Regular paper

Journal of the KIECS. pp. 1171-1180, vol. 14, no. 6, Dec. 31. 2019, t. 98, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2019.14.6.1171

BCI에서 EEG 기반 효율적인 감정 분류를 위한 LSTM 하이퍼파라미터 최적화

Ibrahim Aliyu* · Raja Majid Mahmood** · 임창균***

LSTM Hyperparameter Optimization for an EEG-Based Efficient Emotion Classification in BCI

Ibrahim Aliyu* · Raja Majid Mahmood** · Chang-Gyoon Lim***

요 약

감정은 인간의 상호 작용에서 중요한 역할을 하는 심리 생리학적 과정이다. 감성 컴퓨팅은 감정을 이해하고 조절할 수 있는 인간 인지 인공 지능의 개발하는데 중점을 둔다. 우울증, 자폐증, 주의력 결핍 과잉 행동 장애 및 게임 중독과 같은 정신 질환이 감정과 관련되어 있기 때문에 이러한 분야의 연구가 중요하다. 감정 인식에 대한 노력에도 불구하고, 비정상적인 EEG 신호로부터의 감정 검출은 여전히 높은 수준의 추상화를 요구하기에 정교한 학습 알고리즘이 필요하다. 이 논문에서서는 EEG 기반으로 효율적인 감정 분류를 위해 LSTM을 위한 최적의 하이퍼파라미터를 파악하고자 다양한 실험을 수행하여 이를 분석한 결과를 제시하였다.

ABSTRACT

Emotion is a psycho-physiological process that plays an important role in human interactions. Affective computing is centered on the development of human-aware artificial intelligence that can understand and regulate emotions. This field of study is also critical as mental diseases such as depression, autism, attention deficit hyperactivity disorder, and game addiction are associated with emotion. Despite the efforts in emotions recognition and emotion detection from nonstationary, detecting emotions from abnormal EEG signals requires sophisticated learning algorithms because they require a high level of abstraction. In this paper, we investigated LSTM hyperparameters for an optimal emotion EEG classification. Results of several experiments are hereby presented. From the results, optimal LSTM hyperparameter configuration was achieved.

키워드

BCI(Brain Computer Interface), EEG(Electroencephalography), Emotion, LSTM-RNN
뇌 컴퓨터 인터페이스, 뇌파, 감정, 장단기 기억 구조 순환 신경망

Ⅰ. INTRODUCTION

Emotion is a psycho-physiological process that plays an important role in human interactions and can be expressed either verbally through emotional vocabulary or by expressing non-verbal cues such

* 전남대학교 컴퓨터공학전공 (187282@jnu.ac.kr)
*** 교신저자 : 전남대학교 컴퓨터공학전공
· 접 수 일 : 2019. 10. 26
· 수정완료일 : 2019. 11. 20
· 게재확정일 : 2019. 12. 15
· Received : Oct. 26, 2019, Revised : Nov. 20, 2019, Accepted : Dec. 15, 2019
· Corresponding Author : Chang-Gyoon Lim, Major in Computer Engineering, Chonnam National University
Email : cglim@jnu.ac.kr

as facial expressions intonation of voice and gestures. It is normally caused by the perception of object or situation and often associated with temperament, mood, personality, motivation and disposition [1].

Due to the importance of emotional intelligence in addition to logical intelligence, there has been increasing effort in the area of emotional artificial intelligence in Human-computer interactions (HCIs) and emergence of interdisciplinary research field known as affective computing [2, 3]. Affective computing is centered on the development of human-aware artificial intelligence that can understand and regulate emotions. This field of study is also critical as mental diseases such as depression, autism, attention deficit hyperactivity disorder and game addiction are associated with emotion [4]. However, due to the limited knowledge of the neural mechanisms underlying emotion processing , more effort are needed to improve the efficiency of detecting and measuring emotions in order to improve disease treatment among other benefits [2].

Different methods and models have been proposed to develop human-aware artificial intelligence that has the ability to perceive, understand regulate emotions. EGG-based emotion recognition is one of the popular method in these regards. Jirayucharoensak, et al. proposed the utilization of deep learning network to discover unknown feature correlation between input signals using stacked autoencoder and principal component analysis for feature extraction [5]. Atkinson and Campos proposed a novel feature-based emotion recognition model for emotion classification task by combining mutual information based feature selection methods and kernel classifiers. This technique attempts to address the restriction of small number of emotions classification usually caused by signal's features and noise, EEG constraints and subject-dependent issues [6]. Mohammadi, et al. developed a wavelet-based emotion recognition system using EEG signal. Support vector machine and K-nearest neighbor classifiers were used to detect emotion from features extracted using discrete wavelet transforms [7, 8]. Further literatures that proposed EEG-based emotion detection are reported in [9, 10, 11, 12].

The use of fusion technology to integrate various modalities in describing and detecting emotions has also been proposed [13]. Soleymani, et al. proposed a multimodal emotion recognition method with the goal of recovering affective tags for videos using electroencephalogram (EEG), pupillary response and gaze distance [14]. Zheng, et al. designed a six-electrode placement above the ears to collect electroencephalography (EEG) signals for a multimodal emotion recognition framework called emotion meter to combine brain waves and eye movements [2]. In addition, multimodal method combined with deep learning has been reported to be promising techniques [15, 16, 17, 18]. Three dimensionality reduction methods were investigated to Pincipal analyse the effect of dimensionality reduction methods on Epileptic EEG feature selection and classification [19].

Despite the efforts in emotions recognition, emotion detection from nonstationary EEG signals is still a challenging task as sophisticated learning algorithm that can represent high-level abstraction is required. In this paper, we proposed Recurrent Neural Network for emotion classification. Results of several experiments are hereby presented.

II. MATERIALS AND METHODS

The block diagram of the proposed model is presented in Fig. 1. It consists of 4 stages, which include data acquisition, preprocessing, feature extraction and classifier.

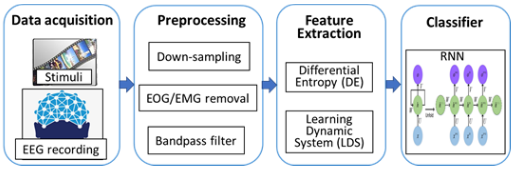


Fig. 1 Block diagram of the proposed model

2.1 EEG data Acquisition

The emotion EEG dataset used for this study is obtained from BCMI laboratory and published in [20]. During data collection, fifteen subjects were exposed to emotional films (positive, neural and negative emotion) containing both scene and audio, which can expose subjects to more real life scenarios and elicit strong subjective and physiological changes. 15 trials were conducted for each experiment. The protocol is shown as in Fig. 2.

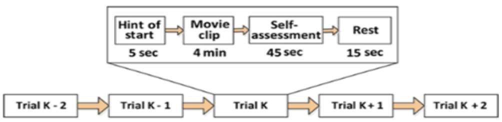


Fig. 2 Protocol of the EEG experiment [19]

2.2 Preprocessing

In preprocessing stage, three operations were conducted, which include downsampling, EOG/EMG removal and bandpass filtering. The EEG data was first downsampled at the rate of 200Hz. Then EMG/EOG contamination were then manually checked and removed. Bandpass filter of 0.3Hz to 50Hz was applied to remove noise and artifacts.

2.3 Feature selection

The features were extracted using different entropy (DE). Since EEG data has higher low frequency energy, DE is suitable for discriminating the low and high frequency energy of the EEG pattern. The DE was applied to construct features in five frequency band (delta: 1-3Hz, theta: 4-7Hz, alpha: 8-13Hz, beta: 14-30Hz, gamma: 31-50Hz) which were extracted using 256-point short Time Fourier Transform with a non-overlapped Hanning window of 1s. The DE is defined as

$$h(x) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp\frac{(x-\mu)^2}{2\sigma^2} \cdot \log\frac{1}{\sqrt{2\pi\sigma^2}} exp\frac{(x-\mu)^2}{2\sigma^2} dx = \frac{1}{2} log 2\pi e\sigma^2 \tag{1}$$

where the time series X obeys the Gauss distribution $N(\mu, \sigma^2)$. In order to further remove irrelevant component and obtain temporal dynamics of emotional state, linear dynamic system (LDS) was employed [21].

2.4 RNN Classifier

Deep learning is simply machine learning method that utilizes many hidden layers to carry out multiple levels of nonlinear operations in neural networks. With several transformations and multiple hidden layers, complex functions can be learned to discriminate the response classes in a classification problem.

Recurrent neural network is deep learning method that was developed for processing sequential data. It consists of feedforward neural networks with cyclic connections. It maps the entire history of input in the network to predict each output by leveraging the temporal relationships between the data at each point in time.

The architecture of a simple RNN is shown in Fig. 3. In the diagram, each node represents a layer of network at each timepoint. It consists of input, hidden, and output layer. The weighted connections from input to hidden layer, hidden to hidden layer and hidden to output layer are represented in the matrix U, W, and V respectively. The final weight matrix is passed through a softmax function to produce a scalar Y value, which is then classified as a binary variable referenced as $\hat{Y}$. The loss function is then applied to compare Y-actual and Y-predicted ($\hat{Y}$). However, the current RNN has a vanishing gradient problem in which a given input influences hidden layer and subsequently on network output [13].

This either decays or explodes exponentially over time as the data goes through transformations on the network. Therefore, two popular solutions have been developed: Long Short-time memory (LSTM) and gate recurrent unit (GRU).

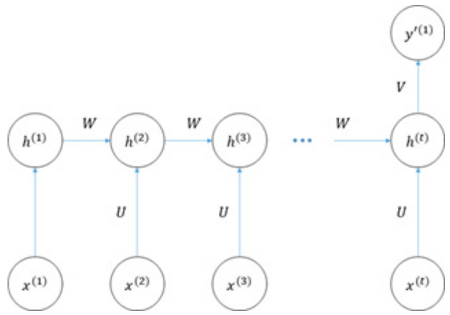


Fig. 3 RNN model. It is a many-to-one type of RNN with t-input nodes, hidden layer (h) and output layer (y)

In this research LSTM-RNN architecture is deployed. It is made up of connected subnetworks called memory block, which remembers inputs for a long time. Blocks contain at least one self-connected accumulator cell and several multiplicative units such as input gate, forget gate and output gate. Information is stored and accessed through the gates by assigning a counter such as 0 and 1. LSTM controls its state update and output by maintaining a hidden vector h and a memory vector m at each time step. The computation at each time step is define as

$$
\begin{aligned}
g^u &= \sigma\left(W^u h_{t-1} + I^U x_t\right) \qquad (2)\\
g^f &= \sigma\left(W^f h_{t-1} + I^f x_t\right)\\
g^o &= \sigma\left(W^o h_{t-1} + I^o x_t\right)\\
g^c &= \tanh\left(W^U h_{t-1} + I^U x_t\right)\\
m_t &= g^f \odot m_{t-1} + g^u \odot g^c\\
h_t &= \tanh\left(g^f \odot m_t\right)
\end{aligned}
$$

where σ is the logistic sigmoid function, $\odot$ represents elementwise multiplication, W^u, W^f, W^o, W^c are recurrent weight matrices and I^u, I^f, I^o, I^c are projection matrices.

In this work, we are focused on recognizing emotion, thus we cast the problem as multilabel classification. Given a series of observations x(1),,....., x(T), we learn a classifier to generate hypotheses $\hat{y}$ of the true labels y. Here, t indexes sequence steps, and for any example, T stands for the length of the sequence. Our proposed LSTM-RNN uses memory cells with forget gates without peephole connections.

Hidden state for each timestep of a neuron is return for every layer in the network. In the first layer 128 hidden neurons and a dropout were applied while subsequent layers consist of 256 hidden neurons. Since our problem is multilabel classification, a dense layer of 3 neurons followed by an element-wise softmax activation function, is implemented at the output layer, while cross entropy is used for loss function. The softmax activation function allows the model to interpret the outputs as probabilities while cross-entropy speeds up the learning process by canceling out the plateaus at each end of the soft-max function.

III. EXPERIMENT RESULTS AND PERFORMANCE ANALYSIS

The experiments were conducted using 7500 data, where 75% of the data was used of training while 25% was used for validation. Separate 1000 test data were used to evaluate the model. Several experiments were conducted to determine the configuration of the model, which includes hyperparameters indicated in Table 1, while other configurations detailed in section 2.3 were maintained.

Table 1. Model hyperparameters for investigation

Method	Parameter
TIME_STEP	1
EPOCHS	100-45
BATCH_SIZE	2-4
DROP_OUT	0.150-0.166
Optimizer	Adagrad, Adam, Nadam, RMSprop

3.1 Experiment I : Dropout experiments against optimizer

The first experiment is to determine the optimal dropout that has a better generalization for our proposed model. The optimizer considered include Adaptive moment (Adam), Stochastic Gradient Descent (SGD), Nesterov Adam optimizer (Nadam), and Adaptive Gradient (AdaGrad). The performance of each optimizer was experimented by varying the dropout from 0.150 to 0.166. Five (5) layers with 100 epochs and 4 batch sizes were used in this experiment. As indicated from the result of the experiment on Table 2, the Adam is observed to have the best generalization of 0.815 accuracy at 0.150 dropout. While all AdaGrad generalizes best at 0.166 dropout with 0.714 accuracy.

Table 2. Dropout experiment against optimizers

NO	Drop out	Accuracy of optimizers (%)			
		Adam	SGD	RMS prop	Ada Grad
1	0.150	0.815	0.336	0.626	0.710
2	0.154	0.671	0.336	0.600	0.693
3	0.158	0.805	0.360	0.609	0.665
4	0.162	0.635	0.348	0.667	0.702
5	0.166	0.682	0.336	0.590	0.714

Generally, from the five experiment, AdaGrad maintain an approximate of 0.70 accuracy across the dropouts. Even though, Adam has the highest accuracy, the performance across the dropout varies with relatively large margin in which the lowest accuracy is 0.635 and highest accuracy is observed to be 0.815. RMSprop maintains an accuracy of about 0.60 with margin difference of 0.08 between maximum and minimum. SGD performs poorly, with highest accuracy of 0.360. Fig. 4 shows the graphical distribution of the performances of the optimizer with dropout.

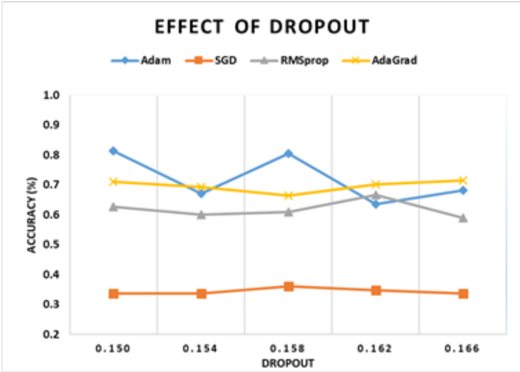
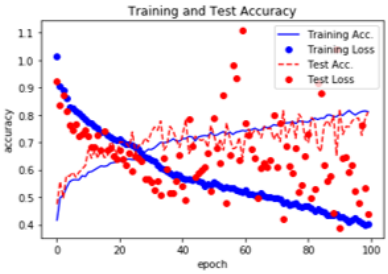


Fig. 4 Optimizer performance at different dropout

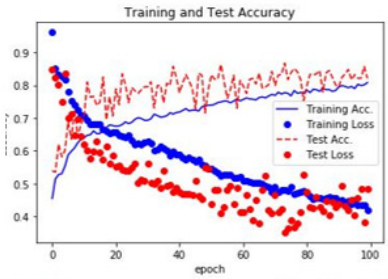
Considering the best performing dropout, Adam performed best with accuracies of 0.815 and 0.805 at the dropout of 0.150 and 0.158, respectively. Fig. 5 shows the training performance of the model during training and validation. At dropout of 0.150, the training accuracy progressively improved throughout the training but the validation(test) accuracy experience overfitting at epoch above 35.

The training loss was gradually decreased through the training, but the validation loss became very unstable at about epoch 30. On other hand, at dropout of 0.158, the training and validation accuracy progressively improved while the training accuracy progressively reduced.

However, the validation loss became unstable after epoch 40 while the testing accuracy begin to experience overfitting around epoch 80, although the performance generally experience bumps. Even though, that accuracy of the dropout 0.150 is higher compared to 0.158, the training performance shows that the 0.158 is more stable.



(a) Dropout 0.150



(b) Dropout 0.158

Fig. 5 Training and Validation Performance (a) Dropout 0.150 (b) Dropout 0.158

3.2 Experiment II : Epoch performance against optimizers

Furthermore, we investigated dropout 0.158 and batch 2 with different epoch while maintain the number of neurons and layers used in the experiment I. From Fig. 5(b), overfitting become more pronounced at about epoch 80. So we experimented from epoch 85 to 45. Also, since the number of epoch has been reduced, we investigated how reduction in batch size will affect the accuracy as the network is updated more with less epoch.

Table 3. Epoch performance against optimizer

NO	Epoch	Accuracy of optimizer (%)			
		Adam	SGD	RMS prop	Ada Grad
1	45	**0.812**	0.316	0.562	0.804
2	55	0.715	0.336	0.586	0.810
3	65	0.794	**0.316**	0.676	0.789
4	75	0.780	**0.316**	0.577	0.785
5	85	0.804	**0.336**	0.557	0.714

Table 3 presents the performance of each optimizer against the epochs under consideration. Epoch 45 with Adam gave the best performance while SGD gave the worst performance at epoch 45, 65 and 75.

For the five experiments conducted in this section, Adam and AdaGrad maintain accuracy of approximately 0.80 with one case on each side below 0.80. RMSprop maintain an average accuracy of 0.59 while SGD performed poorly by maintaining an average accuracy of 0.32. Fig. 6 presents the performances of the optimizer at different epochs.

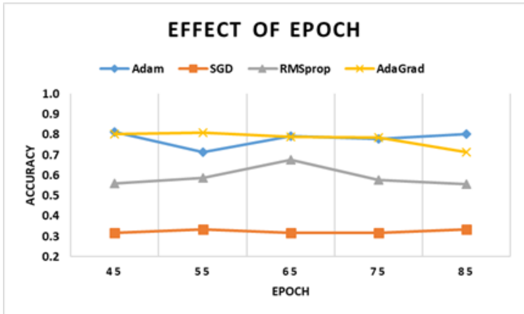


Fig. 6 Optimizer performance at different epochs

3.3 Experiment III : Performance accuracy of optimizer with layers

From experiment II, epoch 45 and 55 gave the best generalization of 0.812 and 0.810 with Adam and AdaGrad, respectively. In this experiment, we further determine how the depth of the network affect accuracy using the configuration of experiment II. The layer is varied between 3 to 7, while all other parameters are maintained. AdaGrad outperformed Adam with accuracy of 0.818 at epoch 55 with 4 layers, while Adam gave the worst result of 0.316 at epoch 55 with 6 layers. Table 4 presents the results of the experiment.

Table 4. Performance accuracy of optimizers with layers

NO	Layer	Accuracy of optimizer (%)			
		Adam		AdaGrad	
		epochs 45	epochs 55	epochs 45	epochs 55
1	3	0.725	0.755	0.691	0.674
2	4	0.761	0.74	0.791	0.818
3	5	0.812	0.715	0.802	0.81
4	6	0.64	0.772	0.666	0.58
5	7	0.475	0.316	0.660	0.585

As shown in Fig. 7, the performance of the model from 3 to 5 layers across all the optimizers progressively improved. However, at experiment 3 with 5 layers' network, epochs 55 for both optimizers experienced a decrease in accuracy. While the highest performances of 0.802 (epochs 45) and 0.818 (epochs 55) accuracy with AdaGrad were obtained with 4 and 5 layers, 7 layers' network gave the worst result of 0.316 with Adam optimizer.

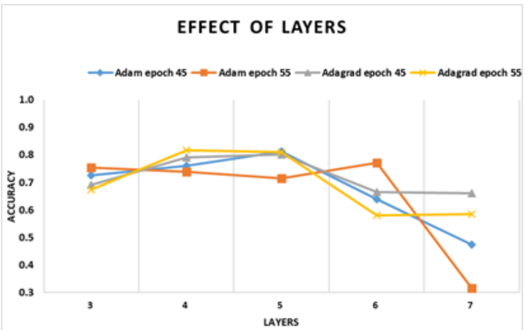


Fig. 7 Performance accuracy of optimizers with layers

Furthermore, after setting up the optimal parameters for our mode, we evaluated its performance against other classifiers. The classifiers considered for the evaluation include Logistic Regression (LR), SVM, Random Forest (RF) and Decision Tree (DT). From Table 5 and Figure 8, our model and LR performed approximately the same at accuracy of 0.82 while RF and DT scored 0.66 and 0.64, respectively.

Table 5. Model performance evaluation.

Classifier	Accuracy
LR	0.82
RF	0.66
DT	0.64
SLSTM	0.82

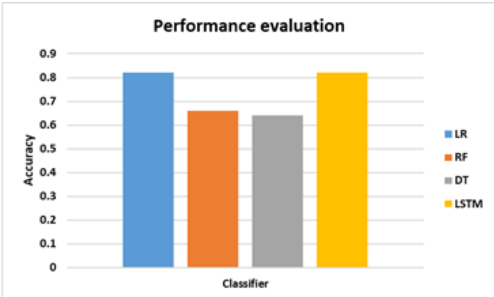


Fig. 8 Model performance evaluation.

Ⅳ. CONCLUSION

In this paper, we developed an LSTM-RNN model for the classification of EEG emotion. Various experiments were conducted to obtained optimal parameters for the model. The best generalization of 0.82% accuracy was obtained with AdaGrad at 0.158 dropout and 4 hidden layers. The model was compared against LR, RF, and DT, and observed to have the best accuracy of 0.82 as that of LR, followed by RF at 0.64 while RF performed the worst at 0.66 accuracy.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2017R1D1A1B03035988).

References

[1] Sander Koelstra, Christian Muhl, Mohammad Soleymanim, Jong-Seok Lee, Ashkan Yazdani, Touradj Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing,* vol. 3, no. 1, 2011, pp. 18-31.

[2] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE transactions on cybernetics,* vol. 49, no. 3, 2018, pp. 1110-1122.

[3] R. A. Calvo and S. D′Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on affective computing,* vol. 1, no. 1, *2010,* pp. 18-37.

[4] Al-Kaysi AM, Al-Ani A, Loo CK, Powell TY, Martin DM, Breakspear M, and Boonstra TW, "Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification," *Journal of affective disorders,* vol. 208, 2017, pp. 597-603.

[5] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, 2014, vol. 2014, pp. 1-10

[6] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert Systems with Applications,* vol. 47, 2016, pp. 35-41.

[7] Z. Mohammadi, J. Frounchi, and M. Amiri, "Wavelet-based emotion recognition system using EEG signal," *Neural Computing and Applications,* vol. 28, no. 8, 2017, pp. 1985-1990.

[8] Y. Kim, "Progressive Image Coding using Wavelet Transform," *J. of the Korea Institute of Electronic Communication Sciences,* Feb. 2014, vol. 9, no. 1, pp. 33-40.

[9] A. E. Vijayan, D. Sen, and A. Sudheer, "EEG-based emotion recognition using statistical measures and auto-regressive modeling," *in 2015 IEEE International Conference on Computational Intelligence & Communication Technology,* Ghaziabad, India, 2015, pp. 587-591.

[10] Y. Zhang, X. Ji, and S. Zhang, "An approach to EEG-based emotion recognition using combined feature extraction method," *Neuroscience letters,* vol. 633, 2016, pp. 152-157.

[11] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, and B. Yan, "Emotion recognition from EEG signals using multidimensional information in EMD domain," *BioMed research international,* vol. 2017, 2017, pp. 1-9.

[12] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Real-time EEG-based emotion monitoring using stable features," *The Visual Computer,* vol. 32, no. 3, 2016, pp. 347-358.

[13] S. K. D′mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Computing Surveys (CSUR),* vol. 47, no. 3, 2015, pp. 43:1-43:34.

[14] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE transactions on affective computing,* vol. 3, no. 2, 2011, pp. 211-223.

[15] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional MKL based multimodal emotion recognition and sentiment analysis," *in 2016 IEEE 16th international conference on data mining (ICDM),* Barcelona, Spain, 2016, pp. 439-448.

[16] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," *IEEE Journal of Selected Topics in Signal Processing,* vol. 11, no. 8, 2017, pp. 1301-1309.

[17] H. Ranganathan, S. Chakraborty, and S. Panchanathan, "Multimodal emotion recognition using deep learning architectures," *in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV),* Lake Placid, NY, USA, 2016, pp. 1-9.

[18] Z. Zhang, F. Ringeval, B. Dong, E. Coutinho, E. Marchi, and B. Schüller, "Enhanced semi-supervised learning for multimodal emotion recognition," *in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Shanghai, China, 2016, pp. 5185-5189.

[19] Tong, I. Aliyu, and C. Lim, "Analysis of Dimensionality Reduction Methods Through Epileptic EEG Feature Selection for Machine

Learning in BCI," *J. of the Korea Institute of Electronic Communication Sciences,* vol. 13, no. 6, Dec. 2018, pp. 1333-1342.

[20] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development,* vol. 7, no. 3, 2015, pp. 162-175.

[21] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," *in 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER),* San Diego, CA, USA, 2013, pp. 81-84.

저자 소개

Ibrahim Aliyu

2014: Federal University of Technology, Minna (B.Eng)
2018: Federal University of Technology, Minna (M.Eng)

※ 관심분야 : Machine Learning, Internet security, Software Defined, networking, Cloud Computing, Internet of Things

Raja Majid Mahmood

2014: Gomal University, Pakistan(B.S.)
2010년: Linmaeus University, Sweden(M.S.)
Chonbuk National University, South Korea (PhD)

※ 관심분야 : Affective computing, Brain computer Interfaces, Information Visualization, Image processing and pattern recognition

임창균 (Chang-Gyoon Lim)

1990 : Chosun Univversity (M.S.)
1997년 Wayne State University, Computer Engineering(Ph. D.)
2011년~2014 광주테크노파크 가전로봇지원센터 센터장
1997년~현재 전남대학교 공학대학 전기전자통신컴퓨터공학부 교수

※ 관심분야 : 인공지능, 지능형로봇, 기계학습, 패턴인식, 클라우드 컴퓨팅, 블록체인