

Detecting Changes in the Variance of Multi-Sensory Accelerometer Data Using MCMC

Alireza Ahrabian, Tarek Elsaleh, Yasmin Fathy and Payam Barnaghi

Institute for Communication Systems
University of Surrey
Guildford, UK

Abstract— An important field in exploratory sensory data analysis is the segmentation of time-series data to identify activities of interest. In this work, we analyse the performance of univariate and multi-sensor Bayesian change detection algorithms in segmenting accelerometer data. In particular, we provide theoretical analysis and also performance evaluation on synthetic data and real-world data. The results illustrate the advantages of using multi-sensor variance change detection in the segmentation of dynamic data (e.g. accelerometer data).

Index Terms— Variance Change Detection, Multivariate Change Detection.

I. INTRODUCTION

Many real world data streams can be characterised by segments in which the statistical parameters of the data are approximately stationary. In particular the detection of changes in variance of univariate time series data has found applications in fields ranging from biomedical engineering [1] to financial time series analysis [2]. With the advent of low cost multi-sensor data acquisition devices, the need for the processing of multi-channel datasets for detecting the joint changes in such sensors has become apparent.

Univariate algorithms for estimating the transition times for sensory data with piecewise constant changes in the variance are broadly separated into two categories: namely i) algorithms based on statistical significance tests [2] and ii) Bayesian methods, that is, identifying a posterior distribution and obtaining the corresponding maximum a posteriori (MAP) estimates of the change point locations and other parameters of interest [3][4]. These techniques have found a wide range of applications in the segmentation of time series data, however, a need had arisen for the development of multi-sensor extensions of such algorithms for more accurate segmentation. The work in [5] proposed to detect the joint changes in the variance from two different sensors by modelling correlations between the channels; where it was assumed that the change point locations for each sensor may occur at differing time instants. While the proposed algorithm is able to exploit channel dependencies in order to estimate change point locations, the method however does not utilise common transition times which increases the number of parameters to estimate. The work in [6] proposed a set of multivariate (multi-sensor) covariance models, ranging from: 1) the independent variance model, that is, identifying segments through changes in the variance of the multi-sensor data, 2) the sparse precision matrix model, that is, identifying changes in the correlation structure of the data.

In this work we seek to segment a 3-axis accelerometer data pertaining to human activity by combining a variation of the multi-sensor independent variance model proposed in [6]

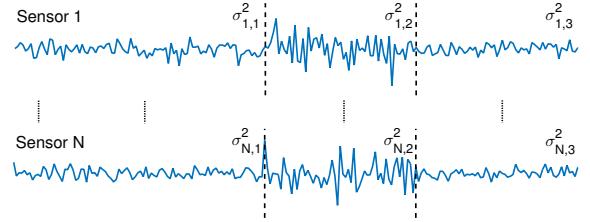


Fig. 1: N sensors with variance $\sigma_{n,k}^2$, where $n = 1, \dots, N$ is the sensor index and $k = 1, 2, 3$ is the segment.

along with the Markov Chain Monte Carlo (MCMC) scheme proposed in [3]. In particular, we propose the common variance model (a variation of the independent variance model), such that both the derivation of the posterior is simplified and theoretical analysis of the common variance model (by extension the independent variance model) can be carried out.

II. MULTIVARIATE CHANGE POINT ALGORITHM

Given a multivariate signal, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_T] \in \mathbb{R}^{N \times T}$, where N corresponds to the total number of sensors, such that each channel vector $\mathbf{x}_t \in \mathbb{R}^N$ is distributed according to a multivariate Gaussian distribution with a covariance Σ_k corresponding to a particular segment k , namely, $\mathbf{x}_t \sim \mathcal{N}(0, \Sigma_k)$. There exists an unknown sequence of common transition time points, $\tau_K = [\tau_1, \dots, \tau_{K-1}]$, such that $\tau_0 = 0$ and $\tau_K = T$, where \mathbf{x}_t belongs to segment index k if $\tau_{k-1} < t \leq \tau_k$ and K is the total number of segments (please refer to Fig. 1 for an example).

The independent variance model in [6] utilises the following covariance, $\Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$, where given a vector, $\text{diag}(\cdot)$ creates a diagonal matrix. In this work we propose to utilise the common variance model in order to simplify the derivation and theoretical analysis of the multi-sensor change point model. That is, we propose to model the covariance in each segment as follows, $\Sigma_k = \sigma_k^2 \mathbf{I}_{N \times N}$, where $\mathbf{I}_{N \times N}$ is the identity matrix. The common variance model assumes that the variance $\sigma_{n,k}^2$ for each sensor and segment is effectively drawn from a distribution with mean σ_k^2 . It should be noted that, the common variance model is more restrictive than the independent variance model. The change in variance between two adjacent segments across all the sensors must either be increasing or decreasing. However, by using the common variance model, we can provide theoretical performance analysis of the multi-sensor model. The estimation of the transition times is based on the Bayesian inference approach proposed in [3], where the posterior distribution for both the number and

locations of the transition points was first derived and then MCMC sampling of the posterior was carried out in order to obtain estimates of the respective parameters. To this end, the posterior distribution of the parameters $\{K, \underline{\tau}_K, \underline{\sigma}_K^2, \lambda\}$, where $\underline{\sigma}_K^2 = [\sigma_1^2, \dots, \sigma_K^2]$, is given by

$$p(K, \underline{\tau}_K, \underline{\sigma}_K^2, \lambda | \mathbf{X}) \propto p(\mathbf{X} | K, \underline{\tau}_K, \underline{\sigma}_K^2) p(K, \underline{\tau}_K | \lambda) p(\lambda) p(\underline{\sigma}_K^2) \quad (1)$$

where $p(\mathbf{X} | K, \underline{\tau}_K, \underline{\sigma}_K^2)$ is the likelihood function, $p(K, \underline{\tau}_K | \lambda)$ corresponds to the prior distribution on both the number of transition times as well as the transition time points, $p(\underline{\sigma}_K^2)$ is the prior on the variance within each segment and $p(\lambda)$ is unit interval uniformly distributed prior on λ . A binomial prior probability is utilised for K and the instance of the transition times $\underline{\tau}_K$, $p(K, \underline{\tau}_K | \lambda) = \lambda^K (1 - \lambda)^{T-K-1}$. An inverse Gamma distribution prior (with $\alpha = 1$ and β a user defined parameter) is utilised for the joint variance for each segment $\sigma_k^2 \sim \text{IG}(\alpha, \beta)$. The posterior density (1) is proportional up to a normalizing constant to the joint probability, $p(K, \underline{\tau}_K, \underline{\sigma}_K^2, \mathbf{X})$, where by integrating out the nuisance parameter $\underline{\sigma}_K^2$ and λ , we obtain an expression for the posterior density for the parameters $\{K, \underline{\tau}_K\}$, that is

$$p(K, \underline{\tau}_K | \mathbf{X}) \propto \left[\prod_{k=1}^{K-1} 2\pi^{-\frac{K}{2}(\tau_{k+1} - \tau_k - 1)} \frac{\beta^\alpha \Gamma(\frac{K}{2}(\tau_{k+1} - \tau_k - 1) + \alpha)}{\Gamma(\alpha)} \right] \times \left(\sum_{j=\tau_k}^{\tau_{k+1}-1} \mathbf{x}_j^T \mathbf{x}_j + 2\beta \right)^{-\left(\frac{K}{2}(\tau_{k+1} - \tau_k - 1) + \alpha\right)} \Gamma(K+1) \Gamma(T-K) \quad (2)$$

The samples were drawn from the posterior distribution (2) by using the Metropolis-Hastings algorithm as outlined in [3] (however we do not update the hyperparameters), where the MAP criterion was then used to determine both the number of segments and change point locations.

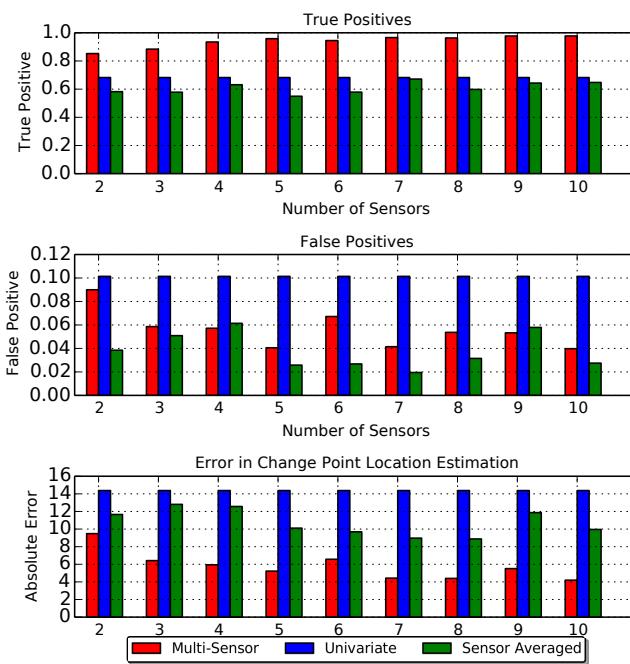


Fig. 2: The performance of the multi-sensor (red bar), sensor averaged (green bar) and single sensor (blue bar) methods. Performance measures include: (upper panel) true positives, (middle panel) false positives and (lower panel) absolute error in change point location estimation.

III. PERFORMANCE ANALYSIS

In this section we provide theoretical justification for the asymptotic performance advantage of utilising the multi-sensor change detection model over averaging across the sensor observations and using the univariate model in [3] (it should be noted that it is trivial to prove the performance improvement of the multi-sensor model against the segmentation of only a single sensor). That is, consider the following scenario: a set of N sensors with the variance in the first segment given by, $\{\sigma_{1,1}^2, \dots, \sigma_{N,1}^2\}$ and the variance in the second segment given by $\{\sigma_{1,2}^2, \dots, \sigma_{N,2}^2\}$, with $\sigma_{n,2}^2 > \sigma_{n,1}^2, \forall n = 1, \dots, N$. That is, each sensor may obtain a variance that is distinct from all other sensors, however, the variance between adjacent segments is strictly increasing (the same argument can be made for the variance decreasing between the two segments). The data points are distributed as follows, $x_{n,k}(t) \sim \mathcal{N}(0, \sigma_{n,k}^2)$ for each segment $k = 1, 2$, with both segments $\sigma_{n,k}^2 \sim \mathcal{D}(\sigma_k^2, \bar{\gamma}_k)$ being drawn from a distribution, \mathcal{D} , such that the sample mean and variance corresponds to the both first moment, σ_k^2 , and dispersion, $\bar{\gamma}_k$, of a distribution respectively (an example of such a distribution is the uniform distribution). The maximum likelihood estimator (that is, the first term in the right hand side of (1)) for the joint variance between the sensors is given by $\hat{\sigma}_k^2 = \frac{1}{N} \sum_{n=1}^N \hat{\sigma}_{n,k}^2$ for each segment $k = 1, 2$, where $\hat{\sigma}_{n,k}^2 = \frac{1}{T-1} \sum_{t=1}^T x_{n,k}^2(t)$, assuming that the data points $x_{n,k}(t)$ are zero mean. While the variance estimator for the average across the sensors is given by, $\hat{\sigma}_{a,k}^2 = \frac{1}{T-1} \sum_{t=1}^T \left(\frac{1}{N} \sum_{n=1}^N x_{n,k}(t) \right)^2$. The corresponding expectation of the respective estimators of the variance conditioned on $\underline{\sigma}_k^2 = [\sigma_{1,k}^2, \dots, \sigma_{N,k}^2]$, are given by

$$\begin{aligned} E\{\hat{\sigma}_k^2 | \underline{\sigma}_k^2\} &= E\left\{ \frac{1}{N} \sum_{n=1}^N \hat{\sigma}_{n,k}^2 \middle| \underline{\sigma}_k^2 \right\} = \frac{1}{N} \sum_{n=1}^N \sigma_{n,k}^2. \\ E\{\hat{\sigma}_{a,k}^2 | \underline{\sigma}_k^2\} &= E\left\{ \frac{1}{T-1} \sum_{t=1}^T \left(\frac{1}{N} \sum_{n=1}^N x_{n,k}(t) \right)^2 \middle| \underline{\sigma}_k^2 \right\} \\ &= \frac{1}{N^2} \sum_{n=1}^N \sigma_{n,k}^2. \end{aligned} \quad (3)$$

As a result, consider the expected difference (using the iterative property of conditional expectation on (3)) between two adjacent segments, that is, $E\{\hat{\sigma}_2^2 - \hat{\sigma}_1^2\} = \sigma_2^2 - \sigma_1^2$ and $E\{\hat{\sigma}_{a,2}^2 - \hat{\sigma}_{a,1}^2\} = \frac{1}{N}(\sigma_2^2 - \sigma_1^2)$, for both the common variance and sensor averaged models, respectively. As the number of sensors increases, $N \rightarrow \infty$, we observe that, $E\{\hat{\sigma}_{a,2}^2 - \hat{\sigma}_{a,1}^2\} \rightarrow 0$, implying that the estimator of the variance using averaging will degrade in performance as the number of sensors increases asymptotically; since it is required that the expected difference between the variance of two adjacent segments needs to be greater than zero for any algorithm to detect a change point.

IV. SIMULATIONS

The performance of the multi-sensor variance change point detection algorithm was verified on both synthetic and real world data. In particular a comprehensive comparison in the performance is carried out with: 1) only a single sensor and 2) averaging across all the sensors univariate change point detection algorithm [3].

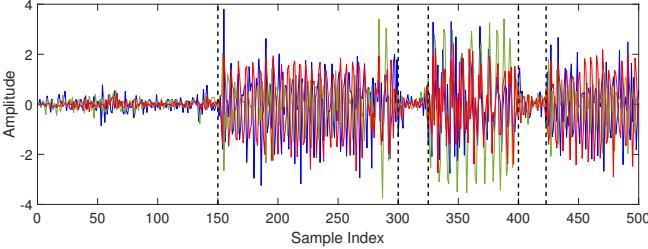


Fig. 3: An example of a segment of the tri-accelerometer data with corresponding changes in variance identified by visual inspection (dashed black lines).

A. Synthetic Data

The first synthetic simulation includes a multi-sensor signal, $\mathbf{x}(n)$, consisting of data points drawn from a multivariate Gaussian distribution with instantaneous changes in the variance at random points in time. In particular, the instantaneous power changes for each segment and sensor were selected randomly along with both the number of samples in the segment and the total signal length; where the total number of samples was selected uniformly between [200, 500] and each segment (common to the multivariate data) length was selected with uniform probability between the interval [50, 100]. The variance for each segment was dependant on the variance of the previous segment; specifically, a scale increase or decrease (for all sensors) was selected with equal probability. Furthermore, the magnitude of the scale decrease was selected with uniform probability between [0.4, 0.65] and for a scale increase [1.4, 2.4] (for each channel both the scale increase and decrease was drawn independently).

We evaluated the performance of the respective algorithms using the following measures: the proportion of false positives and true positives along with the absolute error in the change point location estimate. The following parameters were selected for the respective algorithms: 1) Multi-sensor: $\beta = 5$, 2) single sensor: $\beta = 5$ and 3) sensor averaged: $\beta = 0.00005$. Furthermore, data pre-processing procedures were first applied, that is, data channel de-correlation along with the normalization of the respective channels.

From Fig. 2, it can be observed that the multi-sensor model significantly outperformed, both the sensor averaged and single sensor univariate change detection algorithms. In particular, the multi-sensor model had lower change point location estimation error, when compared with both the sensor averaged and single sensor univariate algorithms. Furthermore, the proportion of true positives was significantly higher than both sensor averaged and single sensor univariate algorithms. While the proportion of false positives was approximately equal to the sensor averaged univariate algorithm.

B. Accelerometer Data

This section assesses the performance of the multi-sensor model on real world tri-axial accelerometer data obtained from UCI data repository [7][8]. We utilised the accelerometer data pertaining to the ‘brush teeth’ activity, where the subject had an accelerometer attached to the wrist, while performing the activity. The pre-processing of the data set consisted of first segmenting the data into non-overlapping windows of sample length 500; where within each non-overlapping window first

TABLE I: Performance of the multi-sensor, single sensor and sensor averaged models, when processing accelerometer data.

Methods	True Positives	False Positives
Multi-Sensor	0.87	0.15
Sensor Averaged	0.83	0.26
Single Sensor	0.72	0.27

differencing, channel wise de-correlation and normalization were then carried out, such that the respective algorithms were then applied to the pre-processed data. As the data did not include labels corresponding to the change in variance, we inferred this from the data sets (example is shown in Fig. 3), where variance-wise segmentation of the data set can be inferred via visual inspection.

The following parameters were selected for the respective algorithms: 1) Multi-sensor: $\beta = 7.5$, 2) single sensor: $\beta = 5$ and 3) sensor averaged: $\beta = 1.5$. Furthermore, we evaluated the performance of the respective algorithms, using: true positives and false positives. From Table I, it can be observed that the multi-sensor model outperformed both the single sensor and sensor averaged univariate algorithms, with respect to both the true positives and more significantly having a lower number of false positives.

V. CONCLUSIONS

In this work we assessed the performance of both univariate and multi-sensor Bayesian change detection algorithms in the segmentation of accelerometer data. Namely, we demonstrated that by exploiting the inter-sensor dependencies of tri-axial accelerometers, more accurate activity segmentation based on the variance can be achieved. In future work we will seek to include labels in the segmentation of the accelerometer data corresponding to activities of interest.

VI. ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Commissions in the H2020 for FIESTA-IoT project under grant agreement no. CNECT-ICT-643943.

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] C. Inclan and G. C. Tiao, “Use of cumulative sums of squares for retrospective detection of changes of variance,” *Journal of the American Statistical Association*, vol. 89, no. 427, pp. 913–923, 1994.
- [3] E. Punskaya, 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.
- [4] M. Lavielle and E. Lebarbier, “An application of MCMC methods to the multiple change-points problem,” *Signal Processing*, vol. 81, pp. 39–53, 2001.
- [5] M. Chabert, J. Tourneret, and M. Coulon, “Joint detection of variance changes using hierarchical Bayesian analysis,” *IEEE Workshop on Statistical Signal Processing*, pp. 613–616, 2003.
- [6] X. Xuan and K. Murphy, “Modeling changing dependency structure in multivariate time series,” *Proceedings of the 24th International Conference on Machine Learning*, p. 10551062, 2007.
- [7] M. Lichman, “UCI machine learning repository,” 2013. [Online]. Available: <http://archive.ics.uci.edu/ml>.
- [8] B. Bruno, F. Mastrogiovanni, A. Sgorbissa, T. Vernazza, and R. Zaccaria, “Analysis of human behavior recognition algorithms based on acceleration data,” *IEEE International Conference on Robotics and Automation*, pp. 1602–1607, 2013.