DivfreqBERT : Encoding Distinct Frequency Ranges of Brain Dynamics Based on the Complexity of the Brain

Sangyoon Bae (stellasybae@snu.ac.kr)
Interdisciplinary Program in Artificial Intelligence, 1, Gwanak-ro
Gwanak-gu, Seoul, 08826 Republic of Korea

Junbeom Kwon (kjb961013@gmail.com)
Department of Psychology, 1, Gwanak-ro
Gwanak-gu, Seoul, 08826 Republic of Korea

Jiook Cha (cha.jiook@gmail.com)
Department of Psychology, 1, Gwanak-ro
Gwanak-gu, Seoul, 08826 Republic of Korea

Shinjae Yoo (sjyoo@bnl.gov)
Computational Science Initiative,
Upton, Shirley, 11967 United States of America

Abstract

The brain functions as a scale-free network, with fMRI BOLD signals demonstrating a power-law distribution. Traditional deep neural networks often overlook this aspect, leading to their underperformance due to a structure ill-suited for the complexity of brain signals. We introduce DivfreqBERT, an end-to-end model tailored for time series data, leveraging scale-free network properties for improved encoding of biological characteristics. Utilizing Lorentzian and multi-fractal functions, it segments whole-brain dynamics into three components, each consistent with the power-law function and displaying distinct small-world connectivity features. This method significantly enhances several downstream tasks, including predicting sex, age, intelligence, and depression in the Adolescent Brain Cognitive Development (ABCD) dataset, encompassing over 11,000 participants aged 9-10, and the UK Biobank (UKB) dataset, with data from over 500,000 participants aged 40-69. During pretraining, DivfreqBERT employs variations in small-worldness across frequencies to order nodes by communicability for masking, facilitating the learning of networks where highly communicable nodes play pivotal roles. Scaling up the model by pre-training on the extensive UKB dataset and fine-tuning on ABCD data markedly improved model performance. Additionally, DivfreqBERT provides interpretability by showing which connections between ROIs within each frequency range influenced the outcome and which ROIs were important. Overall, divfreqBERT demonstrates the potential of neural networks informed by complex system insights, emphasizing the benefits of integrating the brain's complexity into neural network models.

Keywords: Complex system; fMRI; BERT; Knowledge-guided deep neural network

Introduction

The rapid progress of deep learning underscores the need for models that accurately capture the brain's complex characteristics, as traditional models fail to reflect its dynamic, multidimensional nature (Kan et al., 2022; Bedel, Şıvgın, Dalmaz, Dar, & Çukur, 2022). The brain's unique scale-free and small-world properties are essential for adapting to environmental changes and supporting cognitive functions, with hubs forming numerous connections and most nodes being reachable within a few steps, promoting efficient information transfer (van den Heuvel, Stam, Boersma, & Pol, 2008; West & Shlesinger, 1990). This activity generates a 1/f-like power spectrum, indicating a power-law distribution of signal frequencies (Equation 1).

$$log(power) \propto -\beta log(frequency) \tag{1}$$

The β parameter in brain dynamics reflects time-lagged auto-correlation, where a larger β indicates long-range memory, and a smaller β signifies more efficient information processing, with variations linked to different neurological disorders (Eke, Herman, Kocsis, & Kozak, 2002; Tolkunov, Rubin, & Mujica-Parodi, 2010; Maxim et al., 2005; Radulescu, Rubin, Strey, & Mujica-Parodi, 2012). Using the Lorentzian function (equation 2) and multifractal equation (equation 3), the power spectrum is segmented into three parts with distinct β values, uncovering the multifractal nature of brain dynamics as seen in varied resting-state fMRI patterns (He, Zempel, Snyder, & Raichle, 2010; Miller, Sorensen, Ojemann, & Den Nijs, 2009).

$$Power(f) = \frac{A_{ultralow}^2 \cdot f_1^2}{f^2 + f_1^2} \tag{2}$$

$$Power(f) = \begin{cases} A_{low} \cdot f^{\beta_{low}} & \text{if } f_1 < f < f_2 \\ A_{high} \cdot f^{\beta_{high}} & \text{if } f > f_2 \end{cases} \tag{3}$$

We introduced DivfreqBERT, a model that segments the whole brain signal into parts based on a power-law distribution, outperforming other models and enhancing interpretability. It identifies key ROI connections within specific frequency ranges. During pretraining, we utilized variations in small-world characteristics across frequency ranges to prioritize node masking by communicability(Estrada & Hatano, 2008), helping the model understand the dynamics of highly influential nodes.

Results

Model Performances

Table 1: ABCD sex prediction (ROI: HCP MMP1)

Model	AUROC ↑	ACC ↑
Xgboost	0.784 ± 0.005	0.709 ± 0.003
BNT	0.891 ± 0.005	0.800 ± 0.013
BolT	0.883 ± 0.001	0.795 ± 0.001
vanilla BERT	0.883 ± 0.006	0.778 ± 0.027
DivfreqBERT	0.913 ± 0.006	0.834 ± 0.008

Table 2: ABCD fluid intelligence prediction (ROI: HCP MMP1)

Model	MAE ↓	MSE ↓
Xgboost	0.787 ± 0.002	0.987 ± 0.033
BNT	0.793 ± 0.008	1.061 ± 0.093
BolT	0.728 ± 0.004	0.845 ± 0.010
vanilla BERT	0.722 ± 0.005	0.838 ± 0.036
DivfreqBERT	0.670 ± 0.015	0.720 ± 0.024

Methods

Dividing Frequencies

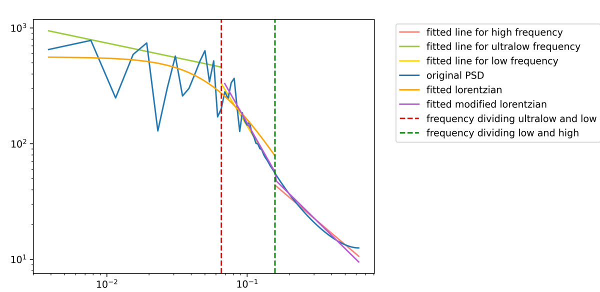

Figure 1: **Examples of the frequency dividing.**
Example of ABCD dataset. f_1 (red dotted line) and f_2 (green dotted line) are fitted by equation 2 and equation 3. They depend on individuals.

Overall Architecture

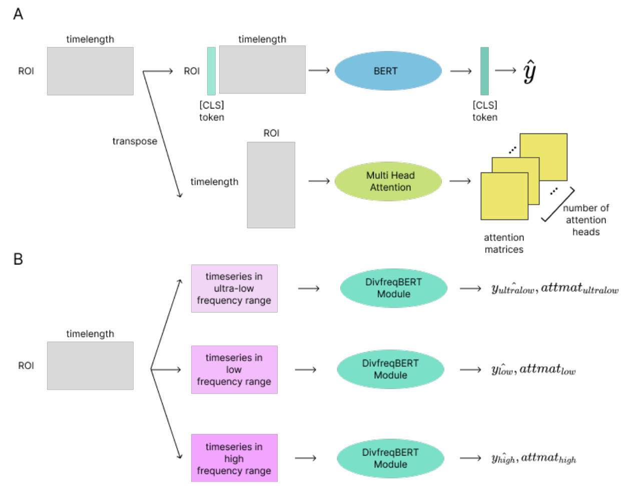

Figure 2: **Overall model architecture.**
A DivfreqBERT module operates within specific frequency ranges on divided timeseries signals. In its temporal module, BERT updates the [CLS] token, which is then processed through a classifier to produce $\hat{y}$. Spatially, signals are transposed and processed through a multi-head attention mechanism to compute attention matrices based on ROIs. Parameter sharing occurs in the temporal module, while the spatial module maintains unique parameters for each frequency, with the output including $\hat{y}$ and attention matrices for each range.

Pretraining

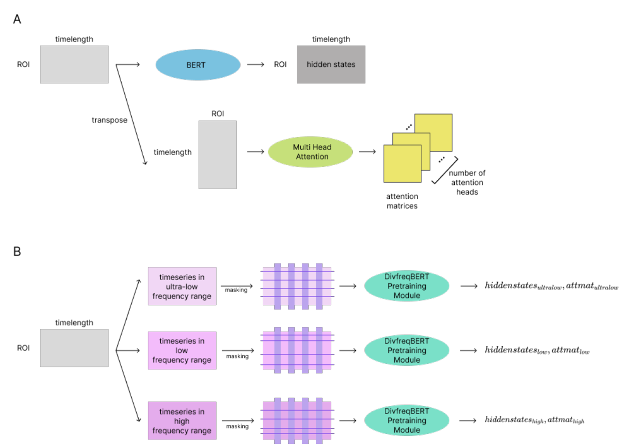

Figure 3: **Pretraining model architecture.**
The DivfreqBERT pretraining module operates on timeseries signals divided by frequency, using BERT's hidden state in the temporal module and a multi-head attention mechanism in the spatial module to calculate attention matrices based on ROIs. Parameters are shared in the temporal module but not in the spatial module. Outputs include the hidden state and attention matrices for each frequency range, with temporal and spatial masking indicated by lavender rectangles and purple lines, respectively.

References

Bedel, H. A., Şıvgın, I., Dalmaz, O., Dar, S. U. H., & Çukur, T. (2022). Bolt: Fused window transformers for fmri time series analysis. *arXiv preprint arXiv:2205.11578*.

Eke, A., Herman, P., Kocsis, L., & Kozak, L. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological measurement*, *23*(1), R1.

Estrada, E., & Hatano, N. (2008). Communicability in complex networks. *Physical Review E*, *77*(3), 036111.

He, B. J., Zempel, J. M., Snyder, A. Z., & Raichle, M. E. (2010). The temporal structures and functional significance of scale-free brain activity. *Neuron*, *66*(3), 353–369.

Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y., & Yang, C. (2022). Brain network transformer. *arXiv preprint arXiv:2210.06681*.

Maxim, V., Şendur, L., Fadili, J., Suckling, J., Gould, R., Howard, R., & Bullmore, E. (2005). Fractional gaussian noise, functional mri and alzheimer's disease. *Neuroimage*, *25*(1), 141–158.

Miller, K. J., Sorensen, L. B., Ojemann, J. G., & Den Nijs, M. (2009). Power-law scaling in the brain surface electric potential. *PLoS computational biology*, *5*(12), e1000609.

Radulescu, A. R., Rubin, D., Strey, H. H., & Mujica-Parodi, L. R. (2012). Power spectrum scale invariance identifies prefrontal dysregulation in paranoid schizophrenia. *Human brain mapping*, *33*(7), 1582–1593.

Tolkunov, D., Rubin, D., & Mujica-Parodi, L. R. (2010). Power spectrum scale invariance quantifies limbic dysregulation in trait anxious adults using fmri: adapting methods optimized for characterizing autonomic dysregulation to neural dynamic time series. *Neuroimage*, *50*(1), 72–80.

van den Heuvel, M. P., Stam, C. J., Boersma, M., & Pol, H. H. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage*, *43*(3), 528–539.

West, B. J., & Shlesinger, M. (1990). The noise in natural phenomena. *American Scientist*, *78*(1), 40–45.