FUSION IN SENSOR NETWORKS: CONVERGENCE STUDY

Elijah C. Liu and José M. F. Moura

Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, PA 15213-3890
{eliu, moura}@ece.cmu.edu

ABSTRACT

In sensor networks, many sensors cooperate and collaborate to monitor overlapping subsets from a set of targets. We consider the important issue of fusing their soft decisions. These soft decisions depend on the sensors' measurements and take the form of probability densities. Consequently, data fusion becomes a problem of probabilistic inference on a factor graph of arbitrary topology, which can be accomplished by belief propagation. This paper studies the convergence of belief propagation when the soft decisions are Gaussian densities, that is, studies the convergence of the variances and means computed by belief propagation. We show that if the spectral radius ρ of a certain matrix is less than one, the means resulting from belief propagation converge to the true means. This extends to general topology sensor networks the results for a fully-connected network of two sensors and m targets in [1].

1. INTRODUCTION

Advances in wireless communications and distributed embedded systems have fueled the growth of research in the field of sensor networks [2]. In this paper, we focus on data fusion in sensor networks by analyzing belief propagation over a general graph topology with Gaussian densities. This paper gives a sufficient condition on the spectral radius of a certain matrix for convergence of the means as computed by belief propagation to the true means.

Sensor networks involve many sensors detecting multiple, perhaps different, targets. Like in [3], we assume that local processors collocated at the sensors map the measurements into probability density functions, which we call soft decisions. For example, given a sensor network with n sensors and m targets, if sensor y_i detects targets x_1, x_3, x_{m-1}, then the corresponding soft decision of the sensor's processor is $p_i^* = p(x_1, x_3, x_{m-1}|y_i)$. From these decisions, we desire to obtain the marginal densities $p(x_i|y_1, \ldots, y_n)$ of the joint density $p(x_1, \ldots, x_m|y_1, \ldots, y_n)$.

This work was supported by DARPA DSO Integrated and Sensing Program through the Army Research Office Grant DAAD19-02-1-0180.

Thus, data fusion in sensor networks can be modeled as probabilistic inference on factor graphs [3]. Furthermore, finding the marginal densities of a factorable joint probability density function amounts to applying belief propagation, an instance of the sum-product algorithm, on a representative factor graph [1, 4, 5]. When the joint density assumes a Gaussian distribution, the messages in belief propagation simply consist of second-order Gaussian statistics [5]. In addition, [6] and [1] have studied the convergence of belief propagation on Gaussian graphs. We recap their results below, and then, explain how our work differs from theirs.

In [6], the authors derive closed-form expressions for the marginal statistics generated by belief propagation on general cyclic graphs and provide a sufficient condition for convergence. Furthermore, they show that when the messages in belief propagation converge to a fixed point, then the marginal means based on these messages converge to the true means. Unfortunately, the given sufficient condition does not easily extend to practice and does not give insight into which Gaussian systems lead to convergence.

The authors in [1] confine their analysis to a factor graph of two factor nodes fully connected to an arbitrary number of variable nodes. Using a different approach than that in [6], they also demonstrate that when belief propagation converges, the estimated marginal means coincide with those of the true marginal densities. Moreover, they state a sufficient condition that can be verified in practice and identify certain structures in the covariance matrices of the soft-decision densities that satisfy the sufficient condition.

However, the graph in [1] cannot model sensor networks well because sensor networks generally involve many sensors, each sensing a collection of targets that could differ from those sensed by others. The resulting bipartite graphs usually contain three or more factor nodes, which we call sensor nodes, that do not fully connect to all variable nodes, which we call target nodes. In this paper, we extend results in [1] to Gaussian graphs of arbitrary topologies.

For notational simplification and due to lack of space, we illustrate our approach with a specific graph. However, we have shown that these results inductively apply to more general factor graphs. In the first section, we present Theorem 1, which demonstrates under reasonable assumptions the convergence of the message variances sent from the sensor nodes to the target nodes. These message variances take the form of diagonal covariance matrices. In the second section, we present Theorem 2, which provides a sufficient condition under the same assumptions for the convergence of the message means and states that the marginal means calculated from the messages converge to the true means. The message means take the form of mean vectors.

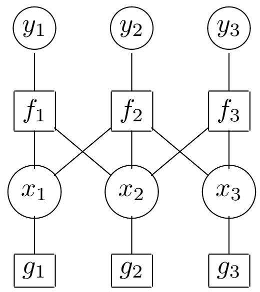

Fig. 1. A Network of Three Sensors and Targets

2. CONVERGENCE OF MESSAGE VARIANCES

Consider three networked sensors where the first sensor y_1 senses targets x_1 and x_2, the second sensor y_2 senses targets x_1, x_2, and x_3, and the third and last sensor y_3 senses targets x_2 and x_3. The graph in Figure 1 models the sensor network. Associated with each sensor node f_i is a soft decision p_i^* such that $p_1^* = p(x_1, x_2|y_1)$, $p_2^* = p(x_1, x_2, x_3|y_2)$, and $p_3^* = p(x_2, x_3|y_3)$. Assume that the soft decisions are Gaussian densities, $p_i^* = \mathcal{N}(\mu_i, \Sigma_i)$, with statistics whose dimension corresponds to the number of edges incident on the sensor node. Associated with each target node x_i is a univariate Gaussian density raised to a power equal to the target node's number of edges minus one. For example, if $p(x_2) = \mathcal{N}(m_2, \sigma_2^2)$, then the density at target node x_2 is $p(x_2)^2 = \mathcal{N}(m_2, \frac{1}{2}\sigma_2^2)$.

As stated earlier, fusing the soft decisions is equivalent to finding the marginals $p(x_i|y_1, y_2, y_3)$ for $i = 1, 2, 3$. Belief propagation attempts to obtain the marginals without calculating the joint density $f = p(x_1, x_2, x_3|y_1, y_2, y_3)$. We denote the joint prior density as $p_0 = p(x_1, x_2, x_3)$.

To facilitate analysis, we generalize and extend the notation and operators used in [1] to our sensor network. Let $q_i^{(k)}(x_j)$ be the message sent from sensor node f_i to target node x_j in the k^{th} iteration. Also, let $q_i^{(k)}$ denote the product of messages sent in the k^{th} iteration from sensor node f_i to its connected target nodes. For example, the collective message sent from node f_1 is $q_1^{(k)} = q_1^{(k)}(x_1)q_1^{(k)}(x_2)$.

From the dynamics of belief propagation, we can relate individual messages in the next iteration to messages in the current iteration. For example,

$$q_2^{(k+1)}(x_2) \quad = \quad \int p_2^* \frac{q_1^{(k)}(x_1)}{p(x_1)} \frac{q_3^{(k)}(x_3)}{p(x_3)} dx_1 dx_3 \quad (1)$$

We introduce the operators π and α in order to write expressions for the collective messages $q_i^{(k)}$ concisely. Again, we follow the notation used in [1] whenever possible. Let π be an operator on a density h that factors a joint density into the product of its marginals. When we apply the operator π on the joint density $f = p(x_1, x_2, x_3|y_1, y_2, y_3)$, we obtain the factored density $\pi(f) = \prod_{i=1}^{3} p(x_i|y_1, y_2, y_3)$. If the joint density f is Gaussian, and $f = \mathcal{N}(\mu, \Sigma)$, then $\pi(f) = \mathcal{N}(\mu, \delta(\Sigma))$, where $\delta(\Sigma)$ denotes a diagonal matrix with the diagonal elements of Σ. In addition, let α be a normalization operator on a density h such that $\alpha h = \frac{h}{\int h(\bar{x}) d\bar{x}}$.

We proceed to define the new operator λ_{ij} not found in [1] in the following paragraph and explain why we need it in the context of the network in Figure 1.

To begin with, from the network of Figure 1, messages sent by sensor node f_3 to its target nodes rely on the messages sent from other sensor nodes to targets x_2 and x_3. We define an operator λ_{ij} that restricts all messages sent by a sensor node f_j to only those messages sent to target nodes shared with another sensor node f_i. For clarity, we interpret the action of λ_{32} on collective message $q_2^{(k)}$ sent by sensor node f_2 to all target nodes x_1,x_2, and x_3. The product of messages sent from sensor node f_2 only to target nodes shared with sensor node f_3 is $\lambda_{32}\left(q_2^{(k)}\right) = \lambda_{32}\left(q_2^{(k)}(x_1)q_2^{(k)}(x_2)q_2^{(k)}(x_3)\right) = q_2^{(k)}(x_2)q_2^{(k)}(x_3)$. Generally, λ_{ij} operates on a density h and keeps only the variables of h that belong to a set of variables shared between densities p_i^* and p_j^*; λ_{ij} marginalizes out the other variables. As another example, if $h = p_0 = p(x_1, x_2, x_3)$, then $\lambda_{12}(h) = p(x_1, x_2)$. If h is Gaussian, the operator λ_{ij} reduces the dimensions of the statistics of h.

With the extended and new operators, we can generically express messages sent in the next iteration from any sensor node f_i to its connected target nodes x_i in terms of the current messages sent from other sensor nodes as

$$q_i^{(k+1)} = \alpha \left(\pi \left(p_i^* \prod_{j \neq i} \frac{\lambda_{ij}\left(q_j^{(k)}\right)}{\lambda_{ij}(p_0)} \right) \prod_{j \neq i} \frac{\lambda_{ij}(p_0)}{\lambda_{ij}\left(q_j^{(k)}\right)} \right) \quad (2)$$

for all $i = 1, 2, 3$.

Ultimately, we want to show that the messages converge to a set of fixed messages as the number of iterations increases, that is, $\lim_{k\to\infty} q_i^{(k)} = q_i^*$, under reasonable assumptions. The symbol $*$ in this paper means a fixed point. First, we assume that the joint prior density $p_0 = p(x_1, x_2, x_3)$ in our example is a multivariate standard normal density, that is $p_0 = p(x_1)p(x_2)p(x_3) = \mathcal{N}(\mathbf{0}, I_0)$, where I_0 is the identity matrix of dimension 3. Also, we assume that the covariance matrices of the soft-decision densities Σ_i for all $i = 1, 2, 3$ are positive definite and contain enough information about the targets, that is, the matrices $\Sigma_i^{-1} - \sum_{j=1, j \neq i}^{3} \xi_i(\lambda_{ij}(I_0))$ are positive definite, too. Notation ξ_i preserves dimensionality and allows us to subtract the matrices $\lambda_{ij}(I_0)$ from Σ_i^{-1}. Sensors' local processors can generate covariances that satisfy these assumptions.

With these assumptions, we can relate the statistics of the messages sent from any sensor node in the k^{th} iteration to the statistics of messages sent from all other sensor nodes through the expression given in (2). As stated earlier, we can bundle together the messages sent from any sensor node to its connected target nodes as $q_i^{(k)}$. For sensor node f_1 in Figure 1, the message bundle sent from f_1 to x_1 and x_2 at the k^{th} iteration is $q_1^{(k)} \sim \mathcal{N}\left(M_1^{(k)}, C_1^{(k)}\right)$, where $M_1^{(k)}$ is a 2×1 mean vector containing the message means and $C_1^{(k)}$ is a 2×2 diagonal covariance matrix containing the message variances.

The covariance of $q_1^{(k)}$ in the next iteration, $C_1^{(k+1)}$, is a function of certain elements in the other two covariances $C_2^{(k)}$ and $C_3^{(k)}$, which are the current message variances sent from sensor nodes f_2 and f_3 to their connected target nodes, respectively. $C_1^{(k+1)}$ depends only on the messages of f_2 and f_3 sent to x_1 and x_2. We represent this mathematically with the function $\mathcal{F}_i$ similar to the one used in [1]. Thus, $C_1^{(k+1)} = \mathcal{F}_1\left(\lambda_{12}\left(C_2^{(k)}\right), \lambda_{13}\left(C_3^{(k)}\right)\right)$.

More generally, function $\mathcal{F}_i$ with covariance matrices $\lambda_{ij}\left(C_j^{(k)}\right), j \neq i$ as arguments returns a covariance matrix $C_i^{(k)}$ embodying the message variances sent by sensor node f_i to its connected target nodes. Using equation (2), we can give a closed-form expression for this function in terms of the covariance matrices $\lambda_{ij}\left(C_j^{(k)}\right)$, and the covariance matrices $\lambda_{ij}\left(I_0\right)$), where I_0 is the covariance matrix of the joint prior density p_0 of variables x_1, x_2, and x_3.

In like manner, we can relate all the message means of a sensor node f_i in the next iteration to the other sensor nodes' message variances and message means in the current iteration by relating $M_i^{(k+1)}$ to all $M_j^{(k)}$ and $C_j^{(k)}$, $j \neq i$. Again, we represent this mathematically with a function $\mathcal{H}_i$ similar to the one used in [1]. For example, the message means of node f_1 in Figure 1 would be expressed as $M_1^{(k+1)} =$

$$\mathcal{H}_1\left(\lambda_{12}\left(M_2^{(k)}\right), \lambda_{12}\left(C_2^{(k)}\right), \lambda_{13}\left(M_3^{(k)}\right), \lambda_{13}\left(C_3^{(k)}\right)\right),$$

which could be expanded out like the function $\mathcal{F}_1$. To preserve clarity and conserve notation, we leave out the closed-form expressions for $\mathcal{F}_i$ and $\mathcal{H}_i$.

Although the message means depend on both the message variances and means, the message variances do not depend on the message means. In our example, we group the covariance matrices containing the message variances into a three-tuple $\left(C_1^{(k+1)}, C_2^{(k+1)}, C_3^{(k+1)}\right)$. This three-tuple is a point in $\mathcal{D}^3 = \mathcal{D} \times \mathcal{D} \times \mathcal{D}$, where $\mathcal{D}$ is the set of positive-definite diagonal matrices. As each coordinate of this point depends on the other coordinates in each iteration of belief propagation, we concisely express this as $\left(C_1^{(k+1)}, C_2^{(k+1)}, C_3^{(k+1)}\right)$
$= \left(\mathcal{F}_1\left(\cdots\right), \mathcal{F}_2\left(\cdots\right), \mathcal{F}_3\left(\cdots\right)\right) = \mathcal{F}\left(C_1^{(k)}, C_2^{(k)}, C_3^{(k)}\right).$
The operator $\mathcal{F}$ with superscript k denotes k applications of corresponding equations in $\mathcal{F}_i$ for the covariance matrices $C_i^{(k)}$ given initial matrices $C_i^{(0)}$ for all $i = 1, 2, 3$. Hence,

$$\left(C_1^{(k)}, C_2^{(k)}, C_3^{(k)}\right) = \mathcal{F}^k\left(C_1^{(0)}, C_2^{(0)}, C_3^{(0)}\right). \quad (3)$$

$\mathcal{F}^k$ behaves as an operator on $\mathcal{D}^3$ possessing certain properties, such as continuity and monotonicity. We can show from these properties that $\mathcal{F}^k$ has a unique fixed point. Indeed, the message variances in belief propagation converge uniquely, as summarized in Theorem 1.

Theorem 1 *The operator $\mathcal{F}$ possesses a fixed point in $\mathcal{D}^3$. Furthermore, denoting this fixed point by (C_1^*, C_2^*, C_3^*)*

$$\lim_{k \to \infty} \mathcal{F}^k\left(C_1^{(0)}, C_2^{(0)}, C_3^{(0)}\right) = (C_1^*, C_2^*, C_3^*)$$

for all $C_1^{(0)}, C_2^{(0)}, C_3^{(0)} \in \mathcal{D}$.

Given the assumptions stated earlier, Theorem 1 says that if we initialize the message variances with positive numbers, the message variances converge in belief propagation. The proof of Theorem 1 is lengthy and follows from the above definitions of operators and functions.

3. CONVERGENCE OF MESSAGE MEANS

Theorem 1 says that the message variances unconditionally converge under reasonable assumptions. However, the message means converge conditionally, and we must manipulate and use the closed-form expressions represented by the functions $\mathcal{H}_i$ in order to find a sufficient condition.

Because we have limited space, we briefly describe our procedure to find the sufficient condition. From linear systems theory, we know that given a sequence $y_{k+1} = T_k y_k + b_k$, where T_k converges to T and $\{\lambda_i\}$ are the eigenvalues of T, if $\rho(T) = \max_i |\lambda_i|$, then y_{k+1} converges. $\rho(T)$ is the spectral radius, or largest eigenvalue magnitude, of matrix T. We collect the mean vectors $M_i^{(k)}$ containing the message means sent from sensor node f_i to its target nodes x_i as one large vector y_k to find an equivalent block matrix T from the closed-form expressions of $\mathcal{H}_i$.

As it turns out from the closed-form expressions of $\mathcal{H}_i$, the sufficient condition for the convergence of the mean vectors $M_i^{(k)}$ containing the message means depends on the spectral radius of the block matrix $T_{\Sigma_1,\Sigma_2,\Sigma_3}$, where

$$T_{\Sigma_1,\Sigma_2,\Sigma_3} = \begin{pmatrix} 0 & \Theta_{1*} & \Theta_{1*} \\ \Theta_{2*} & 0 & \Theta_{2*} \\ \Theta_{3*} & \Theta_{3*} & 0 \end{pmatrix}. \quad (4)$$

Each block Θ_{i*} of the partitioned matrix $T_{\Sigma_1,\Sigma_2,\Sigma_3}$ depends on all the covariance matrices C_i^* of the fixed point and the covariance matrix Σ_i of the soft-decision density p_i^*. We summarize the sufficient condition for the convergence of the message means in the following theorem.

Theorem 2 *If $\rho(T_{\Sigma_1,\Sigma_2,\Sigma_3}) < 1$, then $\exists$ vectors M_1^*,M_2^*, M_3^* such that for any vectors $M_1^{(0)}$, $M_2^{(0)}$, $M_3^{(0)}$ and any matrices $C_1^{(0)}$,$C_2^{(0)}$,$C_3^{(0)} \in \mathcal{D}$,*

i. the sequence $(M_1^{(k)}, M_2^{(k)}, M_3^{(k)})$ converges to (M_1^*, M_2^*, M_3^*).

ii. the marginal means estimated from the message statistics are the true marginal means.

This sufficient condition $\rho(T_{\Sigma_1,\Sigma_2,\Sigma_3}) < 1$ can be verfied in practice because Theorem 1 guarantees that the fixed point of the function $\mathcal{F}$ exists. To verify it, we iterate on $\mathcal{F}$, obtain the fixed point covariance matrices C_i^*, and then, compute the Θ_i^*, which lead to matrix $T_{\Sigma_1,\Sigma_2,\Sigma_3}$. If the spectral radius of $T_{\Sigma_1,\Sigma_2,\Sigma_3}$ is less than one, then the message means converge, from which we conclude that the estimated means for the marginal densities converge to the exact means, applying a result in [6].

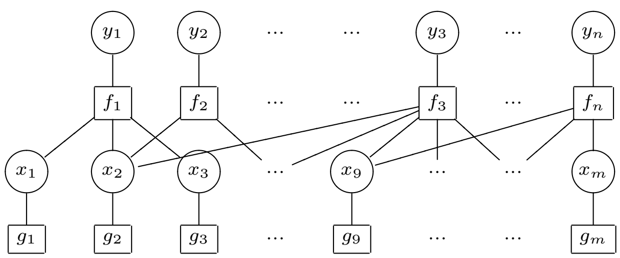

Fig. 2. Sensor Network of General Topology

Arbitrary Topology The degree of a node is the number of edges shared by that node. For example, in Figure 2, the degree of node f_1 is 3 and the degree of node f_2 is 2. We have presented up to now two theorems describing the convergence of the message variances and the message means of the sensor network in Figure 1 for pedagogical reasons. We can in fact show that these results are general and extend to a generic sensor network of arbitrary topology. Contrasting Figure 2 with Figure 1, we see that in the generic topology, we can have n arbitrary sensors monitoring m arbitrary targets. The degree of the sensor nodes and of the target nodes is also arbitrary, that is, the topology is not regular. We note that in this topology, a target can be sensed by several or just a single sensor, as exemplified by target node x_1. The proof for the generic topology entails introducing a few additional operators. Lack of space prevents us from presenting them here.

4. CONCLUSION

We have analyzed data fusion for a partially connected sensor network of three sensor nodes and three target nodes. Specifically, we have shown that we can model the data fusion of the local processors' soft decisions as belief propagation on a factor graph with sensor nodes and target nodes. Furthermore, we have shown that the variances of the messages passed during belief propagation converge unconditionally under reasonable assumptions and that the means of the messages converge to the true means when the spectral radius of a certain block matrix is less than one.

Although we present our convergence results with specific sensor-target interdependencies, our results hold for Gaussian graphs of arbitrary topology. We plan to extend these convergence results even further to Gaussian mixture models and more general probability distributions.

5. REFERENCES

[1] P. Rusmevichientong and B. Van Roy, "An analysis of belief propagation on the turbo decoding graph with gaussian densities," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 745–765, February 2001.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramianiam, and E. Cayirci, "A Survey on Sensor Networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102-114, August 2002.

[3] J. M. F. Moura, J. Lu, and M. Kleiner, "Intelligent sensor fusion: a graphical model approach," *ICASSP'03, IEEE International Conference on Signal Processing*, Hong Kong, April 2003.

[4] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo,CA: Morgan Kaufmann, 1988.

[5] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, February 2001.

[6] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Computation*, vol. 13, pp. 2173–2000, 2001.