Computers in Industry 92 (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 xxx

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

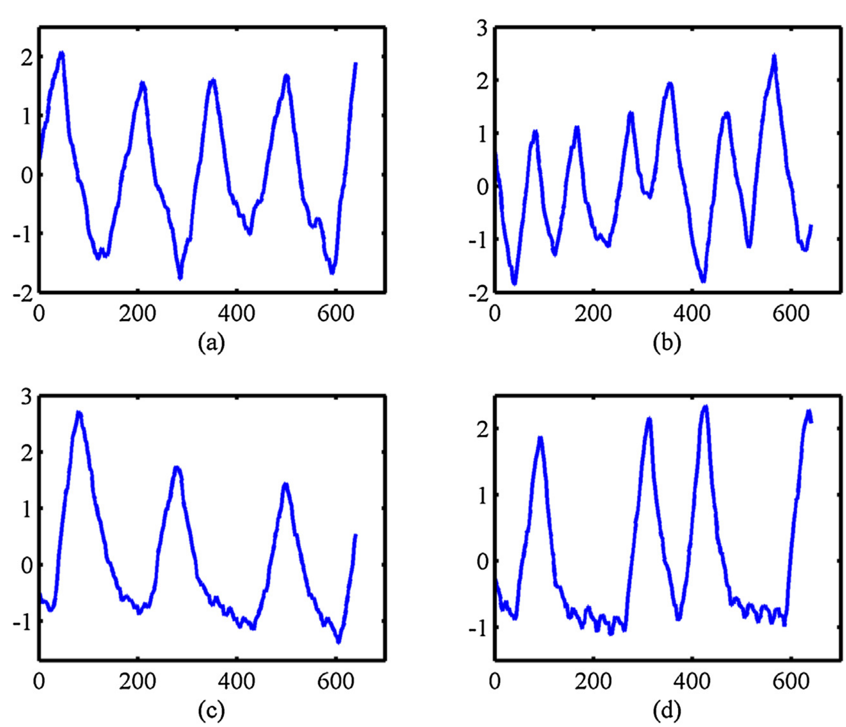

Fig. 1. Four 20-s respiration signal segments under different emotional states: (a) high valence and high arousal, (b) low valence and high arousal, (c) low valence and low arousal, (d) high valence and low arousal. The horizontal axis represents the sampling points while the vertical axis is the locally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