Study of neurophysiological metrics-indexes, in order to predict learning difficulties

Stefanidis V., Papavlasopoulos S., and Poulos M.
Information Science and Informatics,
Ionian University Corfu,
Ioannou Theotoki 72,
Corfu 49100, Greece
{vstefan, sozon, mpoulos}@ionio.gr

Abstract : In this paper we describe a study about predictability of brain reactions during educational procedures. More specifically we research neurophysiological indexes, that are able to describe the scalable difficulty of educational procedures. Using electroencephalographic (EEG) signals we extracted a big amount of feature vectors based on entropies and using proper algorithms and data mining methods we found out good practices in order to predict and describe more efficiently the differentiations of brain reaction in various math problems with scalable difficulty. We use this indexes in order to identify and to predict difficulties in educational procedures. Entropies are indexes that reflects fluctuations in brain activity which is imprinted in brain map. Furthermore we use classification methods in order to find a proper mathematical model to predict learning difficulties.

Key-Words: Brain, Cognition, Educational Difficulties, EEG, Neurophysiology

Introduction

The human brain recognizes and processes educational information from his early childhood. A physiological baby is capable of distinguishing whether an object is bigger than another (comparison), while understanding that five objects are more than two (quantity). The comparison are fundamentals on algorithmic cognitions and they are perceived from human brain before having begun to speak and execute other basic operations.

In the present work we study neurophysiological models capable of describing the different operation of human brain in relation to the difficulty of educational-identification problems. Similar studies have also become in other discipline fields of neurophysiology [2],[3],[4],[9],[10] that have contributed in better comprehension of cortical activity and have opened the horizons for new studies. Within this work we studied the way of reaction of human brain in various learning stimuli with scalable difficulty.

In [28] the authors refer to the differentiation of the signals during the image recognition. The main topic of this research is not only the visualization of the brain potentials, but also does the research refer to the differentiation of the potentials among
the signals produced when the subject tries to distinguish an easy picture from a hard one. This differentiation noticed at the brain activity is visually made from the grand average figure 1, which is at the second page of the research at [8]. The authors don’t use neither a sorting method nor a certain mathematical model. They only do a visual exploration. There is also an effort made in order to clarify the differentiation of the data from the table 1, page 4 at [28], in which the reaction times are presented.
At [29] the authors evaluate the possibility of the cerebral electrical activity’s power spectrum to show differences between people who have normal understanding abilities and others who have learning difficulties. Another goal of this specific research is to correlate this electrical activity with ingenuity, the school activity and the neurophysiological output from several tests. The methods that are used are the classical signal analysis methods like cross validation or the cross correlation of 10 variables and the discriminant functions. The sorting was done with several independent classifiers like replication LD.

There was also correlation made among behavioral variables like Wide Range Achievement.

Method of study, tools and neurophysiological metrics-indexes

In this study participated physiologic individuals (the subjects), which do not suffer from any learning stress. Subjects, were tested in educational procedures (identification tasks) of increasing difficulty (easy figures recognition EFR, hard figures recognition HFR, there is an animal into the figure AIF, there is not animal into the figure NAIF), while simultaneously was recorded their cerebral activity using Electroencephalography method. The data for this study are distributed from researchers under GNU/GPL permission in the follow URL :
http://sccn.ucsd.edu/~arno/fam2data/publicly_available_EEG_data.html

Then we processed the cerebral signals, with final aim to export neurophysiological indexes that are able to describe the difficulty that the human brain faces in his effort to resolve identification procedures. Having in mind the nonlinear and dynamic nature of operation of brain we studied six different entropies; Shannon, Sample, Threshold, Sure, Log Energy and Norm Entropies.

Then, using knowledge data mining algorithms, we studied the possibility of various classifiers predictability of subjects' cerebral reaction in their effort to resolve the educational procedures (identification tasks). In order to evaluate the vectors of indicators (feature vectors) and for knowledge data mining we used the environment of WEKA [41].

After having exported the desirable feature vectors, we evaluated them and exported information about the better prediction model of difficulty of learning operations. We did this procedure using classifiers. A classifier is an algorithm that receives data separated into classes. It is trained to understand, via a dataset, the relation among some features and classes that are assigned to the learning operations, and then, taking some values, it is able to predict in which class can it (the data) be categorized. In fact (in our case) the models constructed from various classifiers, can predict the difficulty that the subject faces, in his trial to solve the learning operation.

The Indexes that we studied in the present work - in order to draw out conclusions via their feature vectors, for their probably correlation with the difficulty of educational-identification problem - are related with entropies of signals.

For the aims of present research were used thirteen classifiers (table 1), which are completely described in the [7].

Table 1. Classification Algorithms.

Classification Aglorithms
BayesNet
AdaBoostM1
BFTree
ClassificationViaRegression
DecisionTable
meta.Decorate
KStar
Logistic
NaiveBayes
NaiveBayesSimple
NBTree
PART
SMO

Results

We received a plethora of feature vectors by executing three different experiments. We used the data from 14 subjects, the 4 tasks and the 6 attributes (entropies) in order to record the cerebral activity. Every feature vector has dimension 1x31, where 31 was the EEG electrodes (channels). We can see 4 feature vectors for Threshold Entropy. Every color is a different math operation. In horizontal axe there are the 31 electrodes, while in vertical axe is Threshold Entropy.

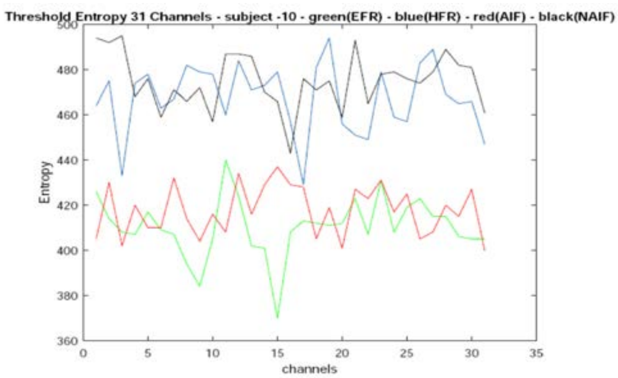

Fig. 1. Threshold Entropy for 31 channels – subject No 10 – green (EFR) – Blue (HFR) – Red (AIF) – Blac (NAIF)

We executed three different experiments:

1st experiment: We used all the available data (from 6 entropies, the 14 subjects and the 4 tasks) without averages. By this way, for classification we got 14 feature vectors (1x31) for each class (4 tasks – math operations) and each attribute (entropy). Summarizing, we totally classified 1083 instances for each math operation.

2nd experiment: At the second experiment we took into account the data from the first experiment with the following difference: We took the mean from the 14 subjects. Thus for classification we got 1 feature vector (1x31) for each class (4 tasks – math operations) and each attribute (entropy). Summarizing, we totally classified 31 instances for each learning operation.

3rd experiment: At the third experiment we took into account the data from the first experiment with the following difference: We took the mean from the 31 channel data. Thus for classification we got 14 feature vectors (1x1) for each class (4 tasks – math operations) and each attribute (entropy). Summarizing, we totally classified 14 instances for each learning operation.

To understand better the total results we can take a look at the confusion matrix of one classification. For example the Naïve Bayes classifier Algorithm confusion matrix for a 2nd type experiment (Fig 2)

```
=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 22  6  2  1 |  a = EFR
  2 28  1  0 |  b = HFR
  0  5 25  1 |  c = AIF
  0  3  1 27 |  d = NAIF
```

Fig. 2. the Naïve Bayes classifier Algorithm confusion matrix

In the confusion matrix of image 1 we can see the following results: for the first math operation (a=EFR) the classification algorithm predicted correctly 22 instances (of 31). Furthermore, it classified incorrectly 6 instances in the second (b=HFR) math operation, 2 in the third (c=AIF) and 1 in the fourth (d=NAIF) and so on.

The corresponding results of each experiment with the prediction of difficulty of educational operation and the equitable classification of corresponding action for each classifier that we used, appear in the three figures below (Fig. 3, 4 and 5):

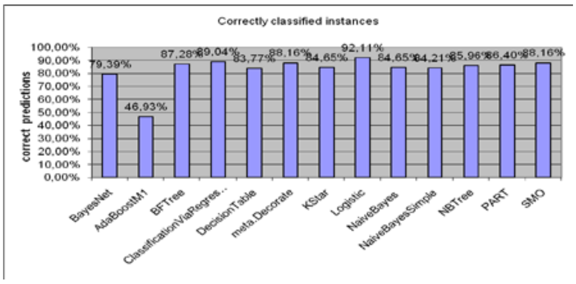

Fig. 3. Classifier Predictability for the 1st experiment

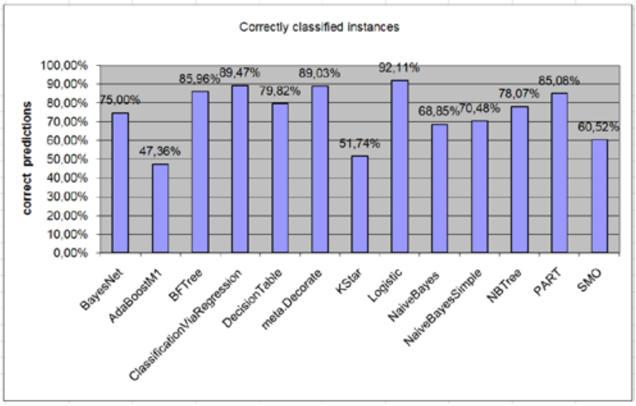

Fig. 4. Classifier Predictability for the 2nd experiment

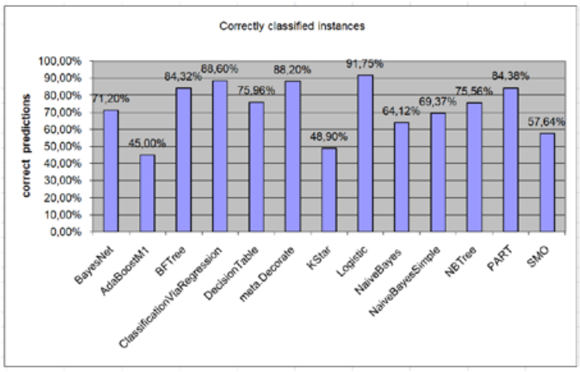

Fig. 5. Classifier Predictability for the 3rd experiment

Discussion

From images 2,3 and 4 we can clearly observe that the classifiers that give the better results are: the Logistic Regression, ClassificationViaRegression and meta.Decorate, getting a good percentage of correctly classified instances (90%). The first and the second are regression algorithms while the third is a meta learning algorithm. Logistic Regression was the best classifier with percentages at about 92% for all the three different experiments.
It is notable that in cases of experiments with means (2nd for subjects means and 3rd for channels means) there is some classifiers (as the kStar from k-neighbourhoods algorithms family and NaiveBayes from Bayesian algorithms family) that lost a big amount of their predictability.
As feature work we think about to examine sets of electrodes separately. For example we would like to study separately the signals from frontal electrodes, or from parietal electrodes, or temporal electrodes, but also combinations of these. For example, how is the variation of the classifiers predictability, when we study only temporal and parietal signals.

Conlusions

We can export very useful conclusions from the previous discussion. Taking into consideration the results from the three experiments of the present research and having the EEG signals during the educational progress we can spot the points where the subject faces difficulties when doing several educational actions. We obtained very good predictions from algorithms Logistic Regression, ClassificationViaRegression and meta.Decorate. We can conclude that, the regression Algorithms give us very good predictability.

As future work we would like to study different nature educational problems and find other algorithms that extract the best results. By this way, we would like to create models with parameters that will change their value depending on the educational problem and will provide the best model each time which will give the best prediction. Also we would like to research other neurophysiologic indexes, like ERPs', Power Spectra, ITCs', EPRSPs', etc. with the aim of extracting the best, each time, algorithms for the prediction of the difficulties that the subject faces when doing an educational procedure.

References

[1] Michael L. Metzker ," *Sequencing technologies — the next generation*" Nature Reviews Genetics 11, 31-46 ,January 2010.

[2] Jacques Cohen, "*Bioinformatics—an introduction for computer scientists*", ACM Computing Surveys, Volume 36 Issue 2,122-158,June 2004

[3] Bein B., «Entropy», *Best practice & Researche Clinical Anaesthesiology* Vol.20, No 1, pp. 101-109, ELSEVIER 2006.

[4] B.J. Stapley et al," *Biobibliometrics: Information retrieval and visualization from co-occurrences of genenames in Medline abstracts*", Pacific Symposium on Biocomputing 5:526-537 (2000)

[5] Martzoukos Y., Papavlasopoulos S., Poulos M., Syrrou M., "*Biobibliometrics & gene connections",*IISA2015

[6] Poulos M. and Stefanidis V. , *"Synchronization of Small Set Data on Stable Period"*, 2nd International Conference Mathematics and Computers in Science and Industry, Malta, 2015.

[7] Baars, B. J., Gage, N. M., (2010), *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience,* Elsevier, 2nd Edition.

[8] Hodgkinson, Colin A., Mary-Anne Enoch, Vibhuti Srivastava, Justine S. Cummins-Oman, Cherisse Ferrier, Polina Iarikova, Sriram Sankararaman, et al. *"Genome-Wide Association Identifies Candidate Metrics That Influence the Human Electroencephalogram."* Proceedings of the National Academy of Sciences 107, no. 19 (2010): 8695–8700.

[9] Dumermuth, G., Molinary, L., (1987), *Spectral Analysis of EEG Background Activity, In Handbook Methods of Analysis of Brain Electrical and Magnetic Signals*, A. S. Gevins (ed.) (Amsterdam: Elsevier).

[10] Sommer BJ, Barycki JJ, Simpson MA. *Characterization of human UDP-glucose dehydrogenase*. J Biol Chem 2004;279:23590.

[11] Marcel, S., Milln, J. del R., (2007), *Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. Pattern Analysis and Machine Intelligence,* IEEE Transactions on 29, 743 - 752.

[12] Dougherty MK, Morrison DK. *Unlocking the code* of 14-3-3. J Cell Sci 2004;117:1875–84.

[13] Hirsch L. J., Richard P., Brenner R.P., (2010), Atlas of EEG in Critical Care, Wiley-Blackwell..

[14] Percival, D. B., Walden, A. T., (1993), *Spectral Analysis for Physical Applications*

Multitaper and Conventional Univariate Techniques, Cambridge, England: Cambridge University Press

[15] Henry Small," *Co-citation in the scientific literature: A new measure of the relationship between two documents*", Volume 24, Issue 4, pages 265–269, July/August 1973.

[16] Fang Li et al.," *Applying Association Rule Analysis in Bibliometric Analysis-A Case Study in Data Mining"* , Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT '09),Huangshan, P. R. China, 26-28,Dec. 2009, pp. 431-434.

[17] Gerald Matthews, Lauren Reinerman-Jones, Julian Abich IV, Almira Kustubayeva, *"Metrics for individual differences in EEG response to cognitive workload: Optimizing performance prediction",* Personality and Individual Differences, Volume 118, 1 November 2017, Pages 22-28

[18] Anderson S. Oliveira,Bryan R. Schlink,W. David Hairston, Peter König, and Daniel P. Ferris, "*Proposing Metrics for Benchmarking Novel EEG Technologies Towards Real-World Measurements*", frontiers in Human Neuroscience

[19] Stefanidis V., Papavlasopoulos S., Poulos M., «*Learning Difficulties Prediction using Multichannel Brain Evoked Potential Data*», International Journal of Computers, Vol. 3, pp. 177-181, October, 2018.

[20] Agatha Lenartowicz, and Sandra K. Loo, "*Use of EEG to Diagnose ADHD*", Curr Psychiatry Repot 2014 Nov; 16(11): 498.

[21] Jasper HH, Solomon P, Bradley C. Electroencephalographic analyses of behavior problem children. *American Journal of Psychiatry*. 1938;95(3):641–658.

[22] «*ENTROPY HANDBOOK Defnitions, Theorems, M-files*» www.uivt.cas.cz/~pklan/eth.pdf

[23] Stefanidis V., Anogianakis G., Evangelou A., Poulos M., *«Bibliometrics associations between EEG entropies and connections between Learning Disabilities and the Human Brain Activity*», IEEE Proceedings, MCSI, pp. 268-272, 2016.

[24] Daube J. R, Rubin D.I, «Clinical Neurophysiology», Oxford University Press, Third Edition, 2009

[25] Bein B., «Entropy», Best practice & Researche Clinical Anaesthesiology Vol.20, No 1, pp. 101-109, ELSEVIER 2006.

[26] Lopez da Silva F. H., Storm Van Leeuwen W., & Remond A., *«Handbook of electroencephalography and clinical neurophysiology: Vol. 2.»,* Elsevier, 1990.

[27] V. Stefanidis, S. Papavlasopoulos, M. Poulos, *Bibliometrics associations between EEG entropies and connections between Learning Disabilities and the Human Brain Activity*, International Journal of Computers, Vol. 3, pp. 177-181, October, 2018, ISSN: 2367-8895

[28] V. Stefanidis, G. Anogiannakis, A. Evangelou, M. Poulos, «*Learning Difficulties Prediction*», International Journal of Biomedical Engineering and Technology, Vol. 21, No2, pp. 176-189, 2016.

[29] V. Stefanidis, G. Anogiannakis, A. Evangelou, M. Poulos, «*Stable EEG Features»,* Optimization, Control, and Applications in the Information Age, Springer Proceedings in Mathematics & Statistics, vol. 130 , pp. 349-357, 2015.

[30] V. Stefanidis, S. Papavlasopoulos, M. Poulos, Bibliometrics *associations between EEG entropies and connections between Learning Disabilities and the Human Brain Activity*, 5th International Conference on ENGINEERING and TECHNOLOGY EDUCATION (ETE '18), London, UK, October 26-28, 2018

[31] V. Stefanidis, M. Poulos, S. Papavlasopoulos, *Bibliometrics EEG metrics associations and connections between Learning Disabilities and the Human Brain Activity*, 12th Joint Conference on Knowledge-Based Software Engineering (JCKBSE 2018), Corfu, Greece, August 27-30, 2018

[32] Poulos M., Stefanidis V., Anogianakis G., Evangelou A. *«Synchronization of Small Set Data on Stable Period»,* IEEE Proceedings, Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI), pp. 263-267, 2016.

[33] Stefanidis V., Anogianakis G., Evangelou A., Poulos M., *«Learning Difficulties Prediction using Multichannel Brain Evoked Potential Data»,* IEEE Proceedings, Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI), pp. 268-272, 2016.

[34] Poulos M. and Stefanidis V. , *« Synchronization of Small Set Data on Stable Period»,* 2nd International Conference Mathematics and Computers in Science and Industry, Malta, 2015.

[35] Stefanidis V. and Poulos M., *«Learning Difficulties Prediction»,* 2nd International

Conference Mathematics and Computers in Science and Industry, Malta, 2015.

[36] G Bokos, NKS Papavlasopoulos, M Avlonitis, *Specific selection of FFT amplitudes from audio sports and news broadcasting for classification purposes*, (2007)

[37] Sozon H. Papavlasopoulos, Marios S. Poulos, George D. Bokos *Classification Control for Discrimination between Interictal Epileptic and Non-Epileptic Pathological EEG Events*, International Journal of Biomedical Sciences 1 (1), 34-41, 2006

[38] S Papavlasopoulos, M Poulos, A Evangelou, *Feature extraction from interictal epileptic and non-epileptic pathological EEG events for diagnostic purposes using LVQ1 neural network*, Mathematical Methods In Scattering Theory And Biomedical Engineering, 390-398,2006

[39] M Poulos, S Papavlasopoulos, N Alexandris, E Vlachos, *Comparison between auto-cross-correlation coefficients and coherence methods applied to the EEG for diagnostic purposes*, Medical Science Monitor 10 (10),

[40] EEGLAB Matlab Toolbox http://sccn.ucsd.edu/eeglab/

[41] WEKA (Data mining environment) http://www.cs.waikato.ac.nz/ml/weka/