Evidence Theory Enabled Quickest Change Detection Using Big Time-Series Data from Internet of Things

Hossein Jafari, Xiangfang Li, Lijun Qian, Alexander Aved, Timothy Kroecker

Abstract—Traditionally in sensor networks and recently in the Internet of Things, numerous heterogeneous sensors are deployed in distributed manner to monitor a phenomenon that often can be model by an underlying stochastic process. The big time-series data collected by the sensors must be analyzed to detect change in the stochastic process as quickly as possible with tolerable false alarm rate. However, sensors may have different accuracy and sensitivity range, and they decay along time. As a result, the big time-series data collected by the sensors will contain uncertainties and sometimes they are conflicting. In this study, we present a framework to take advantage of Evidence Theory (a.k.a. Dempster-Shafer and Dezert-Smarandache Theories) capabilities of representing and managing uncertainty and conflict to fast change detection and effectively deal with complementary hypotheses. Specifically, Kullback-Leibler divergence is used as the similarity metric to calculate the distances between the estimated current distribution with the pre- and post-change distributions. Then mass functions are calculated and related combination rules are applied to combine the mass values among all sensors. Furthermore, we applied the method to estimate the minimum number of sensors needed to combine, so computational efficiency could be improved. Cumulative sum test is then applied on the ratio of pignistic probability to detect and declare the change for decision making purpose. Simulation results using both synthetic data and real data from experimental setup demonstrate the effectiveness of the presented schemes.

Keywords—CUSUM, evidence theory, KL divergence, quickest change detection, time series data.

I. INTRODUCTION

THE Internet of Things (IoT) is the internet-working of many physical devices such as smart sensors and vehicles, and providing them the capability of sending and receiving data. As a result, big time-series data are collected and exchanged. The collected data may be analyzed to detect changes and anomalies in the physical world that can be modeled as stochastic processes mathematically.

Change point detection is a well-developed method to detect a change along a stochastic process [1]. Let us assume a sequence of observation from one sensor with known distribution. At unknown time an event happens and distribution of observation changes. The goal of change point detection is to detect and announce the true change with

H. Jafari, X. Li and L. Qian are with CREDIT Research Center, Department of Electrical and Computer Engineering, Prairie View A&M University, Texas A&M University System, Prairie View, TX 77446, USA (e-mail: hjafari@student.pvamu.edu, xili@student.pvamu.edu, liqian@pvamu.edu, see http://credit.pvamu.edu/).

A. Aved and T. Kroecker are with Information Directorate, US Air Force Research Laboratory (AFRL), Rome, NY 13441, USA (e-mail: alexander.aved@us.af.mil, timothy.kroecker@us.af.mil).

minimum delay subject to constraint that the rate of false alarm be small. This is also called Quickest Detection (QD) [2]. The early works of QD are done in quality control to detect a failure in manufacturing systems [3], [4]. So after detection of change, production line can stop to fix the problem and prevent the loss.

There are many QD applications including intrusion detection in computer networks and security systems [5], [6], tracking, surveillance, health care (patient falling as an example), spectrum sensing in cognitive radio [7] and so on. These applications benefit from data fusion of multi sensors [8], where observations from several sensors aggregated and cooperated in different methods to detect an event.

In reality, sensors are not perfect. In the sensor field, where the sensors installed to monitor an event, sensors provided by several manufacturers have different specifications, accuracy and sensitivity range. And their functionalities decay along the years. So we may not be certain about the value measured by those sensors. Finally besides all sensors related uncertainty metrics, environmental interference and noise can affect their accuracy and measurement.

Based on [3] for sequence of observation from single sensor, cumulative sum test (CUSUM) can detect change in distribution with minimum delay while keeping false alarm rate low. Similarly for the multisensor applications that stream of data generated by sensors, several works are done based on QD to detect the change [9]–[11].

In this paper, we focus on stream of independent observations from N sensors aggregated in data fusion center as shown in Fig. 1. Maximum Likelihood Estimation (MLE) method is applied to estimate the distribution parameters of each sensor independently based on their most recent W observations (W called the sliding window size). Then Kullback-Leibler (KL) divergence method is used to find the distance between estimated distribution with pre- and post-change distributions. Following, related mass function values are calculated based on those distance values for each sensor and then to manage uncertainty, Dempster-Shafer combination rule (DST) is applied to combine the mass values among all N sensors [12]. In the case of high conflict value, Dezert-Smarandache combination rule (DSmT) is applied for more accuracy [13]. However, to improve the computational efficiency, we applied the method to estimate the minimum number of sensors needed to combine. In QD framework, instead of likelihood ratio in computing CUSUM test, we used ratio of related pignistic probabilities. These CUSUM values

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
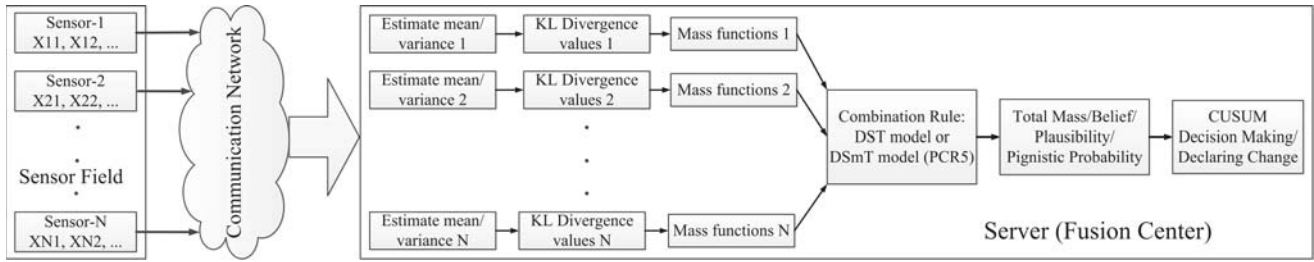Vol:11, No:6, 2017

Fig. 1 Block diagram of presented multisensor change detection using DST/DSmT combination rules

are compared to the predefined threshold value and the result is used as a metric to detect and announce the change point. We show that the presented method outperforms the traditional QD methods when uncertainties are high in practical situations.

We start at Section II with presenting our model, including all the assumptions. In Section III, we give background information about KL divergence, Dempster-Shafer theory of evidence and Dezert-Smarandache theory (DSmT) of plausible and paradoxical reasoning. Then, we explain their related combination rules, belief, plausibility and pignistic probability. Following, we present the modified version of CUSUM test and estimation of minimum number of sensors needed to combine. Section IV shows simulation results for different models and compare their performance. For real data analysis section, Section V explains spectrum sensing application. At the end, Section VI concludes the paper and explains some of our ongoing researches to extend this work.

II. Model

In this work, it is assumed that there are N sensors monitoring an underlying stochastic process and they generate streams of data in parallel. It is also assumed that the observations from each sensor are independent and identically distributed (i.i.d.) with Gaussian distribution. Furthermore, we assume that the observations between different sensors are mutually independent as well.

In this paper, we consider two cases with "ideal sensors" and "practical sensors", respectively. For the first case with "ideal sensors", we assume that sensor readings are accurate and all sensors have large sensitivity range and they are affected by the change [9]. As in most of the works of change detection in the literature [9]–[11], the distributions before change (p_0) and after change (p_1) are assumed to be known. Without loss of generality, p_0 is Gaussian distribution with zero mean and unit variance, $p_0 \sim N(0, 1)$, and p_1 has the same variance but different mean, $p_1 \sim N(1, 1)$. This case will serve as the benchmark to gauge the performance of the presented method for the "practical sensors" case. Although here we assumed that the variance does not change, but as we will explain in Section III, KL divergence considers change in both mean and variance as it is the case in practical applications.

For the "practical sensors", we assume that the sensors are heterogeneous and their readings are inaccurate. Because the empirical data are available and the sensor readings can be calibrated, it is reasonable to assume that the distribution before change (p_0) is known, $p_0 \sim N(0, 1)$. However, due
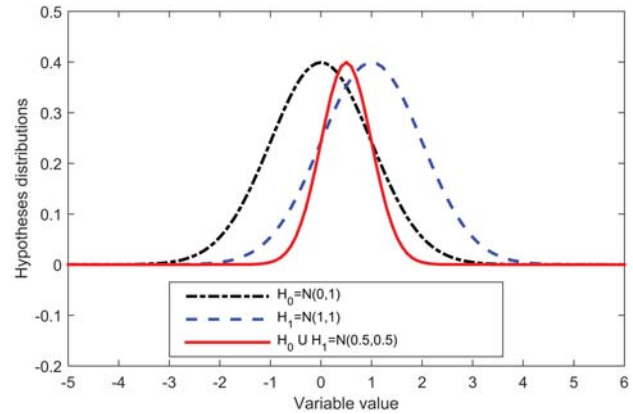

Fig. 2 Hypotheses distributions including null, alternative and uncertainty

to the heterogeneity of the sensors and the uncertainties in sensors' measurements, it is assumed that each sensor n after change may have different mean $p_1 \sim N(\mu_n, 1)$ [11] (and generally different variance), although all the sensors are within their sensitivity range and the mean value μ_n is bounded.

The null hypothesis (H_0) is defined as no change in the distribution of the underlying random process, while the alternative hypothesis H_1 denotes that change happens. The corresponding distributions are $p_0(x) \sim N(\mu_0, \sigma_0^2)$ and $p_1(x) \sim N(\mu_1, \sigma_1^2)$, respectively. In order to facilitate the uncertainties due to "practical sensors", we define $p_{1/2}(x) \sim N(\mu_{1/2}, \sigma_{1/2}^2)$ where $\mu_{1/2} = (\mu_0 + \mu_1)/2$ and $\sigma_{1/2}^2 = (\sigma_0^2 + \sigma_1^2)/4$. This new proposed hypothesis is a union between the two hypotheses H_0 and H_1 ($H_0 \cup H_1$). According to Fig. 2, $p_{1/2}(x)$ maximizes when $p_0(x)$ and $p_1(x)$ are equal and it diminishes when one of $p_0(x)$ or $p_1(x)$ dominates the other one.

III. Proposed Algorithm and Analytical Study

In this section, the detailed procedures of the framework are illustrated and the corresponding analysis are given.

A. Proposed Algorithm

Fig. 1 shows the data flow from sensor field to data fusion center. In the sensor field, N sensors monitor an underlying random process independently and send their observations/readings to the fusion center. In the fusion center, firstly MLE method is applied to estimate mean and variance values. Then distance values are calculated based on

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
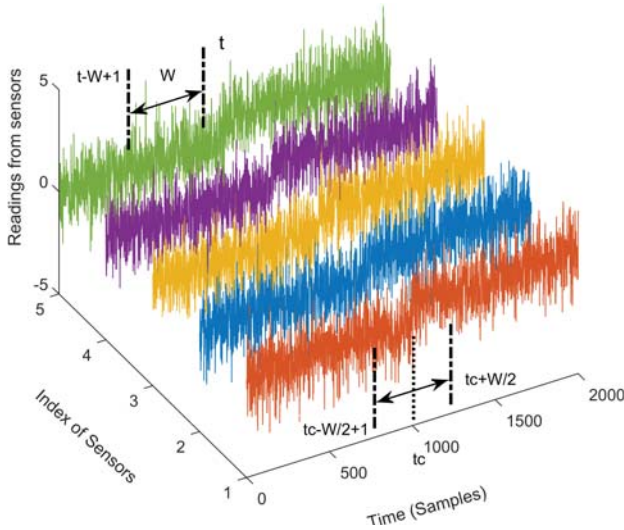Vol:11, No:6, 2017

Fig. 3 Sample readings from five sensors and sliding window

KL divergence. These distance values are used to compute mass functions. Based on the level of conflict, one of the Dempster-Shafer combination rule or PCR5 (Proportional Conflict Redistribution Rule 5, one of the combination rules in DSmT framework) will be chosen to combine the related masses among all N sensors [13]. Following, pignistic probabilities are calculated and finally CUSMU test is applied on the ratio of pignistic probabilities to detect and announce the change.

B. KL Divergence

In information theory domain, relative entropy or Kullback-Leibler divergence (KL divergence) is defined as the distance between two distributions p and q:

$$D(p||q) = \int p(x) log \frac{p(x)}{q(x)} dx \qquad (1)$$

Here $p(x)$ and $q(x)$ are probability density functions (pdf). The value of KL divergence is non-negative. It will equal to zero when the two distributions are the same. As an example, the KL divergence between two normal distributions with pdf of $p(x) \sim N(\mu_p, \sigma_p^2)$ and $q(x) \sim N(\mu_q, \sigma_q^2)$ is:

$$D(p||q) = \frac{1}{2} \left[log \frac{\sigma_q^2}{\sigma_p^2} - 1 + \frac{\sigma_p^2 + (\mu_p - \mu_q)^2}{\sigma_q^2} \right] \qquad (2)$$

Although KL divergence is asymmetric function ($D(p||q) \neq D(q||p)$), we use symmetric version defined by $D_{KL}(p||q)$:

$$D_{KL}(p||q) = D(p||q) + D(q||p)$$
$$= \frac{1}{2} \left[\frac{\sigma_p^4 + \sigma_q^4 + (\mu_p - \mu_q)^2.(\sigma_p^2 + \sigma_q^2)}{\sigma_p^2.\sigma_q^2} \right] - 1 \qquad (3)$$

Sample readings from five sensors out of N sensors and the sliding window are shown in Fig. 3. At each time t, for each sensor i, MLE method is used to estimate the parameters of the observation distribution based on the most recent W data samples between $t - W + 1$ and t, where W is the

sliding window size. W plays a major role and must be chosen carefully to address the tradeoff between estimation accuracy and detection delay. The estimated mean and variance based on W data samples y can be shown as:

$$\hat{\mu}_q = \frac{1}{W} \cdot \sum_{w=1}^{W} y_w \qquad (4)$$

$$\hat{\sigma}_q^2 = \frac{1}{W-1} \cdot \sum_{w=1}^{W} (y_w - \hat{\mu}_q)^2 \qquad (5)$$

C. Combination of Data

As stated in Section II, we consider two hypotheses H_0, H_1, and their union $H_0 \cup H_1$. The union hypothesis, $H_0 \cup H_1$, is called ignorance and it represents uncertainty. In order to handle uncertainties, we used combination rules among sensors belief values. Basic belief assignment (bba) or mass functions will be calculated based on the symmetric KL divergence values.

Let us define $d_i^{l_j}$ as a distance between sensor i and class l_j where $j = 0, 1/2, 1$ and $l_{1/2} = l_0 \cup l_1$. Here class l_j represents the hypothesis H_j. Then for sensor i, the set of distance values is:

$$D_i = \{d_i^{l_0}, d_i^{l_1}, d_i^{l_0 \cup l_1}\} \qquad (6)$$

Small value of distance $d_i^{l_j}$ indicates the probability that the class of sensor i is l_j is higher. Then mass functions can be calculated based on distance values:

$$m_i(l_j) = \frac{1/d_i^{l_j}}{\sum_{j=1}^{m} (1/d_i^{l_j})} \qquad (7)$$

The mass functions for each sensor node i related to hypothesis H_0, H_1 and their union $H_0 \cup H_1$ can be shown as:

$$m_i = \{m_i(l_0), m_i(l_1), m_i(l_0 \cup l_1)\} \qquad (8)$$

In data fusion applications, different data aggregated from multiple sensors are considered as different evidences. Thus, data processing requires combination among the all evidences based on some combination rules. The hypotheses H_0 and H_1 are mutually exclusive and exhaustive, so Dempster-Shafer model can be applied. In D-S framework set of singleton classes called "frame of discernment", $\Theta = \{l_0, l_1\}$. And power set, 2^Θ, define as the set of all subset of frame of discernment including the empty set, $2^\Theta = \{\emptyset, l_0, l_1, l_0 \cup l_1\}$. Combined mass function can be calculated based on the Dempster combination rule [14]:

$$m(l_j) = m_1 \oplus m_2 \oplus ... \oplus m_N \qquad (9)$$

$$m(l_j) = \begin{cases} 0 & l_j = \emptyset \\ \dfrac{\sum\limits_{\cap_{k=1}^{N} l_k = l_j} m_1(l_1).m_2(l_2)...m_N(l_N)}{1 - K} & l_j \neq \emptyset \end{cases} \qquad (10)$$

$$K = \sum_{\cap_{k=1}^{N} l_k = \emptyset} m_1(l_1).m_2(l_2)...m_N(l_N) \qquad (11)$$

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:11, No:6, 2017

Here K is called conflict among all the sources of data. It is used as a normalization factor, $K \in [0,1]$. The higher value of K indicates more conflicting among data sources. In the case of high conflict, we apply Dezert-Smarandache (DSmT) rules instead of Dempster-Shafer combination rule. For instance, PCR5 [13] is one of the most accurate combination rules in the DSmT framework that we used in our simulations. For two sensors, PCR5 has the following form:

$$m_{PCR5}(l_j) = m_{12}(l_j) \tag{12}$$
$$+ \sum_{\substack{l_k \in D^\Theta \backslash \{l_j\} \\ , l_j \bigcap l_k = \emptyset}} \left[\frac{m_1(l_j)^2 . m_2(l_k)}{m_1(l_j) + m_2(l_k)} + \frac{m_2(l_j)^2 . m_1(l_k)}{m_2(l_j) + m_1(l_k)} \right]$$

Here, m_{12} refers to conjunctive consensus:

$$m_{12}(l_j) = \sum_{l_m, l_n \in D^\Theta, l_m \bigcap l_n = l_j} m_1(l_m) . m_2(l_n) \tag{13}$$

Hyper power set, D^Θ, is defined as the set include power set and intersection among classes in the frame of discernment, $D^\Theta = \{\emptyset, l_0, l_1, l_0 \cup l_1, l_0 \cap l_1\}$. In DSmT framework, exclusivity between classes are not essential.

For both DST and DSmT frameworks, the associated belief function is:

$$Bel(l_j) = \sum_{l_k \in 2^\Theta / D^\Theta, l_k \subseteq l_j} m(l_k) \tag{14}$$

The plausibility function can be calculated:

$$Pl(l_j) = \sum_{l_k \in 2^\Theta / D^\Theta, l_k \bigcap l_j \neq \emptyset} m(l_k) = 1 - Bel(\bar{l}_j) \tag{15}$$

where $\bar{x}$ is the complement set of x, $\bar{x} = \Theta - x$. It is clear that $Bel(l_j) \leq Pl(l_j)$. The belief interval, $[Bel(l_j), Pl(l_j)]$, refers to the imprecision on the true probability, where belief function is considered as a lower probability and plausibility function as an upper probability.

Finally, the pignistic probability [15] can be obtained:

$$betP(l_j) = \sum_{l_k \in 2^\Theta / D^\Theta, l_k \neq \emptyset} \frac{|l_j \cap l_k|}{|l_k|} m(l_k) \tag{16}$$

where $|x|$ is the cardinality of x.

D. Probabilistic Information Content (PIC)

Shannon entropy of a probability $P(.)$ over a discrete finite set $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$ is defined by:

$$H(P) = - \sum_{i=1}^{n} P(\theta_i) . log_2(P(\theta_i)) \tag{17}$$

It measures the uncertainty and randomness contained by $P(.)$ and its normalized version is shown in Fig. 4. For the uniform probability distribution with $P(\theta_i) = 1/n$ as a worst case, $H(P) = H_{max} = log_2(n)$. In the case of totally deterministic probability, such that $P(\theta_i) = 1$ and $P(\theta_j) = 0$ for $j \neq i$, $H(P)$ will be zero.

Probabilistic information content (PIC) is the criterion that can be used as a metric for depicting the strength of a critical decision by a specific probability distribution [16]. It is an
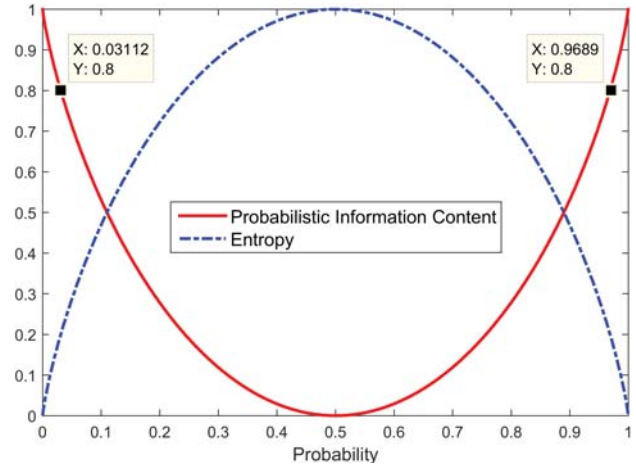

Fig. 4 Entropy and Probabilistic information content

essential measure in any threshold-driven automated decision making system. The PIC is the dual of the normalized Shannon entropy and is defined by [16]:

$$PIC(P) = 1 + \frac{1}{H_{max}} . \sum_{i=1}^{n} P(\theta_i) . log_2(P(\theta_i)) \tag{18}$$

$$PIC(P) = 1 - \frac{H(P)}{H_{max}} \tag{19}$$

According to PIC plot in Fig. 4 while the maximum PIC value of one (zero entropy) indicates the total knowledge to make a correct decision, the minimum PIC value of zero (maximum entropy of one) denotes that the knowledge to make a correct decision does not exist.

We applied the PIC as a metric to find the minimum number of combination needed at each time. Here, the PIC value can be interpreted as a level of reliability or confidence in decision making. Advantage of applying DST-DSmT combination rules in managing uncertainty, our simulations and experiment results show that regardless of the number of sensors, after several combination, the result converges. So it is not necessary to apply DS-DSmT combination rules among all sensors at each time. Instead, minimum number of sensors needed to combine at each time can be estimated prior to combination start. Moreover, combination process could be continuing recursively and it stops whenever the PIC value reaches to predefined threshold value. In this part to simplify our notations instead of pignistic probabilities $betP(l_0)$ and $betP(l_1)$, we will use P_0 and P_1, respectively. So to guarantee the result of decision with $PIC(P_0, P_1) = \alpha$, following conditions need to be satisfied: For before change $P_0 > PIC^{-1}(\alpha)$ or $P_1 < 1 - PIC^{-1}(\alpha)$. Similarly for after change, it is necessary that $P_1 > PIC^{-1}(\alpha)$ or $P_0 < 1 - PIC^{-1}(\alpha)$. Furthermore, for the predefined $PIC = \alpha$, the minimum number of sensors needed to combine for after change can be estimated by:

$$n^* = \arg \min_n \{P_1 | P_1 > PIC^{-1}(\alpha)\} \tag{20}$$

Here $2 \leq n^* \leq N$. Similarly for before change:

$$n^* = \arg \min_n \{P_0 | P_0 > PIC^{-1}(\alpha)\} \tag{21}$$

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:11, No:6, 2017

With substituting distance values based on KL-divergence in pignistic probabilities equation, pignistic probabilities can be calculated directly based on distance values without calculating mass functions:

$$P_1 = \frac{A}{A+B} = \frac{1}{1+B/A} \tag{22}$$

and

$$P_0 = 1 - P_1 = \frac{B}{A+B} = \frac{1}{1+A/B} \tag{23}$$

where:

$$A = \prod_{i=1}^{n} d_{i0} \times \left[\sum \left(\prod_{\substack{i=1 \\ ,j=\{1,x\}}}^{n} d_{ij} \right) + 0.5 \times \prod_{i=1}^{n} d_{i1} \right] \tag{24}$$

$$B = \prod_{i=1}^{n} d_{i1} \times \left[\sum \left(\prod_{\substack{i=1 \\ ,j=\{0,x\}}}^{n} d_{ij} \right) + 0.5 \times \prod_{i=1}^{n} d_{i0} \right] \tag{25}$$

Here, d_{ij} is the KL distance related to sensor i from distributions 0, 1, and x is related to H_0, H_1, and $H_0 \cup H_1$, respectively. For the simplest case without uncertainty, they simplify to:

$$P_1^+ = \frac{\prod_{i=1}^{n} d_{i0}}{\prod_{i=1}^{n} d_{i0} + \prod_{i=1}^{n} d_{i1}} = \frac{1}{1 + \prod_{i=1}^{n} d_{i1} / \prod_{i=1}^{n} d_{i0}} \tag{26}$$

The lower bound for number of sensors needed to combine subject to $PIC = \alpha$ can be calculated based on P_1^+ for after change region by:

$$n^+ = \arg\min_{n} \{ P_1^+ | P_1^+ > PIC^{-1}(\alpha) \} \tag{27}$$

The main goal in change detection is declaring a change as quickly as possible. So it is not necessary to apply DST combination rule among all N sensors at each time t. Simulation results show that it can decrease computation time tremendously.

E. Quickest Detection

In the formulation of quickest change detection (QD) problem, a sequence of observations are generated by a sensor based on underlying process with some distribution. At some unknown time distribution of observations change and the goal is to detect change as quickly as possible, subject to false alarm constraint. If the change happens at unknown time $t = k$, before change observations of each sensor S_i, $y_i(1), y_i(2), ..., y_i(k-1)$, we will follow pre-change distribution H_0. While after change observations of $y_i(k), y_i(K+1), ...$ follow post-change distribution H_1.

QD technique detects and declares the change with minimum delay at time $t = \tau$ ($\tau \geq k$), and tries to keep false alarm rate low. So false alarm rate increases whenever the change is announced before the real changes, $\tau < k$.
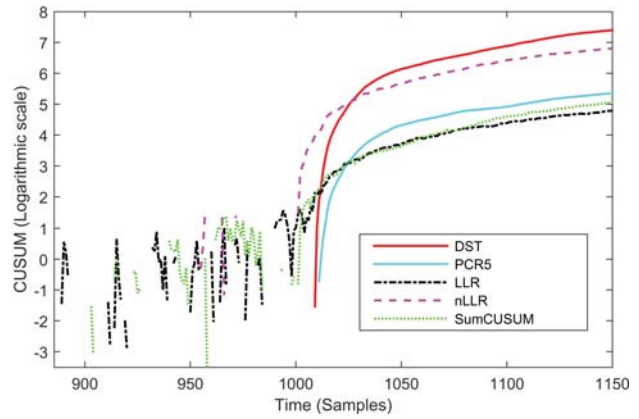

Fig. 5 CUSUM results for test data when distributions are known
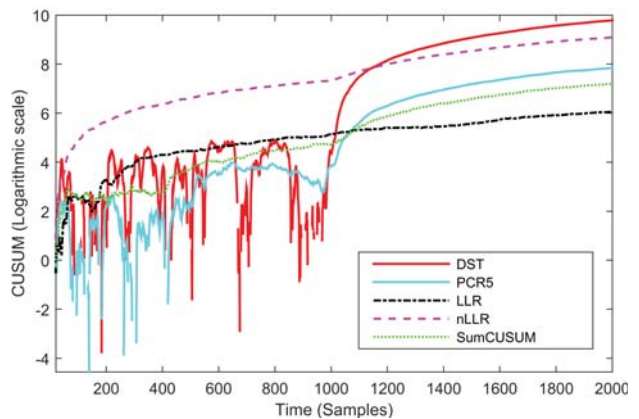

Fig. 6 CUSUM results for test data when distributions are unknown

Cumulative Sum (CUSUM), as an optimal method, could be applied to solve QD problem [17]. For one sensor, CUSUM is calculated recursively by:

$$CUSUM_n(t) = max\{0, CUSUM_n(t-1) + LLR_n(t)\} \tag{28}$$

At each time t, the updated value of CUSUM is compared to predefined probability of false alarm (PFA) as threshold value. It declares the change based on the stopping time rule: $t^* = min\{t \geq 1 | CUSUM_n(t) \geq PFA = \alpha\}$ where $CUSUM_n(0) = 0$ and the log likelihood ratio (LLR) between after change and before change hypotheses for tth observation of sensor n is defined by:

$$LLR_n(t) = log\left(\frac{H_1}{H_0}\right) = log\left(\frac{f^1(x_n(t))}{f^0(x_n(t))}\right) \tag{29}$$

In distributed sensor network, a sequence of observations is generated by multiple sensors and aggregated in fusion center. For multiple sensor, CUSUM could be calculated in different ways. In [9], CUSUM is applied on the sum of the likelihood ratio of individual sensors to determine a change:

$$CUSUM(t) = max\left\{0, CUSUM(t-1) + \sum_{n=1}^{N} LLR_n(t)\right\} \tag{30}$$

In [10], CUSUMs from individual sensors are added together. As a mixture method, [11] used specific generalized

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
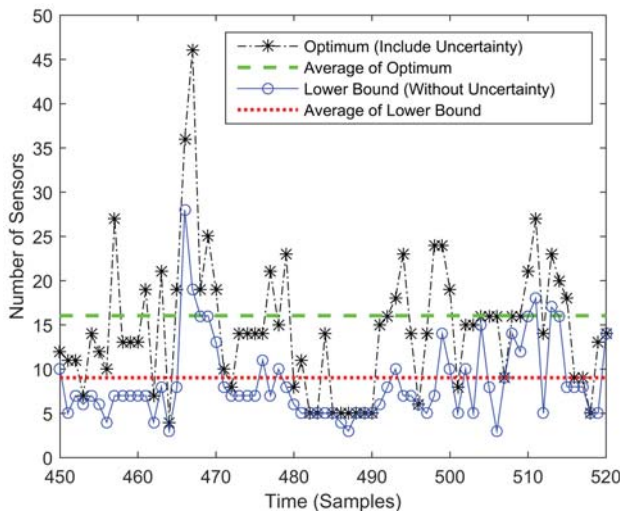Vol:11, No:6, 2017

Fig. 7 Minimum number of sensors needed to combine

TABLE I
SIZE OF PSD-IQ FILES BASED ON NO. OF USRP AND TIME

No. of USRP	Duration	IQ file size	PSD file size	Total size
1	10 min	5GB	2.5GB	7.5GB
1	1 hour	30GB	15GB	45GB
1	1 day	720GB	360GB	1.08TB
8	1 day	5.76TB	2.88TB	8.64TB
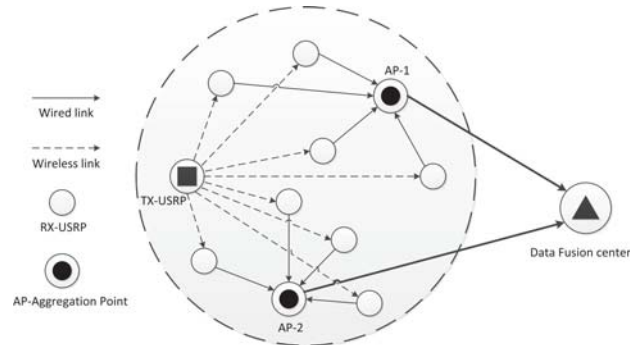
Fig. 8 Cooperative spectrum sensing setup

likelihood ratio to detect a change. All these methods assume Bayesian probability, which does not consider uncertainty or conflict among the observations. We introduce the ratio of pignistic probability to replace with LLR in CUSUM test.

IV. SIMULATION STUDY

We assumed $N = 100$ sensors with total observation of 2000 samples per sensor. Observation of sensors before change follows Gaussian distribution $p_0(x) \sim N(0,1)$ and after change $p_1(x) \sim N(1,1)$ for the ideal-sensors. For practical-sensor case, it is assumed that each sensor n after change may have different mean $p_1 \sim N(\mu_n, 1)$ and μ_n uniformly distributed in $[0.5, 1.5]$.

Here we explain 5 methods and compare their simulation results in Figs. 5-9. In "DST" method, DST combination rule is applied among all sensors and then CUSUM is applied on the ratio of final pignistic probabilities values. In "PCR5" method, PCR5 combination rule is applied instead of DST combination rule. "LLR" method refers to the CUSUM result of one sample sensor on its LLR [17]. In "nLLR", first LLR of all N sensors is added and finally CUSUM is applied on it [9]. And "SumCUSUM" adds CUSUM results from individual N sensors [10]. According to the simulation setting as shown in Fig. 3 for five sample sensors, the real change point happens at time $t_c = 1000$. CUSUM test results based on different methods are shown in Fig. 5 for ideal-sensor case. Although CUSUM is positive, because Fig. 5 is plotted in logarithmic scale so the negative value in figure are related to CUSUM in $(0,1)$. It is clear that "nLLR" method as optimal solution has the best performance in detecting change when the both pre- and post-change distributions are known. Our method with highest slope detects the change after fixed $W/2$ time. Although other methods experience false alarm before real change, our methods based on DST and PCR5 with managing uncertainty guarantee smaller false alarm near zero. Similarly, Fig. 6 shows the simulation result for practical-sensor case, where the post-change distributions are unknown. According

to Fig. 6 our methods, specifically PCR5, outperform others with guaranteeing smaller false alarm.

Simulation results show that regardless of the number of sensors, after several combination, the result converges. So it is not necessary to apply DST-DSmT combination rules among all sensors at each time. Instead, minimum number of sensors needed to combine at each time can be estimated prior to combination start. Moreover, combination process could be continuing recursively and it stops whenever the PIC value reaches to predefined threshold value. For the practical-sensor case, Fig. 7 shows the simulation results to estimate the minimum number of sensors needed to combine subject to PIC=0.9. That means, the minimum sensors are needed to combine so with 90 percent confidence, we make a decision between H_0 and H_1. It includes the lower bound marked with blue circle-line that considers the case without uncertainty. The black asterisk with dash-dot line is related the case with uncertainty. For both cases, the averages are shown with dashed and dotted line, respectively.

V. EXPERIMENTAL DATA SET

Reliable detection of vacant spectrum of licensed bands is critical and fundamental in cognitive radio networks (CRNs) [18], [19]. Secondary users can use spectrum during idle periods to improve the spectrum efficiency. A single spectrum sensing node suffers from multi-path fading, shadowing and hidden node problem. Thus cooperative spectrum sensing is adopted to improve the probability of detection and decrease the false alarm rate by using distributed sensing nodes. In cooperative spectrum sensing, based on different channel condition for each node they receive transmitted signal from the licensed primary user with different signal-to-noise ratio (SNR) [20].

For real data analysis we applied our method to RF spectrum sensing data. We used 8 sets of Universal Software Radio Peripheral (USRP-model 2932) devices from National Instrument.USRP is a Software Defined Radio (SDR). Fig. 8

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
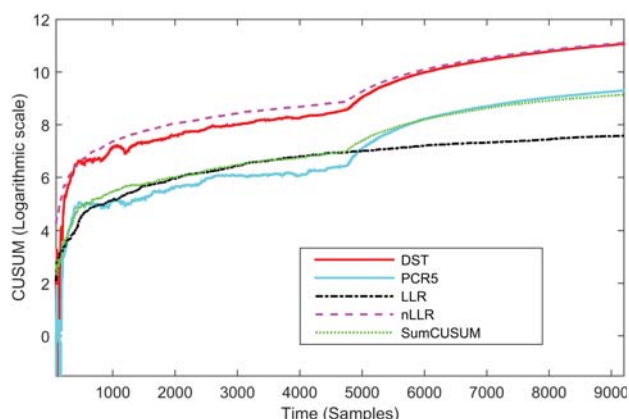Vol:11, No:6, 2017

Fig. 9 CUSUM results for USRP data

shows the setup for cooperative spectrum sensing using USRP devices. Each set of 4 USRP is configured in receiver mode (RX-USRP) and connected to aggregation point that runs the receiver program written in NI-LabVIEW. The 9th USRP is configured as a transmitter (TX-USRP) to send RF signal in specific band and at specific times. Received power spectral density (PSD) and IQ data from 8 USRPs are collected for processing. Table I gives details about the generated data sizes. Because the channel conditions are different for each RX-USRPs due to their locations to the TX-USRP, each receiver collected different data. Fig. 9 shows the CUSUM results in logarithmic scale. PCR5 method outperforms other methods while guaranteeing smaller false alarm rate. The practicality of the presented method to handle real-world big time-series data with uncertainty is also demonstrated.

VI. CONCLUSION

The ever increasing big time-series data demand new data processing methods, especially when the data contain high level of uncertainty. In this study, we presented a detection framework to mitigate the effect of uncertainty using Evidence Theory and KL divergence for distance measures. An algorithm based on sliding window is designed to detect change in the underlying stochastic process and the presented method can pinpoint the real change time after detecting it. Furthermore, to improve the computational complexity we presented a method to estimate and use the minimum number of combination needed. Simulations using synthetic data and experiments using real data are carried out and they demonstrated the practicality of the presented method to handle real-world big time-series data with uncertainty.

ACKNOWLEDGMENT

REFERENCES

[1] Basseville, M. and Nikiforov, I. V. *Detection of Abrupt Change Theory and Application,* Englewood Cliffs, NJ: Prentice Hall, 1993.
[2] Poor, H. V. and Hadjiliadis, O. *Quickest Detection,* New York: Cambridge University Press, 2008.
[3] Page, E. S. *Continuous inspection schemes,* Biometrika, 1954.
[4] Shiryaev, A. N. *On Optimum Methods in Quickest Detection Problems,* Theory of Probability and Its Applications 8: 22–46, 1963.
[5] A. G. Tartakovsky, B. Rozovskii, R. Blazek, and H. Kim, *A Novel Approach to Detection of Intrusions in Computer Networks via Adaptive Sequential and Batch-Sequential Change-point Detection Methods,* IEEE Trans. Sig. Proc., vol. 54, no. 9, pp. 3372–3382, Sep. 2006.
[6] J. S. Baras, A. Cardenas, and V. Ramezani, *Distributed Change Detection for Worms, DDOS and Other Network Attacks,* Proc. American Cont. Conf. (ACC), Boston, MA, pp. 1008–1013, 2004.
[7] Husheng Li, Chengzhi Li and Huaiyu Dai, *Quickest spectrum sensing in cognitive radio,* Information Sciences and Systems. CISS. 42nd Annual Conference on, Princeton, NJ, pp. 203–208, 2008.
[8] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, *Multisensor data fusion: A review of the state-of-the-art,* Information Fusion, vol. 14, no. 1, pp. 28–44, 2013.
[9] Tartakovsky, A. G. and Veeravalli, V. V. *Asymptotically optimal quickest change detection in distributed sensor systems,* Sequential Anal. 27 pp. 441–475, 2008.
[10] Mei, Y. *Efficient scalable schemes for monitoring a large number of data streams,* Biometrika 97 pp. 419–433, 2010.
[11] Yao Xie and David Siegmund. *Sequential multi-sensor change-point detection,* in The Annals of Statistics, Vol. 41, No. 2, pp. 670–692, 2013.
[12] G. Shafer, *Perspectives on the theory and practice of belief functions,* International Journal of Approximate Reasoning, vol. 4, no. 5, pp. 323–362, 1990.
[13] F. Smarandache and J. Dezert, *Advances and Applications of DSmT for Information Fusion,* in American Research Press, Rehoboth, vol. 1-2, 2006.
[14] R. R. Yager and L. Liu, *Classic Works of the Dempster-Shafer Theory of Belief Functions,* in American Research Press, Rehoboth, vol. 2, ch. 1, pp. 3–68, 2006.
[15] P. Smets, *Constructing the pignistic probability function in a context of uncertainty,* Uncertainty in Artificial Intelligence, Vol. 5, pp. 29–39, Aug 2004.
[16] J. Sudano, *Yet Another Paradigm Illustrating Evidence Fusion (YAPIEF),* Proc. of Fusion, Florence, July 2006.
[17] T. L. Lai, *Information bounds and quick detection of parameter changes in stochastic systems,* in IEEE Transactions on Information Theory, vol. 44, no. 7, pp. 2917–2929, Nov 1998.
[18] I. F. Akyildiz, W. Lee, M. C. Vuran, and S. Mohanty, *Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey,* Comput. Netw., vol. 50, no. 13, pp. 21272159, Sep. 2006.
[19] S. Haykin, *Cognitive Radio: Brain-Empowered Wireless Communications,* IEEE JSAC, vol.23, no.2, pp.201-20, Feb. 2005.
[20] T. Yucek and H. Arslan, A survey of spectrum sensing algorithms for cognitive radio applications, in IEEE Communications Surveys & Tutorials, vol. 11, no. 1, pp. 116-130, First Quarter 2009.