ICT for Health Science Research
A. Shabo (Shvo) et al. (Eds.)

doi:10.3233/978-1-61499-959-1-130

EEG Data Fusion for Improving Accuracy of Binary Classification

Inna SKARGA-BANDUROVA[a,1], Tetiana BILOBORODOVA[a], Illia SKARHA-BANDUROV[b], Natalia ZAGORODNA[c], Larisa SHUMOVA[a]
[a]*Volodymyr Dahl East Ukrainian National University,* [b]*Luhansk State Medical University,* [c]*Ternopil Ivan Puluj National Technical University*

Abstract. The paper refers to the problem of classification for multiple medical data. The proposed methodology for EEG data processing consists of seven stages and assumes different variations of the Dempster-Shafer technique as a base instrument for data fusion. Attained accuracy is comparable to other more popular algorithms and can be a promising further basis for real-time data classification.

Keywords. EEG, data fusion, classification, technique

1. Introduction

Successful detection and identification of health state indicators at an early stage can increase the success rate of preventive care and prevent impairment health state. In this context, the recent cutting-edge technologies as a Big Data, Internet of Things (IoT), as well as mobile technologies, and wearable devices step-by-step take up the art of diagnostic to a new level. Data extracted from different points or different diagnostic devices are the rich source of information for knowledge discovery. For example, electroencephalogram (EEG) includes data form several electrodes and can be used not only in clinical research to understand the patient status and their diagnosis, but they are also heavily used in the gaming industry, IoT devices as well as for emotion recognition, military scenarios, etc. Hence, many approaches to machine classification have been proposed in this area. At the same time, when using these techniques, a number of issues associated with the implementation of multi-criteria parameters estimation in real time need to be improved. One of these issues is classifying the human biophysical state by EEG indicators. It is still unclear which machine classifier can be sufficient for clinical application when we have several monitoring data. The ability to group multiple monitoring data into a finite set of classes means that a decision must be made based on several sources of information. To analyze signal structures of very different sizes, we need to perform a multi-sensor analysis on the recorded EEG signal. As a next stage to solve this problem, we propose a classification technique based on the combined multi-criteria probability estimates.

Remaining paper is organized as follows. Section 2 discusses the overview of related works and describes achievements in the practical application of data fusion for medical

[1] Corresponding Author.

tasks. Section 3 presents the proposed approach. Results and conclusions are given in Section 4.

2. Related works

In conducting the EEG patient examination, it is often not possible to detect their health status if only a few parameters are available. Abnormal values of a single variable do not indicate a specific state. Also, it is necessary to take into account the conflict of diagnostic signs and their values, the presence of a large amount of uncertain information. Traditionally, uncertainties are handled by probabilistic methods, such as Bayesian methods and the Dempster-Shafer (DS) theory. Thus in [1], the techniques for classification of uncertain data using the DS function are investigated. The authors [2] presented the various strategies for classifying data with DS approach. In [3], it was proposed to use the cross-correlation method for a multi-component electrocardiogram(ECG) using the DS theory for detecting heartbeats. Experiments with clinical data showed high accuracy with a high signal-to-noise ratio. The canonical decomposition of a sequence describing the change of cardiograms is discussed in [4]. In [5], a new method for processing big time series data was proposed using the DS theory and Kullback-Leibler divergence. The authors of [6] conducted a prediction of data by category using several support vector machines(SVM) and combined the obtained information from several SVM models using the DS approach. The work [7] is aimed at the classification of emotions with EEG signals. The data is processed using various classification algorithms that are combined with the DS.An analysis of the literature has shown that using the DS theory in classifying data has a definite advantage. In the presence of multiple, incomplete, uncertain, or redundant data, the use of elements of the DS theory can improve the efficiency of data classification.

3. Method

The proposed technique to classifying real-time data from several sources involves the following six steps: (0) Data normalization (if necessary); (1) Prediction future points; (2) Analysis of residuals; (3) Check for conflicts; (4) Data fusion; (5) Classification; (6) Estimation of classification accuracy. Normalization of parameter values is carried out using transformations of simple Euclidean distance, since, in general case, various indicators of the human biophysical state are expressed in unequal units.

For this paper we use a binary classification EEG signals with the classes “initial” and “opposite”, which determine the two biophysical states of a person, they are patient with open and closed eyes.

3.1. Prediction future points

This step is performed using the Autoregressive Integrated Moving Average (ARIMA) model [8]. As a result, we receive prediction future points and their residuals. The residuals are uncorrelated over time prediction errors for each step of the time series and are used as the sources of evidence for the fusion method.

3.2. Analysis of residuals

Monitoring the status of n patient's indicators occurs in real time. Their measured values at time step k are defined as defined as $Q_i(k)$, $i = 1, 2, ..., n$. While $Pr_i(k)$ value means their predicted values obtained using the ARIMA model and the last measured values. Residuals are computed by $R_i(k) = abs(Q_i(k) - Pr_i(k))$ and $R_1(k)$, $R_2(k)$, $R_3(k)$ are rated as three biggest residuals corresponding to the three studied indicators of the human biophysical state. The residuals of the three indicators are used as the sources of evidence for the DS fusion method, and the result of the fusion is the probability of an initial or opposite biophysical state of a person.

3.3. Calculation of base probability assignment and conflict resolution

The calculation of basic probability assignment (BPA) is performed using the residuals. The BPA function of a person's normal biophysical state can be determined as follows.

$$m(\{N\}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{F * \sigma^2} \right], \tag{1}$$

where x denotes the remainder of the indicator at a given timestamp; F is a constant; μ is the average value of the remainder of the indicator of the biophysical state of a person; σ is the standard deviation of the remainder indicator of the human biophysical state.

In the case of a binary classification, the probability of the opposite state

$$m(\{C\}) = 1 - m(\{N\}) \tag{2}$$

3.4. Check for conflict of probabilities

At this stage, data are checked in the presence of conflicts in probabilities. If there is no conflict the base DS data fusion technique can be used. Otherwise, we propose to utilize the DS fusion method for time series which combines successive time steps and the weighted average.

3.5. Data fusion using DS approach

3.5.1. Basic DS fusion

The implementation of the basic DS fusion method for n > 2 involves equations (3)-(5). For $m_1, m_2, ..., m_n$ n independent sets of BPA, the combinatorial probability fusion rule is defined as follows:

$$m(C) = m_1 \oplus m_2 \oplus ... \oplus m_n(C) = \begin{cases} 0, & C = \Phi \\ \dfrac{1}{1-K} \sum\limits_{\cap_i C_i = C} \prod\limits_{1 \le i \le n} m_i(C_i), & C \ne \Phi \end{cases}, \tag{3}$$

where $K \in (0,1)$ is a coefficient of normalization that can be considered as a measure of conflict between two sets of evidence. The higher the value of K, the greater the conflict between the two proofs regarding the event of interest (in our case, the opposite biophysical state of a person). A comparison of the probability K coefficient of evidence with K_c coefficient of consistency of evidence is used to determine whether the evidence is in conflict or not. The K coefficient is determined as follows.

$$K = \sum_{\cap_i C_i = \Phi} \prod_{1 \le i \le n} m_i(C_i) . \tag{4}$$

And K_c the coefficient of the evidence consistency is determined as

$$K_c = \sum_{\cap_i C_i \ne \Phi} \prod_{1 \le i \le n} m_i(C_i) \tag{5}$$

3.5.2. Time-series DS fusion

If there is a conflict, the calculation is carried out using successive time steps k and the weighted average method. For the merging via weighted average, the base probability m_i from equations (3)-(5) is converted to m_i^*

$$m_i^* = Crd \times m_i \text{, } i=1,\ 2,\ \ldots,\ n. \tag{6}$$

In this case, the multi-criteria combinatorial rule is determined as

$$m(C) = m_1 \oplus m_2 \oplus \ldots \oplus m_n(C) = \begin{cases} 0, & C = \Phi \\ \dfrac{1}{1-K^*} \displaystyle\sum_{\cap_i C_i = C} \prod_{1 \le i \le n} m_i^*(C_i), & C \ne \Phi \end{cases} , \tag{7}$$

where coefficient K^*

$$K^* = \sum_{\cap_i C_i = \Phi} \prod_{1 \le i \le n} m_i^*(C_i) . \tag{8}$$

Suppose that a probability of an opposite state is $m_{(k)} \oplus m_{k-1}(\{C\})$, where *k, k-1*are the current and previous time steps respectively. And this event occurred. In this case, the final result of a multi-criteria fusion is calculated using the basic fusion equations (3)-(5) at the nearest time steps *k*, *k-1*, *k-2*.If the residues from the three sensitive parameters do not conflict with each other, the basic DS fusion method is used. Otherwise, an additional test is introduced, which can be expressed as follows:

$$m_{Ri}(k) \oplus m_{Ri}(k-1)(\{C\}) > P \text{ and } m_{Ri}(k-1) \oplus m_{Ri}(k-2)(\{C\}) > P$$
$$\text{and } m_{Ri}(k-2) \oplus m_{Ri}(k-3)(\{C\}) > P, \tag{9}$$

where $m_{Ri}(k-j)$ denotes the probability assigned *i*-th residual, the biggest one from the three maximal residuals on the time step *k-j, j* = 0, 1, 2. *P* is the constant threshold used to compare the results of a fusion between two adjacent fragments of a particular human biophysical state and represents consistency of evidence. Usually, a constant threshold value is assumed to be 0.8; 0.9. If there is a sensitive parameter *i* that satisfies the above inequalities (9), then the following equation computed using the DS fusion method is the probability of an opposite biophysical state:

$$m_{R\_i}(k) \oplus m_{R\_i}(k-1) \oplus m_{R\_i}(k-2)(\{C\}) . \tag{10}$$

If (9) is not met, a weighted average approach based on equations (7) and (8) are used to D-S fusion and obtaining the probability of the final event.

3.6. Classification

At this stage, we evaluate the combined utility function generated for each class. The class with maximum BPA value is considered as the best entry for prediction model.

$$E = \underset{i \in [class_1, class_2]}{arg\ max} \{m_{Ri}(k)\} \tag{11}$$

3.7. Estimation of classification accuracy

The classification quality is calculated by the formula (12) as follows.

$$A = (T_P + T_N) / n, \tag{12}$$

where T_P is the number of observations with a truly positive result, T_N is the number of observations with a truly negative result, n is the total number of data elements.

4. Results and conclusion

The experiment was carried out using the open source dataset available in UCI Machine Learning Repository [9]. Multichannel EEG dataset consists of 14980 instances obtained from 14 scalp electrodes in two eye states, open eyes and closed eyes, where 8257 (55.12%) instances with open eyes (initial state) and 6723 (44.88%) instances with closed eyes (opposite state). Detailed description EEG eye state data is given in [10]. These data do not require normalization, since the data are from different sources (electrodes), but have the same scale of measurement. All the rest steps were executed. The optimal ARIMA model is the model order (1, 0, 2), which is selected based on the minimum average Bayesian information criterion values. Then all probabilities values were fused. For the presented estimates of the probabilities of each parameter, the conflict between them was checked separately. As the result classification accuracy 0.93 was obtained. An approach to the classification of human biophysical data in real time is proposed. The result of the EEG data classification showed that the proposed method provides accuracy comparable to other algorithms and provides a suitable basis for real-time data classification. Different variations of the DS model make it possible to use this technique in wearable devices, due to its low computational power and the possibility of using incomplete information.

References

[1] C. Lian, S. Ruan, T. Denœux, An evidential classifier based on feature selection and two-step classification strategy. *Pattern Recognition* **48(7)**(2015) 2318-2327.

[2] Q. Chen, A.Whitbrook, U. Aickelin, C. Roadknight, Data classification using the Dempster–Shafer method. *Journal of Experimental & Theoretical Artificial Intelligence* **26(4)**(2014) 493-517

[3] N.J. Napoli, L.E. Barnes ADempster-Shafer Approach for Corrupted Electrocardiograms Signals. *In FLAIRS Conference*(2016)355-360.

[4] I. Atamanyuk, Yu. Kondratenko,Calculation Method for a Computer's Diagnostics of Cardiovascular Diseases Based on Canonical Decompositions of Random Sequences. In proceedings of the ICTERI-2015. Vol. 1356 (2015) 108-120.

[5] H. Jafari, X. Li, L.Qian, A. Aved, T.Kroecker, Multisensor change detection on the basis of big time-series data and Dempster-Shafer theory. *Concurrency and Computation: Practice and Experience*, **29(17)** 2017 p.e4026.

[6] W. Zhu, Y. Sun, Y. Wu, and J. Liu, A New Forecast Model Based on Dempster-Shafer Theory and Support Vector Machine (2015).

[7] M.Z. Soroush, K. Maghooli, S.K. Setarehdan, A.M., Nasrabadi, A Novel Method of EEG-Based Emotion Recognition Using Nonlinear Features Variability and Dempster–Shafer Theory. *Biomedical Engineering: Applications, Basis and Communications,* **30(04)** 2018. 1850026

[8] Introduction to ARIMA: nonseasonal models https://people.duke.edu/~rnau/411arim.htm

[9] UCI Machine Learning Repository. Available at: https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State

[10] O. Rösler, D. Suendermann. A first step towards eye state prediction using EEG. In proceedings of the AIHLS, 2013.