

Automated real-time classification of functional states: significance of individual tuning stage

Vladimir V. Galatenko, Evgeniy D. Livshitz, Alexander M. Chernorizov, Yury P. Zinchenko, Alexey V. Galatenko, Vladimir M. Staroverov, Sergey A. Isaychev, Vyacheslav V. Lebedev, Galina Ya. Menshikova, Alexey N. Gusev, Ekaterina M. Lobacheva, Rozaliya F. Gabidullina, Vladimir E. Podol'skii, Victor A. Sadovnichy

Lomonosov Moscow State University, Moscow, Russia

Automated classification of a human functional state is an important problem, with applications including stress resistance evaluation, supervision over operators of critical infrastructure, teaching and phobia therapy. Such classification is particularly efficient in systems for teaching and phobia therapy that include a virtual reality module, and provide the capability for dynamic adjustment of task complexity.

In this paper, a method for automated real-time binary classification of human functional states (calm wakefulness vs. stress) based on discrete wavelet transform of EEG data is considered. It is shown that an individual tuning stage of the classification algorithm — a stage that allows the involvement of certain information on individual peculiarities in the classification, using very short individual learning samples, significantly increases classification reliability. The experimental study that proved this assertion was based on a specialized scenario in which individuals solved the task of detecting objects with given properties in a dynamic set of flying objects.

Key words: human functional state, EEG data, automated classification, individual tuning, stress.

1. Introduction

Classification of the human functional state of an individual (Danilova, 1992; Hockey et al., 2003) is an important problem in modern psychology. It is a key subject of the EU FP integrated project “Advanced sensor development for attention, stress, vigilance and sleep/wakefulness monitoring (SENSATION)” (Maglavera et al., 2006), The Research and Technology Organization (RTO of NATO) program “Operator Functional State Assessment” (Hockey et al., 2003; Veltman et al., 2004), and a number of other projects. Applications of the automated classification of functional states include stress resistance evaluation, supervision over operators of critical infrastructure (air traffic controller, power plant operator, engine-driver,

etc.), teaching, and phobia therapy. Such classification is particularly efficient in systems for teaching and/or phobia therapy that include a virtual reality module (e.g., virtual cave) and provide the capability for dynamic adjustment of task complexity: in this case the implementation of feedback based on real-time automated functional state classification can be used for the automated optimal selection of task complexity, and such a selection is a key factor for successful teaching and therapy.

Standard mathematical approaches to the automated classification problem are based on the detection of characteristic patterns of classes, and on the construction of classifier functions that give close values for elements of the same class, and distinguishable values for elements from different classes. If characteristic patterns and classifier functions cannot be revealed using solely specific expert knowledge of the domain, automated and semi-automated machine-learning methods are used: Artificial Neural Networks (Wasserman, 1989), Support Vector Machine (Cortes and Vapnik, 1995), Logistic Regression (Hosmer and Lemeshow, 2000), Random Forest (Breiman, 2001), etc.

However, these methods can be efficiently applied only if the size of learning samples is considerably larger than the dimension of the feature (measurement) arrays used for the classification. In the case of functional state classification based on electroencephalographic (EEG) data, this limitation is violated: the typical size of learning samples here does not exceed a few hundred (in most cases — several dozen), and the number of measurements associated even with short (e.g., one second) time windows exceeds one thousand (and generally exceeds ten thousand). Recently we developed a specialized method for the automated construction of classifier functions for binary functional state classification, which provided reliable classification of stress vs. state of calm wakefulness (also referred to as a normal functional state) despite the relatively small size of learning samples (Galatenko et al., 2012). Along with the efficient procedure for dimension reduction, this method included an individual tuning stage that allowed the involvement of certain information on individual peculiarities in the classification. This stage is non-standard for conventional machine-learning methods. In this paper we show that the individual tuning stage significantly increases the reliability of the classification.

The rest of the paper is organized as follows. In the section “Algorithm scheme” a short description of the algorithm for automated real-time classification of a functional state based on EEG data is given. A detailed description of the algorithm was given (Galatenko et al., 2012). In the section “Experiment design”, an outline of an experiment for evaluating the importance of the individual tuning stage for classification reliability is given. In the section “Results”, experiment results are discussed.

2. Algorithm scheme

2.1 Dimension reduction

Measurements of EEG signals are performed independently by N sensors with frequency F Hz, where a typical value of N is several dozen (up to 256) and the typical value of F is several hundred (up to 5000). Hence, $N \cdot F \cdot L$ measurements

are performed in each L -second time window. However, strong dependencies are present between signals corresponding to closely located sensors, and between data corresponding to close measurements. EEG data contains noise and artifacts that are generally localized in high- and low-frequency domains, or have a good localization in time domain (e.g., artifacts induced by winking).

The procedure for dimension reduction and simultaneous artifact filtration can be described as follows. Measurements associated with different sensors are processed independently. For a given time window (the typical length of a time window here is 30–90 seconds), a signal corresponding to one sensor is decomposed using discrete CDF 9/7 wavelet transform (Daubechies, 1992: Ch. 5, Sec. 5.1; Cohen et al., 1992). Low- and high-frequency terms are discarded (this leads to noise filtration in frequency domain), and for the remaining fragment of the decomposition, the absolute and relative energy values of a signal for each scale component (i.e., for the fixed frequency band) are computed. The typical number K of scale components after the decomposition truncation is 8, so $2K$ energy values are identified (K absolute and K relative values). Median averaging is used in energy computation for the reduction of errors associated with time-localized artifacts.

An array of energy values is a result of dimension reduction for a given sensor and a given time window. Such reduction decreases data array length by a factor of 150–25 000, but keeps the information required for functional state classification.

2.2 Global learning stage

Even after dimension reduction, the length of data arrays associated with each time window is large (its typical value is 1024). At the same time, most of the entries in these arrays carry no information that is essential for functional state classification, and therefore can be excluded from further analysis and discarded. If a learning sample is available, entries that can be efficiently used for binary functional state classification can be detected by the following fully automated procedure, without using any prior expert knowledge of the psychological domain. For a fixed individual, available data corresponding to each functional state is split into non-overlapping time windows with a standard length (e.g. one minute). All data array entries (i.e. energy values) are computed for each time window. A data array entry is called informative for an individual if its values for one functional state turn out to be always higher (or always lower) than for the other functional state. A data array entry is called informative (with respect to a learning sample) if it is informative for a given fraction (e.g., $2/3$) of individuals in this learning sample, with the same inequality sign between values associated with different functional states. Only informative entries are considered in the further analysis. For each informative entry f , its weight (in other words, its significance) $w(f)$ is computed based on the fraction of individuals for which this entry is informative and on the size of the gap between values associated with different functional states for these individuals.

2.3 Individual tuning

If a sufficiently large learning sample is available for an individual, the global learning stage can be avoided and machine learning focusing directly on this individual can be performed. However, in most applications, including teaching and phobia

therapy, learning samples that can be provided for individuals that were not involved in the global learning stage are very small: generally they contain only one fragment of a standard length (e.g., one minute) associated with one functional state (e.g., stress), and a similar fragment associated with another functional state (e.g., state of calm wakefulness). These samples are not sufficient for machine learning, but they can still be efficiently used to increase the classification reliability for an individual. It can be done in the following way. All informative data array entries are computed. Entries that turn out to be higher for one state (while they were lower for this state for most individuals in the learning sample group) are excluded from further analysis for the individual. Other entries are called applicable to the individual, and used in the functional state classification. For entries that are applicable to an individual, the values of these entries for both functional states are recorded as typical values for each state for this individual.

2.4 Individual testing

For a given individual, functional state classification is performed using a voting scheme (Littlestone, 1994). A time window of standard length is taken, all informative entries applicable to the individual are computed, and a vote $V(f)$ is associated with each entry f :

$$V(f) = \max \{-1, \min\{1, v(f)\}\},$$

where

$$v(f) = s(f) \frac{(\log_2 h - (\log_2 h^s + \log_2 h^c) / 2)}{|\log_2 h^s - \log_2 h^c| + 0.05},$$

$s(f)$ equals 1 if for a learning sample, values of f associated with the first functional state (e.g., stress) are generally higher than values of f associated with the second functional state (e.g., state of calm wakefulness), and equals -1 otherwise; h is the value of the entry for the analyzed time window, h^s and h^c are the typical values of this entry for this individual in the first functional state and in the second functional state respectively, 0.05 is a regularization parameter.

The weighted sum

$$S = \sum_f w(f) V(f)$$

of votes for all informative entries applicable to the individual is calculated. If S turns out to be negative, the classification result for the individual and the analyzed time window is set to the second functional state (e.g., calm wakefulness), otherwise it is set to the first functional state (e.g., stress).

3. Experiment design

The study was performed at Moscow State University and was approved by the Ethics Committee of the MSU Faculty of Psychology. The goal of the study was to determine whether the individual tuning stage actually increases the reliability of the classification.

A special scenario for the virtual cave implemented by staff of the MSU Faculty of Psychology was used. In this scenario, individuals solved the task of detecting objects with given properties (shape and color) in the dynamic set of flying objects.

The difficulty of the task was varied by changing time limits for the detection, frequency of object generation, fraction of objects with required properties in the set of all objects, fraction of objects similar to the objects of interest in the set of all objects, etc. The scenario also included disturbing elements such as sudden irritating sounds and punishments (e.g., the threat of mild electric shock) in case of errors (false alarms and misses).

EEG data was measured using BrainAmp equipment, including BrainAmp Standard amplifiers and BrainCap with 64 sensors and frequency 2500 Hz. The segmentation of data and identification of functional states corresponding to segments (stress / calm wakefulness / other) was performed manually by experts of the MSU Faculty of Psychology. Only segments corresponding to stress and calm wakefulness were used for automated classification, and for classification reliability evaluation.

A learning sample at the global learning stage included data corresponding to 19 individuals of both genders (m:f 9:10) aged between 18 and 25, with different levels of stress resistance. For each individual, the learning sample included six non-overlapping 1-minute fragments associated with stress and six non-overlapping 1-minute fragments associated with calm wakefulness. Forty two data array entries were identified as informative with respect to this learning sample. The informative entries included both absolute and relative energy values, but the fraction of relative energy values was considerably higher than the fraction of absolute energy values (73.8% vs. 26.2%). The total number of sensors associated with the informative data array entries was 21. Ten sensors were associated with a single informative entry, 6 — with two informative entries, 3 — with three informative entries, one — with four informative entries, and one — with 7 informative entries.

In further analysis, the dimension of EEG data was reduced by splitting the signal into fragments with duration of approximately one minute, and using informative data entries identified at the global learning stage instead of initial EEG.

Classification reliability was evaluated using a test sample that included data corresponding to 17 individuals of both genders (m:f 8:9) aged between 18 and 25, with different levels of stress resistance. Data corresponding to each individual (associated with both stress and calm wakefulness) was split into several non-overlapping fragments (at least 6 stress fragments and 6 calm wakefulness fragments) of a standard length (approximately 1 minute). Two of these fragments (one corresponding to stress and one corresponding to calm wakefulness) were used for individual tuning (or simply discarded if individual tuning was not performed). For other fragments, automated functional state classification was performed and results were compared with the results of classification performed manually by experts of the MSU Faculty of Psychology. For each individual, automated classification reliability was evaluated as a fraction of fragments for which automated classification and manual classification gave the same result.

In order to assess the significance of the individual tuning stage, a similar evaluation of classification reliability for individuals from the same testing sample was performed, in which the individual tuning stage was avoided, all informative data array entries were used for each individual at the individual testing stage, and values h^s and h^c of these entries, typical to an individual, were replaced by the mean values of these entries for the learning sample.

In this approach, both optimizations provided by the individual tuning stage — selection of informative entries applicable to an individual and identification of typical values of these entries for this individual — were discarded, and weights obtained at the global learning stage were used for the classification. Poor results of the classification could be explained in this case mainly by a poor weights selection: a weight evaluation scheme was optimized for the specialized classification method that was corrupted by avoiding one of its stages (moreover, weights selection in this method is only a secondary result of a the global learning stage, whose main task is the detection of informative data array entries among all entries), and hence classification reliability can be essentially increased by a simple recalculation of weights. In order to test this hypothesis, an additional classification reliability evaluation was performed for a case where an individual tuning stage was avoided and a classic optimization method for the weights selection was used. In this test, the Support Vector Machine (Cortes and Vapnik, 1995) with linear kernel was trained using the same learning sample (in which dimension reduction was performed by a transition to informative data entries detected at the global learning stage) and applied to the same testing sample. The software implementation for this approach used LIBSVM (Chang and Ling, 2011); error-cost parameter C was determined using a standard cross-validation technique (Hastie et al., 2011: Sect. 7.10).

4. Results

The results of classification reliability evaluation are summarized in Table 1. The first column contains identifiers of individuals, the second column contains classification reliability for the classification method with the individual tuning stage, the third column contains classification reliability for a similar method in which the individual tuning stage was discarded, and the fourth column contains classification reliability for an SVM-based classification method.

The results show that the presence of the individual tuning stage, which is non-standard for classical machine learning methods, significantly increases the classification reliability.

In rare cases (e.g., individual M14_A7), individual tuning slightly reduces reliability. This effect is generally associated with individuals without distinct individual peculiarities, and the decrease is negligible. The decrease in reliability by the individual learning stage is explained by the low quality of data fragments used for individual tuning.

However, in most cases individual tuning does not decrease the classification reliability, and for individuals with distinct individual peculiarities this stage essentially increases classification quality (e.g., individuals M5_A1, M18_B5).

A similar situation was revealed in a study focused on binary functional state classification based of the analysis of physiological data (ECG, Galvanic skin response, respiratory parameters, etc.) instead of EEG data. The Gradient Boosted Trees algorithm (Breiman et al., 1984) was used for classifier construction, and decision trees that gave incorrect results for data fragments with a known answer (i.e., associated functional state) were discarded at the individual tuning stage. In a few cases, this individual tuning procedure slightly reduced classification reliability, but in these cases both initial and reduced reliability were relatively high. How-

Table 1. Classification reliability

IndID	Rel_IndTuning	Rel_NoTuning	Rel_SVM
M5_A1	0.88	0.38	0.50
M5_A2	0.69	0.40	0.50
M7_A3	1	1	0.83
M7_A4	1	1	0.86
M11_A5	1	1	0.90
M11_A6	1	1	0.91
M14_A7	0.96	1	0.75
J3_A8	1	0.68	1
J3_A9	1	1	0.75
J12_A10	1	1	0.42
M12_B1	1	1	1
M15_B2	0.74	0.37	0.5
M15_B3	1	1	1
M18_B4	1	1	0.6
M18_B5	1	0.65	0.50
J4_B6	1	1	0.50
J4_B7	1	0.30	0.63

Note. Column IndID contains identifiers of individuals from the testing sample. Column Rel_IndTuning contains classification reliability for the standard implementation of the classification method with the individual tuning stage. Column Rel_NoTuning contains classification reliability for a similar method, in which the individual tuning stage is discarded. Column Rel_SVM contains classification reliability for an SVM-based classification method.

ever, cases where individual tuning reduced the classification quality were rare, and cases where the classification quality was essentially increased by the application of individual tuning (particularly, from a relatively low value to a high value) were essentially more frequent. The details of this study and the analysis of its results are given in the volume dedicated to the V Congress of Russian Psychological Society (Lobacheva et al., 2013).

Data availability

The details of the study, as well as EEG data and additional physiological data, are available upon request from the authors.

Acknowledgments

The research was performed under state contract 07.514.12.4030. The research was partially supported by an RFBR grant (project 11–01–00476) and by Government grants of the Russian Federation (agreement No. 02.G25.31.0030, agreement No. 11.G34.31.0054).

The research used equipment purchased through the MSU Development Program.

The authors would like to thank the staff of the MSU Faculty of Psychology and the MSU Faculty of Computational Mathematics and Cybernetics for their valuable comments and discussions.

References

- Danilova, N.N. (1992). *Psikhofiziologicheskaya diagnostika funktsional'nykh sostoyaniy* [Psychophysiological diagnostics of functional states]. Moscow: Moscow State University.
- Hockey, G.R.J., Gaillard, A.W.K., Burov, O. (2003). *Operator functional state assessment: The Assessment and Prediction of Human Performance Degradation in Complex Tasks*. Amsterdam, IOS Press.
- Maglaveras, S., Maglaveras, N., Lekka, I., Bekiaris, A., Penzel, T., Canisius, S., Jacob, N., Leondaridis, L. (2006). Sensation remote monitoring system for enabling the “anytime, anywhere” monitoring of patients with selected sleep disorders. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 1, 3869–3872.
- Wilson, G., Fraser, W., Beaumont, M., Grandt, M., Varoneckas, G., Veltman, H., Svensson, E., Burov, A., Hockey, B., Edgar, G., Stone, H., Balkin, T., Gilliland, K., Schlegel, R.E., Van Orden, K. (2004). *Operator functional state assessment. RTO Technical Report*. Neuilly-sur-Seine, NATO.
- Wasserman, P.D. (1989). *Neural computing: theory and practice*. New York: Van Nostrand Reinhold.
- Cortes, C., Vapnik, V.N. (1995). Support-Vector Networks. *Machine Learning*, 20, 273–297. doi: 10.1007/BF00994018
- Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression*. Hoboken (New Jersey): Wiley-Interscience. doi: 10.1002/0471722146
- Breiman, L. (2001), Random Forests. *Machine Learning*, 45, 5–32. doi: 10.1023/A:1010933404324
- Galatenko, V.V., Livshitz, E.D., Podol'skii, V.E., Chernorizov, A.M., Zinchenko, Yu.P. (2012). Automated real-time classification of psychological functional state based on discrete wavelet transform of eeg data. *IJAM*, 25 (6), 871–882.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM. doi: 10.1137/1.9781611970104
- Cohen, A., Daubechies, I., Feauveau, J.-C. (1992). Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(5), 485–560. doi: 10.1002/cpa.3160450502
- Littlestone, N., Waumuth, M.K. (1994). The Weighted Majority Algorithm. *Information and Computation*, 108, 212–261. doi: 10.1006/inco.1994.1009
- Chang, C.-C., Lin, C.-J. (2012). LIBSVM: A Library for Support Vector Machines. Retrieved from: <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>.
- Hastie, T., Tibshirani, R., Friedman, J. (2011) *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd edition). New York: Springer.
- Breiman, L., Friedman, J., Olshen, R., Stone, C. (1984). *Classification and Regression Trees*. Belmont: Wadsworth.
- Lobacheva, E.M., Galatenko, Y.N., Gabidullina, R.F, Galatenko, V.V., Livshitz, E.D., Lukashenko, T.P., ... Zinchenko, Y.P. (2013). Automated real-time classification of functional states based on physiological parameters. *Procedia — Social and Behavioral Sciences*, 86, 373–378. doi: 10.1016/j.sbspro.2013.08.582

Received: 19 December 2012

Accepted: 01 March 2013

Available Online: 27 November 2013