Biometric system based on electroencephalogram analysis

Anonymous Author(s)
Affiliation
Address
email

Abstract

Recently, several researchers try to develop reliable biometric systems based on biological signals. Brainwave signals, like Electroencephalograms (EEG), are unique to each person; also, the EEGs are harder to steal and replicate than traditional biometrics like fingerprint or face recognition. Even though, there are many related works, to our knowledge none of them has studied what is the impact of the duration of the recorded signals in user identification accuracy. In order to answer this question, this work presents a method for the development of biometric systems based on EEG signals. The proposed method uses a Discrete Wavelet Transform (DWT) to extract relevant features, and a hyperparameter selection for adjusting the base models following a greedy strategy. In the task of user identification, using five classifiers as base models, the experiments showed that just 2 seconds of recording reaches an accuracy of approximately 90% and with 20 seconds the accuracy increases to 99%.

1 Introduction

Robust biometric systems development is a constant need since it is possible to break security on traditional methods like username/password, face or voice recognition [1]. Brain activity captured by electroencephalograms (EEG) has demonstrated good results as a biometric; EEGs are unique for each person, hence, any disruption in the individuals' behavior can lead to significant changes in the signal, resulting in an authentication failure [2]. However, a question arises: how much EEG recording is necessary to achieve a robust system? This work introduces a method for developing biometric systems, based on EEG signals recorded during an external stimulus [3]. The aim of this work is executing the proposed method to measure the achieved accuracy using different durations of recordings to provide a more proper understanding of how much time is needed to achieve higher precision in user identification seeking a trade-off between high accuracy and short duration.

2 Materials and Methods

The scheme workflow follows data preprocessing, feature extraction, model selection, and user identification. The details of each phase are discussed in the following subsections.

2.1 Data preprocessing

The EEG signals were downsampled to 128Hz, passed through a band-pass frequency filter from 4.0-45.0 Hz, and passed through a common average reference filter to improve the signal-to-noise ratio; furthermore, the electrooculographic (EOG) artifacts were removed. In addition, we applied two different discrete wavelet transforms (DWT). The first one with five levels of decomposition and

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

the second one with four levels of decomposition to making a comparison between them to recognize the best one for biometric systems based on EEG signals.

2.2 Feature extraction

The extracted wavelet coefficients show the energy distribution of the EEG signals in time and frequency in a compact representation. However, the choice of the features represents a critical step in all classification systems because of its direct influence on classification performance. Some authors use the coefficients directly as their feature vectors [4]-[6]. Nevertheless, other authors try decreasing the dimensionality of the feature vectors extracting higher level features such as: the maximum, minimum, mean, standard deviation, entropy or relative energy of the wavelet coefficients in each sub-band [7]-[11]. For the present work, relative wavelet energy was chosen as a feature because it has been shown to be very useful in classification tasks [12].

2.3 User Identification

In this work, five different classifiers establish the pool of base classifiers for selecting the appropriate model to the available data: Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), Random Forest (RF), AdaBoost (AB) and Multi-layer Perceptron MLP. To select the hyperparameters of each classifier, the proposed method uses a greedy search optimization approach. It is based on an exhaustive searching through a manually specified subset of the hyperparameter space of a learning algorithm [13].

3 Experimental Analysis

Two tests were performed. In the first one, we applied the feature extraction process described previously over the entire signals of the dataset, while in the second test we employed the same process for different recording durations. The segmentations were made at 0.25, 0.5, 1, 2, 4, 6, 8, 10, 20, 30, 40, 50, and 60 seconds, all of them starting at the beginning of the recordings. This work uses an open-access dataset of EEG signals recorded during sentiment reaction to YouTube videos provided by Koelstra et al. [3]. This dataset contains 40 recordings of 60 seconds for each of the 32 healthy participants.

3.1 Results and Discussion

The results of the first test shows that MLP reached the highest accuracy of 100% followed by SVM with 99.47%. On the other hand, the worst accuracy was given by K-NN classifier. A T-test using a confidence level of 95% showed that the accuracies between the two different decomposition levels were not statistically different. For this reason, Table 1 corresponds to the results of the second experiment using only a four level DWT because it is computationally cheaper.

Table 1: Comparison between accuracies for different recordings durations

Time (Sec)	**SVM (%)**	**RF (%)**	**KNN (%)**	**AB (%)**	**ANN (%)**
0,25	59,88±2,55	53,25±2,23	37,00±2,40	52,69±2,69	61,97±2,40
0,5	70,25±2,13	59,09±2,24	45,19±3,49	59,84±2,77	73,78±1,79
1	79,34±2,79	70,91±1,58	58,81±1,60	70,38±1,94	83,63±1,31
2	88,31±1,66	79,69±2,42	72,38±2,23	78,84±2,16	92,25±1,92
4	93,75±1,41	87,72±1,77	82,91±1,88	87,31±1,63	96,47±1,37
6	95,44±0,81	90,31±1,55	86,25±1,56	90,59±1,33	97,81±0,93
8	96,69±1,00	91,81±1,56	89,13±0,92	92,31±1,06	98,47±0,45
10	97,66±0,82	93,16±1,12	89,97±1,85	92,44±1,28	98,56±0,37
20	98,94±0,61	95,25±1,10	95,53±0,84	95,06±1,55	99,72±0,23
30	98,88±0,40	96,47±1,04	95,97±1,01	96,22±0,88	99,53±0,37
40	99,31±0,48	96,81±0,94	96,81±0,52	96,84±0,71	99,75±0,25
50	99,47±0,42	97,66±0,68	97,75±0,56	97,88±0,87	100,00±0,00
60	99,50±0,35	97,97±0,89	98,38±0,23	97,88±0,72	99,97±0,1

References

[1] Jain, A., Prabhakara, S. & Pankanti, S. (2003) Biometric recognition: Security and privacy concerns. *IEEE Security Privacy* **1**:33-42.

[2] Armstrong, B., Ruiz-Blondet, M., Khalifian, N., Kurtz, K., Jin, Z. & Laszlo, S. (2015) Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics. *Neurocomputing* **166**:59-67.

[3] Koelstra, S., Mühl, C., Soleymani, M., Lee, J-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A. & Patras, I. (2012) DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing* **3**:18-31.

[4] Orhan, U., Hekim, M. & Ozer, M. (2011) EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications* **38**:13475-13481.

[5] Übeyli, E. (2009) Combined neural network model employing wavelet coefficients for EEG signals classification. *Digital Signal Processing* **19**:297-308.

[6] Saini, R., Kaur, B., Singh, P., Kumar, P., Roy, P., Raman, B. & Singh, D. (2018) Don't just sign use brain too: A novel multimodal approach for user identification and verification. *Information Sciences* **430**:163-178.

[7] Amin, H., Malik, A., Ahmad, R., Badruddin, N., Kamel, N., Hussain, M. & Chooi, W-T. (2015) Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. *Australasian Physical Engineering Sciences in Medicine* **38**:139-149.

[8] Gandhi, T., Panigrahi, B. & Anand, S. (2011) A comparative study of wavelet families for EEG signal classification. *Neurocomputing* **74**:3051-3057.

[9] Jahankhani, P., Kodogiannis, V. & Revett, K. (2006) EEG signal classification using wavelet feature extraction and neural networks. *Proceedings of the IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing*, pp. 120–124. Sofia, Bulgaria: IEEE.

[10] Subasi, A. (2007) EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications* **32**:1084-1093.

[11] Pressel, G., Gareis, I. & Rufiner, H. (2017) Open access database of EEG signals recorded during imagined speech. In E. Romero, N. Lepore, J. Brieva, J. Brieva and I. Larrabide (eds.), *Proceedings Volume 10160, 12th International Symposium on Medical Information Processing and Analysis*, pp. 1–11. Spie.

[12] Koike-Akino, T., Mahajan, R., Marks, T., Wang, Y., Watanabe, S., Tuzel, O. & Orlik, P. (2016) High-accuracy user identification using EEG biometrics. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 854–858. Orlando, FL, USA: IEEE.

[13] Matuszyk, P., Castillo, R., Kottke, D. & Spiliopoulou, M. (2016) A comparative study on hyperparameter optimization for recommender systems. In E. Lex, R. Kern, A. Felferning, K. Jack, D. Kowald and E. Lacic (eds.), *Proceedings of the meeting of the Workshop on Recommender Systems and Big Data Analytics (RS-BDA'16)*. Austria: ACM Digitial Library.