Journal of Social Sciences and Humanities
Vol. 6, No. 4, 2020, pp. 419-422
http://www.aiscience.org/journal/jssh
ISSN: 2381-7763 (Print); ISSN: 2381-7771 (Online)

Construction of Emotion Corpus Based on Electroencephalogram Data

Qin Liu*

School of Foreign Language, Yancheng Teachers University, Yancheng, China

Abstract

This paper takes the emotional corpus based on Electroencephalogram (EEG) data as the research object, and take the accurate analysis of emotional state and the construction of emotional database as the goal. In recent years, more and more attention has been paid to fine-grained sentiment analysis, mainly including opinion holders, and the extraction of opinion objects. With the development of wearable perception, big data and machine learning technology, it brings opportunities for the research and solution of this problem. By collecting physiological signals and other data, we can better analyze the emotional changes of people in different states. We focus on the relationship between EEG data and emotional state, the construction of emotional corpus in small samples and the construction of emotional corpus in big data environment, so as to provide effective support for the construction of emotional corpus. In this paper, we take the accurate analysis of emotional state and the construction of emotional database as the goal. We study the relationship between EEG data and emotional state, the construction of emotional corpus in small samples and the construction of emotional corpus in the environment of big data, so as to provide effective theoretical support for the construction of emotional corpus.

Keywords

Emotion Corpus, Electroencephalogram Data, Small Samples

Received: September 30, 2020 / Accepted: November 25, 2020 / Published online: December 11, 2020

1. Introduction

Currently, emotional corpus has become a hot research field in corpus construction. Its goal is to construct the corresponding relationship between people's emotional state and pragmatic characteristics, so as to analyze the various emotional states implied in utterances, opinions and comments in different situations. Sentiment analysis technology can be divided into two categories: one is based on machine learning, which uses statistical machine learning algorithm to extract features from a large number of annotated and unlabeled subjective corpus, and then carries out text sentiment analysis. The other is based on emotion dictionary, which can analyze text sentiment in different granularity according to the emotional orientation of words provided by emotion dictionary. So far, the construction and tagging of emotional corpora are characterized by multi genre, multi level, multi theory and multi methodology.

Early researchers are mainly engaged in emotion classification. At present, most emotional analysis mainly relies on word bag method and semantic polarity method to divide emotional documents. However, these methods ignore the following important contents: first, emotion classification needs to make a more detailed semantic analysis of attitude expressions with perfect attitude categories and other semantic attributes; second, the atomic structure of this expression is not a single word, but an evaluation block composed of coherent word groups. Therefore, in recent years, more and more attention has been paid to fine-grained sentiment analysis, mainly including opinion holders, and the extraction of opinion objects. With the transplantation of

* Corresponding author
E-mail address: classforyc@163.com

systemic functional linguistics to the field of computer emotion research, corpus based affective research has gradually sprung up. Some affective corpora mainly include Berardinelli's film review corpus, Wiebe's news review corpus and Junko Minato's emotional spoken corpus.

At present, most of emotion discrimination is based on scale, and the biggest challenge is that it is difficult to accurately identify the emotional performance of the assessed. With the development of wearable perception, big data and machine learning technology, it brings opportunities for the research and solution of this problem. By collecting physiological signals and other data, we can better analyze the emotional changes of people in different states. Among them, Electroencephalogram (EEG) reflects the spontaneous and rhythmic electrophysiological activities of neurons in the cerebral cortex, which can accurately determine the state of brain nerve changes. This method has become the focus of clinical brain nerve diagnosis To detect the means.

Based on the dynamic EEG monitoring and data analysis of the subjects, this paper proposes a correlation model between EEG data and emotional state, constructs EEG data processing model for small samples and EEG data processing model for big data, and realizes the construction of emotion corpus for EEG data.

2. Method

This paper includes three levels of research content, namely wearable EEG data acquisition and its correlation with emotional state modeling, emotional corpus construction in small samples and emotional corpus construction in big data environment.

2.1. Correlation Modeling Between EEG Data and Emotional State

The focus of its research is to collect EEG data of subjects by wearable EEG instrument, which is the basis of constructing emotion database. Because the EEG signal characteristics of people are often intrinsically related to their mental state. For example, when people think about problems in negative emotions, the amplitude of error-related negative potential in EEG will be significantly reduced, the latency of event-related potential P300 will be significantly prolonged, and the amplitude will also be reduced, especially in the frontal lobe. Based on the inherent relationship between EEG data and mental state, if we combine the specific characteristics of multiple leads of EEG data, we can accurately judge the mental state of the subjects. For correlation EEG data, if only a single channel feature is examined for classification, it is not of great significance for emotion type recognition. Therefore, it is necessary to quantitatively analyze the correlation of data sets, and fully consider the correlation of EEG data in feature extraction and classification.

In the research process, the actual EEG data belongs to multi-dimensional (multi-channel) non-stationary random signals, and the channels are nonlinear correlated, which brings great challenges to the modeling and solving of EEG data correlation. Therefore, the correlation analysis of high latitude nonlinear data sets is a difficulty in wearable EEG data analysis, and it is a necessary prerequisite to improve the accuracy of emotion type recognition results.

Therefore, the hidden Markov model can be used to estimate the source correlation. Each group of EEG data in discrete sampling can be regarded as a state, each state can produce multiple outputs according to the probability distribution of its characteristics, and the possible related states are regarded as hidden states. According to the observation sequence, the joint probability density of the source is calculated to maximize the probability of the occurrence of the observation sequence.

2.2. Construction of Emotional Corpus Based on Small Sample

The focus of its research is to accurately identify different emotional states by combining the relevance of EEG data in the case of small sample size, and at the same time corresponding to the language and pragmatic situation of the subjects, and combined with the current corpus. Machine learning is used to classify emotional states. The specificity of EEG data was evaluated by constructing appropriate classification features and objective functions.

In the process of research, EEG data acquisition is often accompanied with noise due to the limitation of the cooperation of the tested and the accuracy of the equipment. Therefore, the classification prediction method needs to meet the following requirements: 1) It can effectively deal with high-dimensional nonlinear correlation data; 2) The local optimal solution is the global optimal solution; 3) In the case of small error in the limited training set samples, the independent test set still has small error; 4) The structural risk is minimized, and the model structure and parameters are easy to be optimized.

Therefore, the proposed mam + incremental learning method can be used. Firstly, the multi-channel EEG data is modeled as mam model, and then the normalized local weighted linear regression method is used to solve the model coefficient matrix, and then the coefficient characteristic tensor is obtained. Finally, the model coefficients of training data and test data are projected onto the projection matrix to realize dimension reduction. In the processing of the sample set, the original training set is divided into a small batch of sample sets. After the training, the model coefficient after the training is saved and called in the next training to complete the incremental learning until all the samples are trained, and the final classifier model is formed.

2.3. Construction of Emotional Corpus Based on Big Data

The focus of its research is to use deep learning model to design an efficient recognition network for large amount of data and similar features to achieve the generalization ability of classification. Because the amount of data collected by wearable EEG equipment is not large enough and the data standard is difficult, how to generate enough EEG data and label unlabeled data accurately through a small amount of labeled data is the basis of training deep learning model. On the other hand, due to the lack of obvious distinction between different emotional states, and a small number of deep learning models only consider the temporal and spatial characteristics of EEG data, and completely ignore the spatial and temporal characteristics of EEG data, which makes the classification effect of deep learning models not ideal. What's more, according to the actual test, the deep learning model is prone to over fitting phenomenon, which mainly shows that the effect on the training set is very good, but the effect on the test set is poor. This is because the model takes part of the characteristics of the training samples as the common characteristics of all samples, resulting in the poor generalization ability of the model.

In the process of research, due to the characteristics of large noise, similar characteristics, correlation between channels and non-linear separability, it is necessary to build a suitable generation model to achieve data reconstruction. Secondly, the EEG data not only contains the spatial information represented by the electrode position, but also maintains the inherent time information of the acquisition sequence. The former belongs to multi-channel feature, while the latter belongs to single channel feature. Therefore, it is necessary to design a classification framework based on spatiotemporal fusion to make full use of the information of EEG data. Finally, because EEG big data is not linearly separable, it is necessary to use nonlinear mapping to map the linear inseparable problem in the original feature space to the linearly separable problem in the new feature space, so that the model learning has more distinguishing features.

Therefore, a semi supervised generation network can be designed. Firstly, semi supervised coding is designed and trained by semi supervised training method, so that EEG signal features can be extracted directly, and the data label is generated by converging network. Then, a spatiotemporal information fusion network is designed, in which the single channel time-frequency map of the EEG data sequence with a single electrode is used as the input, while the multi-channel generation model takes the brain topographic map with inherent relative spatial position between multiple electrodes as a multi-channel input. After training, the feature representation of the task related part of the feature vector can be obtained, which can be used as the input of the full connection layer.

3. Conclusion

In this paper, we take the emotional corpus based on EEG data as the research object, and take the accurate analysis of emotional state and the construction of emotional database as the goal. We study the relationship between EEG data and emotional state, the construction of emotional corpus in small samples and the construction of emotional corpus in the environment of big data, so as to provide effective theoretical support for the construction of emotional corpus.

Acknowledgements

This work is supported by Philosophy and Social Science Research Project of Universities in Jiangsu Province (No. 2019SJA1718), Top-notch Academic Programs Project of Jiangsu Higher Education Institutions, Reform of Education and teaching in Yancheng Normal University (No. 2018YCTUJGY047).

References

[1] Kreuzer M, Kochs E F, Pilge S, et al. Construction of the Electroencephalogram Player: A Device to Present Electroencephalogram Data to Electroencephalogram-Based Anesthesia Monitors [J]. Anesthesia & Analgesia, 2007, 104 (1): 135-9.

[2] Hou H , Zhang X , Meng Q . Olfactory EEG Signal Classification Using a Trapezoid Difference-Based Electrode Sequence Hashing Approach [J]. International Journal of Neural Systems, 2020.

[3] Thammasan N, Fukui K I, Numao M. Application of Annotation Smoothing for Subject-Independent Emotion Recognition Based on Electroencephalogram [J]. 2016.

[4] Ohori R, Shinkai D, Nagai Y, et al. Construction of a Model for Discriminating between Electroencephalographic Patterns at the Time of Incorrect Inputs Based on Sensitivity Spectrum Analysis [C]// Symposium on Human Interface. Springer, Berlin, Heidelberg, 2011.

[5] Thammasan N, Moriyama K, Fukui K I, et al. Familiarity effects in EEG-based emotion recognition [J]. Brain Informatics, 2017, 4 (1).

[6] Bum-Jin P, Feng L I, Weon-Eui K. Analysis of HUD Information Based on Driving Simulator [J]. Transportation Standardization, 2012.

[7] Trofimov A G, Skrugin V I, Rodriguez A M H. Extraction and recognition of electroencephalogram dynamic patterns for brain-computer interfaces [C]// Informatica. IEEE, 2013.

[8] Vasuki P , Sambavi B , Joe V . Construction and Evaluation of Tamil Speech Emotion Corpus [J]. National Academy ence Letters, 2020(4).

[9] Sharma R , Pachori R B , Sircar P . Automated Emotion Recognition based on Higher Order Statistics and Deep Learning Algorithm [J]. Biomedical Signal Processing and Control, 2020.

[10] Halim Z , Rehan M . On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning[J]. Information Fusion, 2020, 53:66-79.

[11] Anvesh, Jackson, Udaya, et al. EEG changes in patients on antipsychotic therapy: A systematic review. [J]. Epilepsy & behavior : E&B, 2019.

[12] Hall T H , Ross A A G . Rethinking Affective Experience and Popular Emotion: World War I and the Construction of Group Emotion in International Relations [J]. Political Psychology, 2019, 40(1).

[13] Wang J , Run Y , Shi H . Emotional state representation and detection method of users in library space based on body posture recognition [J]. Digital Library Perspectives, 2020, 36(2):113-125.

[14] Xing X , Li H , Li J , et al. A multicomponent and neurophysiological intervention for the emotional and mental states of high-altitude construction workers [J]. Automation in construction, 2019, 105(SEP.):102836.1-102836.12.

[15] Miranda I M , Aranha C , Ladeira M . Classification of EEG Signals using Genetic Programming for Feature Construction [C]// the Genetic and Evolutionary Computation Conference. 2019.