Learning With Distributed Data in Wireless Sensor Network

Meenakshi Panda*, T Panigrahi†, P M Khilar* and G Panda‡
*Department of Computer Science and Engineering,
National Institute of Technology Rourkela, India-769008
Email: 509cs106@nitrkl.ac.in
†Department of Electronics and Communication Engineering,
National Institute of Technology Rourkela, India-769008
Email: panigrahit@nitrkl.ac.in
‡School of Electrical Science,
Indian Institute of Technology, Bhubaneswar

***Abstract*—Wireless sensor networks (WSN) have been proposed as a solution to environment sensing, target tracking, data collection and others. WSN collect an enormous amount of data over space and time. The objective is to estimate of a parameter or function from these data. Learning is used in detection and estimation problems when no probablistic model relating an observation. This paper investigates a general class of distributed algorithms for data processing, eliminating the need to transmit raw data to a central processor. This can provide significant reductions in the amount of communication and energy required to obtain an accurate estimate. The estimation problems we consider are expressed as the optimization of a cost function involving data from all sensor nodes. Here the distributed algorithm is based on an incremental optimization process. A parameter estimate is circulated through the network, and along the way each node makes a small adjustment to the estimate based on its local data.**

I. Introduction

In wireless sensor networks(WSN) comprising the nodes are employed to collect data like local temperature, wind speed, humidity, or concentration of some materials etc over a geographic area and are envisioned to make a dramatic impact on a number of applications such as, precision agriculture, disaster relief management, radar, and acoustic source localization. In these applications, each node with its computational power, able to send data to a subset of the network nodes, and tries to estimate the parameter of interest [1], [2]. Therefore, there is a great deal of effort to devising algorithms that are able improve the estimate of the parameters of interest in every node with information exchange between nodes [4], [5]. More precisely, in mathematical terms, each node optimizes a cost function that depends on all information in the network. Since DWSN can be be deployed in almost any kind of terrain with a hostile environment, it is preferred over traditional wired network.

In a DWSN architecture, the system performs estimation, detection, classification, localization and tracking tasks etc. In a traditional centralized solution, nodes in the network collect data and transmit them to a central processor for further processing. The central processor estimates the parameter vector using received data and the result is broadcasted back to individual nodes. This mode of operation requires a powerful central processor and large amount of communication between sensor nodes and the central processor. In addition, a centralized solution limits the ability of the nodes to adapt in real time. Therefore, distributed signal processing and optimization are highly desired, especially in large-scale WSN systems [3].

Several methods for estimation problems are proposed in the literature which exploit the physical behavior of the event being measured. The most important characteristic of any event is the correlation between the measurements at different nodes. Recently, several distributed optimization algorithms based on gradient search have been proposed. Due to the large amount of complexity in assuring convergence of distributed gradient search algorithms, the objective function is assumed to be additive and convex.

Distributed signal processing deals with the extraction of information from local data collected at nodes that are distributed over a geographical area. Each node in a network records noisy observations related to the parameter to be estimated. The nodes would then interact with their neighbors in a certain manner, according to the network topology, either in an incremental [4], [7]–[9] or by diffusion [5] approach. A network is more efficient if it requires less communication between nodes to estimate the parameter vector [1], [2].

A. Distributed Optimization Techniques

Distributed wireless sensor networks are characterised by the mode of cooperation i.e. incremental or diffusion. In incremental mode of cooperation, each node transmits its local parameter estimate to the adjacent node and the information flows in a sequential manner. During this time, the nodes act like independent agents and there is a limited interaction among the nodes. This requires less amount of communication and power.

In diffusion mode of cooperation, each node transmits its local parameter estimate to all its neighbors as dictated by network topology. In this mode the nodes have access to more amount of data. But this mode of cooperation requires more amount of communication compared to incremental mode. The amount of communication can be reduced by allowing each node to communicate only with a subset of its neighbors.

In this paper, we focus on the incremental mode of cooperation. Consider a network consisting of N nodes. Each node has access to a local temperature measurement T_i. The objective is to provide each node with information about the average temperature $\hat{T}$. Averages can be viewed as the values minimizing quadratic cost functions. Quadratic optimization problems have solutions which are linear functions of the data. A simple accumulation of parameter estimate leads to a solution. General optimization problems can often be solved using this simple, distributed algorithms.

In general, an optimization problem can be expressed as:

$$f(\theta) = \frac{1}{N}\sum_{i=1}^{N} f_i(\theta) \tag{1}$$

where θ is the parameter to be estimated, and $f(\theta)$ is the cost function which can be expressed as a sum of N *local* cost functions $\{f_i(\theta)\}_N^{i=1}$ in which $f_i(\theta)$ only depends on the data measured at sensor i and is given by,

$$f_i(\theta) = \frac{1}{M}\sum_{j=1}^{M}(x_{i,j} - \theta)^2 \tag{2}$$

where $x_{i,j}$ is the j-th measurement of i-th sensor.

Hence putting the value of $f_i(\theta)$ from (2) into (1),

$$f(\theta) = \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}(x_{i,j} - \theta)^2 \tag{3}$$

In the proposed approach, an estimate of the parameter θ is passed from node to node. Each node updates the parameter to reduce its local cost (2) and then passes the updated parameter to the next node. The flow of information from first node to the last node forms a single cycle. Several cycles through the network are required to obtain a solution. These distributed algorithms can be viewed as incremental subgradient optimization procedures, and the number of cycles required to obtain a good solution can be characterized theoretically. If M and N are large, then a high quality estimate can be obtained using a distributed optimization algorithm with less energy and communications than the centralized approach.

II. Decentralized Incremental Optimization

For a convex differentiable function, $f : \Theta \rightarrow \mathbb{R}$, the following inequality for the gradient of f at a point θ_0 holds for all $\theta \in \Theta$:

$$f(\theta) \geq f(\theta_0) - (\theta - \theta_0)^T \nabla f(\theta_0)$$

In general, for a convex function f, a subgradient of f at θ_0 (observing that f may not be differentiable at θ_0) is any direction g such that

$$f(\theta) \geq f(\theta_0) - (\theta - \theta_0)^T g \tag{4}$$

and the subdifferential of f at θ_0, denote $\partial f(\theta_0)$, is the set of all subgradients of f at θ_0). Note that if f is differentiable at θ_0, then $\partial f(\theta_0) \equiv \{\nabla f(\theta_0)\}$; *i.e.*, the gradient of f at θ_0 is the only direction satisfying (4).

Here a network of N sensors is considered in which each sensor collects M measurements. Let $x_{i,j}$ denote the j-th measurement taken at the i-th sensor. We would like to compute

$$\hat{\theta} = \arg\min_{\theta\in\Theta} f(\theta) \tag{5}$$

where θ is a set of parameters which describe the global phenomena being sensed by the network and $f(\theta)$ is the cost function as defined in (3). The functions, $f_i : \mathbb{R}^d \rightarrow \mathbb{R}$ are convex (but not necessarily differentiable) and Θ is a nonempty, closed, and convex subset of $\mathbb{R}^d$.

The optimization problems can be solved iteratively by using gradient and subgradient methods. The update equation for a centralized subgradient descent approach to solve (5) is

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \alpha\sum_{i=1}^{N} g_{i,k} \tag{6}$$

where $g_{i,k} \in \partial f_i(\hat{\theta}^{(k)}), \alpha$ is a positive step size, and k is the iteration number. In this approach each update step uses data from all the sensors.

In a decentralized incremental approach, each iteration (6) is divided into N subiterations. In nth subiteration, nth sensor node updates its local parameter estimate$f_n(\theta)$. The algorithm can be written as:

$$\psi_0^{(k)} = \hat{\theta}^{(k-1)} \tag{7}$$

$$\psi_i^{(k)} = \psi_{i-1}^{(k)} - \alpha g_{i,k}, \quad i = 1, 2, \ldots, N \tag{8}$$

$$\hat{\theta}^{(k)} = \psi_N^{(k)} \tag{9}$$

where $\hat{\theta}^{(k)}$ is the estimated parameter vector obtained after k iterations and $\psi_N^{(k)}$ is the parameter estimate of Nth node in kth iteration. For analyzing the rate of convergence an arbitrary starting point is assumed.

The energy savings ratio between the use of an incremental optimization algorithm and a centralized optimization algorithm is shown to be [3]

$$R = c_3 M N^{1/d} \epsilon^2 \tag{10}$$

For N nodes with M readings each, a maximum estimation error ϵ, d the number of dimensions the sensor network is deployed in, and c_3 is the ratio between the number of bits required to describe the parameter vector θ and the measurements size in bits. Thus, as the number of readings or nodes in the network increases, it will become more advantageous to use an incremental algorithm for processing.

III. Estimation in Wireless Sensor Network by Distributed Learning

The main challenges in estimating parameters in a wireless sensor network are link failure and impulsive noise. A noise level that fluctuates over a range greater than 10 dB during observation is classified as impulsive. Here we propose an distributed algorithm which is robust to link failure as well as impulsive noise while maintaining faster convergence and low residual mean square error(MSE). Bang et al [6] have proposed a proportional sign algorithm which is robust in the presence of contaminated-Gaussian noise, but this algorithm only involves a fixed nonlinear function [10]. Delouille et al. in [11] have proposed a method that minimizes the mean square error by using an iterative algorithm. Transmission of all data to a central processor and then estimation using techniques such as Wiener Filtering (with complexity $O(N^2)$), requires a large amount of communication. Instead the estimation process is divided among smaller groups of nodes which will interchange their measurements to give an optimal set of estimates. In this paper the robust estimation in incremental approach is shown using a real time estimation problem in sensor network.

Suppose that a sensor network has been deployed over a region to find the average temperature. Each sensor collects a set of M temperature measurements, $\{x_{i,j}\}_{i=1}^{m}, j = 1, 2, \ldots, N$ over some period. At the end of the day the mean temperature

$$\hat{p} = \frac{1}{MN} \sum_{i,j} x_{i,j}$$

is to be calculated. Let us assume that the measurements are i.i.d. and the variance of each measurement is σ^2. However, some fraction say 10% of sensors are damaged or mis-calibrated, so that they give reading with variance $100\sigma^2$. Then the estimator variance will increases by a factor of 10. Ideally, these bad measurements should be identified and discarded from the estimation process. Robust estimation techniques attempt to do so by modifying the cost function.

In a general estimation problem, the classical least-square loss function, $\|x - \theta\|^2$, is used. For robust estimation, the classical least-square loss function is replaced with a different robust function, $h(x, \theta)$. Typically the the robust function $h(x, \theta)$ is chosen to give less weight to data points which deviate greatly from the parameter, θ. So the cost function (2) is modified for a robust estimation and given as

$$f_{robust}(\theta) = \frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M} h(x_{i,j}, \theta)$$

Several robust functions are available in literature. The l_1 distance is one example of a robust function. Another standard robust cost function is the Huber loss function [3] as

$$h(x;\theta) = \begin{cases} \|x-\theta\|^2/2, & \text{for } \|x-\theta\| \le \gamma \\ \gamma\|x-\theta\| - \gamma^2/2, & \text{for } \|x-\theta\| > \gamma \end{cases} \qquad (11)$$

This function acts as usual squared error loss function if the distance between the data point x and γ is within a threshold value γ that means if x close to θ, but gives less weight to points outside a radius γ from the location θ.

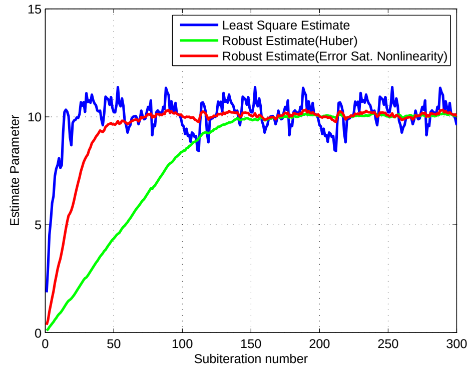
Here another function known as error saturation non-linearity [12], [13] is proposed which is robust against link failure and Gaussian-contaminated impulsive noise. The cost function for error $e = \|x - \theta\|$ is defined as

$$h(e) = \int_0^e \exp[-u^2/2\sigma_s^2]du = \sqrt{\frac{\pi}{2}} erf\left[\frac{e}{\sqrt{2}\sigma_s}\right] \qquad (12)$$

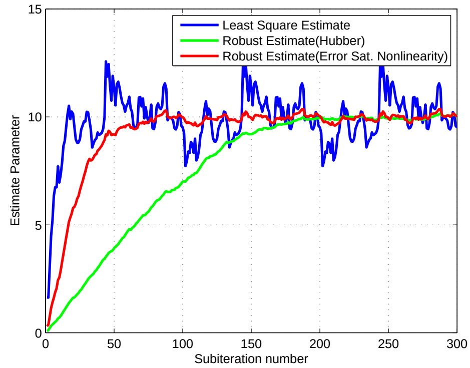
where σ_s is a parameter that defines the degree of saturation.

A distributed robust estimation algorithm is easily attained in the incremental subgradient framework by equating

$$f_i(\theta) = \frac{1}{M} \sum_{j=1}^{M} h(x_{i,j};\theta) \qquad (13)$$



(a) 10% of the sensors are damaged



(b) 50% of the sensors are damaged

Fig. 1. Robust incremental estimation procedures using Hubber function and error saturation nonlinearity when some nodes are damaged and in presence of impulsive noise (a) 10%.(b) 50% of the sensors are damaged

A. Robust incremental estimation during node failure and impulsive noise condition

A contaminated-Gaussian impulsive noise (a two component Gaussian mixture) [14]–[16] is modeled as

$$v(i) = v_g(i) + v_{im}(i) = v_g(i) + b(i)v_w(i) \qquad (14)$$

where $v_g(i)$ and $v_w(i)$ are independent zero mean Gaussian noise sequences with variances σ_g^2 and σ_w^2, respectively; $b(i)$ is a switch sequence of ones and zeros, which is modeled as an i.i.d. Bernoulli random process with probability of occurrence $P_r(b(i) = 1) = p_r$ and $P_r(b(i) = 0) = 1 - p_r$. The variance of $v_w(i)$ is chosen to be much larger than that of $v_g(i)$ so that with $b(i) = 1$, a large impulse is experienced in $v(i)$. The corresponding pdf of $v(i)$ is given as

$$f_v(x) = \frac{1-p_r}{\sqrt{2\pi}\sigma_g} \exp\left(\frac{-x^2}{2\sigma_g^2}\right) + \frac{p_r}{\sqrt{2\pi}\sigma_\Sigma} \exp\left(\frac{-x^2}{2\sigma_\Sigma^2}\right) \qquad (15)$$

where $\sigma_\Sigma^2 = \sigma_g^2 + \sigma_w^2$ and $E[v^2(i)] = \sigma_v^2 = \sigma_g^2 + p_r\sigma_w^2$.

The performance of the robust estimation algorithms in presence of impulsive noise is depicted in Figs. 1(a) and 1(b).

The parameters of the algorithm are taken as follows: the step size $\alpha = 0.1$ and $\sigma_s^2 = 10, \sigma_g^2 = 10^{-3}, \sigma_w^2 = 10^4\sigma_g^2$. The results shows that the incremental robust estimation algorithms are robust to impulsive noise. The simulation results also reveal that the incremental robust estimate using error saturation non-linearity converges faster compared to Hubber loss function.

IV. Conclusion

This paper has investigated a simple distributed algorithm for data processing in a wireless sensor network. The basic operation involves circulation of a parameter estimate through the network, and small adjustments to the estimate at each node based on its local measured data. This distributed algorithm can be viewed as incremental subgradient optimization procedure. A new cost function is proposed which is robust to node failure and impulsive noise. The simulation results show that the robust incremental estimate obtained using error saturation non-linearity is better than the estimate obtained by Huber loss function based method.

References

[1] D. Estrin and M. Srivastav, *Instrumenting the world with wireless sensor networks*,in Proc. Int. conf. Acoustics, Speech, Signal Processing (ICASP), May 2001, pp. 2033-2036.

[2] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, *A survey on sensor networks*, in Communications Magazine, IEEE, Aug 2002, 40(8), pp. 102-114.

[3] M. Rabbat and R. Nowak, *Distributed optimization in sensor networks*,in Proc. 3rd international symposium on Information processing in sensor networks, April 2004, pp. 20-27.

[4] Cassio G. Lopes and Ali H. Sayed, *Incremental Adaptive Strategies Over Distributed Networks*,IEEE Trans on Signal Processing, Vol. 55(8), Aug 2007, pp. 4064-4077.

[5] Cassio G. Lopes and Ali H. Sayed, *Diffusion least-mean squares Over adaptive Networks: Formulation and Performance Analysis*, IEEE Trans on Signal Processing, Vol. 56(7), July 2008, pp. 3122-3136.

[6] S. C. Bang and S. Ann, *A robust adaptive algorithm and its performance analysis with contaminated-Gaussian noise*, Proc. ISPACS, Seol, Korea, Oct 1994, pp. 295-300.

[7] Dimitri P. Bertsekas, "A New Class of Incremental Gradient Methods for Least Squares Problems", *SIAM J. Optim.*, vol. 7, pp. 913-926, Sep. 1997.

[8] Geary, A. and Bertsekas, D.P., "Incremental subgradient methods for nondifferentiable optimization", *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on.*, vol. 1, pp. 907-912, 1999.

[9] Nedic Angelia and Bertsekas D. P., "Incremental Subgradient Methods for Nondifferentiable Optimization", *SIAM J. on Optimization.*, vol. 12, no. 1, pp. 1052-6234, 2001.

[10] S.Koike, Adaptive threshold nonlinear algorithm for adaptive filters with robustness against impulse noise, *IEEE Trans on Acoustics, Speech, and Signal Processing*, Vol. 45(9), Sep 1997, pp. 2391-2395.

[11] V Delouille and R. Neelamani and Richard G. Baraniuk, *Robust Distributed Estimation Using the Embedded Subgraphs Algorithm*, IEEE Trans on Signal Processing, Vol. 55(8), No. 1, Aug 2006, pp. 2998-3010.

[12] N. J. Bershad, *On Error saturation nonlinearities for LMS adaptation*, IEEE Trans on Acoustics, Speech, and Signal Processing, Vol. 36(4), Apr 1988, pp. 440-452.

[13] N. J. Bershad, *On Error saturation nonlinearities for LMS adaptation in Impulsive noise*, IEEE Trans on Signal Processing, Vol. 56(9), Sep 2008, pp. 4526-4530.

[14] X. Wang and H. V. Poor, *Joint channel estimation and symbol detection in Rayleigh flat-fading channels with impulsive noise*, IEEE Communications Letter, Vol. 1(1), Jan 1997, pp. 19-21.

[15] S. C. Chan and Y. X. Zou, *A recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise: Fast algorithm and convergence performance analysis*, IEEE Trans on Signal Processing, Vol. 52(4), Apr 2004, pp. 975-991.

[16] S. R. Kim and A. Efron, *Adaptive Robust Impulsive noise Filtering*, IEEE Trans on Signal Processing, Vol. 43(8), Aug 1995, pp. 1855-1866.