LONG-TERM EEG DATA ANALYSIS

J. Rieger*, L. Lhotska*,V. Krajca** and M. Matousek***

* Czech Technical University/Gerstner Laboratory, Technická 2, Prague, Czech Republic
** Faculty Hospital Na Bulovce/Department of Neurology, Budínova 2, Prague, Czech Republic
*** Prague Psychiatric Center/Center for Neuropsychiatric Studies, Prague, Czech Republic

riegerj@fel.cvut.cz

Abstract: This paper presents the results of experiments with applying methods of quantitative EEG to coma (sleep) long-term EEG records. The training set for this problem domain was created in cooperation with expert and tested on real EEG data. Final classification accuracy is about 80%. The results of classification are presented using comprehensive graphical representation and allow to quickly estimate longterm trend changes in EEG records at first glance. One of the problems that are connected with the evaluation of EEG signals is that it necessitates visual checking of such a recording performed by a physician. In case the physician has to check and evaluate long-term EEG recordings computer-aided data analysis and visualization might be of great help. Software tools for visualization of EEG data and data analysis are presented in the paper.

Introduction

EEG is spontaneous cortical electrical activity recorded from the scalp. Among various methods of brain function monitoring, EEG has a lot of advantages – it is non-invasive, generally available at doctor's surgeries, it is cheap, and has high enough spatial and temporal resolution. In a few last decades there was a fast growth of EEG data amount (daylong records are not exceptions anymore), so it is no longer suitable to analyse records without computer support. Application of quantitative EEG methods is of course more usable for long-term EEG records and in future could help to eliminate subjective influence of expert's opinion. This paper deals with implementation of basic signal processing and pattern recognition methods (adaptive segmentation, feature extraction, classification and visualization). Resulting system is used for real long-term coma (sleep) EEG record analysis.

The Problem Domain – Coma

Coma is a state of brain function. It can be very roughly compared to sleep. However, an individual cannot awaken purposefully from coma, using either internal or external stimulus. Comatose state may have a number of causes, starting from head injury at a serious accident, over cerebral vascular diseases, infectious diseases, brain tumours, metabolic disorders (failure of liver or kidney), hypoglycemia, to drug overdosing, degenerative diseases, and many more. A patient in a state of coma does not manifest any notion of higher consciousness, does not communicate and frequently functions of his/her inner organs are supported by devices. There has been great effort devoted to scaling comatose states into different levels according to seriousness, depth, and to prediction of probable development of patient state. First trial to unify coma classification was the Glasgow classification of unconsciousness (Glasgow Coma Scale - GCS) described in 1974 [1]. GCS has become widely used and reliable scale for classification of coma depth. It is highly reproducible and fast and it is a suitable tool for long-term monitoring of patient coma depth. During next decades, further systems of coma classification have been developed, for example Rancho Los Amigos Scale, Reaction Level Scale RLS85 [2] both classifying into 8 levels, Innsbruck Coma Scale, Japan Coma Scale, etc. Individual systems of coma classification differ in number of levels, way of examination, precision, etc. Experiments with long-term coma (sleep) EEG records showed, that application of quantitative EEG in this domain is possible (coma depth assessment [3], [4], outcome prediction [5]).

Segmentation and Feature Extraction of EEG Signal

In general, EEG signals can be analysed both in temporal and frequency domains. The basic characteristics in temporal domain are the signal variance (square of standard deviation) and mean absolute first derivation. Signal variance is the measure of diversity of samples and their mutual distance, mean absolute first derivation serves for quick estimation of basic frequency characteristics of the signal.

In frequency domain, one of the basic transforms used for signal description is the Fourier transform. We acquire as results both amplitude and phase spectrum. Then it is possible to calculate further signal characteristics, for example correlation function and power spectra. Correlation contains information on relation of signal value $x(t)$ for time $t = t_0$ and value

ISSN: 1727-1983 © 2005 IFMBE

of this signal $x(t)$ (or another signal $y(t)$) for time $t_1 = t_0 + \tau$. Correlation represents optimal way to detect a known waveform in a signal. Autocorrelation is used for a signal correlated with itself. However we must consider the fact that the Fourier transform is not suitable if the signal has time varying frequency, i.e., the signal is non-stationary. In such a case the signal must be divided to stationary segments. There exist several approaches to adaptive segmentation [6], [7] which divide signals to stationary segments. In principle, these methods are based on autocorrelation function (ACF) [6], spectral error measure [6], and generalized likelihood ratio (GLR) [8].

Adaptive segmentation described in [7] was used. Principle of algorithm can be divided into several points:

1) there are two joined windows of the same length, sliding together over the whole signal (Figure 1)
2) same statistical characteristics are calculated in both windows. Selection of those characteristics should be motivated by nature of the signal. In our case, we have used mean absolute amplitude A_W

$$A_W = \frac{1}{WL}\sum_{i=1}^{WL}\left|x_i\right| \qquad (1)$$

and mean absolute first derivation F_W

$$F_W = \frac{1}{WL}\sum_{i=1}^{WL}\left|x_i - x_{i-1}\right| \qquad (2)$$

where WL is the length of the window in samples and x_i is the value of the EEG signal at index i

3) weighted difference G of values A_W and F_W corresponds to similarity of both joined windows

$$G = k_A\left|A_{W1} - A_{W2}\right| + k_F\left|F_{W1} - F_{W2}\right| \qquad (3)$$

where k_A is weight of mean absolute amplitude difference, k_F weight of mean absolute first derivation difference, indexes 1 and 2 correspond to connected windows

4) segment borders are detected as local maximum of G. To avoid segmentation corresponding to small fluctuations of G there is also a threshold T as a minimal possible value of G for a segment border

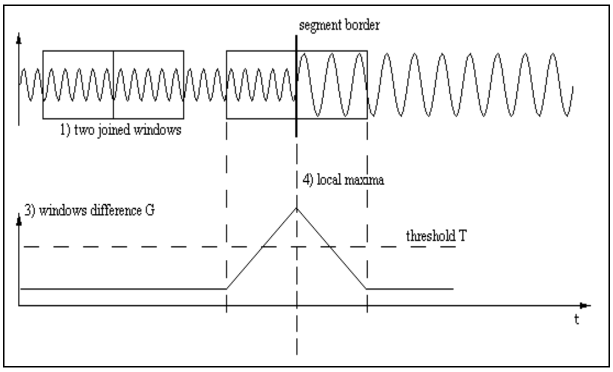

Figure 1: Adaptive segmentation

After segmentation, signal features that will be used for classification have to be selected. For more comprehensive description of the EEG signal, set of features including not only features from frequency domain, but also features characterising signal shape were selected. The following features describing signal are used: average AC amplitude in the segment, variance of AC amplitude, maximum positive and minimum negative values of amplitude, mean value of first derivation, maximum value of the first derivation, maximum value of the second derivation, spectral power values in defined frequency bands (e.g. for EEG in delta, theta, alpha and beta bands). All feature values are normalized. These values serve as input data for cluster analysis module.

Segmented EEG signal is successively classified using feature-based pattern recognition methods. Features describing the object can be arranged into a n-dimensional vector that is called feature vector. Segments are then represented as points in n-dimensional space. The classifier maps the feature space into a set of class indicators. In our study, k-NN (k=5) classifier and RBF neural network have been used for classification.

Development of the Training Set

The core of the developed system is the training set on which practically depends the quality of classification. In the implemented system we have developed simple interface for manual creation of training set during the program execution. However, in the case of classification of comatose EEG we have found out that the classification result is not dependent too much on the type of the classifier but that the most critical point is the training set. Therefore a lot of effort has been paid to its development. It has been developed in a non-traditional way using expert background knowledge. Individual fragments have been acquired from EEG signal. The basic steps in the development of the training set can be described as following:

1) We have saved in total 453 eight-second periods of 18-electrode sleep EEG where the classification into levels 1 thru 10 (provided by professor Milos Matousek, MD) has been known.

2) Since the created training set has shown inacceptable cross-validation error it has been necessary to edit the training set to become acceptable.
3) The segments unsuitable for further processing, for example those containing artefacts, have been excluded from the training set. The number of segments has decreased to 436.
4) The kernel of the training set has been generated by cluster analysis - only such segments have been included for that the classification by cluster analysis has agreed with original classification of professor Milos Matousek. At repeated clustering, it has been searched for such a metrics of the feature space that results in correspondence in classification at the highest number of segments. The core of the training set generated in this way contains 184 segments.
5) Using auxiliary scripts in Matlab realizing classification by nearest neighbour and parallel visual control of results some of the segments excluded in the previous step have been added to the training set, however frequently their classification has been changed by 1 level. The resulting training set has had 349 segments.
6) Using RBF implementation of a neural network the cross-validation error has been computed. Data has been randomly divided in 1:1 ratio into training and testing sets. RBF network has been learned on the training set and the error has been computed using the testing data. This procedure has been repeated many times (in the order of hundreds) for different random distributions training/testing set. The resulting error has been computed as an average error of these distributions. Repeatedly incorrectly classified segments in the second phase of computation have been excluded from the resulting training set.

The resulting training set created in previous steps contains 319 segments classified into levels 1 thru 10 of coma. Average cross-validation error computed using RBF neural network does not exceed the value of 3 percent. Examples of the segments for classes 1, 4, 7 and 10 are on Figure 2.

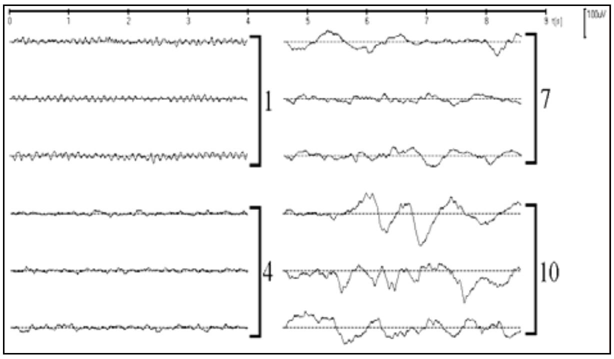

Figure 2: Examples of segments of training set classes 1, 4, 7 and 10

As mentioned in step 4), during cluster analysis optimal feature set and feature space metric have been found. From large set of statistical features (amplitude, derivative, spectral), mean absolute amplitude, deviation, power in delta band (1-4Hz), theta band (4-8Hz), alpha band (8-13Hz) have been selected. Feature weights have been found as well.

Experiment

The approach has been tested on real sleep EEG recording for which the classification has been known (manually classified by expert - prof. Milos Matousek, MD). It is necessary to stress that the comatose EEG is similar to sleep EEG; the main difference is that sleep EEG has much faster trend changes. For segmentation, combination of non-adaptive and adaptive segmentation has been used. Length of the recording has been 2 hours. Length of segments for non-adaptive segmentation has been set up to 32 seconds (at sampling rate of 256 Hz 8192 samples correspond to 32 seconds, this value has been selected with respect to successive computation of FFT). Intervals containing artefacts have been determined by adaptive segmentation and successively excluded from the classification (Figure 3, third segments in electrodes F8T4 and T4T6). As the classifier, k-NN classifier (k = 5) was used. Total number of classified segments has been 204 (only 1 electrode used).

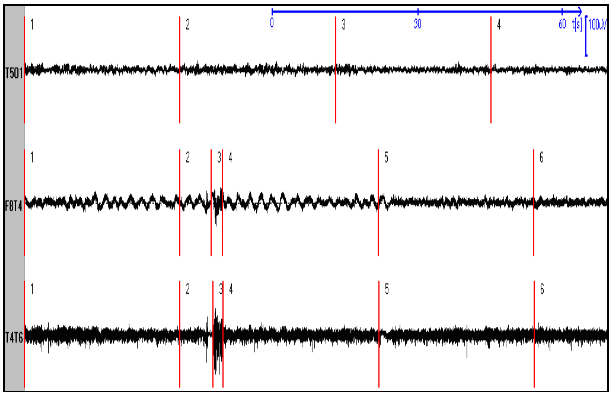

Figure 3: Combination of adaptive and non-adaptive segmentation

Results

Table 1 shows the confusion matrix for the first 7 degrees of coma (sleep). Rows represents classification by expert, columns classification by our system. Number of correctly classified instances was 163 (79,9%), number of incorrectly classified instances was 41 (20,1%). Accuracy of classification about 80% was achieved also in other tested records.

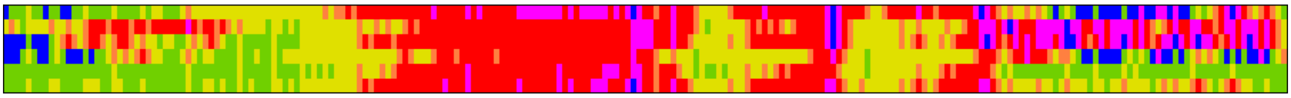

Figure 4: Classification results visualization by color coding. Each of the training set classes has assigned color

Table 1: Confusion matrix

27	6	1	2	0	0	0
1	19	1	2	0	0	1
0	1	14	0	0	0	0
0	1	2	36	2	2	0
0	1	2	1	12	2	0
0	0	2	1	0	41	1
0	1	0	0	2	6	14

For more coarse (but faster) classification result evaluation we use simple visualization [9], when each class of training set has assigned color and segment classified into this class is plotted using that color (Figure 4). Instead of colouring just EEG waveform, we use below showed color bar, when X axis corresponds to time (width of whole bar is 2 hours) and Y axis corresponds to particular electrode (6 electrodes showed). In our example, we use color coding by 7 basic colors of spectrum (class 1 – violet, 2 – blue, 3 – green, 4 – yellow, 5 – orange, 6 – red, 7 – violet, 8 – blue, 9 – green, 10 – red). Then, at first glance we can very coarselly estimate for example this: first hour of record contains EEG signal with slowly growth from degree 3 (green on the left) to degree 7 (violet in the middle), then there is a fast fall to degree 4 (yellow) and so on. Last 30 minutes of record contains a lot of artifacts and classification makes no sense.

Implementation

The developed system consists of several basic modules. Each of the modules enables visualization of data or information it is working with (see Figure 5). The main window serves for visualization of the EEG signal and enables access to all functions of the program. It is possible to set up way of visualization and processing of individual electrodes as needed. For example, it is possible to exclude from visualization (and computation) electrodes that are disconnected, having no signal, or inadequate (ECG, EOG), etc. Scale of the signal of individual electrodes as well as joint scale for all electrodes and several other properties for displaying (grid, time stamps, numbering of segments, isoline, etc.) can be set up. The setup can be saved into a file for repetitive use. The program enables relatively detailed setup of segmentation. Most of the control elements serve for adaptive segmentation. The setup can be saved into a file for repetitive use as well.

The core of the system is the training set used for classification. The user has several options how to create a training set. Therefore the system is flexible and can be used for solving different tasks. The options are the following ones:

- reading the traning set from a file
- generation of the training set by cluster analysis
- generation of the training set manually by moving segments from the main window to corresponding classes of the training set

The user can define required number of classes, add and delete classes, set up their colouring, etc. For individual segments of the training set, it is possible to change their scale, sampling frequency, to move them between the classes, delete, or deactivate (segments are not used for classification).

The system is completed with visualization of classification results (see Figure 4), spectral analysis (see Figure 6), and 3D mapping of the brain activity in time (see Figure 5). The mapping algorithm is based on the existence of 3D model consisting of a set of contour points. The contour points are mutually connected by a system of general n-angles forming an area. The way of interconnection is not important because we are working only with positions of contour points and positions of applied electrodes. Positions of both electrodes and points are normalized in 3D space. The applied algorithm is designed with respect to the real environment where the computation of relatively complex spline curves at each image change could be extremely slow and thus practically unusable. For given placement of electrodes, the values of model coefficients are constant and the computation can be divided into two steps. In the first step (time-consuming), the coefficients of individual electrodes and individual model points are pre-computed. In the next step, the EEG is mapped relatively quickly using the pre-computed coefficients. While the first step is performed only once, the second step is performed for each change of EEG activity. In the application, the most frequent International 10-20 system of electrode placement is considered.

Conclusions

As mentioned above, the EEG signal is more complex, and thus it requires more steps of pre-processing. The first step is the adaptive segmentation that divides the EEG signal into stationary segments. When the signals are segmented, all attributes are calculated, some of them in time domain, some of them in frequency domain. The extracted values are used as input values of a classification system

In cooperation with an expert training set of 10 classes has been developed. During experiment the training set has been used for classification of real EEG record with nearest neighbour classifier.

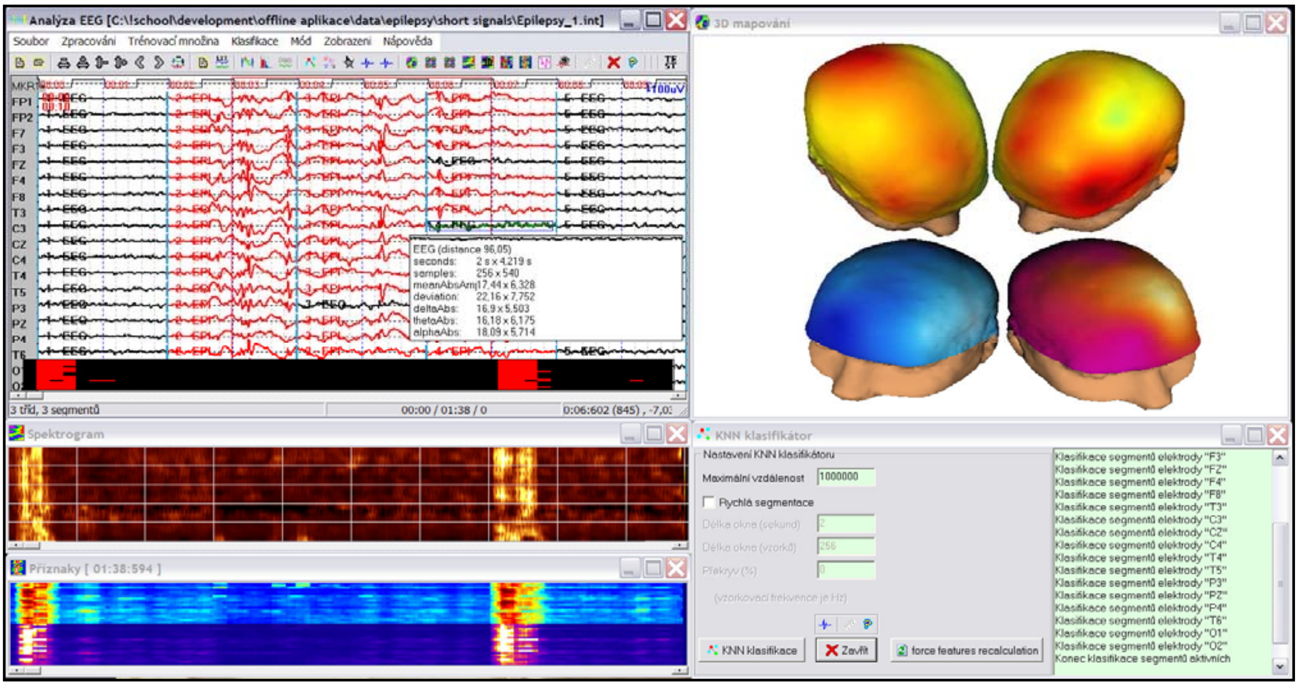

Figure 5: Developed application for longterm EEG record processing. System is designed as modular, allowing to use several signal processing techniques (classification, visualization etc.) at the same time

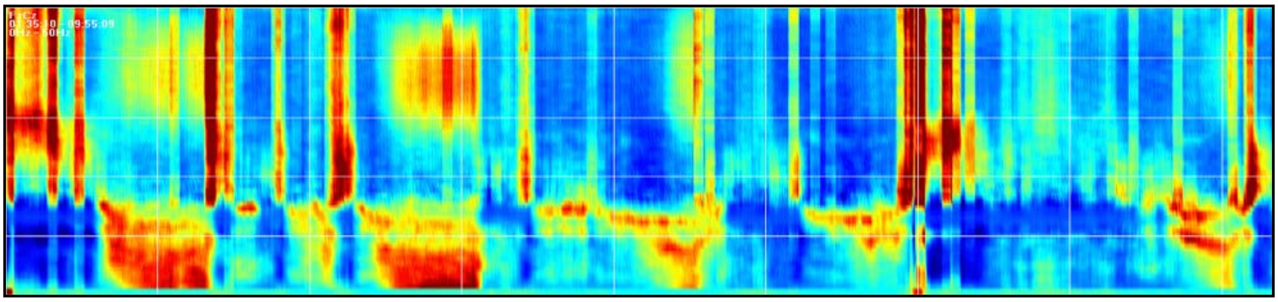

Figure 6: Spectral analysis of long-term sleep EEG. X-axis represents time, Y-axis represents frequency. The highest intensity is represented by red colour; blue colour represents the lowest intensity

Final classification accuracy has been approximately 80%. Using visualization by colour coding of individual classes, we can very quickly obtain a basic knowledge about longterm trend changes in the whole signal.

Higher accuracy of classification could be achieved at first by further editing and enlargement of training set as a critical component of the whole classification system.

Data visualization tools aim at reducing the information overload by intelligent abstraction and visualization of the features of interest in the current situation. This requires context-aware use of knowledge for proper reduction of the complexity of the data displayed without loosing those parts of information, which are essential and critical in the current situation. Newly developed software tools for visualization should support fast comprehension of complex, large, and dynamically growing datasets in all fields of medicine. These tools should lead to a new generation of software tools, which allow health professionals to cope with the increasing amount of detailed patient data.

In this paper, we have described briefly our implementation and by that illustrated possibilities of visualization of the whole process of data processing and evaluation. The example used is from a very complex domain of EEG. The implemented system is simple to use, guiding the user, retrieving all stored data and setups, and executing classification in a minimum number of steps. It presents most of the information in graphical form that is most suitable for perception. The main advantage of the system is that for routine tasks all the setups are defined only once and saved to files. Then the user who may not necessarily know all the subtle details of proper setup starts the corresponding application and evaluates the results. Since it is possible to display classification

ISSN: 1727-1983 © 2005 IFMBE

results of two-hour EEG recording in compressed form on one screen the preliminary visual inspection is very fast. In the next step the user can focus on those segments that indicate serious problems.

Acknowledgements

This work has been supported by the research program „Information Society“ under Grant No. 1ET101210512 „Intelligent methods for evaluation of long-term EEG recordings“.

References

[1] TEASDALE G., JENNET B. (1974): ‘Assessment of coma and impaired consciousness. A practical scale.‘, *The Lancet*, July 1974, ii, pp. 81-84

[2] STARMARK J. E., STÅLHAMMAR D., HOLMGREN E. (1998): ‘The Reaction Level Scale (RLS85). Manual and Guidelines.‘, *Acta Neurochirurgica*, 91, pp. 12-20

[3] THATCHER R. W. et al. (1989): ‘EEG discriminant analyses of mild head trauma‘, *Electroencephalography and Clinical Neurophysiolology*, 73, pp. 94-106

[4] MATOUSEK M. et al. (1996): ‘Quantitative EEG analysis as a suplement to the clinical coma scale RLS85‘, *Acta Anaesthesiologica Scandinavica*, 40, pp. 825-831

[5] KANE N. W. et al. (1998): ‘Quantitative electroencephalographic evaluation of non-fatal and fatal traumatic coma‘, *Electroencephalography and Clinical Neurophysiolology*, 106, pp. 244-250

[6] BODENSTEIN G., PRAETORIUS H. M. (1997): ‘Feature extraction from the electroencephalogram by adaptive segmentation’, Proc. IEEE, 65, pp. 642-652

[7] KRAJCA V. et al. (1992): ‘Multichannel adaptive segmentation – basis for EEG record pre-processing‘, *Lekar a technika*, 23, pp. 28-34

[8] APPEL U., BRANDT A. V. (1983): ‘Adaptive sequential segmentation of piecewise stationary time series’, *Information Science*, 29, pp. 27-56

[9] AGARWAL R. et al. (1998): ‘Automatic EEG analysis during long-term monitoring in the ICU’, *Electroencephalography and Clinical Neurophysiolology*, 107, pp. 44-58

[10] RIEGER J. (2004): ‘Processing of Long-Term EEG Recordings’, Master thesis, CTU Prague (in Czech)