EFFECT-SIZE-BASED ELECTRODE AND FEATURE SELECTION FOR EMOTION RECOGNITION FROM EEG

Robert Jenke, Angelika Peer, and Martin Buss

Institute of Automatic Control Engineering, Technische Universität München

ABSTRACT

Emotion recognition from EEG signals allows the direct assessment of the "inner" state of the user which is considered an important factor in Human-Machine-Interaction. Given the vast amount of possible features from scalp recordings and the high variance between subjects, a major challenge is to select electrodes and features that separate classes well. In most cases, this decision is made based on neuro-scientific knowledge. We propose a statistically-motivated electrode/feature selection procedure, based on Cohen's effect size f^2. We compare inter- and intra-individual selection on a self-recorded database. Classification is evaluated using quadratic discriminant analysis (QDA). We found both feature selection versions based on f^2 yield comparable results. While highest accuracies up to $57{,}5\%$ (5 classes) are reached by applying intra-individual selection, inter-individual analysis successfully finds features that perform with lower variance in recognition rates across subjects than combinations of electrodes/features suggested in literature.

Index Terms— Emotion Recognition, EEG, Feature Selection, Machine Learning.

1. INTRODUCTION

To make Human-Machine-Interaction (HMI) more natural, knowledge about the emotional state of the user is considered an important factor. Emotions are important for both correct interpretation of actions, as well as communication. Beyond its application in HMI, research in emotion recognition can lead to a deeper understanding of emotion mechanisms themselves.

The field of emotion recognition from Electroencephalography (EEG) signals is comparatively new, but is gaining more and more attention in recent years. Bos investigated the influence of different stimuli modalities on extreme classes of the valence/arousal plane using two frequency-based features for linear classification [1]. Schaaff and Schultz compared two sets of features recorded with an in-house developed headband [2]. Classifying three emotional states using support vector machine (SVM), they reached a mean accuracy of 44 and 49%. Li and Lu found gamma-band features to separate two classes (happy and sad) well using common spatial patterns and SVM. Recently, Murugappan et al. published work on a collected dataset from 64 channels over 5 subjects and 4 emotions. They give an overview on earlier work in the field and introduce wavelet domain features [3].

A major challenge in emotion recognition from EEG signals relates to interpersonal variance in both emotion induction and recognition. It is not in general agreed upon which features are most appropriate. Different people show different emotional responses and thus, different features of the EEG data carry the best information for recognition. Additionally, the vast amount of possible features makes it necessary to reduce dimensions in order to avoid over-specification.

Typically, electrode selection is done on the basis of neuro-scientific assumptions. In contrast, we propose a statistically-motivated approach to select electrodes and features. In particular, Cohen's effect size f^2 is used as a measure of separability. This method can be interpreted as a univariate filter-method for feature selection [4]. To assess the generality of commonly used features and to possibly extend the set of electrodes that promise good classification accuracy on an inter-individual basis, we recorded a database from 16 subjects. We investigate differences in classification success rates between features selected based on statistical properties and those given in literature.

The remainder of the paper is organized as follows: the dataset (Sec. 2) and processing steps for artifact removal (Sec. 3) and feature extraction (Sec. 4) are introduced first. Section 5 describes the selection method. Classification and results are given in Sec. 6, followed by discussion and future work in Sec. 7.

2. AFFECTIVE DATABASE

In contrary to most investigations in literature, we aimed for a dataset with enough subjects to draw conclusions on an inter-individual basis. Thus, the recorded dataset consists of 16 subjects, each containing 8 trials of $30s$ EEG recording for 5 different emotions (happy, curious, angry, sad, quiet). A 64-channel EEG cap with g.tec gUSBamp was used for record-

This work is supported in part by the VERE project within the 7th Framework Programme of the European Union, FET- Human Computer Confluence Initiative, contract number ICT-2010-257695.

Table 1. Features typically extracted for emotion recognition from EEG are listed. mean, min, max, and var are used for frequency domain features. Activity, Mobility, and Complexity are known as *Hjorth Features*. Grayed rows are not implemented.

	Feature	Description	Commonly used electrodes	Electrodes from f^2
Frequency domain	Theta-Band	4-8Hz	left anterior sites [5] right posterior sites	Fz, Cz, FT9, F9, CP3, CP1, C5, P9, CP2, CP5
	Alpha-Band	8-12Hz	occipital sites Fp1, Fp2, F7, F8 [2] (for theta-, alpha-, beta-bands) F3, F4 [1]	
	Beta-Band	14-30Hz	F3, F4	
	Gamma-Band	43-68Hz	all electrodes [6]	F4, F8, F6, F2, Fz, FC2, FC6, TP9, CP1, F1, C3, CP3, TP7
	Magnitude Squares Coherence Estimate		P3, P4, T7, T8, C3, C4, F3, F4 [7]	
Time domain	Power of Signal	$P(\mathbf{x}) = \frac{1}{N}\sum_{-\infty}^{\infty} \lvert\mathbf{x}[n]\rvert^2$	main lobe [8]	
	Activity	$A(\mathbf{x}) = \frac{\sum_{n=1}^{N}(\mathbf{x}(n)-\mu)^2}{N}$	CP5, CP6, F3, F4, Afz [9]	
	Mobility	$M(\mathbf{x}) = \sqrt{\frac{\mathrm{var}(\dot{\mathbf{x}})}{\mathrm{var}(\mathbf{x})}}$	CP5, CP6, F3, F4, Afz Pz, P8, Fz, O1, F3 [10]	
	Complexity	$C(\mathbf{x}) = \frac{M(\dot{\mathbf{x}})}{M(\mathbf{x})}$	O2 CP5, CP6, F3, F4, Afz [9]	
	Wavelet domain		all electrodes [3]	

ing at 512Hz. Similar to Schaaff and Schulz [2], the emotions were induced using IAPS pictures [11]. Sets of 4 pre-selected pictures of one emotion were shown for $5s$ each in every trial. For validation of induction, we compared the results of SAM-Tests (Self-Assessment-Manikin Test [12]) with the targeted values of the presented picture sets and received an average correlation coefficient of $r = .545$. Excluding subjects with r-values lower than $.5$ (i.e. 4 subjects), the correlation coefficient increased to $r = .632$.

Pre-studies showed that it takes about 10s for an emotion to be induced using pictures and to last for about 4s. Thus, time intervals between 11-15s after emotion induction onset are considered in the following analysis.

3. ARTIFACT REMOVAL

As a preprocessing step, we investigated the benefits of removing artifacts from the signal. The Independent Component Analysis (ICA) based plug-in ADJUST [13] for the MATLAB Toolbox EEGLAB [14] was used to identify and remove different kinds of artifacts: eyeblinks, eye-movements, drift, and generic discontinuities. The algorithm first transforms EEG data into *Independent Components* (ICs). In the *Selection Process*, artifact-affected components are detected and discarded automatically by a set of rules and thresholds within ADJUST. Finally, the inverse transformation of the remaining components is computed. In addition to the default settings, we extended the detection algorithm to variable thresholds that are adjusted for each user so that at least 95% of the trials have a signal range between $\pm 100\mu V$ after removal, i.e. are within the range of a clean signal.

4. FEATURE EXTRACTION

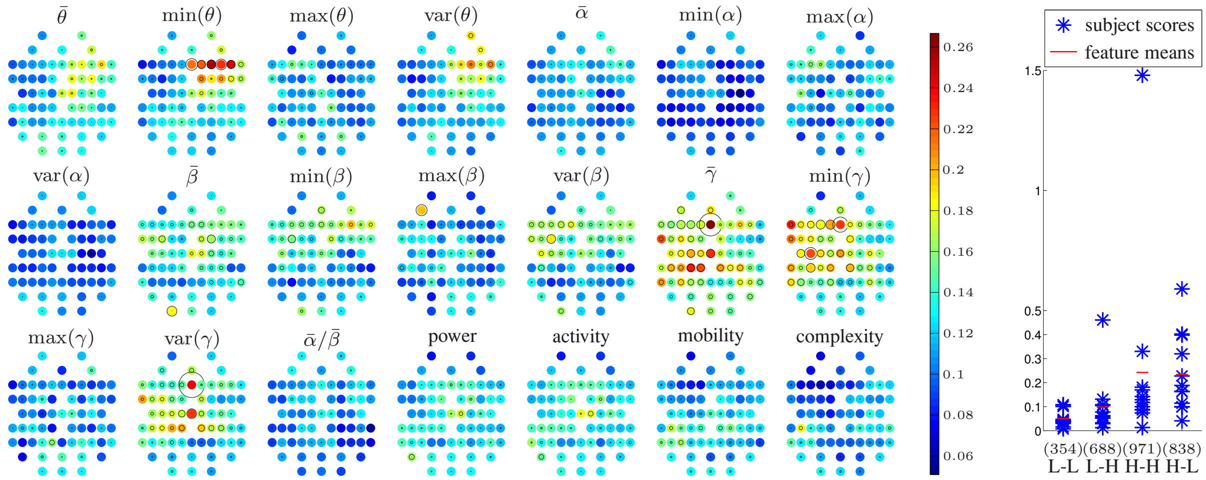
We implemented a variety of EEG features from literature to compare our approach to results obtained by those feature sets. Table 1 lists the considered features from related work. We extracted mean, min, max, and var from the frequency domain features, which are computed using FFT. Both frequency and time domain features were extracted from all 64 electrodes which results in a total of 1344 features. Features are z-normalized to zero mean and standard deviation equal to one. In the equations given in Table 1, $\mathbf{x} \in \mathbb{R}^N$ denotes the vector of the signal of a single electrode, N is the number of time-samples in $\mathbf{x}$, i.e. 2048 in our specific case. A feature of $\mathbf{x}$ is denoted as ξ.

5. ELECTRODE AND FEATURE SELECTION

For electrode and feature selection, we use Cohen's effect size f^2 which is a generalization to more than two classes of Cohen's $d = \left|\frac{\bar{\xi}_1 - \bar{\xi}_2}{\sigma}\right|$ used for the statistical t-test [15]. The spread of the means in the numerator is represented as a standard deviation σ_m. The denominator remains the pooled standard deviation σ of the populations involved. Thus,

$$f = \frac{\sigma_m}{\sigma} \text{ , where } \sigma_m = \sqrt{\frac{\sum_{i=1}^{c}(\bar{\xi}_i - \bar{\xi})^2}{c}} \qquad (1)$$

for equal sample sizes per class. Here, $\bar{\xi}_i$ and $\bar{\xi}$ are the mean of the samples belonging to class i and the overall mean of a feature, respectively. The number of classes is denoted by c, i.e. 5 in our case.



(a) (b)

Fig. 1. (a) Mean (colorbar) and variance (radius of black circle) of f^2 over all subjects; each plot corresponds to one feature, (b) Examples of features (feature identifier in parenthesis) showing low/high mean combined with low/high variance of f^2.

Wilk's Lambda Λ_i [16], which has also been applied for feature selection in a different domain, is directly related to the effect size: $f^2 = (1 - \Lambda_i)/\Lambda_i$. Yet, this measure is not as commonly reported as effect size. Further, significance or test statistics have been proposed for feature selection [4]. We argue, however, that effect size makes more sense since it does not depend on the sample size and is thus useful also for meta-studies.

We implemented two versions of feature selection using f^2 to measure separability of a feature: 1) inter-individual: f^2 is computed for each electrode per feature and is averaged over all 16 subjects for feature/electrode selection, 2) intra-individual: f^2 is computed and used to select electrodes and features individually for each subject.

Fig. 1(a) depicts the mean (color) and variance (radius of black circle) of the effect size of each electrode and feature averaged over all subjects and mapped to the electrode positions on the scalp. Considering scores individually, f^2 reaches values up to 1.47 while averages across subjects are around .2. This means, that in cases where mean and variance of f^2 are high (H-H), a feature is only suitable for few (or even one) subject (see Fig. 1(b)), but does not separate classes well for other subjects. Meanwhile, features scoring high f^2 values and low variance (H-L) are considered generally suitable across subjects. It should be noted that this causes a bias in the case of inter-individual selection. Using the median of f^2 instead would correct this deviation without completely ignoring the outliers. Tests with this method, however, resulted in comparable accuracies while requiring a larger number of features, which is why we kept the mean as inter-subject selection criteria (further discussion in Sec. 7).

In the right column of Table 1 we list the electrodes for each feature that exceed a threshold of .2 for the mean of f^2 measures. Notably, some features, e.g. the complete α-band, score very low values of f^2 for all electrodes.

6. CLASSIFICATION AND EVALUATION

To evaluate the proposed selection method, classification is performed by means of quadratic discriminant analysis (QDA) with diagonal covariance estimates (i.e. Naive Bayes). We use 8-fold cross validation, where 7 folds are used for feature selection and classifier training, the remaining fold (i.e. 1 sample from each class) is used for testing. In inter-individual analysis, f^2 measures are precomputed for each fold and averaged over all subjects, before the classifier is trained with the selected features for each subject individually. Given 5 classes, chance level is at 20%. We found the effect of artifact removal to be marginal and thus, it is not reported here for brevity. The study showed that satisfactory results can also be obtained without cleaning the data.

In Fig. 2 the results from inter-individual analysis are presented. We compared our approach to electrodes and features suggested in [9] (Feature Set FS1: Activity, Mobility, and Complexity from electrodes CP5, CP6, F3, F4, Afz) and [2] (FS2: $\bar{\alpha}$, $\max(\alpha)$, $\bar{\theta}$, $\bar{\beta}$ from electrodes Fp1, Fp2, F7, F8). To be fair, we looked at the same number of dimensions used, i.e. 15 and 16, respectively. FS2 reaches a maximum accuracy of 50% for one of the tested subjects compared to 45% for our proposed approach. However, electrodes and features selected by the effect size measure show generally higher accuracies (see last plot in Fig. 2) and, compared to FS2, smaller variance across subjects.

A comparison of the achieved classification accuracy for inter- and intra-individual selection is depicted in Fig. 3. When accounting for personal differences (intra-individual selection) in emotional responses, classification results improve especially for small numbers of features. Improvement was not so much pronounced for subject $1, 2, 4$, and 9, as for subject $5, 6, 8$, and 10. A maximum of $57{,}5\%$ accuracy is reached by subject 10 using only 2 out of the 1344 possible features. However, since the ideal number of features varies between subjects, average accuracy over all subjects showed only small improvements compared to inter-individual analysis (intra: $\approx 30\%$ vs. inter: $\approx 27\%$). Generally, the *curse of dimensionality* is clearly visible, as accuracy mostly decreases for an increasing number of features.

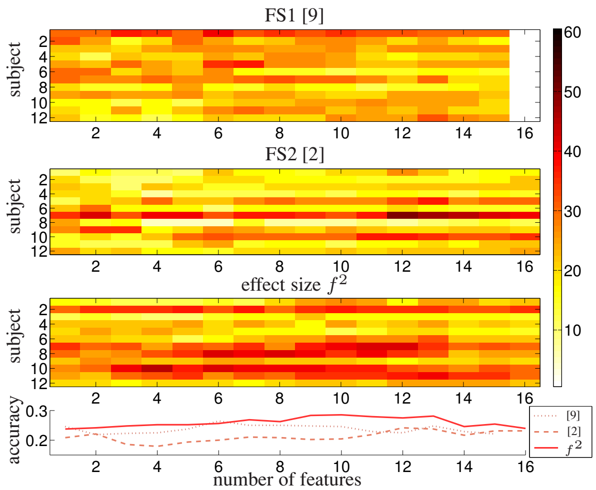


Fig. 2. Classification success rates (color) for 16 features evaluated for multiple subjects compared to those suggested in literature. The bottom graph shows the average accuracies.

7. DISCUSSION AND FUTURE WORK

This paper explored electrode and feature selection for EEG signals based on statistical properties. A univariate feature selection method using Cohen's effect size f^2 from analysis of variance was implemented to investigate automatic electrode and feature selection. The selection of electrodes and features from a statistical point of view proved to be a useful extension to merely relying on neuro-scientific findings. Electrodes and features found by this approach resulted in smaller variance of classification accuracies across subjects and generally higher accuracies when features were selected intra-individually compared to an inter-individual selection. The strong interpersonal variance of emotions is not always obvious from other studies, since many consider small datasets with one or very few subjects only (e.g. [9]). Notably, when comparing to traditionally used features and regions of the brain, our results suggest partially different, better separating features or regions.

Using the median instead of the mean to counter the issue of high variance in f^2 did not improve results. A possible reason for this observation is the likely interaction of features

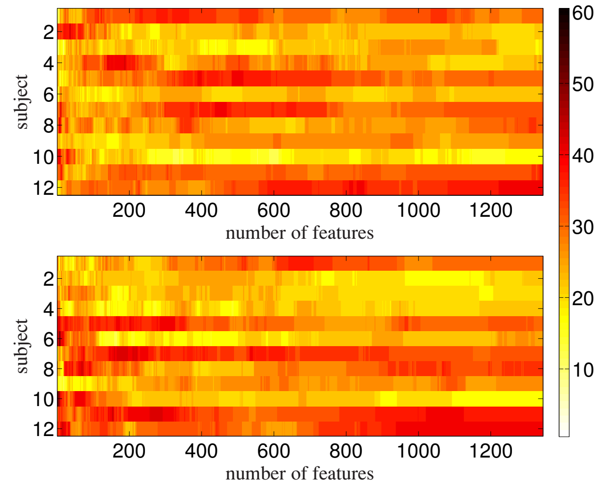


Fig. 3. Classification results based on inter-individual (above) and intra-individual (below) electrode and feature selection.

which is not considered by univariate methods. Hence, ongoing and future work will include the development and comparison of more advanced and multivariate feature selection methods, taking into account dependencies between features. Further, in order to make the recognized emotion computationally easy to process, methods that enable classification in continuous space (e.g. PAD model [17]) should be developed. We are currently also preparing to make this database publicly available to other researchers.

8. RELATION TO PRIOR WORK

The presented work applies automatic, statistically-motivated electrode and feature selection to emotion recognition from EEG signals on a self-recorded dataset with 16 participants and 5 different emotions. Work by Ansari-Asl et al. reports results of the application of synchronization likelihood for electrode selection which was tested on one subject only [9]. Recently, Kroupi et al. did a study on EEG correlates of emotional states labeled in continuous space [18]. Different features were correlated with a self-assessed emotional measure. In agreement with our results, their subject-dependent analysis revealed strong interpersonal differences in the brain activation patterns. Similarly, Li and Lu noticed these differences across subjects, when applying a wrapper method to find optimal frequency bands [6]. Their approach, however, did not consider simultaneous selection of features and electrodes. The presented approach can be extended by including more complex features like Higher Order Crossings as suggested by Petrantonakis and Hadjileontiadis [19]. Finally, Deriche introduced a feature selection method specifically for EEG analysis based on maximizing mutual information [8]. This method has not been applied to emotion recognition, yet, and would be a promising method for future work.

9. REFERENCES

[1] Danny Oude Bos, "EEG-based Emotion Recognition," *Emotion*, vol. 57, no. 7, pp. 1798–806, 2006.

[2] Kristina Schaaff and Tanja Schultz, "Towards emotion recognition from electroencephalographic signals," in *Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*, Sept. 2009, pp. 1–6.

[3] M Murugappan, M. Rizon, R. Nagarajan, and Sazali Yaacob, "Inferring of Human Emotional States using Multichannel EEG," *European Journal of Scientific Research*, vol. 48, no. 2, pp. 281–299, 2010.

[4] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga, "A review of feature selection techniques in bioinformatics.," *Bioinformatics*, vol. 23, no. 19, pp. 2507–17, Oct. 2007.

[5] Gary Garcia Molina, Tsvetomira Tsoneva, and Anton Nijholt, "Emotional Brain-Computer Interfaces," in *Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2009, pp. 138–146.

[6] Mu Li and Bao-Liang Lu, "Emotion classification based on gamma-band EEG," in *IEEE Int. Conf. Engineering in Medicine and Biology Society*, Minneapolis, MN, Sept. 2009, vol. 1, pp. 1323–1326.

[7] Reza Khosrowabadi, Hiok Chai Quek, Abdul Wahab, and Kai Keng Ang, "EEG-based Emotion Recognition Using Self-Organizing Map for Boundary Detection," in *IEEE Int. Conf. on Pattern Recognition*, Aug. 2010, pp. 4242–4245.

[8] Mohamed Deriche and Ahmed Al-Ani, "A new algorithm for EEG feature selection using mutual information," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2001, pp. 1057–1060.

[9] Karim Ansari-asl, Guillaume Chanel, and Thierry Pun, "A Channel selection method for EEG Classification in Emotion Assessment Based on Synchronization Likelihood," in *15th European Signal Processing Conference (EUSIPCO 2007)*, 2007, pp. 1241–1245.

[10] Robert Horlings, Dragos Datcu, and LJM Rothkrantz, "Emotion recognition using brain activity," in *Int. Conf. on Computer Systems and Technologies and Workshop for PhD Students in Computing (CompSysTech)*, Gabrovo, Bulgaria, 2008, p. II.1.

[11] Peter Lang, M. Bradley, and B. Cuthbert, "International affective picture system (IAPS): Affective rating of pictures and instruction manual," *University of Florida, FL*, 2008.

[12] M Bradley and Peter J Lang, "Measuring emotion: The Self-Assessment Manikin and the semantic differential," *Journal of Behavioral Theory and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[13] Andrea Mognon, Jorge Jovicich, Lorenzo Bruzzone, and Marco Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features.," *Psychophysiology*, vol. 48, pp. 229–240, July 2010.

[14] Arnaud Delorme and Scott Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[15] Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Routledge, 1988.

[16] A. El Ouardighi, A. El Akadi, and D. Aboutajdine, "Feature Selection on Supervised Classification Using Wilk's Lambda Statistic," in *Int. Symp. on Computational Intelligence and Intelligent Infarmatics (ISCIII)*, 2007, pp. 51–55.

[17] James Russell and Albert Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, pp. 273–294, 1977.

[18] Eleni Kroupi, Ashkan Yazdani, and Touradj Ebrahimi, "EEG correlates of different emotional states elicited during watching music videos," in *Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2011, pp. 457–466.

[19] Panagiotis C. Petrantonakis and Leontios J. Hadjileontiadis, "EEG-based emotion recognition using hybrid filtering and higher order crossings," Amsterdam, Netherlands, Sept. 2009, pp. 1–6.