Data Augmentation Methods for Machine-learning-based Classification of Bio-signals

Asuka Sakai, Yuki Minoda, and Koji Morikawa
Advanced Research Division
Panasonic Corporation
Tokyo, Japan
{sakai.asuka1, minoda.yuki, morikawa.koji}@jp.panasonic.com

Abstract **— Data augmentation methods for bio-signal classification are proposed. These methods improve recognition performance of human mental states showing intrinsic motivation from brain wave. Conventionally, data augmentation is used to image recognition research. Scaling, rotation, and distortion are applied to the original images to increase examples for machine learning. However, these augmentation methods are not effective for use with biological signals, as they involve spatial manipulation designed to represent the fluctuations of natural images. In the present study, we proposed four novel methods for data augmentation of biological signals. These methods are designed to represent variations inherent to bio-signals, especially for event-related signals. Electroencephalogram (EEG) data from participants engaged in an intrinsic motivation task were utilized to evaluate the feasibility of the proposed data augmentation methods. Our findings demonstrated that the proposed methods are particularly effective for improving prediction accuracy in small datasets.**

Keywords—Data Augmentation, Deep Learning, Bio-signals; Brain wave; EEG;

I. INTRODUCTION

The classification of human physical and mental states based on biological signals is important to design medical or healthcare application. While electroencephalogram (EEG) signals are often used to estimate mental states, there are no clear relationships between mental states and specific patterns of brain activity. Machine learning algorithms, such as those of the deep learning may aid in the extraction of complex relationships from collected data.

When applying machine learning algorithms to biological signals, large amounts of data must be acquired to ensure better performance. However, collecting sufficient amounts of data is often difficult due to the burden it places on both experimenters and participants. In addition to the length of time required for data collection, participants must maintain focus without moving their bodies for long periods of time in order to reduce the impact of noise.

Data augmentation, which expands virtual datasets produced from the collected bio-signals, may represent a viable method for decreasing data collection time in human participants. In image recognition research, data augmentation is commonly used to improve discrimination and generalization performance. Scaling, rotation, shearing, and elastic distortion are applied to artificially generate additional training data and increase the number of image datasets. The application of deep learning algorithms to such expanded datasets has been associated with high generalization performance [1-3]. However, these methods are not effective for use with bio-signals such as those obtained via EEG because the data augmentation methods for image recognition are specifically prepared to simulate fluctuation of the natural images. For example, scaling is used to respond to variable distance between a camera and a given object, while rotation is used to respond to variations in object angle. While object distances and angles are irrelevant to the analysis of bio-signals, the data points nonetheless exhibit variations in time or amplitude.

Biological signals typically exhibit time and amplitude variations such as those observed for the P300 of event-related potentials [4]. In the present study, we proposed a data augmentation method for use with event-related biological signals. The proposed method generates augmented data by shifting the axes of time and amplifying potentials of the measured data. Additional datasets were created using four methods of data augmentation, and the performance of datasets with and without augmentation was compared. Our results demonstrated that data augmentation improved performance, especially within small datasets. Such findings indicate that customized data augmentation may improve the collection of bio-signals in human participants.

II. DATA AUGMENTATION METHODS

To evaluate our proposed method, we developed a machine learning classifier capable of predicting task labels based on measured EEG data. We utilized task labels and EEG data from our previous study regarding intrinsic motivation [5]. Deep learning was applied for classification.

A. Proposed data augmentation method

To design the customized data augmentation methods of the present study, we focused on the inherent fluctuations of time and amplitude associated with bio-signals. Figure 1A includes EEG data for a representative participant obtained over 10 trials (SW: 5 trials, WS: 5 trials, N=1) without data augmentation. Evoked potentials [μV] are represented by the vertical axis, while elapsed time [ms] from the onset of the button press is represented on the horizontal axis.

Four methods of data augmentation were utilized to generate additional datasets in the present study, as follows:

a) Shifting all-time data (All-Shift)

All sampled data were shifted by a specified amount of time. Figure 1B depicts the dataset augmented from that presented in Figure 1A. Since the shifted range was +10 ms and -10 ms, the original dataset was tripled. These time-shifts represent temporal variations of brainwaves observed in the behavioral experiments.

b) Amplifying all-time data (All-Amp)

All sampled data were multiplied by a specified number. Figure 1C depicts the dataset augmented from that presented in Figure 1A. Since the factors used were 90% and 110%, the original dataset was tripled. These multiplications represent variations in the amplitude of brain waves due to factors such as electrode-tissue impedance.

The latter two augmentation methods are derivatives of those presented in a) and b). Each augmentation method was applied only to certain rather than whole periods. In the event-related potential interpretation, differences in mental states often emerge during specific periods (e.g., the peak around 300 ms), rather than at the time of stimulus presentation. Thus, augmentation to periods near the focused peak may be effective. The period was defined as -200 ms to +200 ms from the average time at which the peak occurred.

c) Shifting near-peak data (Peak-Shift)

Only near-peak data were shifted +10 ms and -10 ms. The period ranged from 250 ms to 650 ms (Figure 1D).

d) Amplifying near-peak data (Peak-Amp)

Only near-peak data were multiplied by 90% and 110%. The period ranged from 250 ms to 650 ms (Figure 1E).

As an additional augmentation parameter, the degree of time shift and multiplication factors were altered for comparison. The degree of time shift ranged from ±5 ms to ±50 ms every 5 ms in a) and c). Multiplication factors ranged from ±5 % to ±50 % every 5 % in b). In d), they ranged from 5 % to 50 % every 5 %.

B. Experimental task

The present study utilized two behavioral tasks designed to evaluate intrinsic motivation, as previously described [6]. Participants performed a stopwatch (SW) task and watch-stop (WS) task. In the SW task (i.e., “motivated task”), participants were instructed to press a button within 50 ms of the 5-s time point. A reward was added as their score when they succeeded. In the WS task (i.e., “unmotivated task”), they were instructed to simply press the button when the stopwatch automatically stopped. Success and failure were not defined in this task. The purpose of this classification was to allow for estimation of the participant’s mental state (motivated or unmotivated) based on EEG signals. Participants performed each task 30 times in a pseudo-random order, for a total of 60 trials.

C. EEG data and analysis

Ten healthy adult volunteers with normal or corrected-to-normal vision participated in the present study. All participants provided written informed consent and the protocol was approved by the Ethics Committee of Panasonic corporation.

EEG signals were recorded from two scalp electrode sites (Cz, Pz) based on an extended 10-20 system [7] using sintered Ag/AgCl electrodes. Only EEG data from the Cz site were used for analysis. The sampling rate was 1,024 Hz, and a digital bandpass filter of 1–30 Hz was applied offline. Data obtained between 200 ms before and 1000 ms after the onset of the button press were used for the present analysis. The baseline was aligned to the mean amplitude of the 200 ms pre-stimulus period. The trials in which EEG exceeded ± 60 μV were excluded from the analysis.

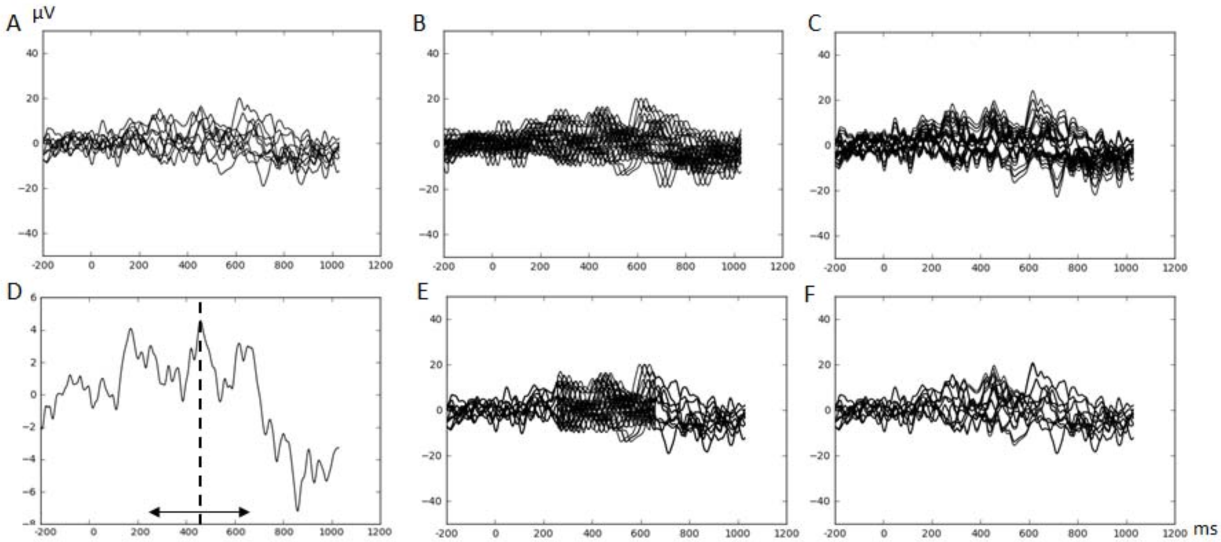

Fig. 1 Examples of training data with/without data augmentation.

D. Classification and experiments

The present study utilized a deep learning architecture with two intermediate layers. All single-trial brainwave data were used as direct input layers (1,230 points). The numbers of intermediate and output layers were 200, 100, and 2, respectively. A total of 100 epochs were utilized, and ten-fold cross-validation was applied for evaluation.

EEG data from five participants were prepared for simulation. The maximum number of trials was 60 (SW: 30 trials, WS: 30 trials). Participants with data obtained over less than 50 trials were excluded from analysis, as their performance could not be accurately compared. Data augmentation was then performed for the samples obtained from the remaining participants.

Four experiments were conducted to investigate the effectiveness of the proposed data augmentation methods. These experiments examined the a) effect of original dataset size, b) effectiveness of the proposed data augmentation methods, c) effectiveness of data augmentation for original datasets of various sizes, and d) compared the four augmentation methods for use with different sizes of augmented datasets. Performance was evaluated based on the accuracy of task prediction.

III. RESULTS

A. Pre-analysis: Effect of original dataset size

Figure 2 shows the average prediction performance of deep learning without data augmentation in five participants. The accuracy values of the classifier learned by all trials (average: 58.5 trials) and 40 trials were similar (73.9% and 74.2%, respectively). However, accuracy decreased to 68.5% when only 20 trials were used. This finding suggests that performance decreases as the size of the original dataset decreases.

B. Performance of data augmentation for small datasets

Figure 3 shows the results of the four proposed data augmentation methods (All-Shift, All-Amp, Peak-Shift, and Peak-Amp) when used with the 20-trial dataset. The performance of the conventional method (without-DA: no data augmentation method) is also shown. Estimated accuracy was averaged over participants and parameters, the amount of shift, and size of the dataset. Accuracy was higher for all data augmentation methods (ranging from 73.0% to 74.2%) than when analyzed without data augmentation (68.5%).

C. Performance of data augmentation for various dataset sizes

Figure 4 shows accuracy results obtained with and without data augmentation for datasets of different sizes. Performance was averaged over parameters and participants. The performance of the proposed method was almost identical to that of the conventional method for the 40-trial and all-trial datasets. This finding suggests that the proposed

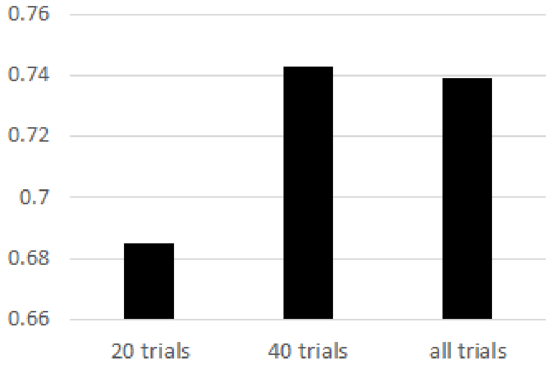

Fig. 2 Performance of classifier without data augmentation

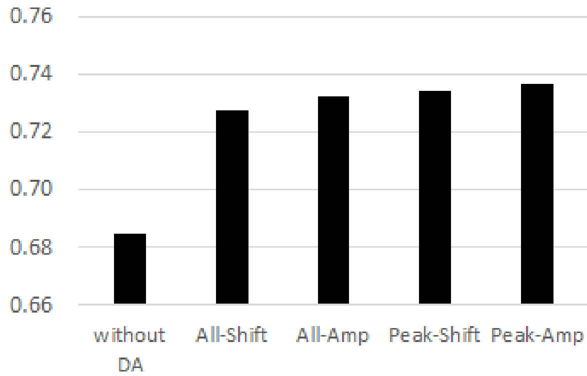

Fig. 3 Performance of classifier learned by training data for 20 trials with/without data augmentation methods.

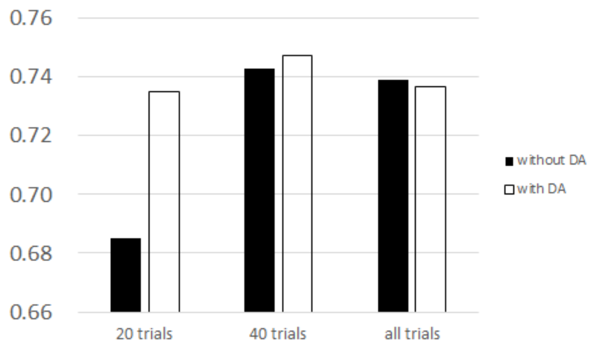

Fig. 4 Performance of classifier learned by training data with/without data augmentation methods

methods improve prediction accuracy, especially in smaller datasets.

D. Comparison of data augmentation methods

Figure 5 shows the performance of the four proposed methods when different augmentation parameters were utilized. The size of the total dataset is depicted on the horizontal axis. The label "x3" indicated that the original dataset has been tripled via augmentation (±5 ms time shift or ±5% change in amplitude). Similarly, the label "x5" indicates that the augmented dataset is five times larger than the original (±10 ms time shift or ±10% change in amplitude relative to the "x3" dataset). Training data for 20 trials were

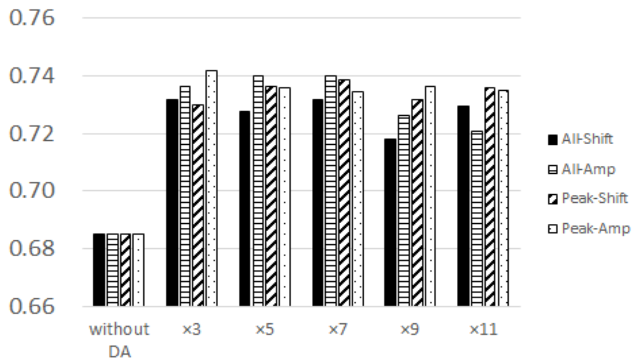

Fig. 5 Performance of classifier learned with/without four data augmentation methods

used. Accuracy was averaged over participants and parameters.

Our analysis revealed that Peak-Amp augmentation of the “x3” dataset was the most effective, exhibiting an 8.30% increase in performance relative to analysis without data augmentation. However, little difference in performance was observed among the four augmentation methods. Interestingly, the effectiveness of the All-Shift and All-Amp augmentation methods tended to decrease as the size of the dataset increased, when compared with the Peak-Shift and Peak-Amp augmentation methods.

IV. DISCUSSION

A. *Summary of findings*

The results of the present study indicate that data augmentation improves prediction accuracy for bio-signals when the dataset is small. Indeed, even when the size of the dataset was decreased by one third, little change in performance was observed when data augmentation was employed.

Biological signals exhibit inherent variations in peak time and amplitude. Thus, the methods proposed in the present study may be effective for small datasets because information regarding this variance cannot be obtained in small datasets. That is, the proposed data augmentation methods may have generated variance by shifting the time and amplitude of bio-signals, allowing the deep learning classifier to learn the variance. In particular, the results presented in Fig. 5 suggest that adding peak time- or amplitude-shifted augmented data enables better performance, while shifts applied across all time periods (All-Shift and All-Amp) are less effective.

In addition, our results indicated that performance accuracy tended to decrease with excessive amounts of data augmentation. In the field of image recognition, the number of images must be increased to a large extent to improve prediction performance [1]. However, when applying data augmentation to biological signals, parameters such as the range of the time or amplitude shift must be adjusted.

Furthermore, performance was also calculated for a single classifier applied to all participants with/without data augmentation. In this case, the dataset was large (413 trials) compared to those for which individual classifiers were utilized (20-60 trials). We observed that the accuracy of the classifier for all participants was lower with data augmentation than without data augmentation. Such findings suggest that augmentation added extra noise when the large dataset was used. Thus, the proposed augmentation methods seem to be adequate for classification based on individual datasets.

B. *Applicability to other bio-signals*

Though EEG data representing motivation were utilized in the present study, our findings suggest that the proposed methods can be applied to other mental states and biological signals. As we focused on ERP due to features such as time points and amplitude, the proposed methods can likely be applied to modalities that retain these features. Classifiers can also be created from other biological signals associated with stimulus events, such as those obtained via electromyography (EMG) and electrocardiography (ECG).

C. *Application example*

Proposed data augmentation is especially effective for small datasets. This finding suggests that it is possible to create classifiers with sufficient performance, even when the size of the dataset is limited, which may be useful when collecting data from participants who experience discomfort or boredom over a large number of trials.

V. CONCLUSIONS

The results of the present study indicate that the proposed data augmentation methods are effective for machine learning and classification based on a small number of biological signals. With data augmentation, performance was improved to almost the same accuracy as that observed for larger datasets without data augmentation.

Further experiments involving more participants are needed to ensure effectiveness of proposed methods.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, G E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Proc. Advances in Neural Information Processing Systems,* pp. 1090-1098, 2012.

[2] D. C. Ciresan, “Deep Big Simple Neural Nets Excel on Hand-written Digit Recognition,” *Neural Computation*, vol. 22, no. 12, 2010.

[3] P Y. Simard, D. Steinkraus, J C. Platt, ”Best Practices for Convolutional Neural Networks,” in *Proc. Seventh Int’l Conf. Document Analysis and Recognition*, vol. 2, pp. 958-962, 2003.

[4] T C. Handy, “*Event-Related Potentials: A Methods Handbook,”* Cambridge, MA, USA: MIT Press, 2004.

[5] Y. Minoda, K. Morikawa, “Identification of event-related potential component for intrinsic motivation between stop-watch and watch-stop task,” *Japanese Journal of Physiological Psychoology and Psychophysiology,* vol. 34, no.2, 2016 (in Japanese).

[6] K. Murayama, M. Matsumoto, K. Izuma, K. Matsumoto, “Neural basis of the undermining effect of monetary reward on intrinsic motivation,” *PNAS*, vol. 107, no. 49, pp. 20911-20916, 2010.

[7] G.H. Klem, H.O. Lüders, H.H. Jasper, and C. Elger, “The ten-twenty electrode system of the International Federation,” *Electroencephalography and Clinical Neurophysiology, Supplement* vol. 52, pp.3-6, 1999.