2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC) | 978-1-6654-0232-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICNISC54316.2021.00187

Stacked Auto-encoder Driven Automatic Feature Extraction for Web-enabled EEG Emotion Recognition

Yixiang Dai[1], Encai Ji[2] and Yong Yao[1, *]
[1]School of Electronics and Information Engineering, Harbin Institute of Technology (Shenzhen), Room 1312, Research building, Road Xiaqing, Shenzhen, China
[2]Shenzhen University, Road Xueyuan, Shenzhen, Shenzhen, China
yixiang.dai@foxmail.com, jec@tsinghua.edu.org, yaoyong@hit.edu.cn

***Abstract*—EEG emotion recognition is able to provide a scientific solution for emotional health assessment. Feature extraction is the fundamental procedure. Traditionally, the future set is generated by the existing theories or rules, which is not convincing and objective enough. Therefore, this paper proposes a data-driven automatic feature extraction methodology for web-enabled EEG emotion recognition based on 2-hidden-layer stacked auto-encoder. Since the web-enabled framework provides large scale of EEG data, emotion-related EEG features can be extracted directly from the time-domain raw wave, which is different from the typical feature extraction methods based on rules and experiences. With the optimal experimental parameters setting, the proposed method extracts typical time-domain distinguishable features from the EEG raw data and obtains relatively low classification error rate. This paper takes a step further towards automatic feature extraction for web-enabled EEG emotion recognition and make the entire framework more impersonal and convincing.**

Keywords-EEG emotion recognition; automatic feature extraction; stacked auto-encoder

I. INTRODUCTION

Nowadays, emotional health draws great public concern. Abnormal emotional states may be the premonition of serious mental diseases. It is important to generate a scientific emotion recognition scheme for emotional health monitoring in daily life. In typical, emotion was studied by psychologists and the research process contained various subjective factors. In the present age, researchers tried to enable lifeless computers or machines to observe and interpret human emotion feature quantitatively to finish the emotional state monitoring and classification tasks.

The basic process of emotion recognition is as following. Firstly, subjects receive emotional stimulations like sound, picture, movie, etc. Meanwhile, multimodal bio signals are recorded including electroencephalography (EEG), heart rate, voice behaviour, facial expression, etc. After that, feature extraction and pattern recognition algorithms are utilized to figure out the emotional labels corresponding to the specific data segments. The final assessment is based on two typical emotion models named discrete model and dimensional model. Discrete model defines several basic emotions and assumes that all the other emotions called complex emotions can be regarded as the combinations of basic emotions. But in this model, the definition of basic emotions is controversial. Dimensional model is more popular in which a valence-arousal plane is proposed to describe human emotions. Different emotions are located in this plane by calculating the arousal and valence values. In specific, valence values indicate the polarity of human emotions while arousal values indicate the level of excitement under different moods [1].

Electroencephalography (EEG) reflects potential changes caused by cognitive activities in human brain, which can avoid intentional disguise in emotion recognition by building the connection between EEG and emotions strictly. Therefore, with the development of sensing technique, EEG based emotion sensing is becoming one of the most popular methods. Chanel et.al selected pictures, recall and games as emotional stimulation sources in [2] [3] [4], respectively. After measuring subjects' EEG signals (also include other physiological signals), they extracted power spectrum density (PSD) features and introduced naive bayes (NB), support vector machine (SVM), and linear discriminant analysis (LDA) classification methods to recognize 3 emotional states based on dimensional model. Other research groups also developed their own emotion recognition frameworks with different alternative stimulations (e.g., sound, movie and video clip), extracted features (e.g., higher order crossings, common spatial pattern and self-organizing map) and classification methods (e.g., quadratic discriminant analysis (QDA) and deep auto-encoder) [5] [6]. In order to expand the application scenarios, some web-enabled EEG emotion recognition frameworks are proposed and help to accumulate big scale of EEG data [7]. However, the feature set for emotion recognition is usually generated by the existing theories or rules.

This paper proposes a data-driven automatic feature extraction methodology for web-enabled EEG emotion recognition based on 2-hidden-layer stacked auto-encoder. Since the web-enabled framework provides large scale of EEG data, emotion-related EEG features can be extracted directly from the time-domain raw wave, which is different from the typical frequency-domain feature extraction. With the optimal experimental parameters setting, the proposed method extracts typical time-

978-1-6654-0232-3/21/$31.00 ©2021 IEEE
DOI 10.1109/ICNISC54316.2021.00187

domain distinguishable features from the EEG raw data and obtains an acceptable recognition accuracy. The main contribution is to take a step towards automatic feature extraction for EEG emotion recognition.

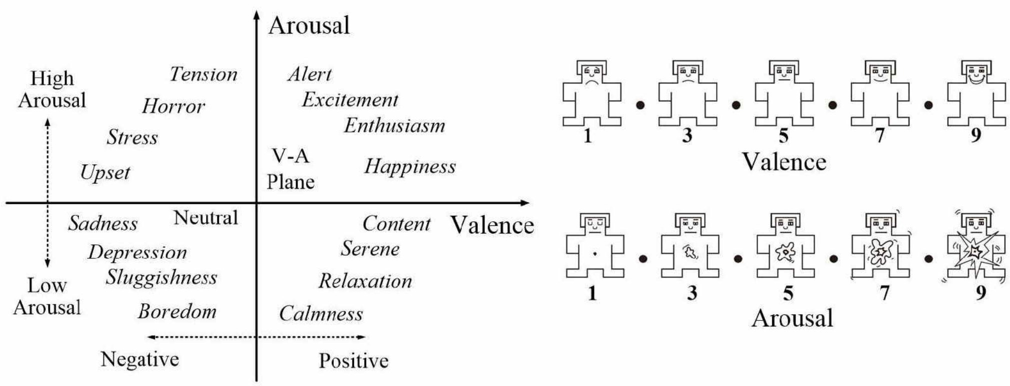
II. METHODOLOGY

A. Dimensional Model

Discrete model defines several basic emotions and regards other emotions as the combination of these basic emotions. However, it is ambiguous to determine the convincing basic emotions. Dimensional model, as shown in Figure 1(a), is more popular and the majority of related works choose this model.

In dimensional model, a hyperspace is proposed to describe human emotions. Typically, the hyperspace is simplified to a two-dimensional valence-arousal plane (sometimes also including other dimensions such as dominance and liking). Different emotions are located in this plane by calculating the valence and arousal values. In specific, valence values indicate the polarity of human emotions while arousal values indicate the level of excitement under different moods.

The emotion labels are determined by subjects themselves according to the SAM questionnaire as shown in Figure 1(b).



(a) (b)

Figure 1. Dimensional model for emotion recognition

B. Web-enabled EEG Emotion Recognition Framework

The data-driven automatic feature extraction method relies on large scale of emotion-related raw data. Therefore, it is necessary to establish a web-enabled emotion recognition structure. In the web-enabled scenarios, when subjects receive emotional stimulations and the EEG data are recorded, the web terminal devices help to deliver the corresponding data and the label information to the remote server. This structure can help to obtain emotion-related physiological effectively and provide more training data for the feature extraction and classification model. Moreover, the proposed stacked auto-encoder automatic feature extraction method requires abundant training data to guarantee an acceptable accuracy.

C. Stacked Auto-encoder Automatic Feature Extraction and Emotion Classification

This method mainly focuses on removing non-objective experiences and rules in feature extraction. It calculated uncertain feature coherences and representations of input signals by a hierarchical feature learning methods based on large data sets so that it can provide a data-driven solution for feature extraction and representing high-level abstraction.

The 2-hidden-layer stacked auto-encoder consisting of 2 sparse auto-encoders. The input is the obtained EEG raw data $\boldsymbol{X}$ in time domain. Select sigmoid function $f(x)=1/\left(1+e^{-x}\right)$ as the activation function. The process of encoding and decoding can be described as $\boldsymbol{H}=f(\boldsymbol{WX})$, $\hat{\boldsymbol{X}}=\boldsymbol{W}^T\boldsymbol{H}$ where $\hat{\boldsymbol{X}}$ denotes the output of the sparse auto-encoder. As an unsupervised learning scheme, sparse auto-encoder tends to utilize a back propagation scheme to approach an identity function as

$$\hat{\mathbf{X}}=h_{\mathbf{W}}(\mathbf{X})\approx\mathbf{X} \tag{1}$$

By adding the sparsity constraint (making neural nodes inactivated in most cases), sparse auto-encoder tries to learn the sparse representation of training set, reflecting the structure of features for classification. The cost function $J^{(A)}$ for training sparse auto-encoder is as following:

$$J^{(A)} = \|h_W(X) - X\|^2 + \lambda \sum_j |h_j| \quad (2)$$

One item is the 2-norm of the difference between the input and the output. The other item is the 1-norm of number of the activated nodes. In the proposed 2-hidden-layer stacked auto-encoder, the output of the first-layer sparse auto-encoder is regarded as the input of the second-layer auto-encoder. After finishing the pre-training process with the cost function talked above, the emotion labels are utilized to carry out a supervised training so that the weight parameters of the 2-hidden-layer stacked auto-encoder can be refreshed (Fine Tunning).

Softmax classifier is the expansion of logistic regression model in multi-class classification. And there are 4 classes of emotion labels according to the descriptions in 3.1. Therefore, it is indeed a problem of multi-class classification. In the softmax classifier, the probability of classify one sample $V^{(i)}$ into class j is

$$P\left(Y^{(i)} = j \mid V^{(i)}\right) = \frac{e^{\theta_j^T H^{(2)(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T H^{(2)(i)}}} \quad (3)$$

where $H^{(2)}$ denotes the output of the second layer. The softmax classifier can figure out the probability above and make the final classification decisions.

After extracting features by the stacked auto-encoder, the next problem is to reduce the redundant features. These redundant features not only decrease the recognition accuracy, but also increase the amount of calculation. A similarity function as following is utilized to evaluate the redundancy:

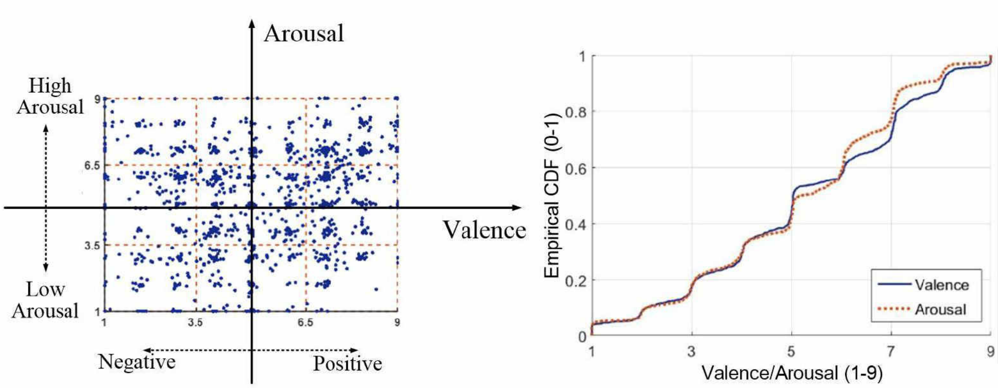
$$similarity(V^{(i)}, V^{(j)}) = \frac{V^{(i)} \cdot V^{(j)}}{\|V^{(i)}\| \cdot \|V^{(j)}\|} \quad (4)$$

When the similarity function values are above the threshold, $V^{(i)}, V^{(j)}$ are too similar to each other and one of them should be removed from the extracted feature set. With the algorithm shown above, the automatic feature extraction, selection and classification for emotion recognition can be realized in which no rules and experiences are utilized.

III. Experiments and Discussions

A. Experiment Setup

Emotion-related EEG data are recorded from 32 subjects to realize a subject-independent emotion recognition with the help of web-enabled sensing structure. Every subject provides 40 data segments with corresponding emotion labels in the V-A plane based on dimensional model. As shown in Figure 2(a), there are 4 emotion classes according to the 4 quadrants in the V-A plane. By analysing the distribution of emotion labels, Figure 2(b) reflects the fact that subjects tend to give valence/arousal scores near the integers.



(a) (b)

Figure 2. Distribution of EEG data segments with labels

B. Feature Extraction and Evaluation

Based on the methodology proposed in section 2.3, the feature number is fixed as 9, 10 and 16 separately, which means 9, 19 and 16 features are extracted from the large scale of EEG raw data respectively. After that, the similarity evaluation and feature selection are carried out 3 times and the corresponding results are displayed in Table 1, 2 and 3.

TABLE 1. 9 FEATURES AUTOMATIC EXTRACTION AND SELECTION

	Similarity (%)								
Feature	1	2	3	4	5	6	7	8	9
1	100								
2	2.7	100							
3	26.1	7.8	100						
4	**81.2**	5.2	71.5	100					
5	8.9	**98.9**	0.6	12.2	100				
6	1.1	1.0	1.3	2.9	1.2	100			
7	75.1	5.1	30.1	27.5	1.4	3.8	100		
8	26.5	1.6	**80.7**	27.9	0.9	7.5	57.4	100	
9	1.5	1.4	0.1	2.7	0.4	**98.9**	2.2	5.9	100

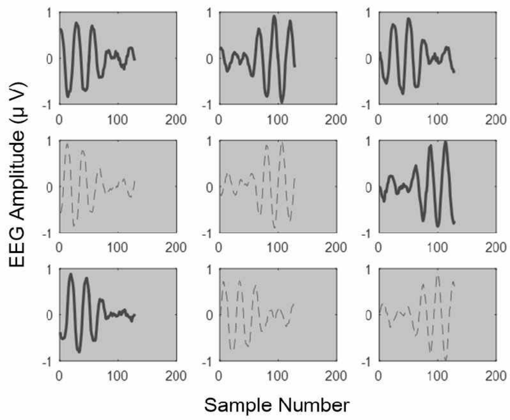
TABLE 2. 10 FEATURES AUTOMATIC EXTRACTION AND SELECTION

	Similarity (%)									
Feature	1	2	3	4	5	6	7	8	9	10
1	100									
2	6.5	100								
3	1.4	1.4	100							
4	0.2	2.8	**98.9**	100						
5	6.0	3.3	2.4	3.9	100					
6	8.6	7.0	3.3	4.9	**99.0**	100				
7	2.9	0.5	4.3	3.9	1.7	0.3	100			
8	**99.1**	8.4	2.5	1.8	3.1	5.8	1.0	100		
9	10.7	**99.3**	3.1	4.7	2.7	6.6	2.2	12.5	100	
10	1.2	0.4	2.8	2.7	2.9	1.5	69.2	0.6	2.8	100

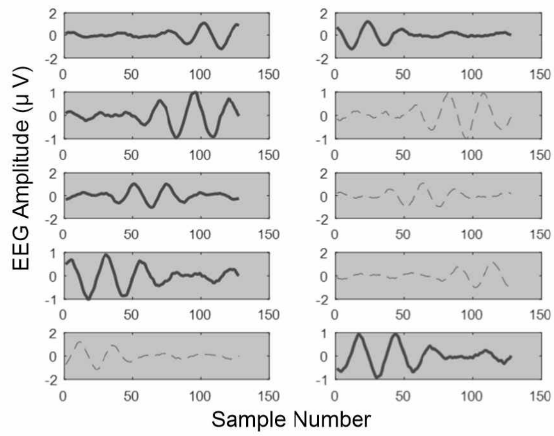
TABLE 3. 16 FEATURES AUTOMATIC EXTRACTION AND SELECTION

	Similarity (%)															
Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	100															
2	1.5	100														
3	27.2	0.9	100													
4	**82.1**	1.6	2.0	100												
5	10.0	11.0	**80.3**	4.5	100											
6	13.1	22.6	19.5	22.1	79.7	100										
7	69.0	2.2	8.7	28.7	0.9	13.3	100									
8	14.8	6.2	61.0	16.0	2.8	1.5	19.4	100								
9	9.2	22.4	3.0	2.6	1.5	28.4	3.2	7.3	100							
10	10.3	5.4	7.8	23.7	2.3	0.4	22.6	25.1	7.9	100						
11	19.9	1.2	13.6	58.4	4.1	0.8	11.8	3.0	5.0	**89.4**	100					
12	0.3	**76.6**	3.1	0.9	23.7	9.6	0.5	0.3	12.5	9.9	4.2	100				
13	27.1	7.5	27.5	13.4	0.8	11.7	9.5	**89.0**	0.4	50.0	24.7	8.9	100			
14	3.7	12.0	4.9	8.6	29.1	69.1	10.7	4.4	**82.6**	1.1	0.7	29.4	1.6	100		
15	0.0	24.0	3.0	11.8	78.0	32.1	5.0	10.8	14.9	4.8	10.6	67.7	18.2	9.2	100	
16	3.7	72.2	10.2	3.2	21.5	13.5	6.3	3.9	73.0	9.6	4.8	26.2	12.6	30.1	5.9	100

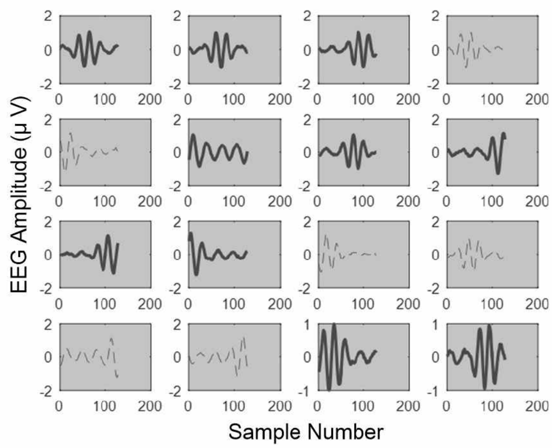
Set the similarity threshold as 70%. Table 1 represents that when 9 features are extracted, feature 4, 5, 8 and 9 are redundant. Therefore, select feature 1, 2, 3, 6, 7 as the input for softmax classifier (as shown in Figure 3(a)) to obtain an accuracy of 54%. Table 2 represents that when 10 features are extracted, feature 4, 6, 8 and 9 are redundant. Therefore, select feature 1, 2, 3, 5, 7, 10 as the input for softmax classifier (as shown in Figure 3(b)) to obtain an accuracy of 67%. Table 3 represents that when 9 features are extracted, feature 4, 5, 11, 12, 13 and 14 are redundant. Therefore, select feature 1, 2, 3, 6, 7, 8, 9, 10, 15, 16 as the input for softmax classifier (as shown in Figure 3(c)) to obtain an accuracy of 65%.



(a)



(b)



(c)

Figure 3. Extracted and selected features

C. Discussions

The automatic extraction method proves to be capable of calculating completely data-driven features. And these efficient features can be further selected by similarity evaluation. The proposed stacked auto-encoder driven automatic feature extraction differs from the existing frameworks for EEG emotion recognition in which features are always extracted in advance based on some rules or experiences. These 3 parameter settings, especially the latter two settings (feature number: 10 and 16), obtain an acceptable emotion recognition accuracy compared to the related works. In addition, the extracted data-driven features provide some describable time-domain phenomena for emotion-related science development. In the future, more data should be accumulated to generate more distinguishable features to improve the recognition accuracy.

IV. CONCLUSIONS

This paper focuses on a data-driven automatic feature extraction methodology for web-enabled EEG emotion recognition based on stacked auto-encoder. With the large scale of EEG data provided by the web-enabled framework, emotion-related EEG features can be extracted directly from the time-domain raw wave. The proposed stacked auto-encoder driven automatic feature extraction differs from the existing frameworks for EEG emotion recognition in which features are always extracted in advance based on some rules or experiences. The main contribution is to take a step further towards automatic feature extraction for EEG emotion recognition and make the entire framework more objective and convincing.

ACKNOWLEDGEMENT

This paper is supported by National Natural Science Foundation of China under Grant #61905150.

REFERENCES

[1] Jatupaiboon N, Pan-ngum S and Israsena P J 2013 *The Scientific World Journal* Article-ID 618649

[2] Chanel G, Kronegg J, Grandjean D and Pun T J 2006 *Multimedia content representation, classification and security* (Berlin Heidelberg: Springer) p 530-537

[3] Chanel G, Kierkels J, J Soleymani M and Pun, T J 2009 *International Journal of Human-Computer Studies* **67** 8

[4] Chanel G, Rebetez C, B´etrancourt M and Pun T J 2011 *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* **41** 6

[5] Xu H and Plataniotis K P 2012 *IEEE 14th International Workshop on Multimedia Signal Processing* (Banff: Canada IEEE) p 299-304

[6] Zhang H J 2020 *IEEE Access* **8** p 164130-164143

[7] Dai Y, Wang X and Li X J 2017 *Measurement* **109C** p 408-424