Automatic Seizure Detection Incorporating Structural Information

Borbala Hunyadi[1,2], Maarten De Vos[3,1,2], Marco Signoretto[1,2], Johan A. K. Suykens[1,2], Wim Van Paesschen[4], and Sabine Van Huffel[1,2]

[1] Department of Electrical Engineering (ESAT), Division SCD, Katholieke Universiteit Leuven, Leuven, Belgium
[2] IBBT-K.U.Leuven Future Health Department, Leuven, Belgium
[3] Neuropsychology Lab, Department of Psychology, University of Oldenburg, Oldenburg, Germany
[4] Department of Neurology, University Hospital Gasthuisberg, Leuven, Belgium
{borbala.hunyadi,maarten.devos,marco.signoretto,
johan.suykens,sabine.vanhuffel}@esat.kuleuven.be
wim.vanpaesschen@uz.kuleuven.ac.be

Abstract. Traditional seizure detection algorithms act on single channels ignoring the synchronously recorded, inherently interdependent multichannel nature of EEG. However, the spatial distribution and evolution of the ictal pattern is a crucial characteristic of the seizure. Two different approaches aiming at including such structural information into the data representation are presented in this paper. Their performance is compared to the traditional approach both in a simulation study and a real-life example, showing that spatial and structural information facilitates precise classification.

1 Introduction

Epilepsy is the second most common neurological disorder after stroke. Over 0.5% of the worldwide population is affected with epilepsy, and approximately 20% of them are not responding to anti-epileptic drugs. The manifestation of this disease is the epileptic seizure. It is an abnormal, synchronous activity of the neurons in the brain. An automatic seizure detection system could help the diagnosis of epilepsy, reducing the workload of clinicians by supporting visual inspection of EEG. Several seizure detection algorithms have been developed in the past decades, applying various methods including time-frequency analysis [5], [4], nonlinear time series analysis [6], feature extraction and machine learning techniques [8], [3], [7].

The drawback of the existing algorithms is the fact that they act on single channel data, however, the spatial distribution and evolution of the ictal pattern are crucial characteristics of the seizure. A two-step system could overcome this issue, where, in the first step a decision is made for each channel by a separate classifier, and in the second step the outputs of these classifiers serve as the input of a combined, final decision procedure. Greene et al. [3] compared such a *late*

integration method to an *early integration method*, where the features extracted from each channel are sorted and stacked into a long feature vector, which is then used to train a single classifier. The *early integration method* is proved to be superior in performance, by "treating the channels as related, exploiting their statistical inter-relationship and the synchronously recorded nature of the EEG" [3]. Shoeb et al. [8] developed a patient-specific seizure detector, which relies on features describing the temporal evolution, the spectral and the spatial structure of the EEG. In order to capture spatial information, the features of each channel are concatenated to form one feature vector. As opposed to the former study, where the sorting operation was intended to remove spatial information, the goal of the stacking in this case is to drive the attention to the locations corresponding to the channels consistently showing seizure activity.

In the present paper a novel alternative solution is investigated. The features extracted from the multichannel data are represented in the form of a matrix as an input to a classifier. The matrix representation of the data helps preserving and exploiting the inherent spatial structure of the multichannel EEG data. Moreover, recent studies ([1], [12]) show that higher-order representation of signals reduces the small sample-size problem, facilitating a precise classification performance even for low number of training points and outperforms traditional vector representation.

We investigate long-term recordings containing EEG data from refractory epilepsy patients undergoing presurgical evaluation. The immediate intervention after seizure onset is necessary to collect information about the seizure and is a key to successful localization of the seizure focus. After sufficient information has been acquired the patient can leave the hospital. Thus, it is essential that the algorithm can learn the seizure pattern after a few occurrences. Moreover, a low number of training points may be provided by seizures of possibly short length. However, the training of a traditional classifier might need a relatively high number of data points. We will show here that the proposed approach performs well when relatively little information is available.

2 Materials and Methods

2.1 EEG data

EEG recordings from 14 patients with refractory partial epilepsy were included in the study. The patients were selected based only on the criterion that at least 4 seizures were recorded during their stay in the epilepsy monitoring unit. Data were sampled at 250Hz, an average referenced electrode montage was used and the electrodes were placed according to the standard 10-20 % 19 electrode system with two additonal electrodes placed over the sphenoidal temporal region.

2.2 Feature Extraction

EEG was segmented into 2s long non-overlapping windows. A total number of 19 features were extracted from each channel of each segment. Thus, one data

point represents the multichannel EEG window in the form of a 19×21 matrix. The features are listed in Table 1 and are selected from the features used in [13].

Table 1. Extracted Features

Frequency domain features	Total power, Peak frequency, Spectral edge frequency (80% , 90% , 95%), Mean and normalized power in the frequency bands (1-3 Hz, 4-8 Hz, 9-13 Hz and 14-20 Hz)
Time domain features	Number of zero crossings, maxima and minima, skewness, kurtosis, root mean square amplitude

2.3 Classification Approaches

Single-channel Classification with Late Integration. Traditional seizure detection systems analyze EEG channels independently and integrate the decision outputs of the single channels into a global decision during a separate step. There are several different strategies to follow. The outputs of the channel classifiers can be binary or probabilistic; post-processing can be performed applying a moving average filter on the outputs from the consecutive epochs [14]; the channel outputs can be integrated via mean, max, or min score, or majority vote [3]. The number of channels contributing to the global score might as well be limited [7]. In the current study the length of the feature vectors corresponds to the number of extracted features. The single-channel feature vectors are fed to a least-squares support vector machine (LS-SVM) [11]. Finally, the binary outputs of single epochs are integrated by a simple OR function.

LS-SVM was chosen because of its low computation costs due to solving a set of linear equations instead of quadratic programming. Moreover, the model is based on all data (all support values are nonzero), which can be beneficial in case of small samples. We use LS-SVMlab toolbox (www.esat.kuleuven.be/sista/lssvmlab, [2]), which performs automatic tuning of the model parameters applying coupled simulated annealing [10].

Including Spatial Information via Early Integration of Feature Vectors. In this approach the feature vectors extracted from each EEG channel are stacked into one long feature vector of length $I \times J$, where I is the number of channels and J is the number of extracted features. One LS-SVM is trained and used for classification. As explained above, the concatenation of the channels in fixed order aims at including spatial information and exploiting the synchronously recorded and inter-dependent nature of multichannel EEG.

In both approaches applying LS-SVM a linear kernel was chosen considering the high dimensionality of input data and the small sample size. Moreover, the choice of linear kernel facilitates a meaningful comparison with the linear model used in the nuclear norm learning approach (see below).

Including Structural Information via Nuclear Norm Regularization. We consider the following model:

$$\hat{y} = \langle A, X \rangle + b, \tag{1}$$

where X is the input pattern, A is a matrix of the same size, $\langle \cdot, \cdot \rangle$ indicates the inner product, and b is a bias term. Decisions are made according to $\text{sign}(\hat{y}) \in \{-1, 1\}$.

Such formulation allows to keep the natural matrix representation of the EEG data: $X \in R^{I \times J}$, where I is the number of channels, and J the number of features. The classifier (namely the pair $(A,\ b)$) is found solving a non-smooth convex optimization problem using a nuclear norm penalty:

$$\min_{(A,b)} F(A, b) = f(A, b) + \mu \|A\|_{\Sigma,1}, \tag{2}$$

where $f(A, b)$ is the quadratic error function accounting for the misclassification. This choice was made specifically because the same loss function is used in LS-SVM classification. Further, μ is a tuning parameter and $||A||_{\Sigma,1}$ is the nuclear norm of the matrix A with singular values σ_i :

$$||A||_{\Sigma,1} = \sum_i \sigma_i. \tag{3}$$

The tuning parameter μ, as well as the tuning parameters of LS-SVM formulation were chosen according to the 5-fold cross-validation of the misclassification error. Regularization via nuclear norm conveys structural information from the matrix by ensuring a low-rank solution. In the current application the low-rank classifier matrix represents the features and spatial distribution characteristic for the patient. Theoretical background and motivation behind the use of nuclear norms as heuristic ensuring low-rank solution, and details of the convex optimization algorithm can be found in [9] and references therein.

3 Results

3.1 Simulation on Randomized Training and Test Set

Performance of the matrix nuclear norm learning (NNL) algorithm was compared to the early integration (EI-LSSVM) and late integration (LI-LSSVM) solution. The test set consisted of 50 % of the available positive data points randomly selected from all segments of all recorded seizures of the given patient, and negative data points randomly selected from all non-seizure segments. The positive to negative ratio was fixed to 1:50 keeping into account the intrinsic unbalanceness of the problem. Classifiers were built based on increasing sizes of training sets, and were all tested on the same fixed test dataset. In total 5 training sets were randomly generated for each of the 14 patients and each training set size, using all available EEG segments during the random selection, excluding the ones in the test set. Performances are reported as the mean area under the

curve (AUC) of the 5×14 trials (ordinate) for each training set size (abscissa) as seen on Figure 1(a). The variability of AUC among the trials is depicted on Figure 1(c).

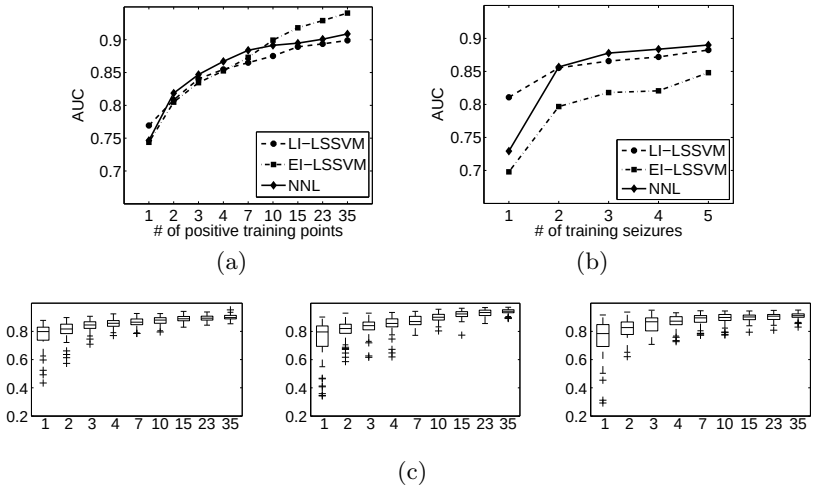


Fig. 1. Results of the simulation on randomized trainingset: (a) mean AUC over all trials in function of the training set size and (c) boxplots of AUC showing the variability in performance between the individual trials. Results of the real-life setting: (b) mean AUC values in function of the number of seizures included in the training set.

NNL approach is able to capture useful information after a few training points, and performs the best for small sizes of training sets. This advantage is not yet seen in case of one training point, although good generalization from only one training point is obviously not feasible for any learning algorithm. On the contrary, EI-LSSVM benefits the most from including additional training points, and it performs the best if greater number of training points are available.

3.2 Real-life Setting

The results of the above simulation are revised in the analysis of the performances of NNL and EI/LI-LSSVM in a real-life setting. A patient-specific seizure detection system first records EEG until the first seizure occurs, and then builds a classifier based on the collected data. Afterwards it goes on with recording and classifying each new data segment in parallel. Once an other seizure occurs, the classifier is updated in order to reach better classification performance based on the additional information.

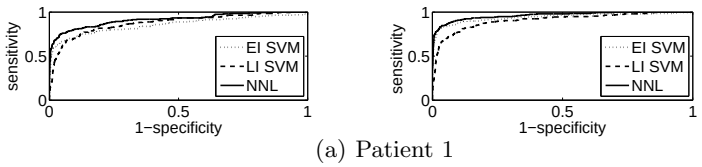
In order to simulate such an environment, the available seizures are ordered based on the time of their occurrence, seizures occurring later on time serve as

test set, together with the appropriate number of non-seizure segments. The first classifier is now built based on the segments of the seizure occurring first in time, then new classifiers are built adding the segments of the consecutive seizures to the training set.

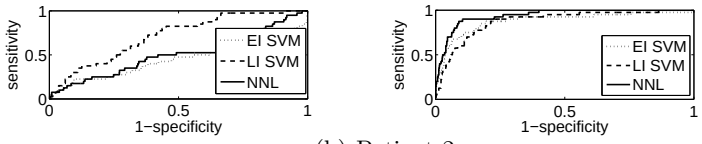
However, in a patient-specific setting, if the first seizure occurs shortly after the start of the recording, there might not be enough diversity of negative training points. Brain activity in different physiological brain state and artifacts have peculiar patterns, and some of them might resemble seizures. In order to include a more complete and representative set of non-seizure segments alpha activity, sleeping and drowsiness patterns, muscle artifacts, chewing artifacts, rapid eye movement and repeated blinking patterns were collected from 29 different patients and were included in a semi-patient-specific training set.

The mean AUC of the three approaches over all the patients with at least five training seizures is depicted on Figure 1(b). NNL proves to be superior when two or more seizures are included in the training set, while LI-LSSVM performs better when only one seizure is available.

Figure 2 illustrates two different scenarios regarding patient-by-patient performance. The receiver operating characteristic (ROC) curves of the different classification approaches are depicted for two patients given one and two training seizures. In the former case NNL and EI-LSSVM are able to capture enough information after one seizure, while in the latter case they require two seizures for their optimal performance. They are outperformed by LI-LSSVM when only one training seizure is available.



(a) Patient 1



(b) Patient 2

Fig. 2. ROC of the different approaches including one and two seizures in the training set (left and right panel respectively)

3.3 Computational costs

The computational costs using the three different approaches are compared in Table 2. The computational times were tested given different sizes of training

sets. Ten positive datapoints correspond to a 20s long seizure, thus the chosen training set sizes represent training sets including increasing number of seizures from one up till five. For small training set sizes EI-LSSVM has the shortest running time, however, its running time increases at a faster rate and exceeds NNL running time given five training seizures. Nevertheless, they both remain within practical limits, unlike LI-LSSVM.

Table 2. Computational times (s)

# of positive datapoints	10	20	30	40	50
EI-LSSVM	2.7	12.6	37.8	79.8	151.1
LI-LSSVM	44.2	248.4	769.7	1652.1	3153.4
NNL	49.8	74.1	96.3	118.0	145.5

4 Discussion

The results acquired in the simulation study show clear superiority of the two approaches incorporating spatial/structural information over the traditional single-channel method. However, EI-LSSVM performance clearly decreases in the real-life experiment. Moreover, LI-LSSVM shows higher mean performance than NNL given one training seizure. The principal difference between the two studies is that data points from different seizures are included in the training set in the simulation study, while the data points of the same seizure are included in the real-life example. Given a patient with certain variability in spatial distribution among the seizures, EI-LSSVM fails to generalize, while LI-LSSVM easily overcomes this problem due to the simple OR function integrating the channel decisions. NNL nevertheless outperforms EI-LSSVM, suggesting that the structural information exploited by its learning algorithm is more flexible than the spatial information encoded in the concatenated feature vector, i.e. the input of EI-LSSVM. Moreover, given multiple seizure patterns, NNL is capable of exploiting additional information and performs slighly better than the independent single-channel LI-LSSVM.

Determining the optimal set of features might improve classification performance, but is beyond the scope of this paper. Furthermore, a future study applying the classifiers as on-line seizure detectors should be carried out and evaluated by clinically relevant measures such as sensitivity, false detection rate over time and alarm delay.

Extensive analysis is to be carried out aiming at defining the exact circumstances under which one classification approach is favorable over the other. A final seizure detection system may be developed, which automatically selects the most appropriate learning and classification technique given the actually available training set.

Acknowledgments. Research supported by Research Council KUL: GOA MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), PFV/10/002 (OPTEC), IDO 05/010 EEG-fMRI, IOF-KP06/11 FunCopt; Flemish Government: FWO G.0302.07 (SVM), FWO G.0427.10N (Integrated EEG-fMRI); IWT: TBM080658-MRI, IBBT; Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, 2007-2011); EU: Neuromath (COST-BM0601)

References

1. Cai, D., He, X., Weng, J.R., Han, J., Ma, W.Y.: Support tensor machines for text categorization. Tech. rep., Computer Science Department, UIUC, UIUCDCS-R-2006-2714 (April 2006)
2. De Brabanter, K., Karsmakers, P., Ojeda, F., Alzate, C., De Brabanter, J., Pelckmans, K., De Moor, B., Vandewalle, J., Suykens, J.: LS-SVMlab toolbox user's guide version 1.7. Tech. rep., ESAT-SISTA, K.U.Leuven (2011)
3. Greene, B., Marnane, W., Lightbody, G., Reilly, R., Boylan, G.: Classifier models and architectures for eeg-based neonatal seizure detection. Physiol. Meas. 29(10), 1157 (2008)
4. Guerrero-Mosquera, C., Malanda Trigueros, A., Iriarte Franco, J., Navia-Vazquez, A.: New feature extraction approach for epileptic eeg signal detection using time-frequency distributions. Med. Biol. Eng. Comput. 48, 321–330 (2010)
5. Meier, R., Dittrich, H., Schulze-Bonhage, A., Aertsen, A.: Detecting epileptic seizures in long-term human eeg: A new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. J. Clin Neurophysiol. 25(3), 119–131 (2008)
6. Polychronaki, G.E., Ktonas, P.Y., Gatzonis, S., Siatouni, A., Asvestas, P.A., Tsekou, H., Sakas, D., Nikita, K.: Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection. J. Neural Eng. 7(4), 046007 (2010)
7. Saab, M., Gotman, J.: A system to detect the onset of epileptic seizures in scalp eeg. Clin. Neurophysiol. 116(2), 427 – 442 (2005)
8. Shoeb, A., Guttag, J.: Application of machine learning to epileptic seizure detection. In: Fürnkranz, J., Joachims, T. (eds.) Proceedings of the 27th International Conference on Machine Learning (ICML-10). pp. 975–982. Omnipress, Haifa, Israel (June 2010)
9. Signoretto, M., De Lathauwer, L., Suykens, J.: Nuclear norms for tensors and their use for convex multilinear estimation. Tech. rep., ESAT-SISTA, K.U.Leuven (2010)
10. Xavier de Souza, S., Suykens, J., Vandewalle, J., Bollé, D.: Coupled simulated annealing. IEEE Trans. Syst. Man Cybern. Part B Cybern. 40(2), 320–335 (2010)
11. Suykens, J.A.K., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Lett. 9(3), 293–300 (1999)
12. Tao, D., Li, X., Wu, X., Hu, W., Maybank, S.J.: Supervised tensor learning. Knowl. Inf. Syst. 13, 1–42 (September 2007)
13. Temko, A., Thomas, E., Boylan, G., Marnane, W., Lightbody, G.: An svm-based system and its performance for detection of seizures in neonates. In: IEEE Eng. Med. Biol. Mag., 2009. EMBC 2009. Annual International Conference of the IEEE. pp. 2643 –2646 (Sept 2009)
14. Thomas, E., Temko, A., Lightbody, G., Marnane, W., Boylan, G.: A comparison of generative and discriminative approaches in automated neonatal seizure detection. In: Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on. pp. 181 –186 (Aug 2009)