NONPARAMETRIC CHANGE DETECTION IN 2D RANDOM SENSOR FIELD

Ting He, Shai Ben-David, and Lang Tong

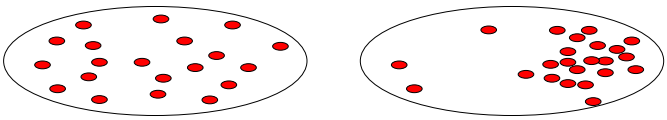
ABSTRACT

The problem of detecting changes from data collected from a large-scale randomly deployed two dimensional sensor field is considered. Under a nonparametric change detection framework, we propose detection algorithms using two measures of change. Theoretical performance guarantee is derived from the Vapnik-Chervonenkis theory. By exploiting the structures of the search domain, we design a suboptimal recursive algorithm to detect the area of largest change which, for M sample points, runs in time $O(M^2 \log M)$ (compared to an $O(M^4)$ required for a straightforward exhaustive search). The lost of performance diminishes as M increases.

Keywords: Non-parametric change detection, Sensor Networks, Detection and estimation algorithms.

1. INTRODUCTION

We are interested in detecting certain phenomenal change in a large-scale randomly deployed sensor field. For example, sensors may be designed to detect certain chemical components. When the sensor measurement exceeds certain threshold, the sensor is "alarmed". The state of a sensor depends on where it resides; sensors in some area are more likely to be in the alarmed state than others. We are not interested in the event that certain sensors are alarmed. We are interested instead in whether there is a change in the geographical distribution of alarmed sensors from data collections at two different time. Such a change in distribution could be an indication of abnormality.



First data collection Second data collection

Fig. 1. Reported alarmed sensors (red) in two collections.

We assume that some (not necessarily all) of the alarmed sensors are reported to a fusion center, either through the use a mobile access point (SENMA [1]) or using certain in-network routing scheme. As illustrated in Fig. 1, suppose that the fusion center obtains two reports of the locations of alarmed sensors from two separate data collections. In the ith report, let S_i be the set of alarmed sensor drawn independently according to some distribution P_i. The identities and the number of alarmed sensors in the two reports may not match. The change detection problem considered in this paper is one of nonparametric hypotheses testing:

$$\mathcal{H}_0 : P_1 = P_2 \quad \text{vs.} \quad \mathcal{H}_1 : P_1 \neq P_2.$$

where we make no prior assumptions about the data generating distributions P_i. We may also be interested in locating areas with significant changes.

Related Work The problem of change detection in sensor field has been considered in different (mostly parametric) settings [2,3]. The underlying statistical problem considered in this paper belongs to the category of two-sample nonparametric change detection. There is a wealth of nonparametric change detection techniques for one-dimensional data set in which data can be ordered uniquely, see [4] and references therein. Most of these techniques do not have natural generalizations to the two dimensional sensor network applications. The classical Kolmogorov-Smirnov two-sample test [4] does apply to the two-dimension sensor problem, but it does not provide the location where changes may occur. In a way, the methods presented in this paper generalize the idea of Kolmogorov-Smirnov test to a different collection of measurable sets using more general forms of distance measures, which allows the formulation of change estimation problem [5].

Summary of Results and Organizations In this paper we present a nonparametric change detection and estimation algorithm based on an application of Vapnik-Chervonenkis Theory [6]. The basis of this approach is outlined in [5] where we have provided a mathematical characterization of changes in distribution. Our focus in this paper is on the use of a relative measure of change and a low complexity algorithm applicable to large scale sensor networks.

We begin a formulation of the change detection problem for sensor networks in Section 2. We then present in Section 3 results that establish a theoretical guarantee of performance. We consider two distance measures in this paper. The first is the so-called $\mathcal{A}$-*distance* (also used in [5]) that measures the maximum change in probability among a collection $\mathcal{A}$ of measurable sets. The second is a relative distance measure—a variation from that in [5]—for cases when the change in probability is concentrated in areas of small probability weight.

Next we derive in Section 4 a practical algorithm applicable to large scale sensor networks. The key to the applicablity of VC Theory is reducing the search in a possibly uncountable collection $\mathcal{A}$ of sets (*e.g.,* planer disks) to one in a *finite* sub-collection $\mathcal{H}(S)$ (a function of the observations $S = S_1 \bigcup S_2$) without affecting the performance. If $M = |S_1 \cup S_2|$ is the total number of data points in two collections, in [5] we have shown that the exhaustive search among the collection of all planer disks has complexity

T. He and L. Tong ({th255,lt35}@cornell.edu) are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. S. Ben-David (shai@cs.waterloo.ca) is with the School of Computer Science, University of Waterloo, Waterloo, Canada.

This work is supported by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

$O(M^4)$. We present a suboptimal search strategy that has complexity $O(M^2 \log M)$. There is a loss of performance, however, but such a loss diminishes as the number of samples increases.

Simulation results are provided in Section 5 in which we compare the performance when two different distance metrics are used.

2. THE PROBLEM STATEMENT

We consider two probability measures P_1 and P_2 on the same measurable space X where $(X, \mathcal{F}, P_i)$ models the ith random collection of alarmed sensors[1]. Denote S_i as the set of locations of alarmed sensors in the ith collection and $S = S_1 \bigcup S_2$. We assume that, in each collection, alarmed sensors are drawn i.i.d. according to P_i and the two drawings are independent. Note that the number and the identities of the collected alarmed sensors may be different in each collection.

We introduce a collection $\mathcal{A} \subseteq \mathcal{F}$ of measurable sets to model the set of geographical areas in which, for practical purposes, events of interest are observed from the sensor field. The collection $\mathcal{A}$ does not have to be countable and is part of the algorithm design. For example, we may be interested in the number of alarmed sensors in a circle centered at some location s in X with some radius r. If we define $\mathcal{A}$ as the collection of all disks in X, for some applications, it may be reasonable to focus on the probabilities of events in $\mathcal{A}$ rather than those in $\mathcal{F}$ of the original probability space. The choice of $\mathcal{A}$ is subjective, and it depends on the application at hand. Other geometrical shapes such as rectangles, and stripes are considered in [7].

Given the probability spaces $(X, \mathcal{F}, P_i)$ and the collection $\mathcal{A} \subseteq \mathcal{F}$, we are interested in whether there is a change in probability measure on $\mathcal{A}$.

3. PERFORMANCE GUARANTEE

To describe "change", we need some notion of distance between two probability distributions. In this paper, we will be consider two distance measures.

$\mathcal{A}$-distance and empirical $\mathcal{A}$-distance [5] Given probability spaces $(X, \mathcal{F}, P_i)$ and a collection $\mathcal{A} \subseteq \mathcal{F}$, the *$\mathcal{A}$-distance* between P_1 and P_2 is defined as

$$d_{\mathcal{A}}(P_1, P_2) = \sup_{A \in \mathcal{A}} |P_1(A) - P_2(A)|. \tag{1}$$

The *empirical $\mathcal{A}$-distance* $d_{\mathcal{A}}(S_1, S_2)$ is similarly defined by replacing $P_i(A)$ by the empirical measure

$$S_i(A) \triangleq \frac{|S_i \bigcap A|}{|S_i|} \tag{2}$$

where $|S_i \cap A|$ is the number of points in both S_i and set A.

The $\mathcal{A}$-distance does not take into account the relative significance of the change. For example, one could argue that changing the probability of a set from 0.99 to 0.999 is less significant than a change from 0.001 to 0.01; the latter amounts to a ten-fold increase whereas the former represents an increase of about 1%. For applications in which small probability events are of interests, we introduce the following notion of *relative $\mathcal{A}$-distance* that takes the relative magnitudes of a change into account.

[1]The notation $(X, \mathcal{F}, P_i)$ is standard: X is the space, $\mathcal{F}$ the σ-algebra, P_i the probability measure.

Relative and Empirical Relative $\mathcal{A}$-distance Given probability spaces $(X, \mathcal{F}, P_i)$ and a collection $\mathcal{A} \subseteq \mathcal{F}$, the *relative $\mathcal{A}$-distance* between P_1 and P_2 is defined as

$$\phi_{\mathcal{A}}(P_1, P_2) = \sup_{A \in \mathcal{A}} \frac{|P_1(A) - P_2(A)|}{\sqrt{\frac{P_1(A)+P_2(A)}{2}}}. \tag{3}$$

The *empirical relative $\mathcal{A}$-distance* is defined similarly by replacing $P_i(A)$ with the empirical measure defined in (2).

The above definition is slightly different from that used in [5]. The proof that the above relative $\mathcal{A}$-distance is indeed a metric follows [8]. With a properly chosen distance measure, we can now specify the class of detectors considered in this paper.

Detector $\delta(S_1, S_2; \epsilon)$: Given two collection of samples S_1 and S_2, drawn i.i.d from probability distributions P_1 and P_2 respectively, and threshold $\epsilon \in (0, 1)$, for hypotheses $\mathcal{H}_0$ vs. $\mathcal{H}_1$, the detector[2] using the $\mathcal{A}$-distance is defined as

$$\delta_{d_{\mathcal{A}}}(S_1, S_2; \epsilon) = \begin{cases} 1 & \text{if } d_{\mathcal{A}}(S_1, S_2) > \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The detector $\delta_{\phi_{\mathcal{A}}}(S_1, S_2; \epsilon)$ using the relative $\mathcal{A}$-distance is defined the same way by replacing $d_{\mathcal{A}}(S_1, S_2)$ by $\phi_{\mathcal{A}}(S_1, S_2)$.

We now provide a theoretical guarantee of the performance by deriving bounds on the miss detection and the false alarm probabilities, both in near exponential form.

Theorem 3.1 *Given probability spaces $(X, \mathcal{F}, P_i)$ and a collection $\mathcal{A} \subseteq \mathcal{F}$, let $S_i \in X$ be a set of n samples drawn according to P_i. The false alarm probabilities for the detectors defined in (4) are bounded by*

$$P_F(\delta_{d_{\mathcal{A}}}) \leq 8(2n+1)^d e^{-n\epsilon^2/32} \tag{5}$$

$$P_F(\delta_{\phi_{\mathcal{A}}}) \leq 2(2n+1)^d e^{-n\epsilon^2/4} \tag{6}$$

where d is the VC-dimension[3] of $\mathcal{A}$.

Furthermore, if $d_{\mathcal{A}}(P_1, P_2) > \epsilon$ and $\phi_{\mathcal{A}}(P_1, P_2) > \epsilon$, the miss detection probabilities satisfy, respectively,

$$P_M(\delta_{d_{\mathcal{A}}}, P_1, P_2) \leq 8(2n+1)^d e^{-n[d_{\mathcal{A}}(P_1,P_2)-\epsilon]^2/32} \tag{7}$$

$$P_M(\delta_{\phi_{\mathcal{A}}}, P_1, P_2) \leq 16(2n+1)^d e^{-n[\phi_{\mathcal{A}}(P_1,P_2)-\epsilon]^2/16} \tag{8}$$

Proof: See Appendix. ■

Note first the decay rates of the error probabilities are different when the two different distance measures are used, with the relative distance measure providing some gain in the decay rate of error probability. Next, the above theorem also provides a way of deciding the detection threshold ϵ for a particular detection criterion. For example, the threshold of the Neyman-Pearson detection for a given size α can be obtained from the bounds on false alarm probabilities. Finally, notice the role of VC dimension d in the error exponent; it appears that d, as a coefficient of $\ln n$, is a minor factor compared with the distance between two probability measures.

[2]We use the convention that the detector gives the value 1 for $\mathcal{H}_1$ and 0 for $\mathcal{H}_0$.

[3]The VC-dimension of $\mathcal{A}$ is defined as the cardinality of the largest set S shattered by $\mathcal{A}$ [9].

4. A LOW COMPLEXITY DETECTOR

The detectors given in (4) are based on the search for maximum distance between empirical probabilities on the (uncountable) collection of measurable sets $\mathcal{A}$. To implement such a detector, it is necessary to reduce, based on the observation data S, the search in $\mathcal{A}$ to a search in a finite collection $\mathcal{H}(S)$. The choice of $\mathcal{A}$ and that of $\mathcal{H}(S)$ dictate the complexity and the performance.

For the collection of disks[4] considered in this paper, it can be shown that the exhaustive search algorithm has a cost in the order of $O(M^4)$ where $M = |S|$ is the total number of data samples collected. This complexity is still prohibitive for large scale networks. The algorithm presented next reduces the complexity to $O(M^2 \log M)$.

Let $\mathcal{A}$ be the collection of two dimensional disks. Given the samples $S = S_1 \bigcup S_2$, consider the finite collection of sample-centered disks $\mathcal{H}_{\text{CD}}(S) \subset \mathcal{A}$ defined by

$$\mathcal{H}_{\text{CD}}(S) \stackrel{\Delta}{=} \{D'(s_i, s_j) : s_i, s_j \in S\} \tag{9}$$

where $D'(s_i, s_j)$ is the disk with s_i at the center and s_j on the boundary. Note that this collection of disks contains M sub-collections, each has one sample at the center and a nested disks defined by the rest of samples in the disk. This particular structure allows the development of the following recursive algorithm that calculates the distance between two empirical probability distributions.

The recursive search for maximum change goes as follows. Fix an s_i and define

$$F_i(j) \quad \stackrel{\Delta}{=} \quad S_1(D'(s_i, s_j)) - S_2(D'(s_i, s_j)) \tag{10}$$

$F_i(j)$ is the change in the empirical probability of $D'(s_i, s_j)$. First sort the sample points into increasing order $s_{j_1}, s_{j_2}, \ldots$ according to their distance to s_i[5], and then computing $F_i(j_k)$ recursively by the update

$$F_i(j_k) = \begin{cases} F_i(j_{k-1}) + \frac{1}{|S_1|} & \text{if } s_{j_k} \in S_1 \\ F_i(j_{k-1}) - \frac{1}{|S_2|} & \text{if } s_{j_k} \in S_2 \end{cases}$$

where $k = 1, 2, \ldots, M-1$. Next we compute

$$j^*(i) \quad = \quad \arg\max_j |F_i(j)|. \tag{11}$$

The optimal disk in $\mathcal{H}_{\text{CD}}$, for fixed center s_i, is given by $D'(s_i, s_{j^*(i)})$, and the maximum difference for disks at center s_i is $|F_i(j^*(i))|$. The search repeats for all possible s_i. Finally, we find the maximum among $|F_i(j^*(i))|, \forall i$, i.e.

$$i_{\max} \quad = \quad \arg\max_i |F_i(j^*(i))|. \tag{12}$$

Then the optimal disk in $\mathcal{H}_{\text{CD}}$ is given by $D'(s_{i_{\max}}, s_{j^*(i_{\max})})$, and the maximum difference is $|F_{i_{\max}}(j^*(i_{\max}))|$.

If $\phi_{\mathcal{A}}(S_1, S_2)$ is needed, it is easy to see that, following the same steps as computing $F_i(j_k)$, the empirical probabilities $S_k(D'(s_i, s_j))$, $(k = 1, 2)$ can be computed, so can $\phi_{\mathcal{A}}(S_1, S_2)$.

The complexity of the proposed algorithm, compared with exhaustive search, is reduced to $O(M^2 \log M)$. The dominating term is the sorting of the sample points according to their distances to a certain sample point, which takes $O(M \log M)$ for each center, and is repeated for M centers. Furthermore, this algorithm can also be modified for other distance metrics that are based on the counting of samples.

It should be noted that the reduction of complexity does come with a cost. The specific choice of $\mathcal{H}_{\text{CD}}$ does not guarantee optimality because $\mathcal{H}_{\text{CD}}$ is not complete [5] with respect to $\mathcal{A}$ in the sense that some type of disks are missing. If, however, probability measures P_i are such that that any disk with a positive area has positive probability, then the loss of performance vanishes asymptotically. Specifically, consider a disk and an arbitrary neighborhood of its center, the strong law of large numbers guarantees that as sample size goes to infinity, there is a sample within this neighborhood of the center almost surely. This implies that the proposed algorithm is asymptotically optimal for search in planar disks.

[4] Other choices are considered in [7].

[5] This sort is at the cost of $O(M \log M)$.

5. SIMULATION

We simulate the case when the distribution of alarmed sensors is a mixture of 2D uniform distributions, one on a $s \times s$ square $\mathcal{D}$ and the other centered at $\mathbf{x}_0 \in \mathbb{R}^2$ with radius r. Specifically, the PDF of the 2D random vector $\mathbf{x}$ is given by

$$p_{\mathbf{x}_0}(\mathbf{x}) \quad = \quad \begin{cases} \frac{p}{\pi r^2 p + (s^2 - \pi r^2) q} & \mathbf{x} \in \mathcal{D}, ||\mathbf{x} - \mathbf{x}_0|| \le r \\ \frac{q}{\pi r^2 p + (s^2 - \pi r^2) q} & \mathbf{x} \in \mathcal{D}, ||\mathbf{x} - \mathbf{x}_0|| > r \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{x}_0$, p, q, and r are parameters, $0 < r << s$ and $0 \le q < p \le 1$. We are interested in detecting whether there is a change in the center of the distribution. The algorithm, of course, is not given the form of P.

We consider a Neyman-Pearson setup in which the size α of the detector is prescribed. From Theorem 3.1, we can choose (n, ϵ) such that

$$\epsilon \ge \sqrt{\frac{32}{n} \log \frac{8(2n+1)^d}{\alpha}} \quad \text{for } \delta_{d_{\mathcal{A}}} \tag{13}$$

$$\epsilon \ge \sqrt{\frac{4}{n} \log \frac{2(2n+1)^d}{\alpha}} \quad \text{for } \delta_{\phi_{\mathcal{A}}}. \tag{14}$$

The decision threshold $\epsilon(n)$ is a measure of detector sensitivity. For a fixed detector size, the smaller the $\epsilon(n)$, the higher the detection power. To this end, the relative distance metric provides approximately three times improvement in detection sensitivity. Fig 2 shows the miss detection probability as a function of sample size. Here we observe a sharp drop in miss detection probability, indicating a threshold value on the number of samples required for the exponential decay of miss detection error. Such a phenomenon is indeed predicted by Theorem 3.1 in which there are minimum sample size n_d^* for $\delta_{d_{\mathcal{A}}}$ and n_ϕ^* for $\delta_{\phi_{\mathcal{A}}}$ such that, from the exponents of the miss detection probability (7-8)

$$\epsilon(n_d^*) < d_{\mathcal{A}}(P_1, P_2), \;\; \epsilon(n_\phi^*) < \phi_{\mathcal{A}}(P_1, P_2).$$

It turns out that n_d^* and n_ϕ^* calculated above match well with the threshold in the simulation shown in Fig 2.

6. CONCLUSION

We presented a nonparametric approach to change detection in a 2D random field. As a by product, the detection algorithm also

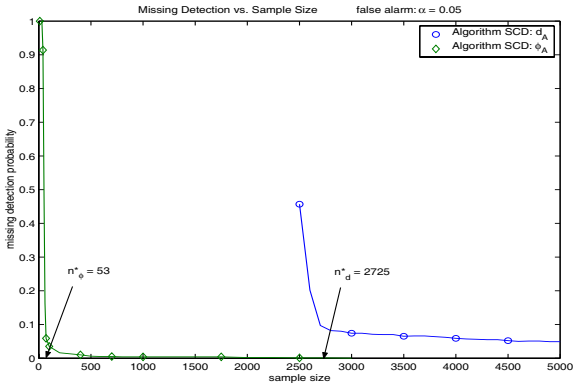


Fig. 2. Miss detection probability as a function of the sample size (simulation results). Here $p = 0.98$, $q = 0.02$, $r = s/12$.

gives an estimate of the location of changes. We provided a theoretical characterization of the miss detection and false alarm probability in the form of exponential decay of error probabilities. We also presented a suboptimal recursive algorithm that reduces the complexity from $O(M^4)$ to $O(M^2 \log M)$. Such a reduction allows the use of of a large number of samples to improve the detection performance.

Words of caution are now in order. Bounds on error probabilities derived here are usually not tight, and the number of samples required for the derived bounds may need to be very large. Such shortcomings come with the advantage of not requiring prior knowledge of the probability distributions.

APPENDIX: PROOF OF THEOREM 3.1

We first prove the theorem for detectors using the $\mathcal{A}$-distance metric $d_{\mathcal{A}}(S_1, S_2) = \sup_{A \in \mathcal{A}} |S_1(A) - S_2(A)|$. From [10], we have

$$\Pr\{\exists A \in \mathcal{A}, ||P_1(A) - P_2(A)| - |S_1(A) - S_2(A)|| > \epsilon\} \leq 8(2n+1)^d e^{-n\epsilon^2/32} \quad (15)$$

Under H_0, $P_1 = P_2$, and the false alarm probability satisfies

$$\begin{aligned} P_F(\delta) &= \Pr\{d_{\mathcal{A}}(S_1, S_2) > \epsilon; \mathcal{H}_0\} \\ &= \Pr\{\exists A \in \mathcal{A}, |S_1(A) - S_2(A)| > \epsilon; \mathcal{H}_0\} \\ &= \Pr\{\exists A \in \mathcal{A}, ||P_1(A) - P_2(A)| \\ &\quad -|S_1(A) - S_2(A)|| > \epsilon; \mathcal{H}_0\} \\ &\leq 8(2n+1)^d e^{-n\epsilon^2/32} \end{aligned} \quad (16)$$

where inequality (16) follows from (15).

For the miss probability, let $A^* = \arg\sup_{A \in \mathcal{A}} |P_1(A) - P_2(A)|$.

$$\begin{aligned} P_M(\delta, P_1, P_2) &= \Pr\{d_{\mathcal{A}}(S_1, S_2) \leq \epsilon; P_1, P_2\} \\ &\leq \Pr\{|S_1(A^*) - S_2(A^*)| \leq \epsilon; P_1, P_2\} \\ &\leq \Pr\{||P_1(A^*) - P_2(A^*)| \\ &\quad -|S_1(A^*) - S_2(A^*)|| \\ &\quad \geq ||P_1(A^*) - P_2(A^*)| - \epsilon|; P_1, P_2\} \\ &\leq 8(2n+1)^d e^{-n[|P_1(A^*) - P_2(A^*)| - \epsilon]^2/32} \end{aligned} \quad (17)$$

Now consider relative distance. The proof for relative distance metric goes line by line as that for the non-relative metric, replacing inequality (15) with the following results from [10],

$$P^{2n}(\phi_{\mathcal{A}}(S_1, S_2) > \epsilon) \leq 2(2n+1)^d e^{-n\epsilon^2/4} \quad (18)$$

$$\begin{aligned} &P^{2n}[|\phi_{\mathcal{A}}(P_1, P_2) - \phi_{\mathcal{A}}(S_1, S_2)| > \epsilon] \\ &\leq 16(2n+1)^d e^{-n\epsilon^2/16} \end{aligned} \quad (19)$$

We have

$$P_F(\delta) \leq 2(2n+1)^d e^{-n\epsilon^2/4} \quad (20)$$

$$P_M(\delta, P_1, P_2) \leq 16(2n+1)^d e^{-n[\phi_{\mathcal{A}}(P_1, P_2) - \epsilon]^2/16} \quad (21)$$

7. REFERENCES

[1] L. Tong, Q. Zhao, and S. Adireddy, "Sensor Networks with Mobile Agents," in *Proc. 2003 Intl. Symp. Military Communications*, (Boston, MA), Oct. 2003.

[2] N. Patwari, A. O. Hero, and B. M. Sadler, "Hierarchical censoring sensors for change detection," in *2003 IEEE Workshop on Statistical Signal Processing*, (St. Louis, MO), pp. 21–24, September 2003.

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

[4] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman & Hall/CRC, 2004.

[5] S. Ben-David, T. He, and L. Tong, "Non-Parametric Approach to Change Detection and Estimation in Large Scale Sensor Networks," in *Proceedings of the 2004 Conference on Information Sciences and Systems*, (Princeton, NJ), March 2004.

[6] V. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencie of events to their probabilities," *Theory of Probability and its Applications*, vol. 16, pp. 264–280, 1971.

[7] T. He, S. Ben-David, and L. Tong, "Nonparametric Change Detection and Estimation in Large Scale Sensor Networks: Algorithms," Sept 2004. preprint.

[8] S. Ben-David, J. Gehrke, and D. Kifer, "Detecting Change in Data Streams," in *Proc. 2004 VLDB Conference*, (Toronto, Canada), 2004.

[9] M. J. Kearns and U. V. Vazirani, *An Introduction to Computational Learning Theory*. Cambridge, MA: MIT Press., 1994.

[10] T. He and L. Tong, "On $\mathcal{A}$-distance and Relative $\mathcal{A}$-distance," Tech. Rep. ACSP-TR-08-04-02, Cornell University, August 2004. `http://acsp.ece.cornell.edu/pubR.html`.