35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

A Boosted Cascade for Efficient Epileptic Seizure Detection

Tingting Ge[1], Yu Qi[1], Yueming Wang[1], Weidong Chen[1], and Xiaoxiang Zheng[2], *Member, IEEE*

Abstract— **Seizure detection from electroencephalogram (EEG) plays an important role for epilepsy therapy. Due to the diversity of seizure EEG patterns between different individuals, multiple features are necessary for high accuracy since a single feature could hardly encode all types of epileptiform discharges. However, a large feature set inevitably causes the increase of the computational cost. This paper proposes a boosted cascade chain to obtain both high detection performance and high computational efficiency. Sixteen features that are widely used in seizure detection are implemented. Considering the sequential characteristics of EEG signals, the features are extracted on each 1-second segment and its former three segments. Thus, a total of 64 features are used to construct a feature pool. Based on the feature pool, Real AdaBoost is used to select a group of effective features, on which weak classifiers are learned to assemble a strong classifier. The strong classifier is transformed to a cascade classifier by reordering the weak classifiers and learning a threshold for each weak classifier. The cascade classifier still has the similar classification strength to the original strong classifier. More importantly, it is able to reject easy non-seizure samples by the first a few weak classifiers in the cascade, thus high computational efficiency can be obtained. To evaluate our method, 90.6-hour EEG signals from four patients are tested. The experimental results show that our method can achieve an average accuracy of 95.31% and an average detection rate of 91.29% with the false positive rate of 4.68%. On average, only about 4 features are used. Compared with support vector machine (SVM), our method is much more efficient with the similar detection performance.**

I. Introduction

Epilepsy is a common neurological disorder that affects 50 million people worldwide. When seizure is onset, epileptiform discharges including slow waves and closely-spaced spikes can be observed in EEG signal. This makes it possible to detect the seizure of epilepsy by signal processing methods and enable a block prevention system for the therapy of epilepsy.

In an epileptic prediction system, one important step is to distinguish the normal states from the seizure states. Various features have been proposed to address this problem, including temporal and frequency methods [1], energy [2], spectral analysis [3], wavelet transformation [4], sub-space analysis [5], Empirical Mode Decomposition [6], entropy [7], and the complexity [8]. Most of them can achieve good performance on a specific sample set. However, since the great diversity of seizure patterns, almost none of them has the ability to detect all types of epileptiform discharges. One solution is to use them all in order to fuse the different strengths of the multiple features. This causes the increase of computational complexity in classification, which should also be seriously considered in a practical seizure detection system.

Thus, to obtain both high detection accuracy and high computational efficiency, it is important to select a few effective features from the multiple features. AdaBoost is a popular algorithm capable of performing feature selection and has been used for epilepsy detection in recent years. Manohar et al. [10] and Amal [11] used AdaBoost for feature selection and achieved high performance in seizure prediction. Although the AdaBoost algorithm can select weak features to learn weak classifiers and combine them to a strong classifier, when there are hard classified samples in the training set and a low false alarm is required, the strong classifier may have a large number of features (weak classifiers). This still leads to expensive computation.

Based on the observation that, although a few hard samples commonly need a large number of weak classifiers to determine the labels, most easy samples can be classified by only a small number of weak classifiers. This inspires a boosted cascade classifier for both high classification accuracy and efficiency. In this study, first we implement 16 different features and extend them to 64 features by considering the adjacent time windows. Then a strong classifier is learned using Real AdaBoost, which combines a sequence of weak classifiers. After that, the weak classifiers are reordered depending on their discriminative abilities and a reject threshold is learned for each weak classifier. This generates the final cascade classifier. The thresholds are conservative enough so that only easy samples are rejected in the beginning stages and hard samples would walk through the whole cascade for final decision. Since most of the EEG segments can be classified within the first several stages, the average detection time is reduced. To test the performance of our method, 90.6 hours of EEG signals from four patients are used. The results show that, the performance of our method is comparable to SVM, while the computational efficiency is improved by a ratio of 90% in seizure detection.

II. Methods

Our method includes three major components, i.e., feature extraction and feature pool construction, strong classifier learning by Real AdaBoost, and cascade classifier learning. Figure 1 shows the framework of our method.

A. Features and Feature Pool

EEG signals are divided to segments by a window with the size of 1 second. The window slides long the time axis with a stride of 0.5 second. So there is a 0.5-second overlap between adjacent segments. Considering the sequential characteristics

* This work was supported by National High Technology Research and Development Program of China (No. 2012AA020408), National Natural Science Foundation of China (No. 61103107), and Research Fund for the Doctoral Program of Higher Education of China (No. 20110101120154).

[1]Tingting Ge, Yu Qi, Yueming Wang, and Weidong Chen are with the Qiushi Academy for Advanced Studies, Zhejiang University, Hangzhou 310027, China. Yueming Wang is the corresponding author (Email: qaas@zju.edu.cn).

[2]Xiaoxiang Zheng is with the Qiushi Academy for Advanced Studies, and also with the Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China. Email: qaas@zju.edu.cn.

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

of EEG signals, sixteen features are extracted on each 1-second segment and its former three segments. These features are used to form a feature pool for detection. The following sixteen features are used.

1) Temporal and frequency analysis. The power feature and spectral analysis are frequently used for seizure detection. In this work, we extract the total signal power (P), frequency edges (F_e) at 80%, and the power in specific bands as three kinds of features. The frequency edge F_e at 80% refers to the frequency that 80% of the total power is contained in frequencies lower than it [13]. The frequency bands are separated to 1-4Hz, 4-8Hz, 8-12Hz, 12-30Hz, 30-58Hz and 62-125Hz. So totally 8 features are computed.

Besides, Empirical Mode Decomposition (EMD) has been proven effective for feature enhancement in [6]. Here, we compute the variance of the first four intrinsic mode functions (VoIMFs) as 4 features.

2) Dynamic characteristics. The permutation entropy (PermH), sample entropy (SampEn), and Lempel-Ziv complexity (LZC) are used to represent dynamic characteristics of EEG. In our experiment, the permutation order n of PermH is set to 3. For SampEn, the matching length m is set to 2 and the tolerance $r = 0.2 * SD$, in which SD is the standard deviation.

3) Morphological characteristics. Epileptiform discharges exhibit slow waves and closely-spaced spikes in morphology. Here, we compute the variance of the range between the upper and lower envelopes of signal (VoE) as a feature. The details of VoE can be found in [14].

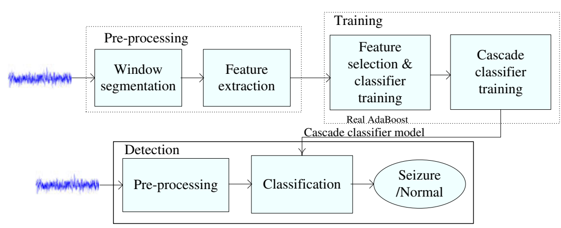

Figure 1. The framework of our method.

B. Learning a Strong Classifier by Real AdaBoost

Based on the feature pool, we use Real AdaBoost to select effective features on which weak classifiers are learned and combined to a strong classifier. Real AdaBoost aims at finding a strong classifier $H_t(x) = \sum_{i=1,2,...,t} c_i(x)$ where x is a test sample and $c_t(x)$ is a weak classifier outputting a real value [15]. The key point is how to learn $H_t(x)$ for seizure detection.

Let $\{(x_1, y_1),...,(x_{a+b}, y_{a+b})\}$ be a training set, where x_i denotes a training sample (segment), y_i is the label in the set of {+1, -1} and +1/-1 denotes the seizure/normal label. First, the algorithm initializes weights $D_0(i) = 1/(2a)$ for positive samples and $D_0(i) = 1/(2b)$ for negative samples. Then an iterative procedure is applied to choose t best features and use them to learn t weak classifiers $c_i(x)$.

On each iterative round, a decision stump is used to construct the weak classifier for each feature. For each feature F_j, the feature values of all samples are partitioned into two parts by a threshold θ and the weak classifier $c_j(x)$ has the form,

$$c_j(x) = \begin{cases} \frac{1}{2}\ln\frac{W_+^1}{W_-^1}, & if \quad f < \theta \\ \frac{1}{2}\ln\frac{W_+^2}{W_-^2}, & else \end{cases} \tag{1}$$

where W_+^k, W_-^k represent the sum of weights of positive samples and negative samples falling into the partitioned parts, respectively. The choice of the threshold θ satisfies the condition of minimizing $Z_j = 2\sum_k \sqrt{W_+^k W_-^k}$. Among all features, the classifier $c_j(x)$ that issues the minimal Z_j is chosen as the best weak classifier $c_t(x)$ on the current round.

At the end of each round, we update the weights of all samples by $D_{t+1}(i) = D_t(i)\exp(-y_i c_t(x_i))$ and apply the normalization operation to make the sums of weights of positive and negative samples equal to 0.5, respectively. After t rounds, we obtain a set of weak classifiers and the strong classifier $H_t(x)$. In practice, we set t to 50 which is able to obtain satisfying performance.

C. Learning a Cascade Classifier

Although the strong classifier $H_t(x)$ is capable of distinguish normal segments from seizure segments effectively, it need to evaluate t features to make a decision, which leads to high computational cost. If we consider relationship between the cumulative sum $H_t(x) = \sum_{i=1,2,...,t} c_i(x)$ and the stage t as shown in Figure 2, we find that some easy samples can be classified in the early stages and only on hard samples, the classifier need to evaluate all weak classifiers. This inspires to construct an efficient cascade classifier based on $H_t(x)$. Similar idea can be found in object detection research [12].

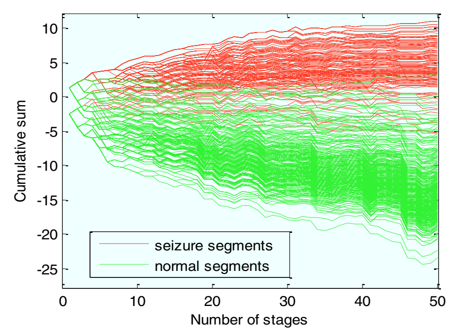

Figure 2. The trace of cumulative sum $H_t(x)$

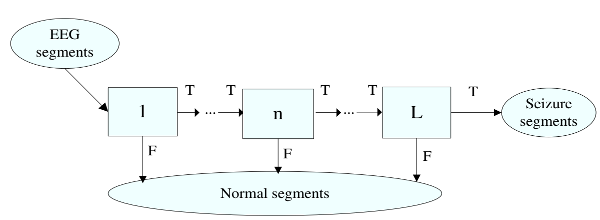

Figure 3. The structure of the cascade classifier

The structure of the cascade classifier is shown in Figure 3. The key points in the cascade classifier is to put the most discriminative weak classifiers to the front of the cascade and assign a threshold r_k to each classifier, so that most easy negative samples can be rejected in the early stages of the cascade, thus high computational efficiency can be obtained. The cascade classifier can be learned on a validation sample set, as shown in algorithm 1.

Algorithm 1

Input:

A validation set $(x_1, y_1), ..., (x_N, y_N)$, $y_i = 0,1$ for negative and positive samples;

The strong classifier $H_t(x) = \sum_{i=1,2,...,t} c_i(x)$ learned via Real AdaBoost;

$m_1, m_2, ..., m_t$, where m_k is the maximum rate of positive samples that can be rejected on the $i-th$ stage.

Initialize:

Cumulative sum of $H_t(x)$ for each sample, $d_{0,i} = 0$

Positive samples rejection rate $p = 0$

For $k = 1, 2, ..., t$

- $p = p + m_k$, $a_k = \#\{i \mid y_i = 1\}$, $b_k = \#\{i \mid y_i = 0\}$;
- select a weak classifier with ID $q(k)$ from $H_t(x)$ that maximizes the separation between the positive and negative samples:
  $q(k) = \arg\max_j \sum_i (d_{k\text{-}1,i} + c_j) y_i / a_k$ -
  $\sum_i (d_{k\text{-}1,i} + c_j)(1\text{-}\, y_i) / b_k$
- update sample traces: $d_{k,i} = d_{k-1,i} + c_{q(k)}(x_i)$;
- determine the threshold r_k as the maximum one that removes no more than p rate of positive samples;
- update $p = p - p_{true}$, where p_{true} is the true positive samples rejection rate under r_k
- remove the sample x_i from sample set, the trace of x_i meet $d_{k,i} \le r_k$

Output:

All output function $c_{q(k)}$ and threshold r_k

The rejection vector $\mathbf{m} = \{m_1, m_2, ..., m_t\}$ is set evenly according to the target detection rate (95% in our experiments). The determination of the thresholds in the training procedure is conservative that guarantees the detection rate does not change, compared with the original strong classifier $H_t(x)$. The false alarm can be reduced in proceeding of the cascade.

In the test process, if the cumulative sum of a sample on stage k is less than threshold r_k , this sample will be rejected and determined as a normal EEG signal. Otherwise, the sample flows into next stage for evaluation until it passes the final stage where it is classified as the seizure segment.

III. Experiment and Result

A. Materials

To evaluate our method, we collect 90.6-hour EEG Data from four patients in the Second Affiliated Hospital, Zhejiang University School of Medicine. A total of 32-channels scalp EEG signals were recorded by NicoletOne amplifier at a sample rate of 256Hz. 50Hz notch filter has been applied in the acquisition. In our experiment, only one channel data is used for epilepsy detection according to doctors' suggestions.

For each patient, the data is divided to 3 parts. The first two parts contain about 150 seizure segments and 10,000 normal segments, which are used for learning the strong classifier and cascade classifier. The rest part is used for testing.

B. Evaluating the number of weak classifiers

This experiment evaluates the effect of the number of weak classifiers in the strong classifier learned by Real AdaBoost. The Receiver Operating Characteristics (ROC) curve reflects the relation between detection rate and false positive rate, and a larger area below ROC curve indicates a better detection performance. We plot the (ROC) curves of strong classifiers with different weak classifiers. As shown in Figure 4, the performance is improved along with the increase of weak classifiers. The improvement from 6 weak classifiers to 30 weak classifiers is significant, while from 30 to 50 weak classifiers, the increase of the performance becomes slow. So 50 weak classifiers should be enough for accurate epilepsy detection.

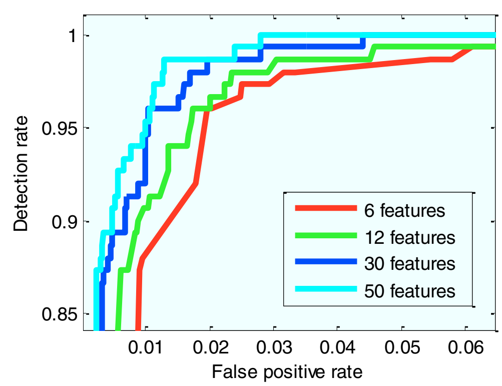

Figure 4. ROC curves of classifiers with different weak classifiers (features).

C. Evaluating the Performance of Cascade Classifier

This experiment evaluates the detection performance of the cascade classifiers. The accuracy (AC), detection rate (DR), and false positive rate (FR) are used for assessment. To verify the classification ability of our cascade classifiers, we compare the performance between our method and SVM. For SVM, the MATLAB version of LIBLIEAR library [17] is used and all 64 features are employed to train the classifier. The L2 loss function is used and the parameters are tuned for optimal performance with cross-validation method.

As shown in TABLE I, the boosted cascade classifier performs well over 4 patients with an average accuracy of 95.31% and an average detection rate of 91.29% with the average false positive rate of 4.68%. Compared with the SVM using all 64 features, the performance of our method is comparable to it.

TABLE I. The comparison between our method and SVM

	Boosted Cascade (Ours)			SVM		
	AC (%)	DR (%)	FR (%)	AC (%)	DR (%)	FR (%)
P1	98.59	91.33	1.39	98.35	85.33	1.63
P2	92.68	92.33	7.31	98.43	86.08	3.32
P3	97.61	96.05	2.39	98.61	84.21	1.38
P4	92.35	85.43	7.63	95.52	81.46	4.47
Average	95.31	91.29	4.68	97.72	84.24	2.7

TABLE II. THE AVERAGE NUMBER OF FEATURES ON THE SIGNALS OF FOUR PATIENTS

	P1	P2	P3	P4	Average
#Stages	1.924	7.474	3.079	6.772	4.812

D. Evaluating the Efficiency of Cascade Classifier

In the previous experiment, our cascade classifier has achieved high detection performance. This experiment assesses the computational cost of our method. Due to the cascade and rejection thresholds in every stage, most non-seizure segments can be rejected on early stages and do not need to pass through all weak classifiers, thus the time cost is reduced. We evaluate the average number of stages and the average time cost for detecting one segment by our cascade classifier.

1. Average number of stages

TABLE II shows the results for the four patients. The average stage number to be used is only about 5, which means lots of feature computation time can be reduced. Note that the SVM classifier needs to evaluate 64 features. In addition, the number of stages used varies over patients because of the diversity of epileptic discharges. For example, for patient 1, only about 2 stages are required on average for classification. While for patient 2, about 7 stages are used on average.

2. Average Time Cost

The experiments are carried out on a common PC with a 2.33GHz Intel Core Duo i3 CPU and 2G RAM. Figure 5 shows the average time cost for detecting one segment. Our boosted cascade classifier only needs 33.6 milliseconds on average to detect a segment, which is less than 10% of the SVM classifier.

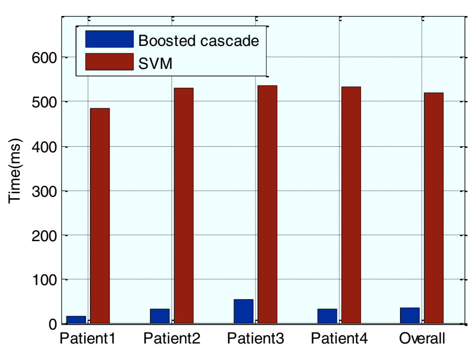

Figure 5. The comparison of time cost of one-segment detection between our method and SVM.

IV. CONCLUSION

In this paper, we proposed a boosted cascade method to detect epileptic seizure. This method is effective with high detection performance comparable to SVM with 64 features, while the average time cost is only about 10% of linear SVM classifier. This method is suitable for both offline seizure segment labeling and online real-time seizure detection tasks.

REFERENCES

[1] K. K. Majumdar, P. Vardhan, “Automatic seizure detection in ECoG by differential operator and windowed variance,” *IEEE Trans. Neural System and Rehabilitation Engineering,* 2011. 19(4): pp. 356-365.

[2] A. G. Correa, E. Laciar, L. Orosco, et al., “An energy-based detection algorithm of epileptic seizures in EEG records,” *in Engineering in Medicine and Biology Society, 2009. EMBC 2009.* Annual International Conference of the IEEE. 2009.

[3] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, “Epileptic Seizure Detection in EEGs Using Time–Frequency Analysis,” *IEEE Trans. Information Technology in Biomedicine*, 2009. 13(5): pp. 703-710.

[4] O.Hasan, “Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy,” *Expert Systems with Applications*, 2009. 36(2, Part 1): pp. 2027-2036.

[5] S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr, “Principal Component Analysis-Enhanced Cosine Radial Basis Function Neural Network for Robust Epilepsy and Seizure Detection,” *IEEE Trans. Biomedical Engineering*, 2008. 55(2): pp. 512-518.

[6] S. Cui, X. Li, X. Gao, et al., “Detection of Epileptic Spikes with Empirical Mode Decomposition and Nonlinear Energy Operator,” *Advances in Neural Networks C ISNN 2005*. 2005, Springer Berlin / Heidelberg. pp. 990-995.

[7] A. S. Zand, G. A. Dumont, M. Javidan, et al., “An entropy-based approach to predict seizures in temporal lobe epilepsy using scalp EEG,” *Engineering in Medicine and Biology Society*, 2009. EMBC 2009. Annual International Conference of the IEEE, pp. 228-231.

[8] D. Abasolo, C. J. James, R. Hornero, “Non-linear Analysis of Intracranial Electroencephalogram Recordings with Approximate Entropy and Lempel-Ziv Complexity for Epileptic Seizure Detection,” *Engineering in Medicine and Biology Society*, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, pp. 1953-1956.

[9] Y. Freund and Robert E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, 1997. 55(1): pp. 119–139.

[10] M. Ayinala, K. K. Parhi, “Low complexity algorithm for seizure prediction using Adaboost,” *Engineering in Medicine and Biology Society (EMBC)*, 2012. Annual International Conference of the IEEE, pp. 1061-1064.

[11] A. Feltane, G. Faye Boudreaux-Bartels and W. Besio, “Automatic Seizure Detection in Rats Using Laplacian EEG and Verification with Human Seizure Signals,” *Annals of Biomedical Engineering*, 2012. pp. 1-10.

[12] L. Bourdev, J. Brandt, “Robust object detection via soft cascade,” *Computer Vision and Pattern Recognition*, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2, pp. 236-243.

[13] C. C. Jouny, G. K. Bergey, “Characterization of early partial seizure onset: Frequency, complexity and entropy,” *Clinical Neurophysiology*, 2012. 123(4): pp. 658-669.

[14] Y. Qi, Y. Wang, X. Zheng, et al., “Efficient epileptic seizure detection by a combined IMF-VoE feature,” *Engineering in Medicine and Biology Society (EMBC)*, 2012. Annual International Conference of the IEEE, pp. 5170-5173.

[15] R. E. Schapire, Y. Singer, “Improved Boosting Algorithms Using Confidence-rated Predictions,” *Machine Learning*, 1999. 37(3): pp. 297-336.

[16] L. Chisci, A. Mavino, et al., “Real-Time epileptic seizure prediction using AR models and support vector machines,” *IEEE Trans. Biomedical Engineering*, 2010. 57(5): pp. 1124-1132.

[17] R. E. Fan, K. W. Chang, C.J. Hsieh, et al., “LIBLINEAR: A Library for Large Linear Classification,” *The Journal of Machine Learning Research*, 2008, vol. 9, pp. 1871-1874.