LSTM/RNN을 사용한 감정인식을 위한 스택 오토 인코더로 EEG 차원 감소

Ibrahim Aliyu* · 임창균**

EEG Dimensional Reduction with Stack AutoEncoder for Emotional Recognition using LSTM/RNN

Ibrahim Aliyu* · Chang-Gyoon Lim**

요약

감성 컴퓨팅은 인간의 상호 작용에서 중요한 역할을 하기 때문에 인간을 인식하는 인공 지능을 통해 감정을 이해하고 식별한다. 우울증, 자폐증, 주의력 결핍 과잉 행동 장애 및 게임 중독과 같은 정신 질환을 잘 이해하고 관리할 수 있을 것이다. 이러한 문제들을 해결하기 위해 감정 인식을 위한 다양한 연구가 수행되었는데 기계학습을 적용하는데 있어서는 알고리즘의 복잡성을 줄이고 정확도를 향상시키기 위한 노력이 필요하다. 본 논문에서는 이러한 노력중의 하나로 Stack AutoEncoder (SAE)를 이용하여 차원 감소하는 방법과 Long-Short-Term-Memory/Recurrent Neural Networks (LSTM / RNN) 분류를 이용한 감성 분류에 대해 연구한 결과를 제시한다. 제안된 방법은 모델의 복잡성을 줄이고 분류기의 성능을 크게 향상시킨 결과를 가져왔다.

ABSTRACT

Due to the important role played by emotion in human interaction, affective computing is dedicated in trying to understand and regulate emotion through human-aware artificial intelligence. By understanding, emotion mental diseases such as depression, autism, attention deficit hyperactivity disorder, and game addiction will be better managed as they are all associated with emotion. Various studies for emotion recognition have been conducted to solve these problems. In applying machine learning for emotion recognition, the efforts to reduce the complexity of the algorithm and improve the accuracy are required. In this paper, we investigate emotion Electroencephalogram (EEG) feature reduction and classification using Stack AutoEncoder (SAE) and Long-Short-Term-Memory/Recurrent Neural Networks (LSTM/RNN) classification respectively. The proposed method reduced the complexity of the model and significantly enhance the performance of the classifiers.

키워드

Electroencephalogram(EEG), Stack AutoEncoder(SAE), LSTM/RNN, Feature Dimensional Reduction

*전남대학교 연구원(cglim@jnu.ac.kr)
**교신저자: 전남대학교 컴퓨터공학전공
·접 수 일: 2020. 05. 28
·수정완료일: 2020. 07. 06
·제재확정일: 2020. 08. 15

Received: May, 28, 2020, Revised: Jul, 06, 2020, Accepted: Aug, 15, 2020

Corresponding Author: Chang-Gyoon Lim
Major in Computer Engineering, Chonnam National University,
Email: cglim@jnu.ac.kr
I. INTRODUCTION

Emotion is a psycho-physiological process in human interactions that can be expressed either verbally through emotional vocabulary, or by expressing non-verbal cues such 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]. EEG signals with several channels analyzed the signals into several frequency bands investigate characteristics of the EEG signals related to the human concentration [2].

EEG is an effective way to recognize emotion as it measures the general response of the electrophysiological activity of brain nerve cells associated with physiological and psychological activities. Since Emotions are two-dimensional continuous space [3], thus we assume that the change of emotional state to be a continuous process. Hence, temporal correlation of EEG signals reflects the current emotional state as being affected by the past emotional state. This makes LSTM/RNN a good candidate for Emotion recognition as it explores the correlations of the time series.

Several efforts have been made in applying LSTM in order to perceive, understand, and regulate emotions. EGG-based emotion recognition is one of the popular methods in these regards. Li, et al. proposed a new method of using Rational Asymmetry (RASM) features of EEG and LSTM as classifier to recognized emotion [3]. Soleymani, et al building an LSTM model that outperformed conditional random fields, support vector regression (SVR) and multi-linear regression (MLR) [4], while Li, et al. combined convolutional neural network and LSTM and obtained a good performance [5]. On the other hand, Xing, et al. applied SAE to decompose source signal from collated EEG signal after which LSTM was employed as a classifier [6].

Despite the efforts in emotions recognition, emotion detection from nonstationary, EEG signals is still a challenging task as a sophisticated learning algorithm that can represent high-level abstraction is required. In this paper, we investigate SAE for feature dimensionality reduction and LSTM classifier for emotion recognition in order to reduce the complexity of features and obtain high accuracy.

II. METHODOLOGY

2.1 Framework

As shown in Fig. 1, our proposed model consists mainly of two sequential parts, including feature extraction and emotion classifier. The details of each part are given below. In the proposed model, SAE was used for feature dimensional reduction in order to reduce the dimension of the collected EEG signals. LSTM was the main component that was used in the emotion timing model to recognize emotion using the EEG feature sequence based on the EEG features produced by SAE.

2.2 EEG Data Acquisition

The emotion EEG dataset used for this study is SEED obtained from BCMI laboratory and
published in [7]. The emotion EEG data were collected from 15 subjects who were exposed to emotional films with real-life scenarios capable of triggering positive, neutral and negative emotions. Fig. 2 presents the experimental protocol for the data collection for 15 trials.

![Fig. 2 Protocol of the EEG experiment [7]](image)

2.3 Pre-processing and Feature Extraction

In Pre-processing stage, three operations were conducted, which include down-sampling, EOG/EMG removal and bandpass filtering. The EEG data was first down sampled to the rate of 200Hz. Then ElectroOculoGraphy (EOG)/ElectroMyoGraphy (EMG) contaminations were then manually checked and removed. The bandpass filter of 0.3Hz to 50Hz was applied to remove noise and artifacts.

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 bands (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

\[
h(x) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx \tag{1}
\]

where the time series \(x\) obeys the Gauss distribution \(N(\mu; \sigma^2)\).

In order to further remove the irrelevant components and obtain temporal dynamics of emotional state, Linear Dynamic System (LDS) was employed. These methods of Pre-processing and feature extraction were performed by the authors [7].

2.4 Feature Dimensional Reduction Based on Stack AutoEncoder

The purpose of this work is to determine an encoder that allows us to effectively reduce the dimension of EEG data from 62 features. SAE is chosen for this task because it consists of multiple layers that can achieve better performance. Fig. 3 shows the SAE structure and hyperparameters we designed for the SAE we designed. It consists of three encoding layers with 123 and 64 neurons for the first two layers while the numbers of neurons for the last layer were investigated to obtain the optimal reduction number to be encoded. It also consists of three encodings with 64, 128 and 62 neurons at each layer, respectively. ReLu activation function was employed at each layer, except for the last layer in which the sigmoid function was used. Adadelta optimizer was used in the experiment. The setup was attained after experiments that proved it to be efficient.

SAE has the same formula as standard autoencoder and is express as:

\[
H = WI \tag{2}
\]

where \(I, H, \text{and } W\) represent input data, hidden layer and transformation matrix of the encoder respectively.
We investigated various reduction numbers which include 12, 22 and 32. We begin with 12 as part of the numbers because studies have shown that EEG source signals came from 12 different functional brain regions based on previous research [5]. The result of the experiment is discussed in the next section.

2.5 Emotion Classifier Based on LSTM/RNN

RNN is a 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 data at each point in time. Similarly, we assumed emotions changes continuously which is reflected in the temporal correlations of EEG.

LSTM is a special type of RNN that 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. Fig. 4 presents the structure of a standard RNN and LSTM. The computation at each time step is defined as

\[
g^i = \sigma(W^h_{hi}h_{t-1} + I^i x_t)
g^f = \sigma(W^h_{hi}h_{t-1} + I^f x_t)
g^o = \sigma(W^h_{hi}h_{t-1} + I^o x_t)
g^c = \tanh(W^h_{hi}h_{t-1} + I^c x_t)
\]

where \( \sigma \) is the logistic sigmoid function. \( \odot \) represents element-wise 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 \( x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(m)} \), we learn a classifier to generate hypotheses 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. The hidden state for each timestep of a neuron is returned for every layer in the network. In the first layer, 128 hidden neurons and a dropout were applied while the subsequent 4 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. “Dropout” was added in the input layer of LSTM to avoid over-fitting[8-12].

III. EXPERIMENTAL RESULTS

The experiments were conducted using 7500 data, where 75% of the data was used for training while 25% was used for validation for both SAE and LSTM. Separate 1000 test data were used to evaluate the model. Several experiments were conducted to determine the configuration of the model. LSTM hyperparameters investigated are indicated in Table 1, while that of SAE configurations discussed in 2. 4 are maintained throughout the experiment.

Table 1. Model hyperparameters for investigation

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME_STEP</td>
<td>1</td>
</tr>
<tr>
<td>EPOCHS</td>
<td>45-55</td>
</tr>
<tr>
<td>BATCH_SIZE</td>
<td>2</td>
</tr>
<tr>
<td>DROP_OUT</td>
<td>0.158</td>
</tr>
<tr>
<td>Optimizer</td>
<td>AdaGrad</td>
</tr>
</tbody>
</table>

3.1 Performance Dimensional reduction with SAE

The SAE model training for all the number of features under consideration were conducted for 50 epochs to reduce the feature to less than 62. Table 2 gave the performance of the SAE for the features. The validation loss performed better than the training loss in all cases except for 32 features as graphically presented in Fig. 5. From the result, the SAE is able to encode and decode the data with very little loss of magnitude 10^-4.

3.2 Performance of LSTM/RNN based Emotion Recognition

In this section, the performance of the LSTM/RNN model for each feature reduction is presented. From the result of the training presented in Fig. 6, both training and validation accuracy for the cases progressively are improved during the training. Likewise, the loss for both training and validation are progressively reduced during the course of the process. However, the validation accuracy experience bumps for 12 and 22 features, while that of 32 features progressively gave a smooth curve. Based on these, 32 feature model is hereby considered to be more stable.

Table 2. SAE Model Training and Validation Loss

<table>
<thead>
<tr>
<th>No. features</th>
<th>Epoch</th>
<th>Training Loss</th>
<th>Validation Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>50</td>
<td>2.2177x10^-4</td>
<td>2.0994x10^-4</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>1.9756x10^-4</td>
<td>1.9161x10^-4</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>2.0085x10^-4</td>
<td>2.0139x10^-4</td>
</tr>
</tbody>
</table>

Fig. 5 SAE model loss performance.
Comparing the validation accuracy attained by the models, 12 features performed poorly with an accuracy of 69.4%, 22 features gave good performance of 84.9% and 32 features gave the best accuracy of 87.5%. The results are presented in Fig. 7.

Furthermore, separate 1000 testing data was used to evaluate the recognition rate of individual emotion under consideration, i.e. Sad, Neutral and Happy. Fig. 9, present the emotion recognition accuracy by comparing the dimensional reduction model with 12, 22, 32 features and without dimensional reduction (62 features). For sad emotion, 32 features gave the best result of 76.6% accuracy while 22 features performed poorly in recognizing sadness with an accuracy of 44.3%. Neutral emotion enjoys a relatively good recognition accuracy among the models with 62 features (with SAE reduction) recording the highest accuracy result of 91.4% while 12 features performed the
least with 82.2% accuracy. Happy emotion enjoyed the highest recognition accuracy of 99.7% with 62 features (with SAE reduction) while 22 features recorded the least accuracy of 78.3%. It clear that the models have difficulty in recognizing sadness and more effective in identifying happiness. Considering all the models with SAE dimensional reduction, 32 features outperformed all the models and thus is chosen for further evaluation.

Finally, we evaluate the overall accuracy of our best model (32 feature) against 62 features (with SAE reduction) and the work Xing, et al. [5] and Zheng and Lu [6] which uses Deep Believe Networks (DBN). Even though Xing, et al evaluated his model accuracy based on Valence and arousal, we consider the mean accuracy of their model because they employ SAE to decomposed source signal from the collected EEG. Other techniques that we compare against include Logistic Regression (LR), Random Forest (RF) and Decision Tree (DT). From the result presented in Fig. 10, our model gave an outstanding overall recognition accuracy of 86.6%, followed by that of Zheng and Lu model with 86.1% accuracy. Xing, et al model obtained a mean accuracy of 77.7% while DT gave the worst result of 64.1% recognition accuracy.

```
IV. CONCLUSION
```

In this paper, we investigated the used of SAE for dimensional reduction in emotion recognition. We obtained the best model with 32 features. Our model outperformed other models which include LSTM (62 features with SAE reduction) model, Xing, et al model, Zheng and Lu model, LR, RT and DT. For future work, four classes of emotion such as neutral, sad, fear and happy can be investigated.

```
Acknowledgement
This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2017R1D1A1B0303388).
```

```
References
```
723


