COSNTRUCTION OF A SCALABLE DECODER FOR A WIRELESS SENSOR NETWORK USING BAYESIAN NETWORKS

Ruchira Yasaratna and Pradeepa Yahampath[*]
Department of Electrical and Computer Engineering
University of Manitoba, Winnipeg, MB R3T 5V6, Canada
Email:{ ruchira,pradeepa}@ee.umanitoba.ca

ABSTRACT

We consider minimum mean square error (MMSE) decoding in a dense sensor network where distributed quantization is used to improve the performance. In view of the exponential complexity of the optimal decoder, we present a framework based on Bayesian networks for designing a scalable, but near-optimal decoder. In this approach, a complexity-constrained factor graph is obtained by an algorithm which constructs an equivalent Bayesian network using the maximum likelihood (ML) criterion, based on a training set of sensor observations. Our simulation results show that, the scalable decoders constructed using the proposed approach preform close to optimal, with both Gaussian and non-Gaussian sensor data.

Index Terms— distributed quantization, scalable decoding, optimal estimation, factor graphs, Bayesian networks

1. INTRODUCTION

Sensor networks deploying a large number of tiny, battery-powered wireless sensors to acquire measurements from a physical process are envisioned to find wide ranging applications. One of the important issues in this context is *distributed source coding* [1], where the high correlation among the sensors which do not communicate with each other, is exploited at a centralized *joint decoder* for efficient communication of the sensor outputs. In general, the joint-decoder, which is optimal under the widely used MMSE criterion, has a complexity that is exponential in the number of sensors [2]. Thus, to realize the full potential of a large-scale sensor network, it is crucial to develop low-complexity and scalable joint-decoders which yield near-optimal reconstructions of the observed physical process.

In previous work, [2] considers a sensor network with jointly Gaussian observations and proposes a procedure for scalable decoding based on running the sum-product algorithm on an appropriately constructed factor graph [3]. The main idea is to factorize the joint probability mass function (pmf) of the quantized sensor outputs using an approximate chain rule, where each factor is only a function of few variables. The best approximation to the joint pmf, which corresponds to a fully connected factor graph is obtained by minimizing the Kullback-Leibler distance (KLD) between the two pmfs. The factor graph is constructed such that the overall decoding complexity is linear in the number of sensor nodes. The simulations results show that, the proposed decoder yields near optimal performance with Gaussian sources. In a subsequent work [4] incorporates the encoder design also into the method of [2], where the encoders exploit the knowledge of source correlations to reduce their transmission rates, by using binning of the quantizer outputs. In order to keep the complexity low, binning is preformed within clusters of sensor nodes, obtained by using a hierarchical algorithm which forms clusters of highly correlated sources, also based on the KLD. Again, Gaussian source statistics are assumed.

In contrast, this paper presents a more general approach to factor graph construction based on representing a factorized multi-variate pmf using a graphical model referred to as a *Bayesian network* (BN) [5]. The optimal factorization is then obtained by ML estimation of the underlying BN structure using a training set of sensor observations, whose distribution is not necessarily Gaussian. Furthermore, in order to preserve the universality of our approach, the sensor clustering for designing binning-based encoders is carried out using an algorithm which uses the *Chi-square* test [6] to measure the degree of dependence between a pair of sensors. The effectiveness of the proposed design procedure is investigated using simulation experiments. In experiments with Gaussian sources, it was observed that the proposed method yielded a system which performs identically to the system obtained by the KLD-based method. Furthermore, it was also observed that for non-Gaussian sources (such as Gaussian mixtures), the proposed design method yielded systems with significantly better performance compared to the KLD-based method.

2. PROBLEM STATEMENT

In the system with N sensors shown in Fig.1, each sensor observes a real valued continuous source sample X_n, $n = 1, \ldots, N$ (in the following, we will use upper case to denote random variables, lower case to denote the realizations, and bold-face to denote random vectors). In view

[*]This work has been funded by NSERC of Canada.

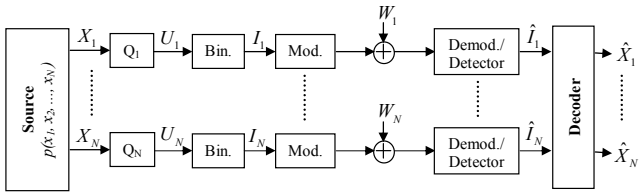

Fig. 1: System model for a sensor network with N nodes.

of the low-complexity requirement, each sensor is assumed to consist of a Lloyd-Max scalar quantizer and a binning function. In the n-th sensor node, X_n is first quantized into $U_n \in \{1, \ldots, 2^{B_n}\}$, where B_n is the resolution (in bits) of the quantizer Q_n, The quantizer output U_n is then mapped to the transmission index $I_n \in \Omega = \{1, \ldots, 2^{R_n}\}$ by using an appropriate binning function, achieving a transmission rate of $R_n \leq B_n$ bits/sample. Each sensor transmits its output (using a suitable modulation scheme) to a central decoder via an interference-free, memoryless additive white Gaussian noise (AWGN) channel. Let the output of the n-th channel after demodulation/detection be $\hat{I}_n \in \{1, \ldots, 2^{R_n}\}$, $n = 1, \ldots, N$. Then it follows that $P(\hat{\mathbf{I}}|\mathbf{I}) = \prod_{n=1}^{N} P(\hat{I}_n|I_n)$, where $\mathbf{I} = (I_1, \ldots, I_n)^T$, $\hat{\mathbf{I}} = (\hat{I}_1, \ldots, \hat{I}_n)^T$, and $P(\hat{I}_n|I_n)$ is the transition probability of the effective discrete channel from the n-th sensor to the decoder.

For simplicity, let us first assume that binning is not used, so that $I_n = U_n$. Let the reconstruction codebook of Q_n be $\{\tilde{x}_n(i_n), i_n \in \Omega\}$. Given the decoder input vector $\hat{\mathbf{i}}$, we define the optimal decoder output of the n-th sensor observation as

$$\hat{x}_n^*(\hat{\mathbf{i}}) = \arg\min_{\hat{x}_n} E\big\{[\tilde{x}_n(I_n) - \hat{x}_n(\hat{\mathbf{i}})]^2\big\}. \quad (1)$$

It is well known that the optimal solution is the conditional mean

$$\begin{aligned} \hat{x}_n^*(\hat{\mathbf{i}}) &= E\{\tilde{x}(I_n)|\hat{\mathbf{i}}\} = \sum_{\forall i \in \Omega} \tilde{x}(i) P(i_n = i|\hat{\mathbf{i}}) \\ &= \frac{1}{P(\hat{\mathbf{i}})} \sum_{\forall i \in \Omega} \tilde{x}(i) \cdot \sum_{\forall \mathbf{i} \in \Omega^N; i_n = i} P(\mathbf{i}) \prod_{k=1}^{N} P(\hat{i}_k|i_k). \quad (2) \end{aligned}$$

This expression can be systematically evaluated by running the sum-product algorithm on a factor graph based on the chain-rule factorization $\prod_{n=1}^{N} P(i_n|i_{n-1}, i_{n-2}, \ldots)$ of the joint pmf $P(\mathbf{i})$ [3]. However, the computational complexity of this task is exponential in N [2]. The complexity can be controlled by using an approximate factorization $\tilde{P}(\mathbf{i}) = \prod_{m=1}^{M} f_m(.)$, where the factors $f_m(\cdot)$ are conditional probabilities involving only a few variables such that a fully connected factor graph is obtained. In the following, we present an approach to constructing such a factorization which ensures a computational complexity linear in N.

3. BAYESIAN NETWORKS

A Bayesian network (BN) [5] is a probabilistic graphical model which can be used to compactly represent the joint pmf of a large number of random variables which exhibit conditional dependencies. That is, a given factor graph can also be represented by a BN. We can thus construct a factor graph with desired complexity by constructing an appropriate BN. More formally, BN $\Lambda(\mathbf{I})$ is a directed graph with N nodes, in which each node represents a random variable and a directed edge (or the lack of an edge) between two nodes indicates a conditional dependence (or independence) between two random variables. Suppose that a training set $\mathcal{S}$ of observations from the true pmf $P(\mathbf{I})$ is available. Then, one approach to obtaining a factorization of $P(\mathbf{I})$ is to start with a BN with no edges, i.e., all variables assumed independent, and then to systematically add edges between nodes. In order to ensure that the resulting pmf is a good approximation to $P(\mathbf{I})$, at each step t, we obtain a new BN $\Lambda_t(\mathbf{I})$, by adding a new edge to $\Lambda_{t-1}(\mathbf{I})$, such that the increase in the likelihood $P(\mathcal{S}|\Lambda_t(\mathbf{I}))$ is maximum. While, this greedy search procedure does not necessarily yield the BN with the overall ML, it is chosen in view of it's computational feasibility. Denote the vector of parent nodes of I_n in $\Lambda(\mathbf{I})$ by $\mathbf{Z}_n \in \{\mathbf{z}_{n1}, \ldots, \mathbf{z}_{nq_n}\}$, $n = 1, \ldots, N$. Let the set of quantization levels of the n-th node be $\{v_{n1}, \ldots, v_{nr_n}\}$. Also, let L_{njl} be the number of times $I_n = v_{nl}$ and $\mathbf{Z}_n = \mathbf{z}_{nj}$ jointly occurred in the training set $\mathcal{S}$. Then, it can be shown that [7, appendix]

$$P(\mathcal{S}|\Lambda) = \prod_{n=1}^{N} \prod_{j=1}^{q_n} \frac{(r_n - 1)!}{(L_{nj} + r_n - 1)!} \prod_{l=1}^{r_n} L_{njl}! \quad (3)$$

where $L_{nj} = \sum_{l=1}^{r_n} L_{njl}$. Rather than starting with the BN corresponding to the case of all independent variables, a better method is to first determine a set of factors according to the degree of statistical dependence among variables. Based on this idea, we propose a two step procedure for constructing a BN to approximate $P(\mathbf{I})$ as follows.

4. DESIGN PROCEDURE

Step 1: Clustering of sensor nodes to obtain an initial factorization for $P(\mathbf{I})$. The grouping of highly dependent sensors in this manner also allows us to design effective binning functions to be used for distributed encoding, in a simplified manner, i.e., for the optimization of binning functions, only those sensors in a cluster are considered.

Step 2: Linking the clusters using the ML-based BN construction to obtain a factor graph which allows us to compute a near-optimal approximation to (2) at a complexity linear in the network size N.

An example of the factor graph obtained by using the proposed procedure is shown in Fig. 2. In this case, Step 1 has been used to form 3 clusters Γ_1, Γ_2, and Γ_3, and Step 2 has then been used to obtain the dependencies $f_1, \ldots, f_5$ between the variables, as shown.

A. Sensor clustering algorithm - The basic approach is to start with an undirected graph in which each node is connected to every other node, and then systematically eliminate

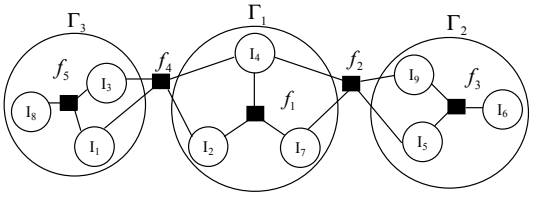

Fig. 2: An example of a factor graph constructed by the proposed approach, for a sensor network with $N = 9$ nodes. The clusters Γ_1, Γ_2, and Γ_3 have been obtained by the given clustering algorithm, with $\mathcal{Z}_{max} = 3$ and $\mathcal{Z}_{min} = 2$. Note that the function nodes correspond to $f_1 = P(I_2, I_4, I_7)$, $f_2 = P(I_5, I_9|I_4, I_7)$, $f_3 = P(I_6|I_5, I_9)$, $f_4 = P(I_1, I_3|I_2, I_4)$, and $f_5 = P(I_8|I_3, I_1)$, and the factor graph corresponds to $\tilde{P}(\mathbf{I}) = f_1.f_2.f_3.f_4.f_5$.

some edges, based on a test of statistical independence of random variables associated with the corresponding node-pairs, until we are left with M unconnected clusters of nodes. Let the maximum and minimum number of nodes to be included in each sensor cluster be $\mathcal{Z}_{\max}$ and $\mathcal{Z}_{\min}$ respectively. On the one hand, it is necessary to limit the cluster size in this manner, in order to control the *degree* d of the final factor-graph. The degree of a factor graph is the maximum number of incidents edges of a function node, and it can be shown that the computational complexity of the sum-product algorithm is exponential in d. Note that, limiting d also simplifies the design of binning functions to be used in the encoders. On the other hand, the lower limit on the cluster size ensures that the correlation among a sufficiently large number of sensors is exploited by each encoder through binning.

As a measure of statistical independence between two discrete random variables I_j and I_k, we use the *Chi-square* test statistic [6]

$$\Upsilon^2_{j,k} = \sum_{i_j, i_k} \frac{[\tilde{P}(i_j, i_k) - \tilde{P}(i_j)\tilde{P}(i_k)]^2}{\tilde{P}(i_j)\tilde{P}(i_k)}, \qquad (4)$$

where $\tilde{P}(i_j, i_k)$, $\tilde{P}(i_j)$ and $\tilde{P}(i_k)$ are the estimates (using training set $\mathcal{S}$) of the joint probability and marginal probabilities of $I_j = i_j$ and $I_k = i_k$. In each iteration of the algorithm, the edge corresponding to a pair of variables with the lowest Υ^2 is eliminated, while ensuring that existing clusters which satisfy both size constraints are not further split. This process is repeated until each remaining cluster contains less than or equal to $\mathcal{Z}_{\max}$ sensor nodes.

B. Factor graph construction using Bayesian networks- Let the set of clusters obtained in the previous step be denoted by $\mathbf{\Gamma}_m$, $m = 1, \ldots, M$. Since there are no links between the clusters, this is a BN which corresponds to the factorization

$$\tilde{P}(\mathbf{I}) = \prod_{m=1}^{M} P(\mathbf{\Gamma}_m). \qquad (5)$$

Our goal is to link each cluster to exactly one other cluster and thereby form a fully connected factor graph, by introducing conditional dependencies between variables in different clusters (see Fig. 2). This is achieved by choosing those conditional dependencies which result in the BN with the highest value of likelihood (3), by using the greedy search procedure outlined in Sec. 3. In each iteration t, the current BN Λ_{t-1} is modified by linking an isolated cluster $\mathbf{\Gamma}_l$ to one of the already connected clusters, such that this results in the largest increase of the likelihood value (3) of the modified BN Λ_t. However, rather than considering the dependency between pairs of variables, we consider the dependency given by the conditional probability $P(\boldsymbol{\alpha}_l|\boldsymbol{\beta}_m)$ of the set of nodes $\boldsymbol{\alpha}_l \subset \mathbf{\Gamma}_l$, given the set of nodes $\boldsymbol{\beta}_m \subset \mathbf{\Gamma}_m$. In order to control the complexity of this task, we specify maximum sizes $|\boldsymbol{\alpha}_l| \leq \lambda$ and $|\boldsymbol{\beta}_m| \leq \mu$ for the two sets $\boldsymbol{\alpha}_l$ and $\boldsymbol{\beta}_m$ respectively. Note that introducing a conditional dependency into the current BN in the above manner is equivalent to replacing the factor $P(\mathbf{\Gamma}_l)$ in (5) by the new factor

$$\begin{aligned} P(\mathbf{\Gamma}_l|\boldsymbol{\beta}_m) &= P(\alpha_l|\boldsymbol{\beta}_m)P(\{\mathbf{\Gamma}_l - \boldsymbol{\alpha}_l\}|\boldsymbol{\alpha}_l, \boldsymbol{\beta}_m) \\ &\approx P(\alpha_l|\boldsymbol{\beta}_m)P(\{\mathbf{\Gamma}_l - \boldsymbol{\alpha}_l\}|\boldsymbol{\alpha}_l), \qquad (6) \end{aligned}$$

where $\{\mathbf{\Gamma}_l - \boldsymbol{\alpha}_l\}$ denotes the set of those nodes in $\mathbf{\Gamma}_l$ which excludes the set $\boldsymbol{\alpha}_l$, and the last approximation is reasonable since the variables within a given cluster are known to be highly correlated. For example, consider the factor graph shown in Fig. 2 (where $\lambda = \mu = 2$). In this case, the cluster $\mathbf{\Gamma}_2 = \{I_5, I_6, I_9\}$ is linked to the cluster $\mathbf{\Gamma}_1 = \{I_2, I_4, I_7\}$ to obtain the factor $P(\mathbf{\Gamma}_2|\boldsymbol{\beta}) = P(I_5, I_9|I_4, I_7)P(I_6|I_5, I_9)$, where $\boldsymbol{\alpha} = \{I_5, I_9\}$, $\boldsymbol{\beta} = \{I_4, I_7\}$, and $\{\mathbf{\Gamma}_2 - \boldsymbol{\alpha}\} = \{I_6\}$. After an appropriate chain rule expansion of the two conditional probabilities in (6), they can be introduced into the BN [5].

5. SIMULATION RESULTS

In order to investigate the performance of scalable decoders constructed with the proposed algorithm, we conducted a number of simulation experiments involving both Gaussian and non-Gaussian sensor observations. In all our experiments with jointly Gaussian sensor observations, the factor-graphs constructed with the proposed approach and the KLD-based approach in [2] and [4] were found to be identical, and hence the performance of the resulting decoders was also identical. However, when the sensor observations were non-Gaussian, the proposed approach lead to designs which were significantly better.

In our experiments with non-Gaussian sources, the correlated observations were generated as follows. Let $\mathbf{Y} = (Y_1, \ldots Y_N)$ be a vector of independent variables, with Y_n having the Gaussian mixture probability density function (pdf) $f_n(y) = \gamma_n \mathcal{N}(y; \mu_{n,1}, \sigma^2_{n,1}) + (1-\gamma_n)\mathcal{N}(y; \mu_{n,2}, \sigma^2_{n,2})$, where $\mathcal{N}(y; \mu, \sigma^2)$ denotes a Gaussian pdf with mean μ and variance σ^2, and $0 < \gamma_n < 1$, $n = 1, \ldots, N$. Then, a vector of sensor observations $\mathbf{X} = (X_1, \ldots, X_N)$ is obtained by the linear transformation $\mathbf{X} = \mathbf{AY}$, where $\mathbf{A}$ is a $N \times N$ constant matrix. We here choose an example with $N = 9$ sensors, in which case it is computationally feasible to estimate the performance of the optimal decoder for comparison. A training set of 100,000 samples was used.

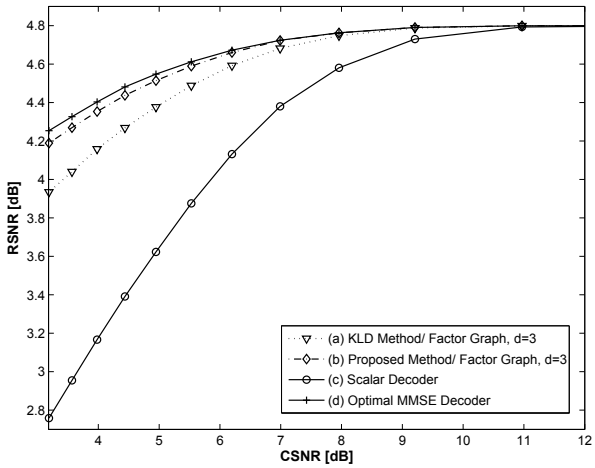

Fig. 3: Performance of several decoders as a function of CSNR. The network has $N = 9$ sensing nodes and each node uses rate 1-bit/sample quantization, without subsequent binning. *Scalar decoder* refers to independent MMSE decoding of sensor observations.

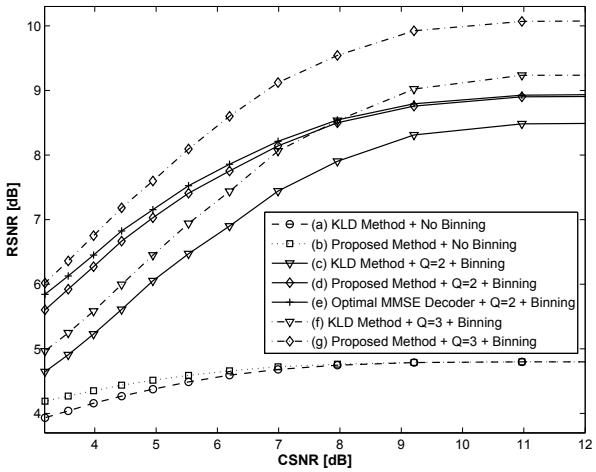

Fig. 4: Performance of various decoders when binning-based distributed quantizers are used in sensors. Each sensor uses a quantizer rate of Q bits/sample and a transmission rate of 1 bit/sample. The curves (a) and (b) are the same as those in Fig. 3. For the system with $Q = 2$, the performance of the optimal decoder is shown in curve (e).

First, we consider the performance of joint decoding when sensors do not use correlation for distributed quantization, i.e., no binning used. Each sensor uses a transmission rate of 1 bit/sample. Fig. 3 shows the reconstruction signal-to-noise ratio (RSNR) of several decoding schemes as a function of the channel signal-to-noise ratio (CSNR). In the case of both KLD-based method and the proposed method of design, the degree d of the factor graph is limited to 3. Fig. 3 also shows, for comparison, the performance of the decoder which optimally reconstructs each sensor output independently of the others (scalar decoder) and the optimal joint decoder. As can be seen, the proposed method gives a decoder which is clearly better. The key difference is that the approximated factorization of the joint pmf constructed by the proposed method is closer to the true joint pmf than that constructed by the KLD-based method.

Next, we investigate the performance of joint decoding when binning is used in each sensor to reduce its transmission bit rate. In this case, the optimal binning function to be used at the output of each quantizer has been designed by the simulated annealing based procedure described in [8]. Recall that, for binning, only those sensors within each cluster (found by the Step 1 of the design algorithm in Sec. 4) are considered. The performance of several systems is compared in Fig. 4. As in the previous case, each sensor uses a transmission rate of 1 bit/sample. However, in order to benefit from binning, each quantizer now uses a rate of $Q > 1$. In Fig. 4, we consider systems with $Q = 2$ and $Q = 3$ bits/sample. The performance improvement due to the use of distributed quantization is of course clear. But more importantly, the decoders designed with the proposed method perform significantly better than those designed with the KLD-based method. In the case of $Q = 2$, we also include the performance of the optimal decoder for comparison, which shows that the performance of the proposed design is within less than 0.25 dB in the entire CSNR range (with $Q = 3$, estimating the optimal decoder performance was computational infeasible). In contrast, the KLD-based design performs more than 1.0 dB below at low CSNR.

6. CONCLUDING REMARKS

The Bayesian networks present an effective framework for constructing factor-graphs for scalable decoding in large-scale sensor networks. Within this framework, it is also quite possible to incorporate models such as Markov random fields or other local dependency structures to further simplify the decoder.

7. REFERENCES

[1] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Mag.*, pp. 80–94, Sept. 2004.

[2] J. Barrõs M. Tüchler, "Scalable decoding on factor trees: A practical solution for wireless sensor networks," *IEEE Trans. Commun.*, vol. 54, no. 2, pp. 284–294, Feb. 2006.

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

[4] G. Maierbacher and J. Barrõs, "Source-optimized clustering for distributed source coding," in *Proc. IEEE GLOBECOM*, 2006, pp. 1–5.

[5] F. V. Jensen, *An Introduction to Bayesian Networks*, Springer, 1996.

[6] Y. M. M. Bishop, S. E. Feinberg, and P.W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, 1975.

[7] F. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309–347, Oct. 1992.

[8] P. Yahampath, "Predictive vector quantizer design for distributed source coding," in *Proc. IEEE ICASSP*, Apr. 2007, pp. III–629–III–632.