Recognizing emotions based on multimodal neurophysiological signals

Journal Item

* Corresponding Author: dawei.song@open.ac.uk

Recognizing emotions based on multimodal neurophysiological signals

Xiang Li[1], Peng Zhang[1], Dawei Song[1,2]*, Yuexian Hou[1]

[1]Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin, China
[2] Department of Computing and Communications, The Open University, Milton Keynes, United Kingdom

In the era of big data, the ability to computationally assess human emotions based on neurophysiological signals, such as brain signals, respiration, and heart rate, is now a possibility. Emotion, sometimes referred to as affect or mood, is an internal experience sometimes caused by external events. Emotion manifests in expressions such as joy, grief, fright, anger, sympathy, and disappointment.

Emotion recognition is a hot topic in cognitive neuroscience and psychophysiology, and has become increasingly relevant to computing and information sciences. Recent neurological studies have emphasized the role of emotion in social interactions, cognition, and rational decision making (*1, 2*). Within the field of artificial intelligence (AI), a new interdisciplinary area called "affective computing" (AC) has emerged. The goal of AC is to empower computer systems to recognize, comprehend, and respond appropriately to human emotions for the sake of natural human–computer interactions (HCIs) (*3*).

More practically, emotion recognition could help in detecting mood-related mental health problems such as depression, which is highly linked to suicide (*4*). The World Health Organization (WHO) estimates that by 2020, major depression will be the second leading cause of disability in the world, just behind ischemic heart disease (*5*). Although physiological and neuroimaging data have become increasingly available, psychiatrists urgently need better tools to utilize the information for diagnosing and prognosing depression in its earliest stages, when there is the highest potential for effective treatment. To address this, machine learning techniques, which can learn from and make predictions using data, may be key for mining reliable diagnostic and prognostic information from these data (*6, 7*).

In the field of education, the Massive Open Online Courses (MOOC) platform has been widely adopted around the globe. However, a major challenge has been how effectively the teachers or the platform can track the level at which a student is paying attention. Neuroscience and psychology research has shown that cognitive processes, such as attention and long-term memorization, are tightly linked to emotions (*1*). If a student gazes at the computer screen at length without displaying signs of interactive pleasure, it is likely that a mood of antipathy has emerged, which can result in low learning efficiency. AC can provide the ability to assess the emotional states of learners in an online environment based on analysis of signals from various sensory apparatuses, and teaching strategies can be automatically adapted to accommodate transitions in emotion, as they occur (*8*).

AC has been adopted for promoting the information retrieval (IR) experience, which underlies applications such as web search engines. Researchers have observed that positive emotions often reflect a user's satisfaction with search results (*9*) and interest in certain information (*10*). Negative emotions, on the other hand, occur more often with a user's dissatisfaction with search results and search strategies (*9*). Capturing the emotional states of users while they are seeking information through search engines can be used as feedback to adjust the search engines' retrieval strategies and to better predict the topical relevance of information being presented to users (*11, 12*).

Most emotion recognition approaches are based on the James-Lange theory, which claims there is a strong correlation between emotions and physiological arousal (*13*). However, to date most studies have concentrated on detecting emotions from a single modality of sensory data. Now, with recent advances in sensing technologies, synchronized detection of neurophysiological responses from different modalities can be acquired. These include measurements of temperature, respiration, electrical conductance of the skin, and electrical activity of the brain and skeletal muscles. Integrating these measurements with advanced machine learning techniques opens up opportunities to develop effective methods for recognizing human emotion.

In previous research, emotion recognition based on multimodal data typically concentrated on the feature-fusion method (which combines features from multiple modalities together directly) or decision-fusion method (which uses a majority vote or a weighted sum of decisions from multiple classifiers for each data modality) (*14, 15*). In these studies, the correlations across different neuro-physiological data modalities have not been effectively exploited. Our published work has shown that the idea of a multimodal deep learning approach is applicable and effective in recognizing emotional states from multiple channels of neurophysiological signals (*16*). In this article, we describe a multimodal fusion framework whereby deep learning techniques can be used to acquire representations across different data modalities, as opposed to just one, and to classify the emotional state of subjects. Specifically, we apply and evaluate our method on a widely used and publicly available benchmarking dataset, namely "A Dataset for Emotion Analysis using Physiological Signals," or DEAP (*17*).

Multimodal deep learning framework

Deep learning is a popular research branch of machine learning, which is inspired by progress in neuroscience and based on studies on information processing and the communication mechanisms of the underlying neural systems. Its hierarchical neural network-based learning architecture encompasses several layers of representations or "nodes," where the values of higher-level nodes are defined based on lower-level ones. Multimodal deep learning aims to gain joint features at an abstract level, by utilizing several pathways of deep learning from correlated data modalities. More recently, deep learning techniques have been successfully applied to feature learning in various pattern recognition tasks (*18*) and multimodal learning applications (*19, 20*).

In our work, we adopted multiple pathways of deep belief networks (DBNs), each of which is built for one data modality. A DBN is a type of deep learning model composed of several stacked restricted Boltzmann machines (RBMs), which are artificial neural networks that consist of two layers of input and output nodes. By stacking several RBMs together, a DBN is able to approximate any mathematical function for data transformation and representation. Usually, the lower RBM's output is regarded as the input of the upper one and it does not allow connections between nodes in the same layer. Each connection is assigned a weight parameter, and its value is decided and adjusted by several rounds of model training by utilizing sample data that is fed through the input layer.

Our multimodal deep learning framework is shown in Figure 1. Each DBN consists of two stacked RBMs, and v1 represents the input layer into which manually extracted features of electroencephalograms (EEGs) and peripheral physiological signals were fed. Further, h1 and h2 are output layers that can be used to extract higher-level abstract representations. Then a discriminative RBM (DRBM), a special kind of RBM acting as a classifier, is put over the combination of h2-layer representations for learning a shared representation in h3 and fulfilling emotion recognition tasks by the label layer, where each node represents one emotional state.

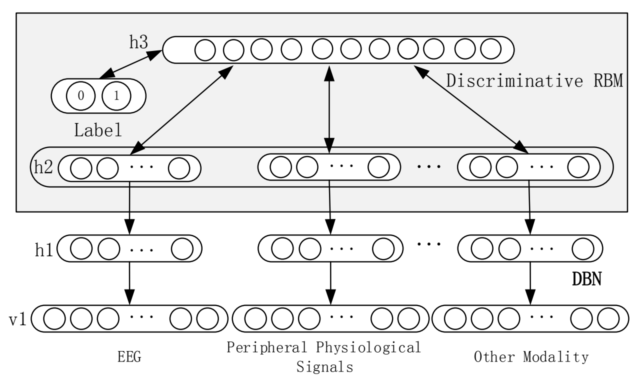

Figure 1. Deep Learning Framework for Multimodal Data.

Empirical evaluation

The framework we describe has been validated on the publicly available DEAP dataset, which was collected through an experimental paradigm designed to elicit different emotions from the subjects through exposure to specifically selected music videos. The subjects' EEG signals, as well as other types of peripheral physiological signals, were continuously recorded while the music video was being viewed. Examples of these peripheral signals include the electrical activity of muscles [electromyogram (EMG)] and eye movement [electrooculogram (EOG)], galvanic skin response (GSR), respiration, and skin temperature. DEAP also provides data from subjects' self-assessment of their emotions, rated according to Russell's valence-arousal scale (*21*). By utilizing these rating values, we divided the samples into two emotional classes (i.e., positive and negative emotional states) as most studies have done for validating emotion recognition performance (17).

Feature extraction is important to machine learning tasks. In addition to the widely used linear features, we opted to extract a variety of nonlinear features (e.g., correlation dimension and Shannon entropy) from neurophysiological signals. We compared the recognition performance of our multimodal deep learning method with a commonly used traditional classification approach. We first applied the k-Nearest Neighbor (kNN) classifier to the features of EEGs. We then applied the same classifier to the combined features of peripheral physiological signals, and to the combined features of EEG and peripheral physiological signals. Through experimental comparisons, we have found that the use of the deep learning approach shown in Figure 1 to analyze multimodal neurophysiological signals can achieve better recognition of emotions than the traditional kNN classifier based on the use of different modalities individually or in combination.

In this article we have summarized how it is possible to recognize emotion using a deep learning approach based on multiple neurophysiological signals. In the future, emotion-related modalities not limited to neurophysiological signals (e.g., speech and functional magnetic resonance imaging) could also be added into the framework as shown in Figure 1. Going forward, it will be interesting to investigate the kind of neurophysiological signals that most strongly contribute to emotion recognition.

References

1. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain.* (Random House, New York, 2008).
2. R. W. Picard, *Proceedings of the SPIE—The International Society for Optical Engineering*, **4299**, 518 (2001).
3. R. W. Picard, *Int. J. Hum. Comput. Stud.* **59**, 55 (2003).
4. S. Chehil, S. P. Kutcher, *Suicide Risk Management: A Manual for Health Professionals.* (Wiley, Chichester, 2012).
5. World Health Organization, *Mental Health: A Call for Action by World Health Ministers.* (World Health Organization, Geneva, Switzerland, 2001).
6. B. Hosseinifard, M. H. Moradi, R. Rostami, *Comput. Methods Programs Biomed.* **109**, 339 (2013).
7. S. Klöppel *et al*., *Neuroimage* **61**, 457 (2012).
8. S. Duo, L. X. Song, *Phys. Procedia.* **24**, 1893 (2012).
9. C. Tenopir *et al*., *Inf. Process. Manag.* **44**, 105 (2008).

10. I. Lopatovska, C. Cool, Paper presented at the ALISE Annual Conference, Phil adelphia, PA, January 8–12, 2008.

11. I. Arapakis, I. Konstas, J. M. Jose, *Proceedings of the 17th ACM International Conference on Multimedia* (ACM, Beijing, 2009), pp.461–470.

12. A. Yazdani, J. S. Lee, T. Ebrahimi, *Proceedings of the 1st SIGMM Workshop on Social Media* (SIGMM, Beijing, 2009), pp. 81–88.

13. C. G. Lange, W. James, *The Emotions* (Hafner, New York, 1967).

14. L. Kessous, G. Castellano, G. Caridakis, *J. Multimodal User In.* **3**, 33 (2010).

15. M. K. Abadi, J. Staiano, A. Cappelletti, *et al., Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (IEEE, Geneva, Switzerland, 2013), pp. 411–416 .

16. X. Li *et al*., paper presented at the SIGIR Workshop/Tutorial on Neuro-Physiological Methods in IR Research, Santiago, Chile, August 13, 2015.

17. S. Koelstra *et al*., *IEEE Trans. Affect. Comput.* **3**, 18 (2012).

18. G. E. Hinton, R. R. Salakhutdinov, *Science* **313**, 504 (2006).

19. J. Ngiam, A. Khosla, M. Kim *et al., Proceedings of the 28th International Conference on Machine Learning* (ICML-11) (ICML, Bellevue, WA, 2011), pp.689–696

20. N. Srivastava, E. Mansimov, R. Salakhutdinov, paper presented at the International Conference on Machine Learning Workshop, Edinburgh, Scotland, June 26–July 1, 2012.

21. J. A. Russell, *J. Pers. Soc. Psychol.* **39**, 1161 (1980).

Acknowledgments

This work was funded in part by the Chinese National Program on Key Basic Research Projects (973 Program) (2014CB744604 and 2013CB329304), the Chinese 863 Program (2015AA015403), and the National Natural Science Foundation of China (61272265 and 61402324).