Multiple-output support vector machine regression with feature selection for arousal/valence space emotion assessment

Cristian A. Torres-Valencia, Mauricio A. Álvarez and Alvaro A. Orozco-Gutiérrez

***Abstract*— Human emotion recognition (HER) allows the assessment of an affective state of a subject. Until recently, such emotional states were described in terms of discrete emotions, like happiness or contempt. In order to cover a high range of emotions, researchers in the field have introduced different dimensional spaces for emotion description that allow the characterization of affective states in terms of several variables or dimensions that measure distinct aspects of the emotion. One of the most common of such dimensional spaces is the bidimensional Arousal/Valence space. To the best of our knowledge, all HER systems so far have modelled independently, the dimensions in these dimensional spaces. In this paper, we study the effect of modelling the output dimensions simultaneously and show experimentally the advantages in modeling them in this way. We consider a multimodal approach by including features from the Electroencephalogram and a few physiological signals. For modelling the multiple outputs, we employ a multiple output regressor based on support vector machines. We also include an stage of feature selection that is developed within an embedded approach known as Recursive Feature Elimination (RFE), proposed initially for SVM. The results show that several features can be eliminated using the multiple output support vector regressor with RFE without affecting the performance of the regressor. From the analysis of the features selected in smaller subsets via RFE, it can be observed that the signals that are more informative into the arousal and valence space discrimination are the EEG, Electrooculogram/Electromiogram (EOG/EMG) and the Galvanic Skin Response (GSR).**

I. INTRODUCTION

Human interaction with the environment is highly influenced for the emotional state [10]. The study of the emotions has become one of the most interdisciplinary investigation fields in recent years with a wide variety of applications [12]. Initially, emotions were classified in a discrete space with a categorization of six basic emotions that are unable to represent the complexity of all the subtle emotional states between the basic emotions [3]. An alternative categorization of emotions can be done in terms of few latent dimensions that characterize the responses of a human being from an affective stimulus [8]. Dimensional spaces for emotion classification allow the representation of a higher range of emotions than the classic discrete space. Arousal and valence dimensions conform the most commonly used dimensional space, where the different emotions are described in terms of a continuous range of states [8]. For the valence dimension that is related with the type of emotion, the range goes form pleasant to unpleasant. In the case of the arousal dimension, the range of characterization of emotions goes from excited to calm and describes the intensity of the emotion [8]. Other latent dimensions related to the description of emotions are the dominance and the liking [6].

The authors are with the Department of Electrical Engineering, Faculty of Electrical Engineering, Universidad Tecnológica de Pereira, Pereira, Colombia, {cristian.torres, malvarez, aaog}@utp.edu.co

Automatic emotion recognition has been developed following unimodal and multimodal approaches. The unimodal works solve the emotion recognition problem from the information of one signal, being more suitable for discrete classification spaces [1]. In the case of multimodal approach, several signals are used to extract information that allows further emotion recognition in discrete and continuous spaces of classification [10]. From all the works developed in dimensional spaces of classification, the different dimensions are analyzed independently following classification or regression problems from the acquired data [5] [10]. Nevertheless, from the study of the arousal and valence dimensional space of classification of emotions, some works have concluded that this two main dimensions are correlated [8].

Unlike the works that develop the emotion recognition in a dimensional space by modeling the outputs independently, this work is focused on the use of the different dimensions as the multiple outputs of one only learning algorithm. In order to accomplish this task, an SVM regressor with multiple outputs is selected. SVM are the state of art learning algortihm with multiple applications in several problems [7]. An approach for combining the regression of multiple variables was originally proposed in [11]. This SVM based regression considering multiple output variables (M-SVR) was first applied to the frequency nonselective channel estimation [11] showing benefits in comparison to previous proposals in that field. Further applications to a biomedical problem were presented in [9].

On the other hand, feature selection is a well known problem in machine learning where a feature space dimension reduction is needed in order to remove redundant data and to avoid the "Curse of dimensionality" problem [2]. Recursive Feature Elimination (RFE) is a embedded method for feature selection based on SVMs proposed by Guyon et. al. in 2002 [4]. An initial study of feature selection applied to the emotion recognition field was presented in [13] using RFE, and shows that several features can be discarded without a significant decreasing of the classification accuracy. In [7] a stage of feature selection based on RFE is applied into a regression problem for emotion recognition in the arousal/valence space, bringing low MSRE and MAE ratings even for reduced subsets of features.

The aim of this work is to develop a methodology for M-SVR using the latent dimensions of dimensional spaces of emotion classification as the outputs with a stage of

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

TABLE I

EXTRACTED FEATURES FROM EEG AND PHYSIOLOGICAL SIGNALS [6]

Signal	*Extracted Features*
GSR	Average skin resistance, average of derivative, 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.
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, 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).
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 each signal.

feature selection using RFE. Several biosignals from a well recognized database for emotion assessment experiments are used for the feature extraction and further dataset building. The main contribution of this work is the development of a strategy of emotion recognition using regression with multiple outputs and the extension of the RFE method to the M-SVR approach.

II. MATERIALS AND METHODS

A. Dataset

The Database for Emotion Analysis using Physiological signals (DEAP) [6] is used as the testbed for the multiple-output regression algorithm with feature selection. The DEAP database contains several experiments of emotion elicitation of 32 subjects watching several videos that induce an emotional state described in terms of some continuous dimensions as arousal, valence and dominance. The signals recorded from each experiment are the Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG), Plethysmograph, Galvanic Skin Response (GSR), Skin Temperature and respiration belt [6].

Several features are extracted form each experiment of the database. An statistical and spectral power analysis in some frequency bands from the signals is developed since it has been reported in some works to bring discriminative information [6]. Some of the features extracted from each signal are presented in Table I (a detailed description of all the extracted features can be found in [6]). For each set of features the assigned levels of arousal and valence are stored as the output vector needed for the regression task.

B. Multiple Output Regression

The introduction of the multiple variables regression will help to use the underlying relationship that the outputs could present [11]. For an observable output vector $\mathbf{y} \in \mathbb{R}^Q$, a multidimensional regression estimation problem should be solved, where a regressor $\mathbf{w}^j$ and b^j ($j = 1, \cdots, Q$) must be found for every output. The generalization of the one-dimensional SVR to the multidimensional case leads to the minimization of equation (1), with $u_i = \sqrt{\mathbf{e}_i^\top \mathbf{e}_i}$, with $\mathbf{e}_i^\top = \mathbf{y}_i^\top - \boldsymbol{\phi}^\top(\mathbf{x}_i)\mathbf{W} - \mathbf{b}^\top$, $\mathbf{W} = \left[\mathbf{w}^1, \ldots, \mathbf{w}^Q\right]$ and $\mathbf{b} = \left[b^1, \ldots, b^Q\right]^\top$. The $\varepsilon-$insensitive loss-function then is extended to multiple dimensions in a L_2-based norm, considering all dimensions in an unique restriction yielding a single support vector for all dimensions.

$$L_P\left(\mathbf{W}, \mathbf{b}\right) = \frac{1}{2}\sum_{j=1}^{Q}\left\|\mathbf{w}^j\right\|^2 + C\sum_{i=1}^{n} L\left(u_i\right), \qquad (1)$$

where $L(u) = 0$ for $u > \varepsilon$, and $L(u) = u^2 - 2u\varepsilon + \varepsilon^2$ for $u \leq \varepsilon$ [11]. ε and C are parameters that need to be tuned. To obtain $\mathbf{W}$ and $\mathbf{b}$ a solution for the weighted least square problem is achieved using the iterative reweighted least square algorithm (IRWLS). From this optimization problem and assuming that the work is developed within the feature space kernel, the best solution of a learning problem can be expressed as a linear combination of the training samples in the feature space, i.e., $\mathbf{w}^j = \sum_i \boldsymbol{\phi}\left(\mathbf{x}_i\right)\boldsymbol{\beta}^j = \boldsymbol{\Phi}^\top\boldsymbol{\beta}^j$. The resultant linear system of equations to solve following the IRWLS procedure over a first-order Taylor expansion of equation (1) is presented in equation (2)

$$\begin{bmatrix} \mathbf{K} + \mathbf{D}_a^{-1} & \mathbf{1} \\ \mathbf{a}^\top\mathbf{K} & \mathbf{1}^\top\mathbf{a} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^j \\ b^j \end{bmatrix} = \begin{bmatrix} \mathbf{y}^j \\ \mathbf{a}^\top\mathbf{y}^j \end{bmatrix} \qquad (2)$$

where $j = 1, \ldots, Q$, $(\mathbf{D}_a)_{ij} = a_i\delta(i-j)$, $\mathbf{y}^j = \left[y_{1j}, \ldots, y_{nj}\right]^\top$ and $(\mathbf{K})_{ij} = k\left(\mathbf{x}_i, \mathbf{x}_j\right)$ is known as the kernel matrix. The search algorithm can be used to compute $\boldsymbol{\beta}^j$. Once the $\boldsymbol{\beta}^j$ have been computed the outputs cannot be directly calculated because they are function of the nonlinear transformation $\phi\left(\cdot\right)$. For each new incoming vector $\mathbf{x}$, the jth output can be computed as $y^j = \boldsymbol{\phi}^\top\left(\mathbf{x}\right)\boldsymbol{\Phi}^\top\boldsymbol{\beta}^j$. Defining a matrix $\boldsymbol{\beta} = \left[\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots, \boldsymbol{\beta}^Q\right]$, the Q outputs can be computed using $\mathbf{y} = \boldsymbol{\phi}^\top\left(\mathbf{x}\right)\boldsymbol{\Phi}^\top\boldsymbol{\beta} = \mathbf{K}_\mathbf{x}\boldsymbol{\beta}$. A detailed description of the method can be found in [11].

C. Recursive Feature Elimination

Recursive feature elimination (RFE) was proposed to select subsets of features using SVM's into DNA microarray classification of pathologies [4]. The main idea of the RFE algorithm is to use the weights of the trained SVM to compute a ranking criterion for the relevance of each feature into the intraclass discrimination. This ranking criterion allows the elimination of one or several features in each iteration. As proposed by Guyon et. al. in [4], the ranking vector $\mathbf{DJ} = \{DJ(i)\}_{i=1}^{D}$ for the non-linear case is computed from the $\boldsymbol{\alpha}'s$ of the SVM training and the change in the cost function following the elimination of one feature following the kernel computing. This method can be extended to the M-SVR case as we propose, by using the $\boldsymbol{\beta}'s$ from each output to compute the ranking criterion as equation (3) shows [4]

$$DJ(i) = \frac{1}{2}\left[\boldsymbol{\beta}_Q^\top \mathbf{H}\boldsymbol{\beta}_Q \;-\; \boldsymbol{\beta}_Q^\top \mathbf{H}(-i)\,\boldsymbol{\beta}_Q\right], \qquad (3)$$

$$\boldsymbol{\beta}_Q = \frac{1}{Q}\sum_{j}^{Q}\boldsymbol{\beta}^j, \qquad (4)$$

where $\boldsymbol{\beta}_Q$ in equation (4) is the average of the $\boldsymbol{\beta}$'s from each output, $\mathbf{H}$ is the matrix with elements $y_h y_k K(\mathbf{x}_h, \mathbf{x}_k)$ and K is a kernel function that measures the similarity between $\mathbf{x}_h$ and $\mathbf{x}_k$. The notation $(-i)$ means that the feature i has been removed [4]. Once the ranking vector $\mathbf{DJ}$ is computed, the feature with the minimum ranking is eliminated from the dataset. The process is repeated iteratively, training the M-SVR and eliminating the feature with less ranking until a desired size is reached.

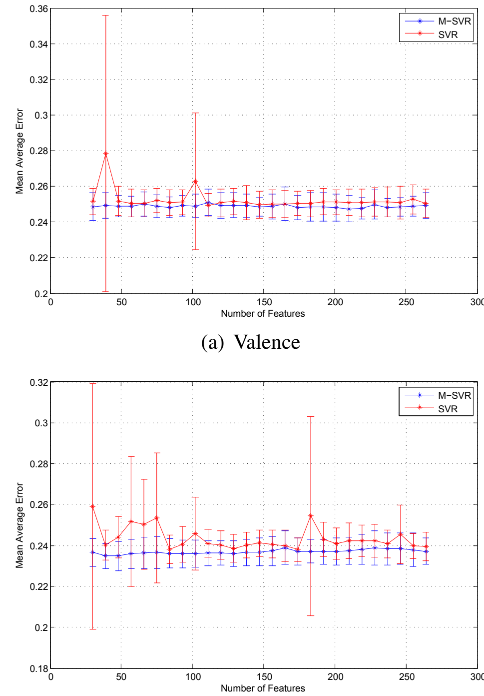
D. Procedure

From the builded dataset, 120 examples are selected for the M-SVR training. A RBF kernel is the mapping function and the hypherparameters γ and C of the kernel are searched into a logarithmic space to obtain the best possible performance for each training set in terms of the determination of the multiple outputs. Another subset of 180 examples are used to test the performance of the M-SVR and a few error metrics are computed as the mean absolute error (MAE), root mean squared error (RMSE) and the coefficient of determination (r2). The selection algorithm RFE takes place to reduce the feature space until the 95% of the original features is removed. The error metrics are computed for each selected subset and the experiment is repeated 10 times with different training and test sets for statistical validation of the results.

The results obtained from the M-RFE-SVR are compared against a single RFE-SVR in each dimension to determine the optimality of the proposed approach and a non-parametric statistical test is performed to determine the difference between the two methodologies. From the feature selection experiments, histograms of occurrence are computed for the analysis of the features that are selected in the smaller sized feature subsets. This histograms are constructed computing the times that a single feature is eliminated in each iteration or the RFE algorithm, giving a description of the percentage of apparition of each feature in the reduced subsets. From this analysis, is possible to determine the signals that bring major information for emotion assessment.

III. EXPERIMENTAL RESULTS

The results from the M-RFE-SVR and RFE-SVR experiments into the emotion recognition problem are presented in Figure 1. It can be noticed that several features can be removed without affecting significantly the RMSE performance for both SVR and M-SVR in the two dimensions. These results are comparable against the state of art for this scope, obtaninig error metrics as MAE around 0.21 which are lower than the reported 0.24 in [7]. For the valence dimension, both of the experiments (M-SVR and SVR) show a similar behavior for the RMSE in terms of mean and



(a) Valence

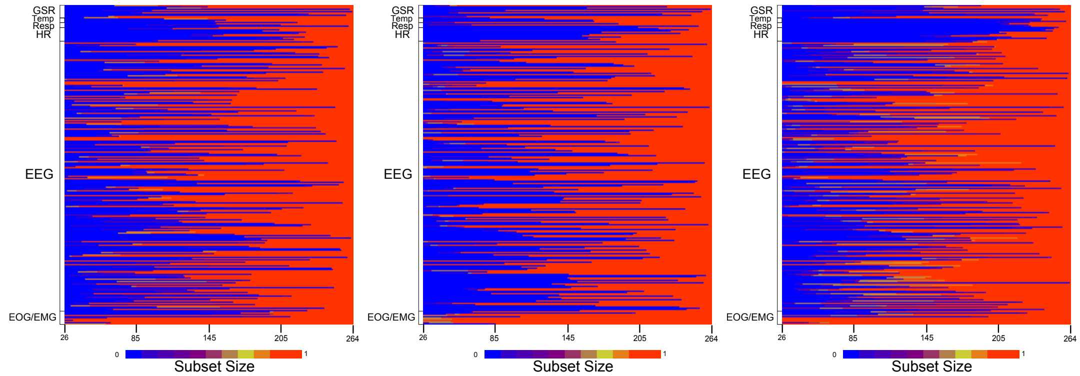
(b) Arousal

Fig. 1. RMSE metric against feature space size from the RFE for M-SVR and SVR experiments in Arousal/Valence space

standard deviation with levels around 0.25 even for small datasets, Figure 1(a). In the case of the arousal dimension, several feature eliminations are developed and the RMSE maintains a level around 0.24 with higher fluctuations for the single SVR, Figure 1(b). The analysis of the features selected in the elimination experiments using RFE is presented in Figure 2. A similar distribution can be observed on the features selected for each iteration of RFE in both SVR and M-SVR experiments.

From the analysis of Figure 2(a), it can be assessed that features from the EEG and the EOG/EMG are eliminated in later stages in comparison to the features from other signals. The selected features in the experiments of single regression, Figures 2(b) and 2(c) comes from the EEG and EOG/EMG, an inclusion of the Temperature and respiratory pattern is also observed. A compilation of the results from all the experiments is presented in Table II for average and bests results respectively. From the multiouputs experiments the metrics are computed for each dimension respectively and the best results corresponds to the minimum RMSE and MAE and the highest coefficient of determination R2.

From an statistical analysis of these results, an equal median test for the RMSE shows that the two modalities (SVR and M-SVR) are different statistically in some subsets of features with a higher performance from the M-SVR method in one subset for valence and in 12 subsets for arousal. For the r2 the statistical test shows that in some feature subsets the regression coefficient of determination has different medians between the SVR and M-SVR, being superior the M-SVR scheme in 10 subsets for valence and



(a) M-SVR (Arousal/Valence)
(b) SVR Arousal
(c) SVR Valence

Fig. 2. Occurrence histograms of the features in different selected subsets from 10 realizations of RFE

TABLE II

AVERAGE AND BEST ERROR METRICS FOR SVR AND M-SVR

Average error metrics for single and multioutput regression			
Experiment	RMSE	MAE	R2
mSVR Valence	$\mathbf{0.252 \pm 0.026}$	$\mathbf{0.213 \pm 0.021}$	$\mathbf{0.038 \pm 0.003}$
mSVR Arousal	$\mathbf{0.240 \pm 0.024}$	$\mathbf{0.203 \pm 0.020}$	$\mathbf{0.025 \pm 0.005}$
Valence	0.261 ± 0.043	0.216 ± 0.035	0.002 ± 0.001
Arousal	0.255 ± 0.129	0.209 ± 0.091	0.013 ± 0.004
Minimum error and maximum r2 for single and multioutput regression			
Experiment	RMSE	MAE	R2
mSVR Valence	$\mathbf{0.248 \pm 0.026}$	0.209 ± 0.021	$\mathbf{0.013 \pm 0.003}$
mSVR Arousal	$\mathbf{0.236 \pm 0.024}$	0.199 ± 0.020	$\mathbf{0.035 \pm 0.005}$
Valence	0.249 ± 0.043	$\mathbf{0.208 \pm 0.035}$	0.005 ± 0.001
Arousal	0.238 ± 0.129	$\mathbf{0.197 \pm 0.091}$	0.027 ± 0.004

in 9 subsets for arousal.

IV. CONCLUSIONS AND DISCUSSION

On the evidence of the results presented in section III, the M-RFE-SVR method proposed combines the feature selection stage with the new M-SVR approach successfully. Several feature eliminations take place while the error metrics for the outputs maintain similar levels compared to the regression using the complete set of features. Figures 1(a) and 1(b) show similar behavior for both dimensions with arousal presenting fewer errors between the estimated and test outputs.

Occurrence histograms in Figure 2 shows that the selected features in smaller subsets belong to similar signals in both multiple and single outputs. The features from the EEG have a higher occurrence in final subsets combined with some features from the EOG/EMG and the GSR signal. The selected features then are coherent in all the experiments giving some important information about the most relevant signals into the HER problem from multimodal information.

ACKNOWLEDGMENT

The authors would like to thank "Universidad Tecnológica de Pereira" for providing the resources needed to complete this work. The development of the work was also financially supported by COLCIENCIAS under the project "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. We also like to thank the authors of the work "SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems" [11] for providing the M-SVR code

REFERENCES

[1] Ma Chang-wei and Guang-Yuan Liu. Feature extraction, feature selection and classification from electrocardiography to emotions. In *Computational Intelligence and Natural Computing, 2009. CINC '09. International Conference on*, volume 1, pages 190–193, 2009.

[2] Kuo-Sheng Cheng, Yu-Shan Chen, and Ting Wang. Physiological parameters assessment for emotion recognition. In *Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on*, pages 995–998, 2012.

[3] Paul Ekman. Universals and cultural differences in facial expressions of emotions. 19:207–283, 1972.

[4] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Mach. Learn.*, 46(1-3):389–422, March 2002.

[5] Kortelainen J., Huang X., Li X., Laukka S., and Pietikainen M y SeppAnen T. Multimodal emotion recognition by combining physiological signals and facial expressions. In *Proc. the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'12), San Diego, CA*, pages 5238–5241, 2012.

[6] S. Koelstra, C. Muhl, M. Soleymani, Jong-Seok 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*, 3(1):18–31, 2012.

[7] Sander Koelstra and Ioannis Patras. Fusion of facial expressions and eeg for implicit affective tagging. *Image Vision Comput.*, 31(2):164–174, 2013.

[8] M.A. Nicolaou, H. Gunes, and M. Pantic. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *Affective Computing, IEEE Transactions on*, 2(2):92–105, April 2011.

[9] Fernando Pérez-Cruz, Gustavo Camps-Valls, Emilio Soria-Olivas, Juan Jos Prez-Ruixo, Anbal R. Figueiras-Vidal, and Antonio Arts-Rodrguez. Multi-dimensional function approximation and regression estimation. In Jos R. Dorronsoro, editor, *ICANN*, volume 2415 of *Lecture Notes in Computer Science*, pages 757–762. Springer, 2002.

[10] T. Pun, M. Pantic, and M. Soleymani. Multimodal emotion recognition in response to videos. *IEEE Transactions on Affective Computing*, 3(2):211–223, 2012.

[11] M. Sanchez-Fernandez, M. de Prado-Cumplido, J. Arenas-Garcia, and F. Perez-Cruz. SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems. *Signal Processing, IEEE Transactions on*, 52(8):2298–2307, Aug 2004.

[12] Jie Shen and M. Pantic. A software framework for multimodal human-computer interaction systems. In *Systems, Man and Cybernetics,IEEE International Conference on*, pages 2038–2045, 2009.

[13] C.A. Torres, A.A. Orozco, and M.A. Alvarez. Feature selection for multimodal emotion recognition in the arousal-valence space. In *Engineering in Medicine and Biology Society (EMBC), 35th Annual International Conference of the IEEE*, pages 4330–4333, July 2013.