Detecting rare events using extreme value statistics applied to epileptic convulsions in children

Stijn Luca[a,*], Peter Karsmakers[a,b], Kris Cuppens[a,b], Tom Croonenborghs[c], Anouk Van de Vel[d], Berten Ceulemans[d,e], Lieven Lagae[e,f], Sabine Van Huffel[a,b], Bart Vanrumste[a,b]

[a] *Department of Electrical Engineering, KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium*
[b] *iMinds Future Health Department, KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium*
[c] *Computer Science Department, KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium*
[d] *University Hospital of Antwerp, Wilrijkstraat 10 2650 Edegem, Belgium*
[e] *Epilepsy Centre for Children and Youth Pulderbos, Reebergenlaan 4, 2242 Zandhoven, Belgium*
[f] *University Hospital Leuven, Herestraat 49, 3000 Leuven, Belgium*

Abstract

Objective: Nocturnal home monitoring of epileptic children is often not feasible due to the cumbersome manner of seizure detection with the standard method of video electroencephalography monitoring. The goal of this paper is to propose a method for hypermotor seizure detection based on accelerometers that are attached to the extremities.
Methods: Supervised methods that are commonly used in literature need annotation of data and hence require expert (neurologist) interaction resulting in a substantial cost. In this paper an unsupervised method is proposed that uses extreme value statistics and seizure detection based on a model of normal behavior that is estimated using all recorded and unlabeled data. In this way the expensive interaction can be avoided.
Results: When applying this method to a labeled dataset, acquired from 7 patients, all hypermotor seizures are detected in 5 of the 7 patients with an average positive predictive value (PPV) of 53%. For evaluating the performance on an unlabeled dataset, seizure events are presented to the system as normal movement events. Since hypermotor seizures are rare compared to normal movements, the very few abnormal events have a negligible effect on the quality of the model. In this way, it was possible to evaluate the system for 3 of the 7 patients when 3% of the training set was composed of seizure events. This resulted in sensitivity scores of 80%, 22% and 90% and a PPV of 89%, 21% and 44% respectively. These scores are comparable with a state-of-the-art supervised machine learning based approach which requires a labeled dataset.
Conclusions: A person-dependent epileptic seizure detection method has been designed that requires little human interaction. In contrast to traditional machine learning approaches, the imbalance of the dataset does not cause substantial difficulties.

Keywords: extreme value theory, unsupervised, unbalanced data, accelerometers, hypermotor seizures, epileptic.

1. Introduction

Video electroencephalography (EEG) monitoring, the gold standard for detecting epileptic seizures [1], is not suitable for home monitoring since electrodes are uncomfortable, need to be in contact with skin at all times, and hamper the patient's sleep. However, there is a need for a long-term home monitoring system for early intervention and prevention of seizure related side effects and giving the neurologist valuable information about the progress of the disease [2]. Although most of the patients can be treated with medication or surgery, in 25% of patients seizures cannot be fully controlled by medication or surgery [3]. Here long-term home monitoring can play a crucial role.

The study presented in this article focuses on hypermotor seizures that are marked by a strong and uncontrolled movement of the arms and legs and can last from a couple of seconds to some minutes. The group of patients that is studied consists of 7 patients with hypermotor seizures, all between the age of 5 and 16 years. Due to the heavy movement, patients can injure themselves during the seizure, which increases the need for an alarm system [4]. The main focus of this article is on pediatric patients, as epilepsy has a higher prevalence in children (as it has in the elderly) than in adults [4]. Furthermore, in children the disorder can still evolve over time.

A home-monitoring system permits a better observation of the patient so that parents or caregivers can be warned to administer medication, reposition the patient or comfort him during or after the seizure. Recently, research has been carried out to improve automatic registration by making use of less intrusive sensors like accelerometers attached to the extremities [2]. The goal is then to (automatically) build a classification model that uses the sensor

*Corresponding author

Preprint submitted at Artificial Intelligence in Medicine on 04/04/2013

data to indicate whether a certain movement is due to an epileptic convulsion.

In the literature, supervised learning algorithms are commonly used (e.g. support vector machines (SVMs) [5–7]) to obtain such classification models and hence annotated examples are required. This annotation of data (every movement needs to be manually evaluated) requires expert (neurologist) interaction and results therefore in a substantial cost. Furthermore, to be able to label the acquired data, EEG has to be recorded, which is uncomfortable for the patient.

In this article the use of an unsupervised approach for estimating seizure detection models is proposed that is based on extreme value theory (EVT)[8]. The method proceeds in three main steps:

(i) The unlabeled sensor data is used to build a model of "normal behavior".
(ii) A threshold border is automatically set in the tails of this model by using EVT.
(iii) An observed movement that exceeds this threshold is classified as a rare event and may be an epileptic convulsion.

The rationale for this approach is two-fold: 1) movements during an epileptic convulsion have characteristics that differ from normal nocturnal movement events and 2) hypermotor seizures are rare so that only a small fraction of seizure events is included when building a model of normality, resulting in a limited effect. The advantage of this method is that it does not require any annotation of data, yet state-of-the-art classification scores are obtained that are comparable to those obtained with a supervised learning algorithm. Furthermore, it does allow to build a patient-specific seizure detector more easily.

The outline of the article is as follows. Section 2 gives an overview of related research. In section 3 a comprehensive review is given of how to use EVT for detecting novelties in general. The dataset and experimental setup are described in section 4. Experimental results can be found in section 5 and are discussed in section 6. Finally, we provide some conclusions and possible directions for future work in section 7.

2. Related research

A lot of work on epileptic seizure detection using EEG signals exists, see e.g. [2] for an overview. In general, seizure detection can be accomplished by classification algorithms that are applied to so-called feature vectors that are extracted from the original sensor signals and capture the main information. In [5] classifications of EEG signals using convolutional networks and SVMs are compared. Other approaches as discrete wavelet transform and entropy are used in e.g. [9, 10].

Given the impracticality of EEG data for home monitoring, accelerometer data are used in this article and hence the focus of this discussion on related research is on epileptic seizure detection using accelerometers. Besides these types of signals, other signals have been used in the past as well. A recent overview can be found in [11].

In 2011 the first commercially available wrist-watch-like detectors for tonic-clonic seizures, the EpiLert and the SmartWatch, were presented, and the first clinical studies were carried out [12, 13]. Kramer et al. [12] included 31 patients in their research. During 1692 hours of monitoring, 20 of 22 seizures were detected with 8 false detections. The detector was worn during night and day. In the research of Lockman et al. [13], 40 patients were included and 7 of 8 tonic-clonic seizures were detected generating 204 non-seizure detections (of which one occurred during the night). Other researchers detected different types of epileptic seizures by means of multiple accelerometers attached to the extremities. Nijsen et al. [14] used 5 accelerometers (arms, legs and sternum) to detect myoclonic seizures. Although the detection rate was good ($> 80\%$), a lot of false positives were detected with a positive predictive value (PPV) that was smaller than 20%. Cuppens et al. [15] used 4 accelerometers (wrists and ankle) to detect frontal lobe seizures with a high sensitivity ($> 90\%$) and specificity ($> 80\%$) in 3 patients. Jallon et al. [16] used 2 accelerometers for the detection of non-specified seizures with a high sensitivity ($> 88\%$) and a PPV of 75% and 55% respectively in two patients. Conradsen et al. [17] used a multi-modal approach using acceleration, angular velocity and surface electromyography on patients with tonic, clonic and versive seizures. Their multimodal approach gave a good sensitivity ($> 90\%$) and specificity (100%).

The weak point of these approaches, however, is that they all require annotation of data. In this article, it is shown that an EVT-based approach can be used to automatically detect hypermotor seizures without the need for an expensive annotation. The use of multivariate extrema for novelty detection in multivariate data was first approached in [18] and [19]. Extrema are used in the classical way as multivariate extrema that occur in one or more dimensions. Thus, novelties are detected by determining events of extremely large or small magnitude in one or more dimensions. Recently Clifton et al. [20] developed a technique that is generally applicable and has the additional advantage of translating a multivariate model of normal behavior to an univariate model of minimal densities, using classical EVT. They applied the approach to a vital-sign monitoring problem.

In this article the approach is used for the first time to develop a detection system for epileptic seizures in an unsupervised manner. The focus is on hypermotor seizures that often involve violent movements. Therefore, the seizure data are suspected to be situated in the tails of the model of normal behavior that can be modelled using EVT.

3. Using EVT to model rare events

In this section the methodology that was recently proposed in [20] is reviewed. In contrast to the commonly used supervised approach, novelty detection using EVT is based on a model of normal behavior. From this point of view one supposes that a d-dimensional dataset $\mathcal{D}$ is given, consisting of data points assumed to be drawn independently from some population X that is distributed according to an unknown probability distribution function (PDF) $y = f(x)$. Novelty detection addresses the question whether a given sample $S = \{x_1, \ldots, x_k\} \subset \mathbb{R}^d$ is drawn from the distribution of X or not.

To obtain an estimate for $f(x)$, a multivariate kernel density estimation (KDE) is performed [21]. Such KDE can be viewed as a weighted sum of Gaussian kernels centered at the data points $x_i \in \mathbb{R}^d (1 \leq i \leq n)$ corresponding to normal behavior:

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(2\pi)^{d/2} |H|^{1/2}} \, e^{-\frac{1}{2}\left((x-x_i)' H^{-1} (x-x_i)\right)}. \quad (1)$$

Here the accent defines vector transposition and $H \in \mathbb{R}^{d \times d}$ denotes a bandwidth matrix which is symmetric and positive-definite. Because our interest is in discerning the general shape of the density, the choice of kernel is not crucial [21]. A Gaussian kernel has the advantage of making the estimation smooth, having all derivatives of all orders.

On the other hand, the choice of the bandwidth matrix H in (1) is crucial in the performance of the estimation. The diagonal elements of H can be viewed as smoothing parameters in the different dimensions, whereas the off-diagonal elements induce smoothing in an oblique direction and can be used to adjust for correlations between the features. If the values are too large, the estimation is oversmoothed and it fails to detect local variations. On the other hand, a too small choice of the elements of H corresponds to an overfitting that induces too much variation on the estimation. See for instance [21] for a more complete elaboration on the subject and [22] for an implementation of the techniques in R [23].

Algorithms for the estimation of H have a high computational cost as the size of the dataset grows [21, 22]. We prefer to apply the following simple and heuristic approach for choosing H, also known as the generalization of Scott's rule of thumb [24]:

$$H = n^{-2/(d+4)} \hat{\Sigma},$$

where $\hat{\Sigma}$ denotes the covariance matrix of the observed data.

Based on this model of normal behavior EVT is applied to the samples S. EVT is used for describing those extremes of samples $S = \{x_1, \ldots, x_k\}$ of length k that have a minimal density value with respect to the KDE $f(x)$. In particular, one considers for each sample $S = \{x_1, x_2, \ldots, x_k\}$ the corresponding sample of density values in $\mathrm{Im}(f) \subset \mathbb{R}$:

$$(y_1, y_2, \ldots, y_k) = (f(x_1), f(x_2), \ldots, f(x_k)).$$

The distribution of Y is strongly related to that of X. Using EVT, one can determine the distribution of minima of samples from Y:

$$y^e := \min_{1 \leq i \leq k} \{f(x_i)\}. \quad (2)$$

The distribution of the population Y^e tells us where the minimal densities of samples S of length k are expected to be located.

According to EVT, the distribution for minimal densities y^e, with k large enough, can be approximated by the Gumbel, Fréchet or Weibull distribution [8]. Fréchet and Gumbel distributions can be used to model minima that are not bounded from below. The Weibull distribution corresponds to a model for minima that possesses a lower bound. The corresponding PDF is given by:

$$w(y) = \frac{a}{b} \left(\frac{y+c}{b} \right)^{a-1} \exp\left(-\left[\frac{y+c}{b} \right]^a \right)$$

where the parameters $a > 0, b$ and c respectively determine shape, scale and position. For minimal densities the location parameter can be set to zero, $c = 0$. The parameters a and b can be found using maximum likelihood estimates.

A sample S of length k is said to be *rare* with respect to f if its minimal density y^e is abnormal with respect to the estimated Weibull distribution. Generally a rare sample S can be situated in the tails of the distribution of f or between the modes of f.

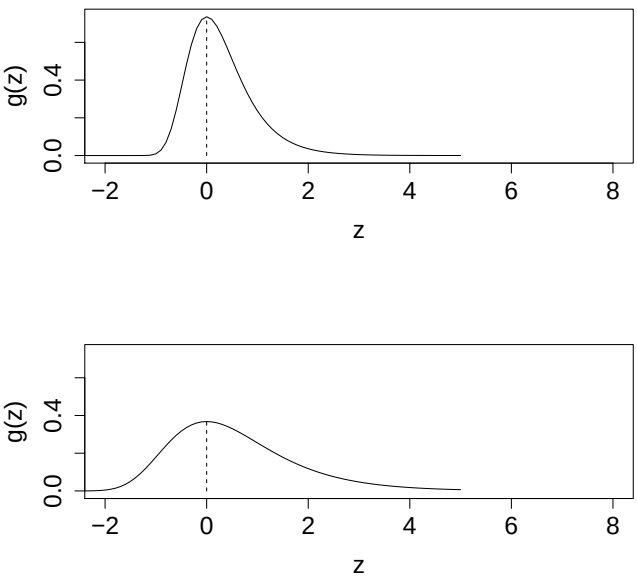

Figure 1: Gumbel density functions centered at $\beta = 0$ and scale parameter $\alpha = 0.5, 1$ from above to below.

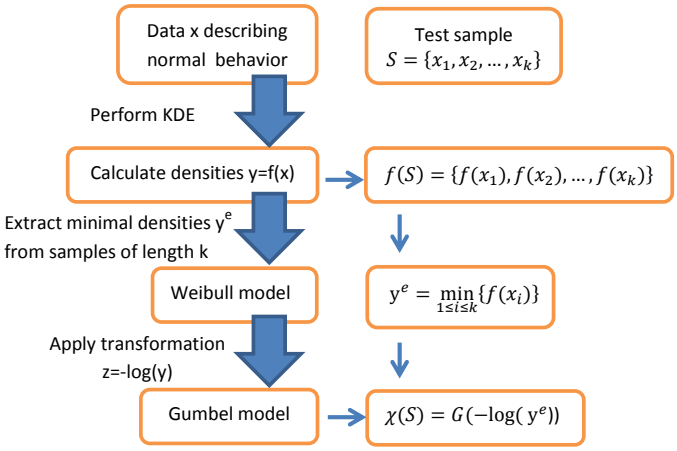

Figure 2: Schematic overview of algorithm for novelty detection using EVT

In our study rare samples have an abnormally small minimal density with respect to the Weibull distribution. These are all pushed into a small interval close to zero. One can use the transformation $z = -\log(y)$ to overcome this skewness. The short tail near zero of the Weibull distribution is then stretched out as the right tail of a Gumbel distribution for maxima, where extremes can be shown in a clearer way. The corresponding cumulative distribution function (cdf) reads:

$$G(z) = \exp(-\exp(-\frac{z-\beta}{\alpha})) \tag{3}$$

where $\alpha = \frac{1}{a} > 0$ and $\beta = -\ln b$. Figure 1 shows Gumbel density functions.

A *n*ovelty score of a sample S can be defined as:

$$\chi(S) := G(-\log(y^e)) \tag{4}$$

where y^e is defined as in (2). The novelty score of a sample increases when the probability of choosing a sample with a higher minimal density increases. One can now threshold these novelties and consider a sample as rare if its corresponding novelty score is above some threshold. From a probabilistic point of view, a typical choice of threshold is e.g. 95%. A schematic overview of the methodology is given in figure 2.

4. Dataset and experimental setup

4.1. Dataset description

In this article the proposed approach is applied to nocturnal detection of hypermotor seizures. The dataset is collected in collaboration with the Pulderbos rehabilitation Center for Children and Youngsters in Zandhoven (Pulderbos), Belgium.

Table 1: Overview of data available in the dataset

patient number	nights of monitoring	hypermotor seizures	normal movements
pat 1	1	2	117
pat 2	2	9	287
pat 3	2	2	439
pat 4	1	2	239
pat 5	5	26	784
pat 6	2	7	381
pat 7	2	3	468
total	*15*	*51*	*2715*

The group of patients that is studied consists of 7 patients with hypermotor seizures, all between the age of 5 and 16 years. The hypermotor seizures are marked by a strong and uncontrolled movement of the arms and legs (e.g. pedaling movement) that can last from a couple of seconds to some minutes. Due to the heavy movement, the patient can injure himself during the seizure, which increases the need for an alarm system. The focus is primarily on pediatric patients, as epilepsy has a higher prevalence in children than in adults. Furthermore, in children the disorder can still evolve over time.

The recordings took place during several nights no more than 2 or 3 months apart, except for patient 5 where the period between successive nights lasted from 5 up to 14 months. An overview of the dataset is given in Table 1. All patients were between the age of 5 and 16 years old, with a mean age of 9 years and a standard deviation of 3 years.

The collected data contain multimodal information: EEG, EMG, EOG, ECG, video, audio and acceleration are registered. The EEG, video and audio signals are used by the neurologists to label the data, which is considered as the golden standard. In this article, seizure detection is based on 12 channel acceleration data, collected by four 3D acceleration sensors that are attached to the extremities and sampled at 250 Hz.

4.2. Preprocessing and experimental setup

The preprocessing step follows the same procedure as in [15, 25, 26] and is briefly reviewed here. Computations were performed using Matlab [27].

The raw accelerometer data are first filtered by a digital low-pass filter with a cut-off frequency of 47 Hz in order to avoid aliasing effect when downsampling from 250 Hz to 125 Hz. This is justified since data analysis has shown that no meaningful frequency content is present above 50 Hz in the considered signals. As a consequence the undesirable electrical network frequency (50 Hz) is also removed in the signal. In order to avoid phase distortion, the data are processed in a forward and reverse direction to obtain a zero-phase filtering.

The gravitational force is present in the accelerometer data as a result of DC offsets in each of the orthogonal axes. These offsets depend on the accelerometer position compared to the gravitational vector (which is always vertically oriented). Since patients are assumed to lie in different positions (e.g. on the back or side) such offsets might give undesirable signal variations and are therefore filtered off using a zero-phase digital high pass filter with cut-off frequency 0.2 Hz.

Given the filtered data, *movement events* are detected when the data of at least one arm or leg show a minimal amount of energy. The following procedure is adopted:

(i) For each limb the L_2 norm is computed of the vector defined by the 3D data given by each individual accelerometer,
(ii) Standard deviations are calculated for each of the 4 resulting signals using a sliding window with a width of 2 seconds with 50% overlap.
(iii) A movement event is detected when the standard deviation exceeds a threshold value.

The sliding window of 2 seconds in (ii) is chosen since:

(a) 2s is short enough to not smooth out meaningful movement events, which are expected to be longer than 2s [25].
(b) 2s is long enough to be robust against measurement noise.

Note that two thresholds are used in (iii), one for the arms and another for the legs. These thresholds were determined by a simulation during which a simulator lies in bed during five minutes, after which he makes small finger and toe movements. The thresholds were set in such a way that these movements were the smallest motions to be detected. This resulted in a threshold of 10 mg for the arms and a threshold of 5 mg for the legs, where g represents the gravitational acceleration which is $9.81 \frac{m}{s^2}$.

Analysis of our annotated data revealed that hypermotor seizure related movement events were less than 30s apart. This result was confirmed by the authors in [4] where it was concluded that hypermotor seizures can occur with interruptions that are typically between 20s and 40s in duration. Therefore, in a final preprocessing step, movement events that occur less than 30s apart are clustered as the duration of the epileptic activity considered here is typically long and sometimes a patient stops moving in the middle of a hypermotor seizure. If the movement events are not clustered, then a single epileptic event can wrongly be considered as two epileptic events.

This filtering causes a substantial reduction in processing time since less data need to be inspected for abnormalities. Typically, 80% to 90% of the data is discarded as it contains no movement.

Each movement event is represented by a specific set of features (collected in a tuple). The extracted features are the same as the ones that are computed in Cuppens et al. [25, 26]. They are common features that are widely used in accelerometer detection [28] and epilepsy research [29, 30]. However, in this article the features are not computed on the complete movement event (except for movement event length) but instead are extracted from 50% overlapping *sliding windows* containing 125 samples, which is the minimum number of samples of a movement event. Hence, in order to have at least a single feature vector per event a length of 125 is set. Remark that our implementation of the movement detector (using the sliding window principle as described above) selects movement events with lengths that are a multiple of 125 samples. As a result a *feature vector set* is obtained per movement event. This allows evaluation of the distribution of features.

Based on fixed-length subsets (of length k) of the feature sets EVT is used to judge whether movement events are related to normal behavior or to rare events including hypermotor seizure.

The following experimental setups are used to validate the proposed detection method:

(i) *Comparison with a supervised learning algorithm:* In order to compare with a supervised learning algorithm, i.e. SVMs [31], fixed partitions of the data in training and test sets are generated. The training sets contained two third and test sets one third of the data. A sufficient amount of seizures is needed to be able to train and validate an SVM model. This is the reason why in the comparison with SVMs in the following section, only results for patients 2, 5 and 6 are given.
(ii) *Clean versus mixed training:* The ultimate goal is to obtain a model of normal behavior in an unsupervised manner. In the considered application this is feasible since we expect that most of the time the observed data will be normal. Abnormal, seizure related, behavior is assumed to be rare. A model of normal behavior is then learned based on *all* available (unannotated) data. This means that some seizure related data might wrongly be assumed as normal. To assess the effect of having seizure related data in the dataset for model estimation two experiments are performed. In a first experiment, indicated by *clean training*, a model estimation is performed on training data that do not contain seizure related data. In this way normal behavior can be modeled accurately leading to high sensitivity scores. In a second experiment, indicated by *mixed training*, seizure related data are added to the training set. As can be expected this leads to lower accuracy and hence lower sensitivity scores. Therefore, the performance scores of the clean training experiment can be viewed as an upper bound to the model accuracy.

5. Results

In this section, our proposed approach is evaluated empirically. The use of EVT for novelty detection is implemented using the statistical software package R [23].

As explained in section 4 the movement data are first partitioned in movement events. From each movement event a set of a feature vectors is extracted. In order to be able to apply EVT, one chooses a feature vector $S = \{x_1, \ldots, x_k\}$ of length k from each movement event. This feature vector is composed of features that are extracted from windows of the event as is explained in more detail in section 4.2. Each x_i denotes a d-dimensional datapoint and represents one window of one movement event. The dataset of all such feature vectors is denoted as $\mathcal{D}$ and the dimension d equals the number of features that are used to represent each window.

Training and testing is done in a m-fold randomization, where $m = 10$ is a typical choice commonly found in literature, see e.g. [32, 33]. Reported results show the average over 10 runs with the corresponding standard deviations. The sample length k needs to be large enough for the Gumbel model to be a valid approximation for minimal densities, but too large values can be time-consuming. Results are reported for the choice $k = 50$, which is considered to be large enough [20]. A fixed threshold of $t = 95\%$ on the novelty scores is a typical choice from a probabilistic point of view and will work fine for most patients whose seizures are extreme with respect to the model.

Results are presented using three evaluation criteria that will be important in our discussion in the next section:

- the sensitivity, i.e. the proportion of actual epileptic convulsions which are correctly identified as such;
- the specificity, i.e. the proportion of normal movement events which are correctly identified;
- the PPV, i.e. the proportion of the seizures predicted by our system that are actual epileptic convulsions.

In our dataset, the ratio of the number of seizures to normal movements roughly does not exceed 3%. This number however varies slightly from one patient to another with a mean of 2% and a standard deviation of 1%. In order to give an indication of the performance measures that would be obtained if no label information is available, a novelty study using EVT is performed such that the ratio of the number of seizures to normal movements is 3%, 2% and 1% respectively in both the training and test set. In order to be able to perform such a study, one needs a sufficient number of seizures. This is the case for patient 2, 5 and 6. For patient 6 the number of normal events in the training data have to be reduced in order to obtain the desired proportions of 1%, 2% and 3% respectively. Performance scores are shown in table 2.

In the next section a comparison of these results with an SVM-classifier that is trained in a patient-specific way is discussed [25, 26]. Table 3 shows the results of this classifier using the same partitions in training and test set during the 10-fold cross-validation to allow for a consistent comparison. For this purpose the training sets contained two third and the test sets one third of the data. The SVM-classifier requires labeled data hence a fairer comparison can be achieved by using table 4 that shows the results of our approach when the training set only consists of normal movement events.

Table 2: Sensitivity (SS), specificity (SP) and PPV for patients 2, 5 and 6 when 3%, 2% and 1% seizures are added to the training data. Thresholds for the novelties are set at 95%. Mean and standard deviation (SD) are calculated over 10 runs in a 10-fold randomization.

EVT		SS		PPV		SP	
		mean	SD	mean	SD	mean	SD
3%	pat2	80.00	17.21	89.28	22.65	99.37	1.40
	pat5	22.22	14.81	20.57	13.21	96.97	1.09
	pat6	90.00	31.62	43.63	9.81	97.01	1.69
2%	pat2	93.33	14.05	81.95	21.00	99.06	1.34
	pat5	25.55	13.90	20.93	1.10	96.64	1.23
	pat6	90.00	31.62	42.00	10.34	96.77	1.84
1%	pat2	100.00	0.00	68.28	15.30	98.33	1.11
	pat5	25.55	12.88	20.92	11.36	96.41	1.23
	pat6	93.33	21.08	40.59	21.75	95.67	19.73

Table 3: Sensitivity (SS), specificity (SP) and PPV for patients 2, 5 and 6 using the SVM-classifier. Mean and standard deviation (SD) are calculated over 10 runs in a 10-fold randomization.

SVMs	SS		PPV		SP	
	mean	SD	mean	SD	mean	SD
pat2	97.14	6.02	85.43	11.12	98.65	1.10
pat5	39.58	31.81	41.18	15.81	95.34	3.45
pat6	84.00	30.98	94.44	8.33	99.76	0.38

Table 4: Sensitivity (SS), specificity (SP) and PPV for patients $1-7$, when no seizures are added to the training data. Thresholds for the novelties are set at 95%. Mean and standard deviation (SD) are calculated over 10 runs in a 10-fold randomization.

EVT *Study*	SS		PPV		SP	
	mean	SD	mean	SD	mean	SD
pat1	100.00	0.00	49.09	37.39	92.18	7.21
pat2	100.00	0.00	60.01	20.04	97.29	2.31
pat3	100.00	0.00	56.33	17.8	97.18	1.43
pat4	70.00	25.81	31.78	25.18	97.29	2.31
pat5	27.77	12.00	20.77	9.96	96.29	1.08
pat6	100.00	0.00	56.65	17.29	97.79	1.47
pat7	100.00	0.00	44.02	9.79	95.77	1.49

6. Discussion

The discussion of the results is organised by successively answering the following questions:

Q1 How does our unsupervised detection method performs? For this purpose, the performance is compared with a state-of-the-art supervised method.

Q2 What is the influence on the performance if the ratio of epileptic convulsions versus normal movements decreases in our training set?

Q3 How does our method performs when no seizure data is present in the training set?

Q4 What is the influence of the values for the parameter k and the novelty threshold and can these values be optimized per patient?

Q5 Is a real-time implementation possible for the proposed approach?

To answer the first question, we use the performance scores achieved by a novelty study using EVT such that the ratio of the number of seizures to normal movements is 3% (Table 2). A comparison with a state-of-the-art supervised method is obtained by using the results of table 3.

Not all seizures are detected when 3% of the training data consists of seizures but this is due to the relatively high percentage of seizure movements that are used in the training set. However, for patients 2 and 6 high sensitivity scores are achieved by our unsupervised method as is the case when using the SVM-classifier. For patient 5, the epileptic convulsions are more similar to the normal movement events. Since the seizures are less extreme, they are also less detectable inducing lower sensitivity and PPV-scores. The SVM-classifier yields better PPV-scores for patients 5 and 6.

Decreasing the ratio of epileptic convulsions versus normal movements in the training set from 3% to 2% or 1% results, as can be expected, in an increase in sensitivity since the model of normal behavior becomes more accurate (table 2). When the ratio of seizures versus normal movements is 1%, the sensitivity achieved by the SVM-approach is lower for patients 2 and 6. This can be of great value since hypermotor seizures often involve violent movements, which increases the need for high sensitivity scores. Hence our unsupervised method can easily be adapted to facilitate continuous learning where precision increases over time. This can be achieved by only incorporating new data if the running system indicates that it is a normal movement event. By doing so, the ratio of seizure data versus normal movement continuously decreases in the training set. Furthermore note that these results indicate the potential performance boost when a labeling for a subset of the data is available.

At the same time the PPV decreases as the number of seizures in the training data decreases. This can be explained mathematically by the fact that a smaller number of seizures in our training data results in a decreased mass in the tails of the distribution. Therefore minimal densities become smaller and novelty scores higher such that false positives have the tendency to occur more often. The SVM-approach yields slightly better PPV-scores when the ratio of seizures versus normal movements in our training set is 1%.

Question 3 is addressed by discussing the results in table 4. Since no seizure data need to be used in the training set, it becomes possible to evaluate our method on patients $1, 3, 4$ and 7 for whom only a small number of seizures is available in our dataset (Table 1). This is in contrast with a supervised method for which examples of both classes are needed in the training phase, making it impossible to validate the supervised SVM-approach on those patients. Moreover, since the SVM-classifier requires a labeled dataset a fairer comparison with our method is obtained using table 4 where the training set only consists of normal movement events.

Note that table 4 also gives an indication of the performance that would be obtained if more data were available. One expects that in comparison to table 2 the sensitivities will further increase and the values of PPV will further decrease. This seems not to be the case for patient 6. However as already indicated in section 5, the number of normal events in the training data had to be reduced for this patient in order to obtain the results in table 2. This results in a lower accuracy of the model and hence a lower PPV in comparison to table 4 where all normal events of the training set can be used.

In order to answer question 4, the influence of the value of the parameter k is discussed first. The previous results were obtained with a value of $k = 50$. As indicated before, the sample length k needs to be large enough because a low value can result in a poor Gumbel approximation leading to biases and unreliable performance scores. This is very natural because the probability of drawing enough samples from the tail decreases for small sample lengths. A too large choice of k can however lead to time-consuming calculations for estimation of Gumbel model parameters.

For $k = 50$, approximations during all runs of the 10-fold randomizations for all patients are satisfactory and the Gumbel model is valid. As an illustration, figure 3 shows some quantile-quantile (Q-Q) plots during 4 runs of the 10-fold randomization of patient 2. The y-axis shows

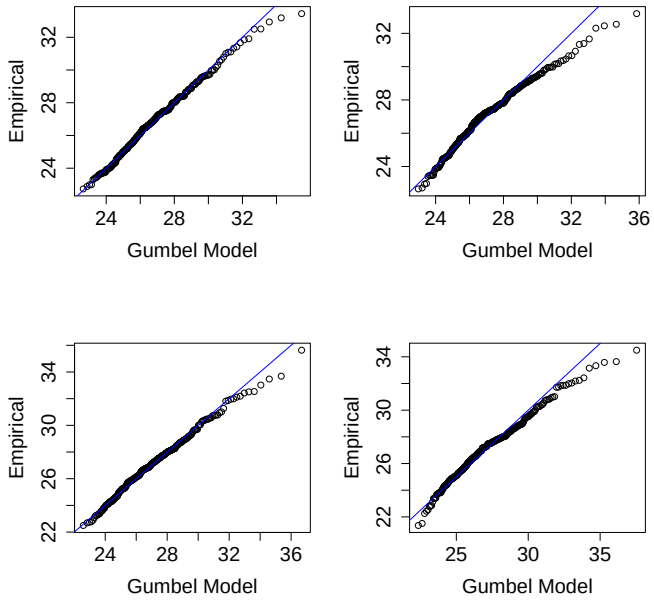

Figure 3: Q-Q plot of empirical quantiles versus Gumbel quantiles for patient 2 during run $2, 4, 6$ and 8 in a 10-fold randomization.

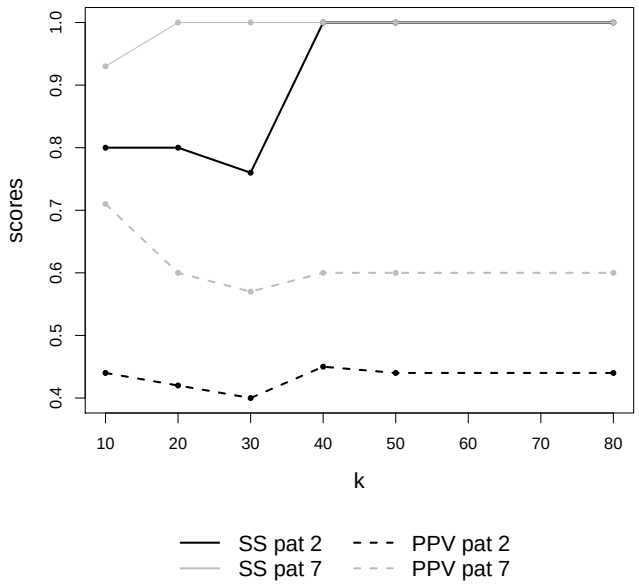

Figure 4: Dependency of sensitivity scores (full lines) and PPV-scores (dashed lines) of patient 2 (gray) and patient 7 (black).

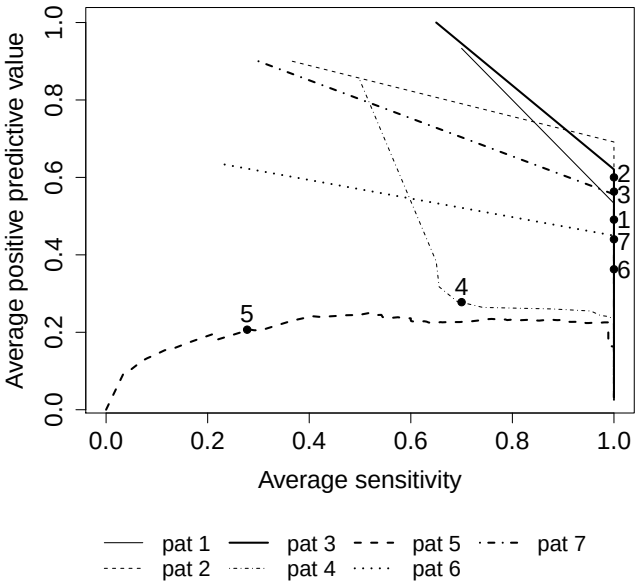

Figure 5: Precision-recall curves for patients $1-7$. The dots indicate the 95% threshold level chosen on the novelty scores.

the empirical quantiles obtained from the data. The x-axis shows the Gumbel quantiles obtained theoretically from (3). If the Gumbel model approximates well during simulation, the points on the graph are expected near the diagonal $y = x$, shown as a solid line. Plots of other runs (90 for table 2 and 70 for table 4) are very similar, following the diagonal $y = x$ very well.

Small values of k can lead to poor Gumbel approximations and thus to unreliable performance scores. To illustrate this fact, the sensitivity and PPV-scores for patients 2 and 7 are shown for different values of k in figure 4. It is clear that for small values of the parameter k performance scores are quite variable. Moreover, the sensitivity scores can show huge jumps when the number of seizures that can be studied in the test sets is small as is the case for the sensitivity scores for patients 7 (Table 1).

To continue the answer to question 4 the influence of the novelty score threshold and the possibility to further personalize our approach per patient is discussed. For this purpose figure 5 shows the precision-recall-curve (PR-curve) for all patients where the PPV and the sensitivity values are shown for different values of the novelty threshold, averaged over the different runs of the 10-fold randomization[1]. Novelty thresholds increase while moving up the PR-curves, starting from the lower right hand corner of the plot. The dots in each curve represent the 95% threshold level that was used in the previous experiments. For clarity, we opt to use the study with 0% seizures in the training data.

[1]We prefer the use of PR-curves over ROC-curves for the (highly) unbalanced datasets [34]

In comparison to the other patients, the seizures that are studied in the test set of patients 4 and 5 have lower novelty scores. Therefore the point on the PR-curve where sensitivity decreases from 100% is situated at a lower threshold level. For patient 4, only two seizures can be be studied in the test set (Table 1), which both result in novelty scores that are fluctuating around the threshold of 95%. Although the PPV can be increased by choosing higher threshold levels, the model is not able to fully detect both seizures in the test set. For patient 5 the PPV cannot be increased by choosing higher threshold levels. When the manifestations of the seizures in patient 5 were studied, it was clear that they were short in duration and subtle in intensity and were clearly different from the seizures in the other patients. Since the seizures are less extreme the model does not succeed in distinguishing them from normal movements resulting in an overall low PPV.

The PR-curves indicate that a 95% threshold level performs reasonably well on average but is suboptimal for each patient individually. For patients $1-3$ and $6-7$ performances of the system can be improved by a slight increase of the threshold, maintaining 100% sensitivity and increasing the PPV. For patient 4 and 5 choosing a lower threshold would result in 100% sensitivity while the PPV will roughly be unaltered. From these results it can be concluded that in principle it is possible to optimize the model parameters in a patient-specific manner if sufficient data are available.

To answer the last question, the CPU-time needed for the calculations on a standard laptop was measured (intel(R) core(TM) i7 Q 720 processor, clock rate 1.60Ghz

and RAM memory of 4GB). The Gumbel parameters are estimated in the training phase by simulating 300 minimal densities from the model of normal behavior. The CPU-time for this simulation does not exceed 2.5 minutes. The testing phase of one set of feature vectors can be done in a fraction of a second (0.48 seconds).

7. Conclusion and future research

In this paper an unsupervised system for the detection of hypermotor seizures is introduced. Since hypermotor seizures are rare, it is not only difficult to gather sufficient training data for traditional machine learning approaches, but the resulting imbalance in the dataset often causes difficulties for these approaches. Our method in contrast turns this imbalance into an advantage since all observed data are used to build a model of "normal behavior". EVT is used to automatically trigger a seizure detection when deviations from this model appear.

Our approach achieves performance scores that are similar to a state-of-the-art supervised machine learning method, such as an SVM-classifier. Although performance decreases slightly when seizure data are included to build the model, the proposed approach still yields reasonable results in the unsupervised setting.

The proposed methodology can be used to build an acquisition system that gains patient specific information in an incremental manner. For this purpose formula (1) can be used to adjust calculated densities to new incoming data. In this way the model of minimal densities can be calculated incrementally without restarting the training process from scratch (saving CPU-time). In order to do this one can choose to keep the covariance matrix H temporarily constant and to update it on a regular basis.

In future research, we would like to validate our approach on a larger dataset and investigate how to automatically optimize the model parameters per patient. Further research is needed to optimize the method in order to integrate the methodology in an alarm system that can be used in the clinical routine. At the moment however it can be seen as a screening tool that can be used to monitor patients during night time and analyze seizures that were not observed by the caregivers outside the EEG lab.

Acknowledgements

We gratefully acknowledge our copy editor Alex Wuyts and our anonymous reviewers for their helpful comments and suggestions.

References

[1] Normark MB, Erdal J, Kjaer TW. Video electroencephalography monitoring differentiates between epileptic and non-epileptic seizures. Danish Medical Bulletin 2011;58(9).

[2] Van de Vel A, Cuppens K, Bonroy B, Milosevic M, Jansen K, Van Huffel S, et al. Non-EEG seizure-detection systems and potential SUDEP prevention: state of the art. Seizure 2013;22(5):345–55.

[3] Dalton AF, Patel S, Chowdhury AR, Welsh M, Pang T, Schachter S, et al. Development of a body sensor network to detect motor patterns of epileptic seizures. IEEE Transactions on Biomedical Engineering 2012;59(11):3204–11.

[4] Tinuper P, Provini F, Bisulli F, Lugaresi E. Hyperkinetic manifestations in nocturnal frontal lobe epilepsy. semeiological features and physiopathological hypothesis. Neurological Sciences 2005;26(3):s210–4.

[5] Mirowski P, LeCun Y, Madhavan D, Kuzniecky R. Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. In: Larsen J, editor. 2008 IEEE Workshop on Machine Learning for Signal Processing. IEEE Workshop on Machine Learning for Signal Processing; IEEE, New York, USA; 2008, p. 244 –9.

[6] Chisci L, Mavino A, Perferi G, Sciandrone M, Anile C, Colicchio G, et al. Real-time epileptic seizure prediction using AR models and support vector machines. IEEE Transactions on Biomedical Engineering 2010;57(5):1124 –32.

[7] Cuppens K, Karsmakers P, de Vel AV, Bonroy B, Milosevic M, Lagae L, et al. Handling unbalanced data in nocturnal epileptic seizure detection using accelerometers. In: Fred A, De Marsico M, editors. Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods. ICPRAM 2013. INSTICC Press, Setubal, Portugal; 2013, p. 447–52.

[8] Coles S. An Introduction to Statistical modeling of Extreme Values. Springer Verlag, London; 2001.

[9] Han L, Wang H, Liu C, Li C. Epileptic seizure detection using wavelet transform based sample entropy and support vector machine. In: SONG Shuang HC, editor. Proceedings of the IEEE International Conference on Information and Automation. IEEE, New York, USA; 2012, p. 759–62.

[10] Uthayakumar R, Easwaramoorthy D. Epileptic seizure detection in EEG signals using multifractal analysis and wavelet transform. Fractals 2013;21(2).

[11] Tzallas AT, Tsipouras MG, Tsalikakis DG, Karvounis EC, Astrakas L, Konitsiotis S, et al. Automated epileptic seizure detection methods: A review study. In: Stevanovic D, editor. Epilepsy - Histological, Electroencephalographic and Psychological Aspects. InTech; 2012, p. 75–99.

[12] Kramer U, Kipervasser S, Shlitner A, Kuzniecky R. A novel portable seizure detection alarm system: Preliminary results. Journal of Clinical Neurophysiology 2011;28(1):36–8.

[13] Lockman J, Fisher RS, Olson DM. Detection of seizure-like movements using a wrist accelerometer. Epilepsy & Behavior 2011;20(4):638–41.

[14] Nijsen T, Aarts R, Cluitmans P, Griep P. Time-frequency analysis of accelerometry data for detection of myoclonic seizures. IEEE Transactions on Information Technology in Biomedicine 2010;14(5):1197–203.

[15] Cuppens K, Lagae L, Ceulemans B, Van Huffel S, Vanrumste B. Detection of nocturnal frontal lobe seizures in pediatric patients by means of accelerometers: a first study. In: He B, editor. 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society Conference Proceedings; IEEE, New York, USA; 2009, p. 6608–11.

[16] Jallon P. A bayesian approach for epileptic seizures detection with 3d accelerometers sensors. In: Sacristan E, editor. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE Engineering in Medicine and Biology Society Conference Proceedings; IEEE, New York, USA; 2010, p. 6325–8.

[17] Conradsen I, Beniczky S, Wolf P, Duun-Henriksen J, Sams T, Sorensen HBD. Seizure onset detection based on a uni- or multimodal intelligent seizure acquisition (uisa/misa) system. In: Sacristan E, editor. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE

Engineering in Medicine and Biology Society Conference Proceedings; IEEE, New York, USA; 2010, p. 3269–72.
[18] Roberts SJ. Novelty detection using extreme value statistics. IEE Proceedings on Vision, Image and Signal processing 1999;146(3):124–9.
[19] Roberts SJ. Extreme value statistics for novelty detection in biomedical signal processing. IEE Proceedings on Science, Technology and measurements 2000;47(6):363–7.
[20] Clifton DA, Hugueny S, Tarassenko L. Novelty detection with multivariate extreme value statistics. Journal of Signal Processing Systems 2011;65:371–89.
[21] Silverman B. Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability 26; Chapman & Hall/CRC, London; 1998.
[22] Duong T. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. Journal of Statistical Software 2007;21(7):1–16.
[23] R Core Team . R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria; 2012. URL: http://www.R-project.org/; (Accessed: 1 September 2013).
[24] Scott DW. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, New York; 1992.
[25] Cuppens K. Detection of epileptic seizures based on video and accelerometer recordings. Ph.D. thesis; KU Leuven; 2012.
[26] Cuppens K, Karsmakers P, Van de Vel A, Bonroy B, Milosevic M, Luca S, et al. Accelerometer based home monitoring for detection of nocturnal hypermotor seizures based on novelty detection. IEEE Journal of Biomedical and Health Informatics 2013;In press. 8p.
[27] The MathWorks Inc . Matlab: The Language Of Technical Computing. Natick, Massachusetts; 2010. Version 7.10.0 (R2010a).
[28] Dargie W. Analysis of time and frequency domain features of accelerometer measurements. In: Gopalan K, editor. Computer Communications and Networks, 2009. ICCCN 2009. Proceedings of 18th Internatonal Conference on Computer Communications and Networks; IEEE, New York; 2009, p. 1–6.
[29] Temko A, Thomas E, Marnane W, Lightbody G, Boylan G. EEG-based neonatal seizure detection with support vector machines. Clinical Neurophysiology 2011;122(3):464–73.
[30] Poh MZ, Loddenkemper T, Reinsberger C, Swenson NC, Goyal S, Sabtala MC, et al. Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. Epilepsia 2012;53(5):e93–7.
[31] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. Springer, New York; 2001.
[32] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Perrault CR, editor. Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2. International Joint Conference On Artificial Intelligence 1995; San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. ISBN 1-55860-363-8; 1995, p. 1137–43.
[33] Borra S, Di Ciaccio A. Measuring the prediction error. a comparison of cross-validation, bootstrap and covariance penalty methods. Computational Statistics and Data Analysis 2010;54:2976–89.
[34] Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: Cohen W, Moore A, editors. The International Conference on Machine Learning 2006. Proceedings of the 23rd international conference on Machine learning; ACM, New York, USA; 2006, p. 233–40.