Investigating the EEG Embedding by Visualization

Yongcheng Wen[1,2], Jiawei Mo[3], Wenxin Hu[1,2], and Feng Liang[1,2](✉)

[1] Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen, China
{1120200244,huwenxin,fliang}@smbu.edu.cn
[2] Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Shenzhen MSU-BIT University, Shenzhen, China
[3] School of Computer Science and Engineering, Central South University, Changsha, China
mojiawei@csu.edu.cn

Abstract. Visualizing EEG data helps clinical doctors and neuroscientists discover potential patterns and abnormalities before further mathematical analysis. Encoding complex EEG data into low-dimension embeddings and visualizing the points in 3-dimension axes with colors can help users quickly recognize some EEG properties. We apply contrastive learning in both self-supervised and supervised manners to extract the time-domain EEG features within different time window sizes. The color points tend to cluster into clouds based on their related classes and graph readers can roughly distinguish people's emotions and identities directly by inspecting the graphs. With self-supervised encoders where the generated embeddings are supposed to be used for general tasks, the visualization method can also uncover the value of the original input features extracted from raw EEG data. The source code is available at:
https://www.github.com/liangfengsid/visContrastive.

Keywords: EEG · Contrastive learning · Latent embedding · Visualization · Self-supervised

1 Introduction

Visualizing the electroencephalogram (EEG) helps users easier to discover patterns in clinical diagnosis [15] and neuroscience studies [9]. Traditional visualization elements of EEG include time-domain signals (such as the voltage) [3], frequency-domain signals (such as the power spectrum density) [4], and the source estimate [13]. However, these elements usually do not directly provide intuitively distinguishable patterns and readers may not easily extract valuable results from the visualization without further analysis. The latent approach for visualization is pervasive in computer vision [1, 5, 14]. Encoding high-dimension EEG features to low-dimension embeddings

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 221–226, 2024.
https://doi.org/10.1007/978-3-031-65126-7_20

increases information density and improves interpretability when visualized [6,7,10,12]. Therefore, we are motivated to explore the visualization effect and the interpretability of EEG embeddings encoded by different methods.

We visualize the EEG embeddings generated by contrastive-learned encoders [2,8] and investigate their ability to provide distinguishable information. The contrastive-learned encoders can be either self-supervised models or supervised ones, depending on whether it is trained in the discovery-driven manner without labels provided or in the hypothesis manner with task-related labels provided. We encode different EEG features in different training manners and visualize a few dimensions of the embeddings, where the colors of points are related to the labels or their temporal information. We find that by inspecting these figures of the EEG embeddings, people can clearly identify clouds of clustered points, where each cloud consists of points whose original EEG features are considered similar. Our study shows that compressing EEG data to low-dimension embeddings by contrastive learning and visualizing only a few dimensions can help EEG readers easily recognize the inherent patterns and relationships.

2 Method

2.1 Dataset and Feature Extraction

We use the SEED [16] dataset, which comprises EEG data from 15 persons (subjects) joining a 3-session testing, with each testing session stimulated by watching 15 movie clips of a total of about 3600 s. The movie stimuli are related to 3 emotions, i.e., positive, neutral, and negative. The EEG signals are collected by 62 electrode channels, down-sampled to 200 Hz, and filtered to bandpass frequency from 0 to 75 Hz. With data grouped by movie clips, we use 90% of the data for training both the encoder and the decoder, and the remaining 10% for testing.

We extract time-domain features from non-overlapping sliding time windows of different sizes, i.e., 0.05, 0.5, 5, and 20 s. For every time window, we extract 5 statistic voltage features for each of the 62 channels, namely the maximum, minimum, mean, median, and standard deviation. The size of each encoder input instance is 310.

2.2 Contrastive Learned Embeddings

The encoder is a non-linear convolutional neural network (CNN) that applies contrastive learning optimizing the NCE loss, which follows a similar procedure as in [11].

For the input features x and y, where y is a positive or negative contrastive sample of x, let $p(x)$ be the probability density function of x, $p(y|x)$ and $q(y|x)$ be the probability density function of the positive and negative samples conditioned on x, respectively. Encoding x and y can be represented by a function f with normalized outputs, and $f(x)$ and $f(y)$ are the normalized latent embeddings, respectively. We use the dot product of $f(x)$ and $f(y)$ adjusted with a temperature parameter τ as the similarity function between these two latent embeddings, which is denoted as $\psi(x,y) = f(x)^T f(y)/\tau$. The objective is to minimize the NCE loss, which is:

$$\mathop{\mathbb{E}}_{\substack{x\sim p(x),y_+\sim p(y|x)\\ y_1,y_2,\dots,y_n\sim q(y|x)}}[-\psi(x,y_+) + log\sum_{i=1}^{n} e^{\psi(x,y_i)}].$$

Positive and negative samples are taken from a minibatch of the training input. In the discovery-driven manner when no label is provided (self-supervised), samples near x along the timeline are positive and those far away from x along the timeline are negative. In the hypothesis manner, specific labels are provided (supervised), samples with the same label as that of x are positive, while those with different labels from x are negative. We train different encoder models without any label, with emotion labels, and with subject labels, respectively.

The encoder is a five-layer 1D convolutional network with skipping connections with each perceptron activated a GELU function. The mini-batch size of the input is 1,024, the learning rate is 0.001. The encoder output dimension is 32, and the number of mini-batch training iterations is 320,000.

2.3 Visualization

If the low-dimension EEG embeddings are invariant and discriminative, only drawing a few dimensions would be enough for revealing intuitive information about their classes. We plot the first three dimensions of the EEG embeddings of the testing set as points along the axes in the figure, with colors to indicate their related temporality or labels.

3 Results and Discussions

Figure 1 shows the results of plotting the first three dimensions of EEG embeddings by self-supervised and supervised contrastive learning methods with time-domain features extracted from different time window sizes. The clusters of color points reveal abundant information on the classes of the EEG samples.

The EEG embedding visualization can be applied to emotion recognition. In the emotion-label-guided case, we cannot find obvious patterns from the point cloud when the time window size is too small, e.g., at 0.05 or 0.5 s. But as the time window size increases, e.g., to 5 s or 20 s, points tend to cluster into 3 clouds, probably corresponding to 3 types of emotions.

Visualizing EEG embeddings can also be used for identity recognition. In the subject-label-guided case, even when the time window size is 0.05 s, the points are roughly clustered into 15 groups, probably corresponding to the 15 subjects, respectively. As the time window size increases, the contours of the point clouds become more apparent, and people can identify the cluster that a point belongs to with high confidence.

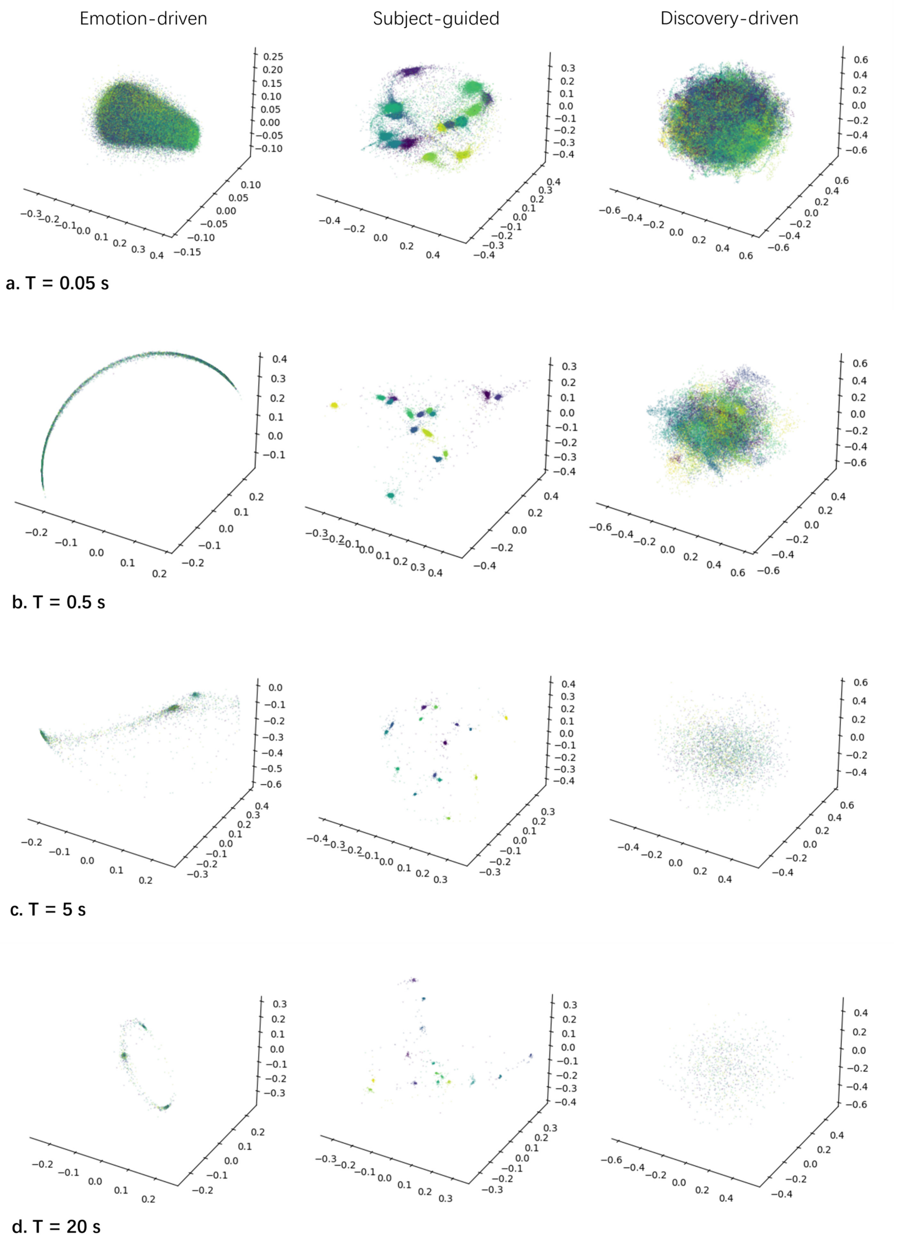

Fig. 1. The first three dimensions of EEG embeddings by discovery-driven, emotion-label-guided, and subject-label-guided contrastive learning, respectively, with features extracted from different time window sizes.

Visualizing the self-supervised learned EEG embeddings can help judge the potential value of the input feature and the effect of the encoder. In the discovery-driven case, the point cloud is in chaos when the time window size is 0.05 s, but some points of the same color start to gather and form blur contours. When the time window size is 0.5 s, we can already see some overlapping color clouds. This indicates that the corresponding input features contain some valuable information and the encoder properly transforms them into identifiable embeddings. Otherwise, if the color cloud is always in chaos, either we need to try other input features, or we need to investigate the effectiveness of the encoder algorithm.

4 Discussion and Future Work

In this paper, we use time-domain EEG features as an example to show that encoding EEG into latent embeddings and visualizing them can greatly improve the interpretability and understandability of EEG and help EEG readers easily discover some patterns. But other traditional EEG features for various tasks can also apply the latent embedding visualization so that clinical doctors and neuroscientists can make a preliminary decision from the EEG results before they go on further analysis.

In the future, we will explore self-supervised and supervised encoding methods with more traditional EEG features and see how visualizing the low-dimension embeddings can help to reveal identifiable patterns. We will also develop other visualization techniques for EEG that make EEG more understandable to clinical doctors and patients, neuroscientists, and biomedical engineers.

Acknowledgment. The work was supported in part by the National Natural Science Foundation of China (under grant 12102267) and the Shenzhen Sustainable Development Special Project (under grant KCXFZ20201221173411032).

References

1. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. Adv. Neural. Inf. Process. Syst. **33**, 9912–9924 (2020)
2. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
3. Daud, S.N.S.S., Sudirman, R.: Pattern of EEG voltage and oscillations under stimulation of Mozart's music and white noise for visual learning process. Biomed. Signal Process. Control **85**, 104986 (2023)
4. Donoghue, T., et al.: Parameterizing neural power spectra into periodic and aperiodic components. Nat. Neurosci. **23**(12), 1655–1665 (2020)
5. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. Adv. Neural Inf. Processing Syst. **27** (2014)

6. Duncker, L., Bohner, G., Boussard, J., Sahani, M.: Learning interpretable continuous-time models of latent stochastic dynamical systems. In: International Conference on Machine Learning, pp. 1726–1734. PMLR (2019)
7. Duncker, L., Sahani, M.: Temporal alignment and latent gaussian process factor inference in population spike trains. Adv. Neural Inf. Process. Syst. **31** (2018)
8. Hyvarinen, A., Sasaki, H., Turner, R.: Nonlinear ICA using auxiliary variables and generalized contrastive learning. In: The 22nd International Conference on Artificial Intelligence and Statistics, pp. 859–868. PMLR (2019)
9. Niso, G., et al.: Open and reproducible neuroimaging: from study inception to publication. In: NeuroImage 119623 (2022)
10. Schirrmeister, R.T., et al.: Deep learning with convolutional neural networks for EEG decoding and visualization. Hum. Brain Mapp. **38**(11), 5391–5420 (2017)
11. Schneider, S., Lee, J.H., Mathis, M.W.: Learnable latent embeddings for joint behavioural and neural analysis. Nature **617**, 360–368 (2023)
12. Song, Y., Zheng, Q., Liu, B., Gao, X.: EEG conformer: convolutional transformer for EEG decoding and visualization. IEEE Trans. Neural Syst. Rehabil. Eng. **31**, 710–719 (2022)
13. Sun, R., Sohrabpour, A., Worrell, G.A., He, B.: Deep neural networks constrained by neural mass models improve electrophysiological source imaging of spatiotemporal brain dynamics. Proc. Natl. Acad. Sci. **119**(31), e2201128119 (2022)
14. Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3733–3742 (2018)
15. Wynn, J.K., Roach, B.J., McCleery, A., Marder, S.R., Mathalon, D.H., Green, M.F.: Evaluating visual neuroplasticity with EEG in schizophrenia outpatients. Schizophr. Res. **212**, 40–46 (2019)
16. Zheng, W.L., Lu, B.L.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans. Auton. Ment. Dev. **7**(3), 162–175 (2015)