DETECTION OF NEWBORN EEG SEIZURE USING OPTIMAL FEATURES BASED ON DISCRETE WAVELET TRANSFORM

Pega Zarjam, Mostefa Mesbah and Boualem Boashash

Signal Processing Research Centre
Queensland University of Technology
GPO Box 2434, Brisbane, QLD 4001, Australia
Email: p.zarjam@student.qut.edu.au

1. ABSTRACT

A new automated method is proposed to detect seizure events in newborns from Electroencephalogram (EEG) data. The detection scheme is based on observing the changing behavior of the wavelet coefficients (WCs) of the EEG signal at different scales. An optimal feature subset is obtained using the mutual information evaluation function (MIEF). The MIEF algorithm evaluates a set of candidate features extracted from WCs to select an informative feature subset. The subset is then fed to an artificial neural network (ANN) classifier that organizes the EEG signal into seizure or non-seizure activity. The performance of the proposed features is compared with that of the features obtained using mutual information feature selection (MIFS) algorithm. The training and test sets are obtained from EEG data acquired from 5 neonates with ages ranging from 2 days to 2 weeks.

2. INTRODUCTION

Seizures represent the most distinctive sign of neuralgic disease in the neonate. The failure to quickly and accurately diagnose the seizure can lead to brain injury or even death. Despite their importance as a diagnosis tool, seizures still elude recognition. This is due to many factors among them are: their delicate and anarchic characteristics, the overlap of their characteristics and those of the background (non-seizure), and their time-varying nature [1]. The widely used method to identify seizures is based on the visual analysis of the EEG data. Its recognition is a time-consuming task which requires highly trained professionals. In order to reduce the cost associated with it, automated detection techniques are needed.

The existing non-parametric methods for EEG seizure detection can be classified into three categories: time-domain based, frequency-domain based, and time-frequency / time-scale based [2, 3, 4, 5]. Since the EEG signals are generally non-stationary, the time-scale and time-frequency methods are usually the most suitable for seizure detection [5].

In this paper, we use the MIEF algorithm to find an optimal feature set based on the discrete wavelet transform (DWT) of the EEG data. The performance of the proposed feature set is evaluated and then compared with that of the feature set, obtained using the MIFS algorithm [6]. The discrimination between seizure and non-seizure states are based on the changing behavior of some statistical quantities of WCs of the EEG.

3. FEATURE EXTRACTION FROM EEG DATA USING THE DWT

The concept of time-scale signal analysis using the WT has been recognized as a potential tool for analyzing non-station -ary signals, such as EEG [7]. Selecting the mother wavelet that matches the shape or frequency characteristics of EEG seizure, is crucial. In this study, we have chosen Daubechies 4-tab wavelet for decomposing the EEG signal, since it gives sufficient resolution and has been proved to be efficient for detection of seizure events [8].

The neonatal EEG signals used in this study were recorded under clinical conditions, by means of MEDELC system, at the Royal Children Hospital in Brisbane, Australia. An expert opinion about the presence of seizure activity in different channels of each of the five neonates was obtained. The well-defined seizure components and non-seizure components were then extracted and used to create training and test databases. Each seizure and non-seizure part of the database was divided into 6 seconds segments. Each segment contains 1536 data points ($f_s = 256$ Hz). The length of the segment was determined by considering the two conflicting issues of reliability in statistical estimation and the stationarity of the signal. EEG segments were then decomposed using the DWT into 9 scales as illustrated in Table 1. Each scale i contains detail components d_i, approximate

This work is supported by an Australian Research Council (ARC) large grant

Wavelet Scale	*Freq. Range*	*No. of Coefficients*
App. Scale	$0 - 0.25$ Hz	3
Scale 9	$0.25 - 0.50$ Hz	3
Scale 8	$0.50 - 1$ Hz	6
Scale 7	$1 - 2$ Hz	12
Scale 6	$2 - 4$ Hz	24
Scale 5	$4 - 8$ Hz	48
Scale 4	$8 - 16$ Hz	96
Scale 3	$16 - 32$ Hz	192
Scale 2	$32 - 64$ Hz	384
Scale 1	$64 - 128$ Hz	768

Table 1. Freq. range corresponding to each wavelet scale, $f_s = 256$.

components a_i, detail coefficients cd_i, and approximate coefficients ca_i.

After examining different statistics of the decomposed EEG segments, the following features were found to be suitable for representing the EEG data:

- The variance of the detail coefficients of the scales 1 to 8 ($Var(cd_i),\ i = 1, 2, ..., 8$) and the detail components of the scales 1 to 5, and 7 ($Var(d_i),\ i = 1, ..., 5, 7$)
- The mean of the detail coefficients of the scales $2, 6$, and 7 ($Mean(cd_i),\ i = 2, 6, 7$) and the detail components of the scales 3 to 9 ($Mean(d_i),\ i = 3, ..., 9$)
- The variance of the approximate coefficients and approximate component of the scale 9 ($Var(ca_9)$, $Var(a_9)$)

The above quantities form a vector F containing $N = 26$ features.

4. OPTIMAL FEATURE SELECTION METHODS

4.1. THE MIFS METHOD

Feature selection is a significant step in classification, as irrelevant and redundant features often degrade the performance of classification algorithm both in speed and accuracy and also increase complexity of the classifier. Therefore, it is desirable to use an optimizing method to reduce the number of selected features, by only keeping the features which provide the greatest contribution to classification performance.

In a previous publication [6], the authors applied the MIFS optimization technique proposed by Battiti [9]. In order to provide an optimal feature set, the MIFS uses the concept of mutual information (MI) to evaluate the information content of each individual feature with regard to the output class. The MIFS algorithm is formalized as follows:

1. Set $F \leftarrow$ “initial set of N features”; $S \leftarrow \{\emptyset\}$.
2. For each feature $f \in F$, compute $I(C; f)$.
3. Find the feature f that maximizes $I(C; f)$;
   set $F \leftarrow F \setminus \{f\}$; set $S \leftarrow \{f\}$.
4. Repeat until $|S| = M$ (M chosen a priori),
   (a) For all couples of variables $(f; s)$ with $f \in F$, $s \in S$, compute $I(f; s)$.
   (b) Choose feature f that maximizes
   $$I(C; f) - \beta \sum_{s \in S} I(f; s)$$
   set $F \leftarrow F \setminus \{f\}$; set $S \leftarrow S \cup \{f\}$.

In the MIFS algorithm, vectors F and S represent the initial and selected feature sets with lengths N and M respectively. Vector C represents the output classes and $I(C; f)$ is the MI between variables C and f. The parameter β regulates the relative importance of the MI between a candidate feature and the already-selected features with respect to the MI of output classes. β is chosen between 0.5 and 1 [9]. This algorithm enables us to find the subset S (with M features) out of a given initial set F (with N features) that maximizes the MI.

4.2. THE MIEF METHOD

In this paper, we propose to use the MIEF optimization method to find the optimal feature subset [10] . The MIEF is a measure of the ability of feature subsets to distinguish between class labels.

A good evaluation function for a specific set of features F and class labels C is $I(C; F)$. However, due to the computational load involved, the MIEF aims to find an evaluation function $g(F)$ whose performance is close to $I(C; F)$ and yet with a limited computational load. The MIEF is formalized as follows:

1. For each $f_i \in F$, choose the feature that has maximal $I(C; f_i)$; set $S \leftarrow f_i$; set $g(S) = I(C; f_i)$.
2. For each feature $f_i \in F,\ f_i \notin S$ compute:
   $$m(f_i) = g(k) + \lambda I(C; f_i)$$
   then choose the feature f_i that maximizes m; set $S \leftarrow S \cup f_i;\ g(S) = m(f_i)$
3. If $|S| < |F|$ go to step 2.

4. $g(F) = g(S)$

In the first step, the function g is initialized to $I(C; f_i)$, which is the maximum MI between a single feature and the class labels. Step two states that the intermediate function m of feature f_i is a summation of its MI with class labels multiplied by the factor λ, which represents information gain, and the latest value of function g. The value of λ ranges between $[0, 1]$, where $\lambda = 0$ means no information is gained. $\lambda \to 1$ if f_i contributes by an amount equals to its information about the class labels when considered alone. After computing m for all non-considered features in the subset, we pick out the features that gives the maximum and substitute its m value into g (step 2). This procedure is repeated until we consider all the features in the subset.

5. CLASSIFIER

Once we have selected the feature vector, an ANN was used to identify seizures and non-seizures in the EEG data. A three layer, fully-connected feed-forward network was used. It employs an adaptive back propagation gradient-decent learning algorithm.

The features related to seizure and non-seizure were separated and normalized to $[-1, 1]$. All seizure data were given a target value of 1 and non-seizure a target value of -1. The ANN was trained in a supervised mode.

6. EXPERIMENTAL RESULTS

6.1. Using the Full Feature Set

All the 26 features explained in section3 have been used here. Two different parameters have been employed to asses the performance of the extracted feature set. Seizure detection rate (SDR) shows the ability of the method to successfully classify seizure activities as seizure, while on-seizure detection rate (NDR) indicates the ability of the method to successfully classify non-seizure activities as non-seizure. From NDR, we can obtain the false alarm rate (FAR $=$ 100% - NDR). The test resulted in an average SDR of 97.48% and FAR of 8.16%. The process is patient and threshold-independent as the ANN performance was achieved using an output threshold of 0.

6.2. Using the Optimal Feature Subsets

We used the MIFS and MIEF algorithms to select M features ($M = 1 : 10$) from the initial set F which contains $N = 26$ features. The parameters β and λ were chosen to be 0.75 and 0.6 respectively. Tables 2 and 3 contain selected features using the MIFS and MIEF algorithms.

No.	*Selected features*	*No.*	*Selected features*
1	$Var(cd_3)$	2	$S_1 \cup Mean(d_4)$
3	$S_2 \cup Mean(d_9)$	4	$S_3 \cup Mean(cd_2)$
5	$S_4 \cup Mean(d_8)$	6	$S_5 \cup Mean(cd_7)$
7	$S_6 \cup Mean(d_5)$	8	$S_7 \cup Mean(d_3)$
9	$S_8 \cup Var(ca_9)$	10	$S_9 \cup Mean(d_6)$

Table 2. Selected features using the MIFS algorithm.

No.	*Selected features*	*No.*	*Selected features*
1	$Var(cd_3)$	2	$S_1 \cup Mean(d_9)$
3	$S_2 \cup Var(cd_6)$	4	$S_3 \cup Mean(d_6)$
5	$S_4 \cup Var(cd_7)$	6	$S_5 \cup Var(cd_4)$
7	$S_6 \cup Var(ca_9)$	8	$S_7 \cup Mean(d_5)$
9	$S_8 \cup Var(d_5)$	10	$S_9 \cup Var(cd_2)$

Table 3. Selected features using the MIEF algorithm.

The selected features, were then fed to the above described ANN. The performance of the two selected techniques in terms of SDR and NDR as a function of the numbers of selected optimal features is shown in Figs. 1 and 2.

There are some fluctuations in the results due to the limited amount of the test data and iterations. Nevertheless, it can be seen that with 10 selected features both methods have good performance.

The MIFS algorithm has picked out the optimal subset containing 10 features which resulted in an average SDR of 94% and FAR of 7.12%.

The MIEF algorithm has selected the optimal 10 features which presented an average SDR of 96.35% and FAR of 6.2%. The selected subset using the MIEF algorithm outperforms that of the MIFS algorithm in terms of SDR. As shown, SDR is about 93.5% with only 4 features selected by the MIEF algorithm. The obtained NDR with a subset of more than 7 features selected by the MIEF is also higher than that of the MIFS.

Comparing with the results given in section 6.1, it is clear that the optimal feature subsets have improved the system performance in terms of FAR. Besides, using the optimal feature subsets can significantly reduce the system complexity for implementation purpose.

7. CONCLUSION

This paper proposed the use of the DWT and ANN for neonatal EEG seizure detection. The MIFS and MIEF algorithms were used to perform feature selection for classifier. The optimal feature subsets decreased the number of inputs to the ANN and resulted in an efficient and less complex structure.

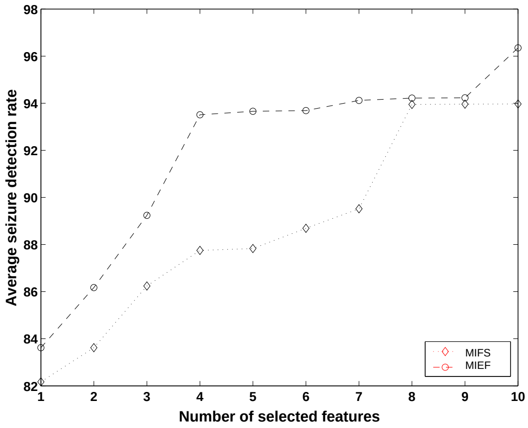

Fig. 1. The average seizure detection rate of two methods as a function of the optimal selected features.

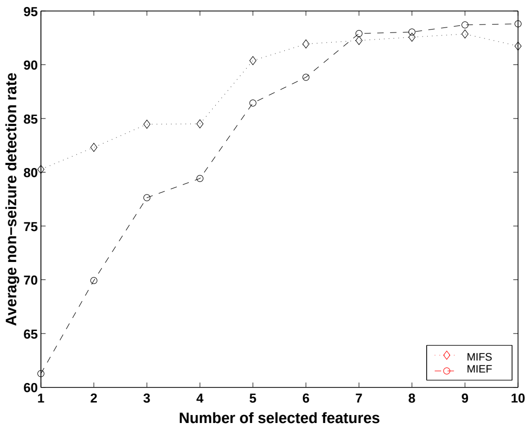

Fig. 2. The average non-seizure detection rate of two methods as a function of the optimal selected features.

The selected subset using the MIEF algorithm outperformed that of the MIFS algorithm particularly in terms of SDR. This is because the MIEF function takes into consideration how features work together and measures the amount on information in a feature subset to distinguish between class labels.

8. ACKNOWLEDGMENT

We would like to acknowledge the assistance of Prof. P. Coldtiz, Dr. Chris Burge and Ms. J. Richmond of the Royal Children Hospital, Brisbane, Australia in obtaining and interpreting the EEG data used in the project. We would also like to thank A. Al-Ani for the discussion related to feature selection algorithms.

9. REFERENCES

[1] P. Kellaway and E. M. Mizrahi, *Neonatal seizures*, Raven Press, 1990.

[2] A. Liu, J. S. Hahn, G. P. Heldt, and R. W. Coen, “Detection of neonatal seizures through computerized eeg analysis,” *Electrenc. and Clin. Neurophysy*, , no. 82, pp. 30–37, 1992.

[3] J. Gotman, D. Flanagah, J. Zhang, and B. Rosenblatt, “Automatic seizure detection in the newborn: methods and initial evaluation,” *Electrenc. and Clin. Neurophysy*, vol. 103, pp. 356–362, 1997.

[4] S. Nagasubramanian, B. Onaral, and R. Clancy, “Online neonatal seizure detection based on multi-scale analysis of eeg using wavelet as a tool,” *IEEE, 19th international conference in Medicine and Biology Society*, vol. 3, pp. 1289–1292, Nov. 1997.

[5] B. Boashash, H. Carson, and M. Mesbah, “Detection of seizures in newborns using time-frequency analysis of eeg signals,” *10th IEEE workshop on Statistical Signal and Array Processing (SSAP), Pocono Manor, Pennsylvania, USA*, pp. 564–568, Aug. 2000.

[6] P. Zarjam, M. Mesbah, and B. Boashash, “An optimal feature set for seizure detection systems for newborn EEG signals,” *Submitted to ISCAS’2003, Bangkok, Thailand.*

[7] S. V. Mehta, R. W. Koser, and P. J. Venziale, “Wavelet analysis as a potential tool for seizure detection,” *Proceeding of the IEEE-SP international symposium in time-frequency and time-scale analysis*, pp. 584–587, Oct. 1994.

[8] H. S. Park, Y. H. Lee, D. S. Lee, and S. Kim, “Detection of epilepticform activity using wavelet and neural network,” *Proceeding of the 19th Annual EMBS conference*, 1997.

[9] R. Battiti, “Using mutual information for selecting features in supervised neural net learning,” *IEEE Transaction on Neural Networks*, vol. 5, no. 4, pp. 537–550, 1994.

[10] M. Al-Ani, A.and Deriche, “Feature selecting using a mutual information based measure,” *International Conference on Pattern Recognition (ICPR),Quebec.*, Aug. 2002.