A Rough Set Approach for User Identification Based on EEG Signals

Wael Khalifa, Mohamed I. Roushdy and Abdel-Badeeh M. Salem

Computer Science Dept., Faculty of Computer and Information Sciences,
Ain Shams University, Cairo, Egypt
Wael.khalifa@cis.asu.edu.eg, miroushdy@hotmail.com, abmsalem@yahoo.com

Abstract

Biometric technologies, referring to those that identify or verify the identity of a person using physiological (e.g. face and fingerprint) or behavioral characteristics (e.g., signature and voice), have the potential to solve many of the security problems. Traditional biometrics, such as facial patterns, fingerprints, eye irises, hand geometry and voice patterns, are well known for person authentication or identification purposes. Despite their widely used, such biometrics have certain limitations. It has been shown in previous studies that the brain-wave pattern for each individual is unique and thus can be used as a biometric. We present in this paper an intelligent rough set approach for user identification based on EEG Signals. We have tested the approach with various features set and a result o 92.6 accuracy was obtained with one of the featuresets.

Keywords: *EEG, human-computer interaction, authentication, machine learning*.

1. Introduction

EEG signals are brain activities recorded from electrodes mounted on the scalp. In comparison to other means for monitoring brain activities, such as magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI), electroencephalography is the most practical and the most used. EEG signals have been used in many brain-computer interfaces whose main goal is to enhance the quality of life of motor-disabled people. For biometrics, EEG signals has several advantages, It is confidential and hard to imitate, since EEG signals are a reflection of individual-dependent inner mental tasks. Moreover, one cannot force a person to give ideal EEG signals as those recorded in normal situations, as brain activity is easy to be influenced by the stress and mood of a person. In this sense, EEG-based biometrics can protect personal safety of its users [1].

An authentication system works by confirming or denying the identity claimed by a person (one to-one matching). While, an identification system tries to find the identity of a given person out of all system know people(one-to-N matching). Authentication and identification share the same preprocessing and feature extraction steps and a large part of the classifier design. Very little work has been done in the area of using EEG as a biometric. Several techniques has been used for identification and authentication such as autoregressive (AR) models with Kohonen's.Vector Quantizer (VQ) for classification, models based on the spectral power of the signal together with a fuzzy Neural Network for the classification. Moreover these techniques has been used, Decision Tree and Neural Network Classifier and hashing of MAR coefficients. [2][3][4][5]

The paper is divided as follows, in section two we discuss the medical aspects of the EEG signal. The used dataset is explained in section 3, The proposed Methodology is discussed in section four. Conclusions are presented in section five.

2. Medical Aspects of EEG

The brain contains about 100 billion neurons and weighs around 1.5 KG. Neurons are responsible of generate electrical signals. The sum of these electrical signals generates an electric field. Fluctuations in the electric field can be measured by devices and this is what we call Electroencephalographic (EEG) [6]. The electrical currents in the brain were discovered in 1875 by an English physician Richard Caton. In 1924 Hans Berger, a German neurologist, used his ordinary radio equipment to amplify the brain's electrical activity measured on the human scalp [7]. He announced that weak electric currents generated in the brain can be recorded without opening the skull, and depicted graphically on a strip of paper. The activity that he observed changed according to the functional status of the brain, such as in sleep, anesthesia, and lack of oxygen and in certain neural diseases, such as in epilepsy. [8]

EEG signals are generated from activities in the neurons[9]. EEG measures mostly the sum of the currents that flow during synaptic excitations of the dendrites of neurons depicted in Figure 1.

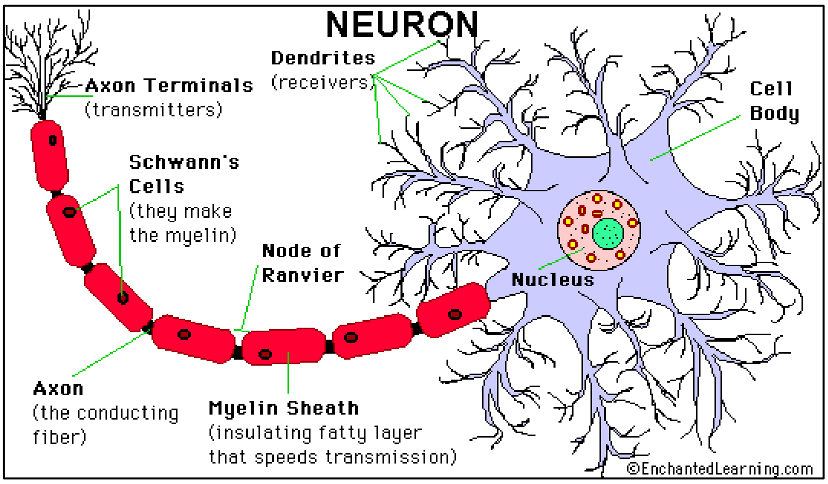

Figure 1. Basic anatomy of a typical cortical neuron, depicting the major input (dendrites), processing center (cell body), and the output region the axon. [10]

A typical EEG Signal capturing device consists of electrodes with conductive media, filters and amplifiers and analogue/digital converters. The internationally standardized 10-20 system is usually employed to record the EEG signals. The electrodes are located on the surface of the scalp, as shown in Figure 2 A. The positions are determined as follows: Reference points are nasion, which is the delve at the top of the nose, level with the eyes; and inion, which is the bony lump at the base of the skull on the midline at the back of the head. From these points, the skull perimeters are measured in the transverse and median planes. Electrode locations are determined by dividing these perimeters into 10% and 20% intervals. Three other electrodes are placed on each side equidistant from the neighboring points, as shown in Figure 2 B. [11]

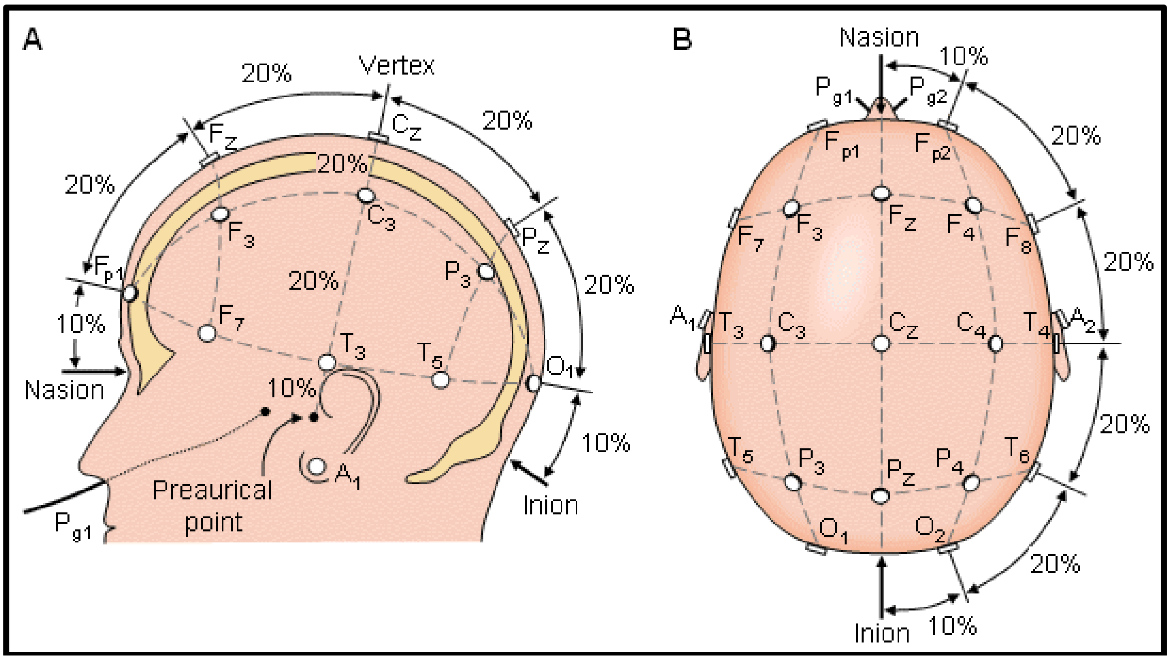

Figure 2 Electrode Placement [12]

3. Dataset

The dataset used is the EEG Motor Movement/Imagery Dataset[13] This dataset was created and contributed to PhysioNet[14][15] by the developers of the BCI2000 [16] instrumentation system, which they used in making these recordings. This data set consists of over 1500 one- and two-minute EEG recordings, obtained from 109 volunteers,

Subjects performed different motor/imagery tasks while 64-channel EEG were recorded using the BCI2000 system. Each subject performed 14 experimental runs: two one-minute baseline runs (one with eyes open, one with eyes closed), and three two-minute runs of each of the four following tasks:

1. A target appears on either the left or the right side of the screen. The subject opens and closes the corresponding fist until the target disappears. Then the subject relaxes.
2, A target appears on either the left or the right side of the screen. The subject imagines opening and closing the corresponding fist until the target disappears. Then the subject relaxes.
3. A target appears on either the top or the bottom of the screen. The subject opens and closes either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears. Then the subject relaxes.
4. A target appears on either the top or the bottom of the screen. The subject imagines opening and closing either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears. Then the subject relaxes.

In summary, the experimental runs were:

1. Baseline, eyes open
2. Baseline, eyes closed
3. Task 1 (open and close left or right fist)
4. Task 2 (imagine opening and closing left or right fist)
5. Task 3 (open and close both fists or both feet)
6. Task 4 (imagine opening and closing both fists or both feet)

The data are provided here in EDF+ format (containing 64 EEG signals, each sampled at 160 samples per second, and an annotation channel). The EEGs were recorded from 64 electrodes as per the international 10-10 system.

4. Rough Sets Approach for User Identification

Any biometric system is divided to two main modules an enrollment module and a verification module. Figure 3 shows the proposed rough set methodology for user identification system. During the enrollment phase, The EEG is captured by the capturing device. Then the captured signal is sent to the preprocessing module where the Nosie is filters, the channels are selected and the data is epcohed, details of the preprocessing methodology is explained in the next section. After the data is cleaned the clean data is sent to the feature extraction module to extract the unique features of the signal. After that the extracted features are sent for the rough set algorithm[19] to generate the rules.

In the verification phase, data is captured, cleaned and the feature are extracted. Based on the stored rules the comparison module will verify the user identity.

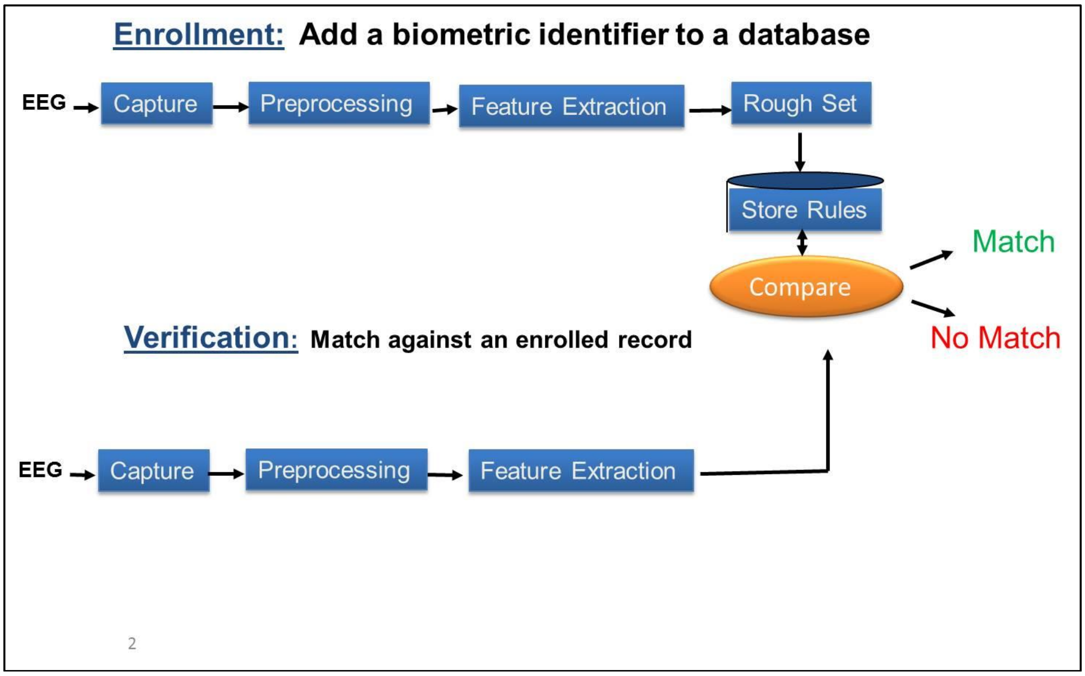

Figure 3 Proposed Identification Methodology

4.1 Data Cleaning

Figure 4 Shows the data cleaning Process. The following is a detailed account of each step.

Step1: Load the EEG data into EEG Lab for signal cleaning.

Step 2: Apply a Band pass filter from 0.5 Hz to 60 Hz.

Step 3: Apply a Notch filter to remove the 50 Hz line noise [17]

Step 4: Artifact Removal, we used Automatic Artifact removal toolbox [18] to remove Artifacts. Electrooculography (EOG) removal using the Blind Source Separation (BSS) algorithm then Electromyography (EMG) Removal using the same algorithm

Step 5: Since the major task performed by the user during the experiment is a motor function we selected the channels FC3, FCZ, FC4, C3,C1,CZ,C2 and C4 [16]. These channels are where the most motor activity data appear. Figure 5 shows the selected channels

Step 6: We focused on the left hand imaginary movement. The events were extracted that refer only to this event type. A single dataset per subject was generated representing the cleaned data with selected events per subject. The file was exported to a CSV format along with the events file to be used by the AIS algorithm.

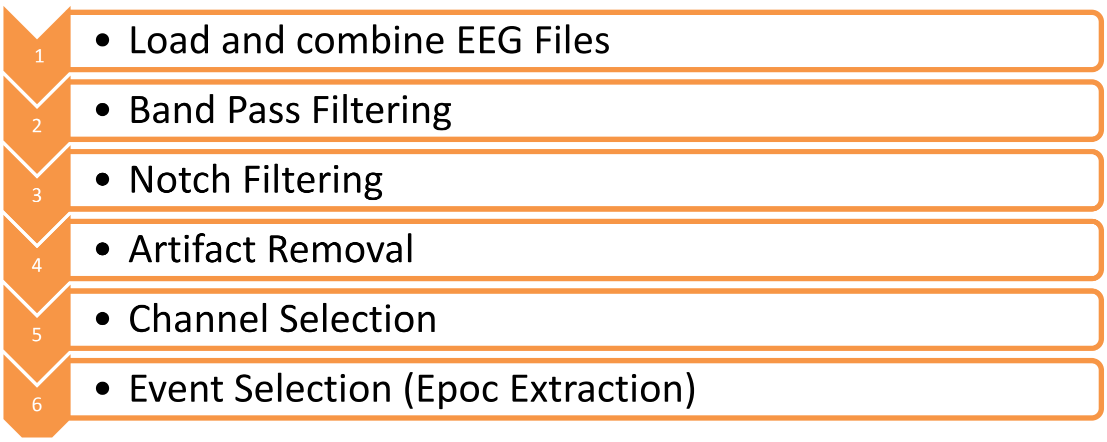

Figure 4 Data Cleaning Process

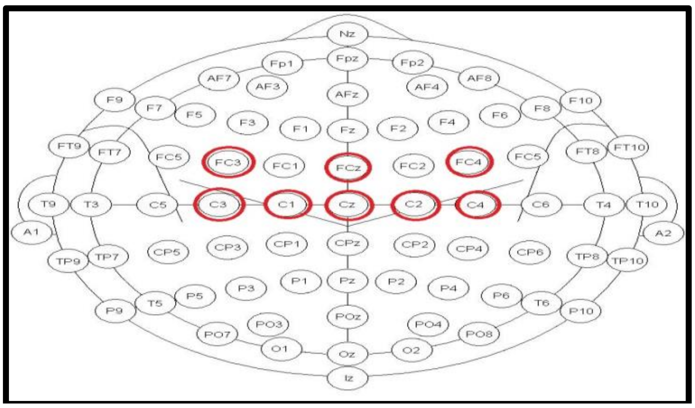

Figure 5 Selected Channels

4.2 Feature Extraction

We tried three different sets of features with rough set algorithm as described in the below sections

Time and Frequency Domain Features

We used the features from both the frequency domain and the time domain namely:

- Mean of each channel
- Average Power at each channel
- Average Power Spectral Density (PSD)
- Mean, Average Power, Average (PSD) at the alpha, beta, theta per channels

That gave us a total of 96 features per event. The data was processed by matlab and loaded in the Rough Set Exploration System(RSES) [20] for processing.

Auto regression coefficients

We used the auto regression coefficients for another trial we generated 3 coefficients per channel per band as the feature vector. This generated 72 features per event. The data was processed by matlab and loaded in the RSES for processing.

Frequency Domain Features

We used the features from only the frequency domain namely: Average Power value, Average Power phase shift, Average (PSD) at the alpha, beta, theta per channels .

That gave us a total of 72 features per event. The data was processed by matlab and loaded in the RSES for processing.

5. Results

The three features sets were used sepratly to test the proposed approach. We ran the three types of features in the RSES and used the cross validation with nine fold for classification testing. Using the Auto regression coefficients only gave the worst classification rate of about 35%. While using the time and frequency domain features not only it made the processing speed much worse but it did not improve the classification accuracy by much at each around we reached an accuract of 55%.

The best performance was with frequency domain features only and the classification rate reached 92.6%. Table 1 shows the summary of the rough set approach results as shown in Table 1.

Table 1 Rough Set Results Summary

Type	Feature Count	Accuracy
Auto Regression Coefficients	72	35%
Time and Frequency Features	96	55%
Frequency Features	72	92.6 %

6. Conclusion

We presented in this paper a new intelligent approach for user idenitification based on Rough Sets. We ran the three types of features namely, auto regression coefficients (72 features), time and freqency features(96 features) and frequency only features(72 features) Using the Auto regression coefficients gave the worst classification rate of about 35%. While using the time and frequency domain features gave an accuracy of 55%. The best accuracy was achieved with the frequency domain features and reached 92.6%. In the feature we will still expriment with various feature set to try to enahnce the system accuracy as well as test the approach on various data sets.

References

[1] Cota Navin Gupta, Ramaswamy Palaniappan Biometric Paradigm Using Visual Evoked Potential, Encyclopedia of Information Science and Technology, Second Edition, Vol1, 2009, pp. pp. 362-368.

[2] Alejandro Riera, Aureli Soria-Frisch, Marco Caparrini, Ivan Cester, and Giulio Ruffini, "Multimodal Physiological Biometrics Authentication", Biometrics: Theory, Methods, and Applications, Wiley Press, 2010.

[3] Ramaswamy Palaniappan,"Multiple Mental Thought Parametric Classification: A New Approach for Individual Identification", Proceeding of International Journal of Signal Processing, Volume 2, PP 222-226, 2006.

[4] Cempírek, M. - Šťastný, J. "The optimization of the EEG-based biometric classification." Applied Electronics., pp. 25-28. 2007.

[5] Shiliang Sun. "Multitask Learning for EEG-Based Biometrics", Proceeding of International Conference on Patter Recognition, , PP 51-55 ,2008

[6] H. L. Atwood, W. A. MacKay., "Essentials of neurophysiology," B.C. Decker, Hamilton, Canada, 1989.

[7] Berger, H., Ueber das Elektroenkephalogramm des Menschen, Archiv für Psychiatrie und Nervenkrankheiten, 1929, Berlin, 87: 527-570.

[8] Teplan, M, "Fundamentals of EEG Measurement," in Measurement Science Review, vol. 2, pp. 1–11, 2002.

[9] Guger C., Schlög A., Neuper C., Walterspacher D., Strein T., Pfurtscheller G., "Rapid Prototyping of an EEG-Based Brain-Computer Interface (BCI)", IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol.9, No. 1, March 2001

[10] http://www.enchantedlearning.com/subjects/anatomy/brain/gifs/Neuron.GIF (26/4/2014)

[11] Jaakko Malmivuo and Robert Plonsey ,Bioelectromagnetism Principles and Applications of Bioelectric and Biomagnetic Fields, Oxford University press, 1995

[12] http://www.bem.fi/book/13/13x/1302ax.gif (26/4/2014)

[13] Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE Transactions on Biomedical Engineering 51(6), 2004, PP 1034-1043.

[14] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23), 2000,e215-e220,

[15] Physionet: http://www.physionet.org/pn4/eegmmidb (26/4/2014(

[16] BCI2000: www.bci2000.org (26/4/2014(

[17] Automatic Artifact Removal Toolbox http://kasku.org/aar/ (26/4/2014)

[18] J. Sleight, P. Pillai, and S. Mohan, Classification of Executed and Imagined Motor Movement EEG Signals, Ann Arbor: University of Michigan, 2009, pp. 1-10

[19] Z. Pawlak. Rough sets. Int. J. Computer and Information Sciences, Volume 11, pp 341-356, 1982.

[20] http://www.mimuw.edu.pl/~szczuka/rses/start.html (26/4/2014)