國 立 交 通 大 學

電信工程學系碩士班
碩士論文

無線感測網路之無參數低複雜度改變檢測

Nonparametric Low Complexity Change Detection in Wireless Sensor Networks

研 究 生：羅楚威　　Student: Chu-Wei Lo

指導教授：李大嵩 博士　　Advisor: Dr. Ta-Sung Lee

中 華 民 國 九 十 六 年 六 月

無線感測網路之無參數低複雜度改變檢測

Nonparametric Low Complexity Change Detection in Wireless Sensor Networks

研 究 生：羅楚威　　Student: Chu-Wei Lo

指導教授：李大嵩 博士　　Advisor: Dr. Ta-Sung Lee

國立交通大學

電信工程學系碩士班

碩士論文

A Thesis
Submitted to Department of Communication Engineering
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in
Communication Engineering
June 2007
Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 六 年 六 月

無線感測網路之無參數低複雜度改變檢測

學生：羅楚威　　指導教授：李大嵩 博士

國立交通大學電信工程學系碩士班

摘要

吾人考慮感測器警報機率分佈改變偵測的問題。在無參數改變偵測的架構下，吾人基於 Rao test 發表一套演算法。吾人也將感測器分群並且估計各群的期望值來做檢測。我們獲得了理論上的效能。我們所提出方法的複雜度為線性，適用於多感測器的情況之下。吾人也考慮在感測器與資料融合中心的連線上有干擾情況時的效能增強。

Nonparametric Low Complexity Change Detection in Wireless Sensor Networks

Student: Chu-Wei Lo Advisor: Dr. Ta-Sung Lee

Department of Communication Engineering

National Chiao Tung University

Abstract

The problem of detecting changes in the distribution of alarmed sensors is considered. Under a nonparametric change detection framework, we present an algorithm based on the Rao test. We also partition sensors into small groups and estimate their mean to perform detection. Theoretical performance guarantees are obtained. Our approach has linear complexity, which is suitable to large number of sensors. We also enhance change detection performance for sensors-to-fusion links with interference.

Acknowledgement

I would like to express my deepest gratitude to my advisor, Dr. Ta-Sung Lee, for his enthusiastic guidance and great patience. I learn a lot from his positive attitude in many areas. Heartfelt thanks are also offered to all members in the Communication System Design and Signal Processing (CSDSP) Lab for their constant encouragement. Finally, I would like to show my sincere thanks to my parents for their invaluable love.

Contents

List of Figures

Acronym Glossary

AWGN	additive white Gaussian noise
BER	bit error rate
BPSK	binary phase shift keying
BSC	binary symmetric channel
CRLB	Cramér-Rao lower bound
IEEE	institute of electrical and electronics engineers
KL	Kullback Leibler
MVU	minimum variance unbiased
NP	Neyman- Pearson
PDF	probability density function
SAR	search in axis-aligned rectangles
SAS	search in axis-aligned strips
SDR	search in diagonal-defined axis-aligned rectangles
SNR	signal-to-noise ratio
VC	Vapnik-Chervonenki

Notations

$\mathcal{A}$	collection set of sensors
α_i	the sensor *i* changes its state at two time slots
D	Kullback Leibler distance
D_f	f -divergences
F	σ -algebra
$\mathcal{H}_i$	hypothesis i
I	The Fisher information
$p_i^{(m)}$	the alarm probability of sensor *i* at time *m*
$P^{(i)}$	probability measures under hypothesis $\mathcal{H}_i$
P_D	detection probability
P_F	false alarm probability
S	number of sensors
T	number of training
u	mean of the Normal distribution
X	sample space
$X_i^{(t)}$	the symbol that the *i*th sensor transmits to fusion center at time t
$Y_i^{(t)}$	the received signal with $X_i^{(t)}$ as the input to the BSC
σ_i^2	variance of the Normal distribution

Chapter 1

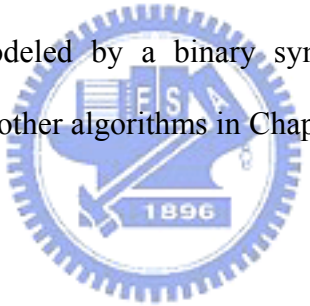
Introduction

Due to remarkable advances in the electronics industry, we have been enabled to construct low-power, low-cost, multifunctional sensors nodes [1] that are small in size and can transmit information in short distances. These sensor nodes are able to perform sensing, communication, and data/signal processing. A sensor network is composed of a large number of sensor nodes that are densely deployed either inside the prescribed phenomenon or very close to it [2]. In order to monitor the change of the environment, we shall seek for methods to detect the occurrence of change efficiently yet rapidly.

We consider the change detection problem in wireless sensor networks. We are interested in the change in the geographical distribution of alarmed sensors at two consecutive time instants. There are many related application of the change detection problem, e.g., health, military, and home surveillance. The problem of change detection in sensor field has been considered in different setting (mostly parametric) [3], [4]. A classical approach is the Kolmogorov-Smirnov two-sample test [5] in which the empirical cumulative distributions are compared, and the maximum difference in the empirical cumulative distribution functions is used as the test statistic. In [6], they focus on the same problem as ours based on the criteria of Vapnik-Chervonenkis dimension [7] and $\mathcal{A}$-distance [8]. There are many other change detection problems

that have been considered [6,9,10], but they all deal with detecting some phenomenon change by one observer over a long time period but not with detecting the geographical change by many observers(sensors) in two time instants. In [6], the complexity of the algorithms is quite intensive when the total number of sensors is large. We find that partitioning sensors into some small groups to perform detection has a lower complexity without degrading the performance.

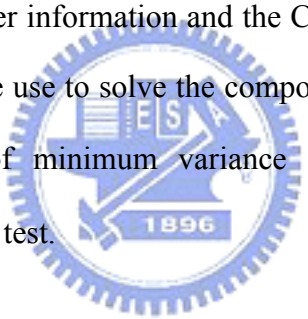
The thesis is organized as follows. Chapter 2 gives some preliminaries. In chapter 3, we develop the main algorithms and evaluate their performance. Our algorithm needs to estimate the probability distribution before changing based on the data samples. Hence, we will calculate the lower bound of the training data number. In Chapter 4 We will consider the problem of performance enhancement when the sensor-to-fusion link is modeled by a binary symmetric channel. Then, we will compare our algorithm with other algorithms in Chapter 5. Finally, chapter 6 concludes the thesis.



Chapter 2

Preliminary

In this chapter, we introduce our system model and some useful theorem in our thesis. We introduce *f*-divergence that is related to Kullback-Leibler distance. It is often useful to think of Kullback Leibler distance as a distance between two distributions. Then, we introduce the Fisher information and the Cramér-Rao lower bound. They are related to the criteria that we use to solve the composite hypothesis problem—the Rao test[11]—and the proof of minimum variance estimator. Finally, we introduce Neyman-Pearson hypothesis test.



2.1 System Model of Change Detection in Sensor Networks

In this section, we define the mathematical model and introduce the notation used through the thesis. There are S sensors deployed in the observation space. Let two probability measures $P^{(1)}$ and $P^{(2)}$ be on the same measure space (X,F) where $(X,F,P^{(t)})$ forms the *t*-th collection the locations of the alarmed sensors. In $(X,F,P^{(t)})$ shown in Figure 2.1, X is sample space and F is a σ-algebra.

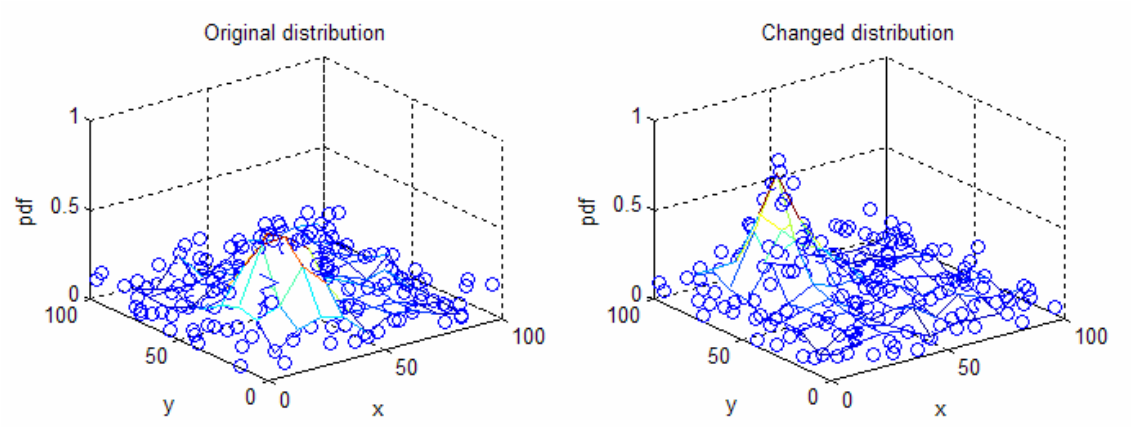


Figure 2.1: Probability distributions for all sensors before and after change

A sensor transmits a symbol designated by 1 to fusion center when it is alarmed, and transmits symbol 0 when otherwise. Let $X_i^{(t)}$ be the symbol that the *i*th sensor transmits to fusion center at time t:

$$X_i^{(t)} = \begin{cases} 0 & \text{(silence)} \\ 1 & \text{(alarm)} \end{cases} \text{ at time } t. \tag{2.1}$$

Figure 2.2 is the system block, in which each sensor-to-fusion link is modeled as a binary symmetric channel (BSC) accounting for, e.g., Rayleigh fading. $Y_i^{(t)}$ is the received signal with $X_i^{(t)}$ as the input to the BSC. The fusion center combines signals from all sensors to decide the change of the geographical distribution of the alarming sensors. Change detection is tested based on the data collected over two consecutive time instants. We define $p_i^{(m)}$ to be the alarm probability of sensor *i* at time *m*. The probability α_i that the sensor *i* changes its state at two time slots is

$$\alpha_i = p_i^{(1)}\left(1 - p_i^{(2)}\right) + p_i^{(2)}\left(1 - p_i^{(1)}\right). \tag{2.2}$$

Note that if $p_i^{(1)}$ is equal to 0.5, α_i is always equal to 0.5, no matter what the value of $p_i^{(2)}$ is.

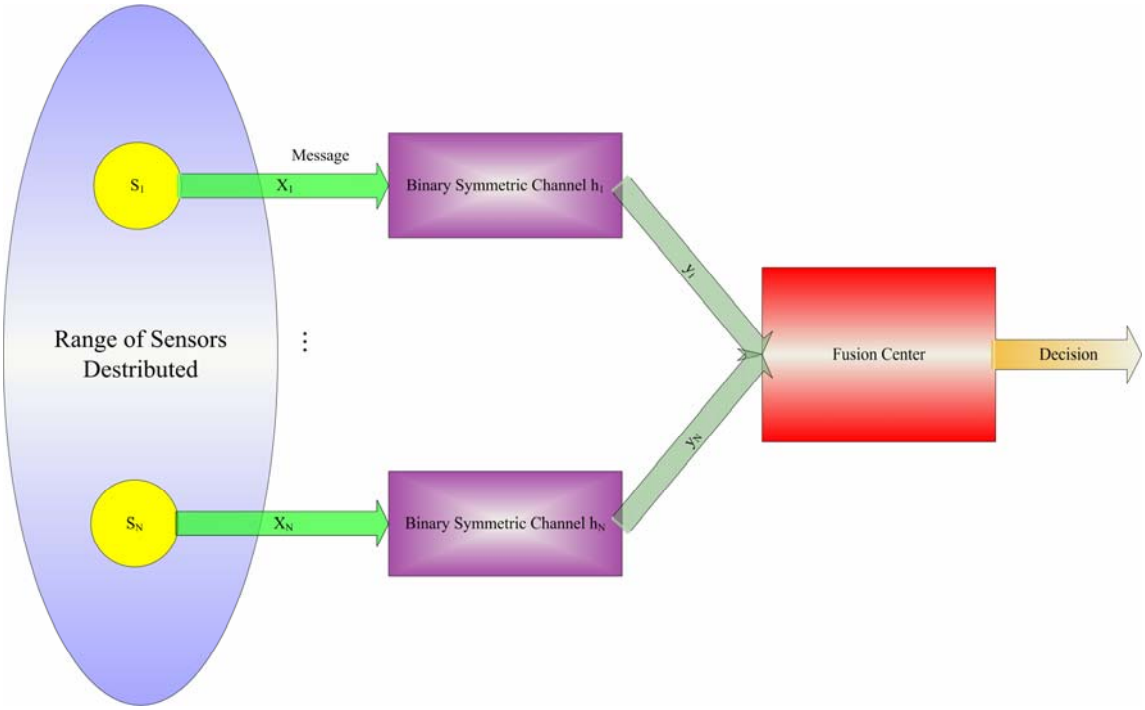


Figure 2.2: Block diagram of sensor networks

2.2 *f*-Divergences

A commonly used measure of the distance between two PDF is the f-divergences proposed by Cisiszár (1967). Let $f:[0,\infty)\to\Re\bigcup\{-\infty,\infty\}$ be a convex function with $f(1)=0$. The f-divergence between two probability measures u and v on $\Re^d$ is defined by

$$D_f(u,v)=\sup_{A=\{A_j\}}\sum_j v(A_j)f\left(\frac{u(A_j)}{v(A_j)}\right), \tag{2.3}$$

where the supremum is taken over all finite measurable partitions A of $\Re^d$. If λ is a measure dominating u and v—i.e., both u and v are absolutely continuous with respect to λ—and $p=du/d\lambda$ and $q=dv/d\lambda$ are the corresponding densities, then the f-divergences may be put in the form

$$D_f(u,v)=\int q(x)f\left(\frac{p(x)}{q(x)}\right)\lambda(x). \tag{2.4}$$

Information divergence is obtained by taking $f(x) = x \log x$:

$$I(u,v) = \sup_{A=\{A_j\}} \sum_j u(A_j) \log\left(\frac{u(A_j)}{v(A_j)}\right). \tag{2.5}$$

$I(u,v)$ is called the Kullback-Leilber (KL) distance. By Jensen's inequality, $D_f(u,v) \geq 0$ and $D_f(u,u) = 0$. Note that $I(u,v) \neq I(v,u)$. We will use KL distance to measure distance between two PDF in the followed chapter.

2.3 Fisher Information and Cramér-Rao Inequality

The Fisher information [12] $I(\theta)$ is the number of information that an observed random variable *X* carries about an unknown parameter θ when the underlying PDF is characterized by *f(x;θ)*. Because the expectation of the score is zero and the variance is simply the second moment of the score, the derivative of the logarithm of the likelihood function with respect to θ is the Fisher information $I(\theta)$ defined by

$$I(\theta) = E_\theta\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]^2. \tag{2.6}$$

It is assumed that the PDF *f(x;θ)* satisfies the "regularity" condition [12]

$$E\left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\right] = 0 \qquad for\ all\ \theta. \tag{2.7}$$

The regularity condition will be satisfied if the order of differentiation and integration can be interchanged. The Fisher information may also be expressed in other different forms. It is usually easy to calculate. It follows form the identity [12]

$$E\left[\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right]. \tag{2.8}$$

The mean squared error of any unbiased estimator T(***X***) of the parameter θ is the lower bounded by the reciprocal of the Fisher information, i.e.,

$$\mathrm{var}(T) \geq \frac{1}{I(\theta)}. \tag{2.9}$$

This lower bound is called the Cramér-Rao lower bound which yields the lowest possible variance for all unbiased estimators. A unbiased estimator T is said to be efficient if it attains the Cramér-Rao lower bound, i.e., $\mathrm{var}(T) = \frac{1}{I(\theta)}$. It gives us a lower bound on the error of estimation in the estimation of θ form the observable data. Also, this allows us to immediately determine if an estimator exists and attains the bound. An estimator attaining the Cramér-Rao lower bound is called minimum variance unbiased (MVU) estimator.

2.4 Central Limit Theorem

Results of many observations demonstrate that the sum of a large number of random variables converages toward the Normal one. The central limit theorem [13] provides a sufficient condition for this to hold.

Let $X_1, X_2, \ldots, X_N$ be a set of N independent random variables and each X_i have an arbitrary probability distribution with finite mean u_i and finite variance σ_i^2. The central limit theorem state that the random variable

$$X = \frac{\sum_{i=1}^{N} x_i - \sum_{i=1}^{N} u_i}{\sqrt{\sum_{i=1}^{N} \sigma_i^2}} \tag{2.10}$$

has a limiting cumulative distribution function which approaches a normal distribution. Note that the central the limit theorem shows the normal convergence of the characteristic function but not as yet the normal convergence of the PDF. Even though the PDF of interest (the difference between the numbers of alarming sensors) is a Bernoulli distribution which is discontinuous whereas the Gaussian is continuous, this result will not be changed. However, the integrals of discrete PDF, for large *N*, behave like integrals of the Gaussian PDF. This is why the distribution function of *X* tends to a

Gaussian distribution function but not necessarily to a Gaussian PDF. The "fuzzy" central limit theorem [13] says that data which are influenced by many small and unrelated random effects are approximately normally distributed.

In our research, each sensor has different probability. Hence, when we observe the whole of phenomenon of sensor, the central limit theorem can help us to model this situation suitably.

2.5 Neyman-Pearson Hypothesis Test

For composite hypothesis testing, we have many different definitions of optimality. Three most common formulations are—Bayes, minimax, and Neyman-Pearson [14]. In the Bayesian formulation, optimality is defined in terms of minimizing the cost function, define as the average risk; in the minimax formulation, prior probabilities are not assumed known and optimality is defined in term of minimizing the maximum of the conditional expected costs under the hypothesis. In practical problems, we often cannot obtain all the information about hypothesis to construct the cost function. In such cases an alternative design criterion, known as Neyman-Pearson criterion, is often utilized.

In testing two possible hypotheses, H_0 versus H_1, there are two types of error that can be made: H_0 can be falsely rejected or H_1 can be falsely rejected. The first one is called “missing” and the second called “false alarm”. Correct acceptance of H_1 is called “detection”. The Neyman-Pearson criterion to place a bound on the false-alarm probability and then to minimize the missing probability subject to this constraint; i.e., the Neyman-Pearson design criterion is

$$\max_{\delta} P_D(\delta) \text{ subject to } P_F(\delta) \le \alpha, \tag{2.11}$$

where δ is our detector, and α is the upper bound which is known as the level of

the test. Thus the goal of the Neyman-Pearson hypothesis test is to find the most powerful α -level test of H_0 versus H_1. Note that the Neyman-Pearson criterion recognizes asymmetry in importance of the two hypotheses.

There are three properties for the Neyman-Pearson design. Assume the hypothesis pair in which probability P_j has density p_j for $j = 0$ and $j = 1$, and suppose that $\alpha > 0$.

(1)Optimality: Let $\tilde{\delta}$ be any decision rule satisfying $P_F(\tilde{\delta}) \le \alpha$, and let $\tilde{\delta}'$ be any decision rule of the form

$$\tilde{\delta}'(y) = \begin{cases} 1 & \text{if } p_1(y) > \eta p_0(y) \\ r(y) & \text{if } p_1(y) = \eta p_0(y) \\ 0 & \text{if } p_1(y) < \eta p_0(y), \end{cases} \tag{2.12}$$

where $\eta \ge 0$ and $0 \le r(y) \le 1$ are such that $P_F(\tilde{\delta}') = \alpha$. Then $P_D(\tilde{\delta}') \ge P_D(\tilde{\delta})$. That is, any size-$\alpha$ decision rule of the above form is a Neyman-Pearson rule.

(2)Existence: For every $\alpha \in (0,1)$ there is a decision rule, $\tilde{\delta}_{NP}$, above form with $r(y) = r_0$, for which $P_F(\tilde{\delta}_{NP}) = \alpha$.

(3)Uniqueness: Assume that $\tilde{\delta}''$ is any α -level Neyman Pearson decision rule for H_0 versus H_1. Then $\tilde{\delta}''$ must be of the above form except possibly on a subset of Γ having zero probability under H_0 and H_1.

2.6 Summary

In this chapter, we introduce system model and many theorems that will be used in the following context. We will discuss system model without noise in Chapter 3; the problem of sensor model with noise will be considered in Chapter 4. We will impose Kullback-Leilber distance which relates to *f*-divergence to find the strategy of detection in the Section 3.2. Fisher information relates to the Rao test in Section 3.1 and

Cramér-Rao lower bound that will be exploited in the proof of minimum variance estimator in Section 3.3.2. The central limit theorem will be used when every sensor has different PDF and we want to model the PDF of sum of sensors information in Section 3.1.2, Section 3.2.2, and Chapter 4. The Neyman-Pearson test is related to the Rao test in Section 3.1.

Chapter 3

Nonparametric Change Detection in Sensors Networks

In this chapter, we consider the change detection problem in wireless sensor networks. We are interested in the change in the geographical distribution of alarmed sensors from data collections at two different times, as in [6]. Assume PDF of alarmed sensors does not change during a period of T time instants, and can thus be estimated by using the data samples collected during this time interval. We assume that the PDF after change is unknown. We will propose a Rao test based solution since this Rao's strategy requires the maximum likelihood parameter estimation priori to change but does not involve those upon the occurrence of change. In the next few sections, we will use the Rao test and Kullback Leibler distance to solve our problem. Finally, we will propose the partition method in our algorithm since we care about the geographical change of probability distribution.

3.1 Rao Test

In this part, we will introduce the method of the Rao test. The Rao test [11] has the asymptotic detection performance as the generalized likelihood ratio test. For finite

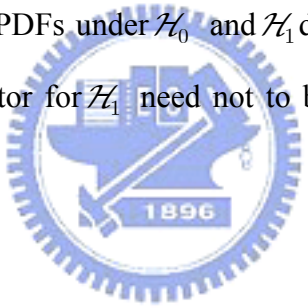
data records, there is no guarantee that the performance will be the same. The main benefit is that this asymptotically equivalent statistic may be easier to compute. This is especially true of the Rao test for which it is not necessary to determine the maximum likelihood estimator for $\mathcal{H}_1$, but only the maximum likelihood estimator for $\mathcal{H}_0$ to be found. The PDF is denoted $p(\mathbf{x};\theta)$. The hypothesis test is

$$\begin{aligned} \mathcal{H}_0 &: \theta = \theta_0 \\ \mathcal{H}_1 &: \theta \neq \theta_0. \end{aligned} \tag{3.1}$$

The Rao test just only needs to know θ_0, and is particularly suitable for the considered scenario. The Rao test decides $\mathcal{H}_1$ if

$$T_R(\mathbf{x}) = \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}^T \Bigg|_{\theta=\theta_0} \mathbf{I}^{-1}(\theta_0) \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} \Bigg|_{\theta=\theta_0} > \gamma, \tag{3.2}$$

where $\mathbf{I}(\theta_0)$ denote the Fisher information matrix, and γ is a threshold. In (3.2) it is implicitly assumed that the PDFs under $\mathcal{H}_0$ and $\mathcal{H}_1$ differ only in the value of θ. The maximum likelihood estimator for $\mathcal{H}_1$ need not to be found for the Rao test. This is advantageous when



3.1.1 Rao Test for Independent and Identically Distributed Sensors

In this part, we consider the simple homogeneous case, i.e., the alarming probabilities among sensors are independent and are identically distributed. We also assume for the moment that the channel is errorless (the cross-over probability of each BSC is zero). Under this condition, the number of sensors that change state x between two collections is Bernoulli distributed:

$$p(x;\alpha) = \binom{N}{x} \alpha^x (1-\alpha)^{N-x}, \tag{3.3}$$

where α is the probability of sensor changing state between two collections from (2.2). The composite hypothesis testing:

$$\mathcal{H}_0 : \alpha = \alpha_0$$
$$\mathcal{H}_1 : \alpha \neq \alpha_0.$$

where α_0 denotes the probability of sensor changing state between two collections before change occurs. The Fisher information for a Bernoulli distribution is given by

$$\begin{aligned}
\mathrm{I}(\alpha) &= -E\left[\frac{\partial^2}{\partial\alpha^2}\ln\left(p(x;\alpha)\right)\right] \\
&= -E\left[\frac{\partial^2}{\partial\alpha^2}\ln\left(\alpha^x(1-\alpha)^{s-x}\frac{N!}{x!(N-x)!}\right)\right] \\
&= -E\left[\frac{\partial^2}{\partial\alpha^2}\left[x\ln(\alpha)+(N-x)\ln\left((1-\alpha)\right)\right]\right] \\
&= -E\left[\frac{\partial}{\partial\alpha}\left[\frac{x}{\alpha}-\frac{N-x}{1-\alpha}\right]\right] \\
&= E\left[\left[\frac{x}{\alpha^2}+\frac{N-x}{(1-\alpha)^2}\right]\right] \\
&= \frac{N\alpha}{\alpha^2}+\frac{N(1-\alpha)}{(1-\alpha)^2} \\
&= \frac{N}{\alpha(1-\alpha)},
\end{aligned}$$

and

$$\frac{\partial}{\partial\alpha}\ln\left(p(x;\alpha)\right) = \frac{x}{\alpha}-\frac{N-x}{1-\alpha}.$$

Substituting above result into the formula of Rao test gives

$$T_R(x) = \frac{\alpha_0(1-\alpha_0)}{N}\left(\frac{x}{\alpha_0}-\frac{N-x}{1-\alpha_0}\right)^2 > \gamma. \tag{3.4}$$

Such that

$$(x-N\alpha_0)^2 > \gamma'. \tag{3.5}$$

Thus the Rao test in our case claims change occurs if the squared difference between the measurement *x* and the mean $N\alpha_0$ exceeds a certain threshold γ'.

3.1.2 Rao Test for Independent and Different Distributed Sensors

In this part, we consider the inhomogeneous case in which the alarming

probabilities of sensors are distinct. This case arises, e.g., when the alarming probabilities of sensors are identical but the BSC's assume different cross-over probabilities across the sensor-to-fusion links. We do not care the changes of probability distribution in each sensor, but in the area. Hence, we combine sensors in the same area to perform detection. By the central limit theorem, when the number of sensors is large and sensors are independent, the probability distribution of the number of sensors alarming looks like the Normal distribution. So, the PDF of number of sensor alarming is denoted by $p(x;u,\sigma^2)$, where p is the Normal distribution, u is the mean, and σ^2 is variance. Consider the composite hypothesis problem

$$\begin{aligned} \mathcal{H}_0 &: u = u_0, \sigma^2 = \sigma_0^2 \\ \mathcal{H}_1 &: u \neq u_0, \sigma^2 \neq \sigma_0^2. \end{aligned} \tag{3.6}$$

This is a two-parameters composite hypothesis problem. The Normal distribution, $N(u,\sigma^2)$ belongs to the exponential family and its log-likelihood function $l(\theta \mid x)$ is

$$-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-u)^2}{2\sigma^2} \tag{3.7}$$

where $\theta = \left(u, \sigma^2\right)$. The Fisher information matrix $\mathbf{I} = -E\left[\frac{\partial U}{\partial \theta}\right]$, where U is given by

$$\left(\frac{\partial l}{\partial u}, \frac{\partial l}{\partial \sigma^2}\right) = \left(\frac{x-u}{\sigma^2}, \frac{(x-u)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right). \tag{3.8}$$

Taking the derivative with respect to θ, we have

$$\frac{\partial U}{\partial \theta} = \begin{pmatrix} \frac{\partial U_1}{\partial u} & \frac{\partial U_2}{\partial u} \\ \frac{\partial U_1}{\partial \sigma^2} & \frac{\partial U_2}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{-1}{\sigma^2} & -\frac{x-u}{\sigma^4} \\ -\frac{x-u}{\sigma^4} & \frac{1}{2\sigma^4} - \frac{(x-u)^2}{\sigma^6} \end{pmatrix}. \tag{3.9}$$

So, the Fisher information matrix **I** is

$$-E\left[\frac{\partial U}{\partial \theta}\right] = \frac{1}{2\sigma^4}\begin{pmatrix} 2\sigma^2 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.10}$$

The Rao test decides $\mathcal{H}_1$, if

$$
\begin{aligned}
T_R(\mathbf{x}) &= \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}^T \Bigg|_{\theta=\left(m_0,\sigma_0^2\right)} \mathbf{I}^{-1}(\theta_0) \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} \Bigg|_{\theta=\left(m_0,\sigma_0^2\right)} \\
&= 2\sigma_0^4 \begin{bmatrix} \dfrac{x-u_0}{\sigma_0^2} & \dfrac{(x-u_0)^2}{\sigma_0^4} - \dfrac{1}{2\sigma_0^2} \end{bmatrix} \begin{bmatrix} \dfrac{1}{2\sigma_0^2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \dfrac{x-u_0}{\sigma_0^2} \\ \dfrac{(x-u_0)^2}{\sigma_0^4} - \dfrac{1}{2\sigma_0^2} \end{bmatrix} \\
&= \frac{(x-u_0)^4}{2\sigma_0^2} + \frac{1}{2} > \gamma
\end{aligned}
\tag{3.11}
$$

The statistic (3.10) is forth moment of $x-u_0$. Compare (3.10) with (3.4), (3.10) is the square of (3.4).

3.2 Kullback Leibler Distance

In probability theory and information theory, the Kullback Leibler distance is a measure of the distance between two distributions. In statistics, it arises as an expected logarithm of the likelihood ratio. The Kullback Leibler distance $D(p \,||\, q)$ is a measure of the inefficiency of incorrectly taking the distribution as *q* when the true distribution is instead *p*. For example, if we knew the true distribution of the random variable, then we could construct a code with average description length $H(p)$. If, instead, we use the code for a distribution *q*, we would need $H(p) + D(p \,||\, q)$ bits on the average to describe the random variable.

The Kullback Leibler distance between two probability mass functions *p(x)* and *q(x)* is defined as

$$
\begin{aligned}
D(p \,||\, q) &= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \\
&= E_p \left[\log \frac{p(x)}{q(x)} \right].
\end{aligned}
\tag{3.12}
$$

The Kullback Leibler distance is always non-negative, it is zero if only if *p*=*q* [12]. However, it is not a true distance measure between distributions because it is not symmetric and does not satisfy the triangle inequality,

`

$$D(p \,||\, q) \neq D(q \,||\, p). \tag{3.13}$$

Nonetheless, it is often useful to think of Kullback Leibler distance as a distance between two distributions. The Kullback Leibler distance remains well-defined for continuous distributions, and furthermore is invariant under parameter transformations.

3.2.1 Kullback Leibler Distance for Independent and Identically Distributed Sensors

In the Section 3.1, we discuss the detection problem from the viewpoint of the Rao test. In this section, we will consider this problem from another viewpoint, the Kullback Leibler distance. First, we discuss the homogeneous case in which the alarming probabilities of sensors are distinct. We also assume that the error probability in the binary symmetric channel is equal to zero. Under above condition, the number of sensors that change state between two collections is a Bernoulli distribution. We assume that p and q are the alarming probabilities before and after change occurs, and then we have:

$$p(x;\alpha) = \binom{N}{x} \alpha^x (1-\alpha)^{N-x} \tag{3.14}$$

and

$$q(x;\alpha') = \binom{N}{x} (\alpha')^x (1-\alpha')^{N-x}, \tag{3.15}$$

where N is the number of sensors, x is number of alarming sensors, α is from (2.2), $\alpha' = \alpha + \Delta$, and $-\alpha \le \Delta \le 1-\alpha$. Because sensors have the same probability distribution, we remove the index word i of α . The Kullback Leibler distance $D(p \,||\, q)$ between p and q is

$$
\begin{aligned}
D(p \,||\, q) &= \sum_{x=0}^{N} \binom{N}{x} \alpha^x (1-\alpha)^{N-x} \log\left(\frac{\alpha^x (1-\alpha)^{N-x}}{(\alpha')^x (1-\alpha')^{N-x}} \right) \\
&= \sum_{x=0}^{N} \binom{N}{x} \alpha^x (1-\alpha)^{N-x} \log\left(\frac{\alpha^x (1-\alpha)^{N-x}}{(\alpha+\Delta)^x (1-(\alpha+\Delta))^{N-x}} \right) \qquad (3.16) \\
&= \sum_{x=0}^{N} \binom{N}{x} \alpha^x (1-\alpha)^{N-x} \log\left(\frac{1}{(1+\frac{\Delta}{\alpha})^x (1-\frac{\Delta}{1-\alpha})^{N-x}} \right).
\end{aligned}
$$

We declare the probability distribution changed if the Kullback Leibler distance $D(p \,||\, q)$ is larger than λ. Otherwise, we declare the probability distribution non-changed. This composite hypothesis test can be written as

$$
\begin{aligned}
& H_0 : D(p \,||\, q) \le \lambda \\
& \text{versus} \qquad (3.17) \\
& H_1 : D(p \,||\, q) > \lambda.
\end{aligned}
$$

Actually, this composite hypothesis problem (3.16) can be rewritten as

$$
\begin{aligned}
& H_0 : \varepsilon \le \Delta \le \eta \\
& \text{versus} \qquad (3.18) \\
& H_1 : \Delta < \varepsilon \ \text{or} \ \Delta > \eta.
\end{aligned}
$$

Hence, we decide H_1 if

$$
\begin{aligned}
& x - N\alpha < \varepsilon \\
& \text{or} \qquad (3.19) \\
& x - N\alpha > \eta.
\end{aligned}
$$

Compare (3.18) with (3.5), and we find that (3.5) is equivalent to (3.18) when $-\varepsilon = \eta$. However, from (3.16) and (3.17), the values of ε and η in (3.18) is difficultly to obtain. We have to use numerical method to get them. This is an obstruction for us to use the composite testing from the viewpoint of the Kullback Leibler distance.

3.2.2 Kullback Leibler Distance for Independent and Different Distributed Sensors

For the inhomogeneous case the central limit theorem implies that, when the

number of sensors is large and sensors are independent, the probability distribution of the number of sensors alarming likes the Normal distribution. So, the PDFs of non-changed p and changed q are approximately given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(x-u_0)^2}{2\sigma_0^2} \right]$$

and (3.20)

$$q(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(x-u_1)^2}{2\sigma_1^2} \right].$$

The Kullback Leibler distance $D(p \,||\, q)$ between p and q is

$$\begin{aligned} D(p \,||\, q) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(x-u_0)^2}{2\sigma_0^2} \right] \log\left(\sqrt{\frac{\sigma_1^2}{\sigma_0^2}} \frac{\exp\left[-\frac{(x-u_0)^2}{2\sigma_0^2} \right]}{\exp\left[-\frac{(x-u_1)^2}{2\sigma_1^2} \right]} \right) dx \\ &= \log \sqrt{\frac{\sigma_0^2 + \Delta_{\sigma^2}}{\sigma_0^2}} \\ &\quad + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(x-u_0)^2}{2\sigma_0^2} \right] \left[\frac{\sigma_0^2\Delta_u^2 - 2\sigma_0^2\Delta_u(x-u_0) - \Delta_{\sigma^2}(x-u_0)^2}{2\sigma_0^2\left(\sigma_0^2 + \Delta_{\sigma^2}\right)} \right] dx, \end{aligned} \tag{3.21}$$

where $\Delta_u = u_1 - u_0$, and $\Delta_{\sigma^2} = \sigma_1^2 - \sigma_0^2$. This case with two variables is difficult to analyze, and we cannot get the variance from only one data sample. Hence, we only consider the variance invariant case. The Kullback Leibler distance $D(p \,||\, q)$ between p and q is

$$\begin{aligned} D(p \,||\, q) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(x-u_0)^2}{2\sigma_0^2} \right] \log\left(\frac{\exp\left[-\frac{(x-u_0)^2}{2\sigma_0^2} \right]}{\exp\left[-\frac{(x-u_1)^2}{2\sigma_1^2} \right]} \right) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-u_0)^2}{2\sigma^2} \right] \frac{-\Delta_u(2x - 2u_0 - \Delta_u)}{2\sigma^2} dx. \end{aligned} \tag{3.22}$$

We declare the probability distribution changes if the Kullback Leibler distance $D(p \,||\, q)$ is larger than λ. Otherwise, we declare the probability distribution does not change. The composite hypothesis test can be written as

$$H_0 : D(p \,||\, q) \le \lambda$$
$$\text{versus} \tag{3.23}$$
$$H_1 : D(p \,||\, q) > \lambda.$$

Also, this composite hypothesis problem (3.22) can be rewritten as

$$H_0 : \varepsilon \le \Delta_u \le \eta$$
$$\text{versus} \tag{3.24}$$
$$H_1 : \Delta_u < \varepsilon \ or \ \Delta_u > \eta.$$

Hence, we decide H_1 if

$$x - u_0 < \varepsilon,$$
$$\text{or} \tag{3.25}$$
$$x - u_0 > \eta.$$

3.3 Algorithm of Nonparametric Low Complexity Change Detection in Wireless Sensor Networks

In this section, we propose our nonparametric low complexity change detection algorithm based on the results of sections 3.1 and 3.2. We propose to partition sensors into several small groups according to their position. If we find that the probability distribution of *at least one among these groups changes*, we declare that the two probability distributions are different. We also have to estimate a parameter in our algorithm. We calculate the lower bound of the number of the data samples for estimation. Then, we discuss the performance of our algorithm and simulations.

3.3.1 Algorithm

We are interested in the change in the geographical distribution of alarmed sensors. Hence, we propose to partition sensors into several small groups according to

their positions to detect geographical distribution change. If we find that the probability distribution of *at least one among these groups changes*, we declare that the two probability distributions are different.

In (3.5) and (3.11), we derive the results of composite hypothesis testing by the Rao test. Alternatively, in (3.19) and (3.25), we got the results of composite hypothesis testing based on the Kullback Leibler distance. However, we have to use numerical methods to find the values of ε and η in (3.19) and (3.25). This increases the difficulty of implementing our algorithm. From (3.5) and (3.11), we propose that if the composite hypothesis testing is:

$$\begin{aligned} &\mathcal{H}_0 : m^{(j)} = m_0^{(j)}, \text{ for all } j = 1,\ldots,K \\ &\mathcal{H}_1 : m^{(j)} \neq m_0^{(j)}, \text{ for some } j. \end{aligned} \tag{3.26}$$

where $m_0^{(j)}$ is the mean of distribution of *j*th partition under $\mathcal{H}_0$ and *K* is the number of partitions, we decide $\mathcal{H}_1$ if

$$\left| x^{(j)} - m_0^{(j)} \right| > r^{(j)}, \quad \text{for some } j, \tag{3.27}$$

where $x^{(j)}$ is the number of sensor changing state and $r^{(j)}$ is a threshold in the *j*th partition. We assume that the PDF under $\mathcal{H}_0$ does not change during a periods of *T* time instants (we call *T* "training number"). Since we do not know the PDF under $\mathcal{H}_0$, we have to estimate the mean $m_0^{(j)}$. We estimate $m_0^{(j)}$ as

$$\hat{m}_0^{(j)} = \frac{1}{T} \sum_{t=0}^{T-1} x^{(j)}[t]. \tag{3.28}$$

This estimator is a minimum variance unbiased estimator for both homogeneous and inhomogeneous cases, and we will proof it later. Then, we design the thresholds of partitions by Neyman-Pearson's criterion that maximizes the detection probability subject to that the false alarm probability is not larger than the threshold.

There are five steps in our algorithm:

Step1. Partition sensors into *K* groups according to their position.

Step2. Record the number of sensor changing state in each partition.

Step3. During *T* pairs of consecutive two time instants, we estimate the mean in each partition by summing average (3.28).

Step4. According to Neyman-Person test, we set threshold in (3.27) for every partition.

Step5. In every two time instants, repeat the test of (3.27).

3.3.2 Lower Bound on Training Number

From previous section, we have designed the algorithm of change detection. In our algorithm, we have to know the original probability distribution via the observe sample sequence (a "training" process). We address the problem: how large the number of data samples is needed for guaranteeing the estimation accuracy to be within a prescribed level? This is an important issue in our research. We have to do a tradeoff between the detection performance and the calculation complexity. First, for the homogeneous case consider the set of observations

$$x[t] = c(t;\alpha), t = 0,1,...,T-1, \tag{3.29}$$

where $c(t;\alpha) \sim$ Binormail distribution $p(x;\alpha) = \binom{N}{x}\alpha^x(1-\alpha)^{N-x}$ and N is the number of sensors. Then, the estimator

$$\hat{\alpha} = \frac{1}{NT}\sum_{t=0}^{T-1} x[t] \tag{3.30}$$

from Section 3.3 is minimum variance unbiased estimator[15]. Because

$$I(\alpha) = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\alpha)}{\partial \alpha^2}\right]$$
$$= -TE\left[\frac{\partial^2}{\partial \alpha^2}\left[\ln\binom{N}{x} + x[t]\ln\alpha + (N - x[t])\ln(1-\alpha)\right]\right]$$
$$= \frac{NT}{\alpha(1-\alpha)}$$

and

$$\frac{\partial \ln p(x;\alpha)}{\partial \alpha} = I(\alpha)(\hat{\alpha} - \alpha).$$

A good approximation for the Binomial distribution is the Normal distribution when $N\alpha(1-\alpha) \gg 1$. In our problem, N is usually large. Hence, the Normal distribution is a good approximation for the Binomial distribution. The probability distribution of $\hat{\alpha}$ is approximated

$$\hat{\alpha} = \frac{1}{NT}\sum_{t=0}^{T-1} x[t] \sim N\left(\alpha, \frac{\alpha(1-\alpha)}{NT}\right). \tag{3.31}$$

We want to know how large T can guarantee the probability that the deviation of the estimated mean from the true mean $N|\hat{\alpha} - \alpha|$ is less than D to be greater than ε. Toward a solution we note that

$$erf\left(\frac{|\hat{\alpha} - \alpha|}{\sqrt{2I(\alpha)}}\right) \geq \varepsilon,$$
$$\frac{(\hat{\alpha} - \alpha)^2 TN}{2\alpha(1-\alpha)} \geq [erfinv(\varepsilon)]^2,$$
$$T \geq \frac{N(erfinv(\varepsilon))^2}{2D^2} \text{(by } \alpha(1-\alpha) \leq \frac{1}{4}\text{)}, \tag{3.32}$$

where $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2} dt$ is the error function and

$erfinv$ is the inverse error function.

Now, we discuss the general inhomogeneous case. Consider the observation

$$x[t] = c(t; m, \sigma^2), t = 0,1,...,T-1, \tag{3.33}$$

where $c(t; m, \sigma^2) \sim$ Normal distribution $p(x; m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$. The

estimator

$$\hat{m} = \frac{1}{T}\sum_{t=0}^{T-1} x[t] \tag{3.34}$$

is the minimum variance unbiased estimator. Because

$$\begin{aligned} p(\mathbf{x};m) &= \prod_{t=0}^{T-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x[t]-m)^2}{2\sigma^2}\right] \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=0}^{T-1}(x[t]-m)^2\right], \\ I(\alpha) &= -E\left[\frac{\partial^2 \ln p(\mathbf{x};m)}{\partial m^2}\right] \\ &= \frac{T}{\sigma^2}, \end{aligned}$$

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x};m)}{\partial m} &= \frac{\partial}{\partial m}\left[-\ln\left[\left(2\pi\sigma^2\right)^{\frac{N}{2}}\right] - \frac{1}{2\sigma^2}\sum_{t=0}^{T-1}(x[t]-m)^2\right]. \\ &= \frac{1}{\sigma^2}\sum_{t=0}^{T-1}(x[t]-m) \\ &= \frac{N}{\sigma^2}(\hat{m}-m) \end{aligned} \tag{3.35}$$

So, the probability distribution of $\hat{m}$ can be written

$$\hat{m} = \frac{1}{T}\sum_{t=0}^{T-1} x[t] \sim N\left(m, \frac{\sigma^2}{T}\right). \tag{3.36}$$

From (3.7),

$$\begin{aligned} & erf\left(\frac{|\hat{m}-m|\sqrt{T}}{\sqrt{2\sigma^2}}\right) \geq \varepsilon, \\ & \frac{(\hat{m}-m)^2 T}{2\sigma^2} \geq [erfinv(\varepsilon)]^2, \\ & T \geq \frac{2\sigma^2 (erfinv(\varepsilon))^2}{D^2}, \\ & T \geq \frac{N(erfinv(\varepsilon))^2}{2D^2}, \ by\ \sigma^2 \leq \frac{N}{4}. \end{aligned} \tag{3.37}$$

We can find that (3.7) is the same with (3.12). Hence, we get a lower bound in the common use.

3.3.3 Performance

In this section, we present the detection probability P_D and false alarm probability P_F for our estimator. From (3.27), the detection probability $P_D^{(j)}$ and false alarm probability $P_F^{(j)}$ in the *j*th partition are

$$P_D^{(j)} = \Pr\left\{ \mid x^{(j)} - m_0^{(j)} \mid > r^{(j)} \mid \mathcal{H}_1 \right\}$$

and (3.38)

$$P_F^{(j)} = \Pr\left\{ \mid x^{(j)} - m_0^{(j)} \mid > r^{(j)} \mid \mathcal{H}_0 \right\}.$$

In the homogeneous case, the PDF $P_j(x^{(j)})$ of number $x^{(j)}$of sensor changing state in *j*th partition is the Binomial distribution $B(S, \alpha)$, where S is the number of sensors, α is probability of sensor changing state. The detection probability $P_D^{(j)}$ and false alarm probability $P_F^{(j)}$ can be determined by

$$\begin{aligned} P_D^{(j)} &= \Pr\left\{ \mid x^{(j)} - m_0^{(j)} \mid > r^{(j)} \mid \mathcal{H}_1 \right\} \\ &= \Pr\left\{ \mid x^{(j)} - S^{(j)}\alpha_0^{(j)} \mid > r^{(j)} \mid \mathcal{H}_1 \right\} \\ &= 1 - \sum_{x=\left\lceil S^{(j)}\alpha_0^{(j)} - r^{(j)} \right\rceil}^{\left\lfloor S^{(j)}\alpha_0^{(j)} + r^{(j)} \right\rfloor} P\left(x; \mathcal{H}_1\right), \end{aligned}$$

and (3.39)

$$\begin{aligned} P_F^{(j)} &= \Pr\left\{ \mid x^{(j)} - m_0^{(j)} \mid > r^{(j)} \mid \mathcal{H}_0 \right\} \\ &= \Pr\left\{ \mid x^{(j)} - S^{(j)}\alpha_0^{(j)} \mid > r^{(j)} \mid \mathcal{H}_0 \right\} \\ &= 1 - \sum_{x=\left\lceil S^{(j)}\alpha_0^{(j)} - r^{(j)} \right\rceil}^{\left\lfloor S^{(j)}\alpha_0^{(j)} + r^{(j)} \right\rfloor} P\left(x; \mathcal{H}_o\right). \end{aligned}$$

where $\lfloor\ \rfloor$ is the floor function and $\lceil\ \rceil$ is the ceil function. When $N\alpha(1-\alpha) \gg 1$, the Normal distribution is a good approximation for the Binomial distribution. Then (3.39) can be approximated by

$$P_F^{(j)} = 2\left(1 - \Phi\left\{\frac{r^{(j)} + S\alpha_0^{(j)}}{\sqrt{S\alpha_0^{(j)}(1-\alpha_0^{(j)})}}\right\}\right),$$

and (3.40)

$$P_D^{(j)} = 1 - \left\{\Phi\left\{\frac{r^{(j)} + S(\alpha_0^{(j)} - \alpha_1^{(j)})}{\sqrt{S\alpha_1^{(j)}(1-\alpha_1^{(j)})}}\right\} - \Phi\left\{\frac{-r^{(j)} + S(\alpha_0^{(j)} - \alpha_1^{(j)})}{\sqrt{S\alpha_1^{(j)}(1-\alpha_1^{(j)})}}\right\}\right\},$$

where Φ is the CDF of the Normal distribution $N(0,1)$ and α_1 denotes the alarming probability after change occurs. From (3.40), when the number of sensors tends to infinity, the false alarm probability approaches zero and the detection probability is approaches unity. So, if we want to improve the detection performance, we have to increase the number of sensors in a partition. However, if we increase the number of sensors in a partition, the number of partitions is decreased (since the total number of sensors is fixed) and this may incur geographical detection performance degradation. There is thus a tradeoff between the number of partitions and the achievable performance.

In the inhomogeneous case, the PDF $P_j(x^{(j)})$ of number $x^{(j)}$ of sensor changing state in *j*th partition is the Normal distribution $N(m, \sigma^2)$. The false alarm probability and detection probability can be written as

$$P_F^{(j)} = 2\left(1 - \Phi\left\{\frac{r^{(j)} + m_0^{(j)}}{\sigma_0^{(j)}}\right\}\right),$$

and (3.41)

$$P_D^{(j)} = 1 - \left\{\Phi\left\{\frac{r^{(j)} + m_0^{(j)} - m_1^{(j)}}{\sigma_1^{(j)}}\right\} - \Phi\left\{\frac{-r^{(j)} + m_0^{(j)} - m_1^{(j)}}{\sigma_1^{(j)}}\right\}\right\}.$$

From above description, we present the detection probability and the false alarm probability in each partition. Now, we will discuss the *total* detection probability P_D and the false alarm probability P_F. From (3.26), we have P_F:

$$
\begin{aligned}
P_F &= \Pr\{H_1 \mid H_0\} \\
&= 1 - \Pr\{H_0 \mid H_0\} \\
&= 1 - \prod_{j=1}^{K} \left(1 - P_F^{(j)}\right).
\end{aligned} \tag{3.42}
$$

We also assume A_j is the *j*th combination of partition with distribution change, and the set of total combinations is $A = \{A_1, A_{2,\ldots,} A_j\}$. We have P_D

$$
P_D = 1 - \sum_{A_j \in A} \Pr\{A_j\} \prod_{a \in A_j} \left(1 - P_D^{(a)}\right) \prod_{b \in \Omega - A_j} \left(1 - P_D^{(b)}\right). \tag{3.43}
$$

3.4 Computer Simulation

In this part, we show our simulation result by the Matlab. In Figure 3.1, we set the number of sensors to be 4096. The (x, y) is the coordinate of sensor, $x \in \{0,1,\ldots,61\}$ and $y \in \{0,1,\ldots,61\}$. Before change occurs, sensors have identical alarming probability equal to 0.8. The change occurs in sensors position at $x \in \{0,1,\ldots,31\}$ and $y \in \{0,1,\ldots,31\}$ with alarming probability decreases to 0.7. Use 100,000 Monte Carlo runs. As our description in Section (3.3.3), if we want to improve the detection performance, we have to increase the number of sensors in a partition.

In Figure 3.2, we set the number of sensors to be 3600. We partition sensors in to nine groups. Before change occurs, sensors have identical alarming probability equal to 0.8. The change occurs in two partitions, with alarming probability decreases to 0.7. We simulate the training number from 11 to 24. Use 50,000 Monte Carlo runs. We have a tradeoff between performance and complexity. In this case, we find that we have performance similar to the optimal optimum as the training number equals 24.

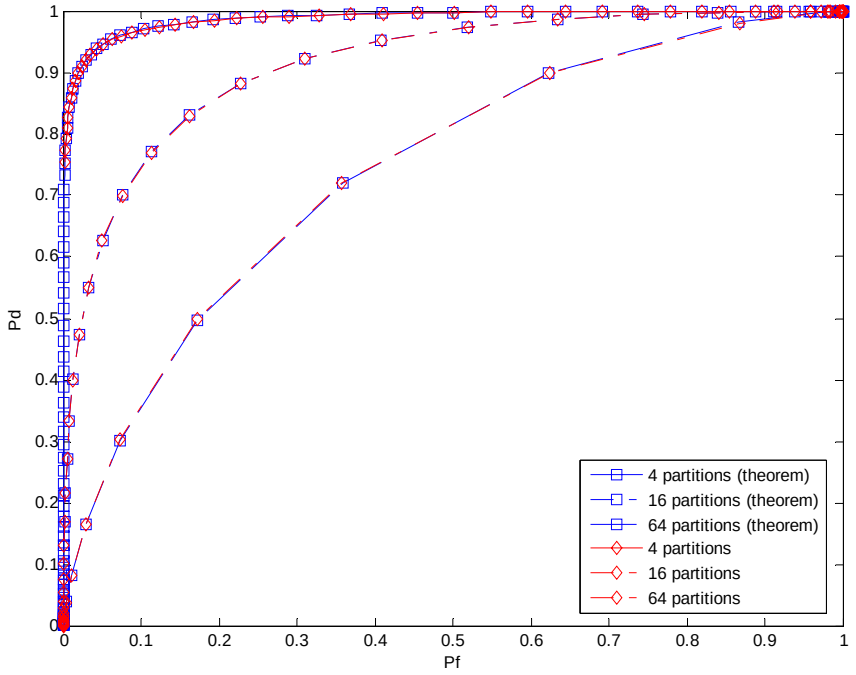


Figure 3.1: Receiver operating characteristics for different partition number

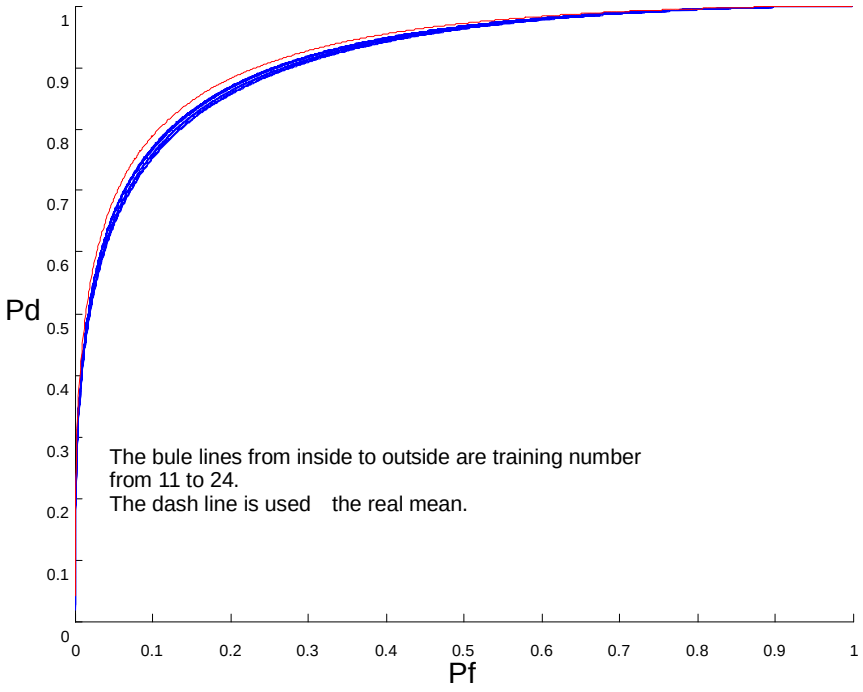


Figure 3.2: Receiver operating characteristics for different training number

3.5 Summary

In this chapter, we consider our detection problem from two different aspects, the Rao's test and the Kullback Leibler distance in Section 3.1 and Section 3.2. We then propose a new change detection algorithm, and then discuss the lower bound on training number required for estimating the ensemble statistics. Finally, we demonstrate via numerical simulation the performance of the proposed method.

Chapter 4

Detection Performance Enhancement Against Binary Symmetric Channel Links

Our previous discussions assume that all the sensor-to-fusion links are error-free. In this chapter we model each communication link as a BSC. We introduce a criterion for characterizing the received data quality, and propose a method for detection performance improvement via dropping sensors yielding bad data quality.

4.1 Definition of Data Quality

Since all the sensor-to-fusion links are modeled as binary symmetric channels with distinct cross-over probabilities, the number of sensor changing state in every partition is Normally distributed according to $N(m,\sigma^2)$. From (3.41), we have

$$P_F^{(j)} = 2\left(1-\Phi\left\{\frac{r^{(j)}+m_0^{(j)}}{\sigma_0^{(j)}}\right\}\right),$$

and

$$P_D^{(j)} = 1-\left\{\Phi\left\{\frac{r^{(j)}+m_0^{(j)}-m_1^{(j)}}{\sigma_1^{(j)}}\right\}-\Phi\left\{\frac{-r^{(j)}+m_0^{(j)}-m_1^{(j)}}{\sigma_1^{(j)}}\right\}\right\}.$$

Under the same false alarm probability in *j*th partition

$$P_F^{(j)} = 2\left(1-\Phi\{k\}\right),$$
$$P_D^{(j)} = 1-\left\{\Phi\left\{\frac{\Delta^{(j)}+k\sigma_0^{(j)}}{\sigma_1^{(j)}}\right\}-\Phi\left\{\frac{\Delta^{(j)}-k\sigma_0^{(j)}}{\sigma_1^{(j)}}\right\}\right\}, \tag{3.44}$$

where $\Delta^{(j)} := m_0 - m_1$, and k is a constant. We want to find the combination α of sensors in jth partition that

$$\max_{\alpha\in A^{(j)}} P_D^{(j)} = \min_{\alpha\in A^{(j)}}\left(\Phi\left(\frac{\sum_{i\in\alpha}\Delta_i^{(j)}+k\sqrt{\sum_{i\in\alpha}(\sigma_{0i}^{(j)})^2}}{\sqrt{\sum_{i\in\alpha}(\sigma_{1i}^{(j)})^2}}\right)-\Phi\left(\frac{\sum_{i\in\alpha}\Delta_i^{(j)}-k\sqrt{\sum_{i\in\alpha}(\sigma_{0i}^{(j)})^2}}{\sqrt{\sum_{i\in\alpha}(\sigma_{1i}^{(j)})^2}}\right)\right) \tag{3.45}$$
$$\left(\text{by } \Delta = \sum_{i\in\alpha}\Delta_i^{(j)} \text{ and } (\sigma_{li}^{(j)})^2 = \sum_{i\in\alpha}(\sigma_{li}^{(j)})^2\right),$$

where $A^{(j)}$ are all possible combinations of sensors in jth partition, $\Delta_i^{(j)}$ is the difference of mean of sensor i at jth partition, $\sigma_{li}^{(j)}$ is the variance of sensor i at jth partition in $\mathcal{H}_l$, $\Delta_i^{(j)}$ and $\sigma_{1i}^{(j)}$ are unknown parameters. While problem (4.2) appears quite difficult to be tackled, it motivates us to develop a method for dropping sensors with bad data quality for detection performance improvement.

In Figure 4.1, the alarming probability of sensor i is changed by BSC. The changed alarm probability $p_i^{'}$ is

$$p_i + (1 - 2p_i)\varepsilon_i. \tag{3.46}$$

The state-changing probability $\alpha_{0i}^{'}$ for sensor i before change occurs is

$$2p_i^{'}\left(1-p_i^{'}\right) = (-8p_i^{\,2} + 8p_i - 2)\varepsilon_i^2 + (8p_i^{\,2} - 8p_i + 2)\varepsilon_i - 2p_i^{\,2} + 2p_i. \tag{3.47}$$

After change occurs, the p_i becomes $p_i + \Delta_i$. The state-changing probability $\alpha_{0i}^{'}$ for sensor i after change occurs is

$$\left(p_i^{'}+\Delta\right)\left(1-p_i^{'}\right) + p_i^{'}\left(1-\left(p_i^{'}+\Delta\right)\right) \tag{3.48}$$

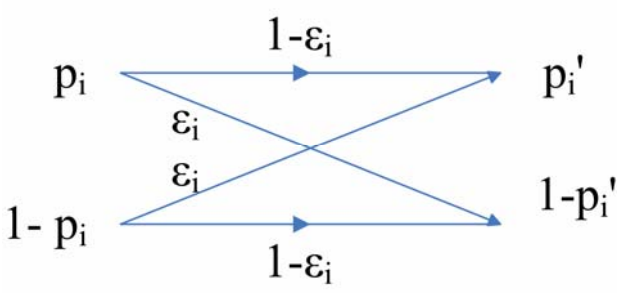


Figure 4.1 The binary symmetric channel with error probability ε_i

When the error probability ε_i approaches to 0.5, $\alpha_{0i}^{'}$ and $\alpha_{1i}^{'}$ approach to 0.5; the difference of mean approaches to minimum (zero); the variance approaches to maximum ($\frac{1}{4}$). If we include this sensor in our detection process, the difference of mean of two distributions will not be increased, but the variances of them will be increased. This would make the detection performance decrease. Hence, we propose to drop sensors with an alarming probability close to 0.5, which is the worst-case choice.

4.2 Adaptive Method

In previous section, we have proposed that we shall drop sensors whose alarming probabilities are near 0.5. However, if we drop too many sensors, the detection performance will be degraded also instead since available number of data samples would be small. We propose the following "adaptive method" for obtaining a satisfactory solution:

Step 1. First, we do not drop any sensors, record the performance, and set zero and unity as the left-point and right-point.

Step 2. Then, we drop sensors with alarming probability $0.5 \pm D$ and record the performance again, where D is zero added with step value *d*. If the latter performs better than the former, we record better performance and increase

D by the same step value. If the latter instead performs worse than the former, jump the process to Step 4.

Step 3. Then, we drop the sensor with alarming probability which distance to 0.5 is less than *new* D and compare its performance with the better performance recorded before. If the new performance is better than the former, we record the better performance and increase D by the same step value again. We repeat Step 3 until the performance is degraded or D is equal to the left-point. If the performance is degraded, we set D as the right-point. We sure *that maximum of performance exists between left-point and right-point.* Else if D is equal to the right-point, we set half of step value as the new one, and set left-point as D, and increase D by new step value, and repeat Step 3.

Step 4. We set half of step value as the new one and decrease D by new step value. Repeat Step 4 until the performance is degraded or D is equal to the left-point. If the performance is degraded, we set D as the right-point, and we sure *that maximum exist between left-point and right-point*, and set half of step value as the new one and decrease D by new step value, and repeat Step 3. Else if D is equal to the left-point, we set half of step value as the new one, and set right-point as D, and decrease D by new step value, and repeat Step 4.

4.3 Computer Simulations

We set the number of sensors to be 3600. We partition sensors in to nine groups. Before change occurs, sensors have identical alarming probability equal to 0.8. The change occurs in two partitions, with alarming probability decreasing to 0.7. The channel model is Rayleigh fading channel. The first step value is 0.02. In Figure 4.2,

SNR is 1dB. In Figure 4.3, SNR is 3dB (10,000 Monte Carlo runs are conducted). From simulations, we can find that as the iteration number increased the detection performance is enhanced.

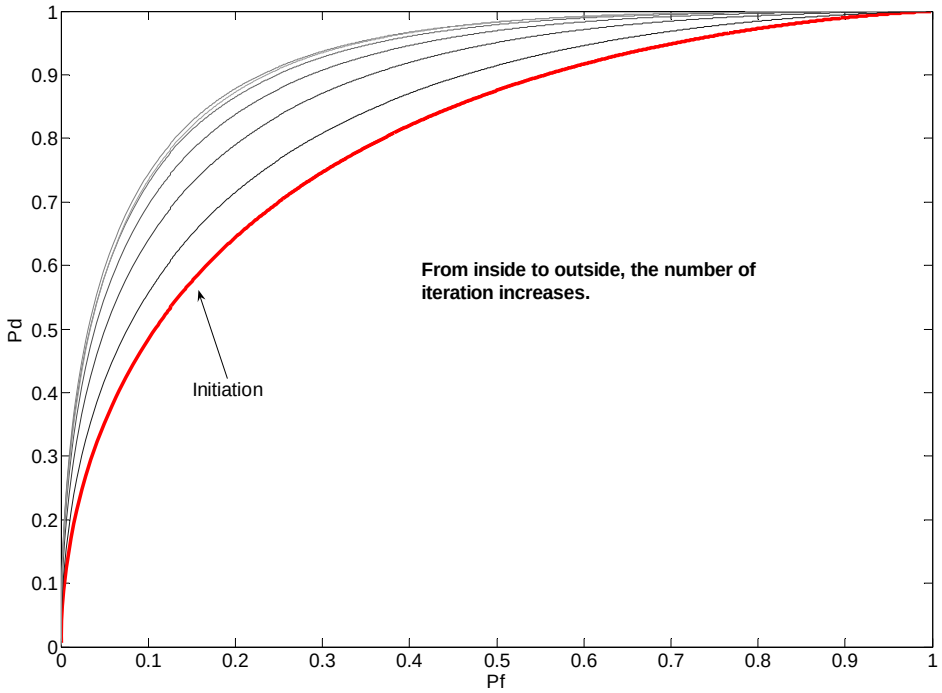


Figure 4.2: Use the adaptive method for BSC with SNR = 1 dB

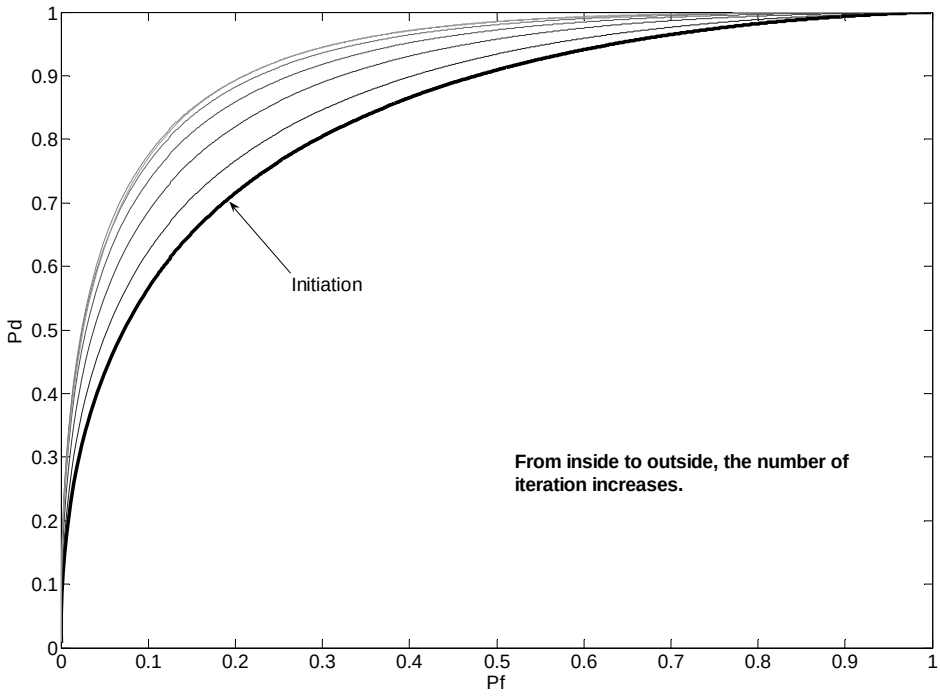


Figure 4.3: Use the adaptive method for BSC with SNR = 3 dB

4.4 Conclusions

In this chapter, we consider the problem of performance degradation since sensor-to-fusion links are no longer error-free. We try to drop some sensors with bad data quality to maximize detection performance. This method we proposed is called adaptive method. From simulations, we find the method is really useful.

Chapter 5

Comparison with Other Methods

In the chapter, we compare our method with [6], which presents several nonparametric change detection and estimation algorithms based on an application of Vapnik-Chervonenkis (VC) theory [7]. The methods in [6] are all based on the $\mathcal{A}$-distance [8] as the measure of distance between two distributions. We will compare our method with three algorithms reported in [6], namely the search in axis-aligned rectangles (SAR) algorithm, the search in axis-aligned strips (SAS) algorithm, and the search in diagonal-defined axis-aligned rectangles (SDR) algorithm, in terms of complexity and the detection performance.

5.1 Comparison of Complexity

In this section, we first compare the complexity of the aforementioned three algorithms. In [6], the authors define the collection $\mathcal{A} \subseteq \mathcal{F}$ of measurable set to model the set of geographical areas that are of practical interest. For example, they may be interested in the number of alarmed sensors in a circle centered at some location with some radius. They are interested in whether there is a change in probability measure on $\mathcal{A}$ and, if there is a change, where the maximum change of probability occurs. The detection problem in [6] is formulated as

$$H_0 : \ P_1(A) = P_2(A) \text{ for all } A \in \mathcal{A}$$

versus

$$H_1 : \ \exists A \in \mathcal{A} \text{ such that } P_1(A) \neq P_2(A).$$

The estimation problem, on the other hand, is to estimate the event $A^* \in \mathcal{A}$ that gives the maximum change. For example, using the $\mathcal{A}$-distance measure

$$A^* = \arg\max_{A \in \mathcal{A}} \left| P_1(A) - P_2(A) \right|.$$

Since the number of the collection of sets $\mathcal{A}$ is possibly infinite, they need to reduce search in $\mathcal{A}$. There are three different search algorithms with different performance and complexity, i.e., the search in axis-aligned rectangles (SAR), the search in axis-aligned strips (SAS), and the search in diagonal-defined axis-aligned rectangles (SDR).

We assume S is the total number of data points in the two collections, T is the number of training samples, and P is the number of partition. In our algorithm, the complexity of training calls for $TS+TP$ additions and P multiplications. The TP additions and P multiplications is from summing average. The complexity of search is $2S+P$ additions. Hence, the complexity of our algorithm is $O(S)$, which is linear in the number of total sensors. In [6], the complexity of SAR is $O(S^3)$, the complexity of SAS is $O(S\log S)$, and the complexity of SDR is $O(S^2)$. Based on these facts our algorithm has lower complexity than those in [6].

5.2 Simulations

In the simulation, the distribution of collected sensors is a mixture of two 2-D uniform distributions, one on a $s \times s$ square $\mathcal{D}$ and the other centered at $\mathbf{x_0} \in \mathcal{D}$ with radius r. Specifically, the probability density function of the 2-D random vector $\mathbf{x}$ is given by

$$p_{\mathbf{x}_0}(\mathbf{x}) = \begin{cases} \dfrac{p}{\pi r^2 p + (s^2 - \pi r^2)q}, x \in D, \| \mathbf{x} - \mathbf{x}_0 \| \leq r \\ \dfrac{q}{\pi r^2 p + (s^2 - \pi r^2)q}, x \in D, \| \mathbf{x} - \mathbf{x}_0 \| > r \\ 0 \qquad\qquad\qquad\qquad , otherwise \end{cases}$$

where $\mathbf{x_0}$, p, q and e are parameters, $0 < r \ll s$ and $0 \leq q < p \leq 1$.

This model corresponds to the scenario when sensors are uniformly distributed in $\mathcal{D}$, and a sensor is alarmed with probability p if it is within distance r from $\mathbf{x_0} \in \mathcal{D}$ and q if it falls outside this distance. We consider the ideal case that data transmission is errorless; hence 1-p is the (uniform) miss detection probability and q is the (uniform) false alarm probability at sensors.

Under hypothesis $\mathcal{H}_0$, two sets of sample points are drawn i.i.d. from the same; under $\mathcal{H}_1$, one set of sample points are drawn from $p_{\mathbf{x}_0}$, and the other set of sample points are drawn independently from $p_{\mathbf{x}_0'}$ for some other center $\mathbf{x}_0'$.

We set s to be 180, $p = 0.98$, $q = 0.02$, and $r = s/12$. We partition sensors in to 100 groups, and the number of training is 100. In Figure 5.1, $\mathbf{x}_0 = (50, 60)$ and $\mathbf{x}_0' = (50, 90)$. In Figure 5.2, $\mathbf{x}_0 = (50, 60)$ and $\mathbf{x}_0' = (50, 61)$ (10,000 Monte Carlo runs are conducted).

In the first case, it is relatively easy to detect whether the change occurs or not. The propose method is comparable to Tong's three algorithms. For the second case, in which the locations of the sets of alarming sensors before and after change occurs almost coincide, the proposed approach significantly outperforms Tong's algorithms. Such a performance advantage comes from (i) our method performs the training process to obtain some knowledge of the distribution of change, (ii) our method resorts to region partition, which can better reflect the actual tendency in distribution caused

by the change effect.

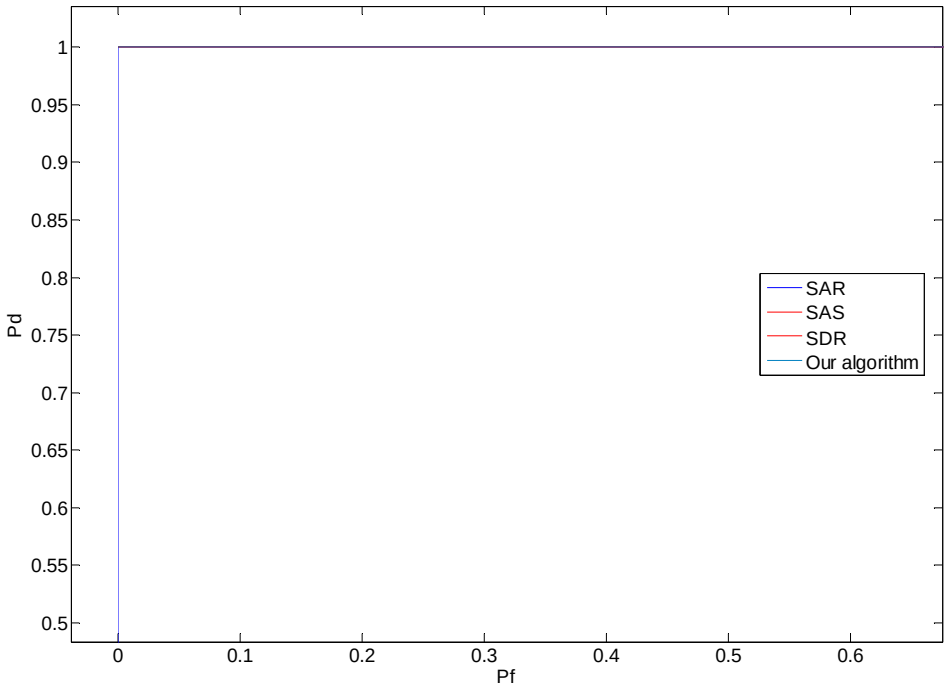


Figure 5.1: Receiver operating characteristics for $\mathbf{x_0} = (50, 60)$ and $\mathbf{x_0'} = (50, 90)$

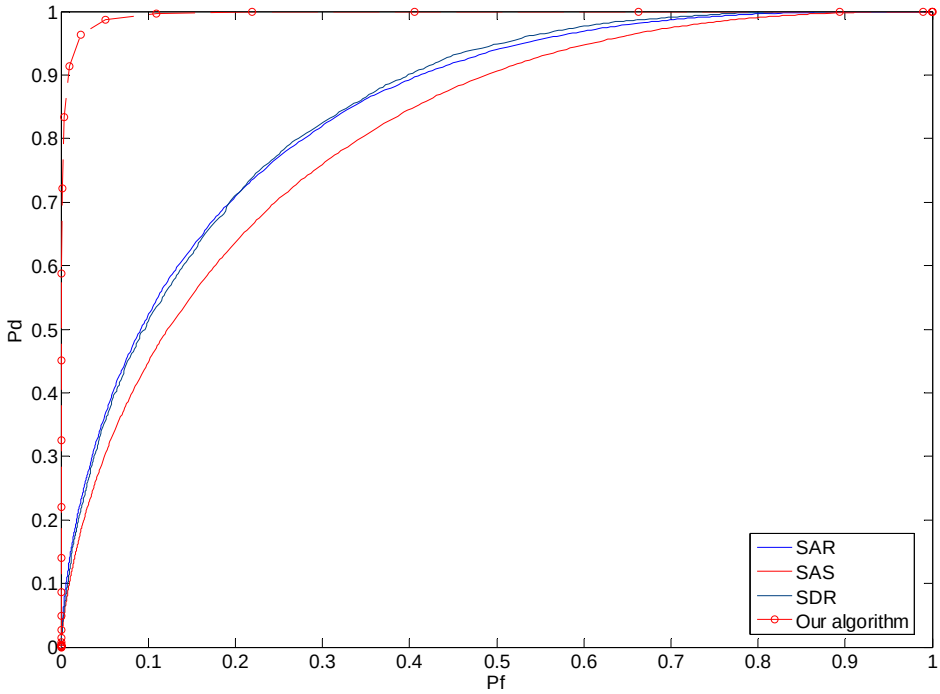


Figure 5.2: Receiver operating characteristics for $\mathbf{x_0} = (50, 60)$ and $\mathbf{x_0'} = (50, 61)$

5.3 Summary

Our algorithm has lower complexity (linear in the total number of sensors) than the three algorithms proposed by Tong in [6]. Simulation results also show that our algorithm exhibits better change detection capability in some critical cases where the locations of the sets of alarming sensors before and after change occurs almost coincide. Although our algorithm has lower complexity and good performance than others, it only applies to the conditions allow training. If training is not allowed, our algorithm does not work.

Chapter 6

Conclusions

In this thesis, we have presented a nonparametric approach to the detection of changes in the distribution of alarmed sensors. The proposed approach relies on partitioning sensors into several groups in the sensing area. Then, we estimate via data training the mean of the related distributions to perform change detection. We exactly study the performance in terms of missing and false alarm probabilities in our approach. Compared with existing methods, our algorithm not only has a low complexity (O(S)) but also yields improved detection performance. We also provide a lower bound of the number of training sample; this serves as a useful guideline for determining the number of training required for guaranteeing prescribed estimation accuracy. We also study the performance enhancement problem when the sensors-to-fusion links are non-ideal and are modeled as BSC. We propose an approach to find a global/local maximum of detection probability under the same false alarming probability. We got performance enhancement in practice by our method.

Bibliography

[1] I.F. Akyildiz et al., ”A Survey on Sensor Networks,” *IEEE Communication Magzine,* pp.102-114, August 2002

[2] C. Intanagonwiwat, R. Govindan, and D. Estrin, “Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks,” *Proc. ACM MobiCom’00*, Boston, MA, 2000, pp. 56–67.

[3] N. Patwari, A. O. Hero, and B. M. Sadler, “Hierarchical censoring sensors for change detection,” in *Proc. 2003 IEEE Workshop on Statistical Signal Processing*, St. Louis, MO, Sep. 2003, pp. 21–24.

[4] Y. Hong and A. Scaglione, “Distributed change detection in large scale sensor networks through the synchronization of pulse-coupled oscillators,” in *Proc. Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Montreal, QC, Canada, May 2004, pp. 869–872.

[5] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. New York: Marcel Dekker, 2003.

[6] T. He and L. Tong, “Nonparametric change detection and estimation in large-scale sensor network ,” *IEEE Trans. Signal Processing*, vol. 54, no. 4, pp. 1204–1217, Apr. 2006

[7] L. Devroye et al., *A probabilistic theory of pattern recognition*, New York: Springer, 1996

[8] T. He and L. Tong, (2004, Aug.) On A-distance and relative A-distance. Tech. Rep. ACSP-TR-08-04-02. [Online]. Available: http://acs.ece.cornell.edu/pubR.html

[9] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*. New York: Wiley Interscience, 1973.

[10] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 3rd ed. London, U.K.: Chapman & Hall/CRC, 2004.

[11] S. M. Kay, *Fundamentals statistical signal processing estimation theory*, Vol.2, New Jersey: Prentice Hall, 1993.

[12] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 1991.

[13] H. Stark and J. W. Woods, *Probability and random processes with applications to signal processing*, 3rd Ed. New Jersey: Prentice Hall, 2002.

[14]N. V. Poor, *An introduction to signal detection and estimation,* 2nd Ed. New York: Springer-Verlag, 1994.

[15] S. M. Kay, *Fundamentals statistical signal processing estimation theory*, Vol.1, New Jersey: Prentice Hall, 1993.

[16] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, pp. 388–404, Mar. 2000.

[17] S. Ben-David, J. Gehrke, and D. Kifer, "Detecting change in data streams," presented at the 2004 Very Large Data Bases (VLDB) Conf., Toronto, ON, Canada, Aug. 29–Sep. 3, 2004.

[18] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.

[19] M. Pollak, "Optimal detection of a change in distribution," *Ann. Statist.*, vol. 13, no. 1, pp. 206–227, 1985.

[20] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*. Englewood Cliffs: Prentice-Hall, 1993.

[21] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1996.

[22] V. V. Veeravalli, "Sequential decision fusion: Theory and applications," *J. Franklin Inst.*, vol. 336, pp. 301–322, Feb. 1999.

[23] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag, 1985.

[24] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, 1986.

[25] M. K. Simon and M-S Alouini, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*, John Wiley & Sons, 2000.

[26] V. V. Veeravalli, "*Decentralized Quickest Change Detection*," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp.1657-1668, May 2001.