Research Article

Epileptic Seizure Prediction by a System of Particle Filter Associated with a Neural Network

Derong Liu,[1] Zhongyu Pang,[2] and Zhuo Wang[2]

[1] *The Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[2] *Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607-7053, USA*

Correspondence should be addressed to Derong Liu, derong.liu@ia.ac.cn

None of the current epileptic seizure prediction methods can widely be accepted, due to their poor consistency in performance. In this work, we have developed a novel approach to analyze intracranial EEG data. The energy of the frequency band of 4–12 Hz is obtained by wavelet transform. A dynamic model is introduced to describe the process and a hidden variable is included. The hidden variable can be considered as indicator of seizure activities. The method of particle filter associated with a neural network is used to calculate the hidden variable. Six patients' intracranial EEG data are used to test our algorithm including 39 hours of ictal EEG with 22 seizures and 70 hours of normal EEG recordings. The minimum least square error algorithm is applied to determine optimal parameters in the model adaptively. The results show that our algorithm can successfully predict 15 out of 16 seizures and the average prediction time is 38.5 minutes before seizure onset. The sensitivity is about 93.75% and the specificity (false prediction rate) is approximately 0.09 FP/h. A random predictor is used to calculate the sensitivity under significance level of 5%. Compared to the random predictor, our method achieved much better performance.

1. Introduction

Epilepsy is a brain disorder in which neurons in the brain produce abnormal signals. One explanation for epilepsy is that neuronal activity of human brain has two patterns. One is the normal pattern which corresponds to normal activities while the other is abnormal pattern in which epilepsy is included. Neuronal activity of epilepsy can cause various abnormal situations such as strange sensations, emotions and behavior and loss of consciousness. Possible reasons causing epilepsy are not unique. Seizure and epilepsy are not completely equivalent. That is, a person having a seizure does not necessarily mean that he/she has epilepsy. According to the medical definition of epilepsy, the condition is that a person with epilepsy should have two or more seizures in a time period.

Based on information from the National Institutes of Health, about 1 in 100, or more than 2 million people in the United States, has experienced an unprovoked seizure or been diagnosed with epilepsy. About 20% of people with epilepsy will continue to experience seizures even with the best available treatment [1].

EEG can be used to record brain waves detected by electrodes placed on the scalp or on the brain surface. This is the most common diagnostic test for epilepsy and can detect abnormalities in the brain's electrical activity. Some nonlinear measurement methods such as dimensions, Lyapunov exponents, and entropies were shown to offer new information about complex brain dynamics and further to predict seizure onset.

Iasemidis et al. [2, 3] were pioneers in making use of nonlinear dynamics to analyze clinical epilepsy. Their method was based on the assumption that there was a transition from normal brain activity to a seizure occurrence. Thus, state changes could indicate seizure occurrence. In 2003 [1], they showed that it was possible to predict seizures minutes or even hours in advance by using the spatiotemporal evolution of shortterm largest Lyapunov exponent on multiple regions of the cerebral cortex, since seizure could be characterized by similarity of chaotical

degree of their dynamical states. Later on, an adaptive seizure prediction algorithm was developed to analyze continuous EEG recordings with temporal lobe epilepsy for the purpose of prediction when only the occurrence of the first seizure is known [4].

There are many researchers who are working in this field and many publications have appeared. Ebersole [5] summarized some seizure prediction methods from the First International Collaborative Workshop on Seizure Prediction (2005). He believed that no seizure forewarning has been realized into the clinic. Hassanpour et al. [6] estimated the distribution function of singular vectors based on the time frequency distribution of an EEG epoch to detect the patterns embedded in the signal. Then they trained a neural network and further discriminate between seizure and nonseizure patterns. Mormann et al. [7] summarized some prediction methods and pointed out some of their pitfalls. They also summarized the current state of this research field and possible future development. In order to improve the performance of an algorithm, a better understanding of the inter-ictal period is necessary and all of its confounding variables should influence the characterizing measures used in the algorithms. They mentioned that a further promising approach would be to model EEG signals to gain insight into the dynamical processes involved in seizure generation [8], [9]. For purpose of comparison, Schelter et al. [10] estimated the performance of a seizure prediction method based on a quantity indicating phase synchronization compared with a Poisson process. Using invasive EEG data of four representative patients suffering from epilepsy, they claimed that two of them have good performance while the other two do not. Therefore, further research in this field is still necessary.

In this work, we use a nonlinear method different from existing ones to predict seizures. We believe that EEG measurements of seizures from epileptic patients can be described as a stochastic process and has a certain probability distribution. Suffczynski et al. [8] investigated the dynamical transitions between normal and paroxysmal state of epilepsy. A Poisson process or a random walk process can be used to simulate the transition between the two states. We found that the characteristic variables from epileptic EEG data can be used to represent the procedure of seizure occurrence. We develop a dynamic model where a hidden variable is involved. Features of the hidden variable can become an indicator of seizure occurrence. The hidden variable is considered to have the property of second order Markov chain. The method of particle filter associated with a neural network is used to estimate the hidden variable. Features of the hidden variable can be extracted and seizure onset can be detected in advance based on these features. As pointed out by Litt et al. [11], during the transition from normal brain activities to a seizure, some regions of the brain have similar activities. This similarity makes it possible for some characteristics detectable during the preseizure period.

Based on a probability distribution, the sensitivity can be reached by a random predictor. It is meaningful only when a predictor has higher sensitivity than the random predictor. We set significance level as 5%. Assume that the random predictor generates alarms following a Poisson process in time without using any information from the EEG [10]. The sensitivity from the random predictor can be obtained. Comparing the two, our prediction results are superior to those from the random predictor.

This paper is organized as follows. In Section 2, we introduce particle filters and neural networks. In Section 3, our method is presented including the dynamic model and the way for solving the hidden variable. In Section 4, experimental data is given. In Section 5, data processing and simulation results are described. Finally, in Section 6, discussion and conclusions are addressed.

2. Particle Filters

Although particle filters, namely, sequential Monte Carlo methods, were introduced much earlier, it became attractive and was further developed in the 1990s since computers can provide more powerful ability of computation. These methods have been very popular over the past few years in statistics and related fields since it can be used to simulate nonlinear non-Gaussian distributions, and they are improved greatly in the implementation [12–17]. Particle filters can approximate a sequence of probability distributions of interest using a set of random samples called particles. These particles are propagated over time following the corresponding distributions by sampling and resampling mechanisms. At any time, as the number of particles increases, particles should asymptotically converge toward the sequence of theoretical probability distribution. In reality, computation time is a very important factor to consider so the number of particles cannot go too big. Thus effective sampling algorithms are key steps to capture a certain probability distribution by a limited number of particles.

The basis of a particle filter is a sequential importance sampling/resampling algorithm [18]. Most sequential Monte Carlo methods developed over the last decade are based on this algorithm. This technique is capable of implementing a recursive Bayesian filter by Monte Carlo simulations. The key idea is to use a sample of random particles to approximate a posterior probability distribution. The sequential sampling is very important in realizing this algorithm. Assume an arbitrary distribution $p(x)$. Samples are supposed to be drawn from $p(x)$, but in many practical cases, $p(x)$ is not a standard probability distribution, for example, Gaussian distribution, and, therefore, it is difficult to draw samples from $p(x)$. Based on the Bayesian importance sampling scheme [19], a sample $x^i, i = 1, \ldots, N$, can be drawn from another probability distribution $q(x)$ called the importance function, which is easy to sample. Thus these particles can approximate the distribution $q(x)$. In order to use these particles to represent the desired distribution $p(x)$, a weighted approximation to the density $p(x)$ is given by

$$\widehat{p}(x) = \frac{\sum_{i=1}^{N} \widetilde{w}^i \delta(x - x^i)}{\sum_{i=1}^{N} \widetilde{w}^i}, \tag{1}$$

where

$$\widetilde{w}^i = \frac{p(x^i)}{q(x^i)}, \tag{2}$$

and $\delta(\cdot)$ is a Dirac delta function defined as

$$\delta(x - x^i) = \begin{cases} 1, & \text{if } x = x^i \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

If the samples are drawn from an importance function $q(x_{1:n} \mid \alpha_{1:n})$, then the weights in (2) are determined as

$$\widetilde{w}^i = \frac{p(x_{1:n}^i \mid \alpha_{1:n})}{q(x_{1:n} \mid \alpha_{1:n})}. \tag{4}$$

Now we can proceed to obtain a recursive updating equation which can keep the previous trajectories of particles when a set of new data is available. At each iteration, samples can approximate the corresponding distribution, for example, $p(x_{1:n-1} \mid \alpha_{1:n-1})$, and then approximate $p(x_{1:n} \mid \alpha_{1:n})$ with a new set of samples. From the Bayesian theory, we can easily obtain

$$q(x_{1:n} \mid \alpha_{1:n}) = q(x_n \mid x_{1:n-1}, \alpha_{1:n}) q(x_{1:n-1} \mid \alpha_{1:n-1}). \tag{5}$$

From (5), we already have samples $x_{1:n-1}^i \sim q(x_{1:n-1} \mid \alpha_{1:n-1})$, and can draw a particle from $x_n^i \sim q(x_n \mid x_{1:n-1}, \alpha_{1:n})$ to augment samples to become $x_{1:n}^i$. The aim is to approximate density function $p(\cdot)$, and $p(x_{1:n} \mid \alpha_{1:n})$ is expressed as follows, based on the Bayesian theory and the Markov properties [20],

$$p(x_{1:n} \mid \alpha_{1:n}) = \frac{p(\alpha_n \mid x_n) p(x_n \mid x_{n-1})}{p(\alpha_n \mid \alpha_{1:n-1})} p(x_{1:n-1} \mid \alpha_{1:n-1}). \tag{6}$$

When particle weights are considered, the updating equation is given by

$$\begin{aligned} \widetilde{w}_n^i &= \frac{p(\alpha_n \mid x_n^i) p(x_n^i \mid x_{n-1}^i) p(x_{1:n-1}^i \mid \alpha_{1:n-1})}{q(x_n^i \mid x_{1:n-1}^i, \alpha_{1:n}) q(x_{1:n-1}^i \mid \alpha_{1:n-1}) p(\alpha_n \mid \alpha_{1:n-1})} \\ &\propto \frac{p(\alpha_n \mid x_n^i) p(x_n^i \mid x_{n-1}^i) p(x_{1:n-1}^i \mid \alpha_{1:n-1})}{q(x_n^i \mid x_{1:n-1}^i, \alpha_{1:n})(x_{1:n-1}^i \mid \alpha_{1:n-1})} \\ &= \widetilde{w}_{n-1}^i \frac{p(\alpha_n \mid x_n^i) p(x_n^i \mid x_{n-1}^i)}{q(x_n^i \mid x_{1:n-1}^i, \alpha_{1:n})}. \end{aligned} \tag{7}$$

Based on the prior distribution, the initial step of the above recursion can be defined for $n = 1$ as

$$\widetilde{w}_1^i = \frac{p(x_1 \mid \alpha_1)}{q(x_1 \mid \alpha_1)}. \tag{8}$$

Thus, particle weights for $n = 1, 2, \ldots$, can recursively be obtained. We can extend the same procedure to all the particles. In (7), the term $p(\alpha_n \mid \alpha_{1:n-1})$ is omitted since it is a value by calculation. Doucet [21] showed that the effect of omission is compensated by normalizing the weights using

$$w_n^i = \frac{\widetilde{w}_n^i}{\sum_{i=1}^N \widetilde{w}_n^i}. \tag{9}$$

The sequential importance sampling algorithm has been developed, but two problems exist in practice. One is the phenomenon of degeneracy and the other is the choice of importance function $q(x)$. In general, all but a few particles will have negligible weights after several iterations and a large computational effort is devoted to updating trajectories whose contribution to the final estimation is almost zero [18]. Liu and Chen [16] introduced a method to measure particle degeneracy. The effective sample size N_{eff} is defined as:

$$N_{\text{eff}} = \frac{N_s}{1 + \text{Var}(w_n^{*i})}, \tag{10}$$

where w_n^{*i} denotes the true weight by calculation directly. It is not easy to calculate N_{eff} from the above equation, so an approximation of N_{eff} can be used as

$$\widehat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^{N_s} (w_n^i)^2}, \tag{11}$$

where w_n^i is the normalized weight obtained from (8). The smaller the $\widehat{N}_{\text{eff}}$, the worse the degeneracy. Generally speaking, increasing the number of particles can reduce degeneracy, but it is impractical. When $\widehat{N}_{\text{eff}} \leq N_{\text{threshold}}$, where $N_{\text{threshold}}$ is usually taken as one third of the particle number, resampling is necessary. Resampling procedures can decrease the degeneracy phenomenon but it introduces practical, and theoretical problems [18]. From a theoretical point of view, the simulated trajectories are no longer statistically independent after resampling so the previous convergence result will be lost. From a practical point of view, it limits the opportunity to parallel computation since all the particles must be combined, although the importance sampling steps can still be realized in parallel.

3. Methods

This section includes three parts. The first part describes our dynamic model. The second one introduces the solution for hidden variable in our model. The last one addresses seizure feature selection and determination.

3.1. Dynamic Model. Energy can be used to represent features of a signal. For epileptic seizures, we find that energy for some specific frequency band (4–12 Hz), which includes theta (4–8Hz) and alpha (8–12 Hz) waves, can be modeled by a similar Poisson process. Other combinations based on delta (0–4 Hz), theta, alpha, and beta (12–30 Hz) waves are also calculated but their characteristics are not as obvious. Our dynamic state model is given by

$$\begin{aligned} x_k &= \alpha x_{k-1} + \beta x_{k-2} + v_k \\ E_k &= A x_k e^{-x_k/B} + w_k, \quad k = 1, 2, \ldots, \end{aligned} \tag{12}$$

where x_k is a random variable and has a normal distribution initially. v_k, w_k are white noise with Gaussian distribution and they are independent. α, β are parameters to be determined. E_k is the energy from specific frequency band. A and B are unknown constants. The process for x_k is actually assumed to be a second-order Markov chain. The hidden variable x_k can represent transition changes and has the ability to indicate seizure occurrence in advance. The process chosen in (12) is based on our study and on the work in [8]. Also, the energy in a frequency band changes continuously and its value is affected by the most recent past values. To the best of our knowledge, no other researchers have developed a model which is used to simulate seizure process behaviors and further to predict their occurrence.

3.2. Solution of the Dynamic Model. We already introduced particle filters in Section 2. In order to improve its performance under small number of particles, we develop a novel algorithm to combine particle filters with neural networks. The strategy of backpropagation neural networks can be used to adjust particles in tail area with low weights in a particle filter.

The basic idea of backpropagation neural networks is to use the steepest descent (gradient) procedure to minimize the error energy at the output layer. The error energy can be denoted as follows:

$$E \triangleq \frac{1}{2}\sum_k (d_k - y_k)^2 = \frac{1}{2}\sum_k e_k{}^2, \tag{13}$$

where $k = 1,\ldots,N$; N is the number of neurons in the output layer. d_k is the target value and y_k is the output of neural network. By using gradient procedure and updating weights of all neurons to train a neural network, proper weights can be found so that the output of the network is close to the desired objective within an assigned error. The activation function in neural networks can be chosen according to actual problems [22].

There are one input, one hidden, and one output layer built in our algorithm. The dimension of input layer is determined adaptively by particle samples in the particle filter. Particles with smaller weights are considered as the input data of a neural network. Their corresponding weights are set as inputs of the neural network, and their particle values as initial weights of the neural network. The weights of the remaining particles are set as biases of corresponding neurons. The neural networks can improve the performance of particle filters, for example, the number of simulation is reduced significantly. The noise w_k in (12) is small since measurements are intracranial EEG data. In general, the computational complexity is $O(N)$, where N is the number of particles. Our algorithm is displayed in Algorithm 1 [23].

3.3. Feature Determination. Based on Algorithm 1, the hidden variable in the dynamic model can be obtained. For a given patient, suppose that the first seizure is known. All the parameters in (12) can be obtained. Parameters α and β can be determined by minimizing errors, based on a known

seizure. A and B can be obtained by minimizing error w_k. One further step is to do regression analysis.

The regression analysis is based on the method of Chatterjee and Hadi [24], expressed by

$$Y = X\xi + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I), \tag{14}$$

where Y is a dependent variable (output), X is an independent variable (input or data), and ϵ is the error. The parameter ξ can be determined using the least square error method and the predicted data can then be obtained from (14).

Normally there is a peak at some time instants before seizure occurrence and x value will be between 270 and 360 during the ictal period. The feature of a "peak" can be described by the mean value (with threshold of $\pm10\%$ of the previous mean value), the variance before it (with threshold of $\pm5\%$ of the mean of previous variance), the peak amplitude (at least 10 more than the previous mean value), and the width of peak (from 1 minute to 6 minutes). The mean value and variance can be calculated for 15–30 minutes before the peak; peak amplitude can be detected by the real peak value, and the width of peak can also be obtained at the same time. We assume that these features will be kept the same at the next seizure onset. All the features can be updated as long as the information of a new seizure is available. Thus the system can adaptively update all related parameters automatically based on available seizure information.

From Figure 1, the hidden variable's value at certain time before seizure occurrence reaches a peak. Before that peak, the variance is small, which means that the curve before the peak is smooth. Figure 1 shows this characteristic. The difference between the time at which seizure is alerted to happen, and seizure actual occurrence is the prediction time. Based on this type of signature, a certain time point before seizure occurrence can be recognized and a seizure alert is provided at that point. For Figure 1, the prediction time is 14 minutes. The minimum intervention time is set to 2 hours in our study. If a seizure appears from 3 to 120 minutes after a seizure is alerted, this prediction is considered to be successful. Otherwise, a false prediction is counted.

4. Experimental Data

The EEG data that we use are invasive EEG recordings of 6 patients with medically intractable temporal lobe epilepsy. The data were recorded during an invasive presurgical epilepsy monitoring at the Epilepsy Center of the University Hospital of Freiburg, Germany. In order to obtain a high signal-to-noise ratio, fewer artifacts, and to record directly from temporal areas, intracranial grid-, strip-, and depth-electrodes were utilized. The EEG data were acquired using a Neurofile NT digital video EEG system with 128 channels, 256 Hz sampling rate, and a 16 bit analog-to-digital converter. For each patient, we were given 4–6 channels of data recorded from temporal areas. The amplitude of data is relative to the real one after sampling them, but all the features will be kept the same.

For each patient, there are datasets called "ictal," and "interictal," with the former containing EEG-recordings with

1. Importance sampling

-For $i = 1, \ldots, N$, sample $\tilde{x}_n^i \sim q(x_n \mid x_{1:n-1}^i, \alpha_{1:n})$, and set $\tilde{x}_{1:n}^i \triangleq (x_{1:n-1}^i, \tilde{x}_n^i)$,
where $q(x_n \mid x_{1:n-1}^i, \alpha_{1:n})$ is a chosen probability density function.
N is the number of particles and n is the current time.
-For $i = 1, \ldots, N$, evaluate the importance weights up to a normalizing constant:
$\tilde{w}_n^i = \tilde{w}_{n-1}^i (p(\alpha_n \tilde{x}_n^i) p(\tilde{x}_n^i \tilde{x}_{n-1}^i)) / q(\tilde{x}_n^i \mid \tilde{x}_{1:n-1}^i, \alpha_{1:n})$, where $p(\alpha_n \mid \tilde{x}_n^i)$, and $p(\tilde{x}_n^i \mid \tilde{x}_{n-1}^i)$
are conditional probability density functions for α_n, and $\tilde{x}_n^i$, respectively.
-For $i = 1, \ldots, N$, normalize the importance weights:
$w_n^i = \tilde{w}_n^i / \sum_{j=1}^N \tilde{w}_n^j$, where w_n^i is the normalized weight.
-At time n, identify particles with high weights, and low weights.
Replace some low weight particles with high ones if needed.
-At time n, adjust particles with low weights by neural networks.
Assign and normalize weights by the aforementioned procedure
-Evaluate $\hat{N}_{\text{eff}}$ using $\hat{N}_{\text{eff}} = 1 / \sum_{i=1}^{N_s} (w_n^i)^2$, where $\hat{N}_{\text{eff}}$ is the threshold parameter.
2. Resampling if necessary
-If $\hat{N}_{\text{eff}} \geq N_{\text{threshold}}$, where $N_{\text{threshold}}$ is a preset threshold, $x_{1:n}^i = \tilde{x}_{1:n}^i$ for $i = 1, \ldots, N$;
-Otherwise, for $i = 1, \ldots, N$, sample an index $j(i)$ distributed according to the discrete distribution
with N elements satisfying $Pr\{j(i) = l\} = w_n^l$ for $l = 1, \ldots, N$;
for $i = 1, \ldots, N, x_{1:n}^i = \tilde{x}_{1:n}^{j(i)}$, and $w_n^{*i} = 1/N$, where w_n^{*i} is an updated weight.

ALGORITHM 1: Importance sampling/resampling particle filter with a neural network.

epileptic seizures, and the latter EEG-recordings without seizure activity. We use all ictal EEG data, and at least 10 hours interictal data for each subject.

For a particle filter, the optimal strategy is to choose $q(x_n \mid x_{n-1}^i, \alpha_n) = p(x_n \mid x_{n-1}^i)$. Therefore, we use linearization technique to linearize the model (12). It now becomes

$$x_k = \alpha x_{k-1} + \beta x_{k-2} + v_k = f(x_{k-1}, x_{k-2}) + v_k, \qquad (15)$$

$$E_k = A f(x_{k-1}, x_{k-2}) e^{-f(x_{k-1}, x_{k-2})/B}$$

$$+ \left[A x_k e^{-x_k/B} - A x_k e^{-x_k/B} / B \right]|_{x_k = f(x_{k-1}, x_{k-2})} \qquad (16)$$

$$\times (x_k - f(x_{k-1}, x_{k-2})) + w_k, \quad k = 1, 2, \ldots.$$

5. Results

5.1. Data Preprocessing. Intracranial EEG data are unprocessed directly from patients. Although they were obtained from intracranial electrode contacts on brain directly, there still exist some unusual values in the recording, for example, very big difference between two close points in the measurement. These points can be replaced with normal ones by interpolation, since there are few of this type of points in our data. Then roll-over windowing technique is applied to them. We choose nonoverlap 5-second window to divide EEG data of a single channel. Wavelet transform "DB4" is used to get the energy of specific band since it can give good performance and it is widely used to analyze EEG data. Compared with energy of different frequency bands, the frequency band of 4–12 Hz shows much better performance and is chosen for use in our model.
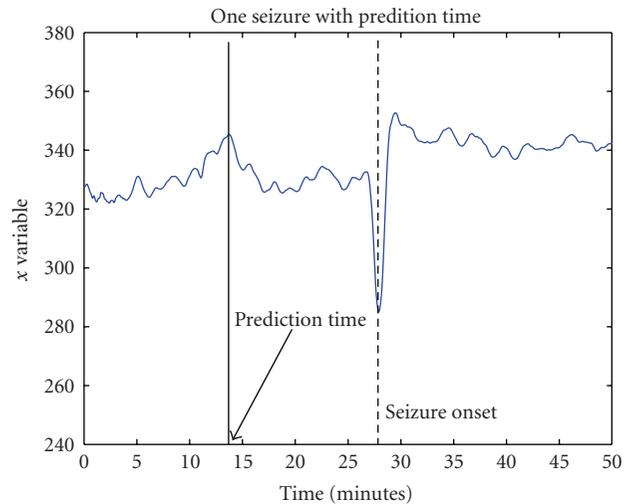

FIGURE 1: One typical figure with prediction time and seizure. The vertical solid line marks prediction time point and the vertical dashed line indicates seizure occurrence.

5.2. Preprocessing. Based on the dynamic model (12), E_k is obtained from the above steps and the hidden variable x_k can be found by particle filter associated with a neural network realized by Algorithm 1. We assume that the initial condition of x_k for the model is a normal distribution $N(300, 5)$. The mean value that we choose is based on initial energy that we calculate. Normally its value is about 300. Thus, less time is needed to run Algorithm 1 at the initial points. Actually this value cannot have any effect on the final result except the running time. v_k, w_k are white noise, and we assume $v_k \sim N(0, 0.6)$ and $w_k \sim N(0, 0.1)$. In the dynamic model (12), A, B are unknown parameters. The number of particles that

TABLE 1: The optimal parameters α and β for six patients

Patient No.	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6
α	0.7972	0.9001	0.8164	0.8775	0.9155	0.8599
β	0.2028	0.0999	0.1846	0.1225	0.0845	0.1401

we use is 200. For each patient, the first seizure is supposed to be known and is used to determine the parameters. The algorithm of minimum least squares error is used to find the optimal parameters under the assumption that the process is steady before the next seizure occurrence. We follow the same procedure when dealing with all seizures of each patient. According to the energy values calculated and parameter optimization, $A = 2800$ and $B = 40$ can be obtained. Table 1 shows the optimal parameters α and β for six patients based on the first seizure occurrence.

Model (12) is a nonlinear model with Gaussian state space. A local linearization technique is applied to nonlinear equations and an approximate linear equation is obtained in (16). A series of values of hidden variable x can be obtained based on Algorithm 1.

5.3. Experimental Results. Intracranial EEG data from six patients are tested using our algorithm. It includes a total of 22 epileptic seizures, and 110 hours of data. Six of them are taken out to determine all the related parameters in model (12) for the subsequent seizures of each patient. After the preprocessing described above, Algorithm 1, namely, the particle filter associated with a neural network, is used to identify the hidden variable x. In order to recognize the general characteristics before seizure onset, the method of linear regression is applied to calculated values of x. This regression process can make clear the tendency of change for the hidden variable x and provide some obvious characteristics which are used to identify seizure occurrence in advance.

Figures 2–7 show the hidden variable x from six patients computed by our algorithm. Each of them includes two figures, one from ictal EEG with one seizure, and the other from interictal EEG without seizure. It is seen for all the ictal EEG that the characteristics occurring some time instants before the seizure can be recognized and used for predicting seizure onset. All patients here have temporal lobe epilepsy. Figure 2 shows an epileptic seizure from a male patient. The prediction time is 42 minutes. After seizure happens, the variable x is on a little high level compared to that before seizure. For the interictal period, the value x is higher than that during ictal period. Figure 3 shows an epileptic seizure from a female patient. Its characteristics are the same as Figure 2 including ictal and interictal transition data. The prediction time is about 11.5 minutes. Data in Figure 4 are from a female patient too. The prediction time is about 30 minutes. The interictal characteristics, which oscillate on the low values, are different from others. Figures 5 and 6 have very similar characteristics: figures for interictal EEG data are on the relative low values smoothly; figures for ictal EEG data are on similar values. Figure 5 is from a young male patient and Figure 6 is from an old female patient. Their
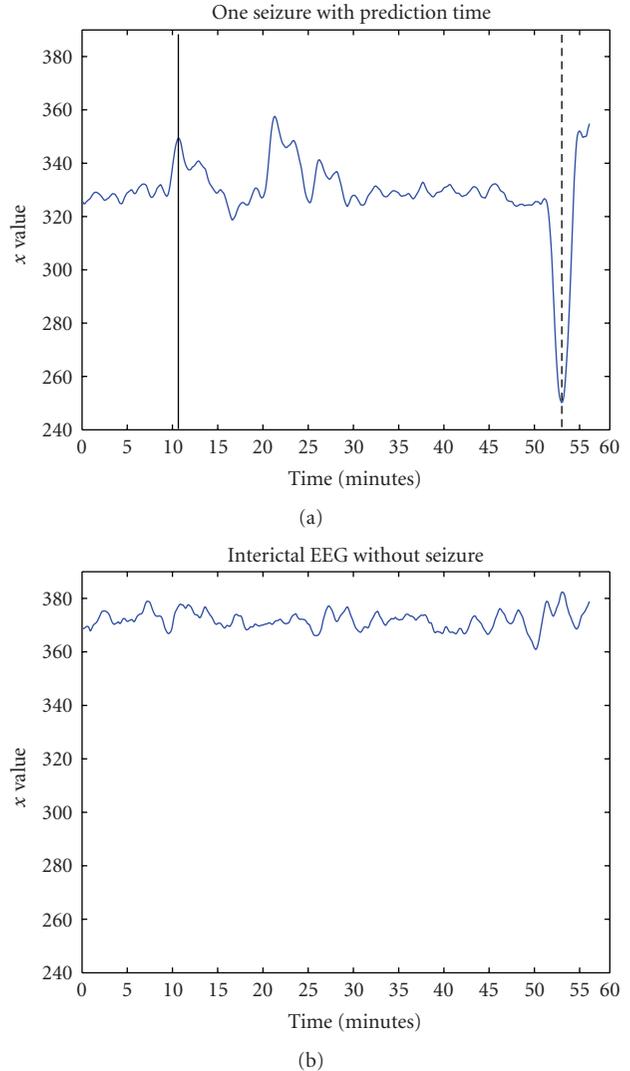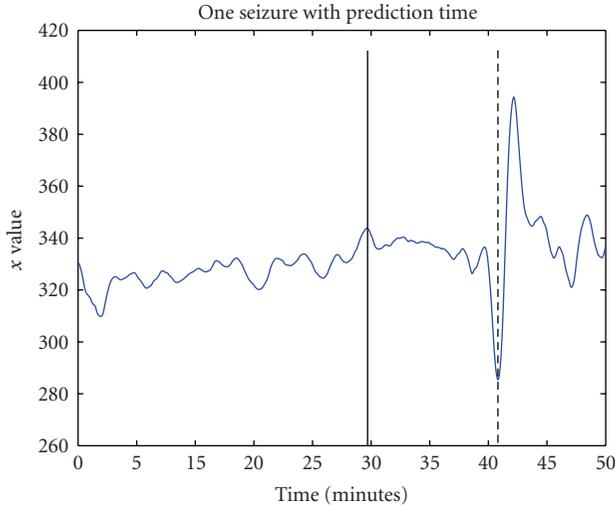

(a)

(b)

FIGURE 2: Ictal and interictal hidden variable x from one patient with epilepsy. The vertical solid line marks prediction time point and the vertical dashed line indicates seizure occurrence.

prediction times are 39 and 38.5 minutes, respectively. Figure 7 comes from a young male patient. Both ictal and interictal values x are relatively low compared to other patients, but its characteristics before the seizure are obvious. This seizure can be known 6.25 minutes in advance.

Totally we tested 16 seizures from these 6 patients. The average prediction time is 38.5 minutes. The longest prediction time is 83.7 minutes and the shortest one is 6.25 minutes. 15 seizures can be predicted successfully. The sensitivity is 93.75%.101 hours intracranial EEG testing data are analyzed by our algorithm and specificity (false-positive rate) is about 0.09 FP/hour.

In order to determine the performance of our method, a random predictor is used to calculate the sensitivity. We assume that the random predictor generates alarms following a Poisson process in time without using any information

(a)
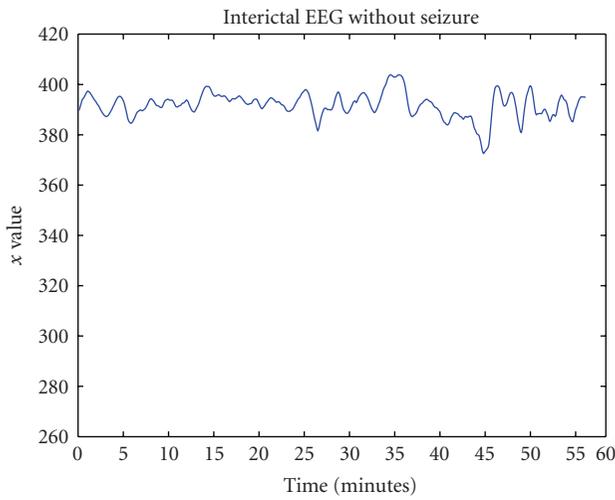
(b)

FIGURE 3: Ictal and interictal hidden variable x from one patient with epilepsy. The vertical solid line marks prediction time point and the vertical dashed line indicates seizure occurrence.
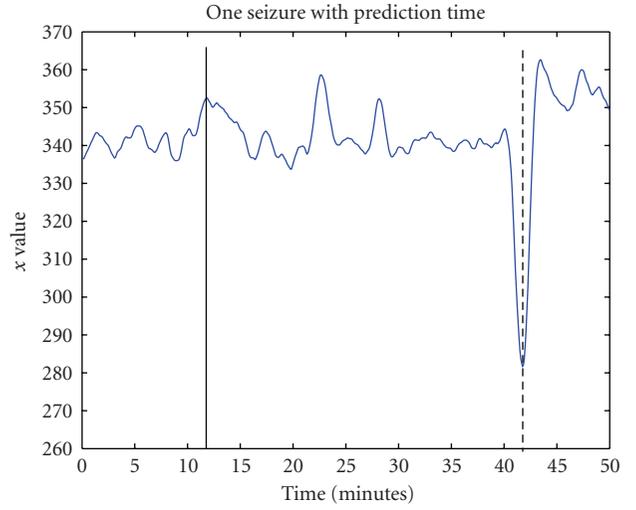

(a)

(b)

FIGURE 4: Ictal and interictal hidden variable x from one patient with epilepsy. The vertical solid line marks prediction time point and the vertical dashed line indicates seizure occurrence.

from the EEG. The probability to raise an alarm in a period of duration can be calculated as [10]

$$P \approx R_{FP} \times P_{SO}, \tag{17}$$

when $R_{FP} \times P_{SO}$ is smaller than one, where R_{FP} is the maximum false prediction rate, which is set as 2 seizures each day, and P_{SO} is seizure occurrence period. In our case, P_{SO} is 2 hours. To decide the statistical significance of sensitivity values, we follow Schelter's method [10] to calculate the probability as

$$P_{\{k;K;P\}} = 1 - \left(\sum_{j<k} \binom{K}{j} P^j P^{K-j} \right)^d, \tag{18}$$

where P is the above probability for the given false prediction rate, and prediction period, and K is the seizure number.

This is the probability of predicting at least k out of K seizures by means of at least one of d independent features correctly. For our case, d is one. The significance level is set at 5%. For 2 seizures, the sensitivity is 100% to meet the significance level. The sensitivity is 67% for 3 seizures and it is 50% for 4 seizures. Our method can detect 15 out of 16 seizures, and the only one missed is from a patient having 4 seizures. For 5 out of the 6 patients, our method has sensitivity of 100%. The sensitivity for the other patient with a missed detection is 75% which is much better than the random predictor (which is only 50%). Therefore, our method has superior performance to the random predictor.

6. Discussions and Conclusions

Although many methods are published for predicting epileptic seizures, none of them has been accepted widely so new
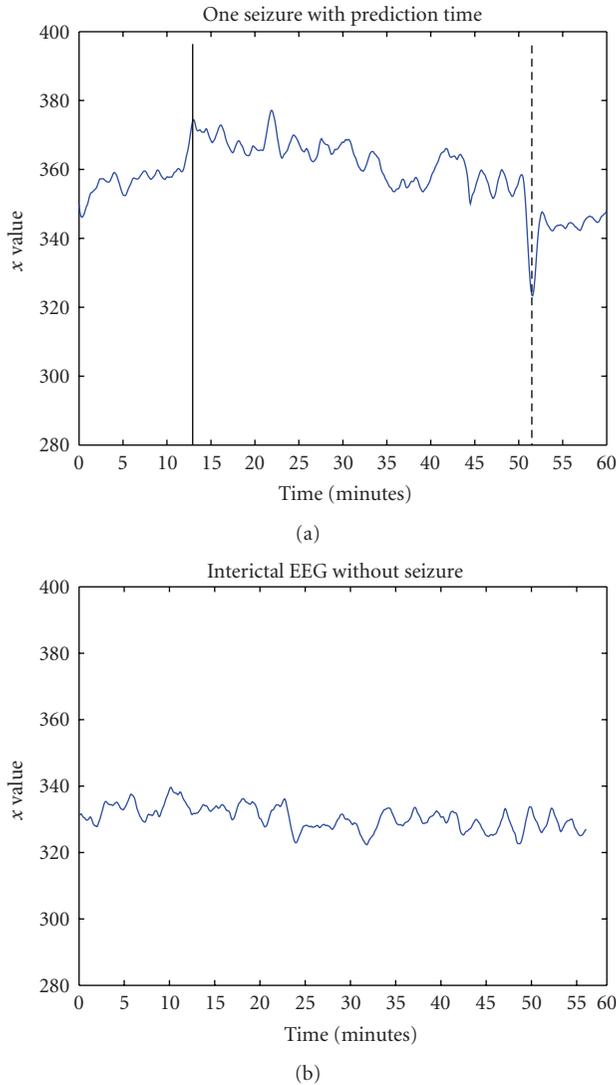
FIGURE 5: Ictal and interictal hidden variable x from one patient with epilepsy. The vertical solid line marks prediction time point and the vertical dashed line indicates seizure occurrence.
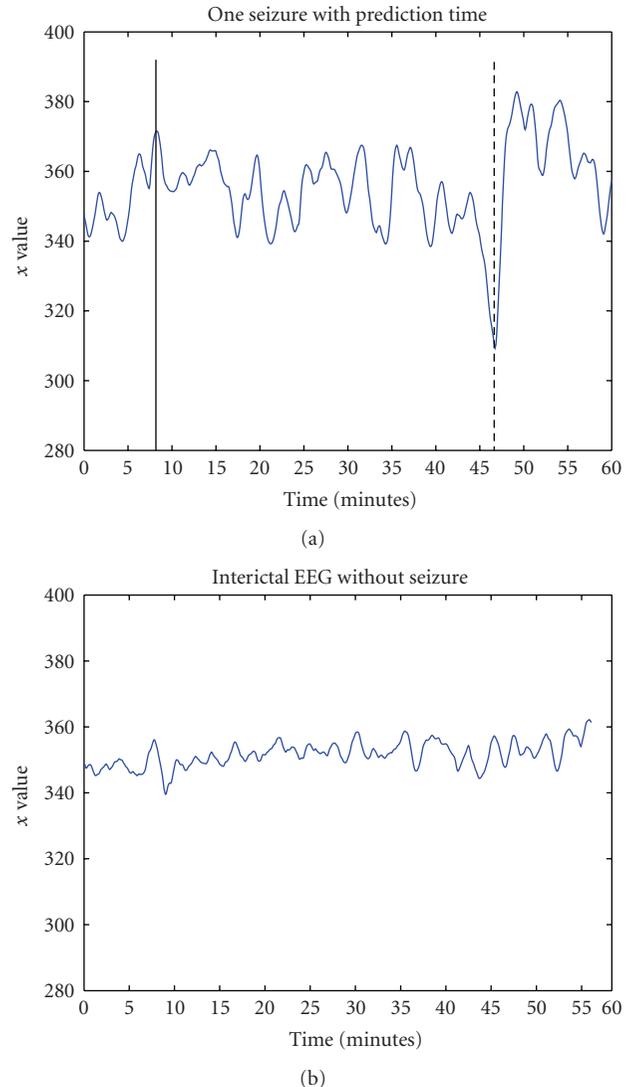

FIGURE 6: Ictal and interictal hidden variable x from one patient with epilepsy. The vertical solid line marks prediction time point and the vertical dashed line indicates seizure occurrence.

methods are necessary to complement or replace current ones. The novel prediction method developed in this paper is different from other current existing methods. The wavelet transform is used to get the energy of specific frequency band of 4–12 Hz in our method. The dynamic model based on energy under frequency 4–12 Hz is used to describe seizure features. A particle filter associated with a neural network is used to solve the hidden variable in the model. Here the important part is to use a neural network, which can improve algorithm performance even with small number of particles. We use 109 hours intracranial EEG data to estimate the performance of this method including 8 hours of data to determine optimal parameters for the second seizure of each patient in the model. 15 out of 16 seizures were successfully predicted, and the sensitivity is 93.75%. The false-positive rate is about 0.09 per hour. Therefore, this algorithm can capture signatures before epileptic seizure onset, and further

can be used to predict them. Our algorithm was applied to a single channel EEG data which represent activities of a certain brain region (temporal areas) since all the 4–6 channels of each patient provided similar EEG data. The results obtained support the thought of modeling EEG signals to gain insight into the dynamical process involving seizure generation [8, 9].

In order to determine the performance of our method, a random predictor under the significance level of 5% is used to obtain the sensitivity. For all six patients, our method has shown superior performance to the random predictor.

The original motivation to predict seizure is to meet the requirement for a successful therapeutic intervention, for example, for drug administration. The time interval between prediction and occurrence of seizure is necessary and useful to the treatment of a patient. In order to meet requirements in clinic, reliability is a key factor for any prediction method,
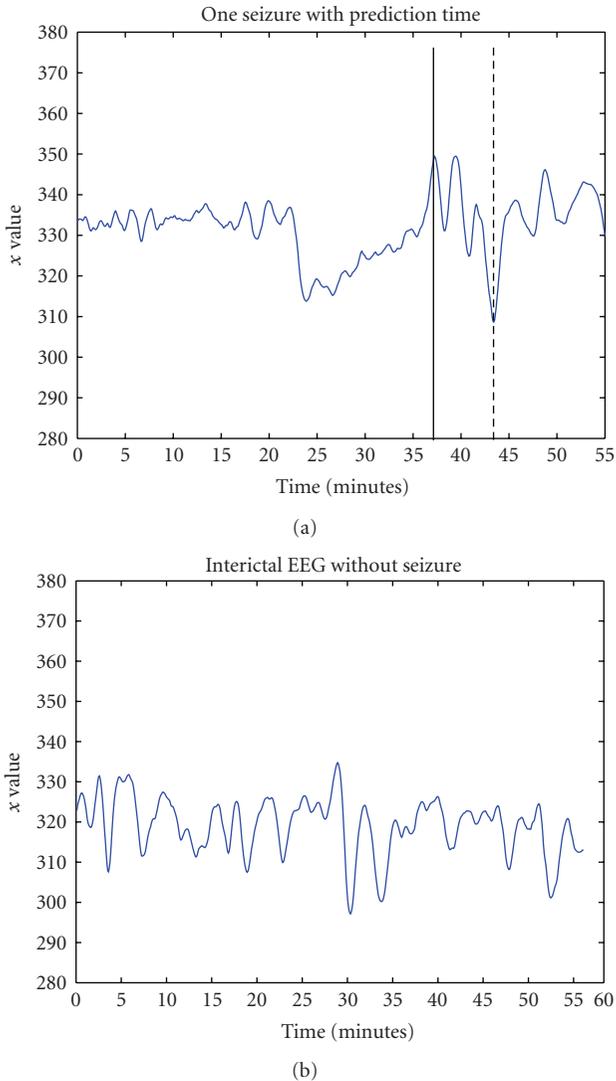
FIGURE 7: Ictal and interictal hidden variable x from one patient with epilepsy. The vertical solid line marks prediction time point and the vertical dashed line indicates seizure occurrence.

There are two important issues in this method. The first one is that noise in EEG data should be low, which can be guaranteed by modern technology. The second one is the choice of channels. In reality, one further step is needed to detect the channel in the brain regions where seizure happens.

This method is promising based on results obtained. Potential applications in clinic for seizure warning need a prior step which is EEG channel selection since channels on different regions of brain have different response to the same seizure. The present algorithm is the first step to apply it to the diagnosis using EEG measurements. It can provide very useful information for doctors and patients.

Acknowledgment

References

[1] L. D. Iasemidis, "Epileptic seizure prediction and control," *IEEE Transactions on Bio-Medical Engineering*, vol. 50, no. 5, pp. 549–558, 2003.

[2] L. D. Iasemidis and J. C. Sackellares, "The evolution with time of the spatial distribution of the largest Lyapunov exponent on the human epileptic cortex," in *Measuring Chaos in the Human Brain*, F. Duke and W. Pritchard, Eds., pp. 49–82, World Scientific, Singapore, 1991.

[3] L. D. Iasemidis, J. C. Sackellares, W. J. Williams, and T. W. Hood, "Nonlinear dynamics of electrocorticographic data," *Journal of Clinical Neurophysiology*, vol. 5, p. 339, 1988.

[4] L. D. Iasemidis, D. S. Shiau, W. Chaovalitwongse, et al., "Adaptive epileptic seizure prediction system," *IEEE Transactions on Bio-Medical Engineering*, vol. 50, no. 5, pp. 616–627, 2003.

[5] J. S. Ebersole, "In search of seizure prediction: a critique," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 489–492, 2005.

[6] H. Hassanpour, M. Mesbah, and B. Boashash, "Time-frequency feature extraction of newborn EEG seizure using SVD-based techniques," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 16, pp. 2544–2554, 2004.

[7] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2007.

[8] P. Suffczynski, F. H. Lopes da Silva, J. Parra, et al., "Dynamics of epileptic phenomena determined from statistics of ictal transitions," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 524–532, 2006.

[9] F. Wendling, F. Bartolomei, J. J. Bellanger, and P. Chauvel, "Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition," *European Journal of Neuroscience*, vol. 15, no. 9, pp. 1499–1508, 2002.

[10] B. Schelter, M. Winterhalder, T. Maiwald, et al., "Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction," *Chaos*, vol. 16, no. 1, Article ID 013108, 2006.

[11] B. Litt, R. Esteller, J. Echauz, et al., "Epileptic seizures may begin hours in advance of clinical onset: a report of five patients," *Neuron*, vol. 30, no. 1, pp. 51–64, 2001.

[12] A. Doucet, N. D. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Pratice*, Springer, Berlin, Germany, 2001.

and specificity and sensitivity are used to assess how well a method works. Sometimes sensitivity of an algorithm is high while its specificity is low, which means there are a lot of false predictions. This situation cannot be allowed in clinic since too many false predictions will lead to impairment due to possible side-effects of interventions or loss of the patients' acceptance of seizure warning [10]. Although our method is tested by a limited intracranial EEG data, it has a reliable performance for all six patients including preictal, interictal, and postictal transition data. Application of our method here focuses on the same type of epilepsy-temporal lobe epilepsy, but its extension to other types of epilepsy is feasible. Also data that we use are intracranial from brain surface directly. Our future research will consider to apply the method to scalp EEG data from patients with epilepsy, and to compare it with results from intracranial ones.

[13] W. R. Gilks and C. Berzuini, "Following a moving target— Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society B*, vol. 63, no. 1, pp. 127–146, 2001.

[14] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings, Part F*, vol. 140, no. 2, pp. 107–113, 1993.

[15] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, New York, NY, USA, 2001.

[16] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, 1998.

[17] M. K. Pitt and N. Shephard, "Filtering via simulation: auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.

[18] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[19] J. Bernardo and A. Smith, *Bayesian Theory*, John Wiley & Sons, New York, NY, USA, 1994.

[20] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[21] A. Doucet, "On sequential simulation-based methods for Bayesian filtering," Tech. Rep., Signal Processing Group, University of Cambridge, Cambridge, UK, 1998.

[22] J. M. Zurada, *Introduction to Artificial Neural Systems*, West Publishing, New York, NY, USA, 1992.

[23] Z. Pang, D. Liu, N. Jin, and Z. Wang, "A Monte Carlo particle model associated with neural networks for tracking problem," *IEEE Transactions on Circuits and Systems I*, vol. 55, no. 11, pp. 3421–3429, 2008.

[24] S. Chatterjee and A. S. Hadi, "Influential observations, high leverage points, and outliers in linear regression," *Statistical Science*, vol. 1, pp. 379–393, 1986.