34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Dynamic Time Warping Based Neonatal Seizure Detection System

Rehan Ahmed[1], Andrey Temko[1], William Marnane[1], Geraldine Boylan[2] and Gordon Lighbody[1]

***Abstract*— Neonatal seizures patterns evolve with changing frequency, morphology and propagation. This study is an initial attempt to incorporate the characteristics of temporal evolution of neonatal seizures into our developed neonatal seizure detector. The previously designed SVM-based neonatal seizure detector is modified by substituting the Gaussian kernel with the Gaussian dynamic time warping kernel, to enable the SVM to classify variable length sequences of feature vectors of neonatal seizures. The preliminary results obtained compare favorably with the conventional SVM. The fusion of the two approaches is expected to improve the current state of the art neonatal seizure detection system**

I. INTRODUCTION

Neonatal seizures are the most common neurological dysfunction in newborns. About one third of all neonatal seizures are clinically visible [1] and many remain undetected in the busy Neonatal Intensive Care Unit (NICU). Failure to detect seizures and the resulting lack of treatment may result in brain damage and in severe cases, death. A system that could automatically detect and annotate seizures on the neonatal electroencephalogram (EEG) would be extremely useful for clinicians in the NICU.

A single neonatal seizure changes in frequency, morphology and propagation. It was shown in [2] that changes in morphology and frequency were present in more than a half of the total neonatal ictal discharges. It can be seen in the example shown in Fig. 1 where a seizure event starts with high amplitude spikes and ends up with very low amplitude spikes. The lack of strong inhibitory factors in immature brain likely contributes to the propensity for spread of the discharges. The changing morphology of the discharges may be the result of a slow recruitment of additional neuronal networks during ictus [2]. Clearly, a detector that can track the seizure temporal evolution characteristics is expected to reach better performance than the classifier that utilizes only static seizure features.

Several ways to incorporate temporal evolution, signal dynamics, contextual information, or sequentiality of the data have been proposed in the literature. Depending on the stage where such information can be introduced, the methods can be divided into the feature, classifier, and decision level categories. The feature level techniques focus on the extraction of short-term data dynamics. An example would be temporal derivatives (delta coefficients) which are widely used in speech processing and are derived from either temporal differentiation or from regression analysis [3]. The classifier level techniques try to model a longer-term temporal evolution. Such methods include Support Vector Machines (SVM) with various sequential kernels [4] which have been used for a number of applications [4], [5], [6]. Hidden Markov Models (HMM) have been widely used to capture sequentiality of the data by uncovering the underlying event structures [7]. The decision making level techniques include smoothing filters such as moving average or median filters, or more sophisticated Viterbi temporal decoding [8] which incorporates the contextual information before making the final decision.

There are striking differences between seizures in neonates and those of older patients in ictal EEG patterns [2]. Although a few attempts have been made to exploit the pattern temporal evolution in the adult EEG [9], to the best of our knowledge, the contextual information for neonatal seizure detection has been either not exploited at all or introduced in the later stages through simple decision smoothing [10].

In this contribution, the previously developed SVM-based neonatal seizure detector [10] is modified by substituting the Gaussian kernel with the Gaussian Dynamic Time Warping Kernel (GDTW) [11] to enable the SVM classify variable length sequences of EEG feature vectors.

II. DYNAMIC TIME WARPING KERNEL

A. Dynamic Time Warping

Dynamic Time Warping (DTW) gives a measure of the similarity between two variable length sequences. Consider two sequences $P = (p_1, ..., p_n)$ and $Q = (q_1, ..., q_m)$ with lengths n and m A local distance $d_l = (p_i - q_j)^2$ between each element of the two sequences can be calculated to get an alignment matrix of size m. A warp path $W = w_1, ..., w_K$ could be constructed in this matrix, where K is the length of warp path. The kth element, $w_k = (i, j)$, of the warp path represents a matching point of two sequences, where (i, j) corresponds to ith and jth indexes of sequences P and Q respectively. An alignment distance D_W could be calculated along a particular warp path W by

$$D_W(P, Q) = \frac{1}{K} \sum_{k=1}^{K} d_l(w_{ki}, w_{kj}) \quad (1)$$

Then by finding the optimal alignment path that gives the shortest distance $D_\phi = min\{D_W(P, Q)\}$, the best measure of similarity between the two sequences can be

*This work was supported by SFI (SFI/10/IN.1/B3036).

[1]Rehan Ahmed, Andrey Temko, William Marnane and Gordon Lightbody are with the Department of Electrical and Electronic Engineering and the Neonatal Brain Research Group, University College Cork, Ireland {rehan,andreyt}@eleceng.ucc.ie, {g.lightbody,l.marnane}@ucc.ie

[2] Geraldine Boylan is with the Department of Pediatrics and Child Health and the Neonatal Brain Research Group, University College Cork, Ireland g.boylan@ucc.ie

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

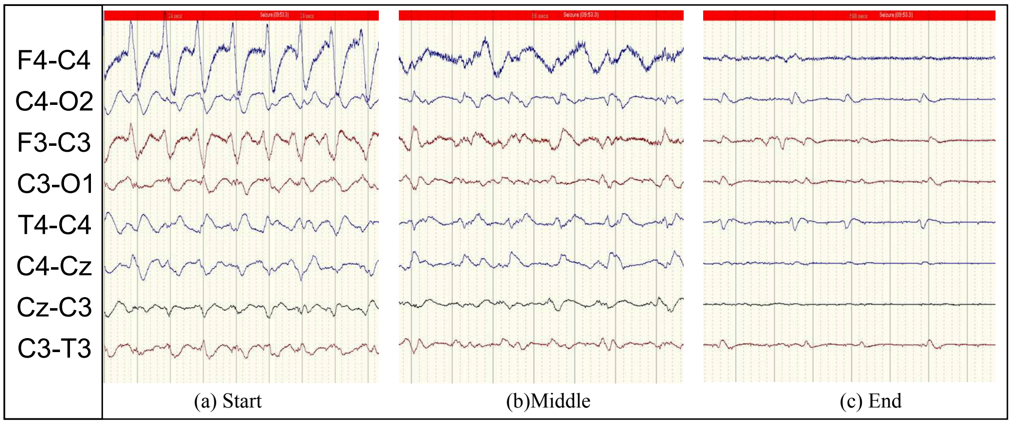

Fig. 1. Evolution of a single seizure event. (a) Slow wave activity at start with sharp/spike components involved high in amplitude, phase reversal. (b) As the seizure progresses the EEG becomes lower in amplitude. (c) only very low amplitude discharges are seen with a flattening of the background.

found. Dynamic programming is used to find the path that gives the shortest DTW distance in the alignment matrix.

B. Dynamic Time Warping in SVM

The SVM has shown good state of the art performance in detecting seizures from the neonatal EEG [10]. SVMs are discriminant classifiers that can transform complex inseparable input data into high dimensional space which can be easily classified using linear discriminant functions. Conventional SVM techniques were developed for a fixed dimension feature vectors which cannot perfectly deal with the dynamic structure and length of neonatal seizures. Although, the feature vectors extracted from each EEG segment are of fixed length but a whole seizure event could result in a variable length sequence of feature vectors. In order to handle this problem, DTW was introduced inside SVM [11].

The classical SVM classifier uses a hyperplane to separate the input data. Consider a two class problem, with a pre-labeled training set $(x_1, y_1), ..., (x_n, y_n)$ where $y_i \in \{-1, +1\}$ and $x_i \in \Re$. In SVM classification, a test vector u is assigned a class y_i by evaluating

$$f(u) = sign\left(\sum_{i}^{n_{sv}} \alpha_i y_i K(u, x_i) + b\right) \quad (2)$$

where α_i are Lagrange multipliers, b is the bias and x_i are n_{sv} support vectors (SV). K is the kernel of the SVM and is used to map the input data into a higher dimensional feature space. A commonly used kernel function is the Gaussian radial basis kernel defined as

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (3)$$

If P and Q are two sequences, then their kernel metric can be found by replacing the euclidean distance in 3 by the DTW distance D_ϕ, [11];

$$K(P, Q) = \exp\left(-\gamma D_\phi(P, Q)\right) \quad (4)$$

This way, SVM with the GDTW kernel will be able to classify the variable length sequences according to their DTW distances and the SV concept is now replaced with support sequences.

III. NEONATAL SEIZURE DETECTION

A. Dataset

The dataset used in this study is taken from the EEG of 18 full term neonates recorded in the NICU of Cork University Maternity Hospital, Cork, Ireland. A Carefusion NicOne video EEG monitor was used to record multichannel EEG at 256Hz using the 10-20 system of electrodes placement. 8 bipolar EEG channels were used (F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3, C3-T3) to annotate the data. Seizures were annotated independently by two experienced neonatal electro-encephalographers. The total length of the all EEG recordings is 816h (hours) which constitutes 1389 seizure events of the duration varying from 10s (seconds) to 40m (minutes). The EEG recordings were not edited and no artifacts have been removed. So, this dataset is true representative of the real time situation in hospitals.

B. SVM Based Seizure Detection System

1) Preprocessing and Feature Extraction: An overview of the whole seizure detection system is shown in Fig.2. The EEG is first down-sampled from 256Hz to 32Hz with an anti-aliasing filter set at 12.8Hz. Each channel of the EEG is then segmented into 8s epochs with a sliding window and 50% overlap. The 55 features used in this study are described in [10]. The usability of these features have been discussed in a number of previous works [10], [12].

2) Classification: After the features are extracted from every epoch, the sequences are formed by grouping the epochs into fixed length sequences of 10 epochs with a shift of one epoch (i.e. a 90% overlap). These sequences are then fed to the GDTW-SVM classifier. All channels are classified separately as shown in Fig. 2. The decision of the computed sequence is averaged with the decision of the one past sequence, which corresponds to the moving average of

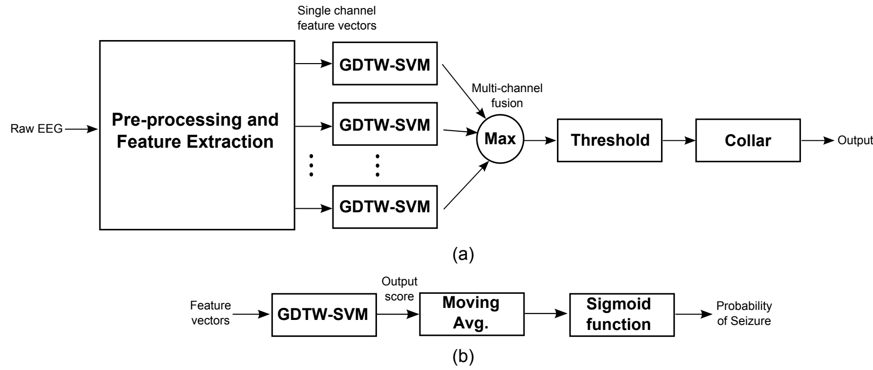

Fig. 2. (a) Overview of the neonatal seizure detection system (b) Overview of the classification stage.

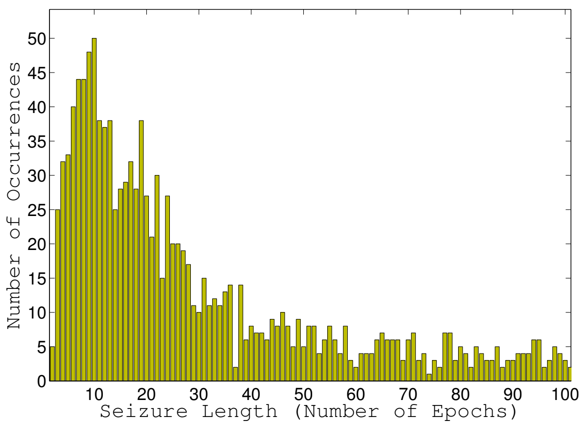

Fig. 3. Histogram showing the length of seizures in the dataset.

two sequence decisions. This means that a decision is made every 4s of the EEG.

3) Post-Processing and Multi-Channel Fusion: The output of the classifier is converted to posterior probabilities as in [10] using Platts method [13]. Then the procedure employed for data annotation is used; if there is a seizure in at least one channel, the whole epoch is marked as seizure, otherwise it is denoted as non-seizure. This corresponds to applying the 'MAX' operator to the probabilistic values of all channels. The output is then compared to a threshold. In order to get the performance curves of the system, the threshold is gradually varied from 0 to 1. Then a binary decision is taken i.e. 1- seizure and 0-non-seizure. A collar technique is applied in which every seizure decision is extended from either side to compensate for the possible difficulties in detecting onset and offset of a seizure event as in [10].

C. Performance Assessment

The classifier was trained on the data of 17 patients from the dataset and tested on the data of one remaining patient that was not used in the training. The length of the EEG recording for testing was 29.7h and had 17 seizure events with a variety of lengths ranging from 17s to 234s. The mean seizure length was 90s.

To quantify the system performance the epoch based sensitivity and specificity, and the event based Good Detection Rate (GDR) and the number of False Detections per hour (FD/h) are reported. Sensitivity is defined as the percentage of correctly detected seizure epochs and specificity is the percentage of correctly detected non-seizure epochs. The Receiver Operating Characteristic (ROC) curve is used to plot the sensitivity and specificity of the system. In contrast to sensitivity, GDR represents the percentage of correctly identified seizure events.

D. Model Selection

In order to find the best GDTW kernel parameter γ and generalization parameter C of SVM, a 2 fold cross validation is applied on the training data of 17 patients. The 20m per patient of EEG for the training data is annotated on per channel basis to get training examples of the seizure class. More details on per channel annotation can be found in [14]. Sequences are formed by concatenating the consecutive epochs that correspond to the same seizure event. As the histogram in Fig. 3 shows that peak is around seizure events of length 10 epochs, therefore a maximum seizure sequence length of 15 epochs was chosen and seizures greater than 15 epochs were chunked down to 10 epochs seizure. Seizure events shorter than 15 epochs were taken directly. Non-seizure sequences were taken with a fixed length of 10 epochs. The total number of seizure and non-seizure sequences used for training the classifier were kept at a 1:1 ratio. After the optimal pair of parameters is found in the internal 2 fold cross validation, the SVM classifier is trained on the 5112 training sequences (2556 seizure, 2556 non-seizure) extracted from the 17 patients.

IV. RESULTS AND DISCUSSION

The performance of the designed GDTW-SVM system and the system proposed in [10] is shown in Fig. 4 and Fig. 5. Quantitatively, the ROC area of 95.41% is obtained by the proposed system in comparison with that of 95.06% obtained by the baseline system [10]. Fig. 5 compares the system in terms of the event-based GDR vs. FD/h metrics. From the figures, it can be seen that the system which exploits temporal evolution characteristics of neonatal seizures reaches

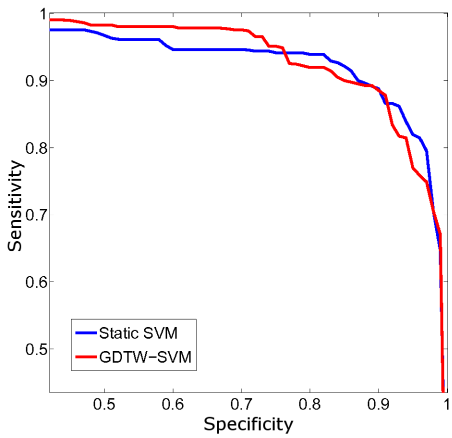

Fig. 4. ROC curves of the GDTW-SVM based and static SVM based neonatal seizure detection system.

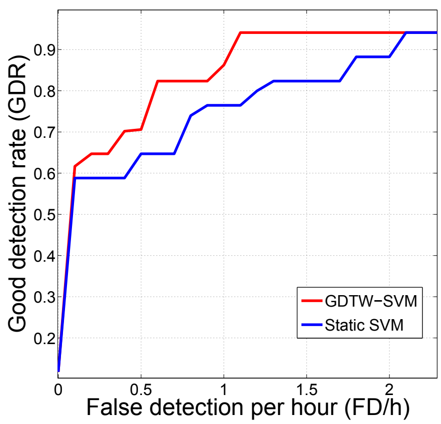

Fig. 5. GDR versus FD/h curve of the GDTW-SVM based and static SVM based neonatal seizure detection system.

similar performance. Both the ROC area and the GDR for the same rate of false detections per hour are improved. With the GDTW-SVM system 86% of the seizure events are correctly identified at the cost of 1 FD/h. At the same cost the static SVM based system reaches a GDR of 78%. It indicates that introducing dynamic kernels increases both the number of detected events and the correctly detected seizure burden (the total time of the neonatal seizure activity).

Fig. 4 shows that the GDTW based system performs better for higher sensitivity values whereas the static SVM based system performs better for higher specificity values. This observed complementary behavior is known to be a good option for successful application of classifier combination techniques. It can be expected that the fusion of the two approaches can significantly improve the current state-of-the art performance in the area of neonatal seizure detection.

In this work, the designed system is able to classify variable length sequences. The proposed system was also tested with the 15 epochs sequences but the results obtained were worse. It is expected that the dynamic choice of testing sequence length will provide an improved detection of both the onset and the offset of the real variable duration seizures.

V. CONCLUSION

A preliminary attempt to incorporate contextual information at the classifier level has been made in this work by means of substituting the static kernels in the SVM with a dynamic kernel. The obtained results are promising. However more thorough evaluation and analysis of GDTW-SVM and other approaches on different levels is necessary to reach solid conclusions. This will be part of our future work.

VI. ACKNOWLEDGMENT

The authors wish to thank Mr. Robert Goulding for providing the figure of evolutionary characteristics of a seizure.

REFERENCES

[1] D. M. Murray, G. B. Boylan, I. Ali, C. A. Ryan, B. P. Murphy, and S. Connolly, "Defining the gap between electrographic seizure burden, clinical expression, and staff recognition of neonatal seizures," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 93, no. 3, pp. F187–F191, 2008.

[2] S. Patrizi, G. L. Holmes, M. Orzalesi, and F. Allemand, "Neonatal seizures: characteristics of eeg ictal activity in preterm and fullterm infants," *Brain and Development*, vol. 25, no. 6, pp. 427 – 437, 2003.

[3] S. Furui, "Speaker-independent isolated word recognition using dynamic features of speech spectrum," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 34, no. 1, pp. 52 – 59, 1986.

[4] A. Temko, E. Monte, and C. Nadeu, "Comparison of sequence discriminant support vector machines for acoustic event classification," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 5, p. V, 2006.

[5] H. Shimodaira, K. I. Noma, M. Nakai, and S. Sagayama, "Dynamic time-alignment kernel in support vector machine.," *Advances in Neural Information Processing Systems 14*, vol. 2, no. 1, pp. 921–928, 2006.

[6] V. Wan and S. Renals, "Speaker verification using sequence discriminant support vector machines," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 2, pp. 203 – 210, 2005.

[7] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257 –286, 1989.

[8] M. Gordan, C. Kotropoulos, and I. Pitas, "A temporal network of support vector machine classifiers for the recognition of visual speech," in *Methods and Applications of Artificial Intelligence* (I. Vlahavas and C. Spyropoulos, eds.), vol. 2308 of *Lecture Notes in Computer Science*, pp. 744–745, 2002.

[9] W. A. Chaovalitwongse and P. M. Pardalos, "On the time series support vector machine using dynamic time warping kernel for brain activity classification," *Cybernetics and Sys. Anal.*, vol. 44, no. 1, pp. 125–138, 2008.

[10] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, "EEG-based neonatal seizure detection with support vector machines," *Clinical Neurophysiology*, vol. 122, no. 3, pp. 464 – 473, 2011.

[11] C. Bahlmann, B. Haasdonk, and H. Burkhardt, "Online handwriting recognition with support vector machines - a kernel approach," in *Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on*, pp. 49 – 54, 2002.

[12] J. Gotman, D. Flanagan, J. Zhang, and B. Rosenblatt, "Automatic seizure detection in the newborn: methods and initial evaluation," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 356 – 362, 1997.

[13] J. C. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," *B Schlkopf C J C Burges A J Smola editors Advances in Kernel Methods Support Vector Learning*, pp. 61–74, 2000.

[14] A. Temko, G. Lightbody, E. Thomas, G. Boylan, and W. Marnane, "Instantaneous measure of eeg channel importance for improved patient-adaptive neonatal seizure detection," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 3, pp. 717 –727, 2012.