Biometric Identification Using An Electroencephalogram Signals Compression Based on KC Function

Azmi Shawkat Abdulbaqi
University of Anbar-College of Computing and Information Technology
Azmi_msc@yahoo.com

Abstract
EEG (Electroencephalogram)is brain waves measure. The brain troubles are evaluated by EEG. It is used to locate of the activity in the brain during a seizure and to consider the patients whose suffer from brain functions problems. These troubles include tumors, coma, confusion and long-term difficulties (such as weakness associated with a stroke). The acquisition of EEG signals requires contact and liveliness and these signals are changes under stress that make so potentially unnecessary if it is acquired under menace. In this paper, an innovative and robust solution for this problem referred to earlier was introduced. To get this goal, the manner depends on models of various data compression models of information-theoretic plus the metrics symmetry related to Kolmogorov complexity. The proposed procedure make a compare between two EEG segments the procedure able to cluster the data in three collections: corresponding record, different participant, self-participant for the stratification of the proposed measure with values close to 0 for the same participant and closer to 1 for the different participants (entrants). The strategy was carried out to determine the participant in the database based on EEG signals. A 1-NN classifier was implemented, using a distance measurement method proposed in this scheme. The classifier was able to correctly identify almost all participants, with 96% accuracy in the underlying database.
Keyword: Kolmogorov Complexity (KC), Electroencephalogram (EEG), Signal Compression, Biometric .

التعريف البيومتري باستخدام ضغط الاشارات الكهربائية للدماغ المستندة لدالة Kolmogorov المعقدة

عزمي شوكت عبدالباقي
جامعة الانبار –كلية علوم الحاسبات وتكنولوجيا المعلومات

الملخص :
EEG (إليكترونسيفالوغرام) هو مقياس النشاط الكهربائي للدماغ . من خلال EEG يتم تقييم مشاكل الدماغ للمرضى حيث يتم تحديد فعالية الدماغ من خلال عينات مختلفة من المرضى الذين يعانون من مشاكل في الدماغ . هذه المشكل تشمل الاورام الدماغية والغيبوبة والارتباك والصعوبات على المدى الطويل مثل ضعف المرتبطة السكتة الدماغية. الاشارات الدماغية يتم الحصول عليها من خلال الاتصال المباشر بالمريض حيث ان هذه الاشارات تتغير تحت تأثير الاجهاد والضغط النفسي. في هذه الورقة، تم تقديم حل مبتكر وقوي لهذه المشكلة المشار إليها سابقا. ولغرض تحقيق الهدف، فإن الطريقة تعتمد على

موديلات لنماذج ضغط مختلفة للبيانات بالإضافة إلى تماثل المقاييس المتعلقة بدالة تعقيد Kolmogorov complexity . الإجراء المقترح يجعل مقارنة بين اثنين من EEG Segments الإجراء قادرا على تجميع البيانات في ثلاث مجموعات: قيود متماثلة وعدد من المشاركين بالاضافة الى المشارك الذاتي للتقسيم الطبقي المقترح، حيث يقاس من خلال القيم القريبة للصفر لنفس المشارك والقيم القريبة ١ لمشاركين متنوعين. وقد نفذت الاستراتيجية لتحديد المشاركين في قاعدة البيانات على أساس إشارات التخطيط الدماغي. وتم تنفيذ مصنف 1-NN باستخدام طريقة قياس المسافة المقترحة في هذا المخطط. وكان المصنف قادرا على التحديد بشكل صحيح لجميع المشاركين تقريبا، مع دقة ٩٦٪ في قاعدة البيانات الأساسية

1. Introduction

There is a necessary need to identification by EEG biometric. EEG has alterations in rhythm or amplitude, because of the circadian cycle or some particular circumstances. To understand EEG waves, the schemes may be strong to fluctuations. The system build to deal with different noise sources such as movement, muscle and so on[1]. The application has been proved a strength and efficiency in classification. This application of parameter-free such as compression algorithm. There are data mining parameters free used in this techniques to overcome the obstacles that affect traditional EEG methods based-on biometrics[2].

The KC used in this study is to conquer most of the problems and limitations. Here, K(x) of The KC of x can be defined as a minimum of software volume that produces the variable x then halt (x is a string of binary digits). The basic characteristic of the KC is that it's a non-mathematical method, to overcome this obstacle, it is usually approximated by some computable measure, such as linguistic, compression-based complexity measures, Lempel-Ziv[3]. Fig (1) show EEG standard signals and fig. (2) show the different normal EEG waves of the normal human. The approximations supply maximum borders of complexity that permit developing the similarity measurements and dissimilarity measurements for recent applications of this notion. The idea in this proposed work is to the determination of EEG Biometrics scheme based on theoretical data models for data compression[4].

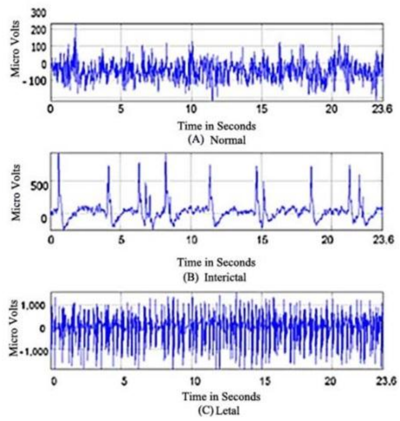


Fig. (1) Sample of EEG Standard Signals

2. Related Works

Recently, EEG signals compression field has been very popular in the area of information technology transfer. The following are some of the researchers who presented their research in that field:

In 2012, N. Sriraam presented a novel and an efficient high performance EEG lossless compression by using WT (wavelet transform) with artificial neural networks predictors. The scheme involves pre-processing by using IWT (Integer Wavelet Transform) , prediction using ANN (Artificial Neural Networks) predictors followed by adjusting the offset of the prediction residue through improved context-based bias cancellation[5].

In 2013, Dr.T.V.U.Kiran presented a proposed a novel method for EEG signal compression based on the DWPT (discrete wavelet packet transform) decomposition. The produced method provides an adaptive approach for optimal EEG signal representation in order to compression and applied to any type of 1-D biomedical EEG signal[6].

In 2015, Ph. Thi. and H. Ng. presents a survey of existing lossy compression algorithms reported in the last two decades and attempt to analyze the algorithms and provide a qualitative comparison among them[7].

4. Kolmogorov Complexity (KC)

KC in the information theory is the length of the shortest computer program (such as text length) that produces the object as output. KC considers computational resources need to appoint the object and also known as KC-Chaitin, program-size complexity, descriptive complexity, or algorithmic entropy. If the string like this 111111 ... which continues to count percent in the same way. The length of the string is 100 characters, but you can write a short program that generates very easily. Let us now consider the string "232046622087638 .." and so on for a hundred numbers[8]. It is supposed to be a random string and it would be very difficult to create a program that could print that was shorter than it is. In other words, there is no way to specify this seemingly random string apart from the name. This observation of the difference between these two chains is what led to the idea of KC. The equation of KC is describe below[9]:

$$Cj(x) = \begin{cases} \min(|p|: f(p) = x) & if\ x \in ran\ f \\ \infty & Otherwise \end{cases}$$

5. Symbolic Aggregate Approximation Algorithm (SAXA)

SAXA is an application algorithm that convert the input time series transforms into a strings. The algorithm is based on the inheriting of the simplicity of the original algorithm and low computational complexity while providing a satisfactory selectivity in range query processing. SAXA transforms a time-series Y of length n into the string of arbitrary length xx, where x«nx«n typically, using an alphabet A of size a>2. PAA approximates a time-series X oflength n into sector X'=(x'1,…,x'M)X'=(x'1,…,x'M) of any arbitrary length M≤nM≤n where each of xi¯xi¯ is calculated as follows[10]:

$$c^{*}i = alpha * j\ ,iif\ ,C' * i\ \in [\beta j - 1 - \boldsymbol{\beta j})$$

$$\boldsymbol{MINDIST(Q^\wedge, C^\wedge) \equiv \frac{\sqrt{n}}{\sqrt{w}} \sqrt{\sum_{i=1}^{w} \big(dist(q * i, c * i)\big)2}}$$

6. Procedure Description

6.2. Quantization

The algorithm based on inheriting the original algorithm simplicity and low computational complexity while providing satisfactory sensitivity and selectivity in range query processing. Patients medical records (PMR) are digital and symbolic records due to they are stored into PC, and for this reason, the compression algorithms are effective when applied to symbolic records, however, the EEG is the real-valued numerical signal. In this case, converting the real EEG signals to the symbolic signal is the pivotal step must be implemented. To complete converting step, SAX algorithm must be used. The converting method consists of divides N segments in the time series and calculates its mean value that is matching any symbol in the new data dimension. The main factor of this method is the size of alphabet and w dimension of the series segments[11].

6.3 Pearson Correlation (Moment Correlation) (PrCo) or (MoCo)

PrCo coefficient is the variance between two-variable and the output is divided by its standard deviations. The above definition format includes the product time, ie the product average of random variables adjusted according to the mean; Hence the product time-modifier in the name. Data sets correlation can be measured of how well they are related. The most discussed measure of the correlation in stats is the (PrCo). The PrCo offer the linear relationship between two data sets. There are mainly two letters are used to represent the PrCo or MoCo[12]:

- Greek letter rho (ρ) for a population.
- The letter "r" for a sample.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x2 - (\sum x)][n \sum y2 - (\sum y)2]}}$$

The principle of algorithm action is the size of the alphabet and the vector w of the string the series segments .

6.4. Compression Method

The measure of similarity between binary objects is the main problems dealt with using the KC theory. A method was suggested, based-on the distance information depends on the length of the shorter binary program. This distance relies on conditional complexities. An algorithm able to collect the data knowledge to be compressed was proposed in this paper in order to understand the basic of KC.

To create an internal model of the data collected, it must be collected statistics. The measure of similarity between binary objects based on the notion of relative compression of two objects by using the information of another object. The NRC (normalized relative compression) of x given y can be describe by :

$$NRC(x,y) = \frac{C(x||y)}{|x|}$$

Where |x| is object size.

6.5. The Statistical Analysis

Each participant has 4 segment compose the training dataset, the residual segment was used as a test dataset. There are two steps to evaluate the proposed algorithm:

1st Step: Results such as mean standard deviation were presented for calculated measurement NRC between records. The results of the collected self-assessment values (the NRC value of record), participant measure, out measure (NRC value computed between the records of the same participant, excluding self-similarity and NRC value of computed between records of different participants).

2nd Step: Performance assessing of the identification method is the basic theory in this research paper that is evaluated by sensitivity means , accuracy and specificity (Eq.2,3 and 4)

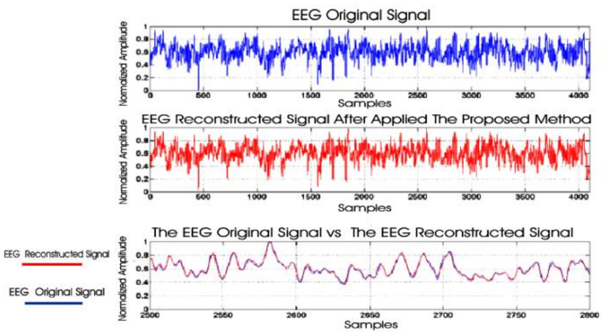
$$sen = \frac{TP}{TP+FN'} \quad \text{……..(2)}$$

$$spec = \frac{TN}{FP+TN'} \quad \text{……..(3)}$$

$$acc = \frac{TP+TN}{TP+FP+TN+FN'} \quad \text{……..(4)}$$

The number of true positives refers to TP, the number of true negatives refers to TN, the number of false positives refers to FP, and the number of false negatives refers to FN. Below fig(3) show the original EEG signals and reconstructed signals.

Fig (3) The EEG Original and Reconstructed Signals



7. Results and Discussion

The CHB-MIT scalp EEG database are used in this study. The healthy entrants are just considered. There are 40 entrants used in this database. Some participants having data collected in different days. The testing depend on training one record for each participant divide it into five segments of 20 seconds. The initial 15 participants are used to optimize model parameters, and the rest participants used for parameters evaluation. The performance of quantization was rated by using reciprocal information, Pearson correlation and the Euclidean distance between EEG signals after quantization.

Table 1 described the process of biometric identification. The NCR is able to distinguish between the three data groups, the training segments, the same or the different participant: (a). EEG training segments with mean value of 0:140; (b). EEG testing segment from the same participant with mean value of 0:300); and (c). EEG testing segments from different participants (with mean value of 0:410.

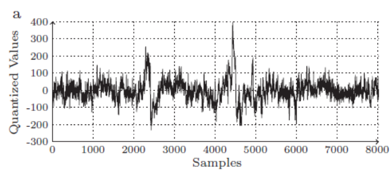
Table (1): Show the performance analysis of EEG depending on PRD,PSNR and MSE

EEG Dataset	Compression Ratio(CR)	Time Duration (sec)	PSNR	MSE	PRD(%)	
					Proposed Method	Other Methods (Average methods)
Dataset(1)	20	1.44	12.989	2.04E+06	9	7
Dataset(2)	30	1.42	17.46	8.94E+05	13	11
Dataset(3)	40	1.42	22.794	7.99E-16	19	16
Dataset(4)	50	1.43	25.336	9.31E+12	23	20

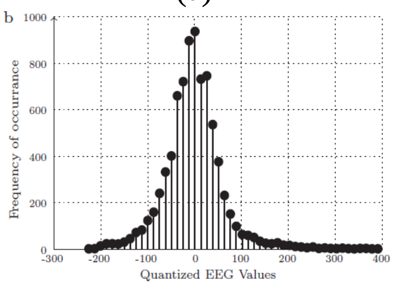
Table (2): Show the quantization levels of different samples of dataset

EEG Dataset	Threshold Level (%)	Quantization Level (1)	Quantization Level (2)	Quantization Level (3)
Data Set(1)	2	10	9	8
Data Set(2)	5	9	8	7
Data Set(3)	9	6	5	4

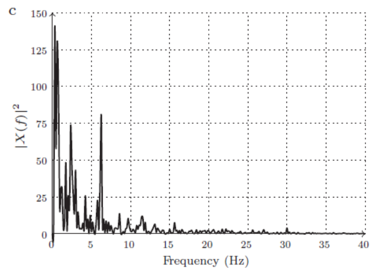
The amplitude distribution and spectral distribution of EEG segment is shown in figures below:



(a)



(b)



(c)

Fig (4): (a) Normal EEG Signal, (b) Amplitude Spectra and (c) Power Spectra of the Signal shown in (a) .

Table (4) A gaining Results from A Tested Samples

Patient ID	P(1)	P(2)	P(3)	P(4)
Primary CR	93.14	93.2	93.1	93.22
Optimized CR	92	91	91	92
Primary PRD	77.41	56.77	56.01	69.86
Optimized PRD	8.8	7.25	8.32	8.46
Stopping Iteration	2	3	2	3

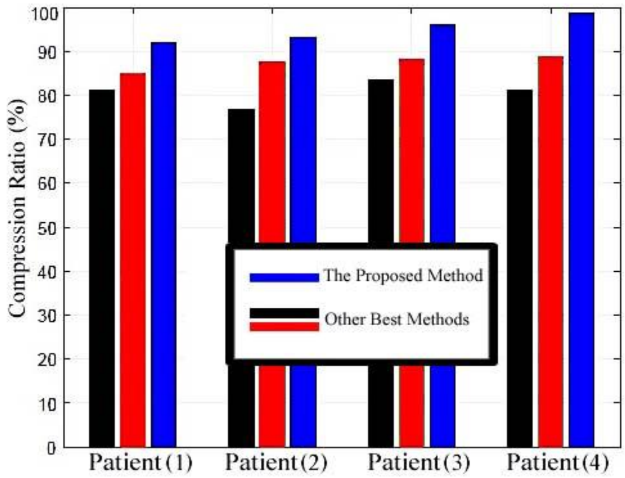


Fig. (5) A Comparison between A proposed Algorithm Using KC Function with some other Algorithms.

8. Conclusion

The proposed model allowed to comparing two EEG segments without applying the traditional approaches that need to brainwaves segmentation. The model was described to be feasible approach to EEG biometric analysis, due to this method presents 99% for accuracy on the database used. The model suffer from some restricted, therefore it need to more improving acceptance rate. There are validated considering for EEG waves method were collected in different days an considering external variability conditions. The method can be implemented not just for identification, but also for authentication. Each new record will be compared to all archived records and only if the NCR value is within the confidence interval of the acceptance to a given person, it will authenticate itself.

References

[1] A. Z. M. et al." State-of-Art Methods and Future Perspectives for Personal Recognition Based on EEG Signals IET Biometrics", Vol. 4, p.p. 179–90, 2015 .

[2] B. K. and K. B. , "Subject Identification from Electroencephalogram (EEG) signals During Imagined Speech Proc. of the 4th IEEE International Conference on Biometrics: Theory Apps. and Sys. p.p. 1–8, 2010.

[3] J.F. ,"Multi feature Biometric System Based on EEG Signals", In Proceedings of the 2nd Int. Conf. on Interaction Sci., P.P. 1341–1345, 2009.

[4] N. S., "A High-Performance Lossless Compression Scheme for EEG Signals Using Wavelet Transform and Neural Network Predictors", International Journal of Telemedicine and Applications(IJTA), Hindawi Publishing Corporation, , Vol.2012, Article ID 302581, 8 pages, 2012.

[5] Dr.T.V.U.Kiran Kumar1, "Generation and reconstruction of EEG signals using MATLAB", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE), Vol. 2, Issue 9, September, 2013.

[6] Ph. Thi.and H. Ng., "Lossy Compression Techniques for EEG Signals", International Conference on Advanced Technologies for Communications (IATC), 2015. 978-1-4673-8374-5/15, IEEE.

[7] A. R. et, al. "Unobtrusive Biometric System Based on Electroencephalogram (EEG) Analysis ", Journal on Advances in Signal Processing", 2008.

[8] M. A. et al. , "PCG Biometric Identification System Based on Feature Level Fusion using Canonical Correlation Analysis", Proceeding of IEEE 27th Canadian Conference of Elec. and Comp. Eng. (CCECE), p.p. 1–6, 2014.

[9] M. Li et al.,"An introduction to KC(Kolmogorov Complexity) and its Applications", Springer Science & Business Media(SSBM), 2009.

[10] R.K. and S.F.,"An Energy Efficient Compressed Sensing Framework for the Compression of EEG (Electroencephalogram Signals)", Sensors, Vol. 14, No.1, p.p. 1474–1496, 2014.

[11] A.M and A.S., "Row-Sparse Blind Compressed Sensing for Reconstructing Multichannel EEG (Electroencephalogram Signals)", Biomedical Signal Processing and Control(BSPC), 2015.

[12] M.R. and D.D.," Lightweight Lossy Compression of Biometric Patterns Via Denoising Auto Encoders", IEEE Signal Processing Letters, Vol. 22, p.p. 2304–2308, 2015.

[13] Buda et al.," Life Time of Correlations and its Applications", p.p. 5-21. , Publisher: Wydawnictwo Niezależne, ISBN: 9788391527290, 2010.