Reproducing a State-of-the-Art paper on Machine Learning applied to Brain Computer Interfaces

Eduard Ramón Aliaga Torrens
aliagaeduard@gmail.com

Alberto Cabellos Aparicio
alberto.cabellos@upc.edu

Abstract

This research addresses the challenge of deciphering user intent from noninvasive EEG signals, a pivotal component of BCIs aimed at assisting individuals with severe speech and physical impairments. The crux of this study lies in enhancing the accuracy of posterior symbol probability estimation during a typing task. Using a novel discriminative model approach, the probability estimation for symbol selection is iteratively refined. The model dynamically incorporates ERP feedback to update these probabilities, continuing until a symbol's confidence surpasses a set threshold. A neural network classifier serves as the backbone of the recursive Bayesian update mechanism, enabling a more precise and efficient interpretation of EEG data. The experimental results, derived from a simulated typing scenario, demonstrate a marked improvement over traditional generative models, highlighting the efficacy of the discriminative methodology presented in this article in BCI applications.

1. Introduction

Electroencephalography (EEG) is a non-invasive method used to record electrical activity of the brain. It has become an integral part of brain-computer interfaces (BCIs), systems that enable direct communication between the brain and an external device, often bypassing traditional pathways like muscle movements [1]. In recent years, significant research has focused on utilizing EEG to infer user intent, pivotal for individuals with severe speech and physical impairments [2]. This line of inquiry is not just academically stimulating but also holds immense potential for real-world applications, particularly in enhancing the quality of life for those with communication and mobility challenges.

Brain-computer interfaces, leveraging EEG, have opened new frontiers in understanding and interpreting human intent. The field of BCI research has expanded rapidly, integrating concepts from neuroscience, engineering, and computer science [4]. One of the key aspects of this research is identifying and decoding user intent from EEG signals, which involves complex signal processing and machine learning techniques. This task is challenging due to the inherent noise in EEG data and the subtlety of signal variations corresponding to different intents.

The research paper in focus here has made significant strides in this field. Titled "Recursive Estimation of User Intent from Noninvasive Electroencephalography Using Discriminative Models" [3], authored by Niklas Smedemark-Margulies et al., it delves into improving the estimation of posterior symbol probabilities in a typing task using EEG. The paper presents a novel approach that employs discriminative models instead of traditional generative models for this purpose. By using a neural network classifier, the study demonstrates improved performance in a simulated typing task, outperforming previous approaches based on generative modeling.

The objective of this paper is to build upon the work of Smedemark-Margulies et al., replicating their results and exploring potential improvements. The approach aims to refine the model's accuracy and efficiency, making it more applicable and beneficial for real-world BCI applications . By incorporating advancements in machine learning and signal processing, the intention is to propose a model that not only aligns with the existing framework but also introduces innovative methodologies to enhance its effectiveness in interpreting user intent from EEG data.

2. Problem Statement & dataset

2.1. Problem Statement

The focus of this research is to replicate and enhance the model developed in "Recursive Estimation of User Intent from Noninvasive Electroencephalography Using Discriminative Models"[3]. This model represents a novel approach in BCI, specifically in the context of EEG-based typing interfaces. Unlike traditional generative models, this model employs discriminative techniques to interpret EEG signals, thereby aiming to improve the accuracy and efficiency in estimating user intent. The challenge lies in accurately classifying EEG signals to predict a user's intended symbol from a set of possibilities, a task which is both complex and critical for enhancing communication aids for individuals with motor impairments. The study explores the model's methodology, seeking to identify and implement potential improvements in its performance and applicability.

2.2. Dataset

The dataset used in the original study consists of EEG recordings from an RSVP typing task [5]. This dataset is characterized by its substantial volume and the richness of its EEG data, which captures the brain's responses to visual stimuli representing different symbols. It includes instances where subjects were asked to focus on specific symbols from a rapidly presented sequence. The EEG data were recorded under controlled conditions, ensuring a high degree of reliability and relevance for the task at hand. This research will use the same dataset to maintain consistency with the original study, allowing for a direct comparison of results. This dataset provides a robust foundation for testing the improved model, ensuring that any observed enhancements in performance are attributable to the model itself and not variations in data quality or type.

3. Reproducing the experimental results

In the process of reproducing the experimental results from the original paper, some issues were encountered due to limited resources that needed methodical problem-solving and adaptations to the approach. This section outlines the primary issues faced and the solutions that were implemented to address them. Followed by the confirmation of the conclusion exposed by the reference paper.

3.1. Multithreading Complications

The code provided by the authors made extensive use of multithreading. While this is typically advantageous for parallel processing and speed, it introduced critical issues for the setup as the resources used in this experiments were limited. The most pressing issue was the instability caused by multithreading, which led to processes being prematurely terminated and bootstrapping failures, making impossible to train the models and obtain the results.

To solve this complications, environment variables were manually set within the code to strategically override the default multithreading. This way the execution was constrained to single-threaded processes. This change significantly stabilized the training process, allowing models to train to completion without the risk of unexpected terminations.

3.2. Large-Scale Data

Another significant challenge was the sheer size of the datasets and models described in the original study. Given the limitations of the available hardware, it was impractical to train the expansive models on the entirety of the dataset without encountering out-of-memory errors and prohibitive training durations.

The solution was to implement a staged training strategy. The training of the models was made on a smaller portion of the dataset. This allowed for rapid prototyping and model refinement without overtaxing the system's resources. Upon identifying the most promising models through this preliminary phase, they allocated the full computational resources to extensively train these select models on the complete dataset.

3.3. Reproduced conclusions

In table 1 it can be seen the results of a comparative study using 50000 balanced samples from . Discriminative models demonstrated superior performance over generative models, aligning with the findings in the provided paper. The 2D CNN model showed the highest balanced accuracy, underscoring the efficacy of discriminative approaches in feature extraction. Conversely, generative models like LDA exhibited lower accuracy, highlighting the advantage of discriminative models in classification tasks.

The information transfer rate varied notably, with the 1D CNN model achieving high ITR values, suggesting that models with a larger number of parameters might be more effective in certain scenarios. However, the limited sample size and balanced nature of the dataset necessitate further investigation into larger and more diverse datasets to confirm the generalizability of the results. Control models, consistently predicting a single class, served as a baseline, confirming the enhanced predictive capability of the discriminative models in this study setting.

Strategy	Model	Parameters	Balanced Acc	ITR
Disc	LogR	3907	0.850	1.934 ± 0.000
Disc	EEGNET	12274	0.808	0.598 ± 0.000
Disc	1D CNN	542210	0.862	3.863 ± 0.000
Disc	2D CNN	1603042	0.867	1.266 ± 0.668
Gen	LDA (Emp Prior)	819401	0.789	3.181 ± 0.682
Gen	LDA (Unif Prior)	819401	0.789	2.500 ± 0.000
Gen	LogR (Emp Prior)	819401	0.792	3.971 ± 0.837
Gen	LogR (Unif Prior)	819401	0.790	3.863 ± 0.000
Control	Always Class 0	0	0.500	0.000 ± 0.0000
Control	Always Class 1	0	0.500	0.000 ± 0.0000

Table 1: Balanced Accuracy, number of parameters and Information Transfer Rate (ITR) for Different Models with 50000 non balanced samples

4. Proposed Method

In this work, a novel methodology for EEG data analysis that synergistically combines spatial and temporal processing stages is proposed. Traditional methods typically employ either 1D or 2D convolutional neural networks (CNNs) that focus on either temporal or spatial features, respectively. Inspired by the interconnected nature of spatial and temporal information in brain activity, the idea is to integrate these two streams into a unified analysis framework.

The inception of the method came from the recognition that cognitive tasks and brain responses are not just isolated to distinct regions of the brain but also evolve over time. Therefore, a comprehensive approach that simultaneously considers where and when brain activity occurs could offer a more complete understanding of EEG data.

To capture the spatial aspect of brain signals, first the strength and location of signals across the scalp is mapped, identifying regions of heightened activity. This step is essential, as it lays the groundwork for understanding the spatial distribution of neuronal activation, which varies with different cognitive functions.

Building upon the spatial foundation, the analysis then shifts focus to the temporal dimension. Here, the progression of brain signals is examined, revealing the dynamic nature of neural responses. By scrutinizing how these signals change and interact as time progresses, the method uncovers temporal patterns that signify the brain's rhythm and response to stimuli.

The integration of spatial and temporal analyses forms the cornerstone of this approach, aiming to encapsulate the complex interplay between the two dimensions. This integrated model is poised to enhance the interpretability of EEG data, paving the way for advancements in brain-computer interfaces and neuroscientific research.

4.1. Combined CNN

To develop the proposed mehthod, the Combined Convolutional Neural Network (CombinedCNN) is presented, a specialized architecture for EEG signal analysis. This model consists of a sequence of 2D and 1D convolutional layers followed by a multi-head attention mechanism, capturing the complex spatial-temporal relationships inherent in EEG data, as depicted in Figure 1.

The initial phase of the CombinedCNN involves spatial feature extraction using 2D convolutional layers. These layers process the EEG signals across the scalp's sensors, identifying signal intensity and location through convolution operations that map brain activity's topographical information. Batch normalization is applied to stabilize the learning process, while ReLU activations introduce non-linearity, vital for modeling complex brain activity patterns. Regularization via dropout of rate 0.5 prevents overfitting, and max pooling layers reduce dimensionality, summarizing the spatial features efficiently.

Transitioning from spatial to temporal analysis, the CombinedCNN employs 1D convolutional layers tailored to capture temporal dependencies. By convolving filters over the EEG signal's time dimension, these layers track the dynamic changes in brain activity over time. The temporal phase also incorporates batch normalization, ReLU activations, and dropout of rate 0.5 for feature refinement, with adaptive average pooling consolidating the temporal information.

At the heart of the CombinedCNN lies the multi-head attention mechanism, which enhances the model's focus on salient features by processing the inputs into queries, keys, and values. The scaled dot-product attention within this mechanism computes alignment scores to weight the values' significance, yielding an integrated representation that combines information from multiple representational subspaces. This capability allows the model to concurrently attend to distinct aspects of the EEG signals.

The feature representations, enriched through attention, proceed to a linear layer designed to match the final output size required for classification. The subsequent fully connected layer further processes these features, leading to the log-softmax layer that outputs the class probability distribution. This end-to-end model architecture, from input to classification, encapsulates a robust approach for interpreting EEG signals. The model effectively delineates the intricate interplay between the spatial and temporal dimensions of brain activity, which is essential for accurate EEG analysis.

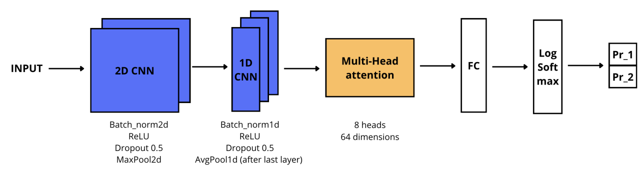

Figure 1: Schematic representation of the Combined Convolutional Neural Network (CombinedCNN) architecture, illustrating the sequential integration of spatial and temporal convolutional layers, followed by the multi-head attention mechanism and classification layers, to process EEG data for intent recognition.

4.2. Model evaluation

In this section the experimental results are described. This can be reproduced using the code found in this repository [1]

The evaluation of each model's performance is meticulously conducted both in the context of individual EEG trials and within a simulated typing task. It is important to note that the RSVP typing task inherently deals with imbalanced classes, as the user is typically focused on selecting one specific symbol from a large alphabet at any given time. To effectively assess performance on unseen single EEG trials, we compute the balanced accuracy, which is the mean of accuracies obtained for each class.

[1] https://github.com/EduardAliaga/Combined-cnn-model-for-EEG-signal-classification

Two iterations of the CombinedCNN are rigorously trained. The first, termed the 'tiny' version, incorporates a singular 2D CNN layer followed by two 1D CNN layers. The more extensive version of the CombinedCNN features two layers of 2D CNNs and three layers of 1D CNNs. They both apply multi-head attention with 8 heads and a model dimension of 64.

To maintain consistency across model evaluations and facilitate direct comparisons, both versions of the CombinedCNN, as well as other architectures such as EEGNet and various simple CNN variants, are trained under uniform parameters. These include an input shape of (62, 63), indicative of the dimensions of the EEG data, and a binary classification objective. The training process is conducted over 25 epochs, with a learning rate set at 0.001. For optimization, the AdamW optimizer is used. Additionally, the training incorporated the prior probability of the target symbol in the query, an essential factor in BCI applications, ensuring a thorough and pertinent approach to training the model.

Strategy	Model	Parameters	Balanced Acc	ITR
Disc	LogR	3907	0.730	0.817 ± 0.047
Disc	EEGNET	12274	0.745	0.930 ± 0.050
Disc	1D CNN	542210	0.782	1.103 ± 0.047
Disc	2D CNN	1603042	0.779	1.153 ± 0.068
Disc	**Combined CNN tiny**	**27890**	**0.784**	$\mathbf{1.137 \pm 0.000}$
Disc	**Combined CNN**	**168370**	**0.820**	$\mathbf{1.299 \pm 0.668}$
Gen	LDA (Emp Prior)	819401	0.509	0.678 ± 0.077
Gen	LDA (Unif Prior)	819401	0.687	0.678 ± 0.077
Gen	LogR (Emp Prior)	819401	0.500	0.218 ± 0.022
Gen	LogR (Unif Prior)	819401	0.694	0.218 ± 0.022
Control	Always Class 0	0	0.500	0.000 ± 0.0000
Control	Always Class 1	0	0.500	0.000 ± 0.0000

Table 2: Balanced Accuracy, number of parameters and Information Transfer Rate (ITR) for Different Models with full dataset compared to the results of the original [paper]

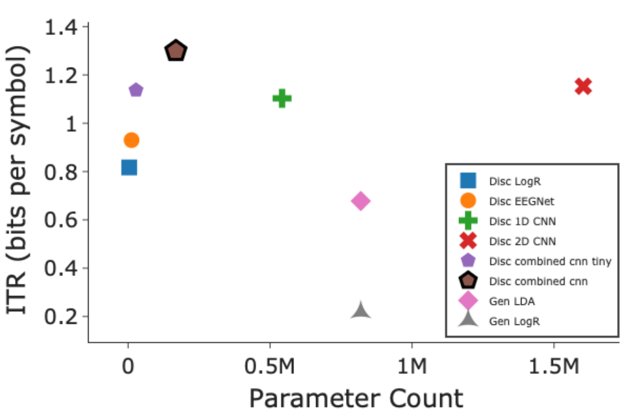

Figure 2: Information Transfer Rate vs Model Size. Even with a tiny model of the combined CNN a better balanced accuracy is obtained.

5. Results

From the data in Table 2 and Figure **??**, it is evident that the Combined CNN model, both in its standard and tiny versions, demonstrates superior performance compared to traditional discriminative models like Logistic Regression (LogR), EEGNet, and both 1D and 2D CNNs. Notably, the Combined CNN model achieves a balanced accuracy of 0.820 ± 0.000 and an Information Transfer Rate (ITR) of 1.299 ± 0.668, significantly outperforming the other discriminative models in terms of efficiency and accuracy. These results are particularly remarkable considering the Combined CNN model uses only 168,370 parameters, a considerably smaller number compared to the 1,603,042 parameters of the 2D CNN model.

Furthermore, the Combined CNN tiny version, designed for even greater efficiency with only 27,890 parameters, still achieves a commendable balanced accuracy of 0.784 ± 0.000 and an ITR of 1.137 ± 0.000. This highlights the model's exceptional ability to maintain high performance despite a significant reduction in complexity, as depicted in Figure 2. The graph illustrates that even with a scaled-down version the model achieves better-balanced accuracy than other larger models.

In contrast, generative models like Linear Discriminant Analysis (LDA) with both empirical and uniform priors, and the LogR models, exhibit lower performance metrics, both in terms of balanced accuracy and ITR. This reinforces the advantage of discriminative models in classification tasks, particularly in scenarios where maximizing ITR is crucial.

These findings challenge the conventional belief that larger, more complex models necessarily yield better performance. The Combined CNN models, through their innovative architecture that synergizes spatial and temporal feature extraction, have demonstrated that efficiency and high performance can coexist in a well-designed model. This not only sets a new benchmark in BCI research but also opens new possibilities for developing more efficient and effective models for EEG data analysis and other applications in neurotechnology.

6. Conclusions

This study enhances the interpretation of user intent from noninvasive EEG signals using a Combined Convolutional Neural Network (CombinedCNN). The research centered on refining the accuracy and efficiency of posterior symbol probability estimations within a typing task context, a crucial aspect of brain-computer interface (BCI) applications.

The implementation of the CombinedCNN, both in its standard and tiny configurations, demonstrated the potential of a discriminative model approach over traditional generative models. The models, trained uniformly across various architectures, showed that a balanced integration of spatial and temporal feature extraction with an attention mechanism can lead to significant improvements in EEG signal classification. This was evident in the performance enhancement observed in a simulated typing scenario, where the models not only replicated but, in some cases, surpassed the baseline results established by previous research.

A key finding of our study is the effectiveness of the tiny version of CombinedCNN. Despite its reduced complexity, it maintained a competitive edge in terms of accuracy and efficiency. This underscores the potential of lightweight models in BCI applications, particularly where computational resources are limited.

Moreover, while the parameters used demonstrated an effective training process, there remains potential for further optimization. Fine-tuning them such as increasing the number of layers, extending the number of epochs, or adjusting the learning rate could potentially enhance model performance. Exploring different configurations and training conditions could lead to even more robust and accurate EEG data interpretation, considering the complex nature of EEG signals with inherent noise and variability.

Additionally, applying a transformer-based architecture to EEG signal analysis presents an intriguing avenue for future research. Transformers, known for their success in natural language processing, could offer significant advantages in capturing the intricate temporal dynamics of EEG data. Their ability to model long-range dependencies and focus on relevant signal segments through self-attention mechanisms may prove particularly beneficial in deciphering complex brainwave patterns. This approach could potentially unlock new levels of accuracy and efficiency in EEG-based BCIs, offering a promising direction for advancing the field.

In conclusion, this research contributes to the field of BCIs by presenting an innovative approach to EEG signal interpretation through the CombinedCNN model. The promising results obtained pave the way for further explorations and potential real-world applications, especially in aiding individuals with severe speech and physical impairments. Future work may focus on exploring more advanced neural network architectures and training methodologies, as well as expanding the application scope to other BCI paradigms.

This study, thus, not only reinforces the importance of discriminative models in EEG-based BCIs but also opens avenues for future research in this dynamic and impactful field.

References

[1] Reza Abiri, Saba Borhani, Eric W Sellers, Yang Jiang, and Xin Zhao. A comprehensive review of eeg-based brain–computer interface paradigms. *Journal of neural engineering*, 16(1):011001, 2019.

[2] Joseph N Mak, Yael Arbel, Jason W Minett, Lynn M McCane, Bettschart Yuksel, Derek Ryan, David Thompson, Luigi Bianchi, and Deniz Erdogmus. Optimizing the p300-based brain–computer interface: current status, limitations and future directions. *Journal of neural engineering*, 8(2):025003, 2011.

[3] Niklas Smedemark-Margulies, Basak Celik, Tales Imbiriba, Aziz Kocanaogullari, and Deniz Erdoğmuş. Recursive estimation of user intent from noninvasive electroencephalography using discriminative models. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[4] Lina Tong, Yihui Qian, Liang Peng, Chen Wang, and Zeng-Guang Hou. A learnable eeg channel selection method for mi-bci using efficient channel attention. *Frontiers in Neuroscience*, 17:1276067, 2023.

[5] Shangen Zhang, Yijun Wang, Lijian Zhang, and Xiaorong Gao. A benchmark dataset for rsvp-based brain–computer interfaces. *Frontiers in neuroscience*, 14:568000, 2020.