

EVALUATING AVERAGE CAUSAL EFFECT USING WIRELESS SENSOR NETWORKS

Mark Coates and Ioannis Psaromiligkos

McGill University
Department of Electrical and Computer Engineering
3480 University St, Montreal, Quebec, Canada H3A 2A7.
{coates,yannis}@ece.mcgill.ca

ABSTRACT

Sensor networks have exciting potential applications in agriculture and medicine, where after the application of treatment, it is beneficial not merely to track the response but to assess the causal impact of the treatment reception. We describe a distributed algorithm for the evaluation of the *average causal effect* of treatment reception upon response. Our procedure applies the expectation-maximization algorithm across a graphical model of the system, using local message-passing techniques. The key collaborative step in the algorithm is simple message aggregation and averaging, which we perform over a tree network topology. Finally, for completeness purposes, we describe a simple framework for the construction and maintenance of the tree topology that provides a robust mechanism for executing the algorithm using spread-spectrum or ultra-wideband communication.

1. INTRODUCTION

Current applications of sensor networks in agricultural maintenance and medical monitoring [1, 2] suggest that there is a tremendous potential for extensive and beneficial use in these application domains. Existing proposals and implementations focus on *passive* sensor networks that simply monitor an environment (or patient). However, after the application of a treatment, it is beneficial not merely to monitor the localized response of the environment but to assess the causal effect of the treatment. Causal assessment is an important step in the development of *active* sensor networks, in which sensor nodes do not just monitor, but make local, autonomous decisions based on their combined measurements and act to modify or influence the environment.

The paper focuses on the scenario where we wish to evaluate the causal effect of treatment *reception* as opposed to treatment assignment or application. This problem has been explored in the context of treatment efficacy assessment in clinical studies where there is the potential for imperfect compliance [3, 4]. In this paper, we consider scenarios in which both the reception of treatment and the response can be expressed using binary variables. We also

assume that other local measurements are being conducted by the sensor node.

The estimation of causal effect using Bayesian graphical analysis techniques has been explored for many years [3–6], and graphical techniques and EM algorithms have also been applied recently in sensor networks [7, 8]. The major contribution of this paper is the application of distributed causal analysis using a sensor network. We describe a distributed algorithm designed to evaluate the *average causal effect*, as defined by Holland [9], of treatment reception upon response. Our procedure applies the expectation-maximization algorithm across a graphical model of the system, using local message-passing techniques. In the absence of strong spatial dependencies between nodes, the key collaborative step in the algorithm is simple message aggregation and broadcasting. In addressing the implementation of the aggregation, we outline a simple method for constructing and maintaining a tree network topology. We describe the nature of the localized message-passing between nodes in the tree, discussing its implementation as a form of spread-spectrum or ultra-wideband communication. The tree-based communication framework has the potential to support a range of sensor network data fusion algorithms (including those presented in [8]).

Section 2 of the paper reviews the concept of causal effect, describes a graphical model for the problem under consideration and outlines the distributed procedure for estimating the average causal effect. Section 3 describes a distributed algorithm for the construction and maintenance of a tree network topology. Section 4 briefly addresses physical layer communication. Finally, Section 5 makes concluding remarks and indicates future research directions.

2. ESTIMATING AVERAGE CAUSAL EFFECT

We address a scenario in which sensor nodes have been distributed in an environment, and a treatment has potentially been assigned at (or in the vicinity of) each sensor node location. A sensor node has knowledge of whether the treatment was assigned and can also determine if the treatment

was received. Finally, the sensor is able to determine if the response is positive.

Following the notation adopted in [4], we represent the assignment of the treatment using the binary variable Z , the reception of treatment by the binary variable D , and the success of the treatment by the binary variable Y . We let $z \in \{z_0, z_1\}$ represent the value of Z , where z_1 indicates that treatment was assigned, and z_0 that it was not. Similarly $d \in \{d_0, d_1\}$ and $y \in \{y_0, y_1\}$ represent, respectively, the values assumed by D and Y , with d_1 indicating that treatment was received, and y_1 indicating a positive observed response. The values z_0 , d_0 and y_0 are the negations of their respective counterparts.

We denote by U all characteristics, both observed and unobserved, that influence the values of D and Y . In general, U will comprise several random variables, both discrete and continuous. Beyond the binary values, each sensor also records a set of measurements W , the nature of which we leave unspecified, save that we assume there is an identifiable dependency between U and W .

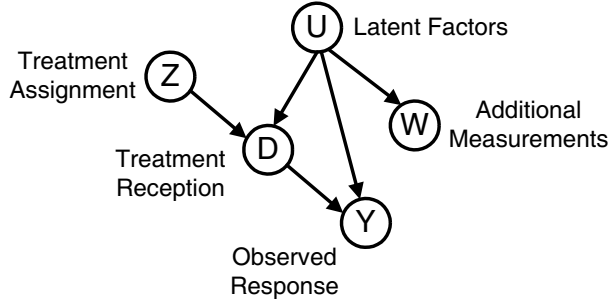


Fig. 1. Graphical model for the variables affecting measurement at an individual sensor node.

The graphical model depicted in Figure 1 represents independence assumptions about the joint probability distribution $p(z, d, y, w, u)$. It specifies that the treatment assignment Z does not directly affect the response Y , but only through the reception of treatment D . The model also asserts that the assignment of treatment does not depend on the latent factors that determine treatment reception and response. A third assertion is the conditional independence between W and Y given the latent variables U .

The graphical model also depicts some causal assumptions. The absence of a direct link from Z to Y implies that $p(y|d, u)$ is the same if d is measured or if its value is enforced exogenously. This permits the prediction of the response to an imposed reception of treatment. In this paper, we are concerned with the average effect on Y due to treatment, as measured by the *average causal effect*, defined as [9]:

$$ACE(D \rightarrow Y) = E_u[p(y_1|d_1, u) - p(y_1|d_0, u)]. \quad (1)$$

The average causal effect is thus the expected difference between the probability of the response being positive when treatment is received and when it is not received. Denoting our data across all sensors by \mathcal{D} , our goal in this paper is to estimate $ACE(D \rightarrow Y)$ by maximizing the likelihood function $\mathcal{L}(\mathcal{D}|ACE(D \rightarrow Y))$.

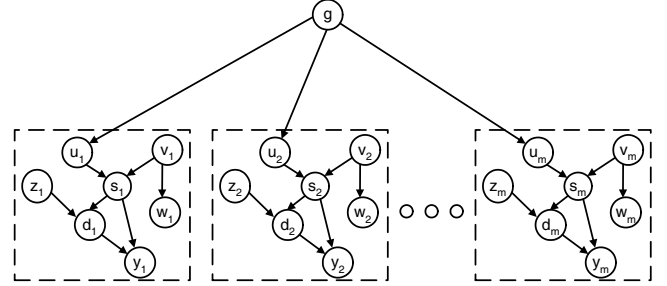


Fig. 2. Graphical model depicting the relationship between the global probability distribution $g(u)$, the local background states u_i and v_i , and the hidden response states s_i for all sensor nodes.

In the clinical trial research of [3,4], the additional measurements W were not present. The authors were able to exploit the observation of [6] that the latent factors U can be replaced by a single, finite state variable such that the resultant model is equivalent with respect to all measurements and manipulations of Z , D and Y . In our case, this replacement disrupts the dependency of W on the latent factors U . In this paper, we decompose the latent factors of the i -th sensor node into two components, U_i and V_i . Now V_i comprises all the *local* latent factors, and the U_i , which are assumed independent of one another, are realizations of finite state variables according to a global probability distribution, $g(u)$. This distribution captures the components of the relationships between treatment assignment, reception and response that are common to all nodes. Together, U_i and V_i determine a single, finite state variable S_i that governs the functional mappings between Z , D and Y , (as described in [4]). Figure 2 depicts the resultant graphical model across the entire system.

The state variable S_i can be considered to consist of sixteen equivalence classes, each of which describes two functional mappings; one from Z_i to D_i , and one from D_i to Y_i . It is convenient to regard each equivalence class as a point in the joint space of two four-valued variables C and R . The variable C determines the mapping from Z to D , and the variable R determines the mapping from D to Y . Using the same mappings as in [4], and denoting the states of S using the notation $cr^{(i,j)}$ with $0 \leq i, j \leq 3$, we can express the average causal effect as:

$$ACE(D \rightarrow Y) = \sum_i [p(cr^{(i,1)}) - p(cr^{(i,2)})]. \quad (2)$$

This is equivalent to the difference, for recipients of the treatment, between the probability of a positive effect and that of a negative effect.

We estimate average causal effect through application of the expectation-maximization (EM) algorithm [10] across the graphical model depicted in Figure 2. This algorithm generates an estimate $\widehat{ACE}(D \rightarrow Y)$ that (locally) maximizes the likelihood function $\mathcal{L}(\mathcal{D}|\widehat{ACE}(D \rightarrow Y))$. The expectation step of the algorithm uses conventional message-passing techniques associated with Bayesian networks to determine the expected values of the state variables s_i , u_i , and v_i [10]. Note that each message in the expectation step is local to one of the nodes (no inter-node communication is necessary).

The maximization step involves the determination of the global distribution $g(u)$ which maximizes the likelihood of observing the u_i across the m sensor nodes. We model $g(u)$ as a multinomial distribution, so this maximization takes the form of an averaging of the expected u_i values. In order to accomplish this, the expected u_i must be aggregated across the network (and divided by the number of nodes). Finally, upon convergence of the EM algorithm, the average causal effect must be estimated by averaging, across all sensor nodes, the evaluations of (2). The collaborative steps in the algorithm, appearing in the maximization step and the generation of the final result, consist solely of aggregations across the network, followed by dissemination (broadcasting) of the result. In the next section, we describe a tree-based communication framework that accomplishes the aggregation and broadcasting task.

3. TREE-BASED IMPLEMENTATION

A communication infrastructure based on a tree topology provides an efficient mechanism for executing the tasks of aggregation and network-wide dissemination. The efficiency is achieved because the tree topology enables hierarchical aggregation of messages, obviating the need to transport all node messages to a central location.

We assume that it is possible to establish connectivity across the network (otherwise aggregation would not be possible). A single node is identified as the root node (the identity of the root can change slowly over time). Data aggregation commences at the leaf nodes. Each leaf node sends its message (the expectation of the state variable u_i) to its parent node. When a parent node has received messages from all its children, it aggregates the received messages and its own message, and relays the result to its parent. The process continues until the root node is reached. The root node performs the maximization (or the estimation in the final step) and begins the dissemination process. In this phase, each parent node broadcasts the message from the root to all of its children nodes. During each iteration of the algorithm

each active node in the network performs at most two communication steps, sending a single message to its parent and broadcasting a single message to its children. In the rest of this section, we outline a simple algorithm for establishing the required tree network topology. For a more sophisticated (and complex) distributed tree-building algorithm the interested reader may wish to consult [11].

We assume that each node is assigned a unique ID that is determined prior to deployment. Initially, the network consists of a multitude of small trees. The proposed algorithm aims to merge all these trees into a single one that spans the whole network and has as a root the highest ID node. To make this possible, the nodes have to store local tree state information which is exchanged with their neighboring nodes. More specifically, each node stores the ID of his parent node, a list containing the IDs of its child nodes, and, finally, the ID of the root of the tree to which it belongs. We note that the latter ID uniquely identifies a tree in the network. The local topology information is transmitted using special “topology” packets which can be of the following four types:

- *Beacon*. Used by an isolated node to broadcast its own ID and attract other nodes to form a tree.
- *Root update*. Used by a parent node to notify its children of a change of the root node of the tree to which they belong.
- *Connect*. Used by a child node to notify its parent to add it to its child node list.
- *Disconnect*. Used by a child to instruct its parent to remove it from its child node list.

In addition, each node periodically transmits the root ID of its tree. To reduce the packet traffic in the network the root ID can be piggybacked on some of transmitted data packets. In the description of the tree formation and maintenance algorithm that follows, \mathcal{T}_n denotes the tree to which node n belongs, and \mathcal{S}_n denotes the subtree with node n as its top-level node.

Immediately after its activation, sensor node n starts scanning for trees in the surrounding area by listening to Beacon topology packets or data packets that contain root IDs. After a random amount of time, the node broadcasts a Beacon topology packet and repeats the process. The scanning process will be interrupted by either of the following two events: (i) the detection of a packet originated from node m belonging to tree \mathcal{T}_m with a higher root ID; or (ii) the detection of a Connect packet sent by another node. In the case of event (i), node n will transmit a Connect topology packet to node m and will attach to \mathcal{T}_m with node n as its parent. Event (ii) will cause the creation of a new tree with node n as the root.

Once node n is part of a tree it will participate in the regular data packet transmission and aggregation process. However, periodically, it will scan the vicinity for trees by listening to Beacon topology packets or data packets that contain root IDs. Packets that contain root IDs smaller than the root ID of \mathcal{T}_n are ignored. If, however, a higher root ID packet is received from node m of tree \mathcal{T}_m , then, the sub-tree \mathcal{S}_n will attach itself to tree \mathcal{T}_m through nodes n and m . This attachment takes place in three steps. First, node n severs its connection with its parent by issuing a Disconnect packet. Then, it transmits a Connect packet to node m in order to connect to tree \mathcal{T}_m . Finally, node n notifies its subtree of the new tree to which they now belong by transmitting a Root update packet to all of its children.

If a sensor node n does not receive any packets from its parent node after a fixed, predetermined, period of time then it assumes that the parent node has failed, and that the subtree \mathcal{S}_n has been disconnected from the network tree. The rest of the nodes are notified of this through Root Update packets that set the root ID to the ID of node n . As described previously, the new tree will eventually merge with one of the neighboring trees and the network tree topology will be reestablished.

4. PHYSICAL LAYER CONSIDERATIONS

The physical layer of a wireless sensor network must support node-to-node data and topology packet transmission. Node and battery size limitations as well as the usually hostile nature of the operational environment call for physical layer designs that are: (i) power efficient; (ii) robust to multipath propagation; (iii) difficult to intercept and eavesdrop upon; and (iv) resistant to unintentional interference as well as jamming. Spread-spectrum (SS) and ultra-wideband (UWB) [12] communications schemes exhibit all four properties and appear to be promising technologies for the physical layer of wireless sensor networks.

In our network, *all* nodes share a common communication channel defined as a common spreading signature (in the case of an SS-based system) or a common time-hopping code (in the case of UWB communications). Child-to-parent message passing is performed using a conventional multiple-access collision-avoidance (MACA) handshaking protocol [13] that involves the exchange of request-to-send (RTS) and clear-to-send (CTS) control packets between the transmitting and receiving node. Parent-to-children communication, however, is handled in a slightly different way. To take advantage of the broadcasting nature of parent-to-children communication we use a variant of MACA: the parent notifies *all* of its children of the forthcoming broadcasting by transmitting a single request-to-broadcast (RTB) control packet to which the children respond with the transmission of a CTS packet. When the parent receives CTS packets

from all of its children it broadcasts the data or topology packet.

5. CONCLUSIONS

In agricultural and medical applications, the role of sensor networks can develop beyond monitoring to active control or localized treatment. This involves decision-making, an arena in which causal analysis plays a vital role. We have presented a distributed algorithm for a specific example of causal analysis using a sensor network, but there remains much to explore and develop.

6. REFERENCES

- [1] K.A. Delin, R.P. Harvey, N.A. Chabot, S.P. Jackson, M. Adams, D.W. Johnson, and J.T. Britton, "Sensor web in Antarctica: developing an intelligent, autonomous platform for locating biological flourishes in cryogenic environments," in *Proc. 34th Lunar and Planetary Science Conf.*, Houston, TX, Mar. 2003.
- [2] D. Myung, B. Duncan, D. Malan, M. Welsh, M. Gaynor, and S. Moulton, "Vital dust: Wireless sensors and a sensor network for real-time patient monitoring," in *Proc. 8th New England Reg. Trauma Conf.*, Burl., MA, Nov. 2003.
- [3] G. Imbens and D. Rubin, "Bayesian inference for causal effects in randomized experiments with non-compliance," *Annals of Statistics*, vol. 25, pp. 305–327, 1997.
- [4] D. M. Chickering and J. Pearl, "A clinician's tool for analyzing non-compliance," in *Proc. 13th Nat. Conf. Artificial Intelligence*, Portland, OR, 1996, pp. 1269–1276.
- [5] D. Rubin, "Bayesian inference for causal effects: the role of randomization," *Annals of Statistics*, vol. 7, pp. 34–58, 1978.
- [6] J. Pearl, *Bayesian Networks and Probabilistic Reasoning*, chapter From Bayesian networks to causal networks, pp. 1–31, Alfred Walter Ltd., London, 1995.
- [7] V. Delouille, R. Neelamani, V. Chandrasekaran, and R. Baraniuk, "The embedded triangles algorithm for distributed estimation in sensor networks," in *Proc. IEEE Workshop Stat. Signal Processing*, St. Louis, MO, Sept. 2003.
- [8] R. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," *IEEE Trans. Signal Proc.*, vol. 51, no. 8, pp. 2245–2253, Aug. 2003.
- [9] P.W. Holland, *Sociological Methodology*, chapter Causal inference, path analysis, and recursive structural equations models, pp. 449–484, American Sociological Association, Washington D.C., 1988.
- [10] M.I. Jordan, Ed., *Learning in Graphical Models*, MIT Press, Cambridge, MA, 1999.
- [11] R. G. Gallager, P. A. Humblet, and P. M. Spira, "A distributed algorithm for minimum weight spanning trees," *ACM Trans. Program. Lang. and Systems*, vol. 5, pp. 66–77, Jan. 1983.
- [12] M. Z. Win and R. A. Scholtz, "Impulse radio: How it works," *IEEE Commun. Lett.*, vol. 2, pp. 36–38, Feb. 1998.
- [13] P. Karn, "MACA: a new channel access method for packet radio," in *Proc. 9th Computer Networking Conf.*, 1990, pp. 134–140.