Dynamic Stopping P300 Speller with Variational Bayesian Neural Network

Abstract

Dynamic stopping P300 spellers have previously compared classifier outputs to a threshold to choose whether or not to stop at an earlier stimulation for each trial. For example, Bianchi et al. proposed using a SVM classifier's hyperplane and grouped the distances from the hyperplane into different regions [Bianchi et al., 2019]. Each region is assigned a confidence value, which can be used to classify the output [Bianchi et al., 2019]. However, there are several limitations. Firstly, there is no guarantee that the confidence values output by the classifier match the actual values. This can lead to an under or overly confident model; both cases are problematic. An under confident model uses more stimulations than the optimal model, and thus has a lower ITR and fails to take advantage of stopping early. Likewise, an overconfident model is less accurate as it uses too few stimulations, which can make the speller unusable due to the number of incorrect classifications. Secondly, there is no intuitive sense behind the confidence values found, as they only represent the values that achieve the optimal accuracy for the speller. It is difficult to both interpret and understand the results. Therefore, grid search must be used to calibrate the model on every subject to find the optimal confidence values. This results in an overall slower model. In an attempt to address these limitations, we propose a novel dynamic stopping method that accounts for uncertainty in the model output and uses an intuitive threshold value while avoiding computationally expensive hyperparameter tuning.

Dataset

The data used in this study was based on a publicly available online dataset [Aricò et al., 2014]. Data was collected using a 6 x 6 row-column (RC) speller. An elastic cap was used with 16 electrodes for all 10 subjects. For each session, each subject completed 3 runs, with 6 trials per run and 8 stimulations per trial. A prediction of the target character was made after each trial. Each intensification lasts for 125 ms with an inter stimulus interval of 125 ms, resulting in a total 250 ms lag between two stimuli. Finally, the data was bandpassed between 0.1 Hz and 20 Hz.

Methods

Non-uniform Dynamic Stopping This study introduces a non-uniform dynamic stopping algorithm to decrease the overall time to make a character prediction. The intent is to use a subset of the total stimulations for each trial, with the model selecting a row and column once the threshold for both are reached. At the start of each trial, the model outputs the probabilities of each row containing the target character for one stimulation and compares the highest probability with the threshold. With the use of a Bayesian model, the probabilities are a closer representation of the true accuracies and therefore can be interpreted as the true probability that the current row or column contains the target character. If the threshold is reached, the row corresponding to the highest probability is selected and the process is repeated to select a column. If the threshold is not reached, then an additional stimulation is used to output the probabilities for each row until the threshold is reached or all row stimulations are used. Then the predicted target character is the intersection of the row and column. Note that the rows and columns may stop at different stimulations and hence is why the algorithm is non-uniform. By optimally choosing the row and column separately, the total number of stimulations required to make a character selection is minimized. For this study, a threshold of 0.9 was used.

Hybrid Bayesian Neural Network Deep neural networks will often under or overfit as it does not account for uncertainty in its predictions. With Bayesian Neural Network (BNN), however, the weights are trained as distributions instead of point-estimates, thus capturing the uncertainty of the model's predictions. The issue with BNN is that the architecture of deep neural networks makes it redundant and costly to account for uncertainty for a large number of successive layers. Therefore, we construct a hybrid BNN architecture with deterministic layers from the EEGNet architecture proposed by [R. T. Schirrmeister, 2017] with multiple layers designed as temporal and spatial filters. This has significantly fewer parameters than other Convolutional Neural Networks (CNN). The EEGNet layers are followed by a probabilistic output layer to capture uncertainty.

Variational Inference The goal is to compute the posterior

$$p(\boldsymbol{\theta}|D) = \frac{p(D_{\boldsymbol{y}}|D_{\boldsymbol{x}},\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(D_{\boldsymbol{y}}|D_{\boldsymbol{x}},\boldsymbol{\theta}')p(\boldsymbol{\theta}')d\boldsymbol{\theta}'}$$

in which D is the training set, Dx are the training features, and Dy are the training labels. The numerator is the product of likelihood and the prior. The denominator is the evidence integral marginalized over all θ. By Bayesian inference, we can predict an output y given an unseen data x.

$$p(y|x,D) = p(y|x,\theta)p(\theta|D)d\theta$$

The integral is intractable as the evidence integral in p(θ|D) is intractable. Therefore, we use a technique called Variational Inference (VI) to obtain the predictive distribution [Blundell, et al, 2015]. We follow the standard procedure [Jospin, 2020] to use an isotropic Gaussian prior for Bayesian Neural Network, which is favored for its mathematical properties and the simple formulation of its log.

Approximate Predictive Distribution As the integral in the predictive distribution is intractable, we use a Monte Carlo approximation. We found 5 samples to achieve good results, so it was used throughout our results.

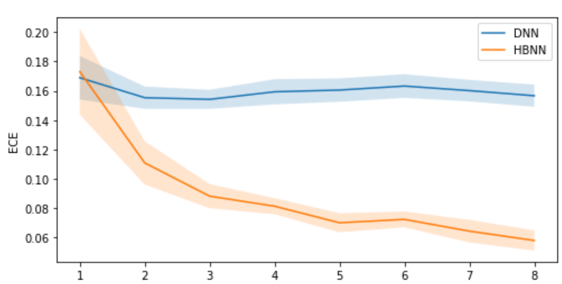
Results

A three-fold cross-validation was used where the training data consisted of two runs while the testing data consisted of 1 run.

Expected Calibration Error (ECE) and Accuracy of DNN v.s HBNN

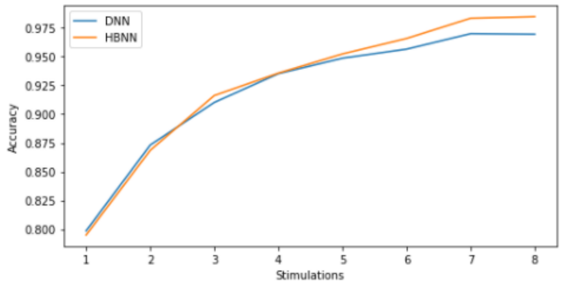
$$ECE = \sum_{b=1}^{B} \frac{n_b}{N} |acc(b) - conf(b)|$$

ECE tells us how well the model is calibrated. n_b is the number of predictions in bin b, N is the total number of data points, and acc(b) and conf(b) are the accuracy and confidence of bin b, respectively. We use 10 bins each with an interval of 0.1.



	DNN	HBNN
Stimulation 1	0.169	**0.173**
Stimulation 2	**0.155**	0.111
Stimulation 3	**0.154**	0.088
Stimulation 4	**0.159**	0.081
Stimulation 5	**0.160**	0.070
Stimulation 6	**0.163**	0.072
Stimulation 7	**0.160**	0.064
Stimulation 8	**0.157**	0.058
Mean	**0.160**	0.090

The figure compares Deep Neural Network (DNN) and HBNN's ECE of classifying the presence of P300 evoked potentials across all subjects at different stimulations. The graph shows HBNN is better calibrated than DNN at all stimulations except the first.

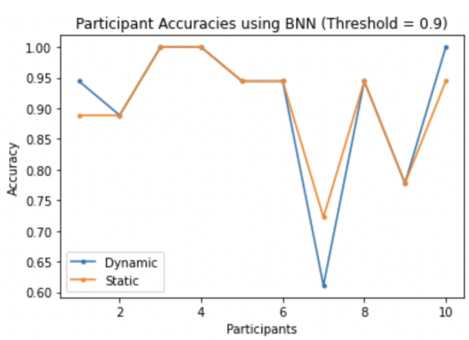


	DNN	HBNN
Stimulation 1	**0.799**	0.795
Stimulation 2	**0.873**	0.869
Stimulation 3	0.910	**0.916**
Stimulation 4	0.935	**0.936**
Stimulation 5	0.949	**0.952**
Stimulation 6	0.956	**0.966**
Stimulation 7	0.970	**0.983**
Stimulation 8	0.969	**0.985**
Mean	0.920	**0.925**

The figure compares DNN and HBNN's accuracy in classifying the presence of P300 evoked potentials across all subjects at different stimulations.

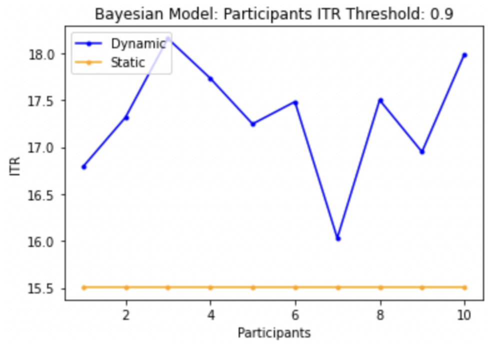
<u>Accuracy, Time, ITR of Dynamic v.s. Static Stopping Method</u>

The dynamic and static methods were compared using the probabilistic model. For the static method, only 6 out of the 8 stimulations were used as it was the minimum needed for the accuracy to reach 90%. However, for the dynamic method, the algorithm was allowed to reach 8 stimulations before stopping and a threshold of 0.9 was used. Overall, results show that the dynamic method achieved greater ITR and faster time without sacrificing accuracy. Accuracies between the dynamic and static methods with the probabilistic model were very close, with both of the methods having reached close to 90% mean accuracy with a threshold of 0.9 for the dynamic method. ITR, however, was superior for the dynamic method as seen in Figure 2. Furthermore, Figure 3 shows that the dynamic method takes on average 2 seconds to make an overall character prediction than the static method.



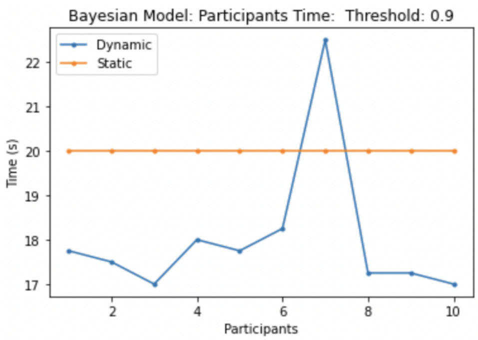
	Static Accuracy	Dynamic Accuracy
Subject 1	**0.944**	**0.944**
Subject 2	**0.944**	0.889
Subject 3	**1.000**	**1.000**
Subject 4	**1.000**	**1.000**
Subject 5	**1.000**	0.944
Subject 6	**0.944**	**0.944**
Subject 7	**0.667**	0.611
Subject 8	**0.944**	**0.944**
Subject 9	**0.889**	0.778
Subject 10	0.944	**1.000**
Mean	**0.928**	0.906

Figure 1: Mean accuracy across patients



	Static ITR	Dynamic ITR
Subject 1	15.509	**16.793**
Subject 2	15.509	**17.318**
Subject 3	15.509	**18.158**
Subject 4	15.509	**17.734**
Subject 5	15.509	**17.248**
Subject 6	15.509	**17.484**
Subject 7	15.509	**16.031**
Subject 8	15.509	**17.502**
Subject 9	15.509	**16.950**
Subject 10	15.509	**17.989**
Mean	15.509	**17.321**

Figure 2: ITR across patients



	Static Time	Dynamic Time
Subject 1	20.000	**17.750**
Subject 2	20.000	**17.500**
Subject 3	20.000	**17.000**
Subject 4	20.000	**18.000**
Subject 5	20.000	**17.750**
Subject 6	20.000	**18.250**
Subject 7	**20.000**	22.500
Subject 8	20.000	**17.250**
Subject 9	20.000	**17.250**
Subject 10	20.000	**17.000**
Mean	20.000	**18.025**

Figure 3: Time (in seconds) to predict the target character for each patient

Discussion

The proposed Bayesian Neural Networks model aims at providing a reliable metric for our non-uniform dynamical stopping algorithm in order to improve the ITR without compromising the prediction accuracy. The accuracy achieved with this model outperforms other deep learning-based P300 RC Spellers such as the Novel P300 PCA-CNN model by [Li et. al, 2020] by 15.5% after 6 stimulations. The dynamic stopping capability of this model also achieved improvements in ITR and is comparable to some of the latest P300 dynamic stopping methods such as Ensemble Dynamic Stopping by [Vo et.al, 2017]. The ensemble method achieved an 88% mean increase in real ITR at 83% mean accuracy compared to our model which achieved a 12% mean increase in real ITR at 90.5% mean accuracy. The improved prediction performance and adaptability makes this model better suited for helping locked-in patients communicate with P300 RC Spellers. An improved ITR means the patient would be able to input sentences faster and more accurately. The dynamical stopping capability allows this model to better adapt to real-world situations with varying headset qualities, individual performances, and other environmental variabilities.