DISTRIBUTED BELIEF PROPAGATION USING SENSOR NETWORKS WITH CORRELATED OBSERVATIONS

Alfonso Cano and Georgios B. Giannakis

Dept. of ECE, University of Minnesota, 200 Union Street SE, Minneapolis, MN, 55455

ABSTRACT

A distributed belief propagation protocol is developed to carry inference and decoding tasks using wireless sensor networks with high-dimensional, correlated observations. Statistical dependencies are modeled using factor graphs. The overall a-posteriori probability is factored so that its factor graph representation can be mapped to the actual communication network. Sum-product message passing updates over the graphical model can thus be mapped to messages among sensors. As an application scenario, distributed spectrum sensing is considered. Simulated tests show that exploiting the correlation present among sensor observations can considerably improve sensing performance.

1. INTRODUCTION

In many sensing applications the phenomena under observation belong to an exponentially-large structured set of hypotheses, such as images or coded messages [1, 2]. Large-scale phenomena can be effectively modeled using probabilistic graphical models whereby statistical dependencies among variables (pixels, bits) are visualized using links in a graph. This representation has helped deriving and understanding simple and effective inference and decoding rules via message passing [3]. In distributed set-ups, however, sensors only obtain local observations (potentially correlated to other sensors' observations) of the phenomena and thus local message passing without collaboration from other sensors may perform poorly. The objective of this paper is to develop distributed message-passing algorithms to carry inference and decoding tasks exploiting multiple sensor observations and the correlations among them.

Approaches for distributed hypothesis testing problems using graphical models are presented in [4,5]. However, these approaches are designed for simple phenomena represented by a discrete scalar random variable. If the problem dimension grows, the number of hypotheses grows exponentially and communication complexity becomes prohibitive. An alternative to these approaches was provided in [1] in the context of distributed decoding, where independence among sensor observations is exploited to derive distributed consensus averaging algorithms that scale linearly with the problem dimension. If sensor observations are correlated, however, this approach is not optimal. Correlated observations modeled via graphical models have been considered in other centralized scenarios such as joint decoding-channel estimation problems [6, 7].

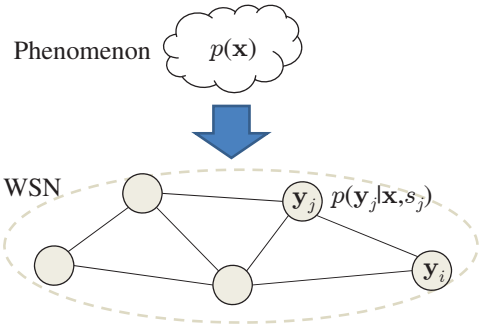

Fig. 1. Sensor network collecting observations.

In this work, the overall inference problem will be represented using two connected graphical models. One representing the statistical dependencies of the phenomena, and a second capturing the dependencies among sensor observations. The latter coincides with the *communication* graph, whereas the former does not have to [8]. The overall graph is re-expressed so that statistical dependency links can be mapped to communication links, and thus sum-product message passing updates over the graphical model correspond to messages among sensors. Compared to the consensus-type iterative algorithms in [1], distributed belief propagation (BP) converges in a finite number of iterations, and exploits correlated sensor observations.

2. SYSTEM MODEL AND PROBLEM STATEMENT

With reference to Fig. 1, consider a WSN consisting of $\mathcal{J} := \{1, \ldots, J\}$ sensor nodes observing the state of the environment represented by the $V \times 1$ vector $\mathbf{x}$. Each entry x_v is a random variable (r.v.) belonging to a finite field $\mathbb{F}_{Q_x}$ of cardinality Q_x representing, e.g., the presence/absence of a target as will be shown in Section 4. Other examples where finite-field environmental variables are considered are: information symbols [1], pixels of an image observed by multiple sensors [9], or robotic-generated maps [2].

Work in this paper was supported by NSF grants CISE-CCF-0830480, 1016605; ECCS-IHCS-0824007, 1002180; and the Spanish Science and Technology postdoctoral fellowship.

Correlation among entries in $\mathbf{x}$ is modeled using a pairwise factor graph with nodes x_v and factors $\phi(x_v)$ (associated with node x_v) and $\psi(x_v, x_w)$ (between nodes x_v and x_w) [3]. The probability mass function (pmf) of $\mathbf{x}$ is thus given as

$$p(\mathbf{x}) = \frac{1}{Z_x} \prod_{v \in \mathcal{V}} \phi(x_v) \prod_{(v,w) \in \mathcal{E}_x} \psi(x_v, x_w) \tag{1}$$

where $\mathcal{E}_x$ is the set of edges connecting entries in $\mathbf{x}$ and Z_x is a normalization constant.

Each sensor $j \in \mathcal{J}$ acquires a $V \times 1$ observation vector $\mathbf{y}_j$ from the state of the environment $\mathbf{x}$. Observations $\mathbf{y}_1, \ldots, \mathbf{y}_J$ are coupled through the unobserved or *hidden* random variables collected in the vector $\mathbf{s} := [s_1, \ldots, s_J]^T$, where $s_j \in \mathbb{F}_{Q_s}$. Variables s_j can represent the unknown fading channel state between the symbol $\mathbf{x}$ and observation $\mathbf{y}_j$ [6].

Assuming conditional independence, the probability of observing $\mathbf{y}_1, \ldots, \mathbf{y}_J$ given $\mathbf{x}$ and $\mathbf{s}$ is given by

$$p(\mathbf{y}_1, ..., \mathbf{y}_J | \mathbf{x}, \mathbf{s}) = \prod_{j=1}^{J} p(\mathbf{y}_j | \mathbf{x}, s_j) \tag{2}$$

where $p(\mathbf{y}_j | \mathbf{x}, s_j)$ is the per-sensor conditional pdf. In the absence of memory, (2) can be further written as

$$p(\mathbf{y}_j | \mathbf{x}, s_j) = \prod_{v=1}^{V} p(y_{jv} | x_v, s_j). \tag{3}$$

In this model different s_j are assumed correlated according to the distance (i.e., connectivity) among sensors. Its pmf $p(s_1, \ldots, s_J)$ is thus assumed factored as

$$p(\mathbf{s}) = \frac{1}{Z_s} \prod_{j=1}^{J} \alpha(s_j) \prod_{(j,i) \in \mathcal{E}_s} \beta(s_j, s_i) \tag{4}$$

where $\mathcal{E}_s$ corresponds to the set of edges connecting sensors; $\alpha(s_j)$ and $\beta(s_j, s_i)$ are factor nodes; and Z_s is a normalization constant. It is worth stressing that unlike $\mathcal{E}_x$, the edge set $\mathcal{E}_s$ does correspond to the *communication* graph of the WSN. For example, closeby sensors experience the same fading effects with higher probability than faraway sensors.

Let $p(\mathbf{x}|\mathbf{y}_1, \ldots, \mathbf{y}_J)$ denote the probability that the state of the environment is $\mathbf{x}$ given that $\mathbf{y}_1, \ldots, \mathbf{y}_J$ was observed. Given this model, the problem of interest is to find maximum a-posteriori probability (APP) of each x_v, namely

$$p_v(x_v | \mathbf{y}_1, ..., \mathbf{y}_J) \propto \sum_{\mathbf{x} \in \mathbb{F}_{Q_x}^V \backslash x_v} \sum_{\mathbf{s} \in \mathbb{F}_{Q_s}^J} p(\mathbf{y}_1, ..., \mathbf{y}_J | \mathbf{x}, \mathbf{s}) p(\mathbf{x}) p(\mathbf{s}) \tag{5}$$

where $\mathbb{F}_{Q_x}^V \backslash x_n$ is the set of vectors $\mathbf{x}$ with fixed x_v. Note also that the term $p(\mathbf{y}_1, \ldots, \mathbf{y}_J)$ has been discarded as it does not depend on $\mathbf{x}$. Finding (5) is challenging because the domains of $\mathbf{x}$ and $\mathbf{s}$ are discrete. An approximate solution can be found via message passing [10] over the factor graph representation of $p(\mathbf{y}_1, ..., \mathbf{y}_J | \mathbf{x}, \mathbf{s}) p(\mathbf{x}) p(\mathbf{s})$. The algorithm involves passing messages from variable nodes to factor nodes, and vice versa [3]. Note however that, in general, this message passing algorithm cannot be efficiently implemented in a distributed fashion since the factor graph describing $p(\mathbf{x})$ does not coincide with the underlying *communication* graph. Sensors cannot attempt to compute (5) locally either, since they only have access to the local posterior $p(\mathbf{y}_j | \mathbf{x}, s_j)$.

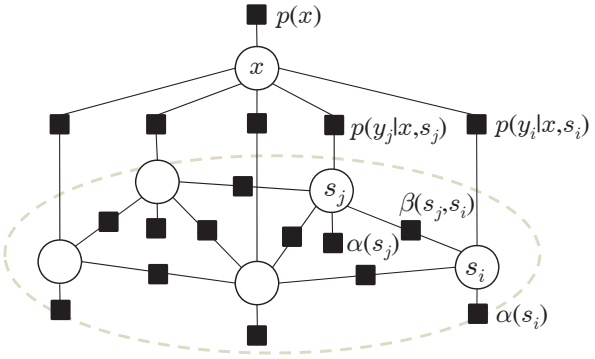

Fig. 2. Factor graph (scalar field).

3. DISTRIBUTED BELIEF PROPAGATION

In this section, the MAP in (5) is reformulated so that it can be efficiently computed in a distributed fashion. A simpler scenario assuming a scalar $\mathbf{x}$ will be tackled first. This will provide intuition for the most general case, pursued afterwards.

3.1. Scalar observations

Assume $\mathbf{x} = x$ is a scalar, and thus $p(\mathbf{y}_j | \mathbf{x}, s_j) = p(y_j | x, s_j)$ in (5). The APP problem in (5) then reduces to [cf. (1), (4)]

$$p(x|y_1, ..., y_J) \propto \sum_{\mathbf{s} \in \mathbb{F}_{Q_s}^J} \prod_{j=1}^{J} p(y_j | x, s_j) p(x) \alpha(s_j) \prod_{(j,i) \in \mathcal{E}_s} \beta(s_j, s_i). \tag{6}$$

The factor graph representation of the term inside the sum is depicted in Fig. 2. As seen, sensor variables are coupled through x, which is common to all. To circumvent this problem, define the per-node-j local variable x_j, and a similarity function $\delta(x_j, x_i) = 1$ whenever $x_j = x_i$, and $\delta(x_j, x_i) = 0$ otherwise. The APP in (6) is equivalent to the following

$$p_j(x_j | y_1, ..., y_J) = \sum_{\mathbf{x} \in \mathbb{F}_{Q_x}^J \backslash x_j} \sum_{\mathbf{s} \in \mathbb{F}_{Q_s}^J} p(x_1, ..., x_J, \mathbf{s} | y_1, ..., y_J). \tag{7}$$

Note the abuse of notation, where $\mathbf{x}$ has been redefined to be of dimensionality $J \times 1$. The posterior inside the sum in (7) can now be factored as

$$p(x_1, ..., x_J, \mathbf{s} | y_1, \ldots, y_J) \propto \prod_{j=1}^{J} p(y_j | x_j, s_j) p^{1/J}(x_j) \alpha(s_j) \prod_{(j,i) \in \mathcal{E}_s} \beta(s_j, s_i) \delta(x_j, x_i). \tag{8}$$

This model effectively decouples global dependencies into local ones. All factors depend on either local variables x_j or s_j, or on local and neighboring variables x_i or s_i. To derive efficient message passing schedules, group local variables by defining $z_j := (x_j, s_j)$, taking values in $\mathbb{F}_{Q_x} \times \mathbb{F}_{Q_s}$. Local and neighboring factors can be likewise grouped by defining $f(z_j) := p^{1/J}(x_j)\alpha(s_j)p(y_j|x_j, s_j)$ and $g(z_j, z_i) := \beta(s_j, s_i)\delta(x_j, x_i)$, respectively. Message passing can now be carried across the factor graph that results from these definitions. Since the factor graph is pairwise, the message-passing algorithm can be simplified to a single node-to-node schedule. Message updates from node j to i take the form

$$\mu_{j\to i}(z_i) \propto \sum_{z_j} f(z_j)g(z_j, z_i) \prod_{k\in N(j)\setminus i} \mu_{k\to j}(z_j) \qquad (9)$$

where $N(j) \setminus i$ is the set of neighboring nodes to j except for i. Message $\mu_{j\to i}(z_i)$ can be seen as a function with Q_xQ_s different values, one for each state of z_j. The algorithm starts by setting all messages to one and runs for a fixed number of iterations. At every node j the approximate marginal distribution $p(x_j|y_1, \ldots, y_J)$ in (7) is computed as

$$b(x_j) = \frac{1}{W} \sum_{s_j\in\mathbb{F}_{Q_s}} \prod_{i\in N(j)} f(z_j)\mu_{i\to j}(z_j). \qquad (10)$$

where W is a normalization constant that ensures (10) is a valid pmf. Since the factor graph contains loops, $b(x_j)$ is not guaranteed to be equal to $p(x_j|y_1, \ldots, y_J)$. However when the factor graph is acyclic the iterates in (9) are guaranteed to converge to the exact per-entry marginals in a number of iterations that equals the diameter of the network.

When $p(\mathbf{s}) = \prod_{j=1}^{J} Q_s^{-1}$; i.e., the entries of vector $\mathbf{s}$ entries are uncorrelated and equally-distributed, the scalar model in this subsection boils down to the model in [5], in which a scalar x is assumed too. The next subsection considers the most general case, not considered in [5] in which x is a vector with correlated entries.

3.2. Vector observations

For V-dimensional $\mathbf{x}$, one can certainly define local vectors $\mathbf{x}_j$ and proceed as in the previous subsection. However, the number of states of $(\mathbf{x}_j, s_j) \in \mathbb{F}_{Q_x}^V \times \mathbb{F}_{Q_s}$ grows exponentially with V, and so message passing becomes impractical. Instead, the idea here is to introduce V variables $\mathbf{s}_v$ and redefine $z_{jv} := (x_{jv}, s_{jv})$. The problem in (5) is now equivalent to

$$p_{jv}(x_{jv}|\mathbf{y}_1, ..., \mathbf{y}_J) = \sum_{\mathbf{x}\in\mathbb{F}_{Q_x}^{JV}\setminus x_{jv}} \sum_{\mathbf{s}\in\mathbb{F}_{Q_s}^{JV}} p(\mathbf{x}_1, ..., \mathbf{x}_J, \mathbf{s}_1, ..., \mathbf{s}_V|\mathbf{y}_1, ..., \mathbf{y}_J). \quad (11)$$

The posterior inside the sum in (11) can be factored as

$$\prod_{j,v} f(z_{jv}) \left(\prod_{v} \prod_{(j,i)\in\mathcal{E}_s} g(z_{jv}, z_{iv}) \right) \left(\prod_{j} \prod_{(v,w)\in\mathcal{E}_x} h(z_{jv}, z_{jw}) \right) \qquad (12)$$

where the functions $f(z_{jv})$, $g(z_{jv}, z_{iv})$, $h(z_{jv}, z_{jw})$ are respectively (re) defined as

$$f(z_{jv}) := \phi^{1/J}(x_{jv})\alpha^{1/V}(h_{jv})p(y_{jv}|x_{jv}, s_{jv})$$
$$g(z_{jv}, z_{iv}) := \beta^{1/V}(s_{jv}, s_{iv})\delta(x_{jv}, x_{iv})$$
$$h(z_{jv}, z_{jw}) := \psi^{1/J}(x_{jv}, x_{jw})\delta(s_{jv}, s_{jw}).$$

Careful examination of (12) reveals that factors can be grouped into $V + J$ subgroups; one subgroup per sensor j and one subgroup per dimension v. An alternating message passing schedule can be derived, involving two steps: 1) local message passing per node j; and 2) distributed message passing per dimension v. These two steps are detailed next.

1) Intra-sensor message passing: The per-sensor-j factors in (12) are $\prod_v f(z_{jv}) \prod_{(v,w)\in\mathcal{E}_x} h(z_{jv}, z_{jw})$, which forms a subgraph. We introduce a prior distribution for $r(z_{jv})$, initialized as $r(z_{jv}) = 1$ and derive message updates

$$\mu_{jv\to jw}(z_{jw}) \propto \sum_{z_{jv}} r(z_{jv})f(z_{jv})h(z_{jv}, z_{jw}) \prod_{ju\in N(jv)\setminus jw} \mu_{ju\to jv}(z_{jv}).$$

After a fixed number of iterations an approximation to the posterior distribution is given by

$$t(z_{jv}) \propto \prod_{ju\in N(jv) r(z_{jv})} f(z_{jv})\mu_{ju\to jv}(z_{jv}) \qquad (13)$$

which is used as a prior for the inter-sensor message passing schedule described next.

1) Inter-sensor message passing: The per-dimension-v factors in (12) are $\prod_v t(z_{jv}) \prod_{(j,i)\in\mathcal{E}_s} g(z_{jv}, z_{iv})$ using the prior $t(z_{jv})$, the following message updates can be derived

$$\mu_{jv\to iv}(z_{iv}) \propto \sum_{x_{jv}} t(z_{jv})g(z_{jv}, z_{iv}) \prod_{kv\in N(jv)\setminus iv} \mu_{kv\to jv}(z_{jv}).$$

After a fixed number of iterations an approximation to the posterior distribution $r(z_{jv})$ is given by

$$r(z_{jv}) \propto \prod_{iv\in N(jv)} t(z_{jv})\mu_{kv\to jv}(z_{jv}) \qquad (14)$$

with which the message passing in step 1) can be run again.

The overall algorithm thus runs iteratively, alternating between steps 1) and 2), and converges if $r(z_{jv}) = t(z_{jv})$, which also yields the final marginal probabilities. Similar alternating schedules can be found in [11]. With loopy factor graphs, this schedule does not generally provide any guarantees on the quality of the approximation or even convergence,

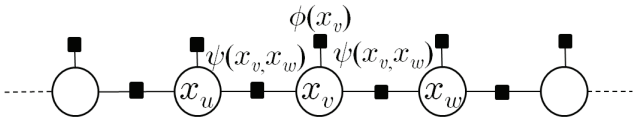

Fig. 3. Correlation model among frequency bands.

even if the individual subgraphs in steps 1) and 2) are trees. Section 4 will test this schedule via simulations.

Iterations in step 1) are carried within each sensor, as all factors used are locally available. Iterations in step 2) involve sending (receiving) messages $\mu_{jv\to iv}(z_{iv})$ ($\mu_{iv\to jv}(z_{jv})$) to (from) neighbouring nodes $iv \in N(jv)$ per dimension v. The overall number of scalars transmitted per sensor j are $|N(j)|VQ_xQ_s$, which grows linearly with V.

4. DISTRIBUTED SPECTRUM SENSING

Consider the WSN is deployed to cooperatively find unoccupied spectral bands by primary users (PUs). Sensors observe V bands and collect K samples per band. Denoting $x_v \in \{0,1\}$ as a binary random variable indicating the presence/absence of a signal at the v-th band, the receive-power in this band per sensor j is given by

$$x_v = 0 : y_{jv} = \eta_{jv}$$
$$x_v = 1 : y_{jv} = h_j P_v + \eta_{jv}$$

where P_v is the PU power, $h_j = h_j(s_j)$ is the *unknown* shadowing effect, assumed to belong to a finite number of states, and η_{jv} is the average receiver noise assumed Gaussian $\eta_{jv} \sim N(0,\sigma_w^2)$ (this defines $p(y_j v|x_v,s_j)$). Samples y_{jv} from nearby bands v,w are assumed correlated, and so is the shadowing $h_j(s_j)$ from nearby sensors j,i. This motivates tackling this problem as in the previous section.

Fig. 2 shows the dependency graph for x_v, which is a Markov chain. Since this model is acyclic, factors $\phi(x_v)$ and $\psi(x_v,x_w)$ can be expressed as $\phi(x_v) := p(x_v)$, $\psi(x_v,x_w) = p(x_w|x_v)/p(x_w)$, where $p(x_v)$ is the prior probability of band v being busy, and $p(x_w|x_v)$ is the probability band w being busy given the neighboring band v is busy [3]. Note that here $Z_x = 1$ in (1).

We simulate a network with $J = 10$ sensors arranged in a circle observing $V = 4$ subbands. Shadowing is assumed to take $h(s_j) = s_j = \{0,1\}$ (on-off) states, with $\alpha(s_j = 1) = 0.65$ and $\beta(s_j = 1, s_i = 1) = q$ and $1-q$ otherwise. Vector $\mathbf{x}$ is generated using $\phi(x_v) = 0.75$ and $\psi(x_v,x_w) = 0.65$. Fig. 4 shows the simulated band detection error probability as a function of q, which is a proxy for fading correlation. For comparison, scalar BP ignoring correlations among entries in $\mathbf{x}$ and $\mathbf{s}$ as in [5] as well as vector BP ignoring correlations in $\mathbf{s}$ are included. As shown in Fig. 4, the proposed distributed BP algorithm efficiently exploits statistical dependencies in $\mathbf{x}$ and $\mathbf{s}$ to dramatically reduce the uncertainty introduced by the noise and shadowing.

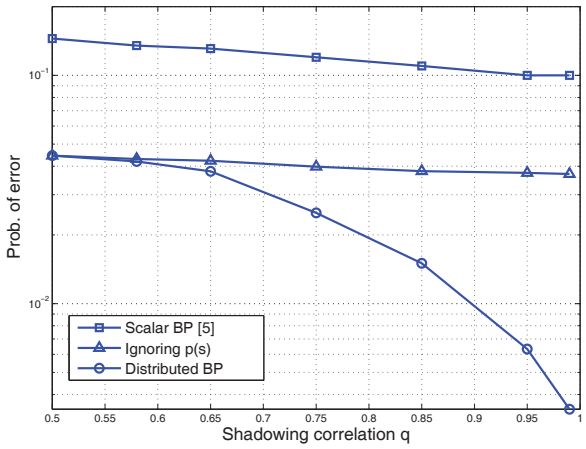

Fig. 4. Error probability as a function of q for a WSN with $J = 10$ sensors observing $V = 4$ spectrum bands.

5. REFERENCES

[1] H. Zhu, G. B. Giannakis, and A. Cano, "Distributed In-Network Channel Decoding," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3970-3983, Oct. 2009.

[2] Y. Rachlin, R. Negi, and P. K. Khosla, "The Sensing Capacity of Sensor Networks," *IEEE Transactions on Information Theory,* vol. 57, no. 3, 1675-1691, March 2011.

[3] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor Graphs and the Sum-Product Algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[4] R. Olfati-Saber, E. Frazzoli E. Franco, and J. S. Shamma, "Belief Consensus and Distributed Hypothesis Testing in Sensor Networks," in *Network Embedded Sensing and Control*, P. J. Antsaklis and P. Tabuada, Editors, pp. 169-182, Springer–Verlag, 2006.

[5] V. Saligrama, M. Alanyali and O. Savas, "Distributed Detection in Sensor Networks with Packet Losses and Finite Capacity Links," *IEEE Trans. on Signal Processing*, vol. 54, pp. 4118 – 4132, Nov. 2006.

[6] P. Schniter, "A Message-Passing Receiver for BICM-OFDM over Unknown Clustered-Sparse Channels," *IEEE Wrkshp. on Signal Proc. Advances in Wireless Comms.*, San Francisco, CA, 2011.

[7] Y. Qi and T. P. Minka, "Expectation Propagation for Signal Detection in Flat-fading Channels," *IEEE Transactions on Wireless Communications,* vol. 6, no. 1, 348-355, 2007.

[8] S. Sardellitti and S. Barbarossa, "Energy Preserving Matching of Sensor Network Topology to Dependency Graph of the Observed Field," *Proc. of 17th Intl. Conf. on DSP,* Corfu Island, Greece, July 6-8, 2011.

[9] E. Kokiopoulou, and P. Frossard, "Distributed Classification of Multiple Observation Sets by Consensus," *IEEE Transactions on Signal Processing,* vol. 59, no. 1, Jan. 2011.

[10] J. S. Yedidia, W. T. Freeman and Y. Weiss, " Understanding Belief Propagation and Its Generalizations" *IJCAI Distinguished Lecture track*, 2001.

[11] K. Rajawat, A. Cano, and G. B. Giannakis, "Network-Compressive Coding for Wireless Sensors with Correlated Data," *IEEE Transactions on Wireless Communications,* submitted June 2011.