TO COOPERATE OR NOT TO COOPERATE: DETECTION STRATEGIES IN SENSOR NETWORKS

Michael A. Lexa, Christopher J. Rozell, Sinan Sinanović and Don H. Johnson

Rice University
Department of Electrical and Computer Engineering
Houston, TX 77005-1892
amlexa,crozell,sinan,dhj@rice.edu

ABSTRACT

This paper is an initial investigation into the following question: Can cooperation among sensors in a sensor network improve detection performance in a simple hypothesis test? We analyze a simple cooperative system using the Kullback-Leibler (KL) discrimination distance and a quantity known as the information transfer ratio which is a ratio of KL distances. We discover that, asymptotically, gain over a non-cooperative system depends on the conditional KL distance. We conclude with an illustrative example which demonstrates that cooperation not only significantly improves performance but can also degrade it.

1. INTRODUCTION

The power constraints placed on any practical sensor network make short distance transmissions between neighboring sensors almost a necessity. That said, a fundamental question to ask is whether or not cooperation among sensors can improve system performance compared to a non-cooperative system. To study this question, we apply the theory of information processing [1] in a standard distributed detection setting. Similar investigations have appeared in the distributed detection literature [2] and even under the heading of 'learning with finite memory' [3], but the approach taken here, to the best of our knowledge, is novel.

We adopt the simple system architecture shown in figure 1. It combines the classic parallel and tandem distributed detection architectures and serves as an initial model for the problem at hand. All detectors share a common decision rule, and each detector collects independent and identically distributed (iid) observations $Y_n, 1 \leq n \leq N$. Based upon these data and their neighbor's decision, they each decide upon the active hypothesis. Specifically, the detection process begins when the first detector makes a decision based on its observation. This first decision is then transmitted to the second detector. Based on the first detector's decision and its own observation, the second detector makes a decision which is passed on to the third detector. The process repeats, and at each stage, the detectors' decisions are transmitted to a fusion center. The fusion center, in turn, makes a final global decision. Because our application here is sensor networks, the fusion center is not considered to be a distantly located processing unit, but simply one of the local sensors involved in the detection process (manager node).

Central to the theory of information processing is the Kullback-Leibler (KL) discrimination distance [4, 5] and a quantity known as the *information transfer ratio* [1]. The information transfer ratio, denoted by γ, is defined as the ratio of the KL distances between the distributions characterizing the input and output of a system. In a very broad sense, the theory of information processing endeavors to quantify how systems process information. Here, we narrow our attention to the implications of Stein's Lemma [5] which states that the KL distance is the asymptotic exponential error rate of an optimal Neyman-Person detector.

We assume the fusion center performs a Neyman-Pearson likelihood ratio test. That is, we assume one of the fusion center's error probabilities is constrained and the other maximized. However, instead of approaching this optimization directly using Lagrange multipliers, we examine the KL distance between the joint distributions under each hypothesis at the *input* of the fusion center. Because of Stein's Lemma, we know the KL distance, computed at the input of the fusion center, is directly linked to the fusion center's (and therefore the entire cooperative system's) performance. The greater the KL distance, the better the system will be able to discriminate one hypothesis from the other asymptotically.

Let $p_i(\mathbf{Y}) = p(Y_1, Y_2, \ldots, Y_N|H_i)$ and $p_i(\mathbf{U}) = p(U_1, U_2, \ldots, U_N|H_i)$ denote, respectively, the joint probability density function of the observations and the joint probability mass function of the decisions under each hypothesis H_i, $i \in \{0, 1\}$, $U_n \in \{-1, 1\}$. We can express the information transfer ratio across the cooperative portion of the system in one of two ways:

$$\gamma_{01} = \frac{\mathcal{D}\left(p_0(\mathbf{U})\|p_1(\mathbf{U})\right)}{\mathcal{D}\left(p_0(\mathbf{Y})\|p_1(\mathbf{Y})\right)} \quad \gamma_{10} = \frac{\mathcal{D}\left(p_1(\mathbf{U})\|p_0(\mathbf{U})\right)}{\mathcal{D}\left(p_1(\mathbf{Y})\|p_0(\mathbf{Y})\right)}$$

where, for two continuous density functions $p_0(x), p_1(x)$, the KL distance between $p_0(x)$ relative to $p_1(x)$ is defined as the expected value of the likelihood ratio with respect to p_0,

$$\mathcal{D}\left(p_0\|p_1\right) = \int p_0(x) \log \frac{p_0(x)}{p_1(x)} dx.$$

We distinguish between the two types of KL distances in the expressions above because, in general, KL distances are not symmetric ($\mathcal{D}\left(p_1\|p_0\right) \neq \mathcal{D}\left(p_0\|p_1\right)$). For the remainder of the paper, however, we suppress the subscripts and focus on γ_{01} because it's the pertinent quantity when we fix the fusion center's false alarm probability and try to maximize its probability of detection. The Data Processing Theorem [5], states that $\mathcal{D}\left(p_0(\mathbf{Y})\|p_1(\mathbf{Y})\right) \geq \mathcal{D}\left(p_0(\mathbf{U})\|p_1(\mathbf{U})\right)$, causing γ to be a number between zero and one. Thus, we interpret γ as being the fractional loss of discrimination distance across the cooperative portion of the system relative to $\mathcal{D}\left(p_0(\mathbf{Y})\|p_1(\mathbf{Y})\right)$. Here, this effectively means that γ represents the relative performance of the cooperative system in figure 1

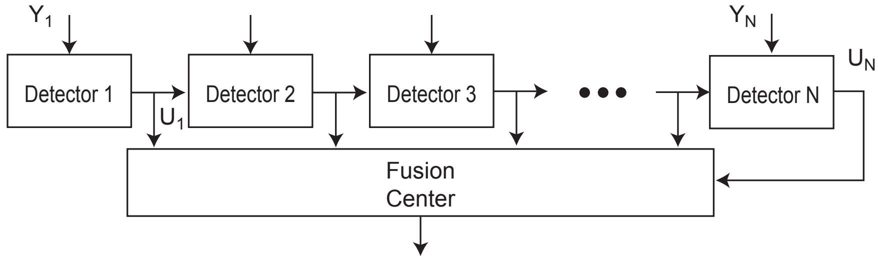

Figure 1: All the detectors, except the first, make their decisions based upon their observations and the preceding detector's decision. The decision of the fusion center is the final, global decision.

to the optimal centralized detector because $\mathcal{D}\left(p_0(\mathbf{Y})\|p_1(\mathbf{Y})\right)$ is the KL distance a centralized detector would see at its input.

2. ANALYSIS AND RESULTS

Because we assume the inputs are iid under each hypothesis, the KL distance between the joint distributions of $\mathbf{Y}$ is simply the sum of the KL distances of Y_n,

$$\begin{aligned}\mathcal{D}\left(p_0(\mathbf{Y})\|p_1(\mathbf{Y})\right) &= \sum_{n=1}^{N} \mathcal{D}\left(p_0(Y_n)\|p_1(Y_n)\right) \\ &= N \cdot \mathcal{D}\left(p_0(Y)\|p_1(Y)\right). \qquad (1)\end{aligned}$$

The subscript on Y is dropped in (1) because $\mathcal{D}\left(p_0(Y_n)\|p_1(Y_n)\right)$ is constant for all n. Because the decision statistics have a Markovian structure [5], the distance between the joint distributions of the binary decisions is

$$\begin{aligned}\mathcal{D}\left(p_0(\mathbf{U})\|p_1(\mathbf{U})\right) &= \mathcal{D}\left(p_0(U_1)\|p_1(U_1)\right) \\ &+ \sum_{n=2}^{N} \mathcal{D}\left(p_0(U_n|U_{n-1})\|p_1(U_n|U_{n-1})\right). \qquad (2)\end{aligned}$$

Here, the conditional distributions $p_i(U_n|U_{n-1}), i \in \{0,1\}$, depend on the decision rules. By definition, the ratio of (1) and (2), is the information transfer ratio,

$$\begin{aligned}\gamma(N) = &\frac{\mathcal{D}\left(p_0(U_1)\|p_1(U_1)\right)}{N \cdot \mathcal{D}\left(p_0(Y)\|p_1(Y)\right)} \\ &+ \frac{\sum_{n=2}^{N} \mathcal{D}\left(p_0(U_n|U_{n-1})\|p_1(U_n|U_{n-1})\right)}{N \cdot \mathcal{D}\left(p_0(Y)\|p_1(Y)\right)}. \qquad (3)\end{aligned}$$

When there is no cooperation (communication) between the detectors, the system in figure 1 degenerates into the standard parallel distributed detection architecture. The output marginal distributions $p_i(U_n),\ i \in \{0,1\}$ become iid and the KL distance between the joint distributions $p_i(\mathbf{U})$ is N times the distance between one of the marginals. Therefore,

$$\gamma = \frac{\mathcal{D}\left(p_0(U)\|p_1(U)\right)}{\mathcal{D}\left(p_0(Y)\|p_1(Y)\right)}. \qquad (4)$$

Note this expression is constant with respect to N. Thus, in terms of γ, adding detectors to the non-cooperative system does not change performance. This fact contrasts with (3) which explicitly depends on N. Asymptotically, though, (3) does reach a limiting value. In the appendix, we prove

$$\lim_{N\to\infty} \gamma(N) = \frac{c_2(1-p) + fc_1}{(1+f-p) \cdot \mathcal{D}\left(p_0(Y)\|p_1(Y)\right)} \qquad (5)$$

where

$$\begin{aligned} p &= P_0\big(U_n = -1|U_{n-1} = -1\big) \\ q &= P_1\big(U_n = -1|U_{n-1} = -1\big) \\ f &= P_0\big(U_n = -1|U_{n-1} = 1\big) \\ g &= P_1\big(U_n = -1|U_{n-1} = 1\big) \\ c_1 &= p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} \\ c_2 &= f\log\frac{f}{g} + (1-f)\log\frac{1-f}{1-g}. \end{aligned} \qquad (6)$$

This result allows us to identify diminishing points of return, i.e. identify the point where adding detectors no longer results in significant gain.

For finite N and a fixed decision rule, the cooperative system will outperform the non-cooperative system if (3) greater than (4). Asymptotically, this condition reduces to $D^* > \mathcal{D}\left(p_0(U)\|p_1(U)\right)$ where

$$D^* = \lim_{N\to\infty} \tfrac{1}{N} \sum_{n=2}^{N} \mathcal{D}\left(p_0(U_n|U_{n-1})\|p_1(U_n|U_{n-1})\right).$$

As the example below demonstrates, this condition in general does not hold. There are cases when cooperation reduces performance (γ decreases). The key is to introduce cooperation (i.e. define decision rules) such that this inequality holds. Unfortunately, but perhaps not unexpectedly, the conditions under which it holds depend on the probability distributions of the observations.

We now present an illustrative example which shows that even simple interactions can produce significant gains. To simplify the calculations, we choose a variant of what Tang *et al.* termed the two-sided constant control strategy (CCS) [6] for the decision rules,

$$\Lambda(Y_1) \underset{H_0}{\overset{H_1}{\gtrless}} t$$

$$\Lambda(Y_n) \underset{H_0}{\overset{H_1}{\gtrless}} ta^{U_{n-1}}, \quad 2 \le n \le N$$

where $\Lambda(\cdot)$ is the likelihood ratio, t is a baseline threshold, and $a \geq 1$ is a parameter. In words, the detectors operate and interact in the following way. The n^{th} ($n \geq 2$) detector collects an observation $Y_n = y_n$ and evaluates the likelihood ratio at that point. If $U_{n-1} = 1$, $\Lambda(y_n)$ is compared to the baseline threshold scaled by a; if $U_{n-1} = -1$, it is compared to t/a. In either case, if $\Lambda(y_n)$ is greater than its threshold then $U_n = 1$, otherwise $U_n = -1$. Tang *et al.* proposed the two-sided CCS as a suboptimal decision strategy for the tandem distributed detection architecture. They showed, however, that this strategy performs nearly as well as the optimal centralized detector in terms of probability of error.

Let the input observations Y_n be Gaussian with mean $m_0 = 0$ (under H_0) and $m_1 = 1$ (under H_0) and variance $\sigma^2 = 1$. In this case, the likelihood ratio is given by $\Lambda(Y_n) = \exp\left(Y_n - \frac{1}{2}\right)$. Working with the sufficient statistic $\Upsilon(Y_n) = Y_n$, the decision rules become,

$$Y_1 \underset{H_0}{\overset{H_1}{\gtrless}} \ln t + \tfrac{1}{2}$$

$$Y_n \underset{H_0}{\overset{H_1}{\gtrless}} \ln t a^{U_{n-1}} + \tfrac{1}{2}, \quad 2 \leq n \leq N.$$

The baseline threshold t is set by specifying the probability of false alarm for a single detector operating in isolation. Letting $P_f = .05$, results in $t = 3.142$. With these rules, the conditional probabilities in (6) are,

$$p = F_0\big(t/a\big), \qquad q = F_1\big(t/a\big)$$
$$f = F_0\big(ta\big), \qquad g = F_1\big(ta\big)$$

where $F_i(\cdot)$ is cdf of the likelihood ratio under hypothesis i. The input KL distance $\mathcal{D}\left(p_0(\mathbf{Y}) \| p_1(\mathbf{Y})\right)$ is $N/2$ [7].

The top panel of figure 2 plots γ as a function of a for various values of N. We interpret this plot in the following way. γ equaling one represents, in some sense, the asymptotic performance of the optimal centralized detector because, as mentioned above, γ is defined relative to $\mathcal{D}\left(p_0(\mathbf{Y}) \| p_1(\mathbf{Y})\right)$. When the system becomes distributed (without cooperation), performance degrades, and the fusion center sees a smaller input KL distance. The relative loss is quantified by the horizontal line. In this example, we can in certain cases, regain some of the loss by cooperating. Anytime the curves rise above the horizontal line, signifies such a case. Because the KL distance is related to the asymptotic exponential error rate of the fusion center, these gains in γ are significant. On the other hand, when the curves dip below the horizontal line, cooperation actually degrades performance.

The bottom plot of figure 2 plots γ when $\sigma = 2$. In this case, the gain, relative to the non-cooperative system, is even greater than before, but the KL distance at the input of the fusion center is smaller.

Interestingly, this example also suggests that there may be no need to have large numbers of cooperating detectors. When $N = 30$, the γ curve is nearly coincident with the asymptotic curve. Adding more detectors simply does not boost performance in any significant manner.

3. CONCLUSION

These results suggest that cooperation may only beneficial (in terms of detection) when the decision rules boost the conditional KL distance at the input of the fusion center for a given pair of input distributions. Perhaps counterintuitively, this statement does not, necessarily, imply the system should be a learning system. In the example above, it is difficult to argue that the individual detectors become "smarter" as the decisions propagate through the system. The decision rules are not adaptive, and upon receiving a decision, a detector's threshold adjusts such that it becomes harder for it to make the same decision. We conjecture that the gain seen in the example stems from the statistical dependencies between $p_0(\mathbf{U})$ and $p_1(\mathbf{U})$ built in by the decision rules. Future work entails investigating this conjecture and finding good, perhaps even optimal, decision rules for a given class of input distributions.

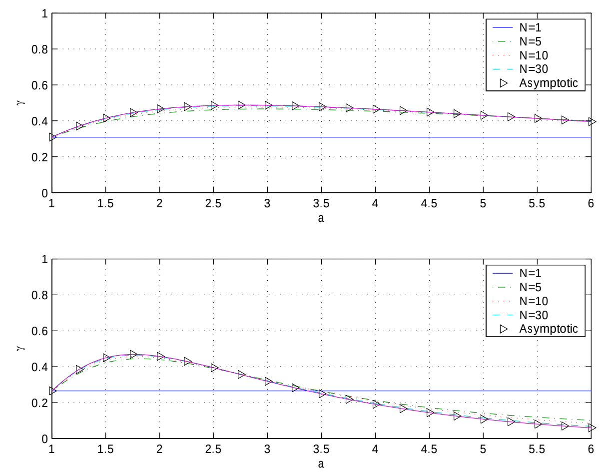

Figure 2: The information transfer ratio is plotted as a function of the linking parameter a when the input observations are iid Gaussian. The top panel shows the case when $\sigma = 1$, the bottom panel $\sigma = 2$.

A. APPENDIX

Taking the limit of (3), we have

$$\lim_{N\to\infty} \gamma(N) = \frac{\lim \frac{1}{N}\sum_{n=2}^{N} \mathcal{D}\left(p_0(U_n|U_{n-1}) \| p_1(U_n|U_{n-1})\right)}{\mathcal{D}\left(p_0(Y) \| p_1(Y)\right)}. \quad (7)$$

From its definition, we can rewrite the conditional KL distance as

$$\mathcal{D}\left(p_0(U_n|U_{n-1}) \| p_1(U_n|U_{n-1})\right) = c_1 h_{n-1} + c_2(1 - h_{n-1})$$

where $h_{n-1} = P_0\big(U_{n-1} = -1\big)$ and c_1, c_2 are defined in (6). Therefore,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=2}^{N} \mathcal{D}\left(p_0(U_n|U_{n-1}) \| p_1(U_n|U_{n-1})\right) = \lim_{N\to\infty}\left[c_1 \frac{1}{N}\sum_{n=2}^{N} h_{n-1} + c_2 \frac{1}{N}\sum_{n=2}^{N}(1 - h_{n-1})\right]. \quad (8)$$

To evaluate the series, we first consider the sequence $\{h_n\}$. Because h_n can be interpreted as a state probability of a two-state, irreducible, and aperiodic Markov chain, we know that its limit exists and is equal to its equilibrium probability [8]. Solving the global balance equations, yields the equilibrium probabilities. Here, we find that $\lim_{n\to\infty} h_n = \frac{f}{1+f-p}$. Now, using the theorem of Cesaro's mean ([5] p. 64), we conclude

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=n_0}^{N} h_n = \frac{f}{1+f-p}.$$

Plugging this result back into (8) and (7) and simplifying gives

$$\frac{c_2(1-p)+fc_1}{(1+f-p)\cdot \mathcal{D}\left(p_0(Y)\|p_1(Y)\right)}. \quad \square$$

B. REFERENCES

[1] S. Sinanović and D.H. Johnson, "Asymptotic rates of the information transfer ratio," *Proceedings of the International Conference on Acoutics, Speech, and Signal Processing (ICASSP'02)*, vol. 2, pp. 1505–1508, 2002.

[2] R. Visvanathan and P. Varshney, "Distributed detection with multiple sensors: Part I-Fundamentals," *Proc. of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan 1997.

[3] M. Hellman and T. Cover, "Learning with finite memory," *The Annals of Mathematical Statistics*, vol. 41, no. 3, pp. 765–782, Jun 1970.

[4] S. Kullback, *Information Theory and Statistics*, Wiley, New York, 1959.

[5] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, 1991.

[6] Z.B. Tang, K.R. Pattipati, and D.L. Kleinman, "Optimization of detection networks: Part I-Tandem structures," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 21, no. 5, pp. 1044–1059, Sep 1991.

[7] D.H. Johnson and G. Orsak, "Relation of signal set choice to the performance of optimal non-Gaussian detectors," *IEEE Trans. on Communications*, vol. 41, no. 9, pp. 1319–1328, Sep. 1993.

[8] J.R. Norris, *Markov Chains*, Cambridge University Press, 1997.