# EEG Data Similarity Using Lempel–Ziv Complexity

Ibrahim Salem Jahan, Michal Prilepok, and Vaclav Snasel

VSB – Technical University of Ostrava,
17. listopadu 15, 708 33 Ostrava – Poruba, Czech Republic
Jahan_nw@yahoo.com, {michal.prilepok,vaclav.snasel}@vsb.cz

**Abstract.** Today still big challenge in world is to find efficient technique for perform recognition on mental tasks and distinguish between them. These allow us to use Brain Computer Interface applications to help disabled people to interact with environment and control external devices such as wheel chair. In this article we used EEG data from National University of Sciences and Technology, Pakistan, which are available online. We made our experiments on signals from one subject performing hand movement task. First we applied Faster Fourier Transformer (FFT), removing the EEG higher frequencies, applying the inverse Fourier transformer then converting EEG data into graphics by turtle graphics, then find the similarity between these trials by Lempel–Ziv complexity, to find maximum similarity between EEG data for same mental task. Our model reached average accuracy up to 52.63%.

**Keywords:** Electroencephalograph (EEG), Neuron, Lempel–Ziv Complexity, Turtle Graphics, EEG Data Similarity.

## 1 Introduction

There are several algorithms to analyze similarity and recognition of mental tasks, but still remains big challenge in world to find very efficient method for analysis and recognition human mental tasks. We will use Lempel–Ziv Complexity technique to finding similarity between EEG data and distinguish between human mental tasks. There are several approaches to classify EEG signal, they include Support Vector Machines (SVM), L1 regularized logistic regression and non-negative matrix factorization (NMF) [1]. In this paper we will give some glance at EEG, Lempel–Ziv Complexity, Turtles graphics and applying proposed method for Electroencephalograph (EEG) data similarity.

## 2 Introduction to EEG

EEG include on complex irregular signals that may provide useful information denote about underlying neural activities of the brain [2]. The human brain electrical activity has been recognized from more than a century. It is defined as that variation of the surface potential distribution on the scalp that reflects functional activities emerging from the underlying brain [3]. This electrical surface potential variation can be

I. Zelinka et al. (eds.), *AETA 2013: Recent Advances in Electrical Engineering and Related Sciences*, Lecture Notes in Electrical Engineering 282,
DOI: 10.1007/978-3-642-41968-3_30, © Springer-Verlag Berlin Heidelberg 2014

recorded by placing set of sensors on the scalp, and measuring the electrical signal between couples of these sensors after that filtered, amplified, and recorded these electrical signals. The final result of these data is called the Electroencephalograph (EEG) [3].

### 2.1 Source of EEG Generating

The EEG signals define as measurements of the currents when flowing during synaptic excitations of the dendrites of multiple pyramidal neuron in the cerebral cortex. When brain cells are activated, the synaptic currents are produced within the dendrites. Normally this current producing a magnetic field can be measurable by electromyogram (EMG) machines and an electrical field over the scalp measurable by EEG techniques. Basically the current in each neuron cell of brain is produced from pumping the positive ions of calcium, sodium, and potassium, and negative ions of chlorine, through the neuron membranes in the direction governed by the membrane potential, as structure of neuron in figure (1) [4].

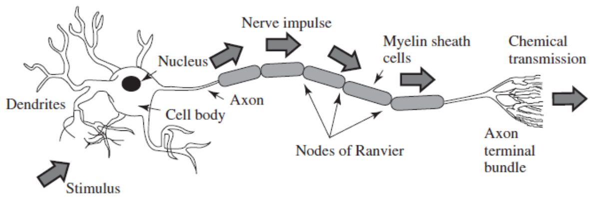


**Fig. 1.** Structure of Neuron [4]

## 3 Faster Fourier Transform (FFT)

A Fourier transform, converting a function from time domain to frequency domain and vice versa. The faster Fourier Transform (FFT) is algorithm that can compute the discrete Fourier transform (DFT) and inverse of DFT, also FFT is an efficient to compute the discrete Fourier transform (DFT). The FFT is important in field frequency analysis, because it takes a discrete signal in time domain and converts this signal to discrete frequency domain representation [5]. FFT and IFFT algorithms have been well know and widely used in several applications due to their efficiency [6].

## 4 Turtle Graphic

The Turtles graphic is an easy way for representation complicated geometric object. Turtle Graphic or L-systems method is used for making the graphics. The basic idea to make graphics by turtle is convert the graphic into sequence of commands, which control turtle and allow to making the specific graphic [7]. Turtle geometry has been used to study and representation many various subjects from simple polygons to complex fractals [8]. To understand turtle geometry, we will explain that it by a virtual turtle. The virtual turtle must know own position, facing direction, and step size, it to follow some commands to change own position, or heading, or notion scale [8]. For example we have a turtle on a plane. The location of turtle on the plane can be represented by a point $A$ given by pair of $(a1, a2)$, also the turtle heading can represent by vector $V$ given by pair of $(v1, v2)$, the length of vector $V$ denote on turtle step size, the pair $(A, V)$ denote on turtle state.

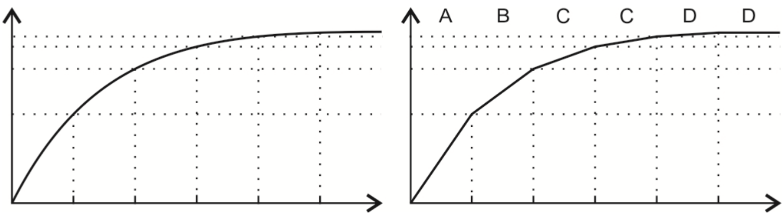


**Fig. 2.** Data conversion

On the left figure on Fig. 2 we can see interpretation in a line chart of the measured raw data. The right figure shows interpretation measured after conversion into turtle graphic command. The command are on the top edge of figure. The final commands sequence for our example is ABCCDD. In this example we have only four commands. The C and D commands doubled, because third and fourth angles have same value. In case of D command is the situation same to command C.

## 5 Lempel–Ziv Complexity

The Lempel–Ziv (LZ) complexity for sequences of finite length was suggested by Lempel and Ziv [9]. It is a non-parametric, simple-to-calculate measure of complexity in a one-dimensional. LZ complexity is related to the number of distinct substrings and the rate of their recurrence along the given sequence [10], with larger values corresponding to more complexity in the data. It has been applied to study the brain function, detect ventricular tachychardia, fibrillation and EEG [11]. It has been applied to extract complexity from mutual information time series of EEGs in order to predict response during isoflurane anaesthesia with artificial neural networks. [12]

LZ complexity analysis is based on a coarse-graining of the measurements, so before calculating the complexity measure $c(n)$, the signal must be transformed into a finite symbol sequence. In this study we have used turtle graphic for conversion measured data into finite symbol sequence $P$.

The sequence P is scanned from left to right and the complexity counter $c(n)$ is increased by one unit every time a new subsequence of consecutive characters is encountered. The complexity measure can be estimated using the following algorithm [9] and [12]:

1. Let $S$ and $Q$ denote two subsequences of $P$ and $SQ$ be the concatenation of $S$ and $Q$, while sequence $SQ\pi$ is derived from $SQ$ after its last character is deleted ($\pi$ means the operation to delete the last character in the sequence). Let $v(SQ\pi)$ denote the vocabulary of all different subsequences of $SQ\pi$. At the beginning, $c(n) = 1$, $S = s(1)$, $Q = s(2)$, therefore, $SQ\pi = s(1)$.
2. In general, $S = s(1)$, $s(2)$, …, $s(r)$, $Q = s(r + 1)$, then $SQ\pi = s(1), s(2), \dots, s(r)$; if $Q$ belongs to $v(SQ\pi)$, then $Q$ is a subsequence of $SQ\pi$, not a new sequence.
3. Renew $Q$ to be $s(r + 1)$, $s(r + 2)$ and judge if $Q$ belongs to $v(SQ\pi)$ or not.
4. Repeat the previous steps until $Q$ does not belong to $v(SQ\pi)$. Now $Q = s(r + 1)$, $s(r + 2)$, …, $s(r + i)$ is not a subsequence of $SQ\pi = s(1)$, $s(2)$, …, $s(r + i - 1)$, so increase $c(n)$ by one.
5. Thereafter, $S$ is renewed to be $S = s(1)$, $s(2)$, …, $s(r + i)$, and $Q = s(r + i + 1)$.

These procedures have to be repeated until $Q$ is the last character. At this time the number of different subsequences in P – the measure of complexity – is $c(n)$.

In our experiment we do not deal with measure of the complexity. From the individual subsequences we create a list of then. One list is created for each data file with turtle commands of the compared commands files.

### 5.1 Comparing Data Using LZ Complexity and Turtle Graphics Commands

The comparison of the LZ sequence lists is the main task. The lists are compared to each other. The main property for comparison is the number of common sequences in the lists. This number is represented by the sc parameter in the following formula (1), which is a metric of similarity between two turtle commands lists.

$$SM = \frac{sc}{\min(c_1,c_2)} \tag{1}$$

Where

- $sc$ Count of common LZ sequences in both dictionaries.
- $c_1, c_2$ Count of LZ sequences in list of the first or the second file.

The SM value is in the range between 0 and 1. If $SM = 1$, then the commands lists are equal and they have the highest difference (have nothing common), when the result value of $SM = 0$.

## 6 EEG Experiments

### 6.1 EEG Data

In our experiment we used EEG data that available online from National University of Sciences and Technology, Pakistan, we choosing Dataset 2 - 2D motion. The EEG data raw was recorded at 500Hz, from a subject male 21 years old using 19 electrodes FP1 FP2 F3 F4 C3 C4 P3 P4 O1 O2 F7 F8 T3 T4 T5 T6 FZ CZ PZ, consisting on several trials of hand and leg movements. In our experiments we used left hand back movement trial - LeftBackward1, left hand back movement trial - LeftBackward2, left hand back movement trial - LeftBackward3, left hand back Imaging movement - Left Backward Imagined.

### 6.2 EEG Data Preparation

The EEG data are prepared in following steps. As a first step separate dataset into individual mental tasks, trails and sensors. We got 122 data parts. In the second step of our process we applied Faster Fourier Transform to transform raw sensor data from time domain into frequency domain. In the frequency domain we removed higher frequencies above than 150Hz. In the next step we applied Inverse FFT to convert data back from frequency domain into time domain. This filtered data we converted using turtle graphics into text format. For the turtle graphic we used 128 commands. Each command represents an angle in the selected first and fourth quadrant. We used only first and fourth quadrat, because the time in data line goes from left to right and the signal does not go backwards.

After that each EEG trial were prepared by LZ complexity to get LZ subsequences from turtle commands list. For each data trial we created a list of LZ subsequences. We compared training and testing lists to find the maximum similarity between EEG trials of same mental task.

### 6.3 Experiment Results

We made similarity between the EEG trials for left hand back movement and imaging left hand back movement task, to find maximum similarity between different EEG trials of the same mental task. Our results are listed in the Table 1. The maximum similarity results of mental tasks by our method reached to 100.00%, minimum similarity was 30.00% and average value of similarity was 52.63%. Our suggested model reached accuracy up to 52.63%.

**Table 1.** Similarity results

| | Minimum | Maximum | Average |
|---|---|---|---|
| **Correctly identified** | 30.00% | 100.00% | **52.63%** |
| **Incorrectly identified** | 0.00% | 70.00% | **47.37%** |
| **True positive rate** | 0.00% | 100.0% | **35.53%** |
| **False positive rate** | 0.00% | 100.0% | **55.26%** |
| **Accuracy** | 30.00% | 100.00% | **52.63%** |

In Figure 3 we can see accuracy of all used sensors. The accuracy of sensor nr. 1 reached to 100%, sensors nr. 5 and 6 reached to 80%, the most accuracy sensors values was between 40% and 60%.

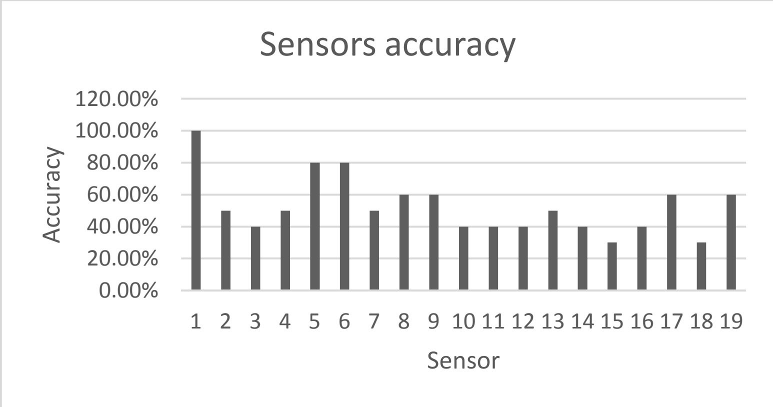


**Fig. 3.** Sensors accuracy

# 7 Conclusion

We made our experiments on EEG signals from one subject performing left hand back movement task in three trials, and other trial for imaging left hand back movement, we applied FFT to EEG data, removing high frequencies, applied Invers FFT, represent EEG data by turtle graphics, then finding the maximum similarity between these trials by LZ complexity. The experiment results on EEG data showed the maximum similarity results of mental tasks by our method reach to 100%, minimum similarity was 30.00% and average value of similarity was 52.63%. Our model reached accuracy up to 52.63%. In future work we will try to collect EEG data using Emotiv EEG neuroheadset, and use this data to find similarity between trials of mental tasks by our proposed method to analysis and recognition on mental tasks.

**Acknowledgements.** This work was partially supported by the Grant of SGS No. SP2013/70, VŠB - Technical University of Ostrava, Czech Republic, and was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073 funded by Operational Program Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic.

# References

1. Shin, B., Oh, A.: Bayesian Group Nonnegative Matrix Factorization for EEG Analysis (December 2012)
2. Bhattacharya, J., Petsche, H.: Universality in the brain while listening to music. Proc. the Royal Society Lond. B 268, 2423–2433 (2001)

3. Ochoa, J.B., Molina, G.G., Ebrahimi, T.: EEG Signal Classification for Brain Computer Interface Applications (2002)
4. Sanei, S., Chambers, J.A.: EEG Signal Processing. John Wiley & Sons Ltd., The Atrium (2007)
5. Garcia, J., Castillo, C.: Frequncy-change Analysis of Nonlinear System Using Spectrum (Fast Fourier Transform) Theory. In: 6th International Conference Electrical Engineering, Computing Science and Automatic Control, CCE, 2009, pp. 1–6 (2009)
6. Perko, M., Fajfar, I., Tuma, T., Puhan, J.: Fast Fourier transform computation using digital CNN simulator. In: 1998 Fifth IEEE International Workshop Cellular Neural Networks and Their Applications Proceedings, London, pp. 230–235 (1998)
7. Prusinkiewicz, P.: Graphical applications of L-system. In: Graphical Interface, pp. 247–253 (1986)
8. Goldman, R., Schaefer, S., Ju, T.: Turtle geometry in computer graphics and computer-aided design, pp. 1471–1482 (2004)
9. Lempel, A., Ziv, J.: On the complexity of finite sequences, pp. 75–81 (1976)
10. Radhakrishnan, N., Gangadhar, B.N.: Estimating regularity in epileptic seizure time-series data, pp. 89–94 (1998)
11. Zhang, X.S., Roy, R.J., Jensen, E.W.: EEG complexity as a measure of depth of anesthesia for patients, pp. 1424–1433 (2001)
12. Abásolo, D., Hornero, R., Gómez, C., García, M., López, M.: Analysis of EEG background activity in Alzheimer's disease patients with Lempel–Ziv complexity and central tendency measure, pp. 315–322 (2006)