Learning to decode human emotions from event-related potentials

Article *in* Neural Computing and Applications · April 2014

DOI: 10.1007/s00521-014-1653-6

6 authors, including:

Olga Georgieva
Sofia University "St. Kliment Ohridski"

Sergey Milanov
Sofia University "St. Kliment Ohridski"

Petia Georgieva
University of Aveiro

Carlos Silva
University of Aveiro

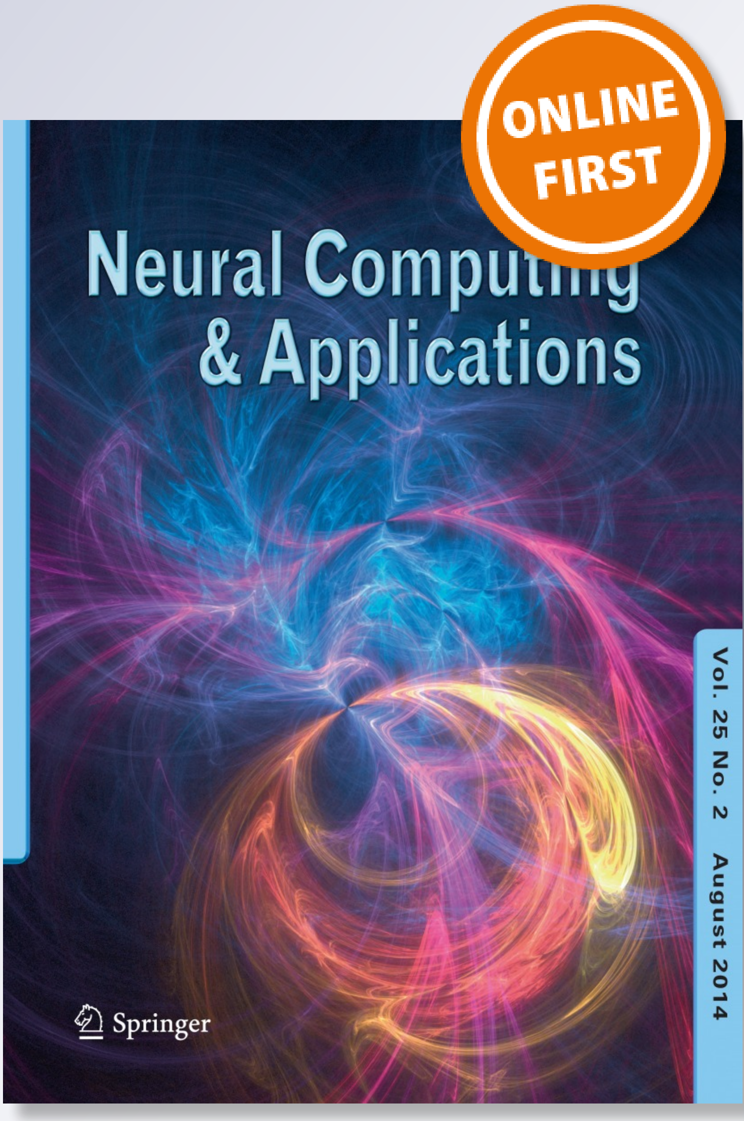
Learning to decode human emotions from event-related potentials

O. Georgieva, S. Milanov, P. Georgieva, I. M. Santos, A. T. Pereira & C. F. Silva

Neural Computing and Applications

ISSN 0941-0643

Neural Comput & Applic
DOI 10.1007/s00521-014-1653-6



Neural Comput & Applic
DOI 10.1007/s00521-014-1653-6

ADVANCES IN INTELLIGENT DATA PROCESSING AND ANALYSIS

Learning to decode human emotions from event-related potentials

O. Georgieva · S. Milanov · P. Georgieva · I. M. Santos · A. T. Pereira · C. F. Silva

Received: 30 November 2013 / Accepted: 28 June 2014

Abstract Reported works on electroencephalogram (EEG)-based emotion recognition systems generally employ the principles of supervised learning to build subject-dependent (single/intra-subject) models. Building subject-independent (multiple/inter-subject) models is a harder problem due to the EEG data variability between subjects. The contribution of this paper is twofold. First, we provide a framework for selection of a small number of basic temporal features, event-related potential (ERP) amplitudes, and latencies that are sufficiently robust to discriminate emotion states across multiple subjects. Second, we test comparatively the feasibility of six standard unsupervised (clustering) techniques to build intra-subject and inter-subject models to discriminate emotion valence in the ERPs collected while subjects were viewing high arousal images with positive or negative emotional content.

O. Georgieva · S. Milanov
FMI, Sofia University "St. Kl. Ohridski", Sofia, Bulgaria
e-mail: olgaig@fmi.uni-sofia.bg

P. Georgieva (✉)
DETI/IEETA, University of Aveiro, 3810-193 Aveiro, Portugal
e-mail: petia@ua.pt

I. M. Santos · A. T. Pereira · C. F. Silva
Dept. de Educação, University of Aveiro, 3810-193 Aveiro, Portugal

I. M. Santos · A. T. Pereira · C. F. Silva
Institute of Biomedical Imaging and Life Sciences (IBILI), Coimbra, Portugal

I. M. Santos · A. T. Pereira · C. F. Silva
Center for Health Technology and Services Research (CINTESIS), Porto, Portugal

Keywords Affective computing · Emotion valence recognition · Feature reduction · Clustering · Event-related potentials (ERPs)

1 Introduction

The quantification and automatic detection of human emotions are the focus of the interdisciplinary research field of affective computing (AC). In [1], a broad overview of the current AC systems is provided. Major modalities for affect detection are facial expressions, voice, text, body language, and posture. However, it is easier to fake facial expressions, posture, or change tone of speech than trying to conceal physiological signals such as galvanic skin response (GSR), electrocardiogram (ECG), or electroencephalogram (EEG). Since emotions are known to be related to neural activity in certain brain areas, affective neuroscience (AN) emerged as a new modality that attempts to find the neural correlates of emotional processes [2]. The major brain-imaging techniques include EEG, magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and positron emission tomography (PET). Among them, the EEG modality [3, 4, 5] has attracted more attention because it is a noninvasive, relatively cheap, and easy to apply technology. A comprehensive list of EEG-based emotion recognition researches is provided in [6]. It is difficult to compare results among them because there are a lot of factors that make different results from different researches including participant, model of emotion, stimulus, feature, temporal window, and classifier. However, the reported works generally employ the principles of supervised learning. Several machine learning algorithms have been used as emotion classifiers such as support vector

Published online: 23 July 2014

machine (SVM) [7–9], Naïve Bayes (NB) [10], K-nearest neighbors (KNN) [11], linear discriminant analysis (LDA) [12], ensemble classifiers such as random forest [13], and simple thresholds [14]. Above researches build an individual classification model for every subject (single/intra-subject model). Building subject-independent (multiple/inter-subject) models is a harder problem due to the EEG data variability between subjects. Supervised AC systems are normally suitable for intra (single) subject emotion recognition and suffer of generalization inaccuracy across multiple subjects [15]. An EEG-based emotion (happy/unhappy) detection system that explicitly considers inter-subject models has been presented recently in [6]; however, it also explores supervised a classification technique (SVM) based on frequency domain features as power spectral density (PSD).

Therefore, the primary goal of the present paper is to prove the feasibility of classical unsupervised learning techniques to decode human affective states based on brain data and more particularly on event-related potentials (ERPs). ERPs are transient components in the EEG generated in response to a stimulus (a visual or auditory stimulus, for example). We tested six standard clustering algorithms (K-means, X-means, Farthest First, fuzzy-C-means, expectation–maximization and hierarchical clustering) to distinguish affective valences encoded into the ERPs collected while subjects were viewing high arousal images with positive or negative emotional content.

Our contribution consists in providing a framework for selection of a suitable low-dimensional feature space based on which standard clustering techniques can discriminate emotion states across multiple subjects.

This work was also inspired by advances in experimental psychology [16, 17] that proved a clear relation between ERPs and visual stimuli with underlined negative content (images with fearful and disgusted faces). Thus, the main focus of the present study was to discover which spatial–temporal patterns (features) in the ERPs indicate that a subject is exposed to stimuli that induce emotions. We apply the correlation-based feature selection (CFS) [18] and the gain ratio (GR) feature evaluation [19] to minimize the number of the relevant spatial temporal patterns and come up with a small number of temporal features (three ERP amplitudes and latencies) that reveal to be sufficiently robust across different subjects.

The paper is organized as follows. In Sect. 2, we briefly describe the data set. The complete feature set extracted from the ERPs is commented in Sect. 3, and the feature reduction methods are discussed in Sect. 4. The standard clustering techniques used in this study are briefly discussed in Sect. 5, and finally in Sect. 6, we present the experimental results showing statistically significant intra- and inter-subject clustering accuracies for the examined ERP data set.

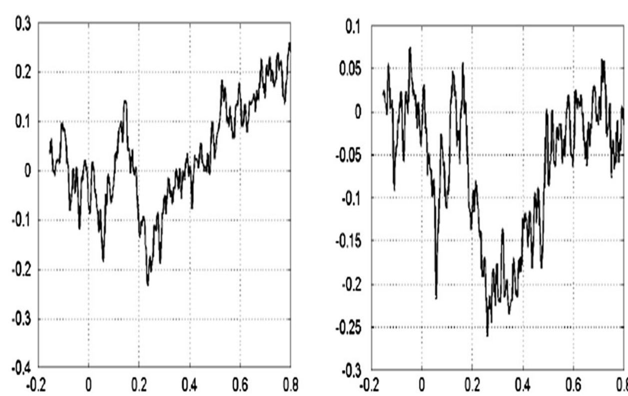
2 Data representation

A total of 26 female volunteers participated in the study. The signals were recorded while the volunteers were viewing high arousal images with positive and negative valence. For each image, signals from 21 EEG channels, positioned according to the 10–20 system, and 2 EOG channels (vertical and horizontal) were sampled at 1,000 Hz and stored. The signals were recorded while the volunteers were viewing pictures selected from the international affective picture system (IAPS) picture repository. A total of 24 of high arousal (IAPS rating >6) images with positive valence ($M = 7.29 \pm 0.65$) and negative valence ($M = 1.47 \pm 0.24$) were selected. Each image was presented 3 times in a pseudo-random order, and each trial lasted 3,500 ms: during the first 750 ms, a fixation cross was presented, then one of the images was presented during 500 ms and at last a black screen appeared during 2,250 ms. The raw EEG signals were first filtered, eye movement corrected, baseline compensated, and segmented into epochs using NeuroScan software. The single-trial signal length is 950 ms with 150 ms before the stimulus onset. The ensemble average for each condition (positive/negative valence) was also computed and filtered using a Butterworth filter of fourth order with passband 0.5–15 Hz. Thus, the filtered ensemble average signals cover the frequency band ranges corresponding to Delta (0.5–4 Hz), Theta (4–8 Hz), and Alpha neural activity (8–12 Hz), see Fig. 1.

The clustering approach we propose consists of three modules: feature extraction, feature reduction, and cluster analysis.

3 Feature space

Many ERP-based affective state detection systems rely on the frequency content of the signals. The features are some



(a) positive valence condition

(b) negative valence condition

Fig. 1 Example of filtered and ensemble-averaged ERP of one channel. **a** Positive valence condition, **b** negative valence condition measures of the ERP energy in certain frequency bands, for example PSD or spectral power asymmetry.

However, a number of studies [3, 16, 20] report evidences about temporal correlates between the processing of emotional stimuli and the occurrence of subsequent positive and negative potentials, known as waves C1, P1, P2, P3, when the stimulus carry very intensive emotions. These waves are associated with specific time of occurrence (early/late waves). For example. the wave C1 has been considered the first response of the primary visual cortex to a stimulus which occurs 60–90 ms after the stimulus onset. It is more expressed in the occipital channel locations. The positive evoked potentials between 100 and 200 ms (wave P1) respond to the valence and the physical characteristics of the stimulus.

Therefore, in the present study, temporal features (amplitudes and latencies) are extracted from the filtered, segmented, and ensemble-averaged ERP data set (Fig. 1). Starting by the localization of the first minimum after *time* = 0 s, the features are defined as a sequence of the local positive and negative picks, and their respective latencies (time of occurrence), see Fig. 2. Twelve temporal features are stored (Table 1) corresponding to the amplitudes of the first three minimums (A_{min1}, A_{min2}, A_{min3}), the first three maximums (A_{max1}, A_{max2}, A_{max3}), and their associated latencies (L_{min1}, L_{min2}, L_{min3}, L_{max1}, L_{max2}, L_{max3}). When this pattern is not found, the feature vector is filled with zeros. As a result, the starting feature data set consists of two matrices corresponding to the positive and negative emotion valences with three dimensions: 21 channels × 12 features × 26 subjects.

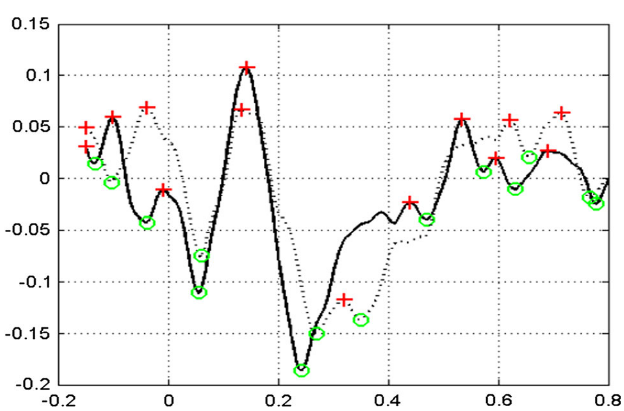


Fig. 2 Filtered signal and extracted features: positive (*line*) and negative (*dot*) valence conditions

Table 1 Complete set of temporal features

#	Feature name	Mean and SD (s)	#	Feature name	Mean and SD (s)
1	A_{min1}	−0.275 ± 0.145	7	L_{min1}	0.077 ± 0.031
2	A_{max1}	0.066 ± 0.165	8	L_{max1}	0.155 ± 0.054
3	A_{min2}	−0.121 ± 0.148	9	L_{min2}	0.220 ± 0.08
4	A_{max2}	0.124 ± 0.142	10	L_{max2}	0.311 ± 0.099
5	A_{min3}	−0.042 ± 0.121	11	L_{min3}	0.398 ± 0.123
6	A_{max3}	0.119 ± 0.130	12	L_{max3}	0.486 ± 0.130

The construction of the feature vectors is not unique. Typically, spatial (channels) and temporal (amplitudes, latencies) or frequency domain (energy) features are considered as one point (sample) in the feature space. This approach leads to a high-dimensional feature space that naturally requires more data. For example, if the temporal features (Table 1) of all channels are considered as one feature vector, the feature space will consist of 252 features (21 channels × 12 features). In this case, besides the necessity of feature reduction, more and longer experiments will be required for data analysis.

We propose here an alternative structure of the ERP feature matrix, where the features of each channel are considered as one sample; thus, we have 42 samples (21 channels × 2 classes) for the intra-subject setting and 1,092 samples (21 channels × 2 classes × 26 subjects) for the inter-subject setting. First, this approach automatically increases the number of samples, but what we found as more beneficial in the experiments was that clustering based on the majority vote (the assigned class is the one that has been chosen by more than 50 % of the channels) makes the learning feasible.

Though the proposed learning data structure is more balanced (less features, more examples, higher degree of freedom), it still has a low ratio between features (12) and samples (42); therefore, further feature reduction is recommendable.

4 Feature reduction

Feature reduction is the process of selecting a relevant subset from the initial feature space. Normally, it is applied to a high-dimensional feature space and a small number of feature vectors (clustering samples). Bad clustering is often due to irrelevant/noisy features or feature redundancy [19].

There are two main groups of feature selection techniques. Model-based methods (Wrapper and Embedded methods), where the trained model is validated against a test set and the model evaluation is with respect to the error, and filter methods (e.g., principal component analysis, correlation-based feature selection) that explore the statistical properties of the data without assuming a particular data model. Common performance measures of the filter methods are the mutual information, correlation coefficient, and inter/intra class distance.

We consider the filter methods as more appropriate for the particular data set, because a data model assumption is avoided. Besides, the model-based methods normally contain the feature selection as part of the learning algorithm, which is less convenient for the explorative study in hand. The following feature reduction filters were applied to the initial feature space.

4.1 Correlation-based feature selection

The *correlation-based feature selection* (*CFS*) method evaluates subsets of features on the basis of the principle that good feature subsets contain features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other [18]. In this study, we use Pearson correlation index to compute r_{Z_i} in order to grade the goodness of the attributes to form a representative subset Z of features:

$$r_{Z_i} = \frac{N\sum_{j=1}^{N} f_{ij}cl_j - \sum_{j=1}^{N} f_{ij}\sum_{j=1}^{N} cl_j}{\sqrt{\left[N\sum_{j=1}^{N} f_{ij}^2 - \left(\sum_{j=1}^{N} f_{ij}\right)^2\right]}\sqrt{\left[N\sum_{j=1}^{N} cl_j^2 - \left(\sum_{j=1}^{N} cl_j\right)^2\right]}}$$

where $f_i = [f_{i1}, \ldots, f_{ij}, \ldots, f_{iN}]$is the vector of the ith attribute, $cl = [cl_1, \ldots, cl_j, \ldots, cl_N]$ is the class vector, and N is the number of samples in the training data set.

The ratio r_{Z_i}estimates the correlation between ith feature vector and the class vector. The higher the value of r_{Z_i}, the greater the correlation between the attribute and the class. The attributes with the highest values of r_{Z_i} are chosen to form the subset Z.

4.2 Gain ratio feature evaluation

Gain ratio (GR) feature evaluation is based on entropy metric [19]. It computes how much information one feature brings to the knowledge of the classes for the training data set S. We have implemented *GR* in previous related works [15] and [21].

CFS and *GR* were implemented by Java Machine Learning Library [22] and WEKA software platform [23]. The feature vectors were first normalized in the interval [−1 1]. Feature reduction was performed for each channel. The results are summarized in Table 2. The numbers in the brackets correspond to the features with highest importance for the respective channel across all subjects. Numbers from 1 to 6 correspond to min/max amplitudes, and numbers from 7 to 12 to their respective latencies. Note that when CFS is applied, features 3 (A_{min2}), 7(L_{min1}), and 10 (L_{max2}) are selected in most channels, and when GR is applied, features 1 (A_{min1}), 3 (A_{min2}), and 10 (L_{max2}) are most frequently selected. Detailed observation of Table 2 reveals that features with longer latency (5, 6, 11, 12) are less frequently selected. This matches closely experimental psychology hypothesis related to temporal dynamics of emotions. According to [24], early waves carry more information about the valence than the arousal of the emotion. Therefore, late waves are less discriminative with respect to the emotion valence.

Table 2 Channel by channel feature reduction

Data channels	CFS method selected features	GR method	
		Selected features	Score
All chs.	[3, 7, 10]	[7, 10, 11]	0.012
Ch 1 (FP1)	[1, 7, 8]	[1, 2, 9]	0.073
Ch 2 (FPz)	[2, 4, 10]	[1, 3, 10]	0.071
Ch 3 (FP2)	[1, 5, 11]	[2, 3, 11]	0.074
Ch 4 (F7)	[1, 3, 7]	[1, 8, 10]	0.070
Ch5 (F3	[3, 4, 10]	[1, 8, 9]	0.080
Ch 6 (Fz)	[3, 7, 10]	[2, 4, 6]	0.081
Ch 7 (F4)	[7, 9, 10]	[4, 7, 9,]	0.060
Ch 8 (F8)	[4, 10, 11]	[1, 4, 9]	0.093
Ch 9 (T7)	[2, 3, 6]	[9, 10, 11]	0.090
Ch 10 (C3)	[3, 6, 10]	[1, 2, 11]	0.093
Ch 11 (Cz)	[3, 6, 10]	[1, 7, 11]	0.086
Ch 12 (C4)	[2, 3, 8]	[1, 8, 11]	0.074
Ch 13 (T8)	[2, 4, 7]	[1, 3, 8]	0.084
Ch 14 (P7)	[1, 4, 10]	[1, 5, 10]	0.107
Ch 15 (P3)	[3, 7, 10]	[1, 10, 11]	0.092
Ch 16 (Pz)	[3, 7, 10]	[1, 3, 11]	0.098
Ch 17 (P4)	[3, 10, 11]	[1, 3, 6]	0.091
Ch 18 (P8)	[4, 8, 10]	[1, 4, 6]	0.089
Ch 19 (O1)	[1, 3, 6]	[7, 9, 10]	0.100
Ch 20 (Oz)	[8, 9, 10]	[7, 9, 10]	0.090
Ch 21 (O2)	[7, 8, 10]	[1, 3, 7]	0.075

The column Score of the GR method represents the impact of the respective channel on the class distribution when the decision is made based on the selected features subset. It is worth noting that the spatial location of features, with stronger impact on the emotional valence, occur in the parietal and occipital regions of the brain (channels P3, Pz, P4, P7, O1, Oz), which fits brain-related theories.

5 Clustering techniques

During the clustering, the selected features, at the previous step, are partitioned into groups of similar observations. Each cluster has to cover the observations related to one emotion state. In the adopted terminology, Cluster 1 needs to accommodate observations conditioned by state P (Positive valence) and Cluster 2 to observations conditioned by state N (negative valence). Six clustering techniques, covering the whole spectrum of clustering paradigms, are included in the present study.

1. *K-means* (KM) clustering aims to partition the feature vectors (samples) into k clusters (initially defined by the user) in which each sample belongs to the cluster with the nearest mean, serving as a prototype of the cluster.
2. *X-means* (XM) is a variation of K-means where the number of clusters is not defined initially, but estimated at each k-means iteration.
3. *Farthest first* (FF) is also a variation of K-means that places each cluster center at the point furthest from the existing cluster centers. This point must lie within the data area. In most cases, FF speeds up the clustering since less reassignment and adjustment are needed.
4. *Fuzzy-C-means* (FCM) algorithm [25] is a variation of K-means in a probabilistic sense. The clusters are described by their center, which is a point in the data space that is most representative for the cluster in probabilistic sense. Every point of the data space belongs to distinct clusters with degree of membership, which is a value between 0 and 1. If the data are close to the cluster center, the membership degree is closer to 1. FCM aims to optimize the following objective function:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} d_{ik}^{2}, \tag{1}$$

where N is the number of the samples; c is the number of clusters; u_{ik} denotes the membership degree of data point x_k, $k = 1,\ldots,N$, and d_{ik} is the Euclidean distance of x_k, to the ith cluster center, $i = 1,\ldots,c$. The coefficient $m \in [1,\infty)$ determines how much clusters may overlap. In this study, $c = 2$ and $m = 2$.

FCM is particularly useful for not well-separated data groups with vague and uncertain boundaries.

5. *Expectation–maximization (EM)* algorithm. Data are represented as a combination of models with different probability distributions; mixture of Gaussian distributions is assumed in this study. EM allows clusters to have different shapes in contrast to k-means, which tends to find clusters of comparable spatial extent defined by the Euclidian distance in the feature space. Gaussian mixture models trained with EM algorithm maintain probabilistic instead of deterministic assignments to clusters.
6. *Hierarchical clustering (HC)* builds a hierarchical tree of clusters, with different level of granularity. The tree is further subdivided into agglomerative (bottom–up) merging small clusters into larger ones and divisive (top–down) splitting large clusters.

KM, XM, and FF belong to the deterministic partitioning clustering; FCM and EM explicitly deal with data uncertainty and ambiguity; and HC represents the hierarchical approach of clustering.

6 Clustering results

The clustering feasibility is assessed with respect to both intra-subject (single subjects) and inter-subject (across all participants in the data set) settings.

6.1 Intra-subject clustering with fuzzy-C-means (FCM)

The independent clustering for each subject is performed by FCM. Based on the feature selection results in Table 2 and taking into account that features with short latency (early waves) are more influenced by the valence of the stimulus, A_{min2}, L_{min1}, L_{max2} are selected as inputs for the clustering step.

6.1.1 Discrimination of P state versus N state

Table 3 is the integrated confusion matrix for all participants in the experiment. Columns Score reflect the rate of correctly clustered data, taking into account that Cluster 1 has to cover observations conditioned by state P and Cluster 2—observations conditioned by state N.

The results clearly show better discrimination rate of state P compared with state N, though the data set is balanced (21 observations for each state). For 23 subjects (out of 26 subjects in total), more than 50 % of the channels associated correctly Cluster 1 with the state P, see column 4 of Table 3. Following the majority vote, the observations of the complete set of channels will be correctly clustered. As with respect to Cluster 2, only for 12 subjects, the clustering will be correct, see column 7 of Table 3. Correct clustering of both states P and N is achieved only for 9 subjects.

The interesting conclusion drawn from this experiment is that the selected features are suitable to discriminate well the P state (88 % rate of correct clustering) and less suitable to discriminate the N state (46 % rate of correct clustering). Detailed observation of Table 1 shows that two of the selected features (A_{min2}, L_{max2}) correspond to temporal window of 200–400 ms after the stimulus. According to [21], during this period, the brain response to images with highly positive valence is stronger. This may explain the higher discrimination rate for the state P.

6.2 Inter-subject clustering with fuzzy-C-means (FCM)

Inter-subject clustering is a challenging task due to the high variability among subjects. Different people may generate

Table 3 Intra-subject FCM clustering

Person	Cluster 1 (accommodates state P)			Cluster 2 (accommodates state N)		
	State P	State N	Score %	State P	State N	Score %
1	14	12	66.67	7	9	42.86
2	**14**	**8**	**66.67**	**7**	**13**	**61.91**
3	**16**	**8**	**76.19**	**5**	**13**	**61.91**
4	**13**	**5**	**61.91**	**8**	**16**	**76.19**
5	15	13	71.43	6	8	38.1
6	20	16	95.24	1	5	23.81
7	15	11	71.43	6	10	47.62
8	6	0	28.57	15	21	100
9	16	16	76.19	5	5	23.81
10	17	17	80.95	4	4	19.05
11	18	11	85.71	3	10	47.62
12	8	6	38.1	13	15	71.43
13	**19**	**9**	**90.48**	**2**	**12**	**57.14**
14	17	16	80.95	4	5	23.81
15	13	13	61.91	8	8	38.1
16	6	0	28.57	15	21	100
17	13	12	61.91	8	9	42.86
18	**17**	**9**	**80.95**	**4**	**12**	**57.14**
19	12	11	57.14	9	10	47.62
20	17	11	80.95	4	10	47.62
21	**21**	**6**	**100**	**0**	**15**	**71.43**
22	12	11	57.14	9	10	47.62
23	**20**	**10**	**95.24**	**1**	**11**	**52.38**
24	**18**	**10**	**85.71**	**3**	**11**	**52.38**
25	**19**	**8**	**90.48**	**2**	**13**	**61.91**
26	16	12	76.19	5	9	42.86

Best results are highlighted in bold

different brain activation given the same emotion state. If common representative models of the affective states across multiple subjects can be learned, they can be used as software sensors to detect emotional states of subjects outside the training set. This problem is relevant to diagnosis of mental processes in patients with brain injuries or to human personality categorization (e.g., discriminate high versus low neuroticism).

The results of FCM clustering are summarized in Table 4 that represent the integrated confusion matrix for all channels.

6.2.1 *Discrimination of P state versus N state*

It is interesting to see that analogous to the inter-subject FCT clustering (Table 3), the state P is again better discriminated than the state N. The best recognition rate is above 80 %. The physical interpretation of the results says for example that the positive valence emotion of 21 subjects (out of 26 in the pool) may be detected only by features extracted from channel F3 alone. As for the state N, the same tendency of less efficient detection is observed. The best discrimination rate for state N is 65 % obtained by clustering features only from channel P8.

Table 4 Inter-subject FCM clustering

Channel	Cluster 1			Cluster 2		
	State P	State N	Score %	State P	State N	Score %
1(FP1)	16	13	61.54	10	13	50
2(FPz)	15	13	57.69	11	13	50
3(FP2)	17	15	65.39	9	11	42.31
4(F7)	18	15	69.23	8	11	42.31
5(F3)	21	16	80.77	5	10	38.46
6(Fz)	**19**	**11**	**73.08**	**7**	**15**	**57.69**
7(F4)	17	14	65.39	9	12	46.15
8(F8)	**16**	**12**	**61.54**	**10**	**14**	**53.85**
9(T7)	15	14	57.69	11	12	46.15
10(C3)	19	13	73.08	7	13	50
11(Cz)	19	16	73.08	7	10	38.46
12(C4)	**18**	**12**	**69.23**	**8**	**14**	**53.85**
13(T8)	**14**	**10**	**53.85**	**12**	**16**	**61.54**
14(P7)	14	14	53.85	12	12	46.15
15(P3)	18	13	69.23	8	13	50
16(Pz)	18	13	70.23	8	13	50
17(P4)	17	13	65.39	9	13	50
18(P8)	**14**	**9**	**53.85**	**12**	**17**	**65.39**
19(O1)	12	10	46.15	14	16	61.54
20(Oz)	13	13	50	13	13	50
21(O2)	16	13	61.54	10	13	50

Best results are highlighted in bold

The results of the two experiments suggest that the P and N emotional states are not correlated with the same features and that people with different personality and psychological characteristics may generate different neuronal activity given the same emotional stimulus.

6.2.2 *Channel discrimination capabilities*

The channel by channel clustering framework facilitates the analysis of the importance of the spatial location of the features. Table 4 shows that the central channels (in particular Fz, Cz and Pz) provide higher clustering accuracy (more than 70 %) for state P. This is in line with the findings in [26, 27] that the stimulus valence is correlated with stronger response in the central cortical line.

6.3 Inter-subject clustering with HC, EM, KM, ZM, and FF

In this section, we analyze the impact of the feature reduction on the global performance of five clustering techniques (HC, EM, KM, ZM, FF) described in Sect. 5.

Besides CFS and GR feature reduction techniques, clustering without feature reduction was also studied. The results, summarized in Table 5, reveal interesting correlations.

As it was expected, both feature reduction techniques benefit the clustering. Clustering not preceded by feature reduction step is not far from a random binary guess (50 %).

Clustering based on CFS selected features (A_{min2}, L_{min1}, L_{max2}) exhibits better accuracy compared with clustering based on GR selected features (A_{min1}, A_{min2}, L_{max2}).

KM and its close variation XM outperform the other techniques and have a good clustering rate (up to 78 % for channel Oz), which is close to the results obtained in the previous section with FCM. This is not unexpected because FCM, KM, and XM explore the statistical characteristics of the data, in contrast to the EM, where the assumption of mixture of Gaussian distributions most probably does not fit well the data in hand. The Hierarchical Clustering does not bring any advantages in this study.

As with the FCM inter-subject clustering, the central channels (in particular Cz, Pz and Oz) provide better clustering accuracy (above 65 %), which is due to the stronger correlation between the stimulus valence and the central cortical zones [26].

7 Conclusions

In this paper, we propose an alternative approach to the challenging problem of human emotion recognition based on brain data. In contrast to most of the recognition systems where supervised learning is used, we apply standard clustering techniques to distinguish the processing of stimuli with positive and negative emotion valence based on ERPs observations. Additionally, we explored the feasibility of training cross-subject clusters to make predictions across multiple human subjects. The core of the present study is the way the features are selected. The combination of a small number of time domain (ERP amplitudes and latencies) and spatial (selected channels) features has the potential to reduce the inter-subject variability and improve the learning of representative models valid across multiple subjects.

Clustering techniques are widely applied in several application areas; however, they are still less known in the framework of affective computing. Based on the results of this study, we believe that the ERP clustering (and particularly the FCM) is a promising approach to extract statistical underlying correlations of the brain activity among subjects and therefore decode human emotional states. Nevertheless, before making stronger conclusions on the

Table 5 Inter-subject clustering

Feature selection clustering channels	None					Correlation FS					Gain ratio				
	HC	EM	KM	XM	FF	HC	EM	KM	XM	FF	HC	EM	KM	XM	FF
1(FP1)	51	61	63	61	51	51	51	63	**67**	51	51	57	61	61	61
2(FPz)	51	49	51	51	51	51	57	57	57	51	51	49	55	49	51
3(FP2)	51	53	55	55	51	49	55	53	53	49	51	59	51	57	51
4(F7)	49	55	53	53	49	51	53	51	55	55	51	55	55	55	59
5(F3)	51	51	51	51	51	51	61	59	59	53	51	51	49	51	53
6(Fz)	51	49	59	53	53	51	59	**65**	61	51	51	59	55	53	51
7(F4)	51	53	61	53	59	51	51	55	57	51	51	57	57	49	61
8(F8)	51	53	53	49	51	51	51	53	49	51	51	53	53	49	53
9(T7)	51	49	49	49	53	51	61	63	63	55	51	51	49	49	59
10(C3)	51	51	53	53	53	51	63	51	51	49	53	63	61	63	51
11(Cz)	51	49	59	49	51	51	**69**	**72**	**71**	57	51	**65**	**66**	**65**	49
12(C4)	51	51	51	51	51	53	57	57	57	61	51	53	55	55	53
13(T8)	51	51	51	51	53	51	63	**68**	**67**	55	51	49	51	51	51
14(P7)	51	53	55	53	53	51	57	63	63	61	51	57	49	57	51
15(P3)	51	49	49	49	53	51	55	55	57	59	51	51	53	53	51
16(Pz)	53	53	53	49	53	53	57	**73**	59	**67**	53	53	61	57	**65**
17(P4)	51	53	51	53	51	51	53	**65**	**65**	55	51	53	**65**	**65**	55
18(P8)	51	51	53	53	49	53	53	53	53	53	51	51	51	51	49
19(O1)	53	55	53	57	55	51	51	59	53	57	53	55	55	51	55
20(Oz)	53	63	61	61	**65**	53	64	**67**	**67**	**69**	53	**67**	**78**	**78**	61
21(O2)	51	49	51	51	49	51	49	53	59	49	49	49	53	53	57

Best results are highlighted in bold

capacity of unsupervised learning to decode emotions, further research is required to answer more challenging questions such as discrimination of more than two emotions. In fact, this is a valid question for all reported works on affective neuroscience [6]. The discrimination is usually limited to two, three, and maximum four valence-arousal emotional classes. An interesting problem is also the human personality clusterization based on EEG, for example distinguishing between high versus low neurotic type of personality.

Also, the number of participants in the experiments is important for revealing stable cross-subject features. In the reviewed references, the average number of participants is about 10–15 and the maximum is 32. We need publicly available data sets to compare different techniques and thus speed up the progress of affective computing.

Acknowledgments The work was supported partially by the project "Data mining methods for development and assessment of software services," 2014 of Science Fund of SU "Kl. Ohridski." It was also supported by the bilateral student Erasmus program between University of Aveiro, Portugal, and Sofia University "St. Kl. Ohridski," Bulgaria.

References

1. Calvo RA, D'Mello SK (2010) Affect detection: an interdisciplinary review of models, methods, and their applications. IEEE Trans Affect Comput 1(1):18–37
2. Dalgleish T, Dunn B, Mobbs D (2009) Affective neuroscience: past, present, and future. Emot Rev 1:355–368
3. Olofsson JK, Nordin S, Sequeira H, Polich J (2008) Affective picture processing: an integrative review of ERP findings. Biol Psychol 77:247–265
4. AlZoubi O, Calvo RA, Stevens RH (2009) Classification of EEG for emotion recognition: an adaptive approach. In: Proceedings of the 22nd Australasian joint conference. Artificial intelligence, pp 52–61
5. Petrantonakis PC, Hadjileontiadis LJ (2010) Emotion recognition from EEC using higher order crossings. IEEE Trans Inf Technol Biomed 14(2):186–194
6. Jatupaiboon N, Panngum S, Israsena P (2013) Real-time EEG-based happiness detection system. Sci World J 2013. Article ID 618649, 12 p
7. Lin YP, Wang CH, Wu TL, Jeng SK, Chen JH (2008) Support vector machine for EEG signal classification during listening to emotional music. In: Proceedings of the 10th IEEE workshop on multimedia signal processing, (MMSP'08), 127–130, Cairns, Australia, Oct 2008
8. Li M, Lu B-L (2009) Emotion classification based on gamma-band EEG. 31st annual international conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2–6, 2009, pp 1323–1326
9. Nie D, Wang X-W, Shi L-C, Lu B-L (2011) EEG-based emotion recognition during watching movies. Proceedings of the 5th international IEEE EMBS conference on neural engineering, Cancun, Mexico, pp 667–670, April 27–May 1, 2011
10. Chanel G, Kronegg J, Grandjean D, Pun T (2006) Emotion assessment: arousal evaluation using EEG's and peripheral physiological signals. In: Gunsel B, Jain A, Tekalp AM, Sankur B (eds) Multimedia content representation, classification and security, vol 4105. Springer, Berlin, pp 530–537
11. Frantzidis Ch A, Bratsas Ch, Klados MA, Konstantinidis E, Lithari ChD, Vivas AB, Papadelis Ch L, Kaldoudi E, Pappas C, Bamidis PD (2010) On the classification of emotional biosignals evoked while viewing affective pictures: an integrated data-mining-based approach for healthcare applications. IEEE Trans Inf Technol Biomed 14(2):309
12. Bos D (2007) EEG-based emotion recognition. http://hmi.ewiutwente.nl/verslagen/capita-selecta/CS-OudeBos-Danny.pdf
13. Tomé AM, Hidalgo-Munoz AR, Pérez ML, Teixeira AR, Santos IM, Pereira AT, Vázquez-Marrufo M, Lang EW (2013) Feature extraction and classification of biosignals emotion valence detection from EEG signals. BIOSIGNALS 2013, international conference on bio-inspired systems and signal processing, Barcelona, February 2013
14. Liu Y, Sourina O, Nguyen MK (2010) Real-time EEG-based human emotion recognition and visualization. In: Proceedings of the international conference on cyberworlds (CW'10), pp 262–269, Singapore, October 2010
15. Georgieva O, Milanov S, Georgieva P (2013) Cluster analysis for EEG biosignal discrimination. IEEE international symposium on innovations in intelligent systems and applications INISTA, Albena, Bulgaria, 19–21 June 2013
16. Santos IM, Iglesias J, Olivares EI, Young AW (2008) Differential effects of object-based attention on evoked potentials to fearful and disgusted faces. Neuropsychologia 46:1468–1479
17. Pourtois G, Grandjean D, Sander D, Vuilleumier P (2004) Electrophysiological correlates of rapid spatial orienting towards fearful faces. Cereb Cortex 14(6):619–633
18. Hall M (1999) Correlation-based feature selection for machine learning. PhD thesis, Department of Computer Science, University of Waikato, New Zealand
19. Ladha L, Deepa T (2011) Feature selection methods and algorithms. Int J Comput Sci Eng (IJCSE) 3(5):1787–1797
20. Stolarova M, Keil A, Moratti S (2006) Modulation of the C1 visual event-related component by conditioned stimuli: evidence for sensory plasticity in early affective perception. Cereb Cortex 16:876–887
21. Milanov S, Georgieva O, Georgieva P (2013) Comparative analysis of brain data clustering. In: Proceedings of doctoral conference in mathematics, informatics and education, Sofia, Bulgaria, pp 94–101, 19–29 September
22. Java Machine Learning Library. http://java-ml.sourceforge.net/
23. Waikato Environment for Knowledge Analysis (WEKA). http://weka.wikispaces.com/
24. Gianotti LRR, Faber PL, Schuler M, Pascual-Marqui RD, Kochi K, Lehmann D (2008) First valence, then arousal: the temporal dynamics of brain electric activity evoked by emotional stimuli. Brain Topogr 20:143–156
25. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York
26. Cuthbert BN, Schupp HT, Bradley MM, Birbaumer N, Lang PJ (2000) Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. Biol Psychol 52:95–111
27. Karegowda AG, Manjunath AS, Jayaram MA (2010) Comparative study of attribute selection using gain ratio and correlation based feature selection. Int J Inf Technol Knowl Manag 2(2):271–277