DEEP ACTIVE LEARNING IN THE BRAIN-COMPUTER INTERFACE SYSTEM FOR MOTION SICKNESS CLASSIFICATION

So-Hyun Han

Department of Artificial Intelligence, Korea University

ABSTRACT

The brain-computer interface (BCI) field is making a lot of progress, including motion sickness. However, electroencephalogram (EEG) requires calibration because it costs a lot for labeling and has a large variability between subjects. In this paper, we apply active learning using a deep neural network (DNN) to reduce calibration data. Previous studies have been conducted using only basic machine learning methods, so it can be a new research direction. In order to perform active learning in brain signals, pseudo labeling using DNN is performed. Here, an uncertain sample close to the decision boundary was selected using uncertainty sampling. Our results show active learning can produce similar overall classification accuracy with 1% source data.

Index Terms— Active learning, brain-computer interface, electroencephalography, motion sickness

1. INTRODUCTION

Electroencephalogram (EEG)-based brain-computer interfaces (BCI) is a system that establishes direct communication between the brain and external devices. It was originally developed for patients with difficulty communicating, such as paralyzed people. EEG is an electrical signal that measures the electrical activity of the brain non-invasively by electrodes attached to the head surface. BCI systems mainly use EEG due to high temporal resolution, user mobility, and relatively low cost. However, EEG requires calibration because it costs a lot to label the data and has a large variability between subjects. Therefore, two motivations can be considered. First, some subject-specific calibration data are required to build the classifier for each subject. Second, one of the main issues in BCI applications is to train models for new subjects as fast as possible and reduce the demand for labeled data.

Active learning has been mainly used with machine learning to classify brain signals in BCI. However, a deep neural network (DNN) was not used in the BCI field due to the characteristics of the EEG signal. DNN is a subcategory of machine learning and has made impressive progress in computer vision [1, 2] and natural language processing [3].

Many EEG studies have been conducted on motion sickness. Motion sickness is a common experience for many people and has been recognized as a problem to be solved, such as virtual reality motion sickness. It is necessary to recognize motion sickness in advance and prevent motion sickness. Motion sickness is mainly evaluated using the SSQ [4]. The purpose of this paper is to predict the degree of motion sickness by effectively reducing calibration data between subjects. The dataset used the MISC [5], which quantifies subject's motion sickness in real-time.

2. RELATED WORKS

2.1. Active learning in BCI

The active learning approach appears to be studied on EEG recordings since 2011 [6]. To prove the use of active learning in the BCI paradigm, Marathe et al. [7] used a combination of QBC [8] and uncertainty sampling by using hierarchical discriminant component analysis (HDCA), common spatial patterns (CSP), and XDawn with Bayesian linear discriminant analysis (XDBLDA) [9] as the committee of classifiers. The EBMAL algorithm [10] was proposed to improve the reliability, representativeness, and diversity of samples selected by a baseline QBC and Expected Model Change Maximization (EMCM) approach. They introduced the limitations of QBC and EMCM [11, 12, 13, 14] and proposed an algorithm using clustering to overcome them.

Wu et al. [15] proposed active semi-supervised transfer learning for offline BCI calibration. ASTL is an algorithm that integrates active learning, weighted adaptation regularization (wAR) [16], and spectral meta-learner (SML) [17]. wAR and SML were used to estimate the pseudo label.

Based on the above-related works, BCI has not conducted much research on active learning. The uncertainty-based approach is mainly used in certain shallow models. EEG is high-dimensional data that DNN models are not well utilized, so this paper aims to reduce rhe data required for calibration by doing active learning with a DNN model in BCI.

2.2. Motion sickness

To assess the participants' level of motion sickness, the 11-point MISC sclae was used. Both before the experiment and at 1-min intervals during the 20 min, the participant in-

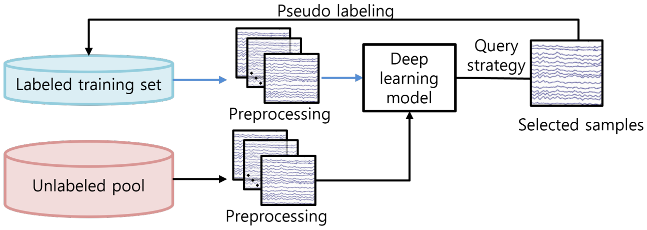

Fig. 1. Proposed framework for deep active learning.

dicated their score on the MISC. The scale is based on the knowledge that nausea, retching, and vomiting as a result of motion sickness are virtually always preceded by initial symptoms such as sweating, yawning, apathy, stomach awareness, and dizziness. [18]

3. PROPOSED METHOD

Since EEG is signal data, unlike image domains, an auto oracle should be used. Here, EEGNet[19] in the BCI domain was used to utilize pseudo labeling. The EEGNet is a compact CNN for the classification and interpretation of EEG-based BCIs. The network use depthwise and separable convolutions for EEG classification. So this network can be used to construct EEG-specific models encapsulating well-known EEG feature extraction concepts.

Fig. 1 shows the overall framework of active learning using EEGNet. First, the deep learning model is trained by preprocessed labeled training set. Using this model, find one of the most uncertain samples in the unlabeled pool. The sample is labeled with EEGNet and added to the labeled training set. Note that this process is pseudo labeling using EEGNet. This is the oracle part that should be designed for active learning. This process is repeated.Table 1 shows definition of notations. Pseudo code for active learning procedure is shown in Algorithm 1.

Active learning is provided with only a few samples labeled for each class to learn the initial decision boundary. With this limitation, we aim to learn this decision boundary using only a few labeled samples. One of the strategies is to randomly select samples. However, there is a possibility that the selected point may not provide important information. Therefore, many other strategies have been proposed to select sufficiently informative samples. Among them, uncertainty sampling was proposed. Uncertainty sampling is a strategy to find the highest uncertain sample that makes the decision boundary most unclear and to clarify the decision boundary with a little data by querying the point about it with a query. In this paper, two of the uncertainty strategies are used. The first is least-confidence strategy, it computes

$$x_{Lc} = \arg\max_{x} 1 - P_\theta(\hat{y}|x) \tag{1}$$

where $\hat{y}$ is shorthand for the model's predicted output for a given instance x, while y indicates the true label for that instance. More specifically, $\hat{y} = \arg\max_y P_\theta(y|x)$, or the class label with the highest posterior probability under the model θ. The second is the Monte Carlo dropout [20]. It is used to estimate predictive uncertainty at test time. It should be noted that since it is a dropout [21], it is applied only to active learning using DNN.

Table 1. Definition of notations

Notation	Description
X_{train}	A small set of labeled instances
X_{pool}	A large pool of unlabeled instances
X_{valid}	Valid data set
X_{test}	Test data set
$budget$	Total number of active learning iterations
k	Index of sample selected by active learning at each iteration

Algorithm 1 Active learning algorithm using EEGNet

Input $X_{train}, X_{pool}, X_{valid}, X_{test}, budget$
Output Accuracy of X_{test} at each iteration
while $budget \neq 0$ **do**

(1) Train EEGNet using X_{train} and calculate for X_{valid}

(2) Use trained classifiers to predict probabilites of X_{pool}

(3) Extract one sample of X_{pool} with most uncertain sample $X_{pool(k)}$

(4) Query the label for $X_{pool(k)}$ from the oracle (pseudo labeling) and merge sample into X_{train}

$$X_{train} = X_{train} \cup X_{pool(k)}$$

$$X_{pool} = X_{pool} - X_{pool(k)}$$

$$budget = budget - 1$$

end while

4. EXPERIMENT AND RESULTS

4.1. Dataset

The dataset is recorded by 23 healthy experimenters in a real driving environment with motion sickness level and 28 channels of EEG data at 500 Hz sampling. All subjects repeated the same experiment twice over two days. Each subject enters their MISC with a keypad every one minute during the 20-minute experiment. To recognize and predict motion sickness, the classification was performed by dividing the MISC of 11 classes into 3 classes. 0 to 1 was set as one class,

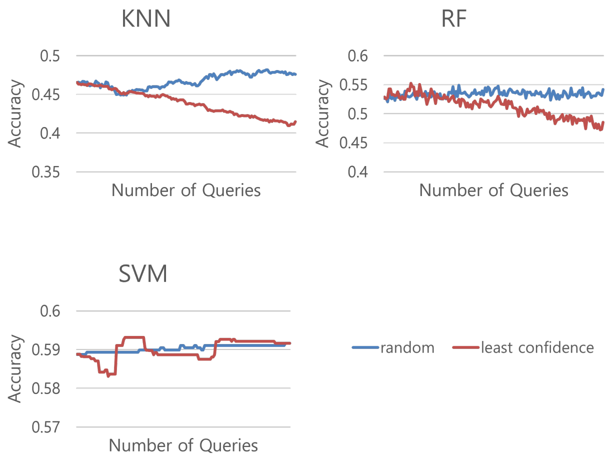

Fig. 2. Example of performance using basic machine learning algorithm

2 to 5 as the second class, and the remaining 6 or more MISC as the third class. For preprocessing the data, a 0.5 Hz high-pass filter was used on the recorded EEG data. Aditionally, the high-frequency noise component was removed from the EEG data using a low-pass filtered at 50 Hz. Then to lower the computational load, the data was downsampled to 250 Hz [22]. In addition, all data were scaled from -10 to 10.

4.2. Experimental setting

One subject was set as target data for active learning setting. Therefore, the remaining 22 subjects were set to source data, 5% of the source data was set to train data, and the rest was set to pool data. Day1 data of the target subject was set as validation data, and day2 data was set as test data. The number of all queries was fixed at 150. To compare active learning using previous machine learning and active learning using DNN, various machine learning methods were experimented. Support vector machine (SVM), k-nearest neighbors (KNN), and random forest (RF) were used as machine learning methods. PCA [23, 24] was then used as an feature extractor to summarize the variances and extract the first few principal components (PCs).

Our approach is implemented with the Pytorch library in Python. We train our DNN with SGD optimizer [25], which has a learning rate of 0.001. All epoch is 100. Step learning rate scheduler was used, and step size was set to 1000 and gamma was set to 0.1.

4.3. Experimental results and discussion

Fig. 2 is an example of performance using basic machine learning. The blue graph represents the account of active learning using random sampling as the number of queries increases. Red is the performance of the least-confidence strategy under the same conditions. It is noteworthy that as the number of queries increases, it can be seen that random sampling results in KNN and RF are better than least confidence. More specifically, the performance of KNN is very poor than that of RF, which may be due to the fact that RF does not work well on high-dimensional and scarce data. In KNN, the performance tends to increase further as random sampling is used. However, when using least-confidence strategy, the performance dropped by about 5%. In RF, the performance of the least-confidence strategy is better when the number of queries is small. This can be said to be because the number of samples with more information was well labeled at the beginning. In SVM, the result of least confidence is more unstable than that of random sampling. Random sampling continues to maintain performance of about 59%, while least-confidence strategy tends to move up and down.

Table 2. Performance of the EEGNet without active learning

sub	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Acc	0.69	0.49	0.64	0.98	0.36	0.66	0.95	0.44
sub	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Acc	0.5	0.33	0.93	0.69	0.73	0.43	0.66	0.93
sub	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**Avg.**
Acc	0.46	0.58	0.2	0.58	0.74	0.67	0.55	0.62

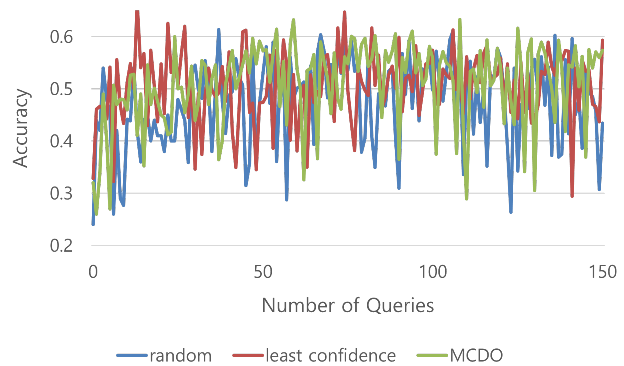

Fig. 3. Performance of active learning with EEGNet

Table 2 shows the performance of EEGNet without active learning. It is noteworthy that this performance is the result of using all-source data. On average, it shows 62% performance. Fig. 3 shows the result of active learning, which performed pseudo labeling using EEGNet. Comparing table 2 and Fig. 3, the average performance is as low as about 55% when active learning is used than when all training data is used. However, this is valuable that 150 queries correspond to only 1% of data in the pool.

These results are also related to the characteristics of motion sickness. Due to the characteristics of motion sickness experiments, the intensity of motion sickness tends to increase in the second half of the experiment. In addition, the data does not stable due to the imbalance in datasets. The cause of the dataset imbalance appears to be the deviation of motion sickness from person to person. Therefore, better performance can be expected if approached after solving this imbalance problem.

Although it does not exceed 62% of EEGNet's performance using all source data, the average performance improves as the number of queries increases when active learning using DNN.

5. CONCLUSION

This paper shows that active learning using DNN can be used in BCI domains. Although performance is lower than that of using all source data, it is a valuable discovery to be able to achieve similar performance using only 1% of data in a pool data. In addition, the importance of using DNN can be more highlighted in that it performs much better than when using basic machine learning algorithms.

6. REFERENCES

[1] Leon A Gatys, Alexander S Ecker, and Matthias Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2414–2423.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[3] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria, "Recent trends in deep learning based natural language processing," *IEEE Computational intelligenCe magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[4] Susan Bruck and Paul A Watters, "Estimating cybersickness of simulated motion using the simulator sickness questionnaire (ssq): A controlled study," in *2009 sixth international conference on computer graphics, imaging and visualization*. IEEE, 2009, pp. 486–488.

[5] Jelte E Bos, Willem Bles, and Eric L Groen, "A theory on visually induced motion sickness," *Displays*, vol. 29, no. 2, pp. 47–57, 2008.

[6] Jakub Šebek, Hana Schaabova, and Vladimir Krajca, "Active learning approach for eeg classification using neural networks: A review," in *2019 E-Health and Bioengineering Conference (EHB)*. IEEE, 2019, pp. 1–4.

[7] Amar R Marathe, Vernon J Lawhern, Dongrui Wu, David Slayback, and Brent J Lance, "Improved neural signal classification in a rapid serial visual presentation task using active learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 3, pp. 333–343, 2015.

[8] H Sebastian Seung, Manfred Opper, and Haim Sompolinsky, "Query by committee," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 287–294.

[9] Bertrand Rivet, Antoine Souloumiac, Virginie Attina, and Guillaume Gibert, "xdawn algorithm to enhance evoked potentials: application to brain–computer interface," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, 2009.

[10] Dongrui Wu, Vernon J Lawhern, Stephen Gordon, Brent J Lance, and Chin-Teng Lin, "Offline eeg-based driver drowsiness estimation using enhanced batch-mode active learning (ebmal) for regression," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016, pp. 000730–000736.

[11] Wenbin Cai, Ya Zhang, Siyuan Zhou, Wenquan Wang, Chris Ding, and Xiao Gu, "Active learning for support vector machines with maximum model change," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2014, pp. 211–226.

[12] Burr Settles and Mark Craven, "An analysis of active learning strategies for sequence labeling tasks," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 2008, pp. 1070–1079.

[13] Burr Settles, Mark Craven, and Soumya Ray, "Multiple-instance active learning," *Advances in neural information processing systems*, vol. 20, pp. 1289–1296, 2007.

[14] Burr Settles, "Active learning literature survey," 2009.

[15] Dongrui Wu, "Active semi-supervised transfer learning (astl) for offline bci calibration," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017, pp. 246–251.

[16] Dongrui Wu, "Online and offline domain adaptation for reducing bci calibration effort," *IEEE Transactions on human-machine Systems*, vol. 47, no. 4, pp. 550–563, 2016.

[17] Fabio Parisi, Francesco Strino, Boaz Nadler, and Yuval Kluger, "Ranking and combining multiple predictors without labeled data," *Proceedings of the National Academy of Sciences*, vol. 111, no. 4, pp. 1253–1258, 2014.

[18] Ouren X Kuiper, Jelte E Bos, Eike A Schmidt, Cyriel Diels, and Stefan Wolter, "knowing what's coming: unpredictable motion causes more motion sickness," *Human factors*, vol. 62, no. 8, pp. 1339–1348, 2020.

[19] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, pp. 056013, 2018.

[20] Yarin Gal and Zoubin Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.

[21] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[22] Kuan-Chih Huang, Alka Rachel John, Tzyy-Ping Jung, Wen-Feng Tsai, Yi-Hsin Yu, and Chin-Teng Lin, "Comparing the differences in brain activities and neural comodulations associated with motion sickness between drivers and passengers," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1259–1267, 2021.

[23] Svante Wold, Kim Esbensen, and Paul Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[24] Jun Li and Dacheng Tao, "Simple exponential family pca," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 453–460.

[25] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton, "On the importance of initialization and momentum in deep learning," in *International conference on machine learning*. PMLR, 2013, pp. 1139–1147.