Quickest Change Detection in Hidden Markov Models for Sensor Networks

Cheng-Der Fuh[1] and Yajun Mei[2]

[1] Institute of Statistical Science, Academia Sinica, and
Graduate Institute of Statistics, National Central University
128, Academia Rd. Sec. 2, Taipei 11529, Taiwan
stcheng@stat.sinica.edu.tw

[2] School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA, U.S.A.
Email: ymei@isye.gatech.edu

”Invited session: Sequential Methods in Sensor Networks. Organizers: Y. Mei and C.-D. Fuh.”

Abstract. The decentralized quickest change detection problem is studied in sensor networks, where a set of sensors receive observations from a hidden Markov model (HMM) $\mathbf{X}$ and send sensor messages to a central processor, called the fusion center, which makes a final decision when observations are stopped. It is assumed that the parameter θ in the hidden Markov model for $\mathbf{X}$ changes from θ_0 to θ_1 at some unknown time. The primary goal of this paper is to investigate how to choose the best stationary quantizers in the context of quickest change detection in sensor networks. A closely related goal of this paper is to report the distribution of the run length to false alarm for HMM in some scenarios.

Keywords. Asymptotic optimality, CUSUM, multi-sensor, quantization, sensor networks, sequential detection.

1 Introduction

Sensor networks have many important applications including environment monitoring, target detection and tracking, and security and surveillance systems [cf. Blum et al. (1997), and Viswanathan and Varshney (1997)]. Since communication costs are usually dominant in sensor networks, especially for wireless sensor networks, it is crucial to reduce communication costs of sensor networks. One naive approach is to transmit all raw data from each sensor node to a central processor, called the *fusion center*, which then undertakes the heavy task of making the relevant decisions. A standard mathematical formulation to transmit the raw data is to require that the sensor messages belong to a finite alphabet (possibly binary).

One of the most important topics in sensor networks is how to take into account complex spatio-temporal structure and relational interactions among sensors, to which hidden Markov model (HMM) is a powerful statistical tool. A hidden Markov model is a doubly stochastic process with an underlying stochastic process that is not directly observable (i.e., hidden) but can be observed only through another set of stochastic processes that produces the sequence of observations. The application to sensor networks can be found in Dogandzic and Zhang (2006), and Huang and Dey (2006).

The primary goal of this article is to investigate decentralized quickest change detection problems in hidden Markov models for sensor networks. It is assumed that the raw sensor observations are characterized by a hidden Markov model with some (possibly vector) parameter θ, and at some unknown (possibly ∞) time ν, the parameter θ changes from one value to another value. Such a change can occur in either the Markov chain or the conditional density function of raw sensor observations or both (although we do assume that the change is detectable by monitoring raw sensor observations, since the true state of HMM is unobservable). The goal is to detect the true change as soon as possible over all possible thresholds λ_k's at the sensors and over all possible detection scheme at the fusion center, under a restriction on the frequency of false alarms. A closely related goal of this paper is to report the distribution of the run length to false alarm for HMM in some scenarios. This is useful to clarify whether the average run length to false alarms are appropriate criterion in HMM models.

2 Problem formulation

Assume that the sensor observations $Y_{k,n}$'s can be characterized by the hidden Markov models (HMM). Specifically, let $\mathbf{X} = \{X_n, n \geq 0\}$ be a discrete time Markov chain on a finite state space $D = \{1, \cdots, d\}$, with transition probability matrix

$$P_\theta = \begin{bmatrix} p_\theta(1,1) & \cdots & p_\theta(1,d) \\ \vdots & \ddots & \vdots \\ p_\theta(d,1) & \cdots & p_\theta(d,d) \end{bmatrix}, \tag{1}$$

where $p_\theta(i,j) = \mathbf{P}_\theta(X_{n+1} = i | X_n = j), 1 \leq i, j \leq d$. We also assume that the initial distribution of the chain $\mathbf{X}$ is the stationary distribution of $\mathbf{X}$

$$\pi_\theta = (\pi_\theta(1), \cdots, \pi_\theta(d))^t, \tag{2}$$

where t denotes transpose.

At time n, it is assumed that the distribution of the sensor observation $Y_{k,n}$ taken at sensor S_k is completely determined by the Markov chain X_n. That is, we assume that the conditional distribution of $Y_{k,n}$ given $X_1 = x_1, \cdots, X_n = x_n$ is $f^{(k)}_{x_n,\theta}(\cdot)$. Moreover, we also assume that given $X_1 = x_1, \cdots, X_n = x_n$, the sensor observations $Y_{1,n}, \cdots, Y_{K,n}$ are conditionally independent.

Definition 1. We call a process $\{Y_{k,n},\ n \geq 0,\ k = 1, \cdots, K\}$ a hidden Markov model (HMM) for sensor networks if there is a Markov chain $\{X_n, n \geq 0\}$ satisfies (1), (2), and $Y_{k,n}$ satisfies the above conditions.

After taking the raw sensor observation $Y_{k,n}$, we assume that each of K sensors S_k quantizes the raw sensor observations and sends the quantized data as a sensor message $U_{k,n}$ at time n to the fusion center, due to data compression and limitations of channel bandwidth. For simplicity, a stationary binary quantizer is used at each sensor S_k :

$$U_{k,n} = \phi_{k,n}(Y_{k,n}) = \begin{cases} 1, & \text{if } Y_{k,n} \geq \lambda_k \\ 0, & \text{if } Y_{k,n} < \lambda_k \end{cases}, \tag{3}$$

where the thresholds $(\lambda_1, \cdots, \lambda_k)$ are constants that needs to be chosen to optimize the network performance. To further reduce communication costs, in practice, $U_{k,n} = 0$ can be represented by the situation when the sensor does not send any sensor messages to the fusion center, i.e., the sensor is silent. The fusion center then uses the stream of messages $U_{k,n}$'s from the sensors as inputs to make a final decision.

To formalize the change point detection problem, let $\mathbf{Y}_n = (Y_{1,n}, \cdots, Y_{K,n})$ and $\mathbf{U_n} = (U_{1,n}, \cdots, U_{K,n})$, and denote by $\mathbf{P}_\infty$ the probability measure when there is no change (i.e., $\nu = \infty$). For each possible change-point $\nu = 1, 2, 3 \cdots$, suppose there is a (new) post-change probability measure $\mathbf{P}_\nu$ on the sample space of $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \cdots$ such that the joint marginal $\mathbf{P}_\nu$-distribution of $\mathbf{Y}_1, \cdots, \mathbf{Y}_{\nu-1}$ is equal to the joint marginal $\mathbf{P}_\infty$-distribution of $\mathbf{Y}_1, \cdots, \mathbf{Y}_{\nu-1}$. Then the decentralized quickest change detection problem can be formally stated as sequentially testing the null hypothesis

$$H_0 : \mathbf{P}_\infty \text{ is true} \qquad \text{(i.e., no change)}$$

against a composite alternative hypothesis

$$H_1 : \mathbf{P}_\nu \text{ is true for some } \nu = 1, 2, \cdots \text{ (i.e., a change occurs).}$$

In many applications with dependent observations, the definition of $\mathbf{P}_\infty$ is obvious, but the definitions of $\mathbf{P}_\nu$'s can be non-trivial since they may depend on specific applications or assumptions. In this paper, we adapt the following formulation. For given random variables $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \cdots$, the likelihood ratio is defined as

$$\frac{d\mathbf{P}_\nu}{d\mathbf{P}_\infty}(\mathbf{Y}_1, \cdots, \mathbf{Y}_n) = \begin{cases} 1, & \text{if } n \leq \nu - 1; \\ \frac{p_{\theta_1}(\mathbf{Y}_1, \cdots, \mathbf{Y}_n)/p_{\theta_1}(\mathbf{Y}_1, \cdots, \mathbf{Y}_{\nu-1})}{p_{\theta_0}(\mathbf{Y}_1, \cdots, \mathbf{Y}_n)/p_{\theta_0}(\mathbf{Y}_1, \cdots, \mathbf{Y}_{\nu-1})}, & \text{if } n \geq \nu. \end{cases} \tag{4}$$

for any $n \geq 1$, where, as conventional, $\mathbf{P}_\theta(\mathbf{Y}_0) = 1$ for $\theta = \theta_0$ or θ_1 and $I\{A\}$ is the indicator function.

Note that the fusion center makes decisions based on the sensor messages (vector observations) $\mathbf{U}_n$. If the sensor quantizers are stationary and pre-determined (i.e., no design issues for the senor messages), then the fusion center faces a classical (centralized) quickest change detection in HMM. In the problem of testing the null hypothesis $H_0 : \mathbf{P}_\infty$ is true, against a composite alternative hypothesis $H_1 : \mathbf{P}_\nu$ is true for some $\nu = 1, 2, \cdots$, it is easy to see that for the first n sensor message vectors, $(\mathbf{U}_1, \cdots, \mathbf{U}_n)$, the logarithm of the corresponding generalized likelihood ratio (GLR) statistic is

$$W_n = \max_{1 \leq \nu < \infty} \log \frac{d\mathbf{P}_\nu}{d\mathbf{P}_\infty}\big(\mathbf{U}_1, \cdots, \mathbf{U}_n\big) = \max\Big(0, \max_{1 \leq i \leq n} \log \frac{d\mathbf{P}_i}{d\mathbf{P}_\infty}\big(\mathbf{U}_1, \cdots, \mathbf{U}_n\big)\Big), \tag{5}$$

since in the change-point problem $\frac{d\mathbf{P}_\nu}{d\mathbf{P}_\infty}\big(\mathbf{U}_1, \cdots, \mathbf{U}_n\big) = 1$ if $\nu > n$. Hence a natural detection scheme is to declare that a change occurs if the log-GLR statistic W_n is too large. That is, it is natural to consider the GLR detection scheme that raises an alarm at the stopping time

$$N := N(c_\gamma) := \inf\{n : W_n \geq c_\gamma\}, \tag{6}$$

where W_n is defined in (5) and c_γ is chosen such that $\mathbf{E}_\infty N(c_\gamma) = \gamma$.

When the probability measures $\mathbf{P}_\nu$'s are those defined in (4), it is easy to see that the statistic W_n in (5) can be written as

$$W_n = \max_{0 \leq i \leq n} (\log S_n^{\mathbf{u}} - \log S_i^{\mathbf{u}}), \tag{7}$$

where the likelihood ratio $S_n^{\mathbf{u}} := \frac{p_n(\mathbf{U}_1, \cdots, \mathbf{U}_n; \theta_1)}{p_n(\mathbf{U}_1, \cdots, \mathbf{U}_n; \theta_0)}$ for $n \geq 1$, $S_0^{\mathbf{u}} = 1$ as conventional, and

$$p_n(\mathbf{U}_1, \cdots, \mathbf{U}_n; \theta) = \sum_{x_0, \cdots, x_n = 1}^{d} \pi_\theta(x_0) \prod_{i=1}^{n} \Big[p_\theta(x_{i-1}, x_i) \prod_{k=1}^{K} f_{x_i, \theta}^{(k,U)}(U_{k,i}) \Big] \tag{8}$$

represents the joint distribution of the first n sensor message vectors, $(\mathbf{U}_1, \cdots, \mathbf{U}_n)$, when the parameter of the HMM is θ and there are no changes. Here $f_{x_i,\theta}^{(k,U)}(U_{k,i})$ is the induced probability mass function of $U_{k,i} = I(Y_{k,i} \geq \lambda_i)$ when $Y_{k,i}$ has a distribution $f_{x_i,\theta}^{(k)}(\cdot)$.

From (7), it is obvious that when the probability measures $\mathbf{P}_\nu$'s are defined in (4), then W_n enjoys the recursive formula of the form

$$W_n = \max\Big(W_{n-1} + (\log S_n^{\mathbf{u}} - \log S_{n-1}^{\mathbf{u}}), 0\Big) \qquad \text{for } n \geq 1, \tag{9}$$

which is similar to that of the classical CUSUM procedure. Hence, in the following we just call the GLR-based detection scheme N in (6) as the CUSUM procedure.

In this paper, we are interested in determining the asymptotically optimal fusion center schemes when the stationary binary quantizers in (3) are given. In this case, the fusion center faces the classical change-point detection problems in hidden Markov models based on the quantized binary vector $\mathbf{U}_n = (U_{1,n}, \cdots, U_{K,n})$. A natural question is how to choose good stationary thresholds. A simple approach is based on the asymptotic properties [cf. Fuh (2003)] as $\gamma \to \infty$, $\overline{\mathbf{E}}_1(\tau) = (1 + o(1))\frac{\log \gamma}{K^{\mathbf{u}}(\lambda)}$, where the $K^{\mathbf{u}}(\lambda)$ is the Kullback-Leibler information number induced on $\mathbf{U}_n$ when the stationary thresholds are $\lambda = (\lambda_1, \ldots, \lambda_K)$. Thus, the Kullback-Leibler information number (or relative entropy) $K^{\mathbf{u}}(\lambda)$ plays an essential role in our setting. By Fuh (2003), and Fuh and Mei (2008) based on $\mathbf{U}_n$, an asymptotic optimal rule at the fusion center is the CUSUM scheme with the thresholds of selecting λ_k which maximize the Kullback-Leibler information number $K^{\mathbf{u}}(\lambda)$.

Given the difficulty of calculating the Kullback-Leibler information number for HMM, it is intractable to find λ that maximizes $K^{\mathbf{u}}(\lambda)$, locally or globally (except some degenerate scenarios). This may not satisfy practitioners or researchers who want to compute a reasonable (not necessarily the best) numerical value of λ to provide a benchmark in their specific application. A natural approach is then to find λ to optimize an *approximation* of $K^{\mathbf{u}}(\lambda)$ that can be *solved* in the sense of generating numerical values. One such idea is presented in Fuh and Mei (2008). A major thrust of this paper is to develop a feasible computational method to simulate the optimal threshold λ.

3 Simulation studies

Motivated by the Gilbert-Elliot model for burst errors in telecommunication, we may assume that there are two states: 0 (a "good" state) and 1 (a "bad" state), with the transition probability matrix

$$\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Conditional on the hidden state $X_n = x_n$ (= 0 or 1), the sensor observations $Y_{k,n}$ can be modeled as

$$Y_{k,n} = \mu_{x_n} + \sigma_{x_n}\epsilon_{k,n}, \qquad 1 \le k \le K,$$

where the $\epsilon_{k,n}$'s are i.i.d. $N(0,1)$. Denote by $\theta = (\alpha, \beta, \mu_0, \mu_1, \sigma_0, \sigma_1)$, and we are interested in detecting a change in θ from θ_0 to θ_1. Assume that the parameters values before a change occurs are $\alpha = 0.01, \beta = 0.1, \mu_0 = \mu_1 = 0, \sigma_0 = 1$ and $\sigma_1 = 5$, i.e., the pre-change parameter of θ is $\theta_0 = (0.01, 0.1, 0, 0, 1, 5)$. In our simulations, we consider the change

- *Change in Markov chain:* only (α, β) changes from $(0.01, 0.1)$ to $(0.5, 0.5)$, i.e., the post-change parameter $\theta_1 = (0.5, 0.5, 0, 0, 1, 5)$.

The remainder of this section is as follows. In Subsection 3.1, we summarize our simulations results on the Kullback-Leibler information in HMM for the above change. In Subsection 3.2, we report our simulation results in the context of quickest change detection. Subsection 3.3 then reports the distribution of run length to false alarms under this scenario.

3.1 Kullback-Leibler information number

Under the above scenario, we present the Kullback-Leibler information number $K(\theta_1, \theta_0)$ of the raw observations $(Y_{1,n}, \cdots, Y_{K,n})$'s and of the quantized observations $(U_{1,n}, \cdots, U_{K,n})$'s, where $U_{k,n} = \begin{cases} 1, \text{ if } Y_{k,n} \ge \lambda \\ 0, \text{ if } Y_{k,n} < \lambda \end{cases}$, for different choices of threshold value λ (we assume all sensors use the same threshold λ).

To derive some candidates of quantized threshold λ, besides some equal-distributed values in the interval of either $[0, 1]$ or $[1, 10]$, we will consider three special values of λ. The first two special values correspond to those λ's that maximize the K-L information numbers of the marginal distributions of $Y_{k,n}$'s with $K = 1$ and 2 sensors, respectively. The third special value λ is motivated from the fact the marginal distribution may be a very poor approximation when $\alpha + \beta \ne 1$ and when $f_{0,\theta}$ and $f_{1,\theta}$ are significantly different. Thus, to derive "better" value of λ, we may simply pretend that $\alpha = \beta = 0.5$ by ignoring the actual exact values of α and β, and then find λ that maximizes the K-L information numbers of the marginal distributions of $Y_{k,n}$'s when $\alpha = \beta = 0.5$. This gives us the third special values of λ.

Since $K(\theta_1, \theta_0)$ can be approximated by $\frac{1}{n}\log S_n$ for large value of n. In our calculation, we simulated $m = 10^3$ values of $\frac{1}{n}\log S_n$ with $n = 10^5$, and then we reported the mean of these $m = 10^3$ repetitions as the estimate of $K(\theta_1, \theta_0)$ (the median of these repetitions is the same as, or very close to, the mean in our examples).

Suppose we are interested in detecting a change in (α, β), the parameters in the transition probability matrix of Markov chain, from $(0.01, 0.1)$ to $(0.5, 0.5)$. Note that when the quantized threshold $\lambda = 0$, then the conditional densities of $U = I(Y > \lambda)$ given $X = 0$ or 1 are identical, and thus it is impossible to detect a change in (α, β) based on $U_{k,n}$'s. Also, by symmetric, it is sufficient to consider when the threshold $\lambda > 0$.

Using the K-L information number of the marginal distributions to approximate the K-L information for HMM, we have three special choices of λ under this scenario: $\lambda = 2.6375$ (one sensor), 2.7354 (two sensors), and 2.3284 (by pretending $\alpha = \beta = 0.5$).

Table 1 summarizes the estimated values of $K(\theta_1, \theta_0)$ for the raw observations $Y_{k,n}$'s and the quantized observations $U_{k,n}$'s with different choices of threshold value λ's. In Table 1, the values in parenthesis are the standard deviations of $m = 10^3$ repetitions. From Table 1, in the scenario when only

Table 1. Estimated Kullback-Leibler information number $K(\theta_1, \theta_0)$ when (α, β) changes from $(0.01, 0.1)$ to $(0.5, 0.5)$

λ	$K(\theta_1, \theta_0)$			
	$K=1$	$K=2$	$K=5$	$K=10$
Raw	0.2509 (0.0014)	0.5411 (0.0022)	1.3193 (0.0039)	1.6870 (0.0048)
1	0.0268 (0.0006)	0.0511 (0.0008)	0.1325 (0.0012)	0.2885 (0.0019)
2	0.0721 (0.0008)	0.1351 (0.0010)	0.3594 (0.0018)	0.7646 (0.0028)
2.3284	0.0786 (0.0008)	0.1449 (0.0010)	0.3819 (0.0017)	0.8192 (0.0028)
2.6375	0.0802 (0.0008)	0.1458 (0.0010)	0.3790 (0.0016)	0.8222 (0.0026)
2.7354	0.0800 (0.0008)	0.1448 (0.0010)	0.3746 (0.0016)	0.8153 (0.0026)
3	0.0781 (0.0008)	0.1401 (0.0010)	0.3573 (0.0015)	0.7836 (0.0024)
4	0.0641 (0.0008)	0.1120 (0.0009)	0.2683 (0.0014)	0.5909 (0.0020)
5	0.0501 (0.0008)	0.0860 (0.0008)	0.1931 (0.0011)	0.4076 (0.0017)

(α, β) changes to $(0.5, 0.5)$, among those choices of λ's, the Kullback-Leibler information number of the quantized observation $U_{k,n} = I(Y_{k,n} \geq \lambda)$ is maximized near $\lambda = 2.6375$, the value that maximizes the K-L information number of the marginal distribution of $U_{k,n}$ with a single-sensor.

3.2 Quickest change detection

To further confirm the above calculations of the Kullback-Leibler information numbers in HMM, in the above scenario, we now run simulations in the context of the quickest change detection problems for the single-sensor system, i.e., $K = 1$ for simplicity. Specifically, let W_n be the CUSUM statistics (i.e., the generalized log-likelihood statistics for HMM) for either raw or quantized observations, and the fusion center uses the CUSUM procedure

$$N(a) = \inf\{n : W_n \geq a\}.$$

We will compare the CUSUM procedure with raw data with three CUSUM procedures with quantized threshold λ's (the specific thresholds depend on the specific kinds of changes).

For these CUSUM procedures, the threshold value was first determined from the criterion on the average run length to false alarm: $\mathbf{E}_\infty(N(a)) \approx \gamma$. A 10^3-repetition Monte Carlo simulation was performed to determine the appropriate values of a to yield the desired average run length to false alarm γ to within the range of sampling error. To speed up the simulation, one efficient algorithm is to run *one* simulation to return the record values of the CUSUM statistics and the corresponding values of run lengths, and then to estimate $\mathbf{E}_\infty(N(a))$ for different a based on these record values.

Next, we simulate the detection delay $\mathbf{E}_1(N(a))$ based on Monte Carlo experiments with 10^3 repetitions, and then report corresponding results of detection delays.

Table 2 summarizes the simulated detection delays of these CUSUM procedures under the above-mentioned change, based on 10^3 repetitions, subject to the constraint $\mathbf{E}_\infty(T(a)) \approx \gamma$, with the values of a in parentheses.

From Table 1, among all quantized observations, the CUSUM procedure with quantized threshold $\lambda = 2.6375$ has the smallest detection delay, which is consistent with the fact that $\lambda = 2.6375$ leads to the largest Kullback-Leibler information number.

3.3 Run lengths to false alarms

Fig. 1 illustrates the histogram and exponential QQ-plots of run lengths to false alarm in the above example. From this plot, the run lengths to false alarm are approximately exponentially distributed.

Table 2. Detection Delays of CUSUM procedures T with $\mathbf{E}_\infty(T) \approx \gamma$ when (α, β) changes from $(0.01, 0.1)$ to $(0.5, 0.5)$

γ	Raw	$\lambda = 1$	$\lambda = 2.6375$	$\lambda = 5$
200	(2.91) 9.6 ± 0.2	(1.88) 45.1 ± 0.8	(2.02) 17.5 ± 0.3	(2.04) 24.8 ± 0.5
400	(3.49) 12.5 ± 0.2	(2.40) 64.1 ± 1.1	(2.90) 24.8 ± 0.4	(2.13) 32.1 ± 0.6
600	(3.90) 14.1 ± 0.2	(2.73) 76.2 ± 1.3	(3.23) 30.2 ± 0.5	(2.71) 38.8 ± 0.7
800	(4.23) 15.4 ± 0.2	(2.94) 84.4 ± 1.3	(3.40) 33.3 ± 0.5	(3.04) 43.6 ± 0.8
1000	(4.43) 16.3 ± 0.2	(3.14) 92.1 ± 1.4	(3.62) 35.7 ± 0.6	(3.24) 47.3 ± 0.8
Theoretical Large γ	$\frac{\log \gamma}{0.2509}$	$\frac{\log \gamma}{0.0268}$	$\frac{\log \gamma}{0.0802}$	$\frac{\log \gamma}{0.0501}$

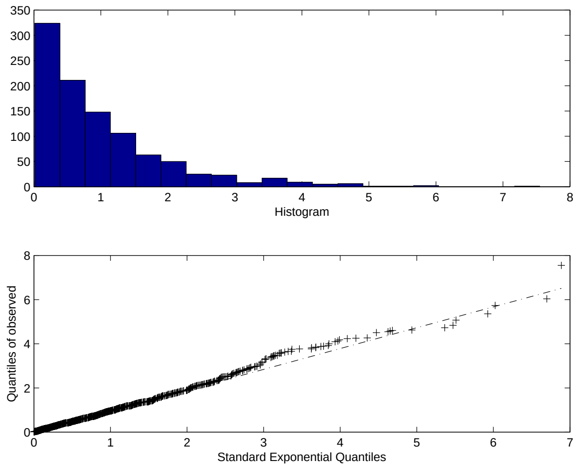

Fig. 1. Example A (change in Markov chain): Histogram and exponential QQ-Plot of Run Lengths to False Alarms with $\mathbf{E}_\infty(N) \approx 1000$.

4 Conclusions

In this article, we examined the decentralized quickest change detection problem in sensor networks. In the detection scheme we considered, each sensor simply compared its current sensor observation to a stationary threshold, and the fusion center used a CUSUM-type scheme based on binary vectors. We also introduced the global and local maximization methods for choosing the "optimal" stationary local thresholds at the sensor level. While the classical quickest change detection problems have been studied for several decades, the decentralized quickest change detection, particularly in HMM models, is still in its infancy. Clearly, there is a lot of research ahead to improve the theory, algorithms and methodology. Hopefully this article can stimulate further research.

References

Blum, R. S., Kassam, S. A. and Poor, H. V. (1997). Distributed Detection with Multiple Sensors: Part II- Advanced Topics. *Proceedings of the IEEE*, **85**, 64–79.

Dogandzic, A. and Zhang, B. (2006). Distributed Estimation and Detection for Sensor Networks Using Hidden Markov Random Field Models. *IEEE Trans. Signal Processing*, **54**, 3200–3215.

Fuh, C-D. (2003). SPRT and CUSUM in Hidden Markov Models. *Ann. Statist.*, **31**, 942–977.

Fuh, C-D and Mei, Y. (2008). Optimal Stationary Binary Quantizer for Decentralized Quickest Change Detection in Hidden Markov Models. Best Paper, *Information Fusion 2008.* http://www.fusion2008.org/bpa.html

Huang, M. and Dey, S. (2006). Dynamic Quantizer Design for Hidden Markov State Estimation via Multiple Sensors with Fusion Center Feedback. *IEEE Trans. Signal Processing*, **54**, 2887–2896.

Viswanathan, R. and Varshney, P. K. (1997). Distributed Detection with Multiple Sensors — Part I: Fundamentals. *Proceedings of the IEEE*, **85**, 54–63.