Online Bayesian Change Point Detection Algorithms for Segmentation of Epileptic Activity

Rakesh Malladi*, Giridhar P Kalamangalam†, and Behnaam Aazhang*‡
*Department of Electrical and Computer Engineering, Rice University, Houston, USA
†Department of Neurology, University of Texas Health Sciences Center, Houston, USA
‡Centre for Wireless Communications, University of Oulu, Oulu, Finland

***Abstract*—Epilepsy is a dynamic disease in which the brain transitions between different states. In this paper, we focus on the problem of identifying the time points, referred to as change points, where the transitions between these different states happen. A Bayesian change point detection algorithm that does not require the knowledge of the total number of states or the parameters of the probability distribution modeling the activity of epileptic brain in each of these states is developed in this paper. This algorithm works in online mode making it amenable for real-time monitoring. To reduce the quadratic complexity of this algorithm, an approximate algorithm with linear complexity in the number of data points is also developed. Finally, we use these algorithms on ECoG recordings of an epileptic patient to locate the change points and determine segments corresponding to different brain states.**

I. Introduction

Epilepsy is a neurological disorder characterized by unprovoked seizures. Nearly 1% of the world's population suffers from this disease and about 30% of them have medically refractory epilepsy (that is, for these patients medication is not effective). Surgical resection is only successful in about 60% of the patients. Inspired by the success in treating movement disorders like Parkinson's disease, electrical stimulation using subdural and depth electrodes is considered as a promising technique to treat epilepsy [1]. However, epilepsy is a dynamic disease in which brain transitions between different states [2]. Each state is defined by a connectivity graph that factorizes the joint probability distribution over the activity at different electrode locations [3] and the activity is measured by electroencephalography (EEG) and electrocorticography (ECoG) recordings. The dynamic behavior observed in EEG and ECoG recordings makes the selection of optimal temporal and spatial locations for stimulation non-trivial [4]. Computational approaches for selecting the optimal electrical stimulation parameters require the complete knowledge of different states and the temporal transitions between them. This paper focuses on temporal transitions between the different states.

Specifically, in this paper we address the problem of detecting time-points in the EEG and ECoG recordings after which the underlying state (represented by a joint probability distribution) changed, henceforth referred to as change points (CP). The problem of detecting change points from a time series is a well-studied problem with applications in domains like finance, engineering, and medicine. Unlike the traditional solutions, any solution to the problem of segmenting epileptic activity should have low complexity, work in online instead of offline mode, and be able to deal with the non-stationary property of EEG and ECoG data. The Bayesian CP detection algorithms presented in this paper satisfy all these requirements.

In this work, we first develop an online Bayesian change point detection algorithm that works for non independent and identically distributed (non-i.i.d.) data. This algorithm, based on the online CP detection algorithm in [5], has quadratic complexity in the number of data sample points. Secondly, an approximate algorithm based on ideas from list decoding [6] is also proposed which has linear complexity in the number of data sample points. The performance of both these algorithms is evaluated and compared with the forward/backward CP detection algorithm [7] on simulated data. Finally, the ECoG activity measured from an epileptic patient is segmented into different states using both these algorithms.

The main contributions of this paper are:

- An online Bayesian change point detection algorithm for the general case of non-i.i.d. data.
- An approximate algorithm with linear complexity, based on ideas from list decoding.
- Segmenting ECoG data to identify segments of activity corresponding to the different epileptic brain states.

The outline of the rest of the paper is as follows. In Section II we describe the system model and the notation. Section III and IV explain the online and the approximate Bayesian CP detection algorithms for non-i.i.d. data respectively. In Section V the likelihood for the linear regression models with Gaussian noise, essential for the algorithms in Sections III and IV, is calculated. In Section VI we demonstrate the performance of our algorithms on simulated data for the data model given in Section V. In Section VII we apply both the exact and the approximate algorithms to segment ECoG data. Final concluding remarks are given in Section VIII.

A. Related Work

Detecting certain events of interest like seizures and spikes from EEG and ECoG is a well-studied problem [8]. To the

This work is funded in part by a NSF grant and through an Academy of Finland research grant. The authors can be reached at {Rakesh.Malladi, aaz}@rice.edu and Giridhar.P.Kalamangalam@uth.tmc.edu.

best of our knowledge, no one has looked into segmenting the activity in an entire observation window to find all the different states. Some of the earliest works in Bayesian change point detection are based on Markov Chain Monte Carlo (MCMC) and its variations [9], [10]. Product partition models (PPM) for change point detection were introduced in [11]. A forward/backward algorithm to solve for CPs in data modeled by PPMs proposed in [7] overcomes the difficulties of convergence in MCMC methods. However all these algorithms work in offline mode and have high complexity. Online Bayesian CP detection algorithms were proposed in [5], [12]. Ideas from re-sampling algorithms for particle filters are applied to the forward/backward algorithm [7] to reduce the complexity without much loss in performance in [12]. On the other hand, [5] focused on casual predictive filtering. These algorithms assume the data within each segment is i.i.d., which is not true for EEG and ECoG data [13].

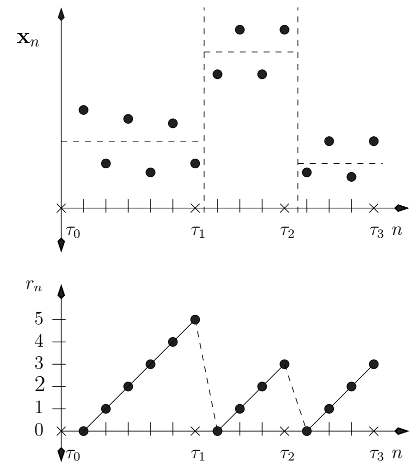


Fig. 1: Example showing a sequence of data samples and the corresponding run-lengths

II. System Model and Notation

Let $\mathbf{x}_{1:N} = (\mathbf{x}_1,\, \mathbf{x}_2, \cdots,\, \mathbf{x}_N)^T$ denote the N data samples observed. Each data sample lies in $\mathbb{R}^d$, i.e., $\mathbf{x}_n \in \mathbb{R}^d$, $\forall\, n = 1, 2, \cdots, N$. Note d is the number of electrodes used for recording EEG and ECoG data. Also, $\mathbf{x}_{i:j}$ represents the data between time indices i and j, i.e., $(\mathbf{x}_i,\, \mathbf{x}_{i+1},\, \cdots,\, \mathbf{x}_j)^T$. Let there be M change points (CPs) in this data sequence, denoted in increasing order by the time indices $\tau_1,\, \tau_2,\, \cdots,\, \tau_M$. By definition, $\tau_0 = 0$ and $\tau_{M+1} = N$. These change points imply, $\mathbf{x}_{\tau_m+1:\tau_{m+1}}$ forms a segment of data drawn from some distribution, $\forall\, m = 0,\, 1,\, \cdots,\, M$ and that this underlying distribution is different in each segment. The objective is to find both the number of change points and their positions.

An auxiliary variable r_n, referred to as 'run-length' at time index n, is defined to help in inferring the change points. Run-length captures the time since the last change point [5]. Since $\tau_0 = 0$ is a change point, $r_1 = 0$. Also $r_n \in [0\,,\, n-1]\,, \forall\, n$. Fig. 1 plots some hypothetical time sample data $\mathbf{x}_n$ with $d = 1$ and the corresponding run-length r_n. There are 2 change points τ_1 and τ_2 in this case and therefore r_n is 0 immediately after these 2 change points.

Given the run-length at a time instant, the run-length at the next time point can either go to 0 or increase by 1, depending on whether a change happens after this time instant or not. The relationship between r_n and $r_{n-1}, \forall\, n$ is given by

$$r_n = \begin{cases} 0 & , \quad \text{if } (n-1) \text{ is a change point} \\ r_{n-1}+1 & , \quad \text{otherwise} \end{cases} \tag{1}$$

The conditional probability of r_n given the run-length at $(n-1)$ is denoted by $\mathrm{P}\,(r_n|r_{n-1})$. In the following subsections, we define the prior probabilities on change points and the general model for likelihood of data within each segment, required for Bayesian inference in Sections III and IV.

A. Prior on Change Points

The change points are assumed to follow a Markov process, where the position of a change point is only dependent on the immediate preceding change point. The conditional probability of the k^{th} CP at some time index j given the $(k-1)^{th}$ change point at i is assumed to depend only on the distance between the change points [7], [11] for $k = 1,\, 2,\, \cdots, M$ and is given by

$$\mathrm{P}\,(\tau_k = j \mid \tau_{k-1} = i) = g\,(j-i)\,,\; 0 \le k-1 \le i < j \le N, \tag{2}$$

where $g\,(\cdot)$ is any discrete probability mass function over the set of natural numbers. Also note when $k = 1, i = 0$ in (2). The prior probability of a time index j being the k^{th} change point for $k = 1,\, 2,\, \cdots, M$, depends on the transition probability and is given by

$$\mathrm{P}\,(\tau_k = j) = \sum_{i=0}^{j-1} g\,(j-i) \cdot \mathrm{P}\,(\tau_{k-1} = i)\,,\; j = 1,\, 2,\, \cdots,\, N. \tag{3}$$

Given $g\,(.)$, the transition probabilities for the run-length r_n given the run-length at the previous time instant is

$$\mathrm{P}\,(r_n = r|r_{n-1}) = \begin{cases} h\,(r_{n-1}+1) & , \quad \text{if } r = 0 \\ 1 - h\,(r_{n-1}+1) & , \quad \text{if } r = r_{n-1}+1 \end{cases} \tag{4}$$

where $h\,(r_{n-1}+1) = \frac{g(r_{n-1}+1)}{\sum\limits_{i=r_{n-1}+1}^{\infty} g(i)}$, $\forall n = 2,\, 3,\, \cdots,\, N$. Also more generally, the probability of a segment whose length is atleast $(r+1)$ is given by

$$\mathrm{P}\,(r_n = r|r_{n-r} = 0) = \prod_{i=1}^{r} (1 - h\,(i))\,, \tag{5}$$

where $r \in [1, n-1]\, \forall n = 2, 3, \cdots, N$. The CP prior information (2) is contained in the run-length transition probabilities (4) and the probability of a segment of some minimum length (5). The algorithms presented in Sections III and IV use (4) and (5), instead of (2).

B. Likelihood for a Data Segment

The data $\mathbf{x}_{1:N}$ is assumed to satisfy the following property "given the positions of change points, the data in different segments is independent" [11]. These models are referred to as product partition models (PPM) [11].

Consider a data segment defined by the run-lengths $r_{n-r} = 0$ and $r_n = r$ at its two end-points. This data segment is assumed to be drawn from some distribution q, where q is an element in the set of some fixed number of distributions $\mathcal{Q}$. To get closed form expressions for likelihood, conjugate priors are defined on the parameters of q. Hyper-parameters are the parameters of the conjugate priors. The likelihood of a segment given a specific model $\mathrm{P}\left(\mathbf{x}_{n-r:n} | r_{n-r} = 0, r_n = r, \mathrm{q}\right)$, is obtained by marginalizing over the parameters of q, but it still depends on the hyper-parameters. The explicit dependence of likelihood on hyper-parameters is dropped for notational convenience in the rest of the paper. The likelihood of the data within the segment coming from this set of distributions $\mathcal{Q}$, denoted by $\mathrm{P}\left(\mathbf{x}_{n-r:n} | r_{n-r} = 0, r_n = r\right)$, is given by

$$\begin{aligned} &\mathrm{P}\left(\mathbf{x}_{n-r:n} | r_{n-r} = 0, r_n = r\right) \\ &= \sum_{\mathrm{q}} \mathrm{P}\left(\mathbf{x}_{n-r:n} | r_{n-r} = 0, r_n = r, \mathrm{q}\right) \pi\left(\mathrm{q}\right), \end{aligned} \tag{6}$$

where $\pi\left(\mathrm{q}\right)$ is the prior on the model space. Typically a uniform prior is used.

The closed form likelihood of a segment given a specific model (the first term inside the summation in (6)) is required to implement the algorithms described in Sections III and IV. The algorithms described in Sections III and IV work with any model whose likelihood can be calculated or approximated. Linear regression models used to generate simulated data and also to model the EEG data are described in Section V.

III. Online Change Point Detection Algorithm

In this section, the online Bayesian CP detection algorithm is described. This algorithm extends the work done in [5] to the general case where the data within each segment is not i.i.d.

The positions and the number of change points are inferred from the posterior distribution of the run-length auxiliary variable. Specifically, the algorithm calculates $\mathrm{P}\left(r_n = r | \mathbf{x}_{1:n}\right), \forall n = 1, 2, \cdots, N$ and $r = 0, 1, \cdots, n-1$ and infers change points from this posterior probability. This is done by calculating the joint probability of the run-length and the data observed upto that point and then by finding the desired conditional distribution. This algorithm has only one forward pass where each new data sample is used to compute a posterior probability of the run-length at that time instant. Change points can then be inferred from this posterior distribution. The proposed algorithm performs inference on a trellis shown in Fig. 2, where each node represents the value of the auxiliary variable r_n. To illustrate this algorithm, we focus on computing the joint probability of r_4 with $\mathbf{x}_{1:4}$ from the trellis in Fig. 2. Going forward, node $r_4 = 0$ can only be reached via one of the 3 nodes, $r_3 = 0$, $r_3 = 1$ and $r_3 = 2$. The joint probability for $r_4 = 0$ is the weighted average of the product of the transition probabilities for moving from the 3 nodes at $n = 3$ to $r_4 = 0$ and the likelihood of the 4^{th} time instant being a segment on its own. The weights are the

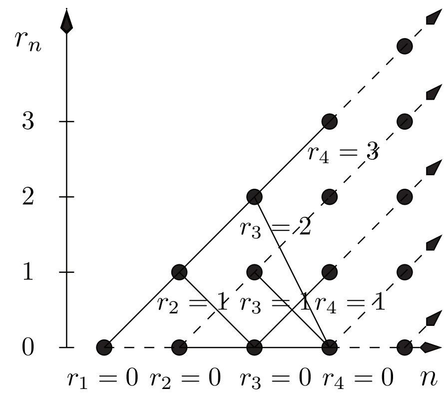


Fig. 2: Online Bayesian CP detection algorithm on trellis

joint probabilities calculated at $n = 3$. Similarly, node $r_4 = 1$ can only be reached from nodes $r_2 = 0$ and $r_2 = 1$ via node $r_3 = 0$. The joint probability for $r_4 = 1$ is similarly calculated by the sum of the product of the joint probabilities at $n = 2$, the transition probability from r_2 to $r_3 = 0$ and from $r_3 = 0$ to $r_4 = 1$ and the likelihood of $\mathbf{x}_{3:4}$ forming a segment. Finally, since there is only one path to reach node $r_4 = 3$, its joint probability is simply the product of transition probability from $r_1 = 0$ to $r_4 = 3$ and the likelihood of $\mathbf{x}_{1:4}$ forming a segment. The posterior distribution of the auxiliary variable is then calculated from Baye's rule. In the following subsection, the precise mathematical steps involved in calculating the posterior probability of run-length are derived.

A. Posterior Distribution of Run-length

First, we will compute the joint probability of the run-length and the data samples observed upto that time at each time instant. Using Markov property of change points (2) and the independence of data in different segments given the change points (PPMs), the joint probability of r_n and $\mathbf{x}_{1:n}$ can be simplified. Since $\mathrm{P}\left(r_1 = 0\right) = 1$ at time index 1, $\mathrm{P}\left(r_1 = 0, \mathbf{x}_{1:1}\right) = \mathrm{P}\left(\mathbf{x}_{1:1} | r_1 = 0\right)$. For each time index $n = 2, 3, \cdots, N$ and $r = 0$, we have

$$\begin{aligned} \mathrm{P}\left(r_n = 0, \mathbf{x}_{1:n}\right) = \sum_{i=0}^{n-2} & \underbrace{\mathrm{P}\left(r_{n-1} = i, \mathbf{x}_{1:n-1}\right)}_{\text{joint probability at } (n-1)} \\ & \cdot \underbrace{\mathrm{P}\left(r_n = 0 | r_{n-1} = i\right)}_{\text{transition probability}} \cdot \underbrace{\mathrm{P}\left(\mathbf{x}_{n:n} | r_n = 0\right)}_{\text{data segment likelihood}}. \end{aligned} \tag{7}$$

Similarly, for $n = 2, 3, \cdots, N$ and $r = n-1$, we have

$$\begin{aligned} \mathrm{P}\left(r_n = n-1, \mathbf{x}_{1:n}\right) = \mathrm{P}\left(\mathbf{x}_{1:n} | r_1 = 0, r_n = n-1\right) \\ \cdot \mathrm{P}\left(r_n = n-1 | r_1 = 0\right). \end{aligned} \tag{8}$$

Finally, for $n = 3, \cdots, N$ and $r = 1, \cdots, n-2$, we have

$$\begin{aligned} &\mathrm{P}\left(r_n = r, \mathbf{x}_{1:n}\right) \\ &= \sum_{i=0}^{n-r-2} \mathrm{P}\left(r_{n-r-1} = i, \mathbf{x}_{1:n-r-1}\right) \mathrm{P}\left(r_{n-r} = 0 | r_{n-r-1} = i\right) \\ &\quad \cdot \mathrm{P}\left(\mathbf{x}_{n-r:n} | r_{n-r} = 0, r_n = r\right) \mathrm{P}\left(r_n = r | r_{n-r} = 0\right). \end{aligned} \tag{9}$$

The posterior distribution of run-length is then calculated from the joint distribution using Bayes' rule.

$$\mathrm{P}\left(r_n = r|\mathbf{x}_1 : n\right) = \frac{\mathrm{P}\left(r_n = r, \mathbf{x}_{1:n}\right)}{\sum_{i=0}^{n-1} \mathrm{P}\left(r_n = i, \mathbf{x}_{1:n}\right)}, \tag{10}$$

for $n = 2, \cdots, N$ and $r = 0, 1, \cdots, n-1$.

B. Inferring the Change Points

The change points are inferred from the posterior probability calculated in (10) by back tracing from the final time index. This procedure is summarized below:

Set $m = 0$ and $\tau_0 = N$.

1) Find $r^\star = \arg\max_r \mathrm{P}\left(r_{\tau_m} = r|\mathbf{x}_{1:\tau_m}\right)$.
2) Increment m by 1, i.e., $m \leftarrow m + 1$.
3) Add one more change point $\tau_m = \tau_{m-1} - r^\star - 1$.
4) If $\tau_m > 0$ go to step (1), else set M = m. The inferred change points are $(\tau_M, \tau_{M-1}, \cdots, \tau_1)$.

Calculating the likelihood of a segment is required to find the posterior of auxiliary variable by (7), (8), (9). The likelihood for the case of linear regression models is given in Section V.

IV. Approximate Change Point Detection Algorithm

The number of different segment likelihoods computed by the CP detection algorithm described in the previous section grows as N^2, where N is the number of data samples. As a result the exact CP detection algorithm described in Section III has $\mathcal{O}\left(N^2\right)$ order complexity. Also the memory requirement to implement increases with N. In this section, we propose a simple approximation scheme that reduces the complexity from quadratic in N to linear in N.

The key idea is to compute the joint probability weights for only a fixed number of nodes N_p, instead of computing these weights at all $N(N-1)/2$ nodes in the trellis. The fixed number of nodes, N_p is constant with time. As a result the number of weights that are computed becomes linear with N. More specifically, at each time index n, we only retain $\mathrm{P}\left(r_n = r_i^\star|\mathbf{x}_{1:n}\right)$ where $r_i^\star \in [0, n-1]$ for $i = 1, 2, \cdots, N_p$. Note that $r_1^\star = 0$ and $r_{N_p}^\star = n - 1$. For the next time instant $(n+1)$, the run-length can be in any one of N_p+1 nodes with r values in the set $\mathcal{R} = \{0, r_1^\star+1, r_2^\star+1, \cdots, r_{N_p-1}^\star, n\}$. The joint probabilities at these $N_p + 1$ nodes are computed and the node with the smallest weight in the set $\mathcal{R} - \{0, n\}$ is discarded. Therefore at each time instant, the weights of only a fixed number of nodes are computed and stored for further calculations. Change points are inferred from the posterior using the procedure described in Section III-B. Simulation results in Section VI demonstrate that this approximation works as well as the exact algorithm.

V. Likelihood Models

The two change point detection algorithms presented in Sections III and IV are applicable for any data model. Both these algorithms use the likelihood of different segments of data in their computations (refer to (7), (8), (9)). In this section, the exact closed form likelihood expression for linear regression data models in white Gaussian noise is derived. The simulated data in Section VI and the real ECoG data in Section VII are modeled with this model.

The data in segment $\mathbf{x}_{n-r:n}$ is modeled as

$$\mathbf{x}_{n-r:n} = \mathbf{H}\boldsymbol{\beta} + \mathbf{n}, \tag{11}$$

where $\mathbf{n}$ is a $(r+1) \times 1$ i.i.d. Gaussian vector distributed with mean 0 and variance σ^2. $\mathbf{H}$ is a $(r+1) \times p$ matrix of basis functions and $\boldsymbol{\beta}$ the corresponding $p \times 1$ regression parameter vector. σ^2 is assumed to have an inverse Gamma prior with hyper-parameters $\nu/2$ and $\gamma/2$. $\boldsymbol{\beta}$ has a Gaussian prior with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{D}$, where $\mathbf{D} = \mathrm{diag}\left(\delta_1^2, \delta_2^2, \cdots . \delta_p^2\right)$. The likelihood of a data segment coming from this model conditioned on hyper-parameters is given by [14]

$$\begin{aligned}&\mathrm{P}\left(\mathbf{x}_{n-r:n}|r_{n-r} = 0, r_n = r, q\right)\\&= \pi^{(r+1)/2}\left(\frac{|\mathbf{M}|}{|\mathbf{D}|}\right)^{\frac{1}{2}} \frac{\gamma^{\nu/2}}{\left(\gamma + \|\mathbf{x}_{n-r:n}\|_K^2\right)^{\frac{(r+1+\nu)}{2}}} \frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)},\end{aligned} \tag{12}$$

where

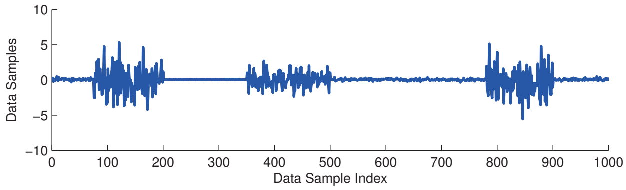
$$\begin{aligned}\mathbf{M} &= \left(\mathbf{H}^T\mathbf{H} + \mathbf{D}^{-1}\right)^{-1}\\ \mathbf{K} &= \left(\mathbf{I} - \mathbf{H}\mathbf{M}\mathbf{H}^T\right)\\ \|\mathbf{x}_{n-r:n}\|_K^2 &= \mathbf{x}_{n-r:n}^T\mathbf{K}\mathbf{x}_{n-r:n}.\end{aligned}$$

The likelihood of data within a segment coming from this model is obtained by substituting (12) into (6). The auxiliary variable posterior probabilities are calculated using the likelihood from (7), (8), (9) for the exact algorithm and from the procedure in Section IV for the approximate algorithm. Change points are then inferred using procedure described in Section III-B.

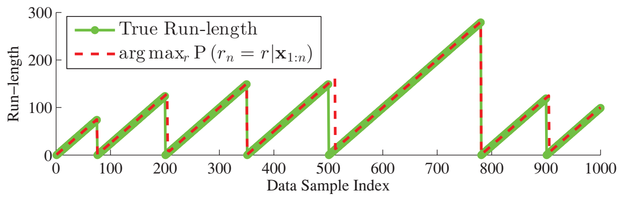
VI. Performance on Simulated Data

In this section, the performance of the two change point detection algorithms in Sections III and IV is evaluated on simulated data. The forward/backward CP detection algorithm in [7] is used as a benchmark to test the performance. The simulated data consists of 1000 data samples, with 6 change points shown in Fig. 3(a). The data in each segment is drawn from Gaussian distribution with mean 0 and some variances. The variance is different in each segment.

The data is assumed to be drawn from the model described in Section V. For this simulation, the values of hyper-parameters are $\nu = 2, \gamma = 2, \delta = 1$. Also the change point transition probabilities (2) are modeled using a geometric distribution with parameter $\lambda = 0.01$, since this leads to uniform probabilities on CP positions. Fig. 3(b) plots the true value of run-length and the maximum of the posterior probability of the run-length for the exact Bayesian CP detection algorithm in Section III. The slight offset between the two curves after change points is due to the time taken for the likelihood to drop because of the changing model. The back tracing described in



(a) Simulated 1000 data samples from Gaussian distribution with different variances



(b) True Run-Length and the Run-Length with Maximum Posterior given by Exact CP Bayesian detector in Section III

Fig. 3: Performance Evaluation on Simulated Data

Section III-B ensures that correct change points are detected. For this data, the exact algorithm, the approximate algorithm with $N_p = 10$ and the forward/backward algorithm [7] detect the same number of change points and the correct locations of all change points without any error.

VII. Epileptic Activity Segmentation

The CP detection algorithms in Section III and IV are applied to ECoG data to identify segments of activity corresponding to the different states of epileptic brain. Data is ECoG recording from a patient with epilepsy. ECoG is recorded from 154 electrodes at 1000 Hz. Fig. 4 shows a snapshot of 10 seconds of activity from 4 channels. In this work, we consider only shorter time windows of 500 samples from all 4 channels, where the channels correspond to 4 electrodes located between the temporal (T) and parietal (P) lobes of the brain. The channels are assumed to be independent and each channel is modeled using the model described in Section V. Also $\nu = 2, \gamma = 2, \delta = 1$ are taken as the values of the hyper-parameters for this simulation. Geometric distribution with $\lambda = 0.001$ is used to model the CP transition probabilities (2), since this leads to uniform probabilities on CP positions (3). The exact algorithm detected about 10 change points, where as the approximate one detected 9 change points. Further work needs to be done to extend the analysis to longer time windows and to incorporate the spatial and temporal correlations in EEG and ECoG data.

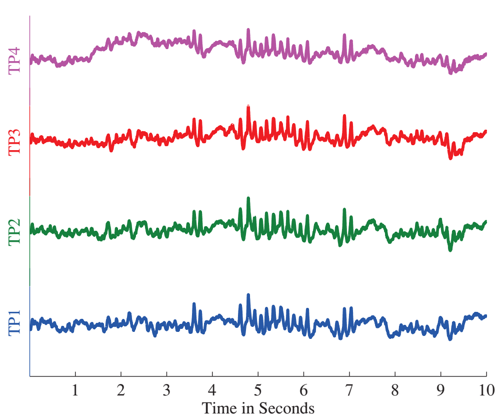


Fig. 4: Snapshot of ECoG activity from 4 channels in a 10 second window

VIII. Conclusions

In this paper, we present an online Bayesian change point detection algorithm with quadratic complexity. This algorithm works for all data models whose likelihoods can be either exactly calculated or approximated. An approximate algorithm with only linear complexity is also developed. These algorithms are then applied to segment epileptic activity into segments of different underlying brain states.

Going forward, the next step will be to jointly learn the graph connectivity model for different states conditioned on the locations of change points. The optimal spatial and temporal parameters of electrical stimulation to treat epilepsy will be determined from this graph connectivity model.

References

[1] C. H. Halpern, U. Samadani, B. Litt, J. L. Jaggi, and G. H. Baltuch, "Deep brain stimulation for epilepsy," *Neurotherapeutics*, vol. 5, no. 1, pp. 59–67, 2008.

[2] J. G. Milton, "Epilepsy as a dynamic disease: A tutorial of the past with an eye to the future," *Epilepsy & Behavior*, vol. 18, no. 12, pp. 33 – 44, 2010.

[3] K. J. Friston, "Functional and effective connectivity in neuroimaging: a synthesis," *Human brain mapping*, vol. 2, no. 1-2, pp. 56–78, 1994.

[4] R. S. Fisher, "Neurostimulation for epilepsy: Do we know the best stimulation parameters?" *Epilepsy Currents*, vol. 11, no. 6, p. 203, 2011.

[5] R. P. Adams and D. J. C. MacKay, "Bayesian online changepoint detection," University of Cambridge, Cambridge, UK, Tech. Rep., 2007.

[6] N. Seshadri and C. E. W. Sundberg, "List viterbi decoding algorithms with applications," *IEEE Transactions on Communications*, vol. 42, no. 234, pp. 313–323, 1994.

[7] P. Fearnhead, "Exact bayesian curve fitting and signal segmentation," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2160–2166, 2005.

[8] S. B. Wilson and R. Emerson, "Spike detection: a review and comparison of algorithms," *Clinical Neurophysiology*, vol. 113, no. 12, pp. 1873–1881, 2002.

[9] P. J. Green, "Reversible jump Markov chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.

[10] E. Punskaya, C. Andrieu, A. Doucet, and W. J. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 747–758, 2002.

[11] D. Barry and J. A. Hartigan, "A bayesian analysis for change point problems," *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 309–319, 1993.

[12] P. Fearnhead and Z. Liu, "On-line inference for multiple changepoint problems," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 4, 2007.

[13] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, no. 14, pp. 77 – 94, 1992.

[14] X. Xuan, "Bayesian inference on change point problems," Master's thesis, University of British Columbia, 2007.