Expert Systems With Applications 42 (2015) 8678–8697

Contents lists available at ScienceDirect

Expert Systems With Applications

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

# Intelligent affect regression for bodily expressions using hybrid particle swarm optimization and adaptive ensembles

Yang Zhang [a], Li Zhang [a,*], Siew Chin Neoh [a], Kamlesh Mistry [a], Mohammed Alamgir Hossain [b]

[a] Computational Intelligence Research Group, Department of Computer Science and Digital Technologies, Faculty of Engineering and Environment, Northumbria University, Newcastle NE1 8ST, UK
[b] Anglia Ruskin IT Research Institute, Faculty of Science and Technology, Anglia Ruskin University, Cambridge CB1 1PF, UK

ARTICLE INFO

*Keywords:*
Bodily expression
Adaptive ensemble regression
Particle swarm optimization
Genetic algorithm
Support vector regression
Mutation distributions

ABSTRACT

This research focuses on continuous dimensional affect recognition from bodily expressions using feature optimization and adaptive regression. Both static posture and dynamic motion bodily features are extracted in this research. A hybrid particle swarm optimization (PSO) algorithm is proposed for feature selection, which overcomes premature convergence and local optimum trap encountered by conventional PSO. It integrates diverse jump-out mechanisms such as the genetic algorithm (GA) and mutation techniques of Gaussian, Cauchy and Levy distributions to balance well between convergence speed and swarm diversity, thus called GM-PSO. The proposed PSO variant employs the subswarm concept and a cooperative strategy to enable mutation mechanisms of each subswarm, i.e. the GA and the probability distributions, to work in a collaborative manner to enhance the exploration and exploitation capability of the swarm leader, sustain the population diversity and guide the search toward an ultimate global optimum. An adaptive ensemble regression model is subsequently proposed to robustly map subjects' emotional states onto a continuous arousal–valence affective space using the identified optimized feature subsets. This regression model also shows great adaption to newly arrived bodily expression patterns to deal with data stream regression. Empirical findings indicate that the proposed hybrid PSO optimization algorithm outperforms other state-of-the-art PSO variants, conventional PSO and classic GA significantly in terms of catching global optimum and discriminative feature selection. The system achieves the best performance for the regression of arousal and valence when ensemble regression model is applied, in terms of both mean squared error (arousal: 0.054, valence: 0.08) and Pearson correlation coefficient (arousal: 0.97, valence: 0.91) and outperforms other state-of-the-art PSO-based optimization combined with ensemble regression and related bodily expression perception research by a significant margin.



## 1. Introduction

In the last decade, automatic emotion recognition has become a new hotspot of AI research as the role played by affect in human life and everyday functioning is well recognized and studied (Izard, Ackerman, Schoff, & Fine, 2002). It not only greatly benefits natural human–computer interaction, but also shows great potential to be applied in a wide variety of applications, such as personalized learning (D'Mello & Graesser, 2010), health monitoring (Lucey et al., 2009), anomalous event detection (Ryan et al., 2009), and interactive computer entertainment (G'Mussel & Hewig, 2013; Savva, Scarinzi, & Bianchi-Berthouze, 2012). Emotions can be expressed and perceived through a wide range of channels, such as face, voice, text, and bodily expressions. However, most efforts conducted so far on automatic emotion recognition have concentrated on the facial (e.g. Li, Chen, Zhao, & Ji, 2013; Owusu, Zhan, & Mao, 2014; Rao, Saroj, Maity, & Koolagudi, 2011) and vocal modalities (e.g. Kostov & Fukuda, 2007; Oudeyer, 2003). Only recently there have been fewer automatic systems that are able to detect emotions based on the bodily modality (e.g. Bernhardt & Robinson, 2007; Kleinsmith, Bianchi-Berthouze, & Steed, 2011). This may be attributed to the complexity of the body itself and the lack of well acknowledged coding models for the body as there are for the face (e.g. the well-established Facial Action Coding System (Ekman, Friesen, & Hager, 2002)).

Recently, the role and importance of body language in the expression and perception of emotions have been revealed by studies in both cognitive neuroscience (e.g. De Gelder, 2009; Van den Stock, Righart, & De Gelder, 2007; Wallbott, 1998) and computer science (e.g. Kleinsmith & Bianchi-Berthouze, 2013; Kleinsmith et al., 2011). Recent studies in cognitive neuroscience (De Gelder, Snyder, Greve, Gerard, & Hadjikhani, 2003; Van den Stock et al., 2007) have further

* Corresponding author. Tel.: +44 1912437089.
*E-mail address:* li.zhang@northumbria.ac.uk (L. Zhang).

http://dx.doi.org/10.1016/j.eswa.2015.07.022
0957-4174/

emphasized that body posture could be the influencing factor over facial expression in cases of incongruent affective displays, and for the discrimination between some emotional states in particular, such as fear and anger, more attention needs to be paid to the bodily display.

Furthermore, a longstanding controversy in cognitive science has concerned whether emotions are better conceptualized in the form of discrete categories (e.g. happy and sad), or continuous dimensions (e.g. valence and arousal) (Hamann, 2012). According to Ekman and Friesen (1967a, 1967b), compared to the face, which is considered to be the foremost modality for expressing discrete emotion categories, the body may perform better for communicating affective dimensions. Recent research (Kleinsmith & Bianchi-Berthouze, 2013) also indicates that by the combination of discrete emotion labels and continuous dimension levels, a more complete and systematic description of the emotional state could be obtained. These highlight the importance of developing a dimensional emotion recognition system based on bodily expressions.

More importantly, current neuroscience studies (Giese & Poggio, 2003; Lange & Lappe, 2007; Vania, Lemay, Bienfang, Choi, & Nakayama, 1990) indicate that our brain utilizes two separate pathways for the recognition of biological information from bodily expressions, one for form information (e.g. a specific configuration of a posture), and the other for motion information (e.g. velocity, acceleration, and frequency). According to Atkinson, Tunstall, and Dittrich (2007), both form and motion bodily signals make their own contributions to affect perception of human behavior. A number of recent developments in computer science (e.g. Kleinsmith et al., 2011; Roether, Omlin, Christensen, & Giese, 2009) further prove that both of them are useful and important for automatic emotion prediction from bodily expressions. Body form and motion information complement each other in conveying emotions, however, they may also become partially redundant or inconsistent in some cases (Kleinsmith & Bianchi-Berthouze, 2013). Thus, it is also significant to identify the roles of both body form and movement information in the automatic regression of different affective dimensions.

This paper aims to address the problem of continuous regression of users' emotional states in a valence and arousal space based on their whole-body expressions. I.e. the proposed system maps subjects' emotional states to a two-dimensional coordinate space spanned by arousal and valence, where each value ranges between $-1$ and 1. Our contributions are summarized as follows:

1. We first of all extract users' static and dynamic bodily features. A particle swarm optimization (PSO) variant is proposed for feature selection and applied to the arousal and valence dimensions respectively. It overcomes early convergence and local optimum trap encountered by conventional PSO by integrating the genetic algorithm (GA) and long-jump mutation techniques of Gaussian, Cauchy and Levy distributions, thus called GM-PSO. In comparison to state-of-the-art PSO research, the proposed algorithm employs the subswarm concept and different strategies for mutations in each subswarm to diversify the search. I.e. in order to improve the balance between exploration and exploitation, GM-PSO divides the original population into two subswarms and employs short and long jumps via the GA operation and the three probability distributions respectively in each subswarm to enhance the subswarm leader and sustain the population diversity to mitigate early convergence. A cooperative strategy is also employed in GM-PSO to enable the above jump-out mechanisms of each subswarm to work in a collaborative manner to increase exploration capabilities of the global swarm leader, sustain the population diversity and achieve a better balance between population diversity and convergence speed. Empirical results indicate that it effectively escapes from the local optimum traps, compares favorably with the state-of-the-art hybrid PSO algorithms and outperforms the conventional PSO and classic GA significantly in terms of finding global optimum and discriminative feature selection.
2. An ensemble regression model with great adaptability is also proposed to robustly predict users' continuous affective dimensions in the valence and arousal space using whole-body expressions. The proposed ensemble model with support vector regressors as the base regressors achieves the best performance and outperforms single model based methods and other related research reported in the literature. Furthermore, it also employs a stand-by regressor to better deal with newly arrived unseen bodily expressions and data stream regression.
3. Continuous and dimensional affective annotation is inherently a challenging task. We present a new annotation method based on inter-annotator correlations and mean value differences to effectively fuse multiple annotations to build ground truth for system evaluation.

The remainder of this paper is organized as follows. Section 2 provides introductions of diverse emotion theories and summarizes related work in automatic affect recognition from bodily expressions. Section 3 presents feature extraction from whole-body expressions and automatic feature selection using the proposed GM-PSO. The proposed adaptive ensemble model for continuous and dimensional affect regression is subsequently discussed in Section 4, together with the other two benchmark single regression methods. In Section 5, we discuss the process of data collection and affective annotation, as well as experimental results in comparison with other state-of-the-art research. Finally, we draw conclusions and identify future directions in Section 6.

## 2. Related work

In this section, we first provide a succinct overview on the conceptualization of emotions universally acknowledged by psychologists. We then briefly discuss recent findings on how both body form and movement perform in affect expression and perception. The state-of-the-art automatic bodily emotion recognition systems and developments are also summarized.

### 2.1. Discrete vs. continuous modeling of emotions

In the literature of psychology, there are mainly two different approaches to structure and differentiate between different emotional states: discrete categories and continuous dimensions. The discrete model argues that the affective state is able to be represented by a number of prototypical emotions or their mixtures. This model has been well adopted and promoted by Ekman et al. (2002) and Izard (1994). According to their studies, there exists a series of basic emotions that can be expressed through corresponding prototypical facial expressions. For example, Fig. 1(a) shows facial expressions for the six basic emotions (i.e. happiness, surprise, fear, anger, sadness, and disgust) (Ekman & Friesen, 1967a, 1967b).

The continuous model argues that emotions are able to be described by certain continuous attributes, and the affective state of each participant could be placed within a continuous low-dimensional space. A representative model proposed by Posner, Russell, and Peterson (2005) employed two orthogonal dimensions: valence and arousal. The valence dimension describes the level of pleasure of an emotion, and it ranges from negative unpleasant feelings to positive pleasant feelings. The arousal dimension refers to the intensity of the emotional experience, and it ranges from apathetic sleepiness to frantic excitement. Fig. 1(b) illustrates the two dimensional emotion model and distributions of some identified emotion categories (Breazeal, 2003; Posner et al., 2005). The dimensional model could be a more flexible and effective way to interpret emotions, especially in the cases of (1) no clear categorical description available for an emotional state; (2) bodily expression-based