CHANNEL OPTIMIZED BINARY QUANTIZERS FOR DISTRIBUTED SENSOR NETWORKS

Biao Chen

Syracuse University
Department of EECS
Syracuse, NY 13244

Peter K. Willett

University of Connecticut
Department of ECE
Storrs, CT 06269-1157

ABSTRACT

Distributed binary quantizer design for sensor nets tasked with a hypothesis testing problem is considered in this paper. Allowing for non-ideal transmission channels, we show that under the conditional independence assumption, the optimum binary quantizer, in the sense of minimizing the error probability, should operate on the likelihood ratio (LR) of the local sensor observations. Necessary conditions for optimality are derived to facilitate finding of optimal LRT thresholds through an iterative algorithm. A design example with binary symmetric channels between local sensors and the fusion center is given to illustrate how the results can be applied in sensor signaling design.

1. INTRODUCTION

The emerging wireless sensor network (WSN) technologies have spurred enormous interest from various research communities. The potential of collecting and jointly processing spatially and temporally distributed information will immensely enhance our ability to understand and evaluate complex systems and environments. Current and future applications include battlefield surveillance, telemedicine, habitat and environmental monitoring and control.

One of the challenges facing many envisioned applications is to communicate locally collected data to a decision making center under strict bandwidth/resource/delay constraints. Data compression at local sensors is inevitable. Motivated by the Slepian-Wolf theorem [1], much of current effort has focused on distributed data compression that exploits the correlation structure among sensor observations [2, 3]. These approaches are largely data centric – the primary goal is to recover the original sensor data at the decision center (also called fusion center, decoder, etc.). For many WSN, the ultimate goal is successful assessment of a certain situation instead of recovering the original data observed at local sensors. A data-centric approach without regard to the underlying inference task will undoubtedly waste significant resources.

In this paper we explore the optimal distributed data compression scheme that caters to a specific inference task, namely distinguishing between two hypotheses. Examples include the detection of a target in surveillance applications and of a hazardous event for environment and security applications. To simplify our analysis and ease our presentation, we assume that a binary local sensor output is desired. A distinctive feature of this work, in addition to being inference-centric, is that the developed binary quantizer is channel aware. Transmission channel characteristics between the local sensors and the fusion center are integrated in the design of local quantizers. Thus the result can be perceived as channel optimized distributed binary quantizers. It thus leads to a promising direction for the applications of joint source channel codes (JSCC) in distributed sensor networks. We note that for point-to-point communications, source quantization and channel error protection have been jointly considered for zero memory signaling (see the early work in [4, 5]). The discussions and techniques developed there, while certainly enlightening, cannot be directly applied to the WSN applications because of the distributed nature. Further, the techniques are designed on the premise that the goal is to recover the original observation (hence the MSE criterion used in coder/decoder design) rather than for any underlying inference problems. Separately in [6], quantization for decentralized hypothesis testing was considered. Jointly optimal quantizer design (in the sense of maximizing the Bhattacharyya distance) at all sensors are obtained via the Lloyd algorithm. The channel effect is, however, idealized in that it only imposes rate constraints.

We show that, under the conditional independence assumption among sensor observations, the optimal binary quantizers amount to a likelihood ratio test (LRT), i.e., the optimal quantization happens on the likelihood ratio (LR) domain rather than the original observation [1]. Perceived from a different viewpoint, our work establishes the optimality of LRT at the local sensors in the presence of non-ideal transmission channels. Notice that such optimality has been established without the consideration of transmission channels [7–9], i.e., the local sensor binary decisions were assumed to be fully accessible at the fusion center.

[1] The latter is true if the distributions under test have monotone likelihood ratios.

The design of local decision rules accounting for possible channel errors has been addressed under the Neyman-Pearson (NP) criterion [10]. There, optimality of the LRT was established under a simple binary symmetric channel (BSC) model between each sensor and the fusion center. In this paper, we assume a general vector channel from the local sensors to the fusion center and investigate the optimality of the LRT for local sensor decisions. The Bayesian criterion is adopted to minimize the error probability at the fusion center. The paper is organized as follows. In the next section, we introduce the problem formulation. In section 3, we present our main result (Theorem 1), along with some discussions and a sketch of proof. A simple design example is given in section 4 and we conclude in section 5.

2. PROBLEM FORMULATION

Consider a hypothesis testing problem with distributed sensors collecting conditionally independent observations, i.e.,

$$p(X_1, \cdots, X_K|H_i) = \prod_{k=1}^{K} p(X_k|H_i) \tag{1}$$

where $i = 0, 1$ and X_k's are local sensor observations, hence K is the total number of sensors. We further assume that the priors on the two hypotheses are given by $\pi_0 = P(H_0)$ and $\pi_1 = P(H_1) = 1 - \pi_0$, respectively. Each local sensor quantizes its observation X_k to one bit:

$$U_k = \gamma_k(X_k).$$

Each U_k is then sent to a channel characterized by $p(Y_k|U_k)$ where Y_k is observed at the fusion center. Hence we have parallel independent channels connecting the sensors to the fusion center, i.e.,

$$P(Y_1, \cdots, Y_K|U_1, \cdots, U_K) = \prod_{k=1}^{K} p(Y_k|U_k) \tag{2}$$

The fusion center is assumed to implement the optimum fusion rule based on the channel output $\mathbf{Y} = [Y_1, \cdots, Y_K]^T$:

$$U_0 = \gamma_0(Y_1, \cdots, Y_K).$$

Thus a decision error happens if U_0 differs from the true hypothesis. Prominent examples of the channel $p(Y_k|U_k)$ include the binary symmetric channel [10] and the fading channel model [11]. A simple diagram illustrating the above fusion network is given in Fig. 1.

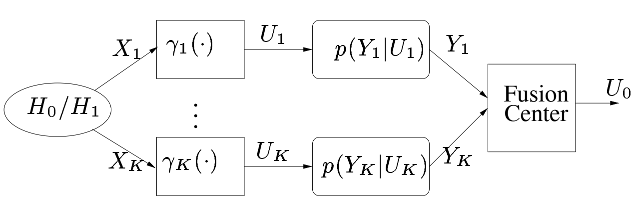

Fig. 1. A block diagram for a wireless sensor network tasked with binary hypothesis testing in the presence of non-ideal transmission channels.

3. OPTIMUM BINARY QUANTIZER DESIGN

We use person-by-person optimization (PBPO) which optimizes the binary quantizer $\gamma_k(X_k)$ for the k^{th} local sensor given fixed quantizers for all other sensors as well as the fusion (decoding) rule. Given the hierarchical structure as in Fig. 1, we have the following result.

Theorem 1 *Assume that the local observations, X_k's, are conditionally independent and that the channels between sensors and the fusion center are characterized by*

$$p(Y_1, \cdots, Y_K|U_1, \cdots, U_K) = \prod_{k=1}^{K} p(Y_k|U_k) \tag{3}$$

If the fusion rule and the k^{th} local decision satisfy

$$P(U_0 = 1|\mathbf{y}^k, U_k = 1) - P(U_0 = 1|\mathbf{y}^k, U_k = 0) \geq 0 \tag{4}$$

$$P(U_0 = 0|\mathbf{y}^k, U_k = 0) - P(U_0 = 1|\mathbf{y}^k, U_k = 1) \geq 0 \tag{5}$$

where $\mathbf{y}^k = [y_1, \cdots, y_{k-1}, y_{k+1}, \cdots, y_K]$, with y_k being the observation of the channel output corresponding to the k^{th} sensor, then the optimum local binary quantizer for the k^{th} sensor amounts to the following LRT

$$U_k = \gamma_k(X_k) = \begin{cases} 1 & \text{if } \frac{p(x_k|H_1)}{p(x_k|H_0)} > \frac{\pi_0}{\pi_1}\eta \\ 0 & \text{if } \frac{p(x_k|H_1)}{p(x_k|H_0)} < \frac{\pi_0}{\pi_1}\eta \end{cases} \tag{6}$$

where η is defined as

$$\frac{\int_{\mathbf{y}^k}[P(U_0 = 1|\mathbf{y}^k, U_k = 1) - P(U_0 = 1|\mathbf{y}^k, U_k = 0)]p(\mathbf{y}^k|H_0)d\mathbf{y}^k}{\int_{\mathbf{y}^k}[P(U_0 = 0|\mathbf{y}^k, U_k = 0) - P(U_0 = 0|\mathbf{y}^k, U_k = 1)]p(\mathbf{y}^k|H_1)d\mathbf{y}^k}$$

Prior to presenting a proof, some remarks are in order.

- Conditions (4) and (5) amount to requiring that the fusion output, conditioned on the same set of channel outputs except for that of the k^{th} sensor, is more likely to be consistent with the k^{th} sensor output than to contradict it. Thus loosely speaking, this amounts to employing a "monotone" fusion rule, hence (4) and (5) are easily satisified by any sensible design.

- The PBPO approach implies that the obtained result is a necessary, not sufficient, condition for optimality. Multiple initializations may be needed to obtain global optimum.

- The test described in (7) for the k^{th} sensor is clearly coupled with the fusion rule as well as all the other sensors' local decision rules. Thus, applying Theorem 1 requires iteration among all the local sensor decision rules and the fusion rule.

- A close inspection of $\gamma_k(\cdot)$ in (7) also shows that the threshold for the LRT is directly affected by the transmission channels (embedded in both $P(U_0|\mathbf{y}^k, U_k)$ and $p(\mathbf{y}^k|H_i)$), indicating that the optimal sensor processing needs to be channel informed.

Sketch of proof

Expanding the error probability with respect to the quantizer rule at the n^{th} sensor using the hierarchical probability structure specified in Fig. 1, we can get, after some algebra,

$$P_{e0} = \int_{x_k} P(u_k = 1|x_k) \left[\pi_0 p(x_k|H_0)A - \pi_1 p(x_k|H_1)B\right] dx_k + C \quad (7)$$

where

$$A = \int_{\mathbf{y}} P(u_0 = 1|\mathbf{y}) \left[\sum_{\mathbf{u}^k} p(\mathbf{y}|\mathbf{u}^{k1}) P(\mathbf{u}^k|H_0) - \sum_{\mathbf{u}^k} p(\mathbf{y}|\mathbf{u}^{k0}) P(\mathbf{u}^k|H_0)\right] d\mathbf{y}$$

$$B = \int_{\mathbf{y}} P(u_0 = 0|\mathbf{y}) \left[\sum_{\mathbf{u}^k} p(\mathbf{y}|\mathbf{u}^{k0}) P(\mathbf{u}^k|H_1) - \sum_{\mathbf{u}^k} p(\mathbf{y}|\mathbf{u}^{k1}) P(\mathbf{u}^k|H_1)\right] d\mathbf{y}$$

$$C = \int_{x_k} \Big[\pi_0 p(x_k|H_0) \int_{\mathbf{y}} P(u_0 = 1|\mathbf{y}) \sum_{\mathbf{u}^k} p(\mathbf{y}|\mathbf{u}^{k0}) P(\mathbf{u}^k|H_0) d\mathbf{y} + \pi_1 p(x_k|H_1) \int_{\mathbf{y}} P(u_0 = 0|\mathbf{y}) \sum_{\mathbf{u}^k} p(\mathbf{y}|\mathbf{u}^{k0}) P(\mathbf{u}^k|H_1) d\mathbf{y}\Big] dx_k$$

with

$$\begin{aligned}
\mathbf{u}^k &= [u_1, \cdots, u_{k-1}, u_{k+1}, \cdots, u_K], \\
\mathbf{x}^k &= [x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_K], \\
\mathbf{u}^{k1} &= [u_1, \cdots, u_{k-1}, u_k = 1, u_{k+1}, \cdots, u_K], \\
\mathbf{u}^{k0} &= [u_1, \cdots, u_{k-1}, u_k = 0, u_{k+1}, \cdots, u_K],
\end{aligned}$$

Clearly C is a constant with regard to U_k. To minimize P_{e0}, one can see from (7) that the optimal quantizer rule for the n^{th} sensor is

$$P(u_k = 1|x_k) = \begin{cases} 0 & \pi_0 p(x_k|H_0)A > \pi_1 p(x_k|H_1)B \\ 1 & \text{Otherwise} \end{cases}$$

If further

$$A > 0 \quad (8)$$
$$B > 0 \quad (9)$$

then the local decision rule amounts to a LRT as in Theorem 1. Further, equation (8) can be rewritten as, after some simplifications,

$$A = \int_{\mathbf{y}^k} \left[P(u_0 = 1|\mathbf{y}^k, u_k = 1) - P(u_0 = 1|\mathbf{y}^k, u_k = 0)\right] p(\mathbf{y}^k|H_0) d\mathbf{y}^k$$

Clearly, $A > 0$ if

$$P(u_0 = 1|\mathbf{y}^k, u_k = 1) - P(u_0 = 1|\mathbf{y}^k, u_k = 0) \geq 0$$

Similarly we can show that (9) is true if

$$P(u_0 = 0|\mathbf{y}^k, u_k = 0) - P(u_0 = 0|\mathbf{y}^k, u_k = 1) \geq 0$$

Thus Theorem 1 is proved.

4. A DESIGN EXAMPLE

In this section, we use a two-sensor example to demonstrate how to obtain the optimal local thresholds for the binary quantizers. Consider the detection of a known signal S in additive Gaussian noises that are independent and identically distributed (i.i.d.) between the two sensors, i.e.,

$$X_k = S + N_k$$

for $k = 1, 2$ with N_1 and N_2 being i.i.d. $\mathcal{N}(0, \sigma^2)$. Without loss of generality, we assume $S = 1$ and $\sigma^2 = 1$. Each sensor quantizes its observation X_k to a binary bit, $U_k = \gamma_k(X_k)$, which is then transmitted through a BSC with identical crossover probability α for both sensors. We first look at the conditions specified in (4) and (5). Using the Gaussian assumption and BSC model, we can derive

$$A = [P(u_0 = 1|y_1 = 1, y_2) - P(u_0 = 1|y_1 = 0, y_2)] (1 - 2\alpha)$$

Thus if 1) $\alpha < 0.5$ and 2) $P(u_0 = 1|y_1 = 1, y_2) > P(u_0 = 1|y_1 = 0, y_2)$, then $A > 0$. Notice condition 2) amounts to using a monotone fusion rule [12]. On the other hand, if $\alpha > 0.5$, one can easily show that the optimal fusion rule should be 'reverse' monotone and we still have $A > 0$. For condition (5), a similar argument can be made. Thus the optimum local decision rule is always a LRT no matter what the parameters are.

Thus Theorem 1 allows us to design the following iterative procedure to obtain optimal quantizers at local sensors.

1. Initialize τ_1 and τ_2.
2. Obtain the optimal fusion rule for fixed τ_1 and τ_2.
3. For fixed fusion rule and τ_2, calculate τ_1 using (6).
4. Similarly for fixed fusion rule and τ_1, calculate τ_2.
5. Check convergence, i.e., if the obtained τ_1 and τ_2 are identical (up to a prescribed precision) to that of the previous iteration. If yes, stop; otherwise, go to 2.

Notice that for the Gaussian problem, the LR is linear in the observation hence the quantizer applies directly on the observation with appropriate threshold translation. Below are two different parameter settings and the respective results.

- $\pi_0 = 0.8$ and $\alpha = 0.1$. For this example, the iteration always converges to the same point $\tau_1 = \tau_2 = 1.0808$, suggesting this may be the global optimum. This is confirmed in Fig. 2 where analytically calculated minimum achievable error probabilities for different τ_1 and τ_2 are plotted that shows a unique minimum point at $(1.0808, 1.0808)$ with $P_{e0} = 0.1909$. Further, the error probability is capped at 0.2. This is sensible given $\pi_0 = 0.8$: one should do no worse than to ignore the local sensor decision and decide H_0.

- $\pi_0 = 0.5$ and $\alpha = 0.1$. This is the equal prior case. For this example, the iteration converges to two different points depending on the initialization: $\tau_1 =$

$\tau_2 = 0.9538$ and $\tau_1 = \tau_2 = 0.0462$. Fig. 3 is the minimum achievable error probability plot for different τ_1 and τ_2. It turns out that both points achieve identical error probability performance at $P_{e0} = 0.3259$, hence both are global minimum.

Not surprisingly, all the local minimum points are symmetric, i.e., $\tau_1 = \tau_2$. Notice that while non-identical optimal local thresholds are possible (see, e.g., [13]), it usually happens only for discrete local sensor observations with carefully selected probability mass functions.

5. CONCLUSIONS

For a sensor fusion network, incorporating transmission channels in the system design may prove useful in resource constrained applications, such as the emerging field of wireless sensor networks. In this paper, we derive the optimal thresholds for joint local binary quantizer design for a hypothesis testing problem under the Bayesian criterion. This design procedure is then applied to a distributed detection example with a known signal and Gaussian noises and binary symmetric channels between sensors and the fusion center. Considered as channel optimized distributed binary quantizers, this work may lead to interesting and novel applications of JSCC in inference-centric wireless sensor networks.

6. REFERENCES

[1] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Information Theory*, vol. 19, pp. 471–480, July 1973.

[2] S.S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," in *Proc. Data Compression Conference*, Snowbird, UT, Mar. 1999, vol. 4, pp. 158–167.

[3] S.D. Servetto, "On the feasibility of large scale wireless sensor networks," in *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing*, Allerton, IL, Oct 2002.

[4] J.J. Spilker, *Digital Communications by Satellite*, Prentice-Hall, Englewood Cliffs, NJ, 1977.

[5] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: an approach to combined source-channel coding," *IEEE Trans. Inform. Theory*, vol. 33, pp. 827–837, Nov. 1987.

[6] M. Longo, T.D. Lookabaugh, and R.M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Trans. Inform. Theory*, vol. 36, pp. 241–255, 1990.

[7] R.R. Tenney and N.R. Sandell Jr., "Detection with distributed sensors," *IEEE Trans. Aerospace Elec. Syst.*, vol. 17, pp. 501–510, July 1981.

[8] A.R. Reibman and L.W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Trans. Aerospace Elect. Sys.*, vol. 23, pp. 24–30, Jan. 1987.

[9] I.Y. Hoballah and P.K. Varshney, "Distributed Bayesian signal detection," *IEEE Trans. Inform. Theory*, vol. 35, pp. 995–1000, Sept. 1989.

[10] S.C.A. Thomopoulos and L. Zhang, "Distributed decision fusion with networking delays and channel errors," *Information Science*, vol. 66, pp. 91–118, Dec. 1992.

[11] B. Chen, R. Jiang, T. Kasetkasem, and P.K. Varshney, "Fusion of decisions transmitted over fading channels in wireless sensor networks," in *Proc. of the 36th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2002, pp. 1184–1188.

[12] P.K. Varshney, *Distributed Detection and Data Fusion*, Springer, New York, 1997.

[13] J.N. Tsitsiklis, "On threshold rules in decentralized detection," in *Proc. 25th IEEE Conf. on Decision and Control*, Athens, Greece, 1986, vol. 1, pp. 232–236.

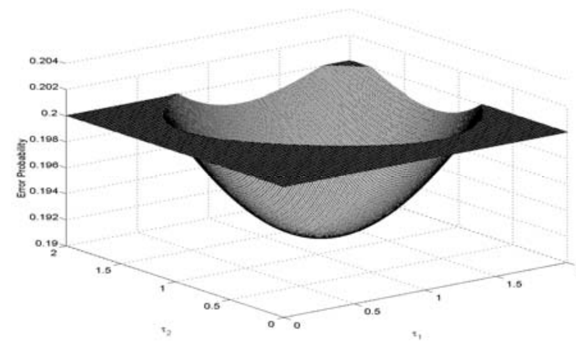

Fig. 2. Minimum achievable error probability for $\pi_0 = 0.8$ as a function of (τ_1, τ_2).

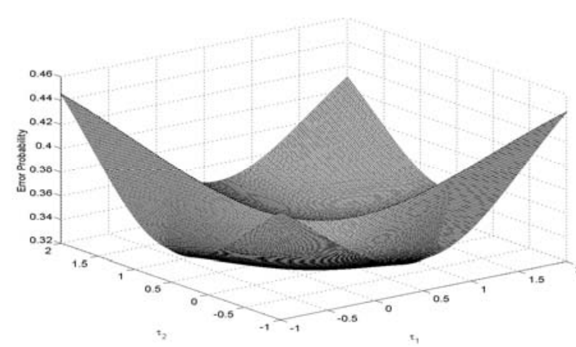

Fig. 3. Minimum achievable error probability for $\pi_0 = 0.5$ as a function of (τ_1, τ_2).