OPTIMAL DISTRIBUTED DETECTION OF MULTIPLE HYPOTHESES USING BLIND ALGORITHM

Aleksandar Jeremić, Kon Max Wong

Dept. of Electrical and Computer Engineering
McMaster University, Hamilton, Canada*

Bin Liu

School of Electrical Engineering
University of Guelph, Guelph, Canada

ABSTRACT

In a parallel distributed detection in order to design the optimal fusion rule, the fusion center needs to have perfect knowledge of the performance of the local detectors as well as the prior probabilities of the hypotheses. Such knowledge is not available in most practical cases. In this paper, we propose a blind technique for the M-ary distributed detection problem. We derive the probability mass function of the local decisions and use this result to develop maximum likelihood estimates of unknown parameters. We also derive analytically the overall detection performance for both binary and M-ary distributed detection and discuss the difference of the overall detection performance obtained using the estimated values of unknown parameters and their true values. Finally, we demonstrate the applicability of our results through numerical examples.

1. INTRODUCTION

Hypothesis testing in distributed signal processing, referred to as distributed detection, is different in essence from that in classical multichannel scenarios. In the latter, usually referred to as centralized detection, observations from all channels are communicated to a central processor, where statistical inference on the hypotheses is conducted. However, many practical difficulties restrict the applicability of centralized detection, such as communication bandwidth, data transmission speed and computational complexity. In addition, observations collected from different channels could be incomparable and a decision on the hypothesis in question based on a mixture of observations may not be reliable [1]. In contrast to centralized detection, each local detector of a parallel distributed detection system preprocesses the observations it collects, makes a local decision and then transmits it to a fusion center where a global inference is made.

Binary distributed detection problems have been extensively studied [1–6]. It has been shown that for a parallel distributed detection system with conditionally independent local detectors, when local decision rules are given, the optimal distributed detection in the sense of minimizing overall error probability can only be achieved if the prior probabilities of the two hypotheses and the parameters describing the performance of local detectors are known. An online adaptive algorithm has been proposed in [3] to estimate the unknown parameters, or the necessary weights may be estimated directly using reinforcement learning as suggested in [4] and [5]. To bypass the aforementioned problems, the authors of [6] found the unknown parameters can be yielded by analytically solving a set of nonlinear equations involving the probabilities of different decision combinations at all local detectors, which although are not available in practical applications either, could be replaced by their corresponding empirical probabilities. The resulting estimator is asymptotically unbiased and substantially more reliable. The distributed detection process with multiple hypotheses, usually referred to as M-ary distributed detection, has recently attracted wide interest [7–9], because a large number of practical problems

In our previous work [10] we derived a blind algorithm for M-ary decision fusion using least-squares estimation and suboptimal hierarchical grouping of hypotheses. However due to its non-optimality the convergence time and/or detection and false alarm errors may be unacceptable in certain applications. In this paper we propose to estimate the unknown parameters (prior probabilities and local detectors' probabilities of false alarm and miss) using statistically optimal maximum likelihood (ML) estimator. The ML estimator accounts for the known parametric form of the likelihood function of local decision combinations, and hence has a better estimation accuracy than the LS estimator. This is especially the case when only a small number of local decisions are used. We then present the analytical expression of overall error probability when the true values of the parameters are given and explore the effect of our blind algorithm to the system detection performance. We illustrate the applicability of our results through numerical examples.

2. OPTIMAL M-ARY FUSION RULE

Let us assume the parallel distributed detection system consisting of N local detectors. In general, the phenomenon changes from time to time, and at each time it could be one of the M possible hypotheses $\{H_0, H_1, \cdots, H_{M-1}\}$ with prior probabilities $P(H_i),\ i = 0, \cdots, M-1$, respectively. For each time, the local detectors $\mathrm{D}_j,\ j = 1, \cdots, N$ make a decision u_j individually according to their own observations z_j. Given the unknown hypothesis H_i, the decision u_j is assumed to be conditionally independent of the decisions from other local detectors. The local detector D_j then sends u_j to the fusion center, where a global decision u_0 on the hypothesis is made based on a particular optimality criterion. For the jth detector,

*This work has been supported by NSERC grant number RGPIN311954-05.

we define the probability of anomaly as

$$\varepsilon_{ik}^{j} \triangleq P(u_j = H_k | H_i \text{ is true}) \tag{1}$$

where u_j is the decision of the jth local detector, $i, k \in \{0, \cdots, M-1\}$ and $i \neq k$.

The fusion rule can be derived by minimizing the probability of error at the fusion center

$$P_e = \sum_{i=0}^{M-1} \sum_{\substack{k=0 \\ k \neq i}}^{M-1} P(H_i) P(u_0 = H_k | H_i \text{ is true}) \tag{2}$$

It has been shown in [13] that minimizing the error probability in Eq. (2) reduces to maximizing the posterior probability

$$\begin{aligned} P(H_i|\boldsymbol{u}) &= P(H_i|u_1, u_2, \cdots, u_N) \\ &= \frac{P(H_i)}{P(\boldsymbol{u})} P(u_1|H_i) \cdots P(u_N|H_i) \end{aligned} \tag{3}$$

where $\boldsymbol{u} = (u_1, \cdots, u_N)$. For $i = 0, \cdots, M-1$, the global decision is therefore [7]

$$\begin{aligned} u_0 &= \arg\max_{H_i} P(H_i|\boldsymbol{u}) \\ &= \arg\max_{H_i} P(H_i) \prod_{j \in \mathcal{S}_0} \varepsilon_{i0}^{j} \cdots \prod_{j \in \mathcal{S}_{M-1}} \varepsilon_{iM-1}^{j} \end{aligned} \tag{4}$$

3. PARAMETER ESTIMATION

In practice, we have no knowledge on prior probabilities of the hypotheses and the local detector performance. We have, in general, only a set of decision sequences $\{u_{jn}\}, j = 1, \cdots, N$ and $n = 1, \cdots, N_d$ available, where N_d is the total number of decisions made by a particular local detector. For a fixed n, the set represents the decisions on the same hypothesis made by N local detectors. It represents the decision sequence made by a particular local detector when j is fixed. In order to apply Eq. (4) to achieve the multiple hypothesis optimality, we need to estimate the unknown parameters, i.e., prior probabilities and probabilities of anomaly using local decisions.

For M hypotheses and N local detectors, let $\boldsymbol{U}$ be the set consisting of all of the possible decision combinations. Clearly, $\dim(\boldsymbol{U}) = M^N$. Let the random variable X_ℓ indicate the number of times the ℓth combination $\boldsymbol{u}_\ell$ occurring with occurrence probability $P(\boldsymbol{u}_\ell)$, where $\ell = 1, \cdots, L$ and $L = M^N$. We refer to X_ℓ as occurrence number. Furthermore, recall the Bayes' rule and conditional independence, the probability of one of the M^N possible combinations can be written as

$$\begin{aligned} p_\ell &= P(\boldsymbol{u}_\ell) \\ &= \sum_{i=0}^{M-1} P(H_i) P(u_1 = s_1, \cdots, u_N = s_N | H_i) \\ &= \sum_{i=0}^{M-1} P(H_i) P(u_1 = s_1|H_i) \cdots P(u_N = s_N|H_i) \end{aligned} \tag{5}$$

where $s_1, \cdots, s_N \in \{H_0, \cdots, H_{M-1}\}$. Substituting the true values of prior and anomaly probabilities into Eq. (5) gives all occurrence probabilities. For a fixed total number of local decisions N_d, the occurrence numbers of all possible decision combinations, namely $\boldsymbol{X} = (X_1, X_2, \cdots, X_L)$, are multinomial distributed with probability mass function

$$P(X_1 = x_1, \cdots, X_L = x_L | N_d) = \frac{N_d!}{x_1! \cdots x_L!} p_1^{x_1} \cdots p_L^{x_L} \tag{6}$$

and var$(X_\ell) = N_d p_\ell (1 - p_\ell)$, cov$(X_s X_\ell) = -N_d p_s p_\ell$ for $s = 1, \cdots, L$ and $s \neq \ell$. We define the vector $\boldsymbol{\varepsilon}$ consisting of $(M-1)MN$ unknown probabilities of anomaly in Eq. (1) and the $(MN+1)(M-1)$ – dimensional vector $\boldsymbol{\theta} = [\boldsymbol{\varepsilon}, P(H_0), \cdots, P(H_{M-2})]$. As illustrated in Eq. (5), the occurrence probability p_ℓ is the nonlinear function of unknown parameters represented by $\boldsymbol{\theta}$, i.e., $p_\ell = f_\ell(\boldsymbol{\theta})$.

For binary distributed detection, the unknown parameters can be obtained by analytically solving a set of nonlinear equations described in Eq. (5). It has also been shown in [6] that the estimates of unknown parameters converge to their true values asymptotically. We propose to extend the approach to its M-ary counterpart. The particular occurrence probability is estimated as time averaging, i.e., empirical probability

$$\begin{aligned} p_\ell &= P(u_1 = s_1, u_2 = s_2, \cdots, u_N = s_N) \\ &\backsimeq \frac{\text{number of } (u_1 = s_1, \cdots, u_N = s_N)}{\text{total number of local decisions } N_d} \end{aligned} \tag{7}$$

where $s_1, \cdots, s_N \in \{H_0, \cdots, H_{M-1}\}$. Let the estimate of the ℓth occurrence probability be y_t and recall the true occurrence probability $p_\ell = f_\ell(\boldsymbol{\theta})$, hence

$$y_\ell = f_\ell(\boldsymbol{\theta}) + e_\ell, \quad \ell = 1, \cdots, L \tag{8}$$

where e_ℓ is the estimation error. We define the vector $\boldsymbol{y} = [y_1, y_2, \cdots, y_L]^T$, $\boldsymbol{f}(\boldsymbol{\theta}) = [f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), \cdots, f_L(\boldsymbol{\theta})]^T$, and $\boldsymbol{e} = [e_1, e_2, \cdots, e_L]^T$, the problem can therefore be formulated as

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{e} \tag{9}$$

To account for the known distribution of local decision combinations, the ML estimation is a very efficient algorithm to apply. As discussed before, the occurrence number of different local decision combinations X_ℓ is multinomial distributed with likelihood function

$$\begin{aligned} &P(X_1 = x_1, \cdots, X_L = x_L | N_d, \boldsymbol{\theta}) \\ &= \frac{N_d!}{x_1! \cdots x_L!} p_1^{x_1} \cdots p_L^{x_L} \end{aligned} \tag{10}$$

where $p_\ell = f_\ell(\boldsymbol{\theta})$, $\ell = 1, \cdots, L$. Once the occurrence numbers are known, the estimate is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} P(X_1 = x_1, \cdots, X_L = x_L | N_d, \boldsymbol{\theta}) \tag{11}$$

4. PERFORMANCE ANALYSIS

In this section we derive the overall probability of error for the proposed M-ary fusion system. In this case, $M-1$ likelihood ratios are necessary to derive the detection scheme. For $\ell = 0, \cdots, M-1$, let us define

$$\Lambda_\ell(\boldsymbol{u}) = \frac{P(\boldsymbol{u}|H_\ell)}{P(\boldsymbol{u}|H_0)} \tag{12}$$

and

$$\phi_\ell(\boldsymbol{u}) = \sum_{\substack{j=0\\ j\neq\ell}}^{M-1} P(H_j)P(\boldsymbol{u}|H_j) \tag{13}$$

The global decision is the hypothesis corresponding to the minimum value of $\phi_\ell(\boldsymbol{u})$ as shown in [13]. Dividing Eq. (13) by $P(\boldsymbol{u}|H_0)$ yields

$$\varphi_\ell(\boldsymbol{u}) = \sum_{\substack{j=0\\ j\neq\ell}}^{M-1} P(H_j)\Lambda_\ell(\boldsymbol{u}) \tag{14}$$

The global decision rule can be written as

$$u_0 = H_\ell \quad \text{if } \varphi_\ell(\boldsymbol{u}) = \min\{\varphi_0(\boldsymbol{u}), \cdots, \varphi_{M-1}(\boldsymbol{u})\} \tag{15}$$

The decision space is an $M-1$ dimensional space spanned by the likelihood ratios $\Lambda_1(\boldsymbol{u}), \cdots, \Lambda_{M-1}(\boldsymbol{u})$. We then define the weight for each likelihood ratio similar to that in binary detection. Let

$$w_0^\ell = \log\frac{P(H_\ell)}{P(H_0)}, \quad \forall\, \ell = 1, \cdots, M-1 \tag{16}$$

and for $\ell = 1, \cdots, M-1$, let w_j^ℓ be the weight of the ℓth likelihood ratio defined in Eq. (12). If the jth local detector makes a decision in favor of H_k, then

$$\begin{aligned} w_j^\ell &= \log\frac{P(u_j = H_k|H_\ell)}{P(u_j = H_k|H_0)} \\ &= \log\frac{\varepsilon_{\ell k}^j}{\varepsilon_{0k}^j}, \quad \forall\, j = 1, \cdots, N \end{aligned} \tag{17}$$

The weight w_j^ℓ is a random variable and can take on M possible values with known probabilities depending only on the decision made at the jth local detector, i.e.,

$$P\left(w_j^\ell = \log\frac{\varepsilon_{\ell k}^j}{\varepsilon_{0k}^j}|H_i \text{ is true}\right) = \varepsilon_{ik}^j$$

Once the decision at the jth local detector is determined, the values of all weights corresponding to the same detector w_j^ℓ, $\ell = 1, \cdots, M-1$ are known. Substituting all weights w_0^ℓ and w_j^ℓ into the global decision rule in Eq. (15) arrives at the global decision. Consequently, the global decision rule for M-ary distributed detection can be written in a compact form

$$u_0 = \begin{cases} H_0, & \text{if } \sum_{j=0}^N w_j^\ell < 0, \forall \ell = 1, \cdots, M-1 \\ H_k, & \text{if } \sum_{j=0}^N w_j^k > 0 \text{ and } \sum_{j=0}^N w_j^k = \\ & \quad \max\{\sum_{j=0}^N w_j^1, \cdots, \sum_{j=0}^N w_j^{M-1}\} \end{cases} \tag{18}$$

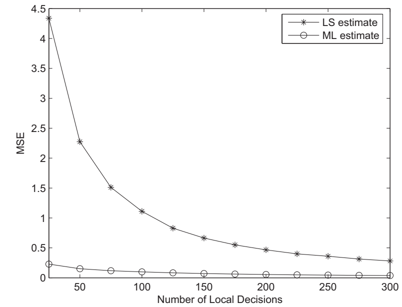

Fig. 1. Comparison of MSE.

Therefore the overall probability of error is

$$\begin{aligned} P_e &= 1 - \sum_{i=0}^{M-1} P(H_i)P(u_0 = H_i|H_i) \\ &= 1 - P(H_0)P\left[\sum_{j=0}^N w_j^1 < 0, \cdots, \sum_{j=0}^N w_j^{M-1} < 0|H_0\right] - \\ &\quad - \sum_{i=1}^{M-1} P(H_i)P\left[\sum_{j=0}^N w_j^i > 0, E_i|H_i\right] \end{aligned} \tag{19}$$

where the event E_i is defined as

$$E_i = \sum_{j=0}^N w_j^i > \sum_{j=0}^N w_j^n, \quad \forall\, n = 1\cdots, M-1 \text{ and } n \neq i \tag{20}$$

Note that the above expression depends on the anomaly probabilities which in blind case are estimated i.e., are random variables. Consequently, the overall probability of error is also a random variable which is further demonstrated in Section 5.

5. NUMERICAL EXAMPLES

In this Section we demonstrate the applicability of our results. In all examples, we assume the distributed detection system consisting of three local detectors and consider a detection scenario with $M = 3$ hypotheses. In Fig 1, we evaluate the estimation accuracy of the proposed algorithms based on 10000 runs. For comparison purposes in addition to ML estimation we performed least-squares (LS) estimation by minimizing the mean square error $\|\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\theta})\|^2$. Decisions from three local detectors are generated according to the true values of prior and anomaly probabilities and the MSE corresponding to LS and ML estimators is illustrated. As expected, the ML estimator outperforms the LS estimator. The MSE of ML estimator is relatively small and does not decrease significantly with the increasing number of local decisions.

In Fig. 2, we demonstrate how overall probability of error evolves with the increasing number of local decisions using two different estimation algorithms for 3-ary distributed detection system. The theoretical value is computed by substituting the true values of both prior and anomaly probabilities

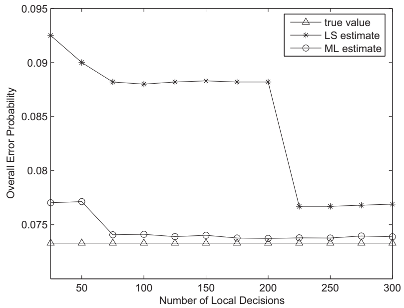

Fig. 2. Overall performance of 3-ary distributed detection system.

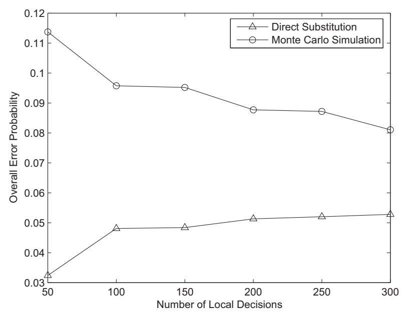

Fig. 3. Comparison of overall error probability using direct substitution and Monte Carlo simulation.

into Eq. (19). The other two overall error probability curves are obtained based on 10000 runs. In general, the overall error probability using ML estimates is smaller and more importantly, converges faster than that using LS estimates. We can also see from the same figure that for a particular set of unknown parameters, there exists a threshold above which increasing number of local decisions will not yield significant improvements in the performance.

As we have stated in Section 4, Eq. (19) gives us the theoretical value of overall error probability only when true values of unknown parameters are known. In blind case the overall detection performance can only be assessed through Monte Carlo simulations. However, direct substitution of the estimates into Eq. (19) gives us a rough idea of the overall detection performance. The difference of overall error probabilities obtained by these two methods is shown in Fig. 3. As it can be seen direct substitution of estimated values in Eq. (19) provides reasonable approximation (less than 5%) after relatively small number of samples $\tilde{1}00$.

6. REFERENCES

[1] Viswanathan, R., and Varshney, P. K., "Distributed detection with multiple sensors : part I– fundamentals", *Proceedings of the IEEE*, 85, 1 (Jan. 1997), 54-63.

[2] Chair, Z. and Varshney, P. K., "Optimal data fusion in multiple sensor detection systems", *IEEE Transactions on Aerospace and Electronic Systems*, 22, 1 (Jan. 1986), 98-101.

[3] Naim, A., and Kam, M., "On-line estimation of probabilities for distributed Bayesian detection", *Automatica*, 30, 4 (Apr. 1994), 633-642.

[4] Ansari, N., Hou, E. S. H., Zhu, B-O., and Chen, J-G., "Adaptive fusion by reinforcement learning for distributed detection systems", *IEEE Transactions on Aerospace and Electronic Systems*, 32, 2 (Apr. 1996), 524-531.

[5] Ansari, N., Chen, J-G., and Zhang, Y-Z., "Adaptive decision fusion for unequiprobable sources", *IEE Proceedings on Radar, Sonar and Navigation*, 144, 3 (June 1997), 105-111.

[6] Mirjalily, G., Luo, Z-Q., Davidson, T. D., and Bosse, E., "Blind adaptive decision fusion for distributed detection", *IEEE Transactions on Aerospace and Electronic Systems*, 39, 1 (Jan. 2003), 34-52.

[7] Baek, W., and Bommareddy, S., "Optimal m-ary data fusion with distributed sensors", *IEEE Transactions on Aerospace and Electronic Systems*, 31, 3 (July 1995), 1150-1152.

[8] Zhang, Q., and Varshney, P. K.,"Decentralized M-ary detection via hierarchical binary decision fusion", *Information Fusion*, 2, 1 (Mar. 2001), 3-16.

[9] Zhu, X., Yuan, Y., Rorres, C., and Kam, M., "Distributed M-ary hypothesis testing with binary local decisions", *Information Fusion*, 5, 3 (Sept. 2004), 157-167.

[10] Liu, B., Jeremic, A., and Wong, K. M., "Blind adaptive algorithm for M-ary distributed detection", in *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawaii, Apr. 2007, 1025-1028.

[11] Saxena, V., Gonzalez-Gomez, I., and Laug, W. E., "A noninvasive multimodal technique to monitor brain tumor vascularization", *Physics in Medicine and Biology* 52, 17 (Sept. 2007) 5295-5308.

[12] So, E. L., "Integration of EEG, MRI, and SPECT in localizing the seizure focus for epilepsy surgery", *Epilepsia* 41, Suppl 3 (2000), S48-54.

[13] Van Trees, H. L., *Detection, Estimation, Modulation Theory, Part I*. New York : John Wiley & Sons, 1968.

[14] Seber, G. A. F. and Wild, C. J., *Nonlinear Regression*. New York : John Wiley & Sons, 1989.

[15] Rao, C. R., *Linear Statistical Inference and its Applications* (2nd ed.). New York : John Wiley & Sons, 1973.

[16] Poor, H. V., *An Introduction to Signal Detection and Estimation* (2nd ed.). New York : Springer-Verlag, 1994.

[17] Gallant, R., *Nonlinear Statistical Models*, New York : John Wiley & Sons, 1987.

[18] Box, M. J., "Bias in nonlinear estimation" *Journal of the Royal Statistical Society. Series B (Methodological)* 33, 2 (1971), 171-201.

[19] Fisher, R. A., *Statistical Methods and Scientific Inference* (3rd ed.). New York : Hafner Press, 1973.

[20] Hero, A. O., Fessler, J. A., and Usman, M., "Exploring estimator bias-variance tradeoffs using the uniform CR bound", *IEEE Transactions on Signal Processing*, 44, 8 (Aug. 1996), 2026-2041.

[21] Eldar, Y. C., "Minimum variance in biased estimation : bounds and asymptotically optimal estimators", *IEEE Transactions on Signal Processing*, 52, 7 (July 2004), 1915-1930.

[22] Evans, M., Hastings, N., and Peacock, B., *Statistical Distributions* (3rd ed.). New York : John Wiley & Sons, Inc. 2000.

[23] Papoulis, A., and Pillai, S. U., *Probability, Random Variables and Stochastic Processes* (4th ed.). New York : McGraw-Hill, 2002.