33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Online EEG Channel Weighting for Detection of Seizures in the Neonate

Andriy Temko, Gordon Lightbody, Geraldine Boylan, William Marnane

Abstract— **A framework for online dynamic channel weighting is developed for the task of EEG-based neonatal seizure detection. The channel weights are computed on-the-fly by combining the up-to-now patient-specific history and the clinically-derived prior channel importance. These estimated time-varying weights are introduced within a Bayesian probabilistic framework to provide a channel-specific and thus patient-adaptive seizure classification scheme. Validation results on one of the largest clinical datasets of neonatal seizures confirm the utility of the proposed channel weighting for the SVM-based detector recently developed by this research group. Exploiting the channel weighting, the precision-recall area can be drastically increased (up to 25%) for the most difficult patients, with the average increase from 81.0% to 84.42%. It is also shown that the increase in performance with channel weighting is proportional to the time the patient is observed.**

I. Introduction

A system that could automatically detect and annotate seizures on the neonatal EEG would be extremely useful to clinicians in the neonatal intensive care unit (NICU). Although a number of methods and algorithms have been proposed previously in an attempt to automatically detect neonatal seizures [1–3], to date their transition to clinical use has been limited due to poor performance.

Clearly, the performance of the seizure detection systems depends on the information included in the corresponding EEG channels. By using as many channels as possible, one minimizes the probability that useful information is missed. On the other hand, it becomes more difficult to automatically find what information is useful in the available channels. Channel selection has been widely used in Brain Computer Interfaces [4–10] mainly for the purpose of computational load reduction. Constant (in time) importance of information captured by a channel can be assumed for these tasks, especially when classification systems targeted are patient-specific, that is, some representation of testing patient data is available beforehand [4], [7–10].

Patient-specific neonatal seizure detection has limited (if any) clinical utility. In fact, samples of testing patient data are never available beforehand in a real-life situation in the NICU. Although neonatal seizure may generalize, many remain focal or multi-focal, that is, (highly) localized in different parts of the brain depending on the patient. Clearly, patient-specific dependencies between a seizure event and its common location can be learnt in a supervised way over the training data available, however these provide little benefit for an unseen testing patient – a new patient may have a different seizure location distribution. Therefore, a channel weighting procedure which can adapt on-the-fly to the testing patient would be very useful. Such channel weighting can be seen as a generalization of channel selection, where the weights can take other than binary forms.

This work aims at improving the neonatal seizure detector, previously developed by this group, by using all available channels and dynamically weighting them to emphasize the relevant information. In particular, this work first formulates the neonatal seizure detection problem in probabilistic terms using a Bayesian framework to help indicate where such weighting terms should be used. A methodology for the estimation of the time-varying channel weights based on the synchronized energy of the classifier probabilistic outputs is proposed. By emphasizing the patient-specific time-varying seizure locations, the detectors manage to self-adapt to every testing patient on-the-fly in an unsupervised manner. The increase in performance confirms the usefulness and consistency of the proposed method when applied to different channel subsets.

II. Neonatal Seizure Detector

A. Dataset

The dataset is composed of EEG recordings from 17 newborns (267.9h, 705 seizures) obtained from Cork University Maternity Hospital, Ireland. Signals from 9 electrodes (T4, T3, O1, O2, F4, F3, C4, C3, and Cz) were recorded using the 10-20 system of electrode placement and the 8 EEG channels in the bipolar montage (F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3, and C3-T3) were used to annotate the data. The dataset contains a wide variety of seizure types including both electrographic-only and electro-clinical seizures of focal, multi-focal and generalized types. The continuous EEG recordings were not edited to remove the large variety of artifacts and poorly conditioned signals that are commonly encountered in the real-world NICU. All seizures were annotated independently by two experienced neonatal electro-encephalographers using video EEG.

Manuscript received 15/04/2011. This work was supported by the Welcome Trust (085249/Z/08/Z) and SFI (10/IN.1/B3036, 07/SRC/I1169).

Andriy Temko, William Marnane, and Gordon Lightbody are with the Department of Electrical and Electronic Engineering and the Neonatal Brain Research Group, University College Cork (UCC), Ireland. andreyt@eleceng.ucc.ie, {l.marnane, g.lightbody}@ucc.ie

Geraldine Boylan is with the Department of Pediatrics and Child Health and the Neonatal Brain Research Group, UCC, Ireland. g.boylan@ucc.ie

B. Automated seizure detection system architecture

The diagram of the previously developed system [11] is shown in Fig. 1. The EEG from the 8 above-mentioned channels was down-sampled from 256Hz to 32Hz with an anti-aliasing filter set at 12.8Hz. The EEG was then split into 8s epochs with 50% overlap between epochs. Fifty-five features were extracted from each channel which represent both time and frequency domain characteristics as well as information theory based parameters. Details on the features can be found in [11, 12]. The features were fed to a SVM classifier and the outputs of the SVM were converted to probability-like values and smoothed with a moving average filter. The averaged value was then compared to a threshold from the interval [0 1]. After comparison, binary decisions were taken per channel: 1 for seizure and 0 for non-seizure. The binary decisions were then fused using logical 'OR'. It has been shown in [12] that the developed system significantly outperformed the existing alternatives. The same set of SVM models as in [11] was used in this work.

C. Performance Assessment and Metrics

The LOO cross-validation method was used to assess the performance of the system for patient-independent seizure detection [1]. This way, all but one patients' data were used for training/development and the remaining patient's data were used for testing. This procedure was repeated until each patient had been a test subject and the mean result was reported. The metric used in the work is the area under the Precision-Recall curve. Unlike ROC curves, PR curves do not present an overly optimistic view of an algorithm's performance if there is a large skew in the class distribution [13] – which is the case for neonatal seizure detection. While recall is the same as sensitivity (accuracy of a seizure class), precision (also known in seizure detection literature as selectivity, relative specificity, and positive predictive value) is defined as a percentage of correctly produced/predicted seizure epochs. Unlike the ROC area, the PR area is not equal to 0.5 for random discrimination but depends on class priors, that is, the number of datapoints in each class.

III. Bayesian Formulation of Seizure Detection

The obtained SVM model can be applied to any EEG channel, thus the developed system (Fig. 1) is channel-independent. Denoting prior/posterior probabilities by P and conditional probabilities/likelihoods by p, using Bayes' theorem, the posterior probability of having a seizure decision S given a feature vector $\mathbf{x}$ can be written as:

$$P(S \mid \mathbf{x}) = \frac{P(S)\, p(\mathbf{x} \mid S)}{P(\mathbf{X})} \tag{1}$$

Equation (1) is already modeled by the existing detection system. If the channel information, c, is taken into account:

$$P(S \mid \mathbf{x}, c) = \frac{P(S, c)\, p(\mathbf{x} \mid S, c)}{P(\mathbf{X}, c)} \tag{2}$$

where $P(S,c)$ is the joint prior probability of having a seizure S and it manifesting itself on channel c. $P(S,c)$ can be decomposed into having a channel-independent prior probability that the seizure occurs across any of the observed channels, $P(S)$, and the probability that the seizure manifests on channel c given that the seizure occurs, $P(c|S)$. Similarly, since the systems described in Section 2 use the designed models which are channel-independent, the likelihood generated by the model is also channel independent, and thus $p(\mathbf{x}|S,c)$ can be decomposed into channel-independent likelihood, $p(\mathbf{x}|S)$, and a data-dependent prior or weighting of channel c, $P(c|\mathbf{x})$. Equation (2) can thus be expanded as:

$$P(S \mid \mathbf{x}, c) = \frac{P(S)\, p(\mathbf{x} \mid S)}{P(\mathbf{X}, c)} P(c \mid S) P(c \mid \mathbf{x}) \tag{3}$$

The two new terms introduced are channel-specific. Additionally, the first prior probability term does not depend on the observation $\mathbf{x}$, that is, it does not change with time. Both terms can be combined to form a final weight for each channel as will be shown in the next section.

IV. Online Channel Weighting

In our case with n=9 electrodes, there are n*(n-1)/2 = 36 possible channels in the bipolar montage. We first model the importance of a given electrode, e, and then obtain the combined weight for channel, c, in bipolar montage.

A. Modelling P(e|x)

The data-driven estimate (at time t) of the 'importance' of the i^{th} electrode $P(e_i|\mathbf{x})$ can be estimated using the probabilistic output of the classifier on a feature vector $\mathbf{x}$ after the moving average filter (Fig. 1). For every electrode e_i, several channels which cover the brain zone around the electrode of interest and share this electrode are selected.

Let $\mathbf{y}_i(r)$ be a vector of probabilistic output at time r of selected channels. The $P_t(e_i|\mathbf{x})$ for electrode i at time t is then expressed as:

$$P_t(e_i \mid \mathbf{x}) = \frac{\sum_{r=1}^{t} \mathbf{y}_i^T(r)\, \mathbf{Q}\, \mathbf{y}_i(r) \Big/ \left|\mathbf{y}_i(r)\right|}{\sum_{j=1}^{N} \left(\sum_{r=1}^{t} \mathbf{y}_j^T(r)\, \mathbf{Q}\, \mathbf{y}_j(r) \Big/ \left|\mathbf{y}_j(r)\right| \right)} \tag{4}$$

where $|\mathbf{y}_i(r)|$ denotes the number of channels associated with the i^{th} electrode, and $\mathbf{Q}$ is a $|\mathbf{y}_i(r)| \times |\mathbf{y}_i(r)|$ square matrix

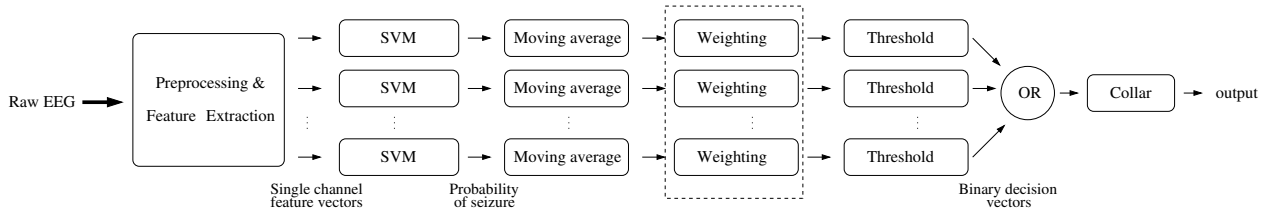

Fig. 1. Neonatal seizure detector system diagram.

$$\mathbf{Q} = \begin{bmatrix} 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ \ddots & & & & \\ 0 & 0 & \dots & & 0 \end{bmatrix} \quad (5)$$

Essentially, the average of the product-moments or cross-correlation at 0^{th} lag between selected channels, which share the same electrode, is calculated here to obtain a measure of agreement between probabilistic activities in a certain electrode at a point in time, r. The cumulative sum in the numerator of (4) represents the integrated synchrony up to the current point in time, t. To assure a partition of unity of $P(e|\mathbf{x})$ over all possible $\mathbf{x}$, the cumulative sum for a particular electrode, e_i, is normalized across all the electrodes.

The integrated synchrony will be high when there is a synchronous rise in probability of seizure in all considered channels that share the chosen electrode. In turn, a synchronized high probability activity is indicative of a seizure. Effectively, this measure emphasizes the electrode/location in the brain which had a history of suspected seizures. Alternatively speaking the measure incorporates the fact that any new seizures are more likely to happen at the location where seizure activity has been observed before. An example of the proposed measure is shown in Fig. 2 for patient 8. For this patient, most seizures are localized in the C4 electrode, which is reflected in the increased integrated synchrony for this electrode. The normalized measure of importance, shown in the middle, indicates that channels which contain electrode C4 will be approximately 2 times more emphasized than the other channels during seizure. It can be seen that the emphasis of the electrode C4 increases during ictal activity and slowly decreases during interictal periods of time.

B. Modelling P(e|S)

The measure $P(e|S)$ is defined as the probability that, given a seizure is occurring, that it is visible in electrode e. It aims at emphasizing a priori, electrodes in which seizures are mostly expected. Unlike the channel-independent prior $P(S)$, which can be modeled based on the training data annotation, the electrode-dependent prior $P(e|S)$ requires per channel annotations which are not available. Thus, $P(e|S)$ is estimated here from the statistics found in the clinical literature. In particular, in [14], it has been shown that as many as 78% of seizures were visible in the C3, C4 zone. In [15], it has been reported that the theoretical visibility of a seizure in the central zone is as high as 94%. In another study [16], it has been shown that 46% of seizures are visible in the Fp1, Fp2 zone, which is close to F3, F4 in our montage. No data were found regarding the visibility of seizures in the temporal or occipital zones, and there are no premises to believe that a seizure is more visible in the occipital and temporal zones than in the frontal zone. Based on the data collected from the literature, the $P(e|S)$ is represented here as 0.46, 0.46, 0.46, 0.46, 0.46, 0.46, 0.78, 0.78, and 0.94 for electrodes T4, T3, O1, O2, F4, F3, C4, C3, and Cz, respectively.

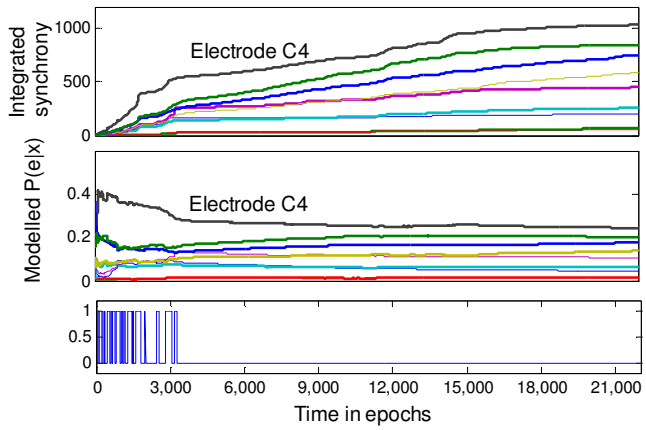

Fig. 2. An example of the proposed measure of electrode importance calculated for patient 1. The top graph shows the integrated synchrony for each electrode. The middle graph plots the resultant $P(e|\mathbf{x})$. The bottom graph plots the ground truth where 1 indicates 'seizure'.

C. Combined Channel Weight

To control the severity of channel weighting, the softmax function is applied here:

$$w_i(t) = \frac{\exp\left(k\, P_t(e_i \mid \mathbf{x})\, P(e_i \mid S)\right)}{\sum_{j=1}^{N} \exp\left(k\, P_t(e_j \mid \mathbf{x})\, P(e_j \mid S)\right)} \quad (6)$$

where t is the current time, N is the number of electrodes, k is a multiplication constant which controls the severity of weighting. If k is large, a single non-zero weight is obtained for the most important electrode and the scheme converges to channel selection, where only those channels which contain this electrode are selected.

The two terms $P(e|S)$ and $P(e|\mathbf{x})$ have been modeled so far for every electrode. Looking at the two electrodes associated with channel c, a final measure per channel in the bipolar montage, at time, t, is calculated as the maximum of measures for the two constituent electrodes:

$$P(c \mid S) P(c \mid \mathbf{x}(t)) \approx \max_{e_i \in c} (w_i(t)) \quad (7)$$

After the weight is calculated in (7), it is used to estimate the final probability in (3). This way, the current probability of a seizure in a channel (evidence) is multiplied by a channel weight (confidence) which is obtained on-the-fly by combining the up-to-now patient-specific history and the clinically-derived prior channel importance.

V. Results and Discussion

A. The effect of channel weighting

To show the effect of the channel weighting for various channel subsets, backward channel elimination is performed in a similar manner to that usually used for feature selection. This procedure effectively provides nested subsets of channels. The performance with and without channel weighting is shown in Fig. 3. The resultant sequence of channels was obtained over the development data and then tested on the testing data. Initially, all 36 channels were included and one channel was eliminated at a time. The

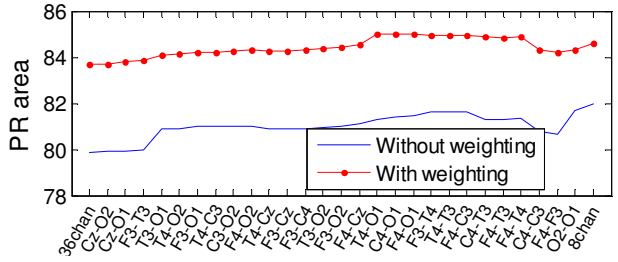

Fig. 3. Performance of the SVM-based seizure detection system with and without channel weighting for various channel subsets

channel was chosen to be eliminated if the difference of performance of a current set of channels with and without it led to the highest improvement or the smallest decrease over the development data. This eventually terminates with the 8 channels which were originally used for data annotation. Having these 8 original channels (which were not considered for elimination in the channel selection routine) as a final point, it was assured that all annotated seizures were potentially detectable. It can be seen that the system with channel weighting consistently outperforms the same system without channel weighting for all subsets of the chosen channels. The average improvement in terms of PR areas is 18% relative, growing from 81.0% to 84.42%.

B. Patient-specific improvement over time

Fig. 4 shows the absolute average difference of the PR areas for 17 patients in the dataset. It can be seen from Fig. 4 that the average increase in the PR area (3.4%) is attributable to significant (up to 25% absolute) performance increases in certain patients. For most patients the channel weighting has almost no effect whereas increases in PR area for patients 1, 2, 5, 7 and 8 are particularly large. In the real clinical situation, it is most difficult and much more important to detect seizures when they are rare and/or focal events (patients 1, 2, 7, and 8), than to miss several seizures in patients which are close to status-epilepticus (patient 3, 4, 9, 14), that is, where seizures happen constantly over time.

The bottom graph of Fig. 4 plots the time evolution of the PR area for patient 8 (as a typical example) as well as the calculated difference with and without weighting. The PR areas and the difference were re-calculated every hour. It can be observed that the difference in performance between the systems with and without channel weighting increases with time. In particular, during the first 4 hours of observation, the difference in performance is only ~5%, whereas at 17 hours it is already ~21%.

Weighting of channels is proposed for improved neonatal seizure detection. It is shown in this work that the largest benefit in performance is expected in the most difficult clinical situations where it is necessary to detect rare focal events in long term monitoring. The proposed measure of channel importance is completely data-driven and computed online in an unsupervised manner, which allows for its usage for other neurological applications which involve EEG monitoring.

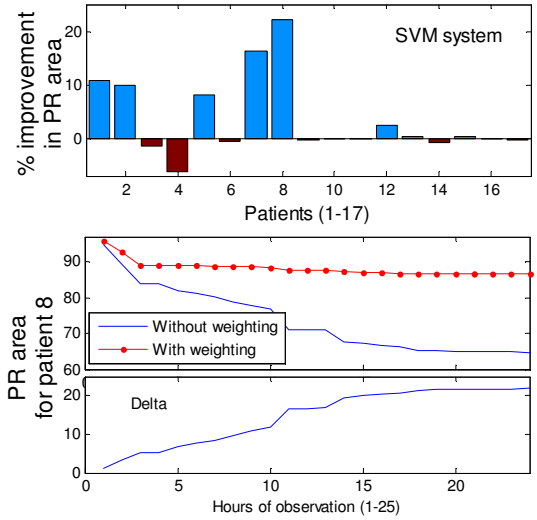

Fig. 4. Per-patient difference in PR areas for the SVM-based system (top), the evolution of the PR area in time for patient 8 with and without weighting (middle) and the calculated delta (bottom).

REFERENCES

[1] J. Mitra, J. Glover, P. Ktonas, A. Kumar, A. Mukherjee, N. Karayiannis, J. Frost, R. Hrachovy, E. Mizrahi, "A Multistage System for the Automated Detection of Epileptic Seizures in Neonatal Electroencephalography", *J Clin Neurophysiol*, v. 26, pp. 1-9, 2009.

[2] P. Celka, P. Colditz, "A computer-aided detection of EEG seizures in infants, a singular-spectrum approach and performance comparison", *IEEE Tran Biomed Eng*, v. 49, pp. 455–462, 2002.

[3] W. Deburchgraeve, P. Cherian, M. Vos, R. Swarte, J. Blok, G. Visser, P. Govaert, S. Huffel, "Automated neonatal seizure detection mimicking a human observer reading EEG", *Clin Neurophysiol*, v. 119, pp. 2447-2454, 2008.

[4] J. Fan et al., "Automatic seizure detection based on support vector machines with genetic algorithms", in *Simulated Evolution and Learning, LNCS*, v. 4247, pp. 845-852, 2006.

[5] M. Schröder et al., "Robust EEG channel selection across subjects for brain-computer interfaces", *EURASIP J App. Sig. Pr.* v.19, 2005.

[6] T. Lan et al., "Salient EEG channel selection in brain computer interfaces by mutual information maximization", in *Proc. IEEE EMBC*, pp. 7064-7067, 2005.

[7] T. Lal et al., "Support vector channel selection in BCI", *IEEE Tr Biomed Eng*, v. 51, i. 6, pp. 1003-1010, 2004.

[8] G. Garipelli et al., "Discriminative channel selection method for the recognition of anticipation related potentials from CCD estimated cortical activity", in *Proc. IEEE MLSP*, p.1-6, 2009.

[9] K. Ansari-Asl et al., "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood", In *Proc. EUSIPCO*, pp. 1241-1245, 2007.

[10] Y. Wang et al., "Common spatial pattern method for channel selection in motor imagery based brain-computer interface", in *Proc. IEEE EMBS,* pp. 5392-5395, 2005.

[11] A. Temko, E. Thomas, W. Marnane, G. Lightbody, G. Boylan, "EEG-based neonatal seizure detection with support vector machines", *Clin Neurophysiol*, v. 122, i. 3, pp. 464-473, 2011.

[12] A. Temko, E. Thomas, W. Marnane, G. Lightbody, G. Boylan, "Performance assessment for EEG-based neonatal seizure detectors", *Clin Neurophysiol,* v. 122, i. 3, pp. 474-482, 2011.

[13] J. Davis, M. Goadrich, "The Relationship between Precision-Recall and ROC Curves", in *Proc. ICML*, pp. 233-40, 2006.

[14] R. Shellhaas et al., "Sensitivity of amplitude-integrated electroencephalography for neonatal seizure detection", *Pediatrics*, v. 120, pp. 770-777, 2007.

[15] R. Shellhaas, R. Clancy, "Characterization of neonatal seizures by conventional EEG and single-channel EEG", *Clin Neurophysiol*, v. 118, i. 10, pp. 2156-2161, 2007.

[16] C. Wusthoff et al., "Limitations of single-channel EEG on the forehead for neonatal seizure detection", *J Perinatol*, v. 29, 2009.