ADAPTIVE STATISTICAL SAMPLING METHODS FOR DECENTRALIZED ESTIMATION AND DETECTION OF LOCALIZED PHENOMENA

Erhan Baki Ermis, Venkatesh Saligrama

Boston University

1. INTRODUCTION

The design and deployment of sensor networks (SNETs) for decision making pose fundamental challenges due to energy constraints and uncertain environments. In this paper we focus on one such problem where minimization of communication costs due to information exchange is required subject to end to end information quality constraints. Specifically, we develop solutions for detection of distributed events, sources, or abnormalities that are localized, i.e., only a small number of sensors in the vicinity of the phenomena are in the field of observation. This problem complements the standard decentralized detection problem, where noisy information about a global event is measured by the entire network. The global phenomena by itself can be one of several different discrete possibilities and researchers have investigated several architectures within this context.

Our objective in this paper is to characterize the fundamental trade offs between global performance (false alarms and miss rate) and communication cost. We develop a framework to minimize the communication cost subject to worst-case misclassification constraints by making use of the false discovery rate (FDR) concept along with an optimal local measure transformation at each sensor node. The preliminary results show that the FDR concept applied in a sensor networks context leads to significant reduction in the communication cost of the system.

2. DISCUSSION

The general problem described here is related to the so called multiple comparison tests (MCPs) in the statistical literature [1] as well as the bio-statistical communities. The setup consists of a collection of sample observations, y, each drawn either with the probability distribution, $f_Y(y/H_0)$, which corresponds to the null hypothesis, H_0, or with $f_Y(y/H_1)$, which corresponds to the positive hypothesis. The problem is to partition the samples in to two bins corresponding to null and positive hypothesis respectively. A general partition can be associated with a decision rule for each sensor node, k, mapping all the observations into the two different hypothesis, i.e.,

$$u_k : y_1^N \mapsto \{0,\, 1\},\ k = 1,\, 2,\, \ldots,\, m \tag{1}$$

The table below summarizes the quantities involved in the MCPs.

	Declared H_0	Declared H_1	Total
True H_0	U	V	m_0
True H_1	T	S	$m - m_0$
Total	$m - R$	R	N

As a possible solution to such problems, Benjamini and Hochberg introduced the FDR procedure [1]. This concept, instead of trying to control the probability of making any false alarms, controls the expected ratio of the number of observations falsely declared as significant, (V), to the total number of observations declared as significant, (R), i.e.,

$$FDR = E(Q) = E\{V/V + S\} = E\{V/R\}$$

It is easy to establish that the false alarm rate [1] is bounded from below by FDR, i.e.,

$$FWER = \text{Prob}\{V \geq 1\} \geq E(Q) = FDR$$

Although it is a weaker notion in terms of false alarm probability, the significant increase in the power of detection makes it a desirable approach in many problems.

FDR procedure is described as follows:

1. Calculate the p values for all the observations
2. Order the p values in ascending order
3. Find the largest index, i_{max}, such that $p^{(i)} \leq \frac{i}{m}\gamma$
4. Declare $p^{(j)}$ significant for $0 \leq j \leq i_{max}$

The following definition of p value of a random variable X is used in this paper:

$$p(X) = \int_X^\infty f_0(t)dt = 1 - F_0(X) \tag{2}$$

where f_0 is the pdf of the observations under H_0, and γ is the FDR constraint. It is obvious that p value of a random variable is also a random variable. Throughout the paper, the p value of the random variable X_0, where X_0 comes from H_0, is denoted by P_0, and similarly for P_1.

This research was supported by ONR Young Investigator Award N00014-02-100362, Presidential Early Career Award (PECASE)

Lemma 2.1 *The random variable P_0 is distributed uniformly in (0,1) regardless of the distribution of X_0.*

Theorem 2.2 *For independent test statistics under null hypothesis, and for any configuration of positive hypotheses, the above procedure controls the FDR at level γ.*

The reader is referred to the original work [1] for a better understanding of the problem context of this approach.

Lemma 2.1 and theorem 2.2 lead to the following proposition.

Proposition 2.3 *False discovery rate constraint is satisfied for FDR procedures applied along with transformations that are measure invariant with respect to the distribution under H_0.*

The FDR procedure suffers from two significant drawbacks that make it unsuitable in our applications:

- The focus of this procedure primarily is to reduce false positives and there is no control over the false negatives.
- The procedure does not lend itself easily to decentralized implementation.

The main contributions in this paper are two fold:

- We develop a new FDR type procedure that increases power of detection through a convenient measure preserving transformation.
- The new procedure lends itself to efficient decentralization.

3. ADAPTIVE VS NON-ADAPTIVE DECISION RULES

An adaptive decision rule allows some form of collaboration between the sensor nodes and requires introduction of time and a "not yet decided state". For this purpose, we allow each sensor to take on three values $u_k(\cdot) \in \{\phi, 0, 1\}$, where ϕ corresponds to undecided state. In the beginning all the sensors are undecided. At time $t+1$, an undecided sensor, k, updates its decision based on its local observation, y_k, and all the messages, $\mathcal{U}^t$ received (note that only positive messages (i.e., equal to 1) are received) from other sensor nodes up to time period t, i.e.,

$$u_k(t+1) : Y_k \times \mathcal{U}^t \mapsto \{\phi, 0, 1\}$$

We are now ready to formalize our problem. The objective is to minimize the number of expected misses subject to an FDR and communication constraint.

$$\text{minimize } E(T) \text{ subject to:}$$
$$E(V/R) \leq \gamma, \ \sum_{i,t} c(u_i(0)...u_i(t)) \leq \alpha$$

where c is the cost of communicating a decision:

$$c(u_i(0)...u_i(t)) = \begin{cases} 1 & \text{if } u_i(\tau < t) \neq 1 \text{ and } u_i(t) = 1 \\ 0 & \text{else} \end{cases}$$

The second constraint limits transmission bit budget to α.

3.1. Distributed Thresholding Strategies

Using the FDR procedure described in section 2 as the basis of decision rule at time t, an algorithm can be established to perform the FDR procedure in a distributed fashion. Although a multi-layered approach can be used to solve more challenging problems, a single layer algorithm will be described for simplicity:

1. Each sensor calculates the p value of its observation, p_i, and tests p_i with $\frac{1}{m}\gamma$
2. The sensor(s) with $p_i \leq \frac{1}{m}\gamma$ declares its observation as significant, and communicates this decision to other sensors by a suitable protocol, (assume l of them declare their observations significant)
3. The decisions of the l sensors are fed back to the system and all the sensors update their threshold to $\frac{l+1}{m}\gamma$
4. Each sensor tests the p value of its data by the new threshold and declares its data as significant accordingly, (assume k more sensors declare their observations significant)
5. The new significant decisions are fed back to the system again, and the new threshold is updated to $\frac{l+k+1}{m}\gamma$
6. Steps 4 and 5 are repeated until when there is no more sensors that declare their observations as significant under the most current threshold, which is when the process terminates.

4. DOMAIN TRANSFORMED FDR PROCEDURE

In many cases the adaptive solution that has been described so far is sufficient. However, in some problems, the distribution of the observations may have characteristics which accentuate the suboptimal nature of the FDR procedure. Specifically, FDR procedure performs best when the p values of the data that come from H_1 are clustered near zero, and that may not necessarily be the case as seen in the following example.

Example Consider two Gaussian random variables with $f_0(x) \sim N(0,1)$ and $f_1(x) \sim N(-4,1)$, and consider the FDR constraint $\gamma = 0.05$. The goal is to detect as many samples of P_1 as possible from a mixture of samples subject to FDR constraint, γ. In this case most of the realizations of the random variable P_1 are close to 1 rather than 0, and FDR procedure described above will not declare them as significant. The issue here is that the procedure terminates before

the necessary threshold is met to detect the significant observations.

To overcome this problem, consider the following transformation on the random variables P_0 and P_1:

$$\hat{P}_i = 1 - P_i, i = 0, 1 \quad (3)$$

Since P_0 is uniformly distributed in (0,1), it is obvious that $\hat{P}_0$ is also uniformly distributed in (0,1). Observe, however, that most of the realizations of $\hat{P}_1$ are close to 0. Therefore, when the FDR procedure is performed on this new set of p values, more of the observations coming from H_1 will be declared as significant, thus the detection power of the test is increased. Furthermore the FDR constraint γ is still satisfied since the transform preserves the $U(0, 1)$ distribution of the p values for the observations coming from H_0.

In this section, a method is developed, which not only solves the early termination problem of the distributed algorithm, but also yields the best performance of FDR procedure, subject to the problem constraints. Conveniently, as a result of this solution, $E(T)$ is also minimized within the capabilities of FDR procedure.

4.1. A Measure Preserving Transform

FDR procedure does not assume knowledge of the distribution of observations under positive hypothesis. However, in many problems, the distribution of the observations under positive hypothesis is known, or can be estimated. Making use of the assumption that the distributions of the observations are known under null and positive hypotheses, a transformation in the p domain is introduced.

The transformation is simple in nature, and is a reorientation of the p domain. Despite its simplicity, it has three very important properties to note:

1. It preserves uniform distribution of p values under null hypothesis,
2. It maps a non-monotonic or monotonically increasing density of p values to a monotonically decreasing one,
3. After the p values are put in ascending order, the plot of p values vs indices looks like a convex function sampled at integer points.

4.1.1. Transformation of p Domain

Let g_0 and g_1 be the probability density functions of P_0 and P_1 respectively. Define the transformation, T_n, on the p domain described as follows:

1. Partition the range of g_1 into n bins of size $\varepsilon = 1/n$, preimages of which induce a partitioning of the p domain,
2. Index the partitions in p domain with a location index i, $i = 1..n$ according to their order of appearance as p ranges from 0 to 1,
3. Index the partitions in p domain with area index j, $j = 1..n$ such that the j^{th} partition has the j^{th} largest area under g_1,
4. Beginning from $j = 1$, rearrange the locations of the partitions so that the location index i of each bin is equal to its area index, j.

Proposition 4.1 *The sequence of transformations $\{T_n\}$ converges to a measurable transformation T.*

The following procedure is referred as the Domain Transformed FDR (DTFDR) procedure throughout the paper:

1. Apply the transformation T to p values of the observations
2. Follow the FDR procedure

The following propositions are direct results:

Proposition 4.2 *T is a measure invariant transformation for samples H_0 in the p-value domain.*

Proposition 4.3 *The DTFDR procedure controls the false discovery rate at the same level as the original FDR procedure.*

Proposition 4.4 *The measure transformation T converts an arbitrary density of p values for H_1 to a monotonically decreasing density over $(0, 1)$.*

4.1.2. Improving the Performance of FDR Procedure

Before proceeding any further, the term "stochastically larger" [4] must be introduced: We say that the random variable X is stochastically larger than the random variable Y, denoted $X \geq_{st} Y$, when $F_X(a) \leq F_Y(a)$ for all a.

Lemma 4.5 *Let $X_1..X_n \in (0, 1)$ be n independent random variables with common density function f_X and let $Y_1..Y_n \in (0, 1)$ be n independent random variables with common density function f_Y. Also, let $X_{(i)}$ and $Y_{(i)}$ denote the i^{th} smallest of $X_1..X_n$ and $Y_1..Y_n$ respectively. If $F_X(t) \geq F_Y(t)\ \forall t \in (0, 1)$, then $Y_{(i)} \geq_{st} X_{(i)}$.*

Theorem 4.6 *For any given data set with known distributions and any integer k, the probability of declaring the first k values as significant is larger under the DTFDR procedure than the FDR procedure.*

4.2. Communication Costs

Apart from the gains in detection power of FDR procedure, the transformation T leads to further advantages in communication costs. When H_1 is a multi modal distribution the linearly increasing FDR threshold can intersect the ordered

p value curve at multiple locations. Hence early termination is possible if an ordered p value above the threshold curve is discovered. However, the DTFDR procedure yields a convex p value curve and overcomes this issue.

Lemma 4.7 *The expected value of the ordered set of p values in the transformed domain are samples of a convex function.*

Proposition 4.8 *The communication cost of the system is minimized by implementation of the DTFDR procedure.*

5. EXPERIMENTAL RESULTS

We have tested the distributed algorithm on a field of sensors of size 128x128, with a Gaussian noise assumption. On the first layer, we set $\gamma_1 = .1$ and for the second layer we set $\gamma_2 = \gamma_1/6$ [2].

The two layer approach was implemented as follows: For the first layer, sensors collaborated with J_1 of their neighboring sensors to generate local estimates. The random variable tested at each sensor was $w_i = y_i - r_i$, where $r_i = \frac{1}{J_1}\sum_{j=1}^{J_1} y_i^j$. Here y_i^j denotes the observations of the neighboring sensors. The p value of these random variables were calculated and tested.

At the second layer, a similar approach was taken, however, only the sensors, who declared their observations to be significant at the first layer, collaborated.

The results of this approach is compared with the implementation of a nonadaptive system, namely Bonferroni procedure, and a related procedure described recently by [3].

6. REFERENCES

[1] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," Journal of the Royal Statistical Society, Series B, vol. 57, pp. 289300, 1995.

[2] Venkatesh Saligrama, Yonggang Shi, and William C. Karl, "Performance Guarantees in Sensor Networks," *29th International Conference on Acoustics, Speech, and Signal Processing,* 2004, Montreal, Canada

[3] R. Nowak and U. Mitra, "Boundary Estimation in Sensor Networks," *2nd International Workshop on Information Processing in Sensor Networks*, 2003, Palo Alto, CA

[4] S. Ross, *Introduction to Stochastic Dynamic Programming*. Academic Press, 1983.

[5] H. L. Royden, "Real Analysis," Prentice Hall, 1988

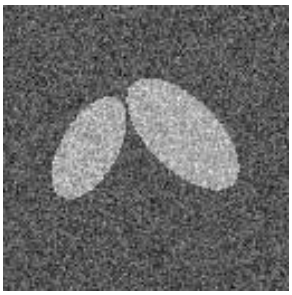
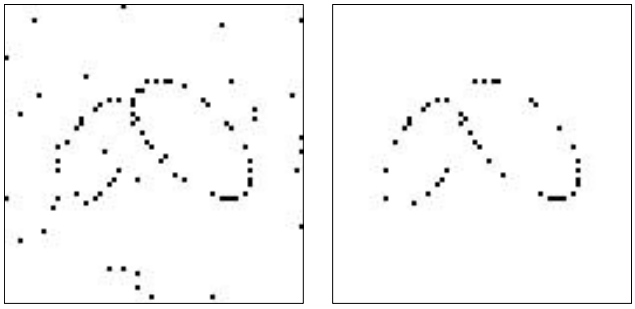
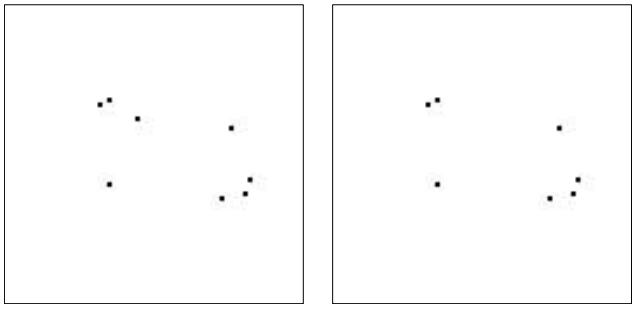


Fig. 1. The sensor field model



(a) Layer 1 (b) Layer 2

Fig. 2. Boundary detection via distributed implementation of FDR procedure



(a) Layer 1 (b) Layer 2

Fig. 3. Boundary detection via distributed implementation of Bonferroni procedure

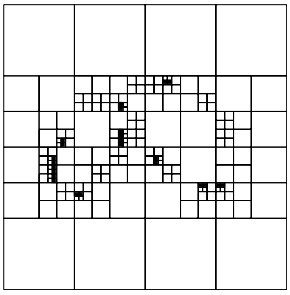


Fig. 4. Boundary detection via the method in [3]