PERFORMANCE GUARANTEES IN SENSOR NETWORKS

Saligrama Venkatesh

Yonggang Shi

William Clem Karl

Information Systems and Sciences Laboratory Electrical and Computer Engineering Department Boston University Email: {srv,yshi,wckarl}@bu.edu

ABSTRACT

Sensor networks for monitoring distributed spatial phenomena has emerged as an area of significant practical interest. In this paper we investigate fundamental issues in detection of spatially distributed phenomena under communication constraints. The novelty of the paper is in providing a tradeoff between global performance and costs involved in communication. In particular we focus our attention on boundary estimation and develop a framework to optimize communication costs subject to worst-case misclassification guarantees. It is shown that the communication cost is primarily a function of two parameters—1) length of the boundary 2) Overall mis-classification error which leads us to the conclusion that wireless sensor network performance is comparable to that obtained with a wired network of sensors.

1. INTRODUCTION

Recent advances in sensor and computing technologies [1, 2] provide impetus for deploying wireless sensor networks – a network of massively distributed tiny devices capable of sensing, processing and exchanging data over a wireless medium. Such networks are envisioned [2] to provide real-time information in such diverse applications as building safety, environmental control, power systems and manufacturing.

Despite its enormous potential the design and deployment of SNETs pose fundamental challenges. This is a direct consequence of three factors, namely, power, ad-hoc networking, and uncertainty. Power limits the range of transmission and overall system lifetime. Ad-hoc networking protocols lead to asynchronous data transmission [3]. Uncertainty implies the unreliability of data received by the decision agent(s). It has been shown recently [4] that the first two factors, namely, power limitation and asynchronous transmission together play a fundamental role in diminishing the average bandwidth per active user in a network. This implies that SNET operation must shift from the conventional *time-driven* to an *event-driven* mode. SNETs have received significant attention within the networking, signal processing and information theory communities [5, 6, 7, 8, 9, 10].

In this paper we focus on the question of how to provide fundamental guarantees for end-to-end information quality under communication constraints. We develop these ideas in the specific context of boundary detection. If the observations are centrally available there is a well-established soluproblems. tion methodology for such Fundamental problems arise when data is distributed and the centralized solutions are no longer feasible due to time and rate constraints (finite bit budget). Our motivation is based on recent work [8] on this topic, where an estimation viewpoint of the problem is presented. The setup in [8] consists of n nodes in a unit square area placed on a square lattice, with each sensor node making noisy estimates of the local sensor field. The sensors then locally collaborate with their corresponding "cluster head" to determine the estimates of the underlying field. The cluster head then makes a local decision either to preserve the distinct measurements or to combine them into a single aggregate parameter. This process is then repeated at different levels to obtain a multi-scale representation of the sensor field. Ultimately, the boundary estimate is provided by those sensors whose measurements are preserved as part of the representation of the sensor field.

Although, this approach provides meaningful tradeoffs between accuracy and the energy required for communication, it has significant drawbacks. First, boundary detection is an indirect outcome of the process of representing the sensor field. Second, the approach does not quantify the performance in terms of boundary detection or mis-classification error probability. Nevertheless, the main drawback is statistical. From a hypothesis testing viewpoint the method at the lowest layer can be interpreted as a variant of the so

This work was partially supported by The Air Force Office of Scientific Research under Grant F49620-03-1-0257, The National Institutes of Health under Grant NINDS 1 R01 NS34189, and The Engineering research centers program of the NSF under award EEC-9986821, and an ONR Young Investigator Grant No. N00014-02-1-0362

called Bonferroni procedure [11]. This scheme pertains to the problem of simultaneous testing of a large number of hypotheses. In the context of boundary detection this setup would view each local sensor observation as an independent hypothesis and thus with L sensors one has L different hypothesis. One is interested in testing between the null hypothesis (absence of a boundary) against a family of hypothesis (presence of a boundary). The sensor-by-sensor inference procedure leads to significant increase in the false positive rate. The Bonferroni procedure controls the overall false alarm rate, α , by setting a *uniform* local threshold, α/L , for each of the L sensors and performing simultaneous inferences at each of the sensors. As the number of sensors increase for a fixed global false alarm specification, the Bonferroni approach would imply a vanishingly small local false alarm rate, leading to a significant decrease in power of detecting a boundary when one exists.

Our paper overcomes these drawbacks by formalizing the problem as a distributed boundary detection problem. Rather than attempting to control the overall false positive rate, we attempt to control the so called false discovery rate (FDR [12, 13]). The false discovery rate is the expected fraction of accepted hypothesis that are erroneously accepted. This weaker notion, as it turns out, leads to significant increase in the power of detection, while still maintaining (in a weak sense) the false alarm rate. We develop a distributed implementation of the FDR procedure appropriate for sensor network problems. We show that the communication costs scale with the length of the boundary and the preset tolerance for the overall FDR.

The organization of the paper is as follows. In Section 2 we provide a description of the problem setup. In Section 3 various solution methodologies are discussed along with accompanying communication costs. Section 4 provides the main results.

2. PROBLEM STATEMENT

We desire a distributed boundary detection approach with limited communication, yet which directly controls performance, in particular the false discovery rate. Such a sensor network is illustrated on the left in Fig. 1, where the center of each sensor node is denoted by a triangle, and a node's region of influence or visibility is indicated with an associated circle.

As discussed above, the typical approach in such distributed detection problems is to collaborate with the nearest cluster head. Each cluster head then simultaneously performs pruning, which amounts to setting a uniform threshold on local detections. This corresponds to the so called Bonferroni strategy. In contrast, here we start with an overall specification of desired false discovery rate and then vary the local thresholds to achieve this performance goal. The objective is to quantify the relationship between size of boundary, desired FDR and the communication costs.

In this preliminary work we restrict ourselves to a two layer sensor network system. At the first and finest layer we apply an FDR-based threshold strategy for local detection of a boundary at each and every sensor node. For the sake of exposition we briefly describe the FDR scheme. Suppose, we have an indexed family of hypothesis, H_{0i} , $i = 1, \ldots, L$. The *p*-statistics p_i for each of the individual hypothesis is computed (we refer to the following section for details). The order statistics, $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(L)}$ for these p-values are then computed. The final step involves computation of the largest integer, K, such that $p_{(K)} \leq (K/L)\alpha$ and accepting those sensors whose p-statistics are smaller than this number and rejecting the rest.

In the context of boundary detection we view each sensor node as observing data x_i . We assume a probabilistic model for mapping the underlying sensor field to the observations. We consider two hypothesis. The null hypothesis, H_{0k} , denotes the absence of a boundary in the neighborhood of sensor, s_k , while, H_1 , denotes the presence of a boundary. In situations where each sensor has a probing radius R it is possible to define a mapping of hypothesis to the actual observations. In this situation we assume a general probability distribution. In situations where the sensor observations are point estimates we assume a Gaussian random field. Neighboring sensors collaborate with a cluster head as in [8], which in turn maps the two hypothesis to the observations. In either case local *p*-statistics corresponding to the presence of a boundary are formed individually at each of the sensor nodes. The question arises as to how to compute the order statistics in a distributed manner. This is accomplished based on a binning argument. There are $\log_2 L$ bins with the kth bin corresponding to the interval $(k\alpha/L, (k+1)\alpha/L]$. We assume that the desirable FDR rate, α , and the number of sensors, L, in the region are globally known. First, all the *j* sensors whose *p*-statistics fall into the first bin are identified as sensors at the boundary. This is accomplished by broadcasting or any other suitable protocol. A counter at each sensor node is then updated to j + 1. Sensors falling within the *j*th bin and excluding the first j sensors are then declared as accepted. The communication terminates when no such sensor is found. It is clear that the communication cost in this procedure scales linearly with the length of the boundary and is a monotonically increasing function of the FDR rate α . It turns out that this relationship can be precisely established.

While, this procedure may suffice in many instances, the detection performance degrades when either the noise increases or when the probing radius is small. To overcome this issue we consider a multi-layered approach. The idea is to set varying FDR rates at different layers to control the overall FDR rate. The first layer will typically have a larger

FDR rate and admits a larger number of false alarms. To further control these false alarms, we perform a second level of inference. In particular, for each sensor node that declared a detection at the finest level, we perform a confirmatory test at the second level, wherein each retained node communicates with its neighbors and performs a second boundary test on a larger spatial scale. These larger scale tests involve more data, occurring over a larger area, and thus are more reliable. The cost is the increased communication necessary to perform such a broader, coordinated test. Our approach limits this expensive second level communication to just a small subset of the entire set of possible nodes. Interestingly, the overall approach can be formulated in terms of an optimization problem that seeks to minimize the communication subject to overall FDR constraints. This is the first time, to our knowledge, that such performance guarantees have been included in a sensor network detection method.



Fig. 1. Illustration of a sensor network. On the left is the collection of sensor centers, denoted by triangles. The local sphere of influence of each sensor node is denoted by the associated circle. When a local detection occurs, the sensor communicates with its neighbors in an attempt to confirm its observation. This process is illustrated on the right. The sensor node with dashed lines has locally declared a detection of a boundary. It is coordinating with its neighbors, illustrated by the lines to refine this decision.

3. THRESHOLD STRATEGIES

We have a set of L sensors with the *i*th sensor denoted by s_i . The sensor s_i observes x_i , which is corrupted by i.i.d zero mean Gaussian noise with variance σ^2 . Each sensor, s_i , can have J_1 local neighboring sensors with which it can collaborate and their measurements are denoted as $x_i^1, \ldots, x_i^{J_1}$.

Our two layered approach for boundary detection works as follows. For the first layer, we choose the FDR to be α_1 . The hypotheses set consists of H_{0i} : { s_i is in a uniform region}. Let $m_i = \frac{1}{J_1} \sum_{j=1}^{J_1} x_i^j$. The random variable we test at each s_i is defined as:

$$w_i = x_i - m_i \tag{1}$$

and its probability distribution is $N(0, \frac{J_1-1}{J_1}\sigma^2)$. We first compute the *p*-value at each s_i : $p_i = 2(1 - F(|w_i|))$ where

F is the cumulative distribution function of w_i , and perform a distributed FDR test.

- Step 1: Sensors with $p_i \leq \frac{1}{L}\alpha_1$ are declared accepted. This is accomplished by some suitable protocol. Each sensor then updates a local counter to keep track of the number of edges, K_s that have been accepted.
- Step 2: If there are no sensors with $p_i \leq \frac{K_s+1}{L}\alpha_1$, the algorithm terminates; otherwise, the process is repeated.

At the second layer, we perform a similar FDR test with, rate α_2 , on the subset of sensors, \mathcal{E} that have been accepted. In this second layer, each element of the subset, \mathcal{E} , collaborates with J_2 neighbors (over a radius larger than the first layer), with mean value $m_{i,j}$, to form more reliable estimates. The random variable v_i is now defined as:

$$v_i = m_i - \frac{1}{J_2} \sum_{j=1}^{J_2} m_{i,j}$$
(2)

and its distribution is $N(0, \frac{J_2-1}{J_1J_2}\sigma^2)$. The same algorithm as in the first layer is then performed for sensors in the set \mathcal{E} . The parameters α_1 and α_2 are picked based on an optimization criterion which minimizes the overall communication costs subject to overall FDR constraints.

4. PRELIMINARY RESULTS

We have tested our algorithm on a field of sensors of size 128×128 as shown in Fig. 2.

On the first layer, we set $\alpha_1 = .1$, and the results of our algorithm is shown in Fig. 3(a). For the second layer, we set $\alpha_2 = \alpha_1/6$ and the final outcome is shown in Fig. 3(b).

We compare these results with a two layer Bonferroni method and results appear in Fig. 4(a) and (b). The result of the method [8] are also shown in Fig. 5. The boundaries detected by the method of [8] are indicated in Fig. 5 as solid squares.

Our two layer, optimization-based FDR approach has identified the correct boundary location and successfully suppressed false detections. The standard Bonferroni-based method does not possess sufficient power under the given false discovery constraint to find the boundary. Similarly, the complexity modified MSE procedure in [8] fails to robustly identify the boundary.

5. CONCLUSIONS

We have presented an approach to distributed processing in sensor networks which optimized communication costs subject to worst-case mis-classification guarantees. Our approach provides a framework in which to rationally include both communication and performance constraints.



Fig. 2. The noisy sensor network measurement.



Fig. 3. Boundary detection results with our method. (a) Layer 1 results. (b) Layer 2 results.



Fig. 4. Boundary detection results with the Bonferroni method. (a) Layer 1 results. (b) Layer 2 results.



Fig. 5. Boundary detection results with the method in [8].

6. REFERENCES

- [1] Seth Edward-Austin Hollar, "COTS dust," M.S. thesis, University of California, Berkeley, 2000.
- [2] G. T. Huang, "Casting the wireless sensor net," *MIT Technology Review*, July-August 2003.
- [3] Hong Xiaoyan, Xu Kaixin, and Mario Gerla, "Scalable routing protocols for mobile ad hoc networks," *IEEE Network Magazine*, pp. 11–21, July-August 2002.
- [4] P. Gupta and P. R. Kumar, "Capacity of wireless networks," *IEEE Transactions on Information Theory*, 2000.
- [5] J. F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, 2003.
- [6] N. Patwari and A. Hero, "Reducing transmissions from wireless sensors in distributed detection networks using hierarchical censoring," in *ICASSP*. IEEE, 2003.
- [7] "Collaborative information processing," *IEEE Signal Processing Magazine*, 2002.
- [8] R. Nowak and U. Mitra, "Boundary estimation in sensor networks," in 2nd International Workshop on Information Processing in Sensor Networks, Palo Alto, CA, April 2003.
- [9] P. Ishwar, R. Puri, S. S. Pradhan, and K. Ramchandran, "On rate-constrained estimation in unreliable sensor networks," in *IPSN 2003*, Palo Alto, CA, USA, April 2003, Information Processing in Sensor Networks, pp. 178–192.
- [10] Sandeep S. Pradhan and Kannan Ramchandran, "Distributed source coding using syndromes (discus): Design and construction," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 626–643, March 2003.
- [11] R.J. Simes, "An improved Bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 73, pp. 751–754, 1986.
- [12] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A pratical and powerful approach to multiple testing," *Journal of the Royal Statistical Society, Series B*, vol. 57, pp. 289–300, 1995.
- [13] X. Shen, H.-C. Huang, and N. Cressie, "Nonparametric hypothesis testing for a spatial signal," *Journal of the American Statistical Association*, vol. 97, pp. 1122–1140, 2002.