Detecting Neonatal Seizures using Sample Covariance Estimation

Aleksandar Jeremic[1] and Dejan Nikolic[2]
[1]*Department of Electrical and Computer Engineering McMaster University, Hamilton, ON, Canada*
[2]*Physical Medicine and Rehabilitation, University Childrens Hospital, Faculty of Medicine, University of Belgrade, Belgrade, Serbia*

Keywords: Seizure Detection, Information Fusion.

Abstract: One of the most frequent of neurological dysfunctions in prematurely born infants is the presence of frequent seizures. As they may be related to serious neurological problems they require immediate detection which is most commonly done using electroencephalography (EEG) systems that enable trained physicians to detect them in the real time. Due to the length of neonatal period (first 28 days) it would be extremely beneficial to have an automated system that is able to detect seizures as it would enable more efficient use of expert time. In this paper we propose a new multichannel technique for detecting seizure in neonates that calculates distance measure using second order statistical properties and Frechet mean. We have demonstrated previously that Frechet mean in certain cases can outperform clustering/detection algorithms that are based on first order distances.

1 INTRODUCTION

A seizure is defined clinically as a paroxysmal alteration in neurologic function, i.e., behavioural, motor, or autonomic function. It is a result of excessive electrical discharges of neurones, which usually develop synchronously and happen suddenly in the central nervous system (CNS). It is critical to recognize seizures in newborns, since they are usually related to other significant illnesses. Seizures are also an initial sign of neurological disease and a potential cause of brain injury (Volpe, 2001).

In a clinical settings physicians are able to detect seizures based on EEG data however the process may be time consuming considering the number of cot-beds in regular size NICU department. To this purpose development of computer-aided diagnosis would be extremely beneficial as such system would be important from both academic and clinical standpoint of view. From the academic stand point automatic recording of seizures and consequently analysis of these data would provide insight into frequency of occurrence and correlate it with the dynamic of neurological development. From clinical standpoint it could be useful tool for adjusting level of medical care based on the neurological state of the brain with respect to seizures.

In our previous work, we proposed several distributed detection algorithms for neonatal seizure detection using some of the commonly used seizure detection algorithms. In this paper we propose new local detectors based on the Frechet mean of the EEG signal covariance calculated using sliding window. First, we present an estimator of the Frechet mean of the covariance matrices on the manifold $\mathcal{M}$ using the different measures of Riemannian distances. Then we introduce the Fréchet mean based on several Riemannian distances and discuss computational algorithms for calculating the proposed distance means. In Section 3 we illustrate applicability of our results using data set of NICU patients. Finally, in Section 4 we discuss future directions.

2 SIGNAL MODEL

2.1 Frechet Mean Local EEG Detectors

We use the notion of Fréchet mean to unify the method of finding the mean of positive definite matrices. The Fréchet mean is given as the point which minimizes the sum of the squared distances (Barbaresco, 2008):

$$\hat{\mathcal{S}} = \mathrm{argmin}_{\mathcal{S}\in\mathcal{M}} \sum_{i=1}^{n} d^2(\mathbf{S}_i, \mathcal{S}) \qquad (1)$$

where $\{\mathbf{S}_i\}_{i=1}^{n}$ represents the symmetric positive defi-

Jeremic, A. and Nikolic, D.
Detecting Neonatal Seizures using Sample Covariance Estimation.
DOI: 10.5220/0007580302460250
In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019)*, pages 246-250
ISBN: 978-989-758-353-7

nite matrices and $d(.,.)$ denotes the metric being used respectively. Therefore the above expression can be interpreted as a way of calculating an averaged sample covariance matrix using a sliding window where S_i represents an $i-th$ window sample covariance estimate. Then the overall estimate of the covariance matrix is calculated using a particular metric. We will use this technique to calculate sample covariance matrix of EEG signal in the absence of seizures assuming that these intervals of no-seizures were properly identified by an expert.

To measure the distance between two $M \times M$ covariance matrices $\mathbf{A}$ and $\mathbf{B}$ on manifold of positive definite matrices $\mathcal{M}$, we consider the metrics which have been developed to measure distance between two points on the manifold itself.

The first metric is obtained by measuring distance between projections on the subspace spanned by unitary matrices (Li and Wong, 2013)

$$d_{R_1}(\mathbf{A},\mathbf{B}) = \sqrt{\mathrm{Trace}(\mathbf{A}) + \mathrm{Trace}\mathbf{B}) - 2\mathrm{Trace}(\mathbf{A}^{\frac{1}{2}}\mathbf{B}\mathbf{A}^{\frac{1}{2}})^{\frac{1}{2}}}$$

In general for any positive definite matrix $\mathbf{A}$ its square root is defined as $\mathbf{A}^{\frac{1}{2}} = \mathbf{S}\sqrt{\mathcal{L}}\mathbf{D}^H$; where $\mathbf{A} = \mathbf{S}\mathcal{L}\mathbf{D}^H$ is the eigenvalue value decomposition of matrix $\mathbf{A}$ with diagonal matrix $\mathcal{L}$ consisting of eigenvalues of $\mathbf{A}$.

The second metric is obtained by measuring the distance between their projections on the subspace spanned by identity matrices. It has been shown (Li and Wong, 2013) that this distance is equivalent to:

$$d_{R_2}(\mathbf{A},\mathbf{B}) = \sqrt{\mathrm{Trace}(\mathbf{A}) + \mathrm{Trace}(\mathbf{B}) - 2\mathrm{Trace}(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}})} \tag{2}$$

Finally, as a last local detector we propose to use the log- Riemannian metric is given as (Moakher, 2005):

$$d_{R_3}(\mathbf{A},\mathbf{B}) = \left\|\log(\mathbf{A}^{-\frac{1}{2}}\mathbf{B}\mathbf{A}^{-\frac{1}{2}})\right\|_2 = \sqrt{\sum_{i=1}^{M}\log^2(\mathcal{L}_i)} \tag{3}$$

where the $\mathcal{L}_i$'s are the eigenvalues of the matrix $\mathbf{A}^{-1}\mathbf{B}$ (Absil et al., 2009).

In order to solve the corresponding minimization problems we presented detailed computational algorithms for calculating these distances in (Jahromi et al., 2015). In all the cases certain iterative procedures are necessary however we demonstrated existence of unique solutions (means) for all the proposed distances.

In order to define local detectors we first identify no-seizure segments and define overlapping windows which are used to calculate covariance matrices in the absence of seizures. using the following algorithm: Let y_i be the i-th sample of inter-arterial pressure measurements. Then the outline of the algorithm is as follows

- within the training data set create windows $\vec{d}_k = [y_{(k-1)*l1+1}, \cdots, y_{k*l1-1}]$ where $l1$ is the length of the window
- within the above window select subwindows of length $l2$ and label them $\vec{d}_k^j$ where $j = 1, l1 - l2 + 1$
- remove the sample mean from the window vectors
- calculate rank 1 sample covariances $\vec{d}_k^{j,T}\vec{d}_k^j$ and average them using Frechet mean instead of commonly used addition

These sample covariances are then used as a cluster of reference covariance matrices in which the centre of the cluster is defined using the above metrics. The threshold is then calculated as a function of predefined probability of false alarm i.e. incorrect seizure detection. Therefore by setting a false alarm ratio to α we can empirically calculate threshold for a particular patient by using event-free segments of EEG recordings.

2.2 Distributed Detection System

Each of the metric detectors v presented in the previous section can be considered as a single channel i.e. local detector. In order to improve the overall performance of a single detectors we propose to combine the existing single detectors and utilize their strengths by extending previous results on blind multichannel information fusion (Liu et al., 2007).

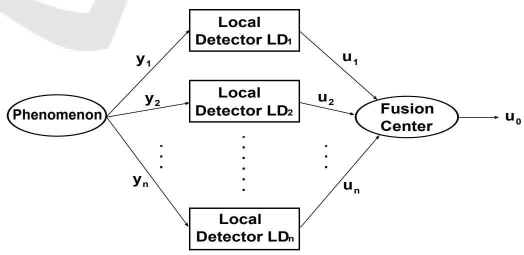

Figure 1: Parallel Distributed Detection System.

Figure 1 shows the structure of a typical parallel distributed detection system with N detectors. The local detectors transmit local decisions u_n based on a particular metric that they are using. Obviously in our case there are three local detectors as we are using three different metrics. All the local decisions are then sent to the fusion centre, where the global decision u_0 is made based on a fusion rule in order to minimize the overall probability of error. Additional detectors can be added into the system whenever more information is required to make final decision.

The local decisions u_n, $n = 1,2,3$ can be expressed as

$$u_n = \begin{cases} 0, & \text{the } n\text{th detector favours } H_0 \\ 1, & \text{the } n\text{th detector favours } H_1 \end{cases} \quad (4)$$

where "favours" should be interpreted as the distance between actual sample covariance estimate and reference covariance estimate is smaller than the empirical threshold for a particular false alarm rate. We use $P(H_0)$ to denote prior probability that the seizures are not present in a particular signal segment. A common assumption used here is the local observations y_n are conditionally independent, given the unknown hypothesis H_i.

After receiving the local decisions, the fusion centre makes the global decision by applying an optimal fusion rule in order to minimize the final error probability. For a binary hypothesis testing problem, the error probability P_e is given by

$$P_e = P(H_0)P(u_0 = 1|H_0) + P(H_1)P(u_0 = 0|H_1) \quad (5)$$

The authors provided the optimality criterion for N local detectors in the sense of minimum error probability in (Varshney, 1986). We recall it here for the case of $N = 3$.

$$u_0 = \begin{cases} 1, & \text{if } w_0 + \sum_{n=1}^{3} w_n > 0 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$\text{where,} \quad w_0 = \log\left(\frac{P_1}{P_0}\right) \quad (7)$$

$$\text{and} \quad w_n = \begin{cases} \log((1-P_n^m)/P_n^f), & \text{if } u_n = 1 \\ \log(P_n^m/(1-P_n^f)), & \text{if } u_n = 0 \end{cases} \quad (8)$$

The probabilities of false alarm and missed detection of the nth local detector are denoted as P_n^f and P_n^m, respectively. The optimal fusion rule tells us that the global decision u_0 is determined by the a priori probability and the detector performances, i.e., P_1, P_n^f and P_n^m. However, they are all unknown in our seizure detection problem, which is usually the case in many other real applications (Mirjalily, 2003),(Liu et al., 2007). In order to make the final decision, we need to utilize the information available to us: the local binary decisions u_n.

Suppose the decision combination $\{u_1 = i,\ u_2 = j \text{ and } u_3 = k\}$ is represented by $\ell = (ijk)_2$, where $i, j, k = 0$ or 1 (Mirjalily, 2003). In our system, the number of all the possible local decision combinations is 2^3 and will be denoted as L in the remainder of this paper. The joint probability of decision $\{u_1 = i,\ u_2 = j \text{ and } u_3 = k\}$ is also the occurrence probability of the ℓth decision combination, given by

$$\begin{aligned} P_\ell &= \Pr(u_1 = i, u_2 = j, u_3 = k) \\ &= P(u_1 = i|H_1)P(u_2 = j|H_1)P(u_3 = k|H_1)P_1 \\ &\quad + P(u_1 = i|H_0)P(u_2 = j|H_0)P(u_3 = k|H_0)(1-P_1) \end{aligned} \quad (9)$$

$$P(u_n = i|H_1) = \begin{cases} 1-P_n^m, & \text{if } i = 1 \\ P_n^m, & \text{if } i = 0 \end{cases} \quad (10)$$

$$P(u_n = i|H_0) = \begin{cases} P_n^f, & \text{if } i = 1 \\ 1-P_n^f, & \text{if } i = 0 \end{cases} \quad (11)$$

In this nonlinear system, only seven out of eight equations are independent since $\sum P_\ell = 1$ and there are seven unknowns P_1, P_n^f and P_n^m, for $n = 1,2,3$. Thus, it can be solved when P_ℓ are known. Although P_ℓ is usually unavailable in practice, it could be replaced by empirical probability defined as

$$\begin{aligned} P_\ell &= \Pr(u_1 = i, u_2 = j, u_3 = k) \\ &\simeq \frac{\text{number of } u_1 = i, u_2 = j, u_3 = k}{\text{number of local decisions } N_t} \end{aligned} \quad (12)$$

where N_t is the number of decisions made by one of the local detectors. The analytical solution to the above nonlinear equations is given in (Mirjalily, 2003).

Note that in a particular setting if the data size is limited and/or the number of events needed for accurate calculation of anomalies is not sufficient we developed a maximum likelihood based algorithm that exploits the multinomial probability mass function describing the decision vector and utilized in order to estimate the anomalies as well as prior probabilities (seizure and no-seizure). We presented the details of these algorithms in (Liu et al., 2014).

3 RESULTS

We evaluate the performance of the proposed algorithms on the data set consisting of preterm infants (GA less than 32 weeks) admitted to the Neonatal Intensive Care Unit at McMaster Hospital. Due to physical limitations we were able to obtain prior expect knowledge on a very limited time length and hence all of the non-seizure epoch were shorter than 400 samples using single C3 channel with minimal motion artefacts.

For illustrational purposes in Figures 2-4 , we illustrate the detection performance as a scatter diagram of windows selected from testing data. Note that in the presence of motion artifacts the actual performance will actually vary significantly. Furthermore

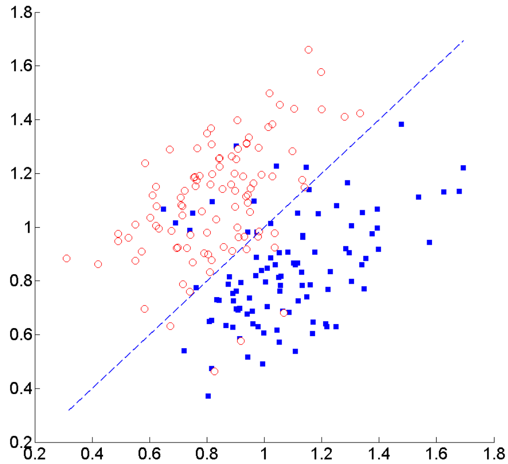

Figure 2: Scatter plot of detection performance using metric d_{R1}.

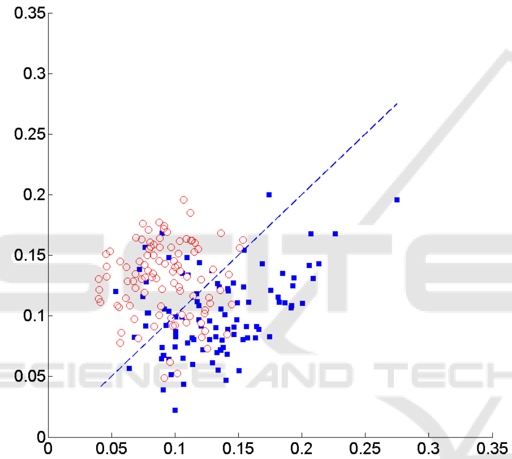

Figure 3: Scatter plot of detection performance using metric d_{R2}.

because the original system design was based on no-seizures the system was calibrated so that the probability of false alarm is controlled. Due to motion artifacts and reaction to pain stimuli during medical procedures in NICU it is quite likely that local detectors will identify these manifestation in EEG as false seizure. The x and y axes represent distances to covariance matrices corresponding to signal segments with and without seizures. Note that in order to test applicability of the proposed techniques we selected signal segments in which the prior probabilities are approximately the same.

Table 1: Average seizure detection performance.

	d_{R1}	d_{R2}	d_{R3}	ML-Fused
false seizures	0.07	0.09	0.12	0.05
missed seizures	0.09	0.08	0.11	0.07

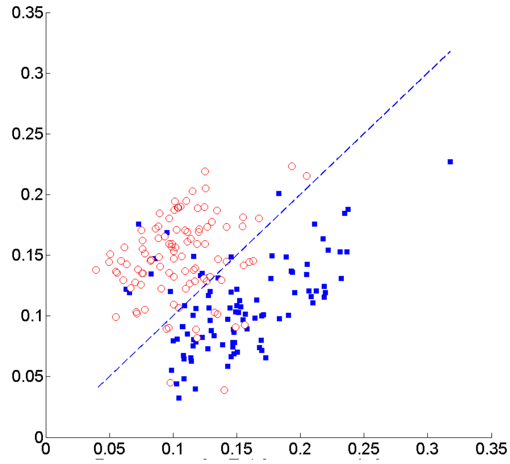

Figure 4: Scatter plot of detection performance using metric d_{R3}.

4 CONCLUSIONS

Automatic systems for seizure detection have been subject of considerable research interest in the past. One of main advantages lies in the fact that expert time is potentially required only during the training session. Furthermore, for newborn patients admitted to NICU such systems enable continuous monitoring of seizure events and hence can provide better insight into neurological development. In recent years significant effort has been placed on developing systems that predict seizures in order to potentially counter them with appropriately generated electrical stimuli. To this purpose in this paper we examine possibility of detecting seizures by measuring different distances between sample covariance matrix estimates. Since different second order distances focus on various structural information we propose to combine their decisions by minimizing overall probability of error. To achieve this goal we define local detectors using empirically determined threshold and fuse their local decisions using our previously developed information fusion algorithm for seizure detection. We demonstrate the applicability of the proposed algorithms using a real data set consisting of multiple NICU patients.

In future work we plan to improve performance by including mean based local detectors as well as instantaneous frequency based detectors as they may account for features that are not accounted for in the proposed covariance based detectors. Furthermore the performance of these detectors should be investigated in scenarios in which priors may have different values as the training of the proposed system depends on seizure occurrence frequency. Finally an effort should be made to evaluate performance when the training set includes epoch intervals that include seizures. In this case we expect that the problem can be easily formulated as a classification problem in which case the Frechet-mean based algorithms often improve performance when combined with mean based algorithms (such as k-means).

REFERENCES

Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.

Barbaresco, F. (2008). Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices & information geometry. *Radar Conference, 2008. RADAR'08. IEEE*, pages 1–6.

Jahromi et al., M. (2015). Estimating Positive Definite Matrices Using Frechet Mean. In *Biosignals 2015*.

Li, Y. and Wong, K. M. (2013). Riemannian distances for EEG signal classification by power spectral density. *IEEE journal of selected selected topics in signal processing*.

Liu, B., Jeremic, A., and Wong, K. (2007). Blind adaptive algorithm for M-ary distributed detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, volume 2.

Liu, B., Jeremic, A., and Wong, K. (2014). Optimal distributed detection of multiple hypotheses using blind algorithm. *IEEE Trand. on Aerospace and Electronic Systems*, 50:1190–1203.

Mirjalily, G. e. (2003). Blind adaptive decision fusion for distributed detection. *IEEE Transactions on Aerospace and Electronic Systems*, 39(1):34–52.

Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747.

Varshney, P. (1986). Optimal data fusion in multiple sensor detection systems. *IEEE Trans. on Aerospace and Electronic Systems*, pages 98–101.

Volpe, J. (2001). *Neurology of the newborn*. WB Saunders Co.