Regularized SVMs for classification of image evoked EEG potentials captured from an observer

Valentina Sulimova[1], Sergey Bukhonov[2], Olga Krasotkina[3], Vadim Mottl[1,4], David Windridge[5]

[1] Tula State University, Tula 300012, 92 Lenin Ave., Russia
[2] MIPT, Moscow, Russia
[3] Markov Processes International, NJ 07901,475 Springfield Ave, Suite 401, Summit, USA
[4] Computing Center of the Russian Academy of Sciences, Moscow 119333, Vavilov St. 40, Russia
[5] Middlesex University, London, UK
vsulimova@yandex.ru, o.v.krasotkina@yandex.ru, vmottl@yandex.ru

Abstract. Last few years a trend of using EEG-based brain-computer interfaces for identifying targets within different image types has emerged. Observing target and non-target images usually evokes different neural response of an observer. So there is the two-class problem of classification of EEG potentials as the problem of detection whether the registered EEG signal contains a response to a target image or not. In this paper the event-related potentials recognition is based on Support Vector Machines (SVM) method.

In this paper we propose to use two types of regularization techniques, that express some a priory preferences about the decision and supply the traditional SVM by the possibility to automatically select the most informative features. Jointly with special preprocessing of EEG signals, such regularizations allow to increase quality of target and non-target signals recognition.

Effectiveness of the proposed approaches is demonstrated on the example of recognition of mammograms with pathology by using brain-computer Interface. The best recognition quality in this is 0.936, what essentially improves the best result, obtained in the previous work for joint training with a number of electrodes.

Keywords: EEG signals, brain-computer interface, RSVP, ERP-potentials recognition, mammograms with pathology recognition, SVM, regularization, feature selection.

1 Introduction

The EEG (Electroencephalogram) signal indicates the electrical activity of the brain [1, 2]. Analyzing EEG signals is very important and is used many years for supporting the diagnosis of brain diseases and for contributing to a better understanding of cognitive process. Also it is actively used as the basis for building systems with a brain-computer interface (BCI) [3]. Traditionally BCI helps to restore sensory and motor functions in patients with motor disabilities. But EEG-based BCI can offer advanced tools of interest for other research and applications fields [4]. So, last few years a trend of using BCIs for identifying targets within different image types has emerged [5-7]. Particularly, some years ago it is first proposed to use BCI in the medical domain as a novel high throughput screening method for mammography [7]. It is provided by the advanced computing power of the human visual system, which has been found to be able to identify and classify the image content as target or non-target with a frequency of more than 10 Hz [7]. This paper focuses on just this medical application.

Today many EEG-based BCIs operate based on the principle of evoking and detecting event-related potentials (ERPs) - the neural response of an observer, associated with the presentation of certain stimuli or events to him [8]. Of the numerous methods of eliciting ERPs in BCI applications, in recent years, the rapid serial visual presentation (RSVP) paradigm has been among the most frequently employed. RSVP involves presenting to the observer a series of visual stimuli at a very fast rate, with one stimulus replaced by the next on a small screen area [9, 10]. Of these stimuli, the ones that are of interest to the subject (depending on the task to be performed) are called the target stimuli. The subject’s cognitive response as the target content is flashed up on the screen leads to the generation or enhancement of certain ERP components in the EEG signal, which is much less prominent in the response to the non-targets. Using RSVP, the classification stage typically reduces to the binary discrimination problem i.e. deciding whether the presented stimulus belongs to the target or non-target class by analyzing the recorded EEG signal and detecting the target ERP components.

However, despite the possible use of ERP detection in RSVP tasks, efficient detection and target stimulus recognition remains a difficult problem that requires efficient signal processing and machine learning techniques and is an active research area [11].

One of the difficulties consists in forming appropriate feature space, which would satisfy the compactness hypothesis [12] underlying any machine learning technique. Some papers try to extract intelligent time-domain features [13-15], but the majority of papers use EEG samples as features [16], which is possible due to EEG data being traditionally split into fixed-length segments. However due to a big number of samples (as a rule, about 1000) and a big number of electrodes (for example, 66 in [7]) are used the over-fitting problem is extremely actual. To extract discriminant and non-redundant features different techniques are used, such as Principal Component Analysis [18], Independent Component Analysis [19], spatial filtering for maximization the signal-to-signal-plus-noise ratio between target and non-target stimuli classes [20], finding optimal linear combination of data from different electrodes [11] and so on [16]. But all of them are applied before learning.

In this paper the event-related potentials recognition is based on Support Vector Machines (SVM) method [21], which is very convenient and proved instrument for pattern recognition in linear spaces . To overcome the indicated over-fitting problem we propose to use two types of regularization techniques, expressing some a priory preferences about the decision rule and helping to the traditional SVM to automatically select the most informative features. Jointly with special preprocessing of EEG signals such regularizations allow to increase quality of target and non-target signals recognition.

2 Classification of image evoked EEG potentials as two class pattern recognition problem

Farwell and Donchin [17] reported the first use of a P300- based brain computer interface, used a positive potential in the EEG about 300 msec after an attended target images. The first benefit of this method is that images can be presented at high temporal rates (~10 per second), faster than that required for fully conscious detection, facilitating a high throughput of image material. Computers are unable to analyze and understand imagery as successfully as humans in such complicated area as breast cancer mammography. Typically during the P300 experiment, participants must classify a series of stimuli which fall into one of two classes: targets and non-targets. Targets appear more infrequently than non-targets. The main hypothesis of brain-computer interface development consist in that target and non-target images usually evokes different neural response of an observer, and the classification of EEG potentials consists in detection whether the registered EEG signal is a response to a target image or not target. From mathematical point of view this is a two-class pattern recognition problem.

Any EEG signal , that is obtained from some electrode, is a real-valued discrete signal $\mathbf{x} = [x_1,...,x_m] \in R^m$. For simultaneous using an information from multiple electrodes we concatenate the respective signals, so the object's feature representation remains to be a vector. We furthermore use the same notification for signals from one and from multiple electrodes and provide some clarification if needed.

Let $(X,Y) = \{\mathbf{x}_j, y_j\}, j = 1,...,N$ be a training set of EEG signals $\mathbf{x}_j \in R^m$, accompanied by a class label $y_j \in \{+1;-1\}$. $y_j = +1$ means that a target image was showed to an participant of the experiment when EEG signal $\mathbf{x}_j$ was registered and $y_j = -1$ means, that it was not presented.

The aim of training is to make a decision function $\hat{y}(\mathbf{x})$, such that for a new EEG signal $\mathbf{x} \notin X$ it detects whether $\mathbf{x}$ contains a response to a target image ($\hat{y}(\mathbf{x}) = +1$) or not ($\hat{y}(\mathbf{x}) = -1$).

We consider here three methods for classification of joint EEG signals: first of all we use plain Support Vector Machine approach as a baseline, then we try to include several types of a priori information about the form of EEG signals to make classification more robust, then we include the regularization into SVM criterion to pick up the parts of EEG signals important for target images classification. Also we try several preprocessing methods for EEG signal to increase signal to noise ratio.

3 Methods for two-class EEG potentials recognition

3.1 Traditional Support Vector Machines (SVM) method

The Support Vector Machines (SVM) method [21] is one of the most convenient and proved instrument for making decisions in a linear feature space.

In this case the decision rule is searched as an optimal linear hyperplane

$$d(\mathbf{x};\mathbf{a},b) = \mathbf{a}^T\mathbf{x} + b \quad \begin{cases} \geq 0 \Rightarrow & \hat{y}(\mathbf{x}) = +1, \\ < 0 \Rightarrow & \hat{y}(\mathbf{x}) = -1, \end{cases}$$

which is completely defined by a direction element $\mathbf{a} = [a_1,...,a_m]^T$ and a bias b. It separates objects of positive and negative classes, but supposes misclassifications δ_j for some training objects $\mathbf{x}_j$:

$$\begin{cases} \sum_{i=1}^{m} a_i^2 + C\sum_{j=1}^{N}\delta_j \to \min(\mathbf{a},b,\boldsymbol{\delta}), \\ y_j(\sum_{i=1}^{m} a_i x_{ij} - b) \geq 1-\delta_j, \quad j=1,...,N, \\ \delta_j \geq 0, \quad j=1,...,N, \end{cases} \qquad (1)$$

where the parameter C defines a degree of influence of training objects misclassifications on a hyperplane.

3.2 Reguralized SVM with the requirement of smoothness of the decision rule

It should be noticed, that in the case of recognition of image evoked EEG potentials there are order relations between objects features, as long as they are elements of EEG signals.

In this connection it is natural to choose a decision rule by such a way, that its coefficients corresponding to neighbor features be close to each other, i.e. the decision rule be a smoothed one.

Introduction of prior information about the searched hyperplane allows us to find a more reliable decision, increasing its generalizing ability.

The respective type of regularization was initially proposed by us in [22]. To supply the SVM with the possibility to choose a smoothed decision rule we add a specific regularization member $\sum_{i=2}^{m}(a_i - a_{i-1})^2$ into the optimization criterion in (1). So, the modified SVM problem with the respective regularization can be formulated in the next form:

$$\begin{cases} \sum_{i=1}^{m} a_i^2 + \gamma \sum_{i=2}^{m} (a_i - a_{i-1})^2 + C \sum_{j=1}^{N} \delta_j \to \min(\mathbf{a}, b, \boldsymbol{\delta}), \\ y_j (\sum_{i=1}^{m} a_i x_{ij} - b) \geq 1 - \delta_j, \quad j = 1, \dots, N, \\ \delta_j \geq 0, \quad j = 1, \dots, N, \end{cases} \tag{2}$$

where $\gamma > 0$ is the smoothness coefficient, which determines the degree of influence of the respective regularization member.

The decision of the problem (2) can be easily found by the gradient descent method.

This approach can be most useful for recognition of image evoked EEG potentials on the basis information from one separate electrode.

3.3 Regularized SVM with the possibility to select most informative EEG elements

It is evident, that a set of electrodes provides us with more information than one. From the other hand, the increase in the number of electrodes leads to an increase in the number of features, the aggravation of the overfitting problem and, accordingly, to the reduction of the generalizing ability of the decision rule. But at the same time, it is obvious that not all information form electrodes is useful for making a decision about presence or absence of the EEG potential. So, to increase the recognition quality we need to have an instrument for automatic selection most informative features. Just for this purpose we introduce the second type of regularization for SVM, which is described in this section.

We apply here a very specific case of generalized probabilistic formulation of the SVM [23], that was initially proposed by us in [24] for the case of multimodal recognition, where each modality is represented by some kernel function and was successfully applied the membrane protein prediction [25]. In this paper we reformulate this approach for the case of recognition in the linear feature space and name it Supervised Selective Features Support Vector Machines (SFSVM).

The training optimization criterion problem for the considered type of regularization can be written as follows:

$$\begin{cases} J_{SFSVM}(a_1, \dots, a_m, b, \delta_1, \dots, \delta_N, C, \mu) = \sum_{i=1}^{m} q(a_i \mid \mu) + C \sum_{j=1}^{N} \delta_j \to \min\left(\mathbf{a},\ b,\ \boldsymbol{\delta}\right), \\ q(a_i \mid \mu) = \begin{cases} 2\mu |a_i| & \text{if } |a_i| \leq \mu, \\ \mu^2 + a_i^2 & \text{if } |a_i| > \mu, \end{cases} \\ y_j \left(\mathbf{a}^T \mathbf{x}_j + b\right) \geq 1 - \delta_j, \ \delta_j \geq 0, \ j = 1, \dots, N. \end{cases} \tag{3}$$

The proposed training criterion is thus a generalized version of the classical SVM that implements the principle of *feature selection*.

The threshold μ is named here ”selectivity” parameter because it regulates the ability of the criterion to enact selection of features. When μ is equal to 0, the criterion is equivalent to the classical SVM (1) with the minimal ability to select features. At the same time, values much greater than zero are equivalent to the Lasso SVM [26] with increasing selectivity as μ grows up to full suppression of all features.

The solution of the problem (3) is equivalent to the solution $(\hat{\xi}_i \geq 0,\ i \in I = \{1,\dots,m\},\ \hat{\lambda}_j \geq 0,\ j = 1,\dots,N)$ of the dual problem

$$\begin{cases} L(\lambda_1,\dots,\lambda_N \mid C,\mu) = \sum_{j=1}^{N} \lambda_j - \sum_{i \in I} \frac{\xi_i}{2} \to \max(\lambda_1,\dots,\lambda_N,\xi_1,\dots,\xi_N), \\ \xi_i \geq 0,\ \ \xi_i \geq \sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l x_{ij} x_{il} \lambda_j \lambda_l - \mu^2,\ \ i \in I = \{1,\dots,m\}, \\ \sum_{j=1}^{N} y_j \lambda_j = 0,\ \ 0 \leq \lambda_j \leq \frac{C}{2},\ \ j = 1,\dots,N, \end{cases}$$

and can be expressed in the form

$$\begin{cases} \hat{a}_i = \sum_{j:\hat{\lambda}_j > 0} y_j \hat{\lambda}_j x_i(\omega_j), \\ \qquad i \in I^+ = \left\{ i \in I: \sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l x_{ij} x_{il} \hat{\lambda}_j \hat{\lambda}_l > \mu^2 \right\}, \\ \hat{a}_i = \hat{\eta}_i \sum_{j:\hat{\lambda}_j > 0} y_j \hat{\lambda}_j x_i(\omega_j), \\ \qquad i \in I^0 = \left\{ i \in I: \sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l x_{ij} x_{il} \hat{\lambda}_j \hat{\lambda}_l = \mu^2 \right\}, \\ \hat{a}_i = 0,\quad i \in I^- = \left\{ i \in I: \sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l x_{ij} x_{il} \hat{\lambda}_j \hat{\lambda}_l < \mu^2 \right\}, \end{cases}$$

where $\{0 \geq \eta_i \geq 1,\ i \in I^0\}$ are additionally computed coefficients.

It should be noticed, that the criterion (3), in contrast to other criteria of feature selection [26,27], explicitly splits the entire set of features into two subsets: “support” features $I^+ \cup I^0$ (which will participate in the resulting discriminant hyperplane) and excluded ones I^-.

As a result, the optimal discriminant hyperplane, which is defined by the solution of the SFSVM training problem (3), can be expressed as

$$\sum_{j:\lambda_j > 0} y_j \lambda_j \left(\sum_{i \in I^+} x_{ij} x_i + \sum_{i \in I^0} \eta_i x_{ij} x_i \right) + b \gtrless 0,$$

where the numerical parameters $\{0 \leq \eta_i \leq 1,\ i \in I^0;\ b\}$ are solutions of the linear programming problem:

$$\begin{cases} 2\mu^2 \sum_{i \in I^0} \eta_i + C \sum_{l=1}^{N} \delta_j \to \min(\eta_i, i \in I^0; b; \delta_1, \ldots, \delta_N), \\ \sum_{i \in I^0} \left(\sum_{l=1}^{N} y_j y_l x_{ij} x_{il} \lambda_l \right) \eta_i + y_j b + \delta_j \geq 1 - \sum_{i \in I^+} \left(\sum_{l=1}^{N} y_j y_l x_{ij} x_{il} \lambda_l \right), \\ \delta_j \geq 0, \quad j = 1, \ldots, N, \quad 0 \leq \eta_i \leq 1, \quad i \in I^0. \end{cases}$$

So, the proposed approach has a very significant qualitative advantage over the other methods – it explicitly indicates a discrete subset of support features within the combination, in contrast to other methods that assign some positive (even if small) weight to each feature and, so, require significantly greater amount of memory.

4 High throughput mammogramms analysis on the basis of image evoked EEG potentials recognition

4.1 Data description

In this paper we show that the proposed regularization techniques allows to improve the quality of image evoked EEG potentials recognition on the example of the high throughput screening for mammography.

In this study we use the same EEG data as in 7 that contains details of the methodology.

We use fragments of EEG signals, that were collected from one mammography expert during watching by them a number of series of mammograms and contains information from 66 electrodes.

For each of 66 electrodes EEG fragments have following characteristics:

1) each EEG fragment is 1100-length signal. That corresponds to watching 11 mammograms (100 ms for a mammogram).

2) for fragments of the target class (target objects) one of 11 mammograms (random from 4 to 7) contains some pathology.

3) for fragments of the non-target class (non-target objects) all of 11 mammograms are free of pathology.

Before cutting into fragments, initial EEG signals from each electrode are filtered with a cutoff frequency of 40 Hz.

Full set of obtained 755 objects was randomly split into train set (98 target and 98 non target objects) and test set (275 target and 284 non target objects).

4.2 EEG fragments preprocessing

In this work to improve the speed and quality of recognition we apply 2 types of preprocessing for each fragment of EEG signal: thinning and moving average filtering a window.

Let $\mathbf{x} = (x_i \in R, i = 1,..,m), \quad m = 1100$ be initial EEG fragment.

In accordance with the Kotelnikov's theorem [28] initial EEG fragments are redundant ones, and can be thinned up to 12.5 times without losing an information. In this connection, we thin out each initial EEG fragment in $step = 11$ times. As a result we obtain 100-length EEG fragments instead of 1100-length ones: $\mathbf{x}' = (x_i' \in R,\ i = 1,..,m')$, $m' = m / step = 1100/11 = 100$. Thinning allows us to reduce the dimensionality of the feature space, the amount of required memory and a number of computations.

To decrease influence of the noise component, we filter signals by the moving average filtering with window size $w = 11$:

$$\begin{cases} x_i'' = \dfrac{1}{w}\sum_{k=i-\lfloor \mathbf{w}/2 \rfloor}^{i+\lfloor \mathbf{w}/2 \rfloor} x_k', \quad i = \lfloor \mathbf{w} / 2 \rfloor + 1, ..., m' - \lfloor \mathbf{w} / 2 \rfloor, \\ x_i'' = x'_{\lfloor \mathbf{w}/2 \rfloor + 1}, i < \lfloor \mathbf{w}/2 \rfloor + 1, \quad x_i'' = x'_{N' - \lfloor \mathbf{w}/2 \rfloor}, \ i > m' - \lfloor \mathbf{w}/2 \rfloor. \end{cases}$$

4.3 Experimental results

First of all, we constructed decision rules for each of electrodes separately for 4 modes: with filtering and without filtering, with taking into account the decision rule smoothness or not. The quality of obtained decision rules was estimated by the area under ROC-curve (AUC) for test set.

The choice of smoothness coefficient γ for the smoothness regularization was made automatically for each electrode using leave-one-out procedure for the training set.

The respective results are presented at the table 1. The best result for each electrode is selected by the bold font.

As we can see from the table 1, in most cases the best result is reached , when we use moving average filtering and the decision rule smoothness regularization. Besides, it should be noticed, that in cases when the best quality is observed for modes without moving average filtering and(or) without the regularization, the decision rule quality, as a rule, is small enough, i.e. such situation in most cases is typical for electrodes that do not contain a lot of useful information. In anothers cases such effect can be explained by inappropriate smoothing window width, which was taken equal to 11 and was not specially chosen for each electrode.

In addition to the table 1 contains results for simultaneous use of all 66 electrodes et once. The table shows, that, as expected, the quality of the decision rule for all electrodes is worse, than the quality of decision rules for a number of separate electrodes. This fact confirms the overfitting problem and necessity to reduce the dimensionality of the feature space.

For the next step of experiments, we choose 7 electrodes with highest recognition quality. There are electrodes 27, 28, 30, 33, 37, 42 and 53. We combined the signals from these electrodes, used the moving average filtering and applied the SVM with decision rule smoothness regularization. Under these assumptions the AUC=0,906 was achieved for the test set. This value exceeds the result for all electrodes and for each individual electrode. It confirms again the overfitting problem and its decrease with decreasing the number of electrodes and, respectively the number of features.

Table 1. AUC-values for different electrodes and different training modes

smoothing		-	-	+	+	smoothing		-	-	+	+
smoothness of the decision rule		-	+	-	+	smoothness of the decision rule		-	+	-	+
Electrode's number	1	0,6858	0,6855	0,657	**0,6985**	Electrode's number	34	0,5012	**0,5036**	0,4781	0,4944
	2	0,6431	0,6192	0,6009	**0,6611**		35	0,5012	**0,576**	0,5493	0,5558
	3	0,6644	0,6758	0,6308	**0,6901**		36	0,6348	0,6711	0,63	**0,6985**
	4	0,6683	0,6473	0,6411	**0,689**		37	0,7203	0,8257	0,7364	**0,8451**
	5	**0,6582**	0,6174	0,6124	0,6167		38	0,6085	0,696	0,6259	**0,6997**
	6	0,6425	0,6791	0,6292	**0,6806**		39	0,7062	0,6931	0,6811	**0,7431**
	7	**0,6091**	0,5954	0,5827	0,5816		40	**0,5897**	0,5857	0,5464	0,547
	8	0,6766	0,6778	0,6242	**0,6873**		41	0,6088	0,655	0,6291	**0,6742**
	9	0,5714	0,6161	0,5782	**0,6267**		42	0,7708	0,8584	0,7815	**0,8672**
	10	**0,6053**	0,555	0,5384	0,5746		43	**0,5378**	0,5366	0,5268	0,5518
	11	0,5509	0,5613	0,5724	**0,576**		44	0,545	0,5642	0,5664	**0,5735**
	12	**0,5929**	0,5478	0,5504	0,5593		45	0,5924	0,6715	0,6281	**0,6961**
	13	0,5368	0,5254	0,5541	**0,5589**		46	0,6428	0,7447	0,6951	**0,7671**
	14	0,5295	**0,5405**	0,5226	0,5397		47	0,5491	**0,6056**	0,5669	0,5759
	15	0,5595	0,6175	0,5876	**0,5891**		48	0,5628	**0,6101**	0,5775	0,5869
	16	**0,5191**	0,5157	0,5171	0,5119		49	0,5713	**0,5866**	0,5371	0,5455
	17	0,5756	0,5609	0,563	**0,5891**		50	0,4918	**0,5419**	0,5268	0,5388
	18	0,5844	0,5601	0,5798	**0,6629**		51	0,5852	**0,603**	0,575	0,6005
	19	0,6222	0,5994	0,5969	**0,6692**		52	0,5622	0,5515	0,5829	**0,5787**
	20	0,5214	0,6301	0,5672	**0,6316**		53	0,6943	0,7688	0,7287	**0,7855**
	21	0,6377	0,6699	0,6202	**0,6944**		54	0,5702	**0,6184**	0,5608	0,5666
	22	0,5103	0,508	**0,5191**	0,5108		55	0,5238	**0,5485**	0,5354	0,5248
	23	0,5471	0,5877	0,5473	**0,5952**		56	0,6548	0,6617	0,6275	**0,6805**
	24	0,616	0,6358	0,585	**0,6594**		57	0,6458	**0,6638**	0,6382	0,6146
	25	0,488	0,4853	**0,5191**	0,5177		58	**0,7606**	0,7137	0,6748	0,7351
	26	0,4931	0,4864	**0,5142**	0,5072		59	**0,7382**	0,7115	0,6733	0,6547
	27	0,748	0,8083	0,7277	**0,8289**		60	**0,7493**	0,7234	0,6854	0,7472
	28	0,7385	0,8305	0,7112	**0,8351**		61	**0,5537**	0,5049	0,5399	0,5127
	29	0,6363	**0,6808**	0,6716	0,663		62	0,6219	**0,6625**	0,5983	0,6011
	30	0,7717	0,8177	0,7439	**0,8252**		63	0,7336	**0,7465**	0,7009	0,7453
	31	0,5568	**0,5786**	0,5422	0,5428		64	**0,6066**	0,6001	0,5831	0,5783
	32	0,6661	0,6812	0,644	**0,7248**		65	0,4933	0,4814	**0,5129**	0,5021
	33	0,7652	0,8308	0,7312	**0,8518**		66	0,6191	**0,6232**	0,5492	0,5475
	all	0,764	0,805	**0,815**	0,811						

At the next stage we try to use the sparse regularization proposed in (section 3.3) for smart dimensionality reduction of the feature space.

In this experiment 13 electrodes were used, that show different recognition quality (small and high) in accordance with the Table 1. The respective electrodes' numbers can be seen at the Figure 1. For this set of electrodes the training in accordance with the proposed SFSVM method (section 3.3) was made a number of times for different selectivity values.

At the Figure 1 for each selectivity value the achieved AUC value is presented. Also features that remained after training are shown as blue bars there.

The Figure 1 shows that increase of selectivity value leads to discarding features, which are in start and end positions of EEG fragments. It seems to be naturally, because these are little informative due to smoothing by a sliding window. Besides, increase of selectivity value till some threshold leads to increase of AUC.

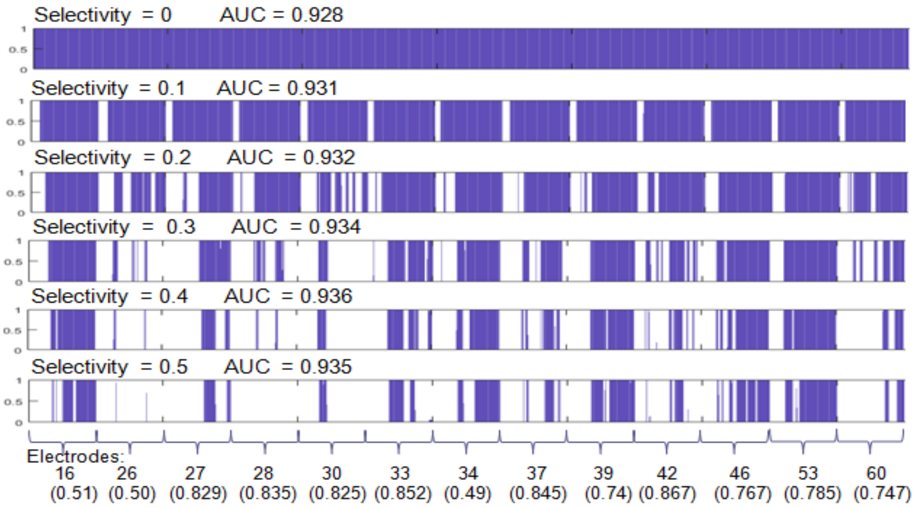

Figure 1. The results of informative feature selection in accordance with SFSVM for different selectivity values

The best recognition quality in this case is 0.936, what additionally improves results, that obtained in previous experiments. Besides, it should be noticed, that results for both types of regularization essentially improve the best result, obtained in [7] for joint training with a number of electrodes, which for the considered data was 0.88.

5 Conclusion

In this paper we propose the special type of preprocessing of EEG signals and two types of regularization for SVM to improve the speed and the quality of recognition of image evoked EEG potentials, captured from an observer.

Effectiveness of the proposed approaches is demonstrated on the example of recognition of mammograms with pathology by using brain-computer Interface. The best recognition quality in this is 0.936, what essentially improves the best result, obtained in the previous work for joint training with a number of electrodes.

The work supported by the Russian Foundation for Basic Research, projects 18-07-01087, 17-07-00993, 17-07-00436.

References

1. Zenkov L. R. Clinical electroencephalography. /3-d edition. — Moscow: MEDpress-inform, 2004. — 368 p. ISBN 5-901712-21-8
2. Teplan, M. 2002. Fundamentals of EEG measurement. Measurement Science Review 2(2), pp. 1-11
3. Wolpaw J.R., McFarland D.J., Neat G.W., Forneris C.A., An EEG-based brain-computer interface for cursor control. Electroencephalography & Clinical Neurophysiology. Vol 78(3), Mar 1991, 252—259
4. Eizagirre A., Vall A. and D. at al. EEG/ERP Analysis: Methods and Applications (2014).
5. Poolman P, Frank R M, Luu P, Pederson S M and Tucker D M 2008 A single-trial analytic framework for eeg analysis and its application to target detection and classification NeuroImage 42 787–98 (2008)
6. Sajda P et al 2014 Evoked neural responses to events in video IEEE J. Sel. Top. Signal Process. 8 358–65 (2014)
7. C.Hope, A. Sterr,P.E. Langovan, N.Geades, D.Windridge, K.Young, K.Wells.High Throughput Screening for Mammography using a Human-Computer Interface with Rapid Serial Visual Presentation (RSVP) / - Proc. SPIE 8673, Medical Imaging 2013: Image Perception, Observer Performance, and Technology Assessment, 867303 (March 28, 2013); doi:10.1117/12.2007557.
8. S. Sur and V. Sinha. Event-related potential: An overview. Industrial Psychiatry Journal 18(1), pp. 70. 2009.
9. C. Keysers, D. Xiao, P. Földiák and D. Perrett. The speed of sight. J. Cogn. Neurosci. 13(1), pp. 90-101. 2001.
10. O. de Bruijn and R. Spence. Rapid serial visual presentation: A space-time trade-off in information presentation. Presented at Advanced Vis. Interf. (2000).
11. Cecotti, H., Eckstein, M.P., Giesbrecht, B. Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering. IEEE Trans. Newral Netw.Learn. Syst. 25, 2030-2042 (2014)
12. Braverman, E. M. Experiments on training a machine for pattern recognition. PhD Thesis. Moscow (1961)
13. V. Abootalebi, M. H. Moradi and M. A. Khalilzadeh. A new approach for EEG feature extraction in P300-based lie detection. Comput. Methods Programs Biomed. 94(1), pp. 48-57. (2009).
14. Z. Amini, V. Abootalebi and M. T. Sadeghi. Comparison of performance of different feature extraction methods in detection of P300. Biocybernetics and Biomedical Engineering 33(1), pp. 3-20. (2013).
15. Tran, L. EEG Features for the Detection of Event-Related Potentials Evoked Using Rapid Serial Visual Presentation. PhD Thesis. 63 p.(2014)

16. Lees S., Dayan N., Cecotti H., McCullagh P., Maguire L., Lotte F., Coyle D.. A review of rapid serial visual presentation-based brain-computer interfaces. J Neural Eng. 2018 Apr;15(2):021001. doi: 10.1088/1741-2552/aa9817. (2018)
17. Farwell LA, Donchin E: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. EEG Clin Neuroph. 1988, 70:510–523.
18. Alpert G. F, Manor R, Spanier A B, Deouel L Y and Geva A B. Spatiotemporal representations of rapid visual target detection: a single-trial EEG classification algorithm IEEE Trans. Biomed. Eng. 61 2290–303 (2014)
19. Kumar S. and Sahin F. Brain computer interface for interactive and intelligent image search and retrieval. High Capacity Optical Networks and Emerging/Enabling Technologies (IEEE), pp. 136–40 (2013)
20. Rivet B et al. xDAWN algorithm to enhance evoked potentials: application to brain–computer interface IEEE Trans. Biomed. Eng. 56 2035–43 (2009)
21. Vapnik, V.N. Statistical Learning Theory / — Wiley-Interscience, 1998. 768 p. (1998).
22. Mottl V.V., Dvoenko S.D., Seredin O.S., Krasotkina O.V. Pattern recognition learning taking into account the criterion of smoothnessof the decision rule. // Control and Inform.: proc. of the chair of autom. and rem. contr. of TSU. (2000)
23. Tatarchuk, A., Mottl, V., Eliseyev, A., Windridge, D.: Selectivity supervision in combining pattern-recognition modalities by feature- and kernel-selective Support Vector Machines. Proc. ICPR (2008).
24. Tatarchuk, A., Urlov, E., Mottl, V., Windridge, D.: A support kernel machine for supervised selective combining of diverse pattern-recognition modalities. In: El Gayar, N., Kittler, J., Roli, F. (eds.) MCS (2010).
25. Tatarchuk A., Sulimova V., Mottl V., Windridge D. Supervised Selective Kernel Fusion for Membrane Protein Prediction. M. Comin et al. (Eds.): Pattern Recognition in Bioinformatics, Lecture Notes in Computer Science Volume 8626, 2014, pp.98-109. (2014)
26. Bradley P., Mangasarian O.: Feature selection via concave minimization and support vector machines. In International Conf. on Machine Learning (1998)
27. Wang, L., Zhu, J., Zou, H.: The doubly regularized support vector machine. Statistica Sinica, 01/2006; 16:589–615 (2006)
28. Burachenko D.l., Kluev N.N., Korjik V.I., Fink V.I. et al. General communication theory. / Fink L.M. eds. – L.: BAC, 1970. – 412c.
29. S. J. Luck, A Introduction to the Event-Related Potential Technique. Cambridge, MA, USA: MIT Press, (2005)
30. K. E. Hild, M. Kurimo, and V. D. Calhoun, “The sixth annual MLSP competition, 2010,” in Proc. IEEE Int. Workshop Mach. Learn. Signal Process., Kittila, Finland, Sep. 2010, pp. 107–111.
31. Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. Journal of Experimental Social Psychology, 40(2), 216-224