}^{l-1}, \quad (1)$$

Orthonormal Clustering Readout (OCRead). Based on the motivation shown in Figure(a), We design a novel readout function to take advantage of the modular-level similarities between ROIs in brain networks, where nodes are assigned softly to well-chosen clusters with an unsupervised process.
Formally, given K cluster centers, each center has V dimensions, $\boldsymbol{E} \in \mathbb{R}^{K\times V}$, a Softmax projection operator is used as the function to calculate the probability $\boldsymbol{P}_{ik}$ of assigning node i to cluster k, $\boldsymbol{P}_{ik} = \frac{e^{\langle \boldsymbol{Z}_{i\cdot}^L, \boldsymbol{E}_{k\cdot}\rangle}}{\sum_{k'}^{K} e^{\langle \boldsymbol{Z}_{i\cdot}^L, \boldsymbol{E}_{k'\cdot}\rangle}}$, where $\langle\cdot,\cdot\rangle$ denotes the inner product and $\boldsymbol{Z}^L$ is the learned set of node embeddings from the last layer of MHSA module. With this computed soft assignment $\boldsymbol{P}$, the original learned node representation $\boldsymbol{Z}^L$ can be aggregated to obtain the graph-level embedding $Z_G = \boldsymbol{P}^\top \boldsymbol{Z}^L$. Besides, we also leverage the Gram-Schmidt process to obtain the orthonormal bases $\boldsymbol{E}$, which can facilitate the learning of clusters and embeddings as shown in Figure(b),



(a) Node features projected to a 3D space with PCA. Colors indicate functional modules.

(b) Orthonormal bases can make indistinguishable nodes in non-orthonormal bases easily distinguishable.

Experiments

Dataset. ABIDE contains brain networks from 1009 subjects, with 516 (51.14%) being Autism spectrum disorder (ASD) patients (positives). The region definition is based on Craddock 200 atlas. ABCD includes 7901 subjects, with 3961 (50.1%) among them being female. The region definition is based on the HCP 360 ROI atlas.
RQ1. How does BrainNetTF perform compared with state-of-the-art models of various types?

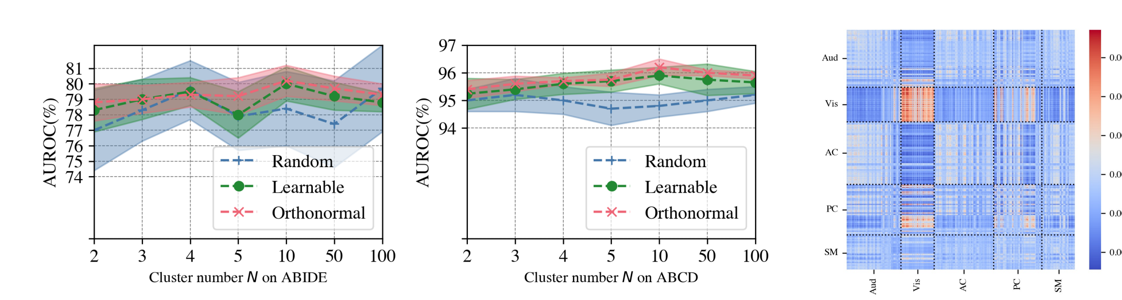
Table: Performance comparison with different baselines (%).

Type	Method	Dataset: ABIDE				Dataset: ABCD			
		AUROC	Accuracy	Sensitivity	Specificity	AUROC	Accuracy	Sensitivity	Specificity
Graph Transformer	SAN	71.3±2.1	65.3±2.9	55.4±9.2	68.3±7.5	90.1±1.2	81.0±1.3	84.9±3.5	77.5±4.1
	Graphormer	63.5±3.7	60.8±2.7	**78.7±22.3**	36.7±23.5	89.0±1.4	80.2±1.3	81.8±11.6	82.4±7.4
	VanillaTF	76.4±1.2	65.2±1.2	66.4±11.4	71.1±12.0	94.3±0.7	85.9±1.4	87.7±2.4	82.6±3.9
Fixed Network	BrainGNN	62.4±3.5	59.4±2.3	36.7±24.0	70.7±19.3	OOM	OOM	OOM	OOM
	BrainGB	69.7±3.3	63.6±1.9	63.7±8.3	60.4±10.1	91.9±0.3	83.1±0.5	84.6±4.3	81.5±3.9
	BrainNetCNN	74.9±2.4	67.8±2.7	63.8±9.7	71.0±10.2	93.5±0.3	85.7±0.8	87.9±3.4	83.0±4.4
Learnable Network	FBNETGNN	75.6±1.2	68.0±1.4	64.7±8.7	62.4±9.2	94.5±0.7	87.2±1.2	87.0±2.5	86.7±2.8
	BrainNetGNN	55.3±1.9	51.2±5.4	67.7±37.5	33.9±34.2	75.3±5.2	67.5±4.7	67.7±5.7	68.0±6.5
	DGM	52.7±3.8	60.7±12.6	53.8±41.2	51.1±40.9	76.8±19.0	68.6±8.1	40.5±29.7	**95.6±4.2**
Ours	BrainNetTF	**80.2±1.0**	**71.0±1.2**	72.5±5.2	**69.3±6.5**	**96.2±0.3**	**88.4±0.4**	**89.4±2.6**	88.4±1.5

RQ2. How does our proposed OCRead module perform with different model choices?

Table: Performance comparison AUROC (%) with different readout functions.

Readout	Dataset: ABIDE			Dataset: ABCD		
	SAN	Graphormer	VanillaTF	SAN	Graphormer	VanillaTF
MEAN	63.7±2.4	50.1±1.1	73.4±1.4	88.5±0.9	87.6±1.3	91.3±0.7
MAX	61.9±2.5	54.5±3.6	75.6±1.4	87.4±1.1	81.6±0.8	94.4±0.6
SUM	62.0±2.3	54.1±1.3	70.3±1.6	84.2±0.8	71.5±0.9	91.6±0.6
SortPooling	68.7±2.3	51.3±2.2	72.4±1.3	84.6±1.1	86.7±1.0	89.9±0.6
DiffPool	57.4±5.2	50.5±4.7	62.9±7.3	78.1±1.5	70.0±1.9	83.9±1.3
CONCAT	**71.3±2.1**	63.5±3.7	76.4±1.2	90.1±1.2	89.0±1.4	94.3±0.7
OCRead	70.6±2.4	**64.9±2.7**	**80.2±1.0**	**91.2±0.7**	**90.2±0.7**	**96.2±0.4**



(a) Influence of two key hyper-parameters for model performance.

(b) Attention heatmap on ABCD.

Figure: The hyper-parameter influence and the heatmap from self-attention.

RQ3. Does the learned model of BrainNetTF exhibit consistency with existing neuroscience knowledge and suggest reasonable explainability?
Answer: Figure(b) displays the self-attention score from the first layer of Multi-Head Self-Attention. The attention scores are the average across all subjects in the ABCD test set. This figure shows that the learned attention scores well match the divisions of functional modules based on available labels, demonstrating the effectiveness and explainability of our Transformer model. More results can be found in our paper.