Semi-Supervised One-Class Transfer Learning For Heart Rate Based Epileptic Seizure Detection

Thomas De Cooman[1,2], Carolina Varon[1,2], Anouk Van de Vel [3], Berten Ceulemans [3,4], Lieven Lagae [4,5], Sabine Van Huffel [1,2]

[1] Department of Electrical Engineering (ESAT), STADIUS, KU Leuven, Belgium
[2] imec, Leuven, Belgium
[3] Department of Paediatric Neurology, Antwerp University Hospital, University of Antwerp, Belgium
[4] Rehabilitation Center for Children and Youth Pulderbos, Belgium
[5] Department of Child Neurology, University Hospital Leuven, KU Leuven, Belgium

Abstract

Automated epileptic seizure detection in a home environment has been a topic of great interest during the last decade. Normally patient-independent heart rate based seizure detection algorithms are used in practice to avoid the necessity of patient-specific data. They, however, lead to mediocre performance due to the large inter-patient heart rate variability. Therefore these algorithms should be adapted to each patient in an efficient way. In this study, a patient-specific algorithm is constructed with only 1 night of not-annotated patient-specific data by using a transfer learning approach. The algorithm was evaluated on 8 pediatric patients with 25 strong nocturnal convulsive seizures. By using only 1 night of patient-specific data, the false alarm rate dropped by a factor of 4 compared to the patient-independent algorithm, leading to on average 0.76 false alarms per night and 88% sensitivity. The results show that the proposed method can quickly adapt to patient characteristics without the requirement of seizure annotations.

1. Introduction

Epilepsy is a neurological disorder that affects around 1% of the people worldwide. During the last decade, increased interest has gone into detecting epileptic seizures in a home environment [1]. One of the options is to use the heart rate for detecting seizures, because most strong convulsive seizures have ictal heart rate changes. In most cases, these changes can be seen as strong heart rate increases caused by sympathic activations [2].

Most algorithms from the literature make one patient-independent (P-I) algorithm for all patients, meaning they do not need any (annotated/seizure) data from the patient they are testing in order to be used [2, 3]. That way, these algorithms can be directly used in practice. However, the performance of these heart rate based P-I algorithms showed to be too low for usage in practice [2].

The main reason for this low performance is the strong patient-dependency of the heart rate features. Heart rate characteristics (both inter-ictal and ictal) vary strongly between patients, and this variance should be removed in order to get better results. However, in order to make a robust patient-specific algorithm using only patient-specific data, at least 5-10 seizures per patient should be acquired [4]. This takes too much time in practice for patients with a low seizure frequency.

It is however possible to adapt to patient-specific characteristics in a more efficient way. Transfer learning is a machine learning technique that allows to solve a new problem (the patient-specific algorithm) with a limited amount of training data by using the known solution of a reference problem (the known patient-independent algorithm) [5]. In this paper, only 1 night of patient-specific data is used in order to train a patient-specific classifier using the transfer learning approach. In order to avoid the necessity of seizure occurrence in this limited amount of training data, a semi-supervised procedure is proposed in this paper.

2. Data

The proposed method is evaluated on data recorded in Pulderbos Revalidation Center for Children and Youth. In total 8 pediatric epilepsy patients with 25 strong convulsive seizures (tonic-clonic and hypermotor seizures with a duration longer than 10s) were monitored during the night for 428.3 hours (37 nights). Seizures were annotated by expert neurologists using the golden standard video-EEG monitoring. The electrocardiogram (ECG) was acquired using lead II with 250 Hz sampling frequency.

3. Methods

3.1. Preprocessing

Only the ECG is used in this paper for automated real-time epileptic seizure detection. The preprocessing steps of this algorithm are based on those discussed in [2], and they are summarized as follows.

The seizure detection algorithm is fully based on heart rate analysis. The R peaks in the ECG are detected in real-time by using an R peak detection algorithm based on dynamic thresholding on the derivative signal.

In a next step, the algorithm automatically detects whether a strong heart rate increase is occurring. This is done by inspecting some features from the strength of the increase, like the slope of the tachogram and the absolute amount of heart rate increase.

Feature extraction occurs whenever such a strong heart rate increase is detected. A previous study showed that using only two features from this heart rate increase already gives good results for patient-specific nocturnal seizure detection: the peak heart rate and the maximal slope of the heart rate during this increase [6]. These two features are extracted and used for further analysis in this study.

3.2. Classification

The extracted features are then classified by a support vector machine (SVM) classifier in order to indicate whether a heart rate increase is caused by a seizure or not. In order to avoid the necessity of annotated patient-specific data, a one-class SVM (OC-SVM) classifier is used here.

An option to implement an unsupervised SVM classifier is by solving the following minimization problem [7]:

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 - \rho + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i \tag{1}$$

s.t.

$$w^T\varphi(x_i) \geq \rho - \xi_i, \qquad \xi_i \geq 0, \qquad i = 1, ..., N \tag{2}$$

with $\varphi(.)$ the applied kernel function and ξ_i the errors of the classifier on data point x_i. However, because the P-S classifier has to be learned with limited data, the normal OC-SVM approach might not lead to a sufficient robust classifier. Therefore, a transfer learning method is applied on the OC-SVM classifier. Transfer learning allows to train a new classifier by using a reference (already trained) classifier which solves a similar problem [5]. In this case, the P-I classifier can be used as a reference, as most of the knowledge contained in this classifier is still applicable per patient (for example, a higher peak heart rate will more likely lead to a seizure, independently of the patient).

The implementation of it in SVM is inspired by the binary method discussed in [5]. This paper proposes a transfer learning adaptation for OC-SVM (called TL OC-SVM from now on). The new minimization problem then becomes:

$$\min_{w,b,\xi} C\frac{1}{2}||w - w'||^2 - \rho + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i \tag{3}$$

s.t.

$$w^T\varphi(x_i) \geq \rho - \xi_i, \qquad \xi_i \geq 0, \qquad i = 1, ..., N \tag{4}$$

with w' is the weight vector obtained from the already trained P-I classifier, defined as

$$w' = \sum_{j=1}^{M}\alpha'_j y'_j x'_j \tag{5}$$

with x'_j data points used in training the P-I classifier with corresponding annotations y'_j and M support vector values α'_j. The P-I classifier is obtained by solving the normal binary SVM problem and is trained using a leave-one-patient-out approach on the evaluated dataset as in [2].

The dual problem is solved in practice, which is formulated as

$$\min_{\alpha} \frac{2C-1}{2C^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j K(x_i, x_j) + \sum_{i=1}^{N}\sum_{k=1}^{M}\alpha_i\alpha'_k y'_k K(x_i, x'_k) \tag{6}$$

s.t.

$$0 \leq \alpha_i \leq \frac{1}{\nu N}, \qquad \sum_{i=1}^{N}\alpha_i = 1 \tag{7}$$

with K the kernel function. The kernel function and kernel parameters of the P-I and P-S need to be the same here.

There are only two remaining parameters that need to be tuned beforehand: C and ν. Hyperparameter C allows to indicate how similar the P-S and P-I classifiers should be. It can be tuned automatically apart from the minimization problem using a leave-one-patient-out procedure.

The ν parameter is the upper bound on the percentage of outliers in the training set [7]. However, as the outliers should be the seizures, this percentage will strongly vary between patients and nights, so fixing a value for all patients would lead to mediocre performance. Hence, we need to make an estimation of ν based on the available patient training data. Here, ν is set as

$$\nu = \frac{\#seiz_{est}}{N} + min(\frac{FP_{allow}}{N}, \frac{\#alarms_{pi} - \#seiz_{est}}{N}). \tag{8}$$

The first term indicates the true percentage of outliers. We assume the patient to give minimal feedback to the system by indicating how many severe seizures he or she had in the training set ($\#seiz_{est}$). The second term can be seen as a safety term. The first term with the minimum function indicates the amount of false alarms FP_{allow} that is allowed in the training set (which can be expressed in false alarms/h). The second term within the minimum function indicates the amount of estimated false alarms ($\#alarms_{pi} - \#seiz_{est}$) caused by the P-I algorithm on the training set. As the P-I classifier is trained so that it results in a high sensitivity but also a high false alarm rate (FAR), the P-S classifier should not lead to more false alarms than the P-I classifier. This safety term thus allows to balance between sensitivity and FAR, and can remove potential influence of a wrong estimation of $\#seiz_{est}$.

In the final stage of the algorithm, an alarm of this P-S algorithm will only cause an alarm if it is also caused by the initial P-I algorithm. This is done in order to avoid too many false alarms if the patient did not show much activity (both epileptic and non-epileptic) during a night.

4. Results & discussion

4.1. Simulation details

The reference P-I classifier is trained using a leave-one-patient-out approach [2]. The P-S classifier is trained by using only one night of data per patient. For each night, the classifier is trained and evaluated by using a leave-one-sample-out approach. The C parameter is optimized for each patient by choosing the value that leads to the best F1 score on the other patients. The ν parameter is automatically set according to (8), assuming for now that $\#seiz_{est}$ is estimated correctly, and FP_{allow} is fixed to 0.2 false positives/h (FP/h) for all patients.

4.2. Results

The P-I seizure detection algorithm results in 92% sensitivity and 3.16FP/night (1 night= 8 hours). If the P-S algorithm is trained by using only 1 night of data, the sensitivity drops slightly to 88%, but the FAR drops to 0.77 FP/night (more than 4 times less false alarms compared to the P-I algorithm). Figure 1 shows the FAR per patient for both approaches. It can be seen that for the three patients with the highest FAR in the P-I approach, the FAR drops strongly towards much lower values. For the ones with already a low FAR, the FARs do not change much due to the fact that these patients typically have very few nocturnal heart rate increases (and thus few data points) and most of these false alarms are caused by R peak detection errors due to too strong noise in the ECG. The inter-patient FAR variability decreases strongly using this P-S approach.

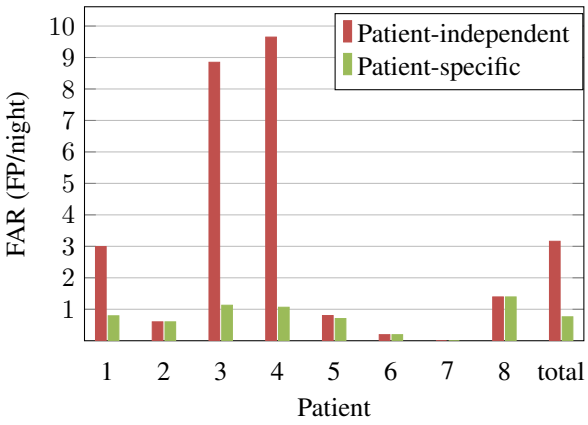

Figure 1. Average FAR per patient and in total for the P-I and P-S seizure detection algorithm.

4.3. Discussion

The added value of transfer learning in TL OC-SVM compared to normal OC-SVM is two-fold. First of all, transfer learning allows to train a robust classifier by using the reference P-I classifier. That way, there is no need for weeks of data (or 10 seizures) in order to get a robust classifier. An extra added value is that by incorporating a binary reference classifier, it is possible to indicate to the OC-SVM the locations of the outliers. Normally, OC-SVM will select these outliers fully based on the minimization problem, but this choice can sometimes be inaccurate without adding extra information. This is illustrated in figure 2, which shows the classifier boundaries for the reference P-I classifier, the P-S TL OC-SVM and normal OC-SVM classifiers, all trained with the same kernel and training data. The normal OC-SVM (without transfer learning) tries to minimize the cost function without any prior knowledge, leading to a poor classifier. However, by incorporating the information from the P-I classifier (f.e. larger peak heart rate and maximal heart rate slope tend to lead to increased possibility of seizures), TL OC-SVM leads to a much better solution, incorporating this information and adjusting the boundary to the available P-S data points. The required percentage of outliers are then found in the area where it makes most sense based on the information of the binary reference classifier. This way, the downside of normal OC-SVM (the correct identification of the true outliers) is removed, and no annotations need to be made in the new P-S data points in order to find them during training. The only input that is still required per patient, is the estimated number of seizures $\#seiz_{est}$ in the training set in order to define the ν parameter.

Despite the fact that the FAR dropped strongly with the proposed method, the FAR is still too high for usage in practice. Most of these false alarms (54% of all false

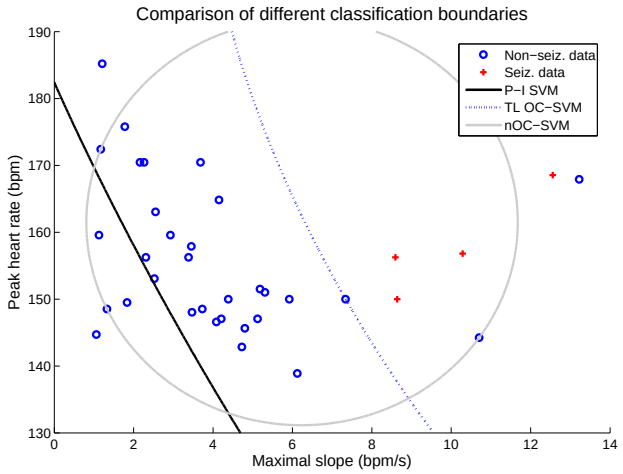

Figure 2. Example of classifier boundaries for the reference P-I classifier (PI-SVM), the P-S OC-SVM with transfer learning (TL OC-SVM) and P-S OC-SVM without transfer learning (nOC-SVM) on 1 night data.

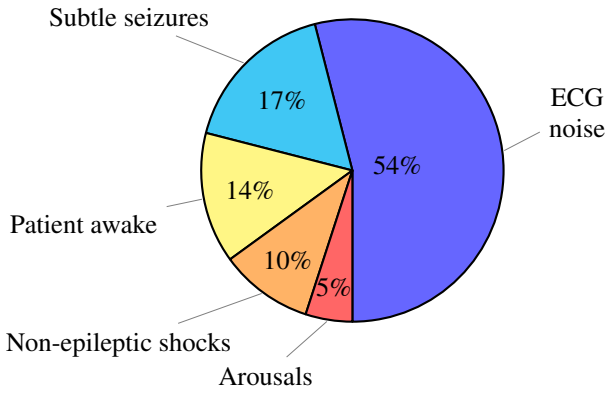

Figure 3. Causes for false alarms in the P-S algorithm.

alarms) are however caused by noisy ECG (see figure 3). These false alarms are caused by R peak detections during strong motion artifacts or when the patient or nurse is adjusting the ECG electrodes. These alarms could be avoided by better noise removal techniques and better sensor settings. In 14% of the cases, the patient was already fully awake and out of the bed, in which case the device could be turned off in practice. Also subtle epileptic activity caused 17% of the false alarms, which in some cases can not be seen as fully unwanted alarms.

5. Conclusion

The proposed semi-supervised OC-SVM classifier approach incorporating transfer learning allows to detect seizures more accurately by using only one night of not-annotated patient-specific data. Future work will focus on automatically estimating the ν parameter based on the data distribution and further evaluate the effect of the learning procedure on long-term data.

Acknowledgements

Research supported by imec: Strategic Funding 2017, ICON: HBC.2016.0167 SeizeIT; European Research Council. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (n 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. Thomas De Cooman is supported by an FWO SBO PhD grant. Carolina Varon is a postdoctoral fellow of the Research Foundation-Flanders (FWO).

References

[1] Van de Vel A, Cuppens K, Bonroy B, Milosevic M, Jansen K, Van Huffel S, Vanrumste B, Cras P, Lagae L, Ceulemans B. Non-eeg seizure detection systems and potential sudep prevention: State of the art: Review and update. Seizure 2016;41:141–153.

[2] De Cooman T, Varon C, Hunyadi B, Van Paesschen W, Lagae L, Van Huffel S. Online automated seizure detection in temporal lobe epilepsy patients using single-lead ecg. International Journal of Neural Systems 2017;1750022.

[3] Osorio I. Automated seizure detection using EKG. International Journal of Neural Systems 2014;24(02):1450001.

[4] Cogan D, Heydarzadeh M, Nourani M. Personalization of noneeg-based seizure detection systems. In Engineering in Medicine and Biology Society, 2016 IEEE 38th Annual International Conference of the. IEEE, 2016; 6349–6352.

[5] Yang J, Yan R, Hauptmann AG. Adapting svm classifiers to data with shifted distributions. In Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007). IEEE, 2007; 69–76.

[6] De Cooman T, Van de Vel A, Ceulemans B, Lagae L, Vanrumste B, Van Huffel S. Online detection of tonic-clonic seizures in pediatric patients using ecg and low-complexity incremental novelty detection. In Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE. IEEE, 2015; 5597–5600.

[7] Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC. Estimating the support of a high-dimensional distribution. Neural computation 2001; 13(7):1443–1471.

Address for correspondence:

Thomas De Cooman
Kasteelpark Arenberg 10 box 2446
3001 Leuven, Belgium
thomas.decooman@esat.kuleuven.be