Proceedings of the International Conference of Machine Vision and Machine Learning
Prague, Czech Republic, August 14-15, 2014
Paper No. 137

Multistage Dynamic Multi-classifier Applied to Biosignal Recognition

Maciej Krysmann
Department of Systems and Computer Networks, Wroclaw University of Technology
Wybrzeze Wyspianskiego 27, Wrocaw, Poland
maciej.krysmann@pwr.edu.pl

Abstract- In this paper, there is proposed successful approach to creation more complex Dynamic Ensemble Systems for classification. As background for this research there is presented task of simultaneous electroencephalography (EEG) and electromyography (EMG) signals recognition. Paper describes whole workflow: from signal acquisition, through feature extraction, selection to classification. Multistage Dynamic Multi-Classifier system is precisely described, with separate sections about DES-DT (Dynamic Ensemble Selection with Dynamic Threshold) system which is component of whole system - finally used on two stages of proposed multistage classifier. There is description of way of operation and creation of dynamic ensemble. In the paper there are presented also results of experimental tests which contained not only complete system evaluation, but also tests which allowed to make a choice which classifier will be the best to be used in stage 2. Finally results of comparison are presented with commentary about possible applications of created system.

Keywords: Dynamic ensemble, Multistage classification, Biosignal recognition.

1. Introduction

Biosignal recognition is one of the hardest tasks in area of signal recognition. Moreover it makes possible to help a lot of people, especially disabled ones. Modern machine learning technics could lead to improvement or even make it possible to create inter alia bioprosthetic limbs, computer interfaces for paralysed and new rehabilitation methods. That is obvious that it means getting back to normal social live, improving independence and comfort of handicapped humans, who now need a lot of or even constant care. Electromyography based palm prostheses are commercially available, however their principles of work are very simple in most cases - as the steering factor there are used levels of myosignal from two muscle groups. Moreover that control schema uses those signals only for open-close command and there is need of use open-open sequence or external switch to change grip type. However many research groups are working on and have really good results (Wolczowski et al., 2004, 2010) in EMG grips recognition. Despite of that good results they are mostly valuable only for healthy or with small disability (for example only palm amputation). For many cases there are bigger amputations, nerve dysfunctions which makes myosignal recognition insufficient. Therefore there is need of taking into account other biosignal and of course it is "source signal" - from electroencephalography.

Recognition of brainwave patterns acquired from EEG signal is young and nowadays very dynamically developing field. There are many approaches to the topic, but still best results are obtained from evoked potentials systems. Principle of that solutions operation is observation of potentials in brainwaves evoked by screen or symbols blinking. Drawbacks of that systems are tiredness of human and need of almost perfect synchronisation. Approach which does not use evoked potentials is however incomparable harder problem, due to high complexity of signal which needs to be analysed, processed and comprehended. Non-evoked case implies several problems which needs to be addressed:

- Location of electrodes used in signal acquisition - there are some approximations of areas responsible for different functions of body, but there is much uncertainty. Moreover using too much electrodes on small area will cause in noise amplification.
- Choice of feature extraction - due to high complexity of brainwave signal there is hard to decide how extract information from signal (because of course raw signal cannot be used in classification). Complexity also makes impossible (for todays computation capabilities) to describe EEG signal as mathematical functions. Therefore there are used variety of methods from signal, sound (and other) description and recognition.
- Feature selection methodology - during the process of creating feature extraction schema there is almost impossible (in that complex problems) to decide what features are valuable and should be chosen to classification. There is need of use some of good methods to select the best ones, in EEG recognition there could be different features subsets used for different signal (electrode), therefore choice of features cannot be made arbitrarily.
- Finding best classification systems, optimisation of it and quality verification - long feature vectors require creation of more complicated systems, to make it possible to classify with good quality. Moreover there is need of dealing with multi source feature vectors which are incomparable between different source electrode. Creating long feature vectors assembled from all electrodes vectors makes task impossible to accomplish, in short time, due to complexity of new feature space.

In research describes task of connecting EEG recognition with EMG recognition. Main focus is placed on classification task, but other issues are also described and they are based on earlier author research.

Simultaneous recognition of electroencephalographic and electromyographic signals is needed in works towards hands rehabilitation systems and more accurate hand prosthesis. Despite ease of EMG classification for healthy people, there should be taken into account that muscles or nerves of disabled people could work wrong way, which induces only EMG recognition system insufficient. That is main motivation to the author research described in this paper - to create EEG and EMG based recognition system, which after further development will be used in rehabilitation even before hand transplant rehab, which is crucial for people suffering congenital amputation. Moreover that kind of system like proposed in paper could be used in creating human-computer interface for better and faster use of computer or even in further perspective in gaming industry.

The paper is organised as follows. In section 2 there is main schema of EMG, EEG processing described, there are addressed feature extraction and selection problems, also there is proposed multistage approach. The next section describes Dynamic Ensemble Selection multi-classifier system used in this research, which was proposed in earlier works in which author was involved. Moreover section 3 proposes multistage DES system which is completely novel approach. Section 4 describes conducted experiments which are discussed in section 5. Section 6 concludes and summarises the paper.

2. Biosignal Recognition

Besides complexity, biosignal need to be processed before classification like any other signal. There is impossible to use raw, continuous signal, so there are used feature extraction technics.

2.1. General Processing Schema

As shown on figure 1 main idea of processing the signals is quite simple. Firstly signals acquired are treated with feature extraction methods. There are used variety of technics, which are described in next subsection. In the second step there is used feature selection methodology to get best and most accurate (in meaning of classification quality) features. Final step is of course classification of data and as a result - number of class, to which according to previous steps classifier made a decision.

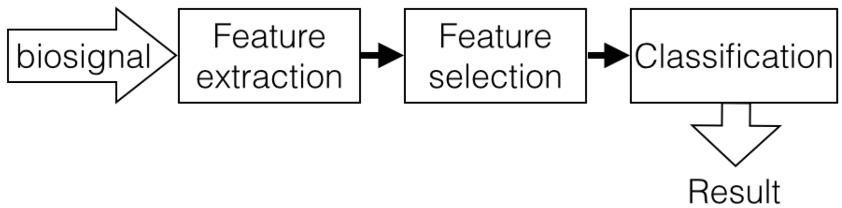

Fig. 1. General Processing Schema

2.2. Feature Extraction and Selection

Variety of features are extracted from raw signal (filtration is made in acquisition hardware) using mostly the MIRtoolbox for MATLAB (Web-1). There are inter alia powers of signal, moments of crossing zero level in window, speed of signal increase and others (see MIRtoolbox: mirfeatures documentation for details).

Nevertheless after extraction there is some uncertainty - if all features are indispensable? Therefore there is used feature selection method. Firstly from over 300 features there are chosen 250 best ones using Information Gain, then features are passed to Principal Component Analysis which as a result gives 100 features of EEG signal and 50 features of EMG signal (number of features was chosen experimentally).

2.3. Multistage Approach

According to figure 1 classification task is divided on two stages. First one is a layer of multi-classifiers designed for each channel (in that case: 8 EEG and 8 EMG) separately. Second stage is classifier which performs fusion of results from stage 1. Use of different classifiers on stage 2 was examined and results are presented in section 5 of this paper.

3. Multi-Classifier System

It is commonly known that for many classification problems multi-classifiers systems outperform single classifiers (Kuncheva 2004), due to inter alia easier overcoming problem of overfitting to the training data. Moreover multiple classifier ensemble built with weaker classifiers could give better results (Kuncheva 2004, Woloszynski 2011), than single stronger classifier, which is furthermore harder to acquire and chose. One of most important problem is to create sufficient ensemble of classifiers. Earlier works showed that dynamically created ensembles exceeds static multi-classifiers (Woloszynski et al., 2010, 2011, Lysiak et al. 2011, Krysmann et al. 2012, 2013).

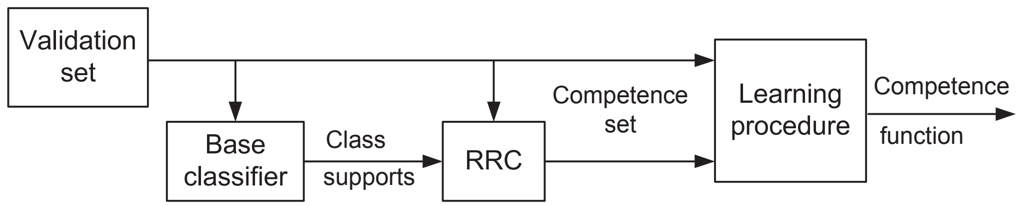

Fig. 2. Use of validation set to perform dynamic ensemble selection

3.1. Dynamic Ensemble Selection

While designing dynamic ensemble type multi-classifier there have to be noticed need of having enough big training (learning) set. That requirement is important because there is need not only to learn base (ensemble candidates) classifiers, but also to perform choice of best classifiers there is need of disjunctive training set, so called validation set ($\mathscr{V}$). Figure. 2 presents way of DES system operating. All candidate classifiers are trained on the same learning set. Let set of candidate classifiers be: $\Psi = \{\psi_1, \psi_2, \dots, \psi_L\}$. Each of them gives vector of class supports $[d_{l1}(x), d_{l2}(x), \dots, d_{lM}(x)]$. Validation set is defined as:

$$\mathscr{V} = \{(x_1, j_1), (x_2, j_2), \dots, (x_N, j_N)\};\ \ x_k \in \mathscr{X},\ \ j_k \in \mathscr{M} \tag{1}$$

Where x_N is feature vector of object form validation set and j_N is correct class of x_N. To make possible dynamic ensemble work there is need of each classifier evaluation, moreover that evaluation must be possible to be done in every point of feature space. Therefore the competence set for each of classifiers, which is build using validation set $\mathscr{V}$, looks as follows, where $C(\psi_l|x_N)$ is value of competence measure for particular classifier in point x_N:

$$\mathscr{C}_l = \{(x_1, C(\psi_l|x_1)), (x_2, C(\psi_l|x_2)), \dots, (x_N, C(\psi_l|x_N))\}. \tag{2}$$

Crucial for dynamic multiple classifiers systems is of course way of obtaining competence set $\mathscr{C}_l$. Efficient method is use of Randomized Reference Classifier, which has proven itself in previous works on DES-DT systems (Woloszynski et al., 2010, 2011, Lysiak et al. 2011, Krysmann et al. 2012, 2013).

3.2. Randomized Reference Classifier

The RRC is considered as stochastic classifier - equivalent to the model of evaluated classifier ψ_L. It is defined using a probability distribution over the product of class supports $[0,1]^M$. In other words, the RRC uses the maximum rule and a vector of class supports $[\delta_1(x), \delta_2(x), \dots, \delta_M(x)]$ for the classification of the feature vector x, where the j-th support is a realization of a random variable (rv) $\Delta_j(x)$.The rvs probability distributions are chosen to satisfy (index l of the classifier ψ_l and its class supports is dropped for clarity, E is expected value operator):
(1) $\Delta_j(x) \in [0,1]$;
(2) $E[\Delta_j(x)] = d_j(x),\ j = 1, 2, \dots, M$;
(3) $\sum_{j=1,2,\dots,M} \Delta_j(x) = 1$,
Above definitions proves that RRC is considered as equivalent of classifier ψ because it produces for feature vector x (on the average) the same vector of class supports.

The RRC performs classification in a stochastic manner, so it is possible to calculate the probability of classification an object x to the i-th class:

$$P^{(RRC)}(i|x) = Pr[\forall_{k=1,\dots,M,\ k\neq i}\ \Delta_i(x) > \Delta_k(x)]. \tag{3}$$

In particular, if the object x belongs to the i-th class, from (3) we simply get the conditional probability of correct classification $Pc^{(RRC)}(x)$.

Crucial for presented above approach is the choice of probability distributions for the rvs $\Delta_j(x), j \in \mathscr{M}$ to satisfy 1-3 conditions. In this paper beta probability distributions are used with the parameters $\alpha_j(x)$ and $\beta_j(x)$ ($j \in \mathscr{M}$). The justification of the choice of that distribution can be found in (Woloszynski et al. 2011).

To get the probability of correct classification of RRC at a point $x_k \in V$, we apply the RRC to a validation point x_k and assume for (3) $i = j_k$:

$$Pc^{(RRC)}(x_k) = \int_0^1 b(u, \alpha_1(x_k), \beta_1(x_k)) [\prod_{j=2}^{M} B(u, \alpha_j(x_k), \beta_j(x_k))] du, \tag{4}$$

where $B(.)$ is a beta cumulative distribution function.

The RRC according to the assumptions is considered as equivalent to the base classifier $\psi_l \in \Psi$, we can use the probability (4) as the competence of the classifier ψ_l at the validation point $x_k \in \mathscr{V}$, i.e.

$$C(\psi_l|x_k) = Pc^{(RRC)}(x_k). \tag{5}$$

As result of previous calculations there is competence set $\mathscr{C}_l$ (2) created. Because values of competence now exist only for objects (points) form validating set $\mathscr{V}$ (1), to make classifier evaluation for particular object which will be classified, there is need of generalisation (approximation) of competence for considered point x_m using all validation points $x_k \in \mathscr{V}$. Previous works (Krysmann et al. 2013) describes research on competence generalisation methods, so called methods of classifier competence learning. Various methods was considered resulting in choice of potential function (6) as both fast and accurate.

$$C(x_m) = \frac{\sum_{k=1}^{N} C_k * K(x_m, x_k)}{max(C_k)}$$

where

$$K(x_m, x_k) = e^{-d(x_m, x_k)} \tag{6}$$

3.3. Classifier Selection

In order to perform selection of classifiers to final ensemble, there is need of decision is classifier ψ_l competent enough in point x_m to be in final ensemble. Instead of proposed in base article (Woloszynski et al. 2011) static threshold of competence, there is used DES Dynamic Threshold method described wider in (Krysmann et al. 2012). The DES-DT methodology works as follows:

1. Set minimal level of competence (called further threshold) needed to allow classifier to ensemble at level 0.9.
2. Check how many of classifiers are competent enough for object x_m to be in ensemble.
3. If there are less that 3 classifiers in ensemble, lower threshold by 0.1 and go to step 2, else go further.
4. Perform classification of object x_m using created ensemble of classifiers.

3.4. Multistage DES-DT based Classification

Complexity of the task forces multistage approach in order to preserve quality and short time system response. Alternative to multistage could be folding feature vectors created from each channel (electrode) into a very long vector. As result there would be vectors with around 1200 features - that would make classification problem very hard to achieve and also time-consuming.

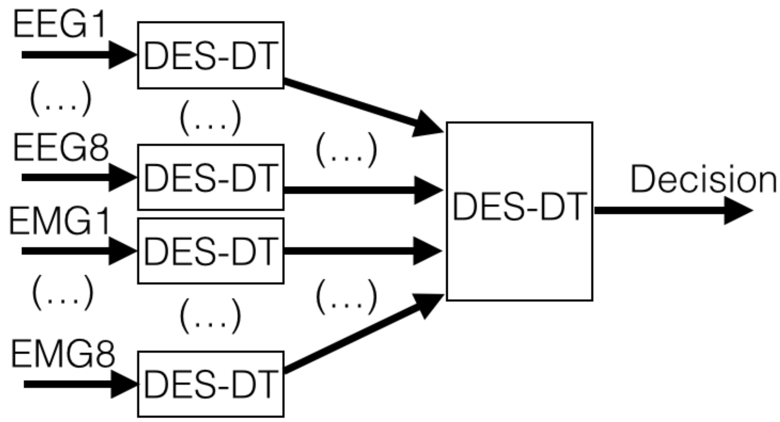

Fig. 3. Diagram of proposed Multistage Dynamic Ensemble System

Therefore approach proposed by author is as follows - also presented on Figure 3 (EEG1, EEG8 and other similar - are feature vectors after selection):

1. Create multi-classifier DES-DT separately for each of signals (channels / electrodes) and execute learning process using data for that channel and correct class number (the same as in final result).

2. Get values of support for each class from each of multi-classifiers from stage 1.

3. Create new feature vector according to that schema: $\{d_1^1, d_2^1, d_1^2, d_2^2, ..., d_1^c, d_2^c\}$, where d is value of support (d_1^1 - support for class 1, d_2^1 - support for class 2) and c is number of classifier from stage 1. Examined problem which is presented in the paper have only two classes, but this approach could be used also in multiple-class problem.

4. Perform learning process using new feature vectors and correct class numbers.

Choice of DES-DT system on stage 1 was motivated by earlier authors works (Woloszynski et al. 2011). Utilising DES-DT on stage 2 was tested and measured against other commonly known classifiers and the results of quality measurements are presented later in the paper.

4. Experiments

Proposed system was tested using experimental benchmark data, to evaluate quality and ensure that using and developing Multistage Dynamic Multi-Classifier is correct path for dealign with multiple source biosignal data.

4.1. Evaluated Stage 2 Classifiers

Besides evaluation of whole Multistage DES-DT system, there were tested several classifiers used in stage 2. DES-DT used in stage 1 could be treated as completely separate classifiers, so prof of their superiority over other known classifiers is described in previous works (Woloszynski et al., 2010, 2011, Lysiak et al. 2011, Krysmann et al. 2012, 2013). Evaluated classifiers:

- DES-DT
- Parzen Density Classifier
- Nearest-Neighbour (5 and 15 size neighbourhood)
- Support Vector Machine
- Nearest Mean Classifier
- Artificial Neural Network (1 hidden layer 10, 20 neurons and 2 hidden layers each 20 neurons)

4.2. Experimental Data

Data was collected during experiments, using software created in Department. That was needed to assure that data, which came from two different acquisition hardware is collected simultaneously. There were recorded 8 channels form EEG and 8 channels from EMG, both with sampling frequency 1000 Hz. There were collected sample trials made by 4 people, each trial was 1 minute long, repeated 4 times. There were performed palm grasp gestures and hand relaxation - those were 2 classes.

5. Results

Results obtained in experiments are presented in Table 1, there were performed 2-class classification. Explanation of classifiers acronyms are in section 4.1. Results presented were obtained as mean of multiple repetitions of 10-fold cross-validation test for all data. Data before averaging was tested using post hoc Holm step-down procedure to ensure that differences between results are statistically significant level (significance level $p = 0.05$).

Table 1. Qualities (% of correct classification) of classifiers in stage 2 (Fuzzers)

Classifier	DES-DT	Parzen	5-NN	15-NN	SVM	NMC	NN 10	NN 20	NN 20-20
Quality in %	84.21	76.56	73.01	72.38	73.62	72.84	63.44	73.37	64.51

It can be easily seen that system with DES-DT in stage 2 outperforms systems with other classifiers as fusion methods. Other methods works on the same level of quality in meaning of statistic results - therefore they can be called statistically the same for that task (except of 2 neural networks with worst qualities). Results have proven that using multistage dynamic multi-classifiers is good approach to classification of different biosignals (different type), also there is space for improvement, because quality of best system could possibly could be better, but it need lot of more research on many parts of system.

6. Conclusion

The paper present successful approach to creating classification system which will perform different biosignal recognition. In paper there is proposed complete workflow for EEG and EMG classification. In that hard task of recognition simultaneous signals from electroencephalography and electromyography results obtained are satisfactory. However there is a place for improvement, but use of multistage dynamic multi-classifier seems to be correct path to achieving good classification systems. Moreover this paper has proven that DES-DT system is really good as component for more complex systems, such as proposed multistage dynamic multi-classifier system. Thanks to that approach there is possible to accomplish classification basing on data which comes from different sources with different nature, without rising the time of classification. That happens because of complete independence of classifiers on stage 1 - they can work in parallel, which follows nowadays trends to make as much as it possible concurrently.

There is important to point out possible applications of created system. Main authors motivation was to create system towards rehabilitation for people suffering congenital amputation. Moreover proposed system could be used as human-computer interface. That kind of interface will allow paralysed people to use computer easily if there will be created correctly adapted menu system.

Acknowledgments

This work is financed from Grant For Young Scientists and PHd Students Development, under agreement: B30067/K0402.

References

Kuncheva I. (2004). “Combining Pattern Classiffiers: Methods and Algorithms” Wiley-Interscience

Krysmann M.,Kurzynski M. (2012). “On new methods of dynamic ensemble selection based on randomized reference classifier” Journal of Medical Informatics Technologies, vol. 20, 101107.

Krysmann M.,Kurzynski M. (2013). “Methods of Learning Classifier Competence Applied to the Dynamic Ensemble Selection”, Advances in Intelligent Systems and Computing Volume 226, 151-160.

Lysiak R., Kurzynski M., Woloszynski T.(2011). “Probabilistic approach to the dynamic ensemble selection using measures of competence and diversity of base classiffers” Lecture Notes in Artiffcial Intelligence 6679, 229-236

Wolczowski A., Kurzynski M. (2010) “Human-machine interface in bio-prosthesis control using EMG signal classiffication”, Expert Systems 27, 53-70

Wolczowski A., Kurzynski M. (2004) “Control of artificial hand via recognition of EMG Signals”, LNCS 3337, Springer Verlag 356-367.

Woloszynski T., Kurzynski M.(2010) “A measure of competence based on randomized reference classiffier for dynamic ensemble selection” In: 20th Int. Conf. on Pattern Recognition, IEEE Computer Press, 4194-4197

Woloszynski T., Kurzynski M. (2011) “A probabilistic model of classiffier competence for dynamic ensemble selection” Pattern Recognition 44 2656-2668.

Website References

Web-1: https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox, consulted 25 Mar. 2014.