Cross-validation of a classification method applied in a database of sEMG contractions collected in a body interaction videogame

M F Montoya Vega¹, O A Henao²

¹Universidad Tecnológica de Pereira, Faculty of Engineering, Master’s in Electrical Engineering, 27 Street #10-02 Pereira, Colombia.
²Universidad Tecnológica de Pereira, Faculty of Health Sciences, 27 Street #10-02 Pereira, Colombia.

mf.mv@utp.edu.co,²oscarhe@utp.edu.co

Abstract. paper presents the evaluation and cross-validation of four pattern recognition classifiers, with the objective of finding the best one for classify surface electromyography (sEMG) signals combined with information extracted from videogame’s variables. The classifiers, a linear classifier, a quadratic classifier, a k-nearest-neighbor classifier, and a support vector machine, were computed on a data matrix created with the recorder signal collected from 12 subjects in a body interaction videogame that used a sEMG as a control strategy for upper limbs virtual rehabilitation. Although the classifiers of sEMG signal had a widespread study, there is not evidence of how to deal with the information of sEMG signals combined with game variables. The classification task is related with discern each subject as “good player” or “bad player”, looking to following the performance of the videogame’s users through the game sessions. A cross-validation of 10 iterations was computed, PCA and Relief were used as feature extraction and selection methods. The evaluation was developed using the percentage of accuracy, defined as the well-predicted points. The best accuracy in the classification task was found using a SVM with a misclassification parameter of 400 and an RBF kernel regularization parameter of 60. Base on this result, the SVMs showed to be the appropriate classifier to be used on sEMG signal combined with videogame variables and should be implemented to follow the user’s performance.

1. Introduction

The development of interactive technologies has allowed a wide diversification of the devices and techniques that can help to complement conventional rehabilitation processes. A more interesting and less explored form in the use of physiological signals in the Human-Computer Interaction (HCI) is the bio-cybernetic adaptation. Here the characteristics extracted from the physiological signals allow describing specific psychological states of the user (stress, fatigue) [1]. These states are used to physiologically adapt the system and assist the user with multiple purposes. These systems, called computational physiological systems [2] can encourage therapy progress by adapting intelligently and progressively to the needs of the user in an environment that is challenging, attractive and patient-friendly. In particular, the surface electromyography signal (sEMG) has been used in this type of systems to train aspects of motor learning. Recently, the rise of physiological sensors has facilitated the use of sEMG signals in applications such as ergonomics, sports training, and rehabilitation [3].
Along with virtual reality systems, alternative therapies for motor disorders and motor rehabilitation have been implemented, which has proven to be a more motivating and adherent strategy for patients.

On the other hand, the use of classifiers of the signal sEMG has been widely used to study patterns of neuromuscular disorders [4], as well as for the diagnosis of fatigue states [5] in combination with other optimization methods, and for the creation of human-computer intuitive interfaces. The tool most used by researchers is the support vector machines (SVMs), which have demonstrated high precision in the classification of this type of signals for differentiation of hand and arm gestures [6].

Three basic models to solve classification problems are the linear classification (CL), the quadratic classification (CQ) and the k-nearest-neighbor (KNN) classification. The goal in classification is to take a database X and to assign it to one of K discrete classes. In the most common scenario, the classes are taken to be disjoint, so that each input is assigned to one and only one class. The input space is thereby divided into decision regions whose boundaries are called decision boundaries or decision surfaces [7]. Particularly, a well-known robust method is the SVM. An important property of SVMs is that the determination of the model parameters corresponds to a convex optimization problem, and so any local solution is also a global optimum [7].

Figure 1. Force Defense, a body interaction videogame controlled by sEMG contractions. Superior left corner: the user’s life and the players awarded points. Inferior right corner: The user’s contraction level.

The non-stationary and stochastic nature of sEMG signals is especially considered in classification task. Nevertheless, in isometric contractions it has been shown that in short intervals of the signal, this can be assumed quasi-stationary [3], [8], [9]. Therefore, the temporal and frequential features are used to describe the behavior of this signal, and features that can better characterize patterns of muscular activity play a key role classification task. Feature selection and extraction can be used to speed up the learning process, boost the classification accuracy, improve model generalization capability [7]. The extraction of this relevant features can be developed through different methods. The Principal Component Analysis PCA is one of the most used, which seeks a space of lower dimensionality, known as the principal subspace [7]. The goal is to represent data in a space that best describes the variation in a sum-squared error sense [10]. On the other hand, RELIEFF (a Relief-based feature selection algorithm) is considered one of the most successful algorithms for assessing the quality of features due to its simplicity and effectiveness. Is a classical supervised feature selection algorithm in the filter model [11]. The main goal is to estimate attributes according to how well their values distinguish among instances that are near each other. For that purpose, relieff for a given instance
searches for its two nearest neighbors: one from the same class (called nearest hit) and the other from a different class (called nearest miss) [11]. The evaluation of the classifiers is often used to find the best one. This evaluation is made using the accuracy, defined as how many times the classification method was right with the prediction of the classes [7], [12].

Thereby, in previous works a videogame called Force Defense was designed (figure 1.)[13], [14]. It is a computational physiology system with adaptation to muscle fatigue detected in the sEMG signal obtained in line with the low-cost sensor Myo armband. While each subject interacted with the videogame the state of muscular fatigue of the biceps was analyzed, if this state was positive the difficulty of the video game decreased, on the contrary, if the state of fatigue was negative, the difficulty of the game increased. The interaction of each subject ended when he won, lost, or was unable to continue. sEMG signals and muscle fatigue states were pre-processed and recorded in .csv files.

Although the classifiers for sEMG signal had been worldwide studied, there is no evidence of which type of classifier could be the best for sEMG muscle fatigue descriptors in together with videogame variables taken under protocols of virtual rehabilitation. That is why we want to find the best classifier for our database. The database chosen for this study was the data collected in a physiological experiment carried out with 12 subjects. In the experimental protocol, the subject was asked to play the videogame previously described. The classification task is related to classifying the subjects as “good player” or “bad player”, due to the videogame was designed to motor rehabilitation and the performance of each subject should be monitored.

2. A Materials and Methods
A super-matrix was created with the data of each subject collected in the muscle fatigue experimental protocol already described. The information available for the study was the sEMG signal pre-processed, the muscle fatigue index, and the game variables, such as the point awarded, the attacks receive, the difficulty changes, and the subject’s life lost (figure 1.). This information was divided by contractions. The samples of every subject were added to the previous one until completing the entire super-matrix. Using Matlab as processing software, data matrix X was created, where every row, related to the features, was integrated as follow:

\[ x_n = \text{FatigIndex}(1:1001) + \text{points}(1002:2001) + \text{Life}(2002:3001) \]  (1)

From row number 1 to row 1001, the fatigue index, from row number 1002 to row 2001 the points awarded by each subject during one contraction, and from row number 2002 to row number 3001 the life lost by each subject during one contraction.

The samples \( x_n \) were separated in every work period, a contraction and the respective rest, of all the signal of each subject. Having in mind that each subject had a different number of samples due to each one had different times of interaction with the game, for instance, some subjects lost the game faster than others, the total number of samples was of 198. Finally, the matrix had a dimension of 198x3001, as can be seen in equation (2):

\[
X = \begin{bmatrix}
x_{n0,1} & \cdots & x_{n0,1001} & \cdots & x_{n0,2001} & \cdots & x_{n0,3001} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
x_{n1,1} & \cdots & x_{n1,1001} & \cdots & x_{n1,2001} & \cdots & x_{n1,3001} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
x_{n2,1} & \cdots & x_{n2,1001} & \cdots & x_{n2,2001} & \cdots & x_{n2,3001} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
x_{n\text{tot},1} & \cdots & x_{n\text{tot},1001} & \cdots & x_{n\text{tot},2001} & \cdots & x_{n\text{tot},3001}
\end{bmatrix} 
\]  (2)

Where \( n1 \) is the number of samples of subject 1, \( n2 \) the number of samples of subject 1 plus subject 2, and so on until \( n\text{tot} \) is the number of total samples, 198 for our case. For instance, subject 1 had 24...
samples, so $n_0=1$ and $n_1=24$, then subject 2 had 6 samples, so $n_2=30$, and so on until 198 total samples.

A label vector was created as follows: due to the performance of every user is evaluated as the ratio between the points awarded and the sum between the points awarded and the life lost (equation 3), this performance index was extracted in every contraction of each user. The mean value of this vector was extracted, and the values above this mean were labelled as 2 “good player” and the values below the mean were labelled as 1 “bad player”, and thereby the T label vector had a size of 198x1.

$$\text{Performance index} = \frac{\text{PointsAwarded}}{\text{PointsAwarded}+\text{LifeLost}}$$  \hspace{1cm} (3)

Using the software Matlab, the four classifications tasks previously named were carried out with the respective evaluation. The training set was of 70% of the data, and an analysis of selection and extraction of features was previously develop using PCA and Relieff. The cross-validation was made with 10 iterations and an analysis of the average of the set of iterations was perform with box diagrams. Moreover, a nested cross-validation with 10 iterations was computed in the last two classifiers, in order to find the optimal parameters for our database. The KNN parameter that was changed was the number of k-neighbors using 1, 2,3,5,7,9,11 neighbors. In the SVM a RBF kernel was used, and the parameters that were changed were the kernel regularization parameter using 20, 40, 60, 70, 80, 90, 100, 120; and the misclassification error parameter using 1, 10, 100, 200, 400, 600, 800, 1000. To reduce the computational cost, the SVM was trained with the number of characteristics with better performance obtained using the KNN classifier. The named classifiers were evaluated using the minimum distance of the accuracy to a target accuracy of 100% and 0% of standard deviation.

3. Results and Discussion

A first approach to the selection and extraction of features was developed using a correlation analysis and a distance analysis. These two matrices can be seen in figure 2 and 3. The correlation matrix shows a low correlation between the characteristics, this could mean that a low number of it could be ignored two reduce the dimensionality of our database. The Euclidean distance matrix shows a high similarity between most of the features, therefore the variance is low.

![Figure 2: Correlation matrix of the features of the videogame database. The colour bar represents the values of the correlation index.](image1.png)

![Figure 3: Euclidean distance matrix of the features of the videogame database. The colour bar represents the distance values between features.](image2.png)
A PCA and Relieff analysis were computed to notice the relevance of features and avoid redundancy. As the variance of this matrix is low, the PCA method could be better than the Relieff. In figure 4 the projected space with PCA can be seen. In this region, the classes seem to be separable. The PCA relevance for each feature shown in figure 5a) evidence that the awarded points and the lost life have higher weights than the fatigue index.

![Figure 4](image-url)  
**Figure 4.** Projected space using the PCA method for the features of the videogame database. Z1, Z2 and Z3 are the coordinates in the projected space.

![Figure 5](image-url)  
**Figure 5.** a) PCA relevance of features for the videogame database. a) The Relieff relevance of features for the videogame database.

The relevance is shown in figure 5b) was extracted using the Relieff method. According to this, some fatigue index features and the points awarded features are the ones with higher weights. The cross-validation of the Linear, Quadratic and KNN classifiers can be seen in figure 6 and 7, where they evidence a better performance with a higher number of characteristics. In both cases, the worst classifier was the Linear and the best was the KNN. The results are summarized in table 1.
Table 1. Best accuracy using cross-validation of the Linear, Quadratic and KNN Classifiers. Number of features with the best performance, the percentage of accuracy and the respective standard deviation.

| CLASSIFIER | RELIEFF | | | | PCA | | | |
|---|---|---|---|---|---|---|---|---|
| | Number of Features | Accuracy (%) | Standard Deviation (%) | Number of Features | Accuracy (%) | Standard Deviation (%) | Number of Features | Accuracy (%) | Standard Deviation (%) |
| Linear | 2827 | 69.33 | 9.244 | 2948 | 69.33 | 6.806 | | | |
| Quadratic | 2633 | 61.00 | 5.273 | 1740 | 63.99 | 7.944 | | | |
| KNN | 2196 | 67.66 | 6.452 | 1804 | 69.00 | 4.116 | | | |

Figure 6. Box diagram of the percentage of accuracy of the classifiers using a cross-validation with the extraction of features using PCA. Black line: Linear classifier, black dot: best accuracy. Dark grey line: Quadratic classifier, dark grey dot: best accuracy. Light grey line: KNN Classifier, light grey dot: best accuracy.

Figure 7. Box diagram of the percentage of accuracy of the classifiers using a cross-validation with the selection of features using Relieff. Black line: Linear classifier, black dot: best accuracy. Dark grey line: Quadratic classifier, dark grey dot: best accuracy. Light grey line: KNN Classifier, light grey dot: best accuracy.

As the best performance for KNN was found with 1804 features with PCA, these parameters were used to train an SVM. The result can be seen in the boxplot in Fig. 8, where the median was 81.66% with a standard deviation of 6.33%. The misclassification error parameter, chosen as the mode in the nested cross-validation, was 400, and the kernel regularization parameter, chosen as the mode, was 60.

4. Conclusions
The classification task in pattern recognition is one of the most used techniques to analyze databases. In our case, the task to develop is to classify the users of a videogame as “good player” or “bad player” knowing their muscle index fatigue and their games variables. To have a better understanding of our database, first was developed a correlation analysis and distance analysis, where was found that the features had a low correlation and the dimensionally of the database will not be reduced too much, and that a PCA extraction of features could be more convenient due to the low variance and that the projected space show separable classes. Based on the results, this was confirmed, due to the three classifiers showed a better percentage of accuracy with PCA than with Relieff, and that the number of features using PCA was lower than using Relieff (table 1). Even though the Linear classifier with PCA had better accuracy than the KNN with PCA, this last one was chosen as the best due to the number of features used was lower at least for 1000 characteristics.
Figure 8. Cross-validation for SVM classifier using PCA with the same number of relevant features than the best performance for KNN. The median was 82.66% with a standard deviation of 6.33%.

The SVM using the same number of features than KNN with PCA evidenced a higher percentage of accuracy, almost 10% more than the KNN. The KNN is sensitive to bad feature selection because of the variance of the data; as was explained before, our database has low variance so was expected that the SVM had a better performance. In the other hand, the SVM method is known for being more robust than the others, is better with high dimensional data, since it will only use the most relevant points to find a linear separation.

The higher accuracy of the SVM for sEMG contractions is accorded with the literature [15], [16], where SVMs were used for classification tasks of muscle fatigue sEMG signal. Based on the results, the SVMs also showed to be the appropriate pattern recognition method to classify information of sEMG signals combined with videogame variables. This classifier seems to be the best to be implemented on the videogame to provide information about the performance of each player during the interaction, and to record the progress therapy through the game’s sessions.

5. Acknowledgments
The authors gratefully acknowledge the support of the Master’s program of Electrical Engineering of the Universidad Tecnológica de Pereira, particularly to the teachers who kindly transmit their knowledge, and they also thank the Applied Physiology and Neuroscience group of the Universidad Tecnológica de Pereira, for the constant advisory and the provisional stance for HCI-group.

References
[1] Dix A 2009 Human-computer interaction Encyclopedia of database systems (Springer) pp 1327–1331.
[2] Fairclough S H and Gilleade K 2014 Advances in physiological computing (Springer).
[3] Cram J R 2011 Cram’s introduction to surface electromyography (Jones & Bartlett Learning).
[4] Yousefi J and Hamilton-Wright A 2014 Characterizing EMG data using machine-learning tools (Computers in biology and medicine vol. 51) pp 1–13
[5] Wu Q et al. 2016 Hybrid BF–PSO and fuzzy support vector machine for diagnosis of fatigue status using EMG signal features (Neurocomputing vol. 173) pp 483–500.
[6] Alkan A and Güney M 2012 Identification of EMG signals using discriminant analysis and SVM classifier (Expert Systems with Applications vol. 39 n.ª 1) pp 44–47.
[7] Bishop C M 2006 Pattern recognition and machine learning (Information science and statistics) (springer-verlag New York).
[8] De Luca C J 1997 The use of surface electromyography in biomechanics (Journal of applied biomechanics vol. 13 n.° 2) pp 135–163.

[9] Cifrek M et al 2009 Surface EMG based muscle fatigue evaluation in biomechanics (Clinical Biomechanics vol. 24 n.° 4) pp 327–340.

[10] Duda R O and Hart P E 1973 Pattern classification and scene analysis (New York: A Wiley-Interscience Publication).

[11] Kononenko I 1994 Estimating attributes: analysis and extensions of RELIEF (European conference on machine learning) pp 171–182.

[12] Murphy K P 2012 Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning. MIT press).

[13] Montoya M F, Muñoz J and Henao O A 2015 Surface EMG based muscle fatigue detection using a low-cost wearable sensor and amplitude-frequency analysis-Detección de la fatiga muscular a través de un sensor wearable de bajo costo y análisis de amplitud y frecuencia de la señal EMG superficial (Actas de Ingeniería vol. 1) pp 29–33.

[14] Montoya M F, Henao O A and Muñoz J 2017 Muscle fatigue detection through wearable sensors: a comparative study using the myo armband (ACM: Proceedings of the XVIII International Conference on Human Computer Interaction) p 30.

[15] Savur C and Sahin F 2015 Real-time american sign language recognition system using surface emg signal (Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on) pp 497–502.

[16] Wu Q et al 2017 Classification of EMG Signals by BFA-Optimized GSVCM for Diagnosis of Fatigue Status (IEEE Trans. Automation Science and Engineering vol. 14 n.° 2) pp 915–930.