ROBUST EEG EMOTION CLASSIFICATION USING SEGMENT LEVEL DECISION FUSION

Viktor Rozgić, Shiv N. Vitaladevuni, Rohit Prasad
Speech Language and Multimedia Technologies, Raytheon BBN Technologies
vrozgic,svitalad,rprasad@bbn.com

ABSTRACT

In this paper we address single-trial binary classification of emotion dimensions (arousal, valence, dominance and liking) using electroencephalogram (EEG) signals that represent responses to audio-visual stimuli. We propose an innovative three step solution to this problem: (1) in contrast to the typical feature extraction on the response-level, we represent the EEG signal as a sequence of overlapping segments and extract feature vectors on the segment level; (2) transform segment level features to the response level features using projections based on a novel non-parametric nearest neighbor model; and (3) perform classification on the obtained response-level features. We demonstrate the efficacy of our approach by performing binary classification of emotion dimensions on DEAP (Dataset for Emotion Analysis using electroencephalogram, Physiological and Video Signals) and report state-of-the-art classification accuracies for all emotional dimensions.

***Index Terms*—** EEG, emotion recognition

1. INTRODUCTION

Non-invasive methods for detecting health disorders have the potential for revolutionizing the field of medicine. As sensors such as electroencephalogram (EEG) become less intrusive and more affordable, their adoption in healthcare applications becomes more pervasive. One such application is detection of psychological health disorders such as mild Traumatic Brain Injury (mTBI), Post-traumatic stress disorder (PTSD), Depression, etc.

Emotions in particular provide salient cues into an individual's psychological status and have been an active area of research in modeling and analysis of human behavior. Research has focused on detecting emotions from a variety of sensory data including speech [1], text [2], facial expressions [3], physiological signals [4] and EEG [5-12].

We present a novel approach for classifying emotion dimensions in EEG responses to audio-visual stimuli. First, instead of the accepted practice in EEG-based emotion recognition studies [5-12], where EEG features are extracted from the full response-to-stimuli, we segment EEG responses into multiple overlapping segments and extract features for each segment separately. This way each EEG response-to-stimuli is represented with multiple segment-level feature vectors. We motivate this choice by dynamic nature of emotions and the fact that only parts of stimuli, and consequently EEG responses, are relevant for classification of emotion dimensions. Therefore, averaging introduced by feature extraction on the full response level can reduce discriminative potential of the features.

Concatenation of the segment-level features into a response-level feature vector can lead to prohibitively high-dimensional feature vectors, and poses an additional assumption that order of segments is important. Alternative segment fusion by calculation of statistical functionals of segment-level features over all segments-within-a-response introduces a loss of information when the features are highly non-stationary.

In order to overcome the above limitations, we transform multiple segment-level feature vectors to a single response-level feature vector in a manner that retains information relevant for classification task. First, for each segment feature-vector we find its $2K$ nearest neighbors (K in each class assuming binary classification problem) in the set of segment feature-vectors that do not correspond to the same response. Second, we transform the original segment-level feature vectors to $2K$-dimensional multinomial distribution vectors, with probabilities obtained from $2K$ nearest neighbor distances using radial basis kernel with appropriate width. Third, we fuse all multinomial distribution vectors that correspond to a single response into a response-level feature vector in four ways: (1) Generating histograms of votes for the nearest neighbor class on the segment-level; (2) Calculating geometric means of all probability vector entries that correspond to distances to neighbors that belong to the same class; (3) Generating normalized histograms of values that probability vector coordinates take over all segments; and (4) Estimating parameters of Dirichlet distribution that generates all segment-level probability vectors. Finally, we concatenate all four features obtained by fusion into a response-level feature vectors and perform classification.

Relation to prior work: From the application perspective work presented in this paper is related to several other research works on EEG-based estimation of emotion categories and dimensions [5-12]. From this set the most similar are studies [5, 6] that deal with the same dataset and the study [12] that employs audio-visual stimuli. As previously discussed, while these studies suggest feature extraction on the response-level, we propose a conceptually different feature extraction on the response segment-level.

The method we propose for transformation of the original segment-level features to the probability feature vectors is related to the Naïve Bayes nearest neighbor (NBNN) method [13]. Namely, restricting calculation of the geometric means (the second fusion feature) to probabilities corresponding only to the nearest neighbors, and picking a class with a smaller geometric mean is equivalent to NB-NN classification criterion.

Finally, first two steps in the proposed transformation from segment to response-level feature vectors effectively perform projection of the original segment-level features using a non-parametric nearest neighbor model. This relates proposed method to the design of NB-NN kernel presented in [14]. However, while the NB-NN kernel measures distance between responses without notion of the response-level feature vector we fuse segment-level projections to the response-level vectors explicitly.

We conducted all experiments on the DEAP [4]. We performed single-trial binary classification for each of four emotional dimensions (arousal, valence, dominance and liking),d averaged classification accuracies over 32 subjects in the database and obtained the best reported accuracies on DEAP for each emotion dimension: arousal (68.4%), valence (76.9%), dominance (73.9%) and liking (75.3%).

The remainder of the paper is organized in the following way. Section 2 provides a description of the DEAP dataset and segment-level features we use. In Section 3 we describe the proposed method for segment-to-response feature transformation. In Section 4 we list used classification algorithms. Section 5 describes our experiments and summarizes key results. In Section 6 we conclude the paper and provide directions for future research.

2. DATASET AND FEATURES

We tested the proposed scheme for binary classification of emotion related categories (arousal, valence, dominance and liking) on DEAP (**D**ataset for **E**motion **A**nalysis using electroencephalogram, **P**hysiological and Video Signals) [4]. DEAP contains 32-channel EEG, multiple peripheral physiological signals (galvanic skin response, blood pressure, breathing and heart rate, skin temperature and facial electromyography signals) and frontal facial videos recorded for 32 participants while they were watching 40 one-minute long music videos. After the presentation of each stimulus, participants rated its content in terms of arousal, valence, likability, dominance (on scale from 1 to 9) and familiarity (on scale 1 to 5).

In order to match experimental conditions for emotion dimension classification on DEAP in previously reported studies [5,6], we: (a) applied the same set of EEG signal pre-processing steps as in [5] (down-sampling to 128Hz, removing eye-blinking artefacts, bandpass filtering each channel to 4-45Hz interval and averaged channels to a common reference); and (b) transformed ratings for arousal, valence, dominance and liking to two categories corresponding respectively to rating intervals [1,5) (class "0") and [5,9] (class "1"). In our experiments we used only features extracted from EEG signals.

2.1 Segment-level EEG features

We segment EEG response signals into multiple overlapped segments and extract the following features for each segment: spectral power in theta (4-8 Hz), slow alpha (8-10 Hz), alpha (8-12 Hz), beta (12-30Hz) and gamma (30+ Hz) bands for each channel and spectral power differences between symmetric channel pairs for the same set of frequency bands.

3. SEGMENT-LEVEL TO RESPONSE-LEVEL FEATURE TRANSFORMATION

In this section we propose a three step method for transformation of segment to response-level feature vectors. For this purpose we introduce the following notation. Let $x_i = \{x_{i,j}\}_{j=1}^{N_{seg}}$ be a set of segment feature vectors for the response $i(\ i = 1, .., N_{res})$. We assume that emotion labels assigned to a response apply to all segments within that response.

In the first step, for each segment $x_{i,j}$we find its $2K$ nearest neighbors (K in class"1" and K in class"0") in the set of segments that do not belong to the same response. Let us denote distances (in ascending order) to neighbors in class "1" and "class "0", respectively with $\left(d_{i,j}^{1,k}\right)_{k=1}^{K}$ and $\left(d_{i,j}^{0,k}\right)_{k=1}^{K}$.

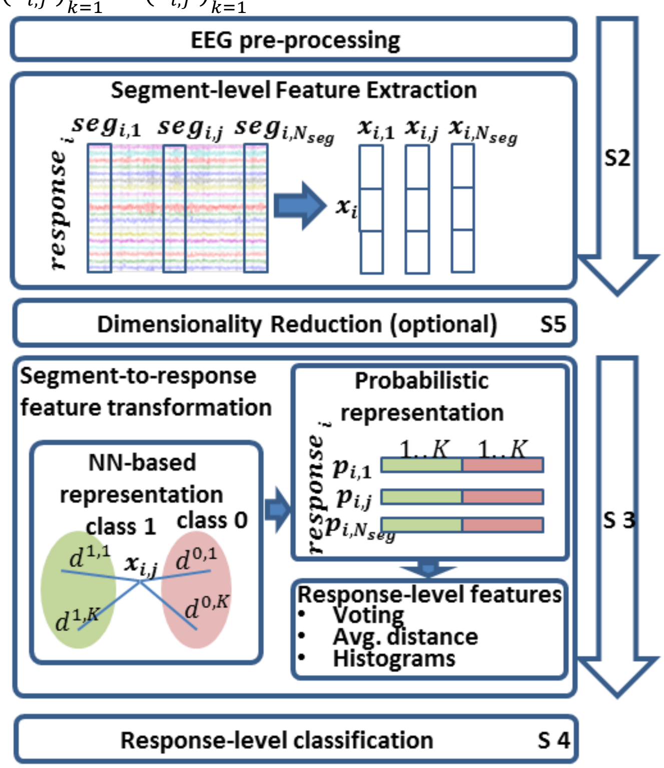

Figure 1 Schematic representation of the proposed methodology. S1-S5 denote sections where the corresponding blocks are described.

In the second step we transform the original segment-level feature vector $x_{i,j}$ to a multinomial distribution vector $p_{i,j} = [p_{i,j}^{1,1}, .., p_{i,j}^{1,K}, p_{i,j}^{0,1}, ,,, p_{i,j}^{0,K}]$, where $p_{i,j}^{\cdot\cdot} \sim e^{-d_{i,j}^{\cdot\cdot\,2}/\varepsilon}$. We set the distance kernel width ε to the mean segment-to-nearest-neighbor distance calculated over all segments that do not belong to i^{th} response [15]. Vector $p_{i,j}$ preserves information about both class and distance properties of the neighborhood of the segment $x_{i,j}$.

The probability vector $p_{i,j}$ and can be interpreted as a projection of the segment-level feature vector to a non-parametric nearest neighbor class model. Since in EEG applications we usually operate with high dimensional feature vectors, the proposed non-parametric model based projection is much more appealing than projections based on parametric models with high number of parameters (e.g. Gaussian mixture model).

In the third step we fuse all segment-level probability vectors from a single response to a response-level feature vector. We propose to do fusion creating 4 types of response-level features, each with particular properties.

(1) NN voting histogram: This feature represents relative number of segments which NNs belong to class "1". The feature assures that classifier using it, when properly trained, should not perform worse than classifier that counts nearest neighbor votes over all segments within a response. Formally, we calculate this feature as in Equation 1.

$$m_i^{1-NN} = {}^{1}/_{N_{seg}} \sum_{j=1}^{N_{seg}} I\left(p_{i,j}^{1,1} \geq p_{i,j}^{0,1}\right) \tag{1}$$

(2) Average segment-to-class distance: This 2-dimensional feature (Equation 2) contains average distances from segments to their *K* nearest neighbors in the class "1" and *K* nearest neighbors in class "0". If we compare distances to only *1*-NN instead of *K*-NNs we get exactly decision criteria of the Naïve Bayes NN technique that is successfully used on image classification tasks [13].

$$m_i^{NB-NN} = \left[\left(\prod_{\substack{k=1..K \\ j=1..N_{seg}}} p_{i,j}^{1,k} \right)^{\frac{1}{KN_{seg}}}, \left(\prod_{\substack{k=1..K \\ j=1..N_{seg}}} p_{i,j}^{0,k} \right)^{\frac{1}{KN_{seg}}} \right] \tag{2}$$

(3) Histograms: We obtain this feature by calculating *B*-bin normalized histograms for each dimension of the segment-level probability vectors over all segments within a response. We denote this *2BK*-dimensional vector with m_i^H.

(4) Generating Dirichlet distribution: Under an assumption that the segment-level multinomial distributions are i.i.d. samples from a Dirichlet distribution (Equation 3) with parameters $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,2K})$ we use these parameters as features.

$$p(p_{i,j}|\alpha_{i,1}, \dots, \alpha_{i,2K}) = \frac{\prod_{k=1}^{2K} \Gamma(\alpha_{i,k})}{\Gamma\left(\sum_{k=1}^{2K} \alpha_{i,k}\right)} \prod_{k=1}^{2K-1} p_{i,k}^{\alpha_{i,k}-1} \tag{3}$$

It is possible to estimate the Dirichlet distribution parameters in the maximum likelihood sense [16], however for simplicity we have used the moment matching approximations [17]. Finally, we use we denote this *2K*-dimensional vector of parameters as m_i^D.

4. RESPONSE-LEVEL CLASSIFICATION

We represent each response with a single feature vector obtained by concatenation of features derived by fusion of segment level features, $m_i = [m_i^{1-NN}, m_i^{NB-NN}, m_i^H, m_i^D]$. Further, we test three classifiers on the set of all response-level feature vectors: (1) NB-NN classifier implemented by comparison of coordinates corresponding to the m_i^{NB-NN} vector; (2) *"NN voting"* classifier (compares numbers of segment level votes assigned to different classes, obtained by comparing the coordinate corresponding to m_i^{1-NN} with 0.5; and (3) support vector machine classifier [18] with radial basis kernel (RBF-SVM) operating on the full response-level feature vectors $m_i (i = 1, .., N_{sti})$.

5. EXPERIMENTAL RESULTS

After pre-processing steps listed in Section 2.1, and in order to experiment with different segment lengths, we segmented EEG signals using 1s, 2s, 4s and 8s windows with 1s shift creating respectively 60, 59, 57 and 53 segments per response. For each segment we extracted spectral powers in theta (4-8 Hz), slow alpha (8-10 Hz), alpha (8-12 Hz), beta (12-30Hz) and gamma (30+ Hz) bands for all channels and spectral power differences between symmetric channels (14 pairs) for the same frequency bands. This creates 5(32+14)=230-dimensional feature vector per segment.

In all experiments we evaluated classification accuracies in a single-trial setup for each subject separately. In other words, we used leave-one-response-out cross validation scheme to obtain single subject accuracy. We report accuracies averaged over all subjects with standard deviations for four emotional categories: arousal (ARO), valence (VAL), dominance (DOM) and liking (LIK). We present results and findings for three experiments.

In our first experiment we compared segment-level classification accuracies for different segment durations. Our goal was to see whether the classification accuracy benefits from the longer segments. We performed the segment level classification using combination of kernel principal component analysis (K-PCA) [19] dimensionality reduction and 1-NN classifier. We selected K-PCA dimension and the kernel width using cross-validation on the set of segments that belong to the training set with average 1-NN classification accuracy as the optimization criterion. For this purpose we searched the dimension-kernel width grid $\{\{230\} \cup \{20k: k = 1, ..7\}\} \times \{2^{-3}, .., 2^3\}$.

Interestingly, when we substituted 1-NN classifier with the linear SVM obtained results were effectively the same. Results presented in Table 1 show that we did not benefit from longer segment durations.

Table 1 Segment level classification accuracy averaged over subjects and standard deviation. Used method: combination of K-PCA and 1-NN classifier.

	VAL[%]	ARO[%]	DOM[%]	LIK[%]
1s	63.1(4.6)	59.5(7.6)	62.0(9.3)	64.1(11.1)
2s	64.4(5.2)	60.6(8.1)	62.3(10.4)	64.0(10.1)
4s	61.1(7.9)	60.3(8.0)	61.3(9.8)	64.5(12.0)
8s	60.9(7.4)	59.8(8.3)	61.1(10.6)	64.1(12.6)

In the second experiment we fused segment-level features into response-level feature vectors by calculating statistical functionals (mean, standard deviation, min, max, range, mode, median, skewness and kurtosis) for each segment-level feature vector dimension over all segments in a response. Further, we classified the obtained 2070-dimensional ($2070 = 230 \cdot 9$) response-level feature vectors using combination of K-PCA and 1-NN classifier. As in the previous experiment we have optimized K-PCA parameters using cross-validation on the training set. In this case we used average response-level 1-NN classification accuracy as the optimization criterion. We present results for this experiment in Table 2 and consider them a baseline for comparison with the proposed method.

Table 2 Response level classification accuracy averaged over subjects and standard deviation. Used method: fusion of segment-level feature vectors using statistical functionals followed by combination of K-PCA and 1-NN classifier

	VAL[%]	ARO[%]	DOM[%]	LIK[%]
1s	54.7(10.6)	58.8(11.6)	59.7(15.4)	60.1(15.2)
2s	54.4(10.7)	59.5(8.5)	60.0(13.9)	60.2(14.2)
4s	53.8(10.8)	58.8(11.7)	61.3(15.0)	58.7(11.9)
8s	54.9(8.2)	59.5(11.0)	59.1(14.0)	59.8(11.4)

In the third experiment, we tested performance of three classifiers (Section 4) based on the proposed segment-to-response transformation method (Section 3). We used K-PCA dimensionality reduction as a preprocessing step for each classifier. For the first two classifiers, NB-NN and "*NN voting*" we optimized K-PCA parameters in the same way as for the first two experiments using cross-validation classification performances on the training set as the optimization criterion. For the third classifier RBF-SVM we have used the same set of parameters as for the *"NN voting"*. We made this sub-optimal choice to simplify the training since for the RBF-SVM classifier we also had to optimize the parameter C that controls trade-off between training errors and the SVM margin size and the kernel width γ. We transformed original segment-level feature vectors to the response-level feature vectors (Section 4) using *K=5* nearest neighbors and optimized C and γ parameters on the grid $\{2^{-6},..,2^{6}\} \times \{2^{-6},..,2^{6}\}$ using cross-validation on the training set. We present classification accuracies averaged over all subjects in Table 3.

From the results in Table 3 we can see that the RBF-SVM classifiers outperform NB-NN, *"NN voting"* and the baseline classifier from the second experiment (Table 2) when applied with the identical segment duration. On all but LIK classification task for 2s segments and ARO classification task for 8s the RBF-SVM is better than the second best classifier (*"NN voting"*) for the same segment duration with at least 5% significance. Overall, the best performance on ARO, VAL and DOM categories is achieved by RBF-SVM classifiers on 1s and 2s segments. Additional t-tests confirm that these classifiers outperform all other classifiers, independently of the segment duration, with 5% significance. Most importantly, all classifiers based on the proposed segment-to-response transformation perform significantly better than: (a) the baseline classifier based on a more basic scheme for fusion of segment-level feature vectors (Table 2); and (2) the best reported results [4] on DEAP (VAL(57.6%), ARO(62.0%), LIK(55.4%)) obtained by the response-level feature extractions.

Table 3 Response level classification accuracy averaged over subjects and standard deviation. Used methods: (A) K-PCA followed by NB-NN; (B) K-PCA followed by *"NN voting"*; (C) K-PCA followed by segment-to-response feature transformation and RBF-SVM

		VAL[%]	ARO[%]	DOM[%]	LIK[%]
1s	A	62.9(6.4)	64.5(11.6)	62.9(15.9)	65.0(12.7)
	B	74.5(8.2)	66.0(12.4)	68.7(16.8)	72.7(12.1)
	C	76.9(6.4)	68.4(12.1)	73.9(11.1)	75.3(10.6)
2s	A	68.0(8.7)	66.3(11.6)	65.0(16.6)	66.0(13.5)
	B	73.1(9.2)	66.4(11.9)	68.8(17.3)	71.5(13.3)
	C	76.0(6.6)	68.9(12.0)	73.2(12.6)	72.7(12.4)
4s	A	68.4(7.8)	64.8(11.0)	65.6(17.3)	67.6(13.3)
	B	68.5(9.8)	66.2(12.1)	66.3(17.3)	70.7(13.1)
	C	73.0(9.1)	69.1(10.5)	71.3(14.3)	72.7(13.2)
8s	A	63.8(8.3)	63.6(11.9)	62.8(15.6)	66.9(13.9)
	B	66.3(10.1)	64.6(11.4)	65.3(13.0)	68.1(14.7)
	C	68.0(10.2)	65.6(10.9)	70.4(12.8)	73.6(12.0)

6. CONCLUSIONS AND FUTURE WORK

In this paper we presented state-of-the-art results for binary emotion dimension classification tasks (valence, arousal, dominance, liking) on DEAP. These results imply that the proposed segment-level feature extraction and the segment-to-response level feature transformation methods represent an appealing choice for EEG emotion recognition.

Future work will focus on two directions that can bring improvements in EEG emotion recognition: (1) Identification of the most informative task-dependent filter banks; and (2) Discovering the most relevant stimuli/response segments by correlating audio-visual stimuli content and EEG responses, or by correlating EEG responses produced by multiple subjects to the same stimuli.

7. ACKNOWLEDGEMENTS

This paper is based on work supported by the DARPA DCAPS program. The views expressed here are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

8. REFERENCES

[1] B. Schuller, A. Batliner, S. Steidl, and D. Seppi. "Recognising Realistic Emotions and Affect in Speech: State of the Art and Lessons Learnt from the First Challenge." Speech Communication, 53(10), 2011.

[2] C. O. Alm, D. Roth, and R. Sproat. "Emotions from text: machine learning for text-based emotion prediction." In Proc. Human Language Technology and Empirical Methods in Natural Language Processing, (HLT/EMLP) 2005.

[3] Busso, Carlos, et al. "Analysis of emotion recognition using facial expressions, speech and multimodal information." In Proc. International Conference on Multimodal Interfaces (ICMI), 2004.

[4] F. Nasoz et al. "Emotion recognition from physiological signals using wireless sensors for presence technologies." Cognition, Technology & Work 6.1 (2004): 4-14.

[5] S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. "DEAP: A Database for Emotion Analysis using Physiological Signals, In IEEE Trans. on Affective Computing, Special Issue on Naturalistic Affect Resources for System Building and Evaluation, in press

[6] S. Koelstra, A. Yazdani, M. Soleymani, C. Muehl, J.-S. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras. "Single Trial Classification of EEG and Peripheral Physiological Signals for Recognition of Emotions Induced by Music Videos." In Proc. Conference on Brain Informatics, 2010.

[7] P. C. Petrantonakis and L. J. Hadjileontiadis. "Emotion Recognition from Brain Signals Using Hybrid Adaptive Filtering and Higher-Order Crossings Analysis." IEEE Trans. on Affective Computing, accepted for publication.

[8] R. Khosrowabadi, C. Q. Hiok, A. Wahab, and K. A. Kai. "EEG-based Emotion Recognition Using Self-Organizing Map for Boundary Detection." In Proc. Pattern Recognition (ICPR), 2010

[9] Y. Liu, O. Sourina, and M. K. Nguyen. "Real-Time EEG-based Emotion Recognition and Applications." In Trans. on Computational Science, 256-278, 2011

[10] K. Takahashi. "Remarks on SVM-based emotion recognition from multi-modal bio-potential signals," In Proc. Robot and Human Interactive Communication (ROMAN) 2004.

[11] Y-P. Lin, C-H. Wang, T-P. Jung, T-L. Wu, S-K. Jeng, J-R. Duann and J-H. Chen."EEG-Based Emotion Recognition in Music Listening." In Trans IEEE on Biomedical Engineering, 57(7), 2010

[12] D. Nie, X-W. Wang, L-C. Shi, and B-L. Lu. "EEG-based Emotion Recognition during Watching Movies." In Proc EMBS Conference on Neural Engineering, 2011

[13] O. Boiman, E. Shechtman, and M. Irani. "In defense of nearest-neighbor based image classification." Computer Vision and Pattern Recognition, 2008. CVPR 2008

[14] T. Tuytelaars, M. Fritz, K. Saenko, T.Darrell. "The NBNN kernel." In Proc. International Conference on Computer Vision (ICCV), 2011

[15] S. Lafon. "Diffusion Maps and Geometric Harmonics." Ph.D. thesis, Yale University, 2004.

[16] T. P. Minka. "Estimating a Dirichlet distribution." Tech. Report, Microsoft Reasearch, 2000

[17] N. Wicker, J. Muller, K. R. Kalathur, O. Poch. "A maximum likelihood approximation method for Dirichlet's parameter estimation." Journal of Comput. Stat. and Data Anal., 52(3), 2008

[18] V. Vapnik. "The Nature of Statistical Learning Theory." Springer-Verlag, New York, 1995.

[19] B. Schölkopf, A. Smola, and K-R. Müller. "Kernel principal component analysis." Artificial Neural Networks, ICANN'97