35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Dynamic physiological signal analysis based on Fisher kernels for emotion recognition

Hernan F. Garcia , Álvaro A. Orozco and Mauricio A. Álvarez

Abstract— **Emotional behavior is an active area of study in the fields of neuroscience and affective computing. This field has the fundamental role of emotion recognition in the maintenance of physical and mental health. Valence/Arousal levels are two orthogonal, independent dimensions of any emotional stimulus and allows an analysis framework in affective research. In this paper we present our framework for emotional regression based on machine learning techniques. Autoregressive coefficients and hidden markov models on physiological signals, based on Fisher Kernels characterization are presented for mapping variable length sequences to new dimension feature vector space. Then, support vector regression is performed over the Fisher Scores for emotional recognition. Also quantitatively we evaluated the accuracy of the proposed model by acomplishing a hold-out cross validation over the dataset. The experimental results show that the proposed model can effectively perform the regression in comparison with static characterization methods.**

I. Introduction

Human emotion recognition plays an important role in applications designed for people with disabilities, physiological medicine, people with some difficulty in recognizing emotions or interface development of intelligent environments [4]. Recently, there has been a growing interest in improving the interaction between humans and computers (Human Computer Interfaces HCI). This emerging field has been an interest research for several scientific areas, i.e., computer science, engineering, psychology, and neuroscience [20]. Over the past years, neuropsychological research has produced various theories regarding the processing of emotion: the basic prototype emotion [5], right hemisphere (RH) [14] and valence/arousal models [9].

Emotions have been conceptualized as action dispositions that vary along valence and arousal dimensions [13]. Valence refers to the pleasant/unpleasant quality of a stimulus and ranges from negative to positive, whereas arousal refers to the intensity of a stimulus and ranges from dull to arousing [8]. Using this bi-dimensional or circumplex model, one can see how emotions are defined. For example, anger and sadness are both negative in valence, but anger is high in arousal, whereas sadness is low in arousal [14]. Facial expressions and physiological signals, provide the building blocks to understand emotion. In order to effectively use facial expressions or physiological signals, it is necessary to understand how to interpret these signals, and it is also important to study what others have done in the past.

Current ongoing research on emotion recognition is focused on the investigation of specific neurophysiological signatures for each emotion [2]. Electrophysiological measurements that assess the human brain activity, such as EEG seem to be sensitive to emotional states, since they capture alterations of the brain activity derived from specific neural networks that play a crucial role in the occurrence of emotional states such as fear [19]. Yohanes in [21], proposes to use discrete wavelet transform (DWT) coefficients as features for emotion recognition from EEG signals. The proposed feature extraction method fully utilizes the simultaneous time-frequency analysis of DWT by preserving the temporal information in the DWT coefficients.

The most commonly employed strategy in automatic dimensional affect recognition from visual signals is to reduce the recognition problem to a two-class problem (positive vs. negative or active vs. passive classification [17]); or a four-class problem (classification into the quadrants of 2D arousal/valence (A-V) space [6]). Currently, there are also a number of works focusing on dimensional and continuous prediction of emotions from the visual modality. The work by Gunes and Pantic focuses on dimensional prediction of emotions from spontaneous conversational head gestures. The prediction is carried out by mapping the amount and direction of head motion, and occurrences of head nods and shakes into arousal, expectation, intensity, power and valence level of the observed subject using support vector regression (SVRs) [7].

Kipp and Martin in [11] investigated (without performing automatic prediction) how basic gestural form features (e.g., preference for using left/right hand, hand shape, palm orientation, etc.) are related to the single pleasure, arousal, dominance (PAD) dimensions of emotion. The work by Nicolaou et al. focuses on dimensional and continuous prediction of emotions from naturalistic facial expressions within an Output-Associative relevance vector machine (RVM) regression framework by learning non-linear input and output dependencies inherent in the affective data [16].

In this paper, we propose an emotion recognition system based on machine learning techniques, in which the signals that are being analyzed are a physiological response to multimodal sources (EEG, EOG, plethysmograph, EMG, GSR, Respiration belt and Temperature). A novel framework is formulated under the autoregressive hidden markov model (AR-HMM) that implies the probabilistic dependency

H. F. García, A. A. Orozco and M. A. Álvarez are with the Department of Electrical Engineering, Faculty of Engineering, Universidad Tecnológica de Pereira, Pereira, Colombia. {hernan.garcia, aaog, malvarez}@utp.edu.co

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

between sequential biosignals target [18]. The motivation for its conception was to capture the temporal dynamics by employing a temporal window. Then a method to map a variable length sequence to a new fixed dimension feature vector space is introduced [1]. The mapping is obtained by the derivatives of the parameters of an underlying generative model. This new feature space is called the Fisher score space on which, any discriminative classifier can be used to perform discriminative training. The main idea of Fisher kernels is to combine generative models with discriminative classifiers to obtain a robust classifier which has the strengths of each approach. Since each Fisher score space is based on a single generative model, then feature space is assumed to be suitable for binary classification problems in nature. Finally, dimensional and continuous prediction of emotions is a relatively unexplored area in the field of affective computing, and which prediction method is best suited to the task is still unknown. Therefore, we introduced a Support Vector Regression (SVR) to enable the learning of such correlations and generate more substantiated predictions by embedding in the model an initial output estimation (AR-HMM) together with the Fisher Scores.

The rest of the paper is arranged as follows. Section II provides a detailed discussion of probabilisic model AR-HMM. Section III presents our emotion characterization method using Fisher Scores. Sections IV and V discuss the experimental setup and results respectively. The paper concludes in Section VI, with a summary and discussion for future research.

II. Auto Regressive model

An autoregressive (AR) process models the linear dependency that may exist in a given time series. It models the signal as the output of a linear system driven by white noise of zero mean and unknown variance [10].

Let the time series training data be: $\mathbf{x} = \{\langle \mathbf{x}_1 y_1\rangle, \langle \mathbf{x}_2 y_2\rangle, \ldots, \langle \mathbf{x}_n y_n\rangle\}$, $\mathbf{x}_n \in M\Re$ and $\mathbf{y}_n \in \{1, 2, \ldots, C\}$. Here $\mathbf{x}_n = [x_n(1), x_n(2), \ldots, x_n(M)]^T$ is the nth time series of length M, y_n is the corresponding class label and C is the number of classes.

Using an AR model with order P, the value of time series $\mathbf{x_n}$ at discrete time t can be represented as:

$$x_n(t) = -\sum_{n=1}^{P} a_{np} x_n(t-p) + e_n(t) = \hat{x}_n(t) + e_n(t) \quad (1)$$

where $e_n(t) \sim N\left(0, \sigma^2\right)$ is the zero mean white noise with σ^2 as variance, and $\mathbf{a}_n = [a_{n1}, a_{n2}, \ldots, a_{nP}]^T$ are the AR coefficients.

The autocorrelation function (ACF) of $\mathbf{x}_n$ at lag p is estimated using $r_{np} = \sum_t x_n(t) x_n(t+p), \quad p = 1, \cdots, P$ and represented as $\mathbf{r}_n = [r_{n1}, \ldots, r_{nP}]^T$.

The variance of the time series, r_{n0}, estimated using $\sum_t x_n(t) x_n(t)$ gives its instantaneous characteristic.

Since $e_n(t) \sim N\left(0, \sigma^2\right)$, the probability density function (pdf) of $\mathbf{x}_n$ can be written as:

$$\begin{aligned} p\left(\mathbf{x}_n | \mathbf{a}_n, \sigma^2\right) &= \left(2\pi\sigma^2\right)^{-M/2} \exp\left(-0.5\sigma^2 \sum_{t=1}^{M} e_n^2(t)\right) \\ &= \left(2\pi\sigma^2\right)^{-M/2} \exp\left(-0.5\sigma^2 \mathbf{a}_n^T \Sigma_n \mathbf{a}_n\right) \end{aligned} \quad (2)$$

where the autocorrelation matrix, Σ_n, is defined as

$$\Sigma_n = \begin{pmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{p-1} & r_{p-2} & \cdots & r_1 & 1 \end{pmatrix} \quad (3)$$

The AR coefficients $\mathbf{a_n}$ can be derived from the autocorrelation function $\mathbf{r_n}$ and the autocorrelation matrix Σ_n as $\mathbf{a_n} = \Sigma_n^- 1\mathbf{r_n}$.

Therefore, (2) can be written as:

$$p\left(\mathbf{x}_n | \mathbf{r}_n\right) \propto \exp\left(-\frac{1}{2} \mathbf{r}_n^T {\Sigma_n}^{-1} \mathbf{r}_n\right) \quad (4)$$

III. Fisher Kernels

A mapping function, ϕ, that is capable of mapping variable length sequences to fixed length vectors enables the use of discriminative classifiers for variable length examples [1]. Fisher kernel defines such a mapping function and is designed to handle variable length sequences by deriving the kernel from a generative probability model. The gradient space of the generative model is used for this purpose.

A. Fisher kernels for HMMs using continuous density mixture of Gaussians

In emotion recognition problems, HMMs are extensively used and have proven successful in modeling affective status. Among different HMM architectures, left-to-right models with no skips are shown to be superior to other HMM architectures.

In this work, we have used the AR coefficients as observations in a left to right HMM with no skips. The parameter of such an architecture are, prior probabilities of states π_i, transition probabilities, a_{ij} and observation probabilities, $b_i\left(O_t\right)$ which are modeled by mixture of M multivariate Gaussians

$$b_i(O_t) = \sum_{m=1}^{M} w_{im} N(O_t; \mu_{im}, \Sigma_{im}) \quad (5)$$

where O_t is the observation at time t and w_{im}, μ_{im} and Σ_{im} are weight, mean and covariance of the Gaussian component m at state i, with a total of M Gaussians components. Hence, Fisher scores are computed from HMM parameters. The methodology used in this work for Fisher kernel and scores space is depicted in [1].

IV. Experimental Setup

A. Database

In this work we used a multimodal dataset for the analysis of human affective states called *DEAPdataset* [12]. The electroencephalogram (EEG) and peripheral physiological signals (Electromyography (EMG), Electrooculography (EOG), Galvanic Skin Response (GSR), Respiration belt, Plethysmograph and Temperature) of 32 participants were recorded as each watched 40 one-minute long excerpts of music videos. Participants rated each video in terms of the levels of arousal, valence, like/dislike, dominance and familiarity. The data was downsampled to $128Hz$ for further processing.

B. AR-HMM learning and Fisher Scores Derivation

In this step, an HMM is trained for every subject trial. Autoregressive coefficients are computed for each physiological signal and used as observations for HMMs. For a left-to-right HMM, the prior probability matrix is constant since the system always starts with the first state with $\pi_i = 1$. We use the HMM Toolbox written by Kevin Murphy [15]. This toolbox supports inference and learning for HMMs with discrete outputs (dhmm's), Gaussian outputs (ghmm's), or mixtures of Gaussians output (mhmm's). Due to a large data processing, the Gaussians used was diagonal. Therefore, 1280 HMMs are trained. Each HMM has five states, and a two mixtures of Gaussians is used in each state. Fisher score spaces are calculated for each HMM and the discriminative regression is done via SVR. The SVR runs are performed with the LIBSVM toolbox [3].

In the experiments, we used training sets for SVR parameter and kernel selection, and an independent test set to assess the generalization performance of our method. To perform subject independent experiments, we applied fifty fold, hold-out cross-validation. For each regressor in the experiments, we performed fifty trainings and obtained results on the validation set, where the average and standard deviation of root mean square error (RMSE) are reported. All the decisions for parameter (epsilon, cost, gamma and degree) and kernel (RBF or Polynomial) selection are given with respect to the accuracies on the validation set. The test set is completely independent and never used either during training. The proposed method is compared against SVM regression using static features computed over the same dataset [12].

V. Results

The results in Table I show RMSE values obtained from Fisher scores for the parameters of the HMMs. The abbreviations m (Number of Gaussians) and e (Number of States) refers to the HMM parameters. This method shows better results when a combination of three states and two Gaussian mixtures in HMM training is used.

Regressions were performed with the goal of recovering three modalities (Positive-Negative emotions, Pleassant-Unpleasant Valence and Active-Passive Arousal) with a hold-out cross-validation scheme. Table II shows the results obtained for the entire regression experiment over the Fisher scores and the static features. The results show that when the dynamic framework for regression was used, biosignals sets achieve lower RMSE values in comparison with the regression over the Static Features. The active-passive arousal scheme proved to be more accurate on the arousal regression. Additionally, it can be seen that regression models trained with all biosignals provide the lowest RMSE for the arousal and valence dimension, and the regression models trained using the EEG and EMG cues provide the lowest RMSE for the arousal dimension.

TABLE I

Effect of HMM parameters on the emotion recognition performance (RMSE and std values).

Biosignal Set	HMM $3e$, $2m$	HMM $5e$, $3m$
All Bioset	0.6794 ± 0.0337	0.6991 ± 0.0165
EEG	0.6716 ± 0.0429	0.6788 ± 0.0330
EMG	0.6681 ± 0.0373	0.6787 ± 0.0320
EOG	0.6738 ± 0.0449	0.6878 ± 0.0371
GSR	0.6737 ± 0.0438	0.6791 ± 0.0475
Respiration belt	0.6724 ± 0.0436	0.6953 ± 0.0352
Plethysmograph	0.6748 ± 0.0414	0.6801 ± 0.0414
Temperature	0.6719 ± 0.0429	0.6953 ± 0.0366

In Figure I, we also provide an illustrative comparison between the RMSE values computed by SVR regarding the RBF and Polynomial Kernels.

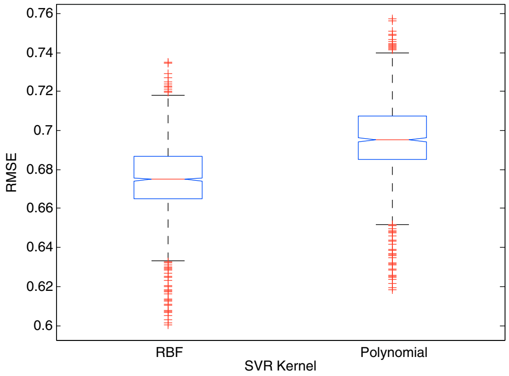

Fig. 1. RMSE values for different kernels used in regression process.

VI. Conclusions and Future Work

This paper presents a novel method for emotion regression by mapping dynamic physiological signals to a Fisher score space based on HMMs. HMMs provide a robust method for recognizing valence arousal levels, by modeling and processing dynamic data. However, the performance of the regression model is improved by combining discriminative models with HMMs which are more suitable in regressions problems.

The results show a better performance in the regression of the dynamic features of Fisher scores. EEG signals proved to be more relevant in the regression process of valence-arousal levels, which leads to an accurate emotion regression process. However, other physiological signals such a EMG,

TABLE II

EMOTION REGRESION PERFORMANCE OF DIFFERENT SCHEMES.

Biosignal Set	Valence/Arousal				Valence				Arousal			
	Positive		Negative		Pleasant		Unpleasant		Active		Passive	
	Dynamic	Static	Dynamic	Static	Dynamic	Static	Dynamic	Static	Dynamic	Static	Dynamic	Static
All Bioset	0.6806	0.6555	0.7108	0.6874	0.6682	0.6745	0.6002	0.6743	0.6621	0.6745	0.6117	0.6539
EEG	0.7020	0.7061	0.6413	0.6355	0.6799	0.6799	0.6459	0.6629	0.6747	0.6768	0.6410	0.6598
EMG	0.6994	0.7038	0.6417	0.6334	0.6744	0.6765	0.6406	0.6596	0.6750	0.6739	0.6412	0.6570
EOG	0.7050	0.7033	0.6420	0.6329	0.6764	0.6756	0.6426	0.6587	0.6750	0.6754	0.6413	0.6585
GSR	0.7047	0.7058	0.6427	0.6353	0.6764	0.6784	0.6426	0.6615	0.6726	0.6761	0.6390	0.6592
Respiration belt	0.7032	0.6679	0.6416	0.6011	0.6740	0.6365	0.6403	0.6206	0.6741	0.6250	0.6404	0.6094
Plethysmograph	0.7041	0.7060	0.6455	0.6354	0.6762	0.6798	0.6424	0.6628	0.6753	0.6768	0.6416	0.6599
Temperature	0.7020	0.7043	0.6461	0.6339	0.6763	0.6784	0.6424	0.6615	0.6757	0.6743	0.6419	0.6575

EOG, GSR, plethysmograph, temperature and respiratory, showed significant results. The performance improvement using multimodal techniques leads to the conclusion that by adding other modalities such as facial expressions and speech, accuracy and robustness should further improve.

Due to the dynamic analysis framework for the physiological changes that presents a specific person in their valence-arousal levels, the proposed methodology has a great potential in applications derived from emotional regression.

Due to the dimensionality problem of Fisher scores it would be of high interest to carry out dimensionality reduction techniques. Some dimensional reduction methods depicted in the state of art are principal component analysis, linear discriminant analysis and recursive feature elimination. These methods aim to maximize the variance and the class separability in the new feature space.

ACKNOWLEDGMENTS

This research is developed under the projects: “Desarrollo de un sistema efectivo y apropiado de estimación del volumen de tejido activo cerebral para el mejoramiento de los resultados terapéuticos en pacientes con enfermedad de Parkinson intervenidos quirúrgicamente”, with code 111056934461 and “Desarrollo de un sistema automático de mapeo cerebral y monitoreo intraoperatorio cortical y profundo aplicación a la Neurocirugía”, with code 111045426008, financed by Colciencias.

REFERENCES

[1] O. Aran and L. Akarun, “A multi-class classification strategy for fisher scores: Application to signer independent sign language recognition,” *Pattern Recognition*, vol. 43, no. 5, 5 2010.

[2] L. Brown, B. Grundlehner, and J. Penders, “Towards wireless emotional valence detection from eeg,” in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 30 2011-sept. 3 2011, pp. 2188 –2191.

[3] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, p. 1:27, 2011, software available at url=http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[4] A. R. Daros, K. K. Zakzanis, and A. C. Ruocco, “Facial emotion recognition in borderline personality disorder,” *Psychological Medicine*, vol. FirstView, pp. 1–11, 10 2012.

[5] P. Ekman, *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*, 2nd ed. 175 Fifth Avenue, New York: Owl Books, 2007.

[6] D. Glowinski, A. Camurri, G. Volpe, N. Dael, and K. Scherer, “Technique for automatic emotion recognition by body gesture analysis,” *Computer Vision and Pattern Recognition Workshop*, vol. 0, pp. 1–6, 2008.

[7] H. Gunes and M. Pantic, “Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners,” in *Proceedings of the 10th international conference on Intelligent virtual agents*, ser. IVA'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 371–377.

[8] K. M. Heilman, “The neurobiology of emotional experience.” *Journal of Neuropsychiatry and Clinical Neurosciences*, no. 9, pp. 439–448, 1997.

[9] W. D. S. Killgore and D. A. Yurgelun-Todd, “The right-hemisphere and valence hypotheses: could they both be right (and sometimes left)?” *Social Cognitive and Affective Neuroscience*, vol. 2, no. 3, pp. 240–250, 2007.

[10] B. V. Kini and C. C. Sekhar, “Large margin mixture of ar models for time series classification,” *Appl. Soft Comput.*, vol. 13, no. 1, pp. 361–371, Jan. 2013.

[11] M. Kipp and J.-C. Martin, “Gesture and emotion: Can basic gestural form features discriminate emotions?” in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, sept. 2009, pp. 1–8.

[12] S. Koelstra, C. Muhl, 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,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[13] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, “Emotion, attention, and the startle reflex.” *Psychological review*, vol. 97, no. 3, pp. 377–395, July 1990.

[14] M. Mneimne, A. S. Powers, K. E. Walton, D. S. Kosson, S. Fonda, and J. Simonetti, “Emotional valence and arousal effects on memory and hemispheric asymmetries,” *Brain and Cognition*, vol. 74, no. 1, pp. 10 – 17, 2010.

[15] K. Murphy, “Hidden markov model toolbox,” 2005. [Online]. Available: http://www.ai.mit.edu/ murphyk/Software/hmm.html

[16] M. Nicolaou, H. Gunes, and M. Pantic, “Output-associative rvm regression for dimensional and continuous emotion prediction,” in *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, march 2011, pp. 16–23.

[17] M. A. Nicolaou, H. Gunes, and M. Pantic, “Audio-visual classification and fusion of spontaneous affective data in likelihood space,” in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ser. ICPR '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 3695–3699.

[18] D. W. Park, J. Kwon, and K. M. Lee, “Robust visual tracking using autoregressive hidden markov model,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, june 2012, pp. 1964 –1971.

[19] H. Silva, A. Fred, S. Eusebio, M. Torrado, and S. Ouakinin, “Feature extraction for psychophysiological load assessment in unconstrained scenarios,” in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 28 2012-sept. 1 2012, pp. 4784 –4787.

[20] M. Wallmer, M. Kaiser, F. Eyben, B. Schuller, and G. Rigoll, “Lstm-modeling of continuous emotions in an audiovisual affect recognition framework,” *Image and Vision Computing*, no. 0, 2012.

[21] R. Yohanes, W. Ser, and G.-B. Huang, “Discrete wavelet transform coefficients for emotion recognition from eeg signals,” in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 28 2012-sept. 1 2012, pp. 2251 –2254.