Computers in Industry 92–93 (2017) 84–90

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Respiration-based emotion recognition with deep learning

Qiang Zhang[a,b], Xianxiang Chen[a], Qingyuan Zhan[c], Ting Yang[c], Shanhong Xia[a,*]

[a] Institute of Electronics, Chinese Academy of Sciences, China
[b] University of Chinese Academy of Sciences, China
[c] China-Japan Friendship Hospital, China

ARTICLE INFO

Article history:
Received 2 December 2016
Accepted 24 April 2017
Available online 3 July 2017

Keywords:
Emotion recognition
Deep learning
Wearable computing
Respiration
Arousal-valence theory

ABSTRACT

Different physiological signals are of different origins and may describe different functions of the human body. This paper studied respiration (RSP) signals alone to figure out its ability in detecting psychological activity. A deep learning framework is proposed to extract and recognize emotional information of respiration. An arousal-valence theory helps recognize emotions by mapping emotions into a two-dimension space. The deep learning framework includes a sparse auto-encoder (SAE) to extract emotion-related features, and two logistic regression with one for arousal classification and the other for valence classification. For the development of this work an international database for emotion classification known as Dataset for Emotion Analysis using Physiological signals (DEAP) is adopted for model establishment. To further evaluate the proposed method on other people, after model establishment, we used the affection database established by Augsburg University in Germany. The accuracies for valence and arousal classification on DEAP are 73.06% and 80.78% respectively, and the mean accuracy on Augsburg dataset is 80.22%. This study demonstrates the potential to use respiration collected from wearable deices to recognize human emotions.

1. Introduction

With great advance of artificial intelligence, it is promising to conduct affective computing with physiological signals. Many physiological signals are detected based on wearable devices, such as electrocardiogram (ECG), electroencephalography (EEG), electromyogram (EMG), blood volume pressure (BVP), galvanic skin response (GSR), temperature (TEMP), respiration pattern (RSP), photoplethysmogram (PPG). There is increasing evidence that these signals contain information related to human emotions [1–6]. Emotion recognition based on physiological signals is promising because these signals are involuntary manifestation of human body and people cannot control them intentionally. Moreover, continuous emotion assessments can be obtained through measurements of physiological signals.

Human emotions can be affected by many factors [7,8], and different emotions usually have fuzzy boundaries. Recent studies developed different kinds of emotion recognition models and tested them on their own dataset. In 2001, Professor Picard applied artificial intelligence to recognize human emotional states given physiological signals [9]. They extracted statistical time and frequency values and achieved 81% recognition accuracy on eight emotional classes. After that, more complicated features have been extracted. Duan et al. proposed differential entropy to represent EEG feature related to emotional states and achieved average accuracy of 81.17% [10]. Giakoumis et al. [11] introduced the Legendre and Krawtchouk moments to extract biosignal features. Yannakakis and Hallam used the approximate entropy feature [12] and preference learning [13]. Lin et al. applied machine learning algorithms to categorize EEG signals and obtained an average classification accuracy of 82.29% for four emotions [14]. Wang et al. systematically compared three kinds of EEG features (power spectrum feature, wavelet feature and nonlinear dynamical feature) for emotion classification [15].

Although different features have been tried to describe emotion-related characteristics of physiological signals, manual feature extraction inherits some primary limitations. First of all, performance of hand-crafted feature largely depends on the signal type and human experience. Poor domain knowledge may lead to an inappropriate feature that cannot capture the characteristics of certain signals. Second, there is no general guarantee that any feature selection algorithms will end to the optimal feature set. Third, the most manual features are statistical and can't depict signal details, which means a loss of information.

* Corresponding author.
E-mail address: shxia@mail.ie.ac.cn (S. Xia).

http://dx.doi.org/10.1016/j.compind.2017.04.005
0166-3615/

Distinctively, deep learning can automatically derive features from the raw signals, as opposed to manually pre-designed statistical features. Deep learning allows automatically feature selection and bypasses the computational cost of feature selection. Recently deep learning methods have been tried to process physiological signal like EEG and skin resistance, and achieved comparable results in comparison with other conventional methods [16–18]. In 2013, Martinez et al. introduced Convolutional Neural Network (CNN) to establish physiological models of affect [16]. To the best knowledge of the authors, this is the first attempt to use deep learning for computational modeling of affect. Since that, some studies on deep emotion recognition have been published [19–21]. For example, Zheng trained a Deep Belief Network (DBN) to classify two emotional categories (high and low valence) from EEG data, and Jirayucharoensak implemented a sparse auto-encoder whose input features are from 32-channel EEG signals [24]. To detect sleep stage, Martin et al. compared the manual features and a model combining DBN and hidden Markov model (HMM) [25].

After choosing deep learning as the feature extraction method, we move forward to talk about physiological signals. Different physiological signals are of different origins and may describe different functions of human body. For instance, the ECG and BVP relate to the cardiovascular system, while the EMG describes electrical activities of muscle. It is important to investigate the dynamics of every signal in order to clearly figure out its feasibility and limitation in assessing psychological activity. To this end, this work investigates RSP signals alone.

Respiratory pattern contains rich information about emotional states. Respiration velocity and depth usually varies with human emotion. For example, deep and fast breathing shows excitement that is accompanied by happy, angry or afraid emotion; shallow and fast breathing shows tension; relaxed people often have deep and slow breathing; shallow and slow breathing shows a calm or negative state. When in calm states, people usually breathe about 20 times per minute while in excitement, people breathe 40 to 50 times per minute. From RSP data, we selected four segments corresponding four kinds of emotions, as shown in Fig. 1.

As the RSP signal contains a wealth of emotional information, and can be easily detected in wearable devices, therefore in this paper, we focus on emotion recognition via respiration signals. To help recognize emotions, we used the Russel's Circumplex theory of emotion [26]. Specifically, each emotion is seen as a linear combination of two affective dimensions: arousal and valence. Fig. 2 shows the general architecture of the deep learning framework. We used a deep sparse auto-encoder (SAE) to extract hidden features of RSP. Two logistic regression categorize the features, with one for the arousal classification, and the other for the valence classification.

To validate the efficacy of the SAE-based approach, an emotion classification experiment was carried out using the DEAP database, which is the largest, most comprehensive physiological signal-emotion dataset publicly available to date. To further evaluate the proposed method on other people, after model establishment, we used the affection database established by Augsburg University in Germany. The paper is organized as follows: Section 2 introduces the arousal-valence theory, Section 3 describes the deep learning framework that consists of sparse auto-encoder and logistic regression, Experiment data, setting and results are presented in Section 4, and discussion and conclusion are shown in Sections 5 and 6, respectively.

2. Arousal-valence emotion theory

In this study, we used the Russel's Circumplex theory to help emotion recognition. This theory indicates that emotional states are distributed in a two-dimensional circular space, with arousal and valence dimensions [26]. Arousal is the vertical axis and

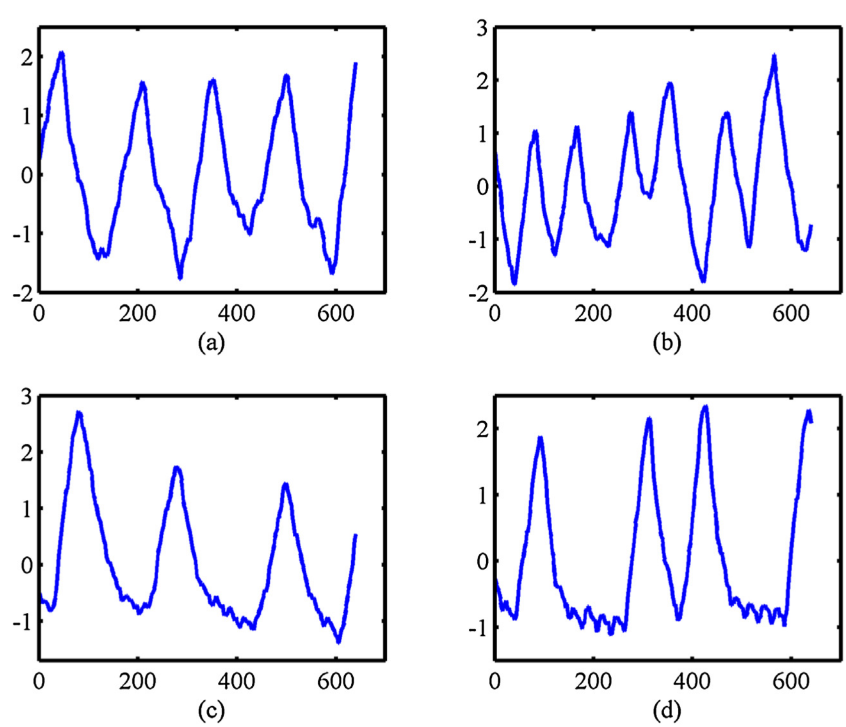

Fig. 1. Four 20-s respiration signal segments under different emotional states: (a) high valence and high arousal, (b) low valence and high arousal, (c) low valence and low arousal, (d) high valence and low arousal. The horizontal axis represents the sampling points while the vertical axis is the locally normalized magnitude of respiration signals.

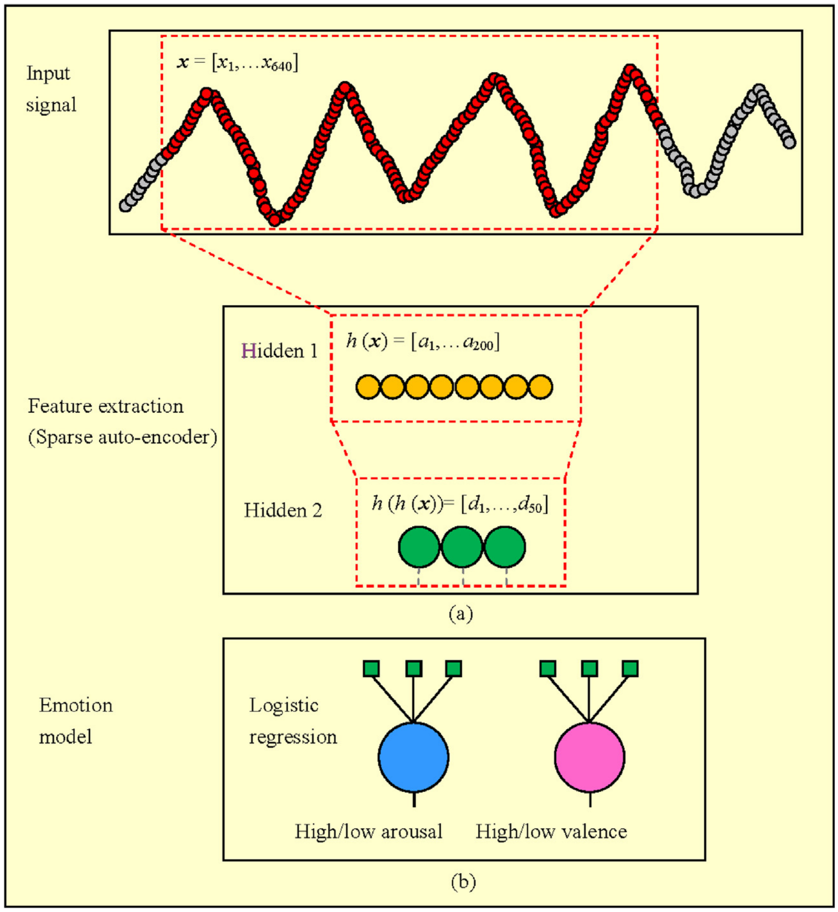

Fig. 2. Example of general architecture of the deep learning framework. The architecture contains: (a) a sparse autoencoder (SAE) with two hidden layers, and (b) two logistic classifiers. In the illustrated SAE, the first hidden layer (200 neurons) processes a respiration signal with path length of 640 samples. A second hidden layer (50 neurons) processes the 200 neurons from the first hidden layer. The final 50 neurons (features) form the output of the SAE which provides a number of extracted (learned) features which feed the input of the logistic regression.

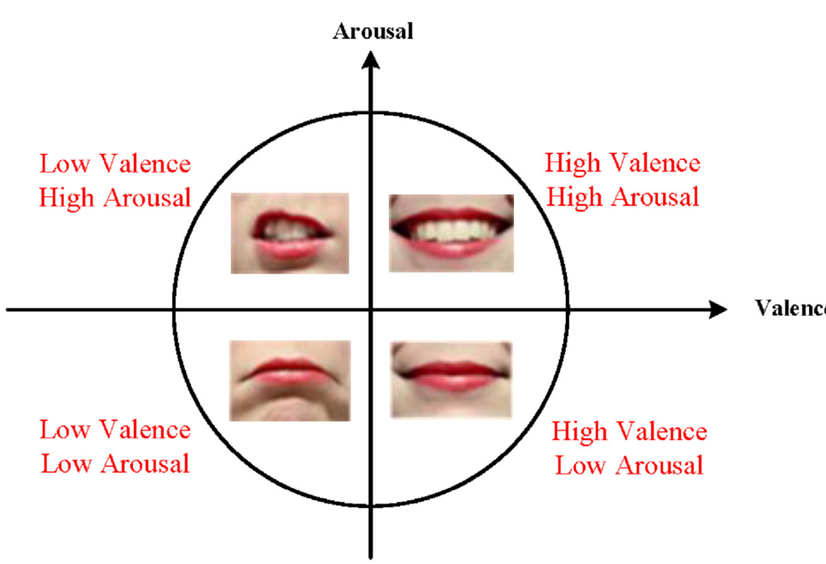

Fig. 3. Arousal-valence theory of emotions. Y-axis represents the arousal while the x-axis represents the valence.

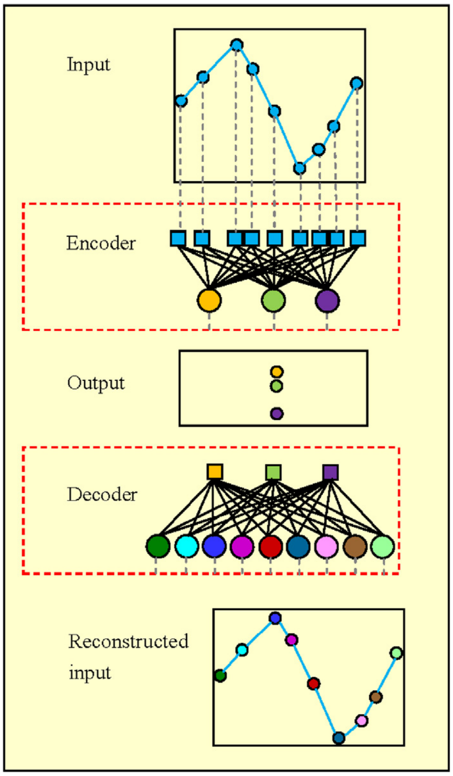

Fig. 4. The structure of an auto-encoder. The encoder generates thelearned representation (extracted features) from the input signals.The features are fed to adecoder that attempts to reconstruct the input.

measures the intensity of emotional activation, while valence is the horizontal axis and describes the negative or positive mind. The arousal-valence emotion theory suggests that a common and interconnected neurophysiological system is responsible for all affective states [27].

We can see in Fig. 3 how the emotions are classified by this theory. As we can see the "High Valence, Low Arousal" is mapped in the lower right quadrant, whereas the "Low Valence, High Arousal" is mapped in the higher left quadrant. The "High Valence, High Arousal" is somewhere in the higher right quadrant, while the "Low Valence, Low Arousal" is in the lower left quadrant. By determining the arousal valence value of certain emotion, we can recognize emotions as the four classes. In this way, the emotion recognition task is divided into two binary classification tasks.

3. Deep learning framework

We investigated an effective method that maps signals of user behavior to affective states. As shown in Fig. 2, we use a deep learning framework consisting of the sparse auto-encoder that transforms the raw signals into extracted features and two logistic regression that predict affective states. Our hypothesis is that automatic feature extraction via deep learning will yield physiological affect detectors of higher predictive power, which, in turn, will deliver affective models of higher accuracy.

3.1. Sparse auto-encoder

Sparse auto-encoder is a method of deep learning that is used to automatically learn features from unlabeled data. As the basic element of SAE, an auto-encoder (AE) converts the input data to the hidden representation (extracted features) by the encoder (see Fig. 4). The AE also learns how to map back the features to the input space by a decoder. The original and the reconstructed input data are as similar as possible, despite of small reconstruction error on the training examples. The whole structure takes the extracted features of one AE as the input of another AE, to find general representation from the input hierarchically [28–30].

In the following procedures, we focused on pre-training an AE, which is taken as the first SAE hidden layer. The techniques can also be applied to train the second and other hidden layers. Once pre-trained, the first layer will not be changed during training other hidden layers. Generally, the encoder maps an input example $x \in R^n$ to a hidden representation $h(\boldsymbol{x}) \in R^m$

$$h(\mathbf{x}) = f(\mathbf{W}_1 * \mathbf{x} + \mathbf{b}_1), \tag{1}$$

where $\boldsymbol{W}_1 \in R^{m \times n}$ is a weight matrix, $\boldsymbol{b}_1 \in R^m$is a bias vector, and $f(z)$ is a non-linear activation function, typically a sigmoid function $f(z) = 1/(1+\exp(-z))$. The decoder maps the hidden representation back to a reconstructed input $\tilde{\boldsymbol{x}} \in R^n$:

$$\tilde{\mathbf{x}} = f(\mathbf{W}_2 * h(\mathbf{x}) + \mathbf{b}_2), \tag{2}$$

where $\boldsymbol{W}_2 \in R^{n \times m}$ is a weight matrix, and $\boldsymbol{b}_2 \in R^n$ is a bias vector and f is the same function as that of the encoder.

To minimize the dissimilarity between the original and the reconstructed input, we consider decreasing the reconstruction gap $\sum_{i=1}^{D} \| \boldsymbol{x}^{(i)} - \tilde{\boldsymbol{x}}^{(i)} \|$.Given the training dataset with D examples $\boldsymbol{x}^{(i)}$, $i = 1, \ldots, D$, the weight matrices $\boldsymbol{W}_1$and $\boldsymbol{W}_2$ and the bias vector $\boldsymbol{b}_1$and $\boldsymbol{b}_2$ are adjusted by back-propagation. Further, we constrain the expected activation of the hidden units to be sparse by adding a penalty term [31]. Thus it turns out to be the following optimization problem:

$$\min \sum_{i=1}^{P} \| \mathbf{x}^{(i)} - \tilde{\mathbf{x}}^{(i)} \|^2 + \beta \sum_{j=1}^{m} \mathrm{SP}(\rho \parallel \hat{\rho}_j), \tag{3}$$

where

$$\mathrm{SP}(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_j} \tag{4}$$

is a sparse penalty term, $\hat{\rho}_j = \frac{1}{D}\sum_{i=1}^{D} h_j(\boldsymbol{x}^{(i)})$ is the average activation of hidden unit j, ρ is a sparsity level, and β leverages the weight of the sparsity penalty term. We minimized the cost function by the L-BFGS [32], so as to obtain the optimal $\boldsymbol{W}_1$ and $\boldsymbol{b}_1$that is used to find the inner features $h(\boldsymbol{x})$.

3.2. Logistic regression

After extracting features via the sparse auto-encoder, we fed the features into two logistic regression (LR) to recognize the arousal and valence, respectively. LR processes the binary classification problem: $y^{(i)} \in \{0, 1\}$. To train the logistic regression, we consider the following procedure. We have a training set $\{(\boldsymbol{x}^1, y^1), \ldots, (\boldsymbol{x}^D, y^D)\}$ of D labeled examples, where the input features are $\boldsymbol{x}^{(i)} \in R^n$. Given an input $\boldsymbol{x}^{(i)}$, we expect our hypothesis to estimate the probability $P(y^{(i)} = k | \boldsymbol{x}^{(i)})$ for each class. Thus, our hypothesis will output a 2-dimensional vector (whose elements sum to 1) giving us two probabilities. Concretely, our hypothesis $h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})$ takes the form:

$$h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) = \frac{1}{1+\exp(-\boldsymbol{\theta} * \mathbf{x}^{(i)})}. \tag{5}$$

Here $\boldsymbol{\theta} \in R^n$ is the vector of parameters of LR. The parameters were trained by minimizing the cost function:

$$J(\boldsymbol{\theta}) = -\frac{1}{D}\sum_{i=1}^{D}\frac{1}{2}\left(h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) - y^{(i)}\right)^2. \tag{6}$$

We optimized the cost function by the L-BFGS method so as to obtain the optimal parameters of LR.

4. Experiment

4.1. Experiment data

To test the efficacy of our method, we adopted an international database known as Dataset for Emotion Analysis using Physiological signals (DEAP). DEAP consists of physiological signals from the spontaneous reactions of 32 participants in response to 40 one-minute long music video clips. These 32 participants age between 19 and 37 (mean age 26.9) and half of them are females. In each of the 40 trials, signal recordings include electroencephalogram, electromyography, electrooculogram, galvanic skin response, temperature, respiration pattern, plethysmography, audio and video signals. Participants were asked to do self-assessment by assigning values from 1 to 9 to five different status, namely, valence, arousal, dominance, liking, and familiarity.

In order to realize fast on-site emotion detection, we tried to evaluate RSP segments with different time length between 5 s and 60 s. Some choices result in bad performance, some other choices maybe contain more information, but they are too long for on-site detection. We gave consideration to both the efficacy and efficiency and finally decided to use 20-s segments. The RSP signal was sampled at 32 Hz. From the start of each record, we cut the RSP data in 20-s segments and the cutting window moves every second. I.e., in two adjacent segments, the later one starts from the 2rd second of the former segment, as a result there is 19-s overlapping between adjacent segments. So in a 60-s trial, we could obtain 41 segments. In total of 32 participants, with 40 trails in every participant and 41 segments in each trial, we have 52,480 (41 segments $\times$ 40 trails $\times$ 32 participants) data samples. Every segment was not normalized using global mean and standard deviation of the 60-s trial; instead, they were normalized by the local mean value and standard deviation within each segment, making learned features only sensitive to variation within the segment and insensitive to the baseline level.

To further evaluate the proposed method on other people, after model establishment, we used the affection database established by Augsburg University. The database consists of 4 affective states over 25 days. Musical induction method spontaneously leads subjects to real emotional states. The RSP signal was also sampled at 32 Hz. Following the abovementioned procedure, the 120-s RSP data wad cut into 101 20-s segments. In total of 25 days, with 4 emotions in every day and 101 segments in each trial, we have 10,100 (101 segments $\times$ 4 emotions $\times$ 25 days) data samples.

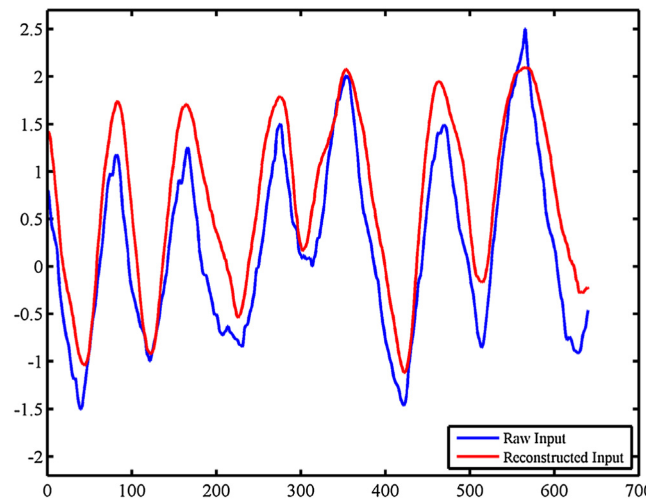

Fig. 5. The original (Blue) and the reconstructed (Red) RSP. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Experiment setting

We trained several sparse auto-encoders to extract features for the RSP signal and across all affective states in the dataset. The topologies of the auto-encoders were selected after preliminary experiments with 1- and 2-layer SAE and 20-step trial neuron numbers in each hidden layer. To test the efficacy of our method with previous results, we divided the value of arousal and valence into two binary classes according to the assigned values. The threshold we chose is 5, and the tasks can be treated as binary classification problems, namely, high/low valence and high/low arousal. Among all of the DEAP data, 80% samples were used as training data and the rest 20% samples were used as test data. After the emotion recognition model establishment, the whole dataset of Augsburg University is used to evaluate its generalization capability on other people.

4.3. Experiment results

We have systematically tested critical parameters of SAE (e.g., the number of layers, the numbers of neurons in each layer, and the sparsity penalty). The number of neurons in each hidden layer was searched with step of 20 in the range of [10:500]. The learning rate was 0.01 and the momentum regarding the weight update was used to prevent being struck in local minima. In our experiment, suppose the inputs are the respiration signal values from a 20-s segment (32 Hz, 640 points) so $\boldsymbol{x} \in R^{640}$, we used 200 neurons in the first hidden layer and 50 neurons in the second hidden layer, so $h_1(\boldsymbol{x}) \in R^{200}$ and $h_2(h_1(\boldsymbol{x})) \in R^{50}$.

Table 1
Comparison with representative studies in the area of physiological signal-based emotion recognition in recent years.

	Signal Type	Method	Classification Accuracy	
			Valence	Arousal
Zhuang et al. [34]	EEG	Support Vector Machine	70.9	67.1
Torres-Valencia et al. [35]	RSP, GSR, EEG, TEMP	Hidden Markov Models	58.75	75.00
Martinez et al. [36]	BVP, SC	Convolutional Neural Network	63.3	69.1
Xu et al. [37]	EEG	Deep Belief Networks	66.88	69.84
Liu et al. [38]	EOG, EEG	Multimodal Deep Learning	**85.2**	80.5
Our work, 2017	RSP	Deep Sparse Auto-Encoders	73.06	**80.78**

After each layer of SAE was pretrained, we fine-tuned the deep learning framework include the SAE and the logistic regression. We demonstrated results on the optimal SAE that has two hidden layers. As shown in Fig. 1, the first and second hidden layer contains 200 and 50 neurons, respectively. To show the reconstruction of SAE, as in Fig. 5, the same RSP signal (blue) with Fig. 1(b) is fed into the SAE. The red line represents the output (reconstructed input) of SAE. We can indicate that the output is able to track changes of input even when the input RSP signal is complex.

Table 1 compares our method with some representative studies that are also tested on the DEAP dataset. Our work achieved similar accuracy with these methods. The classification accuracy of valence and arousal is 73.06% and 80.78%, respectively. Further, we evaluated the proposed method using the affection database established by Augsburg University. The established model achieved 85.89% accuracy for arousal classification and 83.72% for valence classification. The mean accuracy of four types of emotions is 80.22%.

5. Discussion

In this study we propose an effective emotion recognition system based on RSP alone. The main contributions of this paper can be described as the following aspects. First, considering the feature extraction and feature selection properties of deep learning, we introduced the sparse auto-encoder to extracted respiration features for emotion recognition. Second, the experiment results indicate that the respiration signal is so powerful in affective state recognition that wearable devices do not have to fully rely on EEG. Deep learning is of computational complexity in model training but takes little time in testing. On average, it takes our algorithm 2.81 s to test one data sample on the Matlab 2014b in the 32-bit Windows 7 operating system with an Intel i5 CPU and 4G RAM. At least, we believe the 2.81 s is an acceptable response time in practical emotion recognition applications.

In Fig. 1, RSP signals under four emotional states are compared. We can indicate that emotions with high valence have more steady respiration with uniform RSP magnitude than those with low valence. As for difference in arousal, high arousal has large respiration frequency while low arousal has small frequency. Low arousal has fast sniffing and slow breath with long tail.

About signal types, most studies use EEG, this is mainly based on the consideration that the brain is primarily responsible for emotion processing and corresponding responses. Consequently, an increasing amount of researchers are dedicated to detect affective states using neurosignal paradigms such as EEG. However, brain activity nature of emotional functions is partially localized in both space (some cortical and subcortical regions) and frequency (mostly upon the neural oscillation bands) [33]. Although EEG provided high temporal resolution, it suffers from its poor spatial resolution and raised susceptibility to noise. More attention should be transferred from the mainstreaming EEG to other psycho-physiological signals.

As for the classifiers, Table 1 clearly shows the ever growing proliferation of deep learning in emotion classification. Different deep models (CNN, DBN) has gradually replaced shallow models (SVM, HMM) and achieved superior performance. This is mainly due to several facts. First of all, deep models are more flexible in feature extraction and eliminate the assumptions used by current feature extraction algorithms, such as the linearity assumption in PCA. Some irregular respiration activities, such as whimper and guffaw, may lead to complex signal waveform and cessation of respiration, which couldn't be fully characterized by statistical features. Deep learning can automatically derive features, which may contain information that manually pre-designed statistical features don't involve. Therefore we selected deep learning with the hope of seizing the intrinsic characteristics of respiration activities. Second, deep learning makes use of the unlabeled data and reduce the number of labeled samples needed for the learning process where labeled data can guide a learning algorithm towards faster hyper-parameter selection. As a result, the emotion recognition model is fully-data driven and application specific.

Although studies in Table 1 all used the DEAP dataset, they were evaluated in different ways, such as different fold cross validation and different partition of test dataset, so the results of these studies cannot be compared directly. Nevertheless, we can still roughly find that our method almost achieved the same performance as the best study did.

Respiration signals have a simple structure that is similar to a sinusoidal function or a triangular function, which made it very easy to capture the inner features. Moreover, respiration signals are of non-invasiveness, cost effectiveness and simple acquisition, thus is promising in wearable applications, such as wrist-based emotion detection. It is now possible to implement RSP-based emotion recognition from laboratories to real-world applications.

6. Conclusion

This paper investigates the potential of respiration signals alone as reliable tools for emotion recognition. One of the widespread dimensional affect theories, the arousal-valence theory, is adopted to help classify four types of emotions. A sparse auto-encoder is applied to extract and select emotion-related features. Two logistic regression detect a high/low arousal and a positive/negative valence. Test results show that the proposed deep learning framework achieved acceptable performance. It can be concluded that respiration signals have a great potential to recognize emotions. Future wearable devices are expected to monitor human emotions without interrupting ongoing activities.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61302033), National Key Research and Development Project 2016YFC1304302 and Key Project of Beijing Municipal Natural Science Foundation Z16003.

References

[1] R.W. Picard, R. Picard, Affective Computing, vol. 252, MIT Press, Cambridge, 1997.
[2] P. Ekman, R.W. Levenson, W.V. Friesen, Autonomic nervous system activity distinguishes among emotions, Science 221 (4616) (1983) 1208–1210.
[3] F. Hönig, A. Batliner, E. Nöth, Real-time recognition of the affective user state with physiological signals, Proceedings of the Doctoral Consortium, Affective Computing and Intelligent Interaction (2007).
[4] J. Anttonen, V. Surakka, Emotions and heart rate while sitting on a chair, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2005, pp. 491–499 April.
[5] C.M. Jones, T. Troen, Biometric valence and arousal recognition, Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, ACM, 2007, pp. 191–194 November.
[6] R.L. Mandryk, K.M. Inkpen, T.W. Calvert, Using psychophysiological techniques to measure user experience with entertainment technologies, Behav. Inform. Technol. 25 (2) (2006) 141–158.
[7] T. Vogt, E. André, Improving automatic emotion recognition from speech via gender differentiation, Proc. Language Resources and Evaluation Conference (LREC 2006), Genoa, 2006 May.
[8] A. Mill, J. Allik, A. Realo, R. Valk, Age-related differences in emotion recognition ability: a cross-sectional study, Emotion 9 (5) (2009) 619.
[9] R.W. Picard, E. Vyzas, J. Healey, Toward machine emotional intelligence: analysis of affective physiological state, IEEE Trans. Pattern Anal. Mach. Intell. 23 (10) (2001) 1175–1191.
[10] R.N. Duan, J.Y. Zhu, B.L. Lu, Differential entropy feature for EEG-based emotion classification, 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, 2013, pp. 81–84 November.
[11] D. Giakoumis, D. Tzovaras, K. Moustakas, G. Hassapis, Automatic recognition of boredom in video games using novel biosignal moment-based features, IEEE Trans. Affect. Comput. 2 (3) (2011) 119–133.

[12] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. U. S. A. 88 (6) (1991) 2297–2301.
[13] G.N. Yannakakis, J. Hallam, Entertainment modeling through physiology in physical play, Int. J. Hum. Comput. Stud. 66 (10) (2008) 741–755.
[14] Y.P. Lin, C.H. Wang, T.P. Jung, T.L. Wu, S.K. Jeng, J.R. Duann, J.H. Chen, EEG-based emotion recognition in music listening, IEEE Trans. Biomed. Eng. 57 (7) (2010) 1798–1806.
[15] X.W. Wang, D. Nie, B.L. Lu, Emotional state classification from EEG data using machine learning approach, Neurocomputing 129 (2014) 94–106.
[16] H.P. Martinez, Y. Bengio, G.N. Yannakakis, Learning deep physiological models of affect, IEEE Comput. Intell. Mag. 8 (2) (2013) 20–33.
[17] W.L. Zheng, J.Y. Zhu, Y. Peng, B.L. Lu, EEG-based emotion classification using deep belief networks, 2014 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2014, pp. 1–6 July.
[18] K. Li, X. Li, Y. Zhang, A. Zhang, Affective state recognition from EEG with deep belief networks, 2013 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2013, pp. 305–310 December.
[19] X. Zhu, W.L. Zheng, B.L. Lu, X. Chen, S. Chen, C. Wang, EOG-based drowsiness detection using convolutional neural networks, IJCNN, (2014) , pp. 128–134 July.
[20] W.L. Zheng, B.L. Lu, Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks, IEEE Trans. Auton. Ment. Dev. 7 (3) (2015) 162–175.
[21] W.L. Zheng, H.T. Guo, B.L. Lu, Revealing critical channels and frequency bands for emotion recognition from EEG with deep belief network, 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, 2015, pp. 154–157 April.
[24] S. Jirayucharoensak, S. Pan-Ngum, P. Israsena, EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation, Sci. World J. 2014 (2014).
[25] M. Längkvist, L. Karlsson, A. Loutfi, Sleep stage classification using unsupervised feature learning, Adv. Artif. Neural Syst. 5 (2012).
[26] J.A. Russel, A circumplex model of affect, J. Pers. Soc. Psychol. (1980).
[27] J. Posner, J.A. Russell, B.S. Peterson, The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology, Dev. Psychopathol. 17 (03) (2005) 715–734.
[28] Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.
[29] I. Goodfellow, H. Lee, Q.V. Le, A. Saxe, A.Y. Ng, Measuring invariances in deep networks, Advances in Neural Information Processing Systems, (2009) , pp. 646–654.
[30] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, Adv. Neural Inform. Process. Syst. 19 (2007) 153.
[31] H. Lee, C. Ekanadham, A.Y. Ng, Sparse deep belief net model for visual area V2, Advances in Neural Information Processing Systems, (2008) , pp. 873–880.
[32] D.C. Liu, J. Nocedal, On the limited memory BFGS method for large scale optimization, Math. Program. 45 (1–3) (1989) 503–528.
[33] Emotion Discrimination Using Spatially Compact Regions of Interest Extracted from Imaging EEG Activity.
[34] X. Zhuang, V. Rozgic, M. Crystal, Compact unsupervised eeg response representation for emotion recognition, 2014 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), IEEE, 2014, pp. 736–739 June.
[35] C.A. Torres-Valencia, H.F. Garcia-Arias, M.A.A. Lopez, A.A. Orozco-Gutiérrez, Comparative analysis of physiological signals and electroencephalogram (eeg) for multimodal emotion recognition using generative models, 2014 XIX Symposium on Image, Signal Processing and Artificial Vision (STSIVA), September, 2014, pp. 1–5 July.
[36] H.P. Martínez, Advancing Affect Modeling via Preference Learning and Unsupervised Feature Extraction, IT University of Copenhagen, Center for Computer Cames Research, 2013.
[37] H. Xu, K.N. Plataniotis, EEG-based affect states classification using deep belief networks, Digital Media Industry & Academic Forum (DMIAF), IEEE, 2016, pp. 148–153 July.
[38] W. Liu, W.L. Zheng, B.L. Lu, Emotion recognition using multimodal deep learning, International Conference on Neural Information Processing, Springer International Publishing, 2016, pp. 521–529 October.