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

Feature Selection for Multimodal Emotion Recognition in the Arousal-Valence Space

Cristian A. Torres, Álvaro A. Orozco and Mauricio A. Álvarez

Abstract— **Emotion recognition is a challenging research problem with a significant scientific interest. Most of the emotion assessment studies have focused on the analysis of facial expressions. Recently, it has been shown that the simultaneous use of several biosignals taken from the patient may improve the classification accuracy. An open problem in this area is to identify which biosignals are more relevant for emotion recognition. In this paper, we perform Recursive Feature Elimination (RFE) to select a subset of features that allows emotion classification. Experiments are carried out over a multimodal database with arousal and valence annotations, and a diverse range of features extracted from physiological, neurophysiological, and video signals. Results show that several features can be eliminated while still preserving classification accuracy in setups of 2 and 3 classes. Using a small subset of the features, it is possible to reach 70% accuracy for arousal and 60% accuracy for valence in some experiments. Experimentally, it is shown that the Galvanic Skin Response (GSR) is relevant for arousal classification, while the electroencephalogram (EEG) is relevant for valence.**

I. INTRODUCTION

Emotion is a psycho-physiological process that affects the behavior of an individual with respect to a particular situation, and plays an important role in human communication. Emotions affect the responses of different biological systems, including facial expressions, muscles, voice, activity of the Nervous System and the Endocrine System [6] [10]. Various discrete categorizations of emotions have been proposed in [3] and [7]. Other dimensional scales of emotion have also been proposed, like the valence-arousal scale by Russell [8]. In the valence-arousal space each emotional state can be placed on a two-dimensional plane with arousal and valence as the horizontal and vertical axes.

Emotion assessment is often carried out through analysis of a user's emotional expressions and/or physiological signals. So far, most of the studies on emotion assessment have focused on the analysis of facial expressions and speech to determine a person's emotional state. Physiological signals are also known to include emotional information that can be used for emotion assessment but they have received less attention [10]. Recent advances in emotion recognition have motivated the creation of novel databases containing emotional expressions in different modalities. There are a few publicly available multi-modal emotional databases which include both physiological responses and facial expressions, those are the enterface 2005 emotional database, MAHNOB HCI [10] and the DEAP database [6] that contains the arousal, valence and dominance index annotations. There have been recent studies on multimodal emotion recognition over the arousal and valence space like [11], where three physiological signals were used to extract a set of features and classification was performed by a Quadratic Discriminant Classifier (QDC). A preliminary study on emotion classification in the arousal-valence space presented at the EMBC 2012, shows classification performance results of 54.5% for arousal and 38.0% for valence in a multiclass problem using a k-nearest neighbors (KNN) classifier [5].

The classification task in multimodal systems combines different features from different signals in order to recognize the kind of emotion that someone is expressing. Finding a way to reduce the dimensionality of the feature space to overcome the risk of "overfitting" is a well known problem in pattern recognition and machine learning in general. Several Support Vector Machine (SVM) based approaches for feature selection like *Penalty-based methods*, *Feature scaling methods* and *Wrapper methods* have been proposed. Penalty-based methods formulates the optimization problem of the SVM in order to set a large number of weights to zero with the drawback that the number of features chosen is restricted by the penalty parameter included in the optimization of the SVM. The idea of feature scaling methods is that feature rankings can be generated from scaling factors. The magnitude of the weights of a linear discriminant function is a scaling factor of the inputs and non-linear discriminant functions can incorporate scaling factors into the kernel. Finally wrapper methods are computationally intensive methods that searches for an optimum subset of m features by trying all the possible combinations of m features. The combination that yields best classification performance is selected [4].

Another approach to feature selection is the feature ranking, where each feature is ranked based in the contribution to the separation between classes. In 2002 Guyon et al. propose a Recursive Feature Elimination (RFE) algorithm based in feature ranking for SVM, to reduce the dimension of the dataset in DNA microarrays. Since there are just a few examples of the pathology i.e. leukemia, against the high number of genes, a selection of the most discriminant genes is necessary to yield better classification performance. SVM-RFE performs feature selection by

Cristian A. Torres, A. A. Orozco and M. A. Álvarez are with the Department of Electrical Engineering, Faculty of Engineering, Universidad Tecnológica de Pereira, Pereira, Colombia. {cristian.torres, aaog, malvarez}@utp.edu.co

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

computing the change in the cost function when a single feature is eliminated, then the feature that brings the less change in the cost function is discarded. The elimination of features is carried out recursively as an instance of backward feature elimination. For computational reasons, it may be more efficient to remove several features at a time, at the expense of possible classification performance degradation [4]. In first instance RFE was proposed in a biclass context. Further extensions to multiclass problems have been proposed in recent years. The basic scheme is to convert the multiclass problem into several biclass problems using different approaches like One vs One (OvO) and One vs All (OvA). Works like [9] and [12] applied these schemes using RFE to solve multiclass problems combining the feature ranking of each biclass problem to generate a final ranking which leads the feature selection.

The main objective of this work is to determine the most discriminating features in multimodal emotion classification. The DEAP database is used as the testbed for the feature extraction and classification, 299 features are extracted from the signals of each realization from the database. Using a discrete categorization of emotions based on the valence-arousal space, RFE was implemented for the feature selection task.

II. METHOD

A. Datasets

The multimodal database used as the testbed in this work is the DEAP database which contains the electroencephalogram (EEG) and peripheral physiological signals as Galvanic Skin Response (GSR), Respiratory Pattern, Blood volume pressure, Skin Temperature, electromyography and electrooculogram signal of 32 participants. The signals were recorded as each patient 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. Frontal face videos were also recorded for 22 of the 32 participants [6]. From the database each signal was individually analyzed for each realization and a set of features were extracted and organized as Table I shows.

Starting with the complete DEAP database, we selected a subset of it (we only included those patients with available video signal), and organized different datasets to be used in this work. The datasets include 265 features from the EEG signals and the physiological signals, and also 33 features from the video, to bring a total number of 299 for each realization. These video features correspond to the mean shape that was extracted analyzing fiducial face points of 2 frames per second over all the secuence and obtaining the mean shape of all the points [1]. The corresponding indexes for the features of the dataset are organized as Table II shows.

TABLE I
EXTRACTED FEATURES FROM EEG, PHYSIOLOGICAL AND VIDEO SIGNALS [6]

Signal	*Extracted Features*
GSR	Average skin resistance, average of derivative, average of derivative for negative values only, proportion of negative samples in the derivative vs. all samples, number of local minima in the GSR signal, average rising time of the GSR signal, 10 spectral power in the $[0-2.4]$Hz bands, zero crossing rate of Skin conductance slow response (SCSR) $[0-0.2]$ Hz, zero crossing rate of Skin conductance very slow response (SCVSR) $[0-0.08]$Hz, SCSR and SCVSR mean of peaks magnitude.
Skin Temperature	Average, average of its derivative, spectral power in the bands ($[0-0.1]$Hz, $[0.1-0.2]$Hz).
Respiration pattern	Average respiration signal, mean of derivative (variation of the respiration signal), standard deviation, 10 spectral power in the bands from 0 to 2.4Hz.
Blood volume pressure	Average and standard deviation of HR, HRV, and inter beat intervals, energy ratio between the frequency bands $[0.04-0.15]$Hz and $[0.15-0.5]$Hz, spectral power in the bands ($[0.1-0.2]$Hz, $[0.2-0.3]$Hz, $[0.3-0.4]$Hz), low frequency $[0.01-0.08]$Hz, medium frequency $[0.08-0.15]$Hz and high frequency $[0.15-0.5]$Hz components of HRV power spectrum.
EEG	theta, slow alpha, alpha, beta, and gamma Spectral power for each electrode. The spectral power asymmetry between 14 pairs of electrodes in the four bands of alpha, beta, theta and gamma.
EMG and EOG	Eye blinking rate, energy of the signal, mean and variance of the signal.
Video	Mean shape (shape of the face in all the frames from each video).

Using the arousal and valence ratings of the participants for each realization, datasets with two and three classes were organized. For the arousal and valence indexes, the realizations with rating levels of 1 to 2, and 7 to 8 were selected to form a dataset with two classes. Then, a third class is added using the realizations with rating levels between 4 and 5 to form another dataset. We refer to these datasets as DX, where X is a number between 1 and 4. A detailed description of the datasets is showed in Table III.

B. Recursive Feature Elimination

RFE is a feature selection method based in feature ranking for SVMs. For linear classification problems, the ideal objective function is the expected value of the error, that is the error rate computed on an infinite number of examples. For the purpose of training, this ideal objective is replaced by a cost function J computed on training examples only. In the non-linear kernel case for SVMs, the idea is to calculate the change of the cost function $DJ(i)$ caused by removing the i feature.

$$DJ(i) = (1/2)\,\alpha^T\,[H - H(-i)]\,\alpha, \tag{1}$$

Where H is the matrix with elements $y_h y_k K(x_h, x_k)$,

TABLE II
FEATURE INDEXES OF THE D DATASET

FEATURES	INDEXES
GSR	1:11
Skin Temperature	12:15
Respiratory pattern	16:19
Blood Volume Pressure	20:29
EEG	30:253
EOG-EMG	254:265
Video	266:299

TABLE III
DATASETS

Dataset	Level	Description
$D1$	Arousal	2 classes (1 =Passive, 2 =Active)
$D2$	Arousal	3 classes (1 =Passive, 2 =Active and 3 =Neutral)
$D3$	Valence	2 classes (1 =Unpleasant, 2 =Pleasant)
$D4$	Valence	3 classes (1 =Unpleasant, 2 =Pleasant and 3 =Neutral)

K is a kernel function that measures the similarity between x_h and x_k. To compute the change in cost function by removing feature i, the α that determine the solution of the SVM remains unchanged and matrix H is re-computed. This is computing $K\left(x_h\left(-i\right), x_k\left(-i\right)\right)$, yielding matrix $H\left(-i\right)$, where $\left(-i\right)$ means that feature i has been removed. RFE then eliminates the feature on the basis of the small change in the cost function, that is, the feature corresponding to the smallest $DJ(i)$ shall be removed [4]. The procedure can be iterated by following these steps:

1) Train the Classifier
2) Compute the ranking criterion for all features ($DJ(i)$)
3) Remove the feature (or several features) with smallest ranking criterion.

For the development of the RFE algorithm, the PRTOOLS toolbox for MATLAB was used [2]. Since Radial Basis Function (RBF) kernel was documented to give the best intraclass discrimination for the multimodal emotion problem [10]. A RBF-SVM is used for computing the H matrix following the kernel calculation in (2), to solve the cost function in (1).

$$K\left(u, v\right) = exp\left(-\frac{1}{2\gamma^2}\left\|u - v\right\|\right). \qquad (2)$$

The γ parameter in the RBF-kernel function (2) and the regularization parameter C of the SVM, were selected by searching into a logarithmical space in each iteration of the RFE algorithm. The parameters that yield the best classification accuracy over different training and test sets are selected. Then, a cross validation scheme was used with 10 repetitions for every experiment. The RFE algorithm is set to remove a number n of features in each iteration until a subset of a fixed number of features is reached. At each iteration the accuracy of the RBF-SVM over the test set was calculated.

III. EXPERIMENTAL RESULTS

Several experiments are carried out for the feature selection problem. The first one corresponds to the RFE over $D1$ and $D2$ datasets, that is for the arousal levels with two and three classes. In the first iteration, the RFE algorithm is set to remove 55 features for the biclass problem and 65 features for the multiclass problem. The number of features to eliminate is downsampled in each iteration, until it eliminates one feature at the time. The recursive procedure is carried out leaving only the 0.05% of the original feature space. The mean and standard deviation of 10 realizations of RFE for $D1$ and $D2$ are presented in Figure 1.

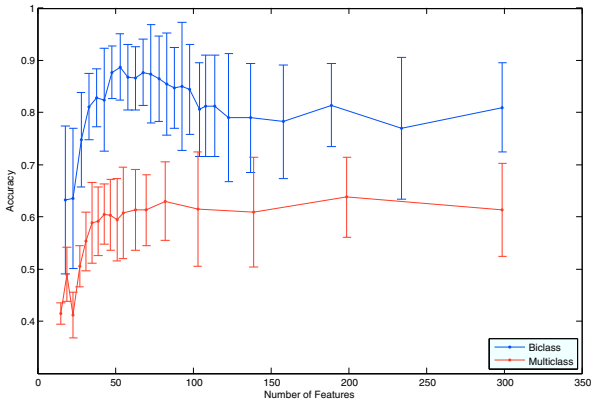

Fig. 1. RFE - biclass and multiclass problem for Arousal level. Starts with the complete set of 299 to a final subset of 13 features

An histogram containing the percentage of occurrence of individual features in the final subset of features is also presented in Figure 2. The percentage of occurrence is calculated from all the realizations of RFE over $D1$ and $D2$ datasets. The x axis corresponds to the feature location into the dataset D as Table II shows and the y axis represents the occurrence of a individual feature in the final subset of features selected by RFE.

The next experiment is carried out for $D3$ and $D4$ datasets, that correspond to valence levels. RFE operates in the same way as for the arousal datasets. The results for the mean and standard deviation over several RFE realizations are showed in Figure 3. An histogram is also presented for the features with more occurrence in the final subset as Figure 4 shows.

A detailed relation of the classification performance from different subsets of features is presented in Table IV.

IV. DISCUSSION AND CONCLUSIONS

As Table IV shows, the classification performance of the selected subsets of features only decreases significantly for feature subsets that contains less than 35 features.

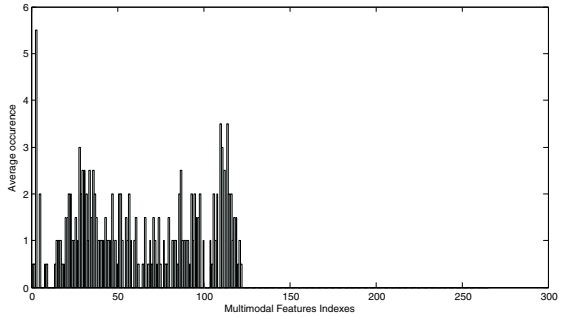

Fig. 2. Histogram for the remaining features in the final subset for arousal index over several RFE realizations with 2 and 3 classes

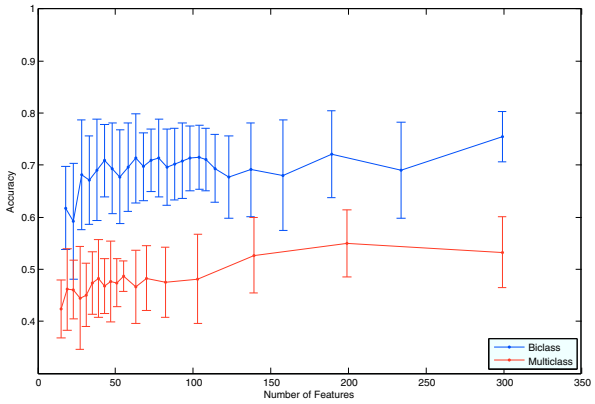

Fig. 3. RFE - biclass and multiclass problem for Valence level. Starts with the complete set of 299 to a final subset of 13 features

That is, RFE discards features that do not contribute to intraclass separation. Figures 1 and 3 confirms that several features can be discarded without affecting the classification performance in the multimodal emotion recognition.

The histograms in Figures 2 and 4 show the features that are most discriminant in every classification problem for arousal and valence. For the arousal labels, after RFE was performed the feature with more occurrences over the final subset in all the realizations was the second feature from the GSR signal (average of the derivative). For the valence labels the features that were retained across all the realizations were the features from the EEG, blood volume pressure and skin temperature signal. The video features were never used for classification. This is due to the fact that a mean shape of the all forms of the face over all the video may contain not only the expression for the emotion represented by the arousal and valence index, but also the shape of the face in other emotion states during all the sequence.

Future work could be oriented to compare RFE with other feature selection methods discussed in previous sections of this paper. The classification performance could be also improved by including nonlinear dynamics into the analysis of the physiological signals.

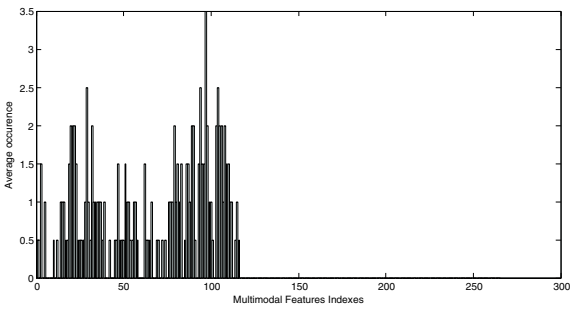

Fig. 4. Histogram for the remaining features in the final subset for valence index over several RFE realizations with 2 and 3 classes

TABLE IV
CLASSIFICATION PERFORMANCE (%) FROM DIFFERENT SUBSETS OF FEATURES SELECTED BY RFE

	Number of Features					
Dataset	299	150	82	35	25	13
$D1$	81.00	81.43	85.71	84.14	72.29	57.14
$D2$	61.32	61.49	63.03	58.85	41.53	41.47
$D3$	75.49	68.06	71.25	64.44	58.33	64.24
$D4$	53.26	52.69	47.52	47.41	49.61	42.41

ACKNOWLEDGMENTS

This research is developed under the projects: “Desarrollo de un sistema automático de mapeo cerebral y monitoreo intraoperatorio cortical y profundo: aplicación neurocirugía” and “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”, financed by Colciencias with codes 111045426008 and 111056934461 respectively. We like to thank the Control and Instrumentation research group and the Technological University of Pereira for bringing the resources needed to complete this work.

REFERENCES

[1] T. Cootes, C. Taylor, and M. M. Pt, “Statistical models of appearance for computer vision,” 2004.
[2] R. P. W. Duin, “Prtools version 3.0: A matlab toolbox for pattern recognition,” in *Proc. of SPIE*, 2000, p. 1331.
[3] P. Ekman, W. V. Friesen, M. O’Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, and P. E. Ricci-Bitti, “Universals and cultural differences in the judgments of facial expressions of emotion.” *Journal of personality and social psychology*, vol. 53, no. 4, pp. 712–717, Oct. 1987. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/3681648
[4] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Mach. Learn.*, vol. 46, no. 1-3, pp. 389–422, Mar. 2002. [Online]. Available: http://dx.doi.org/10.1023/A:1012487302797
[5] K. J., H. X., L. X., L. S., and P. M. . S. T., “Multimodal emotion recognition by combining physiological signals and facial expressions: a preliminary study.” in *Proc. the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC’12), San Diego, CA*, 2012, pp. 5238–5241.
[6] 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.
[7] W. Parrott, *Emotions in Social Psychology: Essential Readings*, ser. Key Readings in Social Psychology. Psychology Press, 2001.
[8] J. Russell, “A circumplex model of affect,” *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
[9] M.-D. Shieh and C.-C. Yang, “Multiclass svm-rfe for product form feature selection,” *Expert Syst. Appl.*, vol. 35, no. 1-2, pp. 531–541, July 2008.
[10] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A multimodal database for affect recognition and implicit tagging,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.
[11] G. Valenza, A. Lanata, and E. Scilingo, “The role of nonlinear dynamics in affective valence and arousal recognition,” *Affective Computing, IEEE Transactions on*, vol. 3, no. 2, pp. 237–249, 2012.
[12] X. Zhou and D. P. Tuck, “Msvm-rfe: extensions of svm-rfe for multiclass gene selection on dna microarray data.” *Bioinformatics*, vol. 23, no. 9, pp. 1106–1114, 2007.