

SEGMENT PARAMETER LABELLING IN MCMC MEAN-SHIFT CHANGE DETECTION

Alireza Ahrabian¹, Shirin Enshaeifar¹, Clive Cheong-Took² and Payam Barnaghi¹

¹Institute for Communication Systems, University of Surrey

²Department of Computer Science, University of Surrey

Email: {a.ahrabian, s.enshaeifar, c.cheongtook, p.barnaghi}@surrey.ac.uk

ABSTRACT

This work addresses the problem of segmentation in time series data with respect to a statistical parameter of interest in Bayesian models. It is common to assume that the parameters are distinct within each segment. As such, many Bayesian change point detection models do not exploit the segment parameter patterns, which can improve performance. This work proposes a Bayesian mean-shift change point detection algorithm that makes use of repetition in segment parameters, by introducing segment class labels that utilise a Dirichlet process prior. The performance of the proposed approach was assessed on both synthetic and real world data, highlighting the enhanced performance when using parameter labelling.

Index Terms— Mean-Shift Change Detection, Markov Chain Monte Carlo, Dirichlet Process, Nonparametric Bayesian.

1. INTRODUCTION

The partitioning of time series data into segments of piecewise stationary statistics is important in many fields ranging from the analysis of accelerometer data corresponding to human physical activity to the analysis of biomedical data [1]. Bayesian approaches that consider the change point transition times and number of segments as statistical parameters are particularly interesting. Such techniques enable the modeller to incorporate uncertainty in the statistical parameters being estimated, thereby providing the means of controlling the model complexity with respect to the model fit.

The work in [2] proposed a fully hierarchical Bayesian model where the uncertainty in the relevant parameters of interest were captured using appropriate prior probabilities. A reversible jump Markov Chain Monte Carlo (MCMC) technique was then used in order to obtain estimates of the relevant parameters. While closed form analytical expressions to the posterior derived in [2] is not possible, the work in [3] proposed an efficient recursive solution in order to compute such an estimate. More recently, a multivariate extension that captures changes in the dependency structure of data was proposed in [4], while the work in [5] proposed a

nonparametric Bayesian method for detecting changes in the variance of data.

Many change point detection algorithms often assume that parameters from different segments are distinct, however, infinite hidden Markov models (IHMM) and Dirichlet process mixture models (DPMM) assume that data at each time point can be generated by a parameter that belongs to a potentially infinite number of states or classes that is determined from the data set [6] [7]. In particular these methods have been introduced into change point detection algorithms [8]. However, such work often assigns a parameter (belonging to a particular state) label to each time point and not the parameters corresponding to a given segment.

In this paper we propose a mean-shift (that can be generalised to parameters of a statistical model) Bayesian change point detection algorithm that exploits repetition in segment parameters for more robust segmentation. This is achieved by extending the change point detection algorithm introduced in [2], by including parameter class labels that employ a Dirichlet process prior for identifying the number of distinct segment parameters. Simulations on synthetic and real world data verify the efficacy of the proposed method.

2. BACKGROUND

2.1. MCMC Change Point Detection

Given a set of N data points \mathbf{x} and transition times $\tau_K = [\tau_1, \dots, \tau_K]$ where $\tau_0 = 1$ and $\tau_{K+1} = N$, there exist $K + 1$ segments such that for each segment the following functional relationship between the data points (within the time indices $\tau_i + 1 \leq \tau \leq \tau_{i+1}$) and the statistical parameter ϕ_i is satisfied, that is

$$\mathbf{x}_{\tau_i+1:\tau_{i+1}} = f(\mathbf{x}_{\tau_i+1:\tau_{i+1}}, \phi_i) + \mathbf{n}_{\tau_i+1:\tau_{i+1}} \quad (1)$$

for $i = \{0, \dots, K\}$, where $\mathbf{n}_{\tau_i+1:\tau_{i+1}}$ is a set of i.i.d. zero mean Gaussian noise samples (with a specific variance of σ_i^2). The hierarchical Bayesian model introduced in [2], derived a posterior distribution such that the target parameters of interest were both K and τ_K , namely

$$p(K, \tau_K, \phi, \mathcal{H}|\mathbf{x}) \propto p(\mathbf{x}|\phi, K, \tau_K)p(\phi, K, \tau_K|\mathcal{H})p(\mathcal{H}) \quad (2)$$

where ϕ corresponds to the vector of segment parameters (that is integrated out of the posterior) and \mathcal{H} represents the

This work was supported by European Commissions in the H2020 for FIESTA-IoT project under grant agreement no. CNECT-ICT-643943 and EU H2020 IoT-Crawler agreement no. 779852.

set of hyperparameters for the relevant prior probabilities. The likelihood function used in the posterior distribution (2) assumes that the parameters ϕ_i are distinct for each segment [2], that is

$$p(\mathbf{x}|\phi, K, \tau_K) = \prod_{i=0}^K p(\mathbf{x}_{\tau_i+1:\tau_{i+1}}|\phi_i) \quad (3)$$

2.2. Dirichlet Process Mixture Model

A finite mixture model (FMM) assumes that the data is drawn from a weighted combination of distributions from the same parametric family with differing parameters [6] [9]. That is, $p(\mathbf{x}) = \sum_{v=1}^V \pi_v f_p(\mathbf{x}|\theta_v)$, where V is the number of classes, π_v is the mixing coefficient and θ_v corresponds to the class parameter/s of the probability distribution $f_p(\cdot)$. For a given data set, determining the number of classes V in a systematic way can be a challenging task. In order to overcome this problem, we must first consider the generative model of the FMM, where the i^{th} class indicator random variable c_i is introduced, that is

$$\begin{aligned} x_i|c_i, \boldsymbol{\theta} &\sim f(x_i|\theta_{c_i}) \\ c_i|\boldsymbol{\pi} &\sim \text{Discrete}(\pi_1, \dots, \pi_V) \\ \theta_v &\sim G_0 \\ \boldsymbol{\pi} &\sim \text{Dir}(\alpha/V, \dots, \alpha/V) \end{aligned} \quad (4)$$

where G_0 corresponds to the prior distribution of the parameters, $\boldsymbol{\theta} = [\theta_1, \dots, \theta_V]$ and $\boldsymbol{\pi} = [\pi_1, \dots, \pi_V]$. The mixing coefficients $\boldsymbol{\pi}$ in the model (4) govern the likelihood of selecting a given class. By employing a Dirichlet distribution prior on the mixing coefficients and taking the limit $V \rightarrow \infty$, results in the Dirichlet process (DP) mixture model [9], which is often written as

$$\begin{aligned} x_i|\theta_i &\sim f(x_i|\theta_i) \\ \theta_i &\sim G \\ G &\sim \text{DP}(G_0, \alpha) \end{aligned} \quad (5)$$

where G is drawn from the Dirichlet process with base measure G_0 . Inference of the DP mixture model is generally carried out using Gibbs sampling, where the state of the Markov chain consists of all the parameters and class labels. A particularly interesting outcome of the DP mixture model is in the inference of the class parameters (which does not require the direct specification of the number of classes). That is, the conditional posterior probability of assigning a data point to an existing class is given by

$$p(c_i = v | \mathbf{c}_{-i}, x_i, \boldsymbol{\theta}) \propto \frac{n_{-i,v}}{N-1+\alpha} L(x_i|\theta_v) \quad (6)$$

where $n_{-i,v}$ is the number of data points (excluding x_i) assigned to class v and \mathbf{c}_{-i} is a vector of class labels excluding c_i . While the conditional class posterior probability for assigning the data point x_i to a new class is given by

$$\begin{aligned} p(c_i \neq c_l \text{ for all } l | \mathbf{c}_{-i}, x_i) \\ \propto \frac{\alpha}{N-1+\alpha} \int L(x_i|\boldsymbol{\theta}) dG_0(\boldsymbol{\theta}) \end{aligned} \quad (7)$$

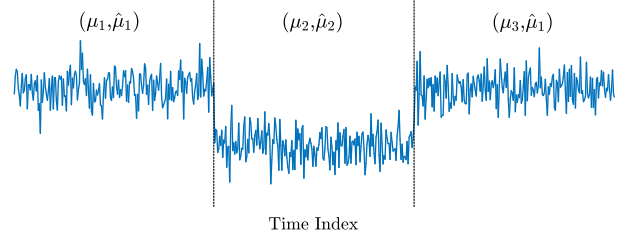


Fig. 1: Figure illustrating the difference between distinct mean μ_i and class mean $\hat{\mu}_i$ within each segment.

3. PROPOSED WORK

We propose a novel mean-shift change detection algorithm by including parameter class labels in the technique proposed in [2]. That is, given a set of data points \mathbf{x} and transition times τ_K , where the functional relationship defined in (1) for mean-shift change detection (assuming distinct parameters in each segment [2]) is given by $f(\mathbf{x}_{\tau_i+1:\tau_{i+1}}, \mu_i) = \mu_i \mathbf{1}_{1:\tau_{i+1}-\tau_i}$, with μ_i corresponding to the mean of the segment and $\mathbf{1}_{1:\tau_{i+1}-\tau_i}$ is a vector with elements equal to 1. Each segment has a distinct parameter μ_i ; however, data often exhibit parameters that repeat across different segments, therefore the model in [2] does not efficiently use the similarity in the segment parameters.

We propose to include parameter class label c_i , such that the mean parameters $\boldsymbol{\mu} = [\mu_0, \dots, \mu_K]$, are generated by the following Gaussian mixture model

$$p(\boldsymbol{\mu}|\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2, \boldsymbol{\pi}) = \sum_{v=1}^V \pi_v \mathcal{N}(\boldsymbol{\mu}|\hat{\boldsymbol{\mu}}_v, \hat{\boldsymbol{\sigma}}_v^2) \quad (8)$$

where $\hat{\boldsymbol{\mu}} = [\hat{\mu}_1, \dots, \hat{\mu}_V]$ and $\hat{\boldsymbol{\sigma}}^2 = [\hat{\sigma}_1^2, \dots, \hat{\sigma}_V^2]$ correspond to the class parameters. For each segment the functional relationship in (1) is given by $f(\mathbf{x}_{\tau_i+1:\tau_{i+1}}, \hat{\mu}_i) = \hat{\mu}_i \mathbf{1}_{1:\tau_{i+1}-\tau_i}$, where different segments now can be assigned to the same class of parameters (see Fig. 1). As a result, data in segments with the same parameter labels are combined for more robust segmentation.

3.1. Bayesian Model

The following model formally states the proposed mean-shift change point algorithm that includes parameter labelling

$$\begin{aligned} \mathbf{x}_{\tau_i+1:\tau_{i+1}}|\hat{\mu}_i, \sigma_i^2 &\sim f_j(\mathbf{x}_{\tau_i+1:\tau_{i+1}}|\hat{\mu}_i, \sigma_i^2) \\ \sigma_i^2 &\sim G_{\sigma^2} \\ \mu_i|\hat{\mu}_i, \hat{\sigma}_i^2 &\sim \mathcal{N}(\mu_i|\hat{\mu}_i, \hat{\sigma}_i^2) \\ (\hat{\mu}_i, \hat{\sigma}_i^2) &\sim G \\ G &\sim \text{DP}(G_0, \alpha) \\ \mathbf{x}_{\tau_i+1:\tau_{i+1}}|\tau_K, \mu_i &\sim f_j(\mathbf{x}_{\tau_i+1:\tau_{i+1}}|\mu_i) \\ \tau_K, K &\sim \text{Bin}(\tau_K, K|\lambda) \end{aligned} \quad (9)$$

where $\text{Bin}(\cdot)$ corresponds to a Binomial distribution, G_0 is the joint prior distribution of the class mean and variance,

G_{σ^2} is the prior distribution of the variance of the data points with the same class label and $f_j(\cdot)$ corresponds to the joint Normal distribution.

The state of the Markov chain consists of the following parameters, $\{\tau_K, K, \mathbf{c}_K, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2, \boldsymbol{\sigma}^2\}$, where $\mathbf{c}_K = [c_0, \dots, c_K]$ and $\boldsymbol{\sigma}^2 = [\sigma_1^2, \dots, \sigma_V^2]$. Inference of the parameters is carried out by using a Metropolis-Hastings-within-Gibbs sampling scheme. The Gibbs moves are performed on each parameter in the set, $\{\mathbf{c}_K, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2, \boldsymbol{\sigma}^2\}$, while a variation of the Metropolis-Hastings algorithm is used to obtain samples for the parameters $\{\tau_K, K\}$.

The Gibbs sampling procedure requires the conditional posterior distributions for all class means $\hat{\mu}_v$ and variances $\hat{\sigma}_v^2$, along with the conditional posterior variance of the data points within the same class σ_v^2 and each class label c_i . The exact derivation of these conditional posterior distributions were carried out in both [2] and [6]; where we have assumed the following class mean prior, $p(\hat{\mu}_v | \lambda_\mu, \delta) \sim \mathcal{N}(\lambda_\mu, \delta \sigma_v^2)$, that is dependent on σ_v^2 . Furthermore, the class variance has an inverse Gamma prior given by, $p(\hat{\sigma}_v^2 | \beta, \omega) \sim \mathcal{IG}(\beta, \omega)$.

The conditional posterior distribution of the parameters $\{\tau_K, K\}$ is given by, $p(\tau_K, K | \lambda, \mathbf{c}_K, \hat{\boldsymbol{\mu}}, \boldsymbol{\sigma}^2, \mathbf{x})$. Owing to the selection of the appropriate conjugate priors, we can integrate out the nuisance parameters $\{\hat{\boldsymbol{\mu}}, \boldsymbol{\sigma}^2, \lambda\}$. This is carried out by considering the following posterior distribution (for convenience we have omitted the hyperparameters $\{\lambda_\mu, \delta, \nu, \gamma\}$)

$$p(\hat{\boldsymbol{\mu}}, \boldsymbol{\sigma}^2, \tau_K, K, \lambda | \mathbf{c}_K, \mathbf{x}) \propto p(\mathbf{x} | \hat{\boldsymbol{\mu}}, \boldsymbol{\sigma}^2, K, \tau, \mathbf{c}_K) \times p(K, \tau | \lambda) p(\lambda) \prod_{v=1}^V p(\hat{\mu}_v | \lambda_\mu, \delta) p(\sigma_v^2 | \nu, \gamma) \quad (10)$$

where $p(\tau_K, K | \lambda) = \lambda^K (1 - \lambda)^{T-K-1}$, $p(\sigma_v^2 | \nu, \gamma) \sim \mathcal{IG}(\nu, \gamma)$ and $p(\lambda)$ has uniform probability between $[0, 1]$. The likelihood function is given by

$$p(\mathbf{x} | \hat{\boldsymbol{\mu}}, \boldsymbol{\sigma}^2, K, \tau_K, \mathbf{c}_K) = \prod_{v=1}^V \prod_{i: c_i=v} p(\mathbf{x}_{\tau_i+1: \tau_{i+1}} | \hat{\mu}_v, \sigma_v^2)$$

where data within segments with the same parameter label v are combined for potentially more accurate parameter estimation (in the mean squared error sense). Integration of (10) with respect to the parameters $\{\hat{\mu}_v, \sigma_v^2, \lambda\}$ results in the following expression for the conditional posterior distribution of the parameters $\{\tau_K, K\}$

$$p(\tau_K, K | \mathbf{c}_K, \mathbf{x}) \propto \prod_{v=1}^V \frac{2^{\frac{K}{2}}}{\Gamma(\frac{\nu}{2})} \Gamma(K+1) \Gamma(N-K+1) \left(\frac{\gamma}{2}\right)^{\frac{K}{2}} \times \Gamma\left(\frac{d_v + \nu}{2}\right) \pi^{-\frac{d_v}{2}} [\gamma + Y_v^T P_v Y_v]^{-\frac{d_v + \nu}{2}} (d_v + \delta^{-1})^{-\frac{1}{2}} \quad (11)$$

where Y_v is the concatenated vector of all data points with the same segment label v , d_v is the number of data points with label v , and $P_v = (\mathbf{I}_{d_v} - \mathbf{1}_{1:d_v} M_v \mathbf{1}_{1:d_v}^T)$, with $M_v = (d_v + \delta^{-1})^{-\frac{1}{2}}$. Finally, we note that there are some challenges

from drawing samples from (11) due to the dependence on \mathbf{c}_K that we have addressed in the next section.

3.2. MCMC Sampling

Samples for parameters, $\{\mathbf{c}_K, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2, \boldsymbol{\sigma}^2\}$ are obtained by drawing samples from the following posterior densities: $p(\hat{\mu}_v | \boldsymbol{\mu}, \mathbf{c}_K, \hat{\sigma}_v^2, \sigma_v^2)$, $p(\hat{\sigma}_v^2 | \boldsymbol{\mu}, \mathbf{c}_K, \hat{\mu}_v)$, $p(c_i | c_{-i}, \boldsymbol{\mu}, \hat{\mu}_v, \hat{\sigma}_v^2)$ and $p(c_i \neq c_l \text{ for all } i \neq l | c_{-i}, \boldsymbol{\mu})$ where details on exact distribution form and sampling are found in [6]. While $p(\sigma_v^2 | \mathbf{x}, \tau_K, \mathbf{c}_K)$ is found by concatenating all the data points that have the same segment label v . More details on exact distribution form and sampling can be found in [2].

In order to evaluate the conditional posterior distribution $p(\tau_K, K | \mathbf{c}_K, \mathbf{x})$ we use a modification of the Metropolis-Hastings algorithm outlined in [2] that incorporates segment labels \mathbf{c}_K . Given the current state of the Markov chain $\{\tau_K, K\}$, we select one of the steps with the following probabilities:

- birth of a change point with probability, b
- death of a change point with probability, d
- update of change point positions with probability, u

where $b = d = u$ for $0 < K < K_{max}$, and $b + d + u = 1$ for $0 \leq K \leq K_{max}$.

A birth move consists of proposing a new change point τ_{prop} with uniform probability from the existing time indices $[2, N-1]$ excluding the time indices τ_K . The proposed set of change points including τ_{prop} is given by τ_{K+1} , where the segment between the time indices $[\tau_i, \tau_{i+1}]$ ($\tau_i < \tau_{prop} < \tau_{i+1}$) with the class variable c_i is split into two new segments with two new class variables $\{\hat{c}_i, \hat{c}_{i+1}\}$. As we have not yet inferred the new class labels from the conditional class posterior distributions, we assume that the two classes $\{\hat{c}_i, \hat{c}_{i+1}\}$ are distinct and thus independent from all other segments, to circumvent the lack of information we have for assignment to an existing class. The proposed transition time is accepted with the following probability, $\alpha_{birth} = \min\{1, r_{birth}\}$, where

$$r_{birth} = \frac{p(\tau_{K+1}, K+1 | \mathbf{c}_{K+1}, \mathbf{x}) q(\tau_K | \tau_{K+1}) q(K | K+1)}{p(\tau_K, K | \mathbf{c}_K, \mathbf{x}) q(\tau_{K+1} | \tau_K) q(K+1 | K)}$$

with $q(\tau_{K+1} | \tau_K) = \frac{1}{N-K-2}$, $q(K+1 | K) = b$, $q(\tau_K | \tau_{K+1}) = \frac{1}{K+1}$ and $q(K | K+1) = d$.

The death move proposes to remove a transition time τ_{prop} , by choosing with uniform probability from the set $[\tau_1, \dots, \tau_K]$. That is, the segments $\tau_i + 1 \leq \tau \leq \tau_{i+1}$ and $\tau_{i+1} + 1 \leq \tau \leq \tau_{i+2}$ where $\tau_{i+1} = \tau_{prop}$, are combined into one segment $\tau_i + 1 \leq \tau \leq \tau_{i+2}$. Furthermore, the class labels $\{c_i, c_{i+1}\}$ are combined into one segment with a new class label (utilising the argument used for the birth of a change point, that is $\hat{c}_i \neq c_j$ for all $j \neq i$). The removal of τ_{prop} is accepted with probability $\alpha_{death} = \min\{1, r_{birth}^{-1}\}$.

The update of the change points is carried by first removing the time index τ_j in τ_K and proposing a new change point at some new location. That is, the death move is first applied followed by a birth move, for all $j = \{1, \dots, K\}$.

4. SIMULATIONS

4.1. Synthetic Data

We evaluated the performance of the proposed algorithm under two different scenarios. Namely, the first scenario considered that every segment was produced with a randomly generated mean parameter (with fixed variance), while the second scenario assumed that each segment has a mean parameter (with fixed variance) drawn from a fixed number of repeating classes. Furthermore, each segment length was drawn with uniform probability between $[20, 70]$ samples, while the number of segments were also selected with a random probability (each realisation on average had approximately 7 change points). The proposed method was compared with the following algorithms: MCMC [2], Group Fused LASSO [10] and PELTS [11]. The following parameters were selected for the proposed method: $\alpha = 2$, $\lambda_\mu = 0$, $\delta = 1$, $\nu = 2$, $\gamma = 20$, $\beta = 0.01$ and $\omega = 200$ (in general we fix all the parameters except α and γ); we note that parameters for each method were selected such that the proportion of false positives were as close as possible for each method. We evaluated the performance of the respective algorithms using the following measures: the proportion of true positives along with the absolute error in the change point location estimate.

From Table 1 it can be observed that the proposed method outperformed both the MCMC in [2] along with the Group Fused LASSO, with respect to the number of true positives and change point location estimation, for time series data with random mean parameter assignment. Furthermore, Table 2 shows that the proposed method was able to significantly outperform the MCMC and Group Fused LASSO algorithms with respect to the proportion of true positives when segmenting data with repeating mean parameters. However, the PELTS algorithm was able to outperform the proposed method with respect to the proportion of true positives and change point location estimation (as shown in both Table 1 and 2). The PELTS algorithm represents the state of the art for detecting changes with respect to a fixed statistical parameter (using dynamic programming), whereas the proposed Bayesian change point detection algorithm can be formulated in a parameter-less fashion with both estimated (from the data points) and noninformative hyper priors, along with the prediction of change points using the predictive posterior distribution which is not possible with PELTS.

Table 1: Random mean parameter assignment.

Methods	True Positives	False Positives	Error
Proposed Method	97.3%	5.9%	1.9
MCMC [2]	84.6%	9.6%	3.16
G. F. LASSO [10]	63.4%	50.6%	5.52
PELTS [11]	99.9%	6.2%	1.4

4.2. Real World Data

We also evaluated the performance of the proposed method when analysing power data pertaining to a smart plug connected to an individual’s desktop machine. The data used

Table 2: Repeating mean parameter assignment.

Methods	True Positives	False Positives	Error
Proposed Method	85.9%	5.7%	3.16
MCMC [2]	36.3%	7.5%	5.38
G. F. LASSO [10]	41.6%	31.6%	4.6
PELTS [11]	93.2%	5.9%	3.35

for this simulation was collected from the University of Surrey’s smart building testbed at the Institute for Communication systems. An interesting feature of the power data (shown in Fig. 2) is the repetitive nature of the mean parameter, corresponding to similar patterns of activity carried out by the individual.

The segmentation of the power data using the proposed method is shown in Fig. 2 (upper panel), where it can be observed that the proposed method was able to effectively segment the data (PELTS had similar performance to the proposed method; however, owing to limited space we have not included the results). In particular it should be noted that, for data points following time indices $t > 300$, the relative difference between the mean parameter of adjacent segments was much smaller. As a result, the method in [2] was not able to effectively separate the segments as shown in Fig. 2 (lower panel). Therefore demonstrating the advantage of using parameter labelling for each segment.

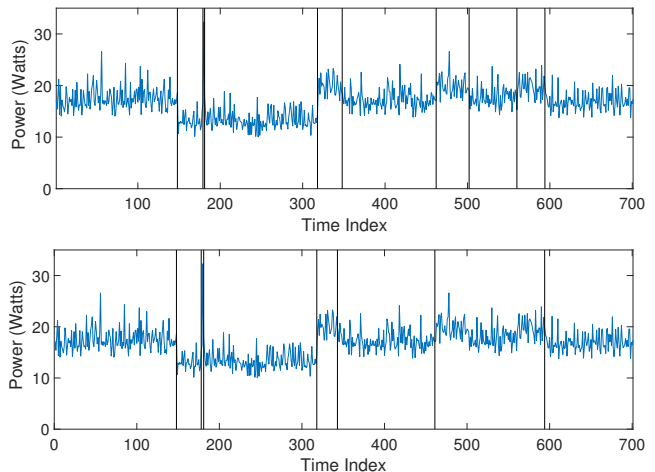


Fig. 2: Detecting changes in power data with repeating patterns, using the proposed method (upper panel) and MCMC method in [2] (lower panel).

5. CONCLUSIONS AND FUTURE WORK

This work proposes a mean-shift change point detection algorithm that captures parameter repetition for more robust time series segmentation. This was achieved by labelling parameters for a given segment and by employing a Dirichlet process prior. We have shown the advantage of the proposed method on synthetic and real world data. Future work will extend the proposed method for a wider class of time series models as well as including hyperparameter updates.

6. REFERENCES

- [1] E. Sejdic, C. M. Steele, and T. Chau, "Segmentation of dual-axis swallowing accelerometry signals in healthy subjects with analysis of anthropometric effects on duration of swallowing activities," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1090–1097, 2009.
- [2] E. Punsakaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 747–758, 2002.
- [3] P. Fearnhead, "Exact Bayesian curve fitting and signal segmentation," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2160–2166, 2005.
- [4] X. Xuan and K. Murphy, "Modeling changing dependency structure in multivariate time series," *International Conference on Machine Learning*, pp. 1055–1062, 2007.
- [5] A. A. Hensley and P. M. Djuric, "Nonparametric learning for hidden Markov Models with preferential attachment dynamics," *International Conference on Acoustics, Speech and Signal Processing*, pp. 3854–3858, 2017.
- [6] C. E. Rasmussen, "The infinite Gaussian mixture model," *Advances in Neural Information Processing Systems 12*, pp. 554–560, 2000.
- [7] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "Bayesian nonparametric inference of switching dynamic linear models," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1569–1585, 2011.
- [8] S. I. M. Ko, T. T. L. Chong, and P. Ghosh, "Dirichlet process hidden Markov multiple change-point model," *Bayesian Analysis*, vol. 10, no. 2, pp. 275–296, 2015.
- [9] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, vol. 2, no. 2, pp. 249–265, 2000.
- [10] K. Bleakley and J.-P. Vert, "The group fused LASSO for multiple change-point detection," *arXiv:1106.4199*, 2011.
- [11] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 500, no. 107, pp. 1590–1598, 2012.