OPTIMAL LINEAR COOPERATION FOR SIGNAL CLASSIFICATION

Zhi Quan[1], *Muyang Ye*[1,2], *Zhi Ding*[2], *and Shuguang Cui*[2,3]

Department of Electrical and Electronic Engineering, Southern University of Science and Technology, China[1]
School of Information Science and Technology, ShanghaiTech University, China[2]
Department of Electrical and Computer Engineering, Texas A&M University, USA[3]

ABSTRACT

In distributed inference, cooperation among networked agents can be exploited to enhance the performance of each individual agent. In this paper, we consider signal classification over a network of agents, where each agent observes a certain signal under a particular signal-to-noise ratio (SNR). Each agent produces a statistic that summarizes its observations over a time period and then forwards it to a fusion center for identifying the type of signal in a global manner. A linear cooperation strategy for signal classification is formulated as maximizing the classification probability subject to constrained misclassification probabilities. We show that this problem can be transformed into a convex problem under some conditions and linear cooperation is a simple but effective strategy that can greatly enhance the performance of signal classification over networked agents.

Index Terms— M-ary hypothesis testing, signal classification, data fusion, distributed inference, convex optimization.

1. INTRODUCTION

Signal classification has been developed into a widely practiced field with numerous applications, including cognitive radios [1], image analysis and processing [2], speech recognition, fingerprint identification, seismic signal analysis, radar target classification, and medical diagnosis, etc. Signal classification is defined as the categorization of input data into identifiable pattern classes using the extraction of important data from a background of impertinent details [3].

Distributed inference is the task of signal detection/classification, parameter estimation, and target tracking by a network of agents, based on cooperation. There have been many works in literature on distributed binary detection [4, 5, 6] in the last two decades, as well as its applications for wireless sensor networks [7] and cognitive radios [8]. However, the distributed solution for signal classification is relatively limited.

This work was supported by FRG-SUSTC1501A-08 and NSFC grant 61540044. Email: quanz@sustc.edu.cn.

In this paper, we focus on cooperation strategies for distributed inference that enables signal classification by multiple agents over the network. The problem of signal classification was approached through M-ary hypothesis testing, where a network of spatially distributed agents observe the signal independently, calculates local statistics, and then forward them to a fusion center for making a global decision[9]. We here propose a linear cooperation strategy that uses the weighted sum of the local statistics for the M-ary hypothesis testing. The distributed signal classification with linear cooperation is formulated as maximizing the classification probability with the misclassification probabilities being constrained. We show that the problem can be transformed into a convex optimization problem. The design of such a signal classification network requires careful analysis and optimization in order to provide optimal classification and assess the system performance.

2. SYSTEM MODEL

Consider a signal generated from M possible hypotheses $\{\mathcal{H}_0, \mathcal{H}_1, ..., \mathcal{H}_{M-1}\}$ with equal *prior* probabilities. A network of K agents are deployed spatially over the field to sense the transmitted signal. Under hypothesis $\mathcal{H}_i$, the signal received by the k-th agent at the n-th time instant is given as

$$x^{(k)}(n) = h_k s_i(n) + v^{(k)}(n), \quad n \in \{0, 1, ..., N-1\} \tag{1}$$

where $s_i(n)$ is the signal under hypothesis $\mathcal{H}_i$, h_k the channel gain, and $v^{(k)}(n)$ the additive white Gaussian noise (AWGN), i.e., $v^{(k)}(n) \sim \mathcal{N}\left(0, \sigma^2\right)$. It is assumed that h_k remains unchanged within each operation period. For simplicity, denote $\boldsymbol{v} = [v^{(1)}(n), v^{(2)}(n), ..., v^{(K)}(n)]^T$, i.e., $\boldsymbol{v} \sim N(0, \boldsymbol{\Sigma_v})$, and $\boldsymbol{x} = [x^{(1)}(n), x^{(2)}(n), ..., x^{(K)}(n)]^T$. Without loss of generality, the noise samples are assumed to be independent and identically distributed (i.i.d) over time and across agents.

In the classification problem (1), we should choose $\mathcal{H}_i$ for which $p(\boldsymbol{x}|\mathcal{H}_i)$ is maximized. The optimal local solution at agent k is the minimum distance receiver [10] and thus we choose $\mathcal{H}_i$ if

$$T_i^{(k)} = \sum_{n=0}^{N-1} x^{(k)}(n) s_i(n) - \tfrac{1}{2}\varepsilon_i \tag{2}$$

is the maximum among $\{T_0^{(k)}, T_1^{(k)}, ..., T_{M-1}^{(k)}\}$, where $\varepsilon_i = \sum_{n=0}^{N-1} |s_i(n)|^2$ is the energy of signal $\boldsymbol{s_i}$. Correlation based methods have also been shown to be optimal for signal classification in the frequency domain over cognitive radios networks [1, 11].

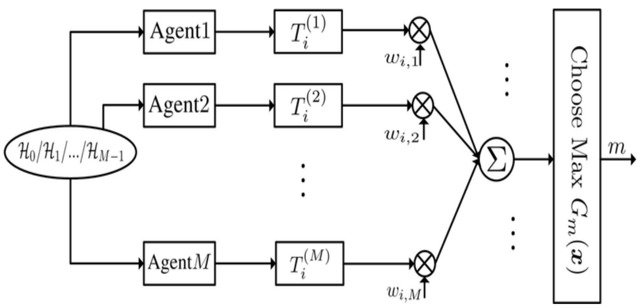


Fig. 1. Schematic representation of linear cooperation for M-ary hypothesis testing. Since $T_i^{(k)}$ is transmitted to the fusion center through coded bits, we don't need to consider the channel distortion between agents and the fusion center.

In this paper, instead of making a local decision discussed above each agent forwards its local statistic $T_i^{(k)}$ to the fusion center for a global decision. We propose a linear cooperation method based on the minimum distance receiver (2) to combine the statistics from the K agents, as illustrated in Fig. 1. That is,

$$G_i = \sum_{k=1}^{K} w_{i,k} T_i^{(k)} = \boldsymbol{w_i}^T \boldsymbol{T}_i \tag{3}$$

where $w_{i,k}$ is the weight coefficient of the k-th agent for $\mathcal{H}_i$, $\boldsymbol{w_i} = [w_{i,1}, w_{i,2}, ..., w_{i,K}]^T$ and $\boldsymbol{T}_i = [T_i^{(1)}, T_i^{(2)}, ..., T_i^{(K)}]^T$.

The weight of a particular agent stands for its contribution to the global decision. Intuitively, the agent with a better chance (usually under a higher SNR) to make the decision correctly should be assigned a lager weighting coefficient. In this paper, the global decision rule for linear cooperation is to choose $\mathcal{H}_m$ if G_m is the maximum among $\{G_0, G_1, ..., G_{M-1}\}$, i.e,

$$m = \arg \max_{0 \le i \le M-1} G_i. \tag{4}$$

3. PROBLEM FORMULATION

For each hypothesis $\mathcal{H}_i$, we obtain a vector $\boldsymbol{w}_i$ for calculating G_i in (3). Then we apply (4) to decide the most likely hypothesis $\mathcal{H}_m$.

To determine the error probability associated with (4) is generally difficult since an error occurs if any of the $M-1$ statistics exceeds the one associated with the true hypothesis. In this paper, we apporach the error probability in a different manner. We use a threshold τ_i to decide $\mathcal{H}_i$ if

$$G_i \ge \tau_i. \tag{5}$$

Correspondingly, the classification probability $P_{C_i} = P(\mathcal{H}_i|\mathcal{H}_i)$ is given by

$$P_{C_i} = P\left(G_i \ge \tau_i|\mathcal{H}_i\right) \tag{6}$$

and the misclassification probability $P_{M_{i,j}} = P(\mathcal{H}_i|\mathcal{H}_j)$

$$P_{M_{i,j}} = P\left(G_i \ge \tau_i|\mathcal{H}_j\right), \quad j \ne i. \tag{7}$$

Our objective is to maximize the classification probability P_{C_i} while the individual error probability $P_{M_{i,j}}$ is constrained by a specified value $\epsilon_{i,j}$, $j \ne i$. That is,

$$\begin{aligned} \max_{(\boldsymbol{w}_i, \tau_i)} \quad & P_{C_i} \\ \text{st.} \quad & P_{M_{i,j}} \le \epsilon_{i,j}, \\ & \qquad \forall j \in \{0, 1, ..., M-1\}, \ \ j \ne i. \end{aligned} \tag{8}$$

Recall the decision rule (4) based on linear corporation (3). Under hypothesis $\mathcal{H}_i$, the statistic G_i should be the maximum one among $\{G_j\}_{0 \le i \le M-1}$, but the noise and channel distortion might cause misclassification. The weight vector $\boldsymbol{w}_i$ plays an important role in minimizing misclassification errors. That is said, the distributed system provides the flexibility to separate G_i away from $\{G_j\}_{j \ne i}$ in signal space by manipulating $\boldsymbol{w}_i$. This flexibility greatly improve the reliability and robustness of the system, even where some agents might be fooled by the noise, channel, or device malfunction. Practically, the design of such a distributed signal classification system should have $\epsilon_{i,j} < 0.5$. We will show in the next section how these considerations can simplify the analysis.

4. OPTIMAL SOLUTION

To solve (8) for each $\mathcal{H}_i$, we first look into the two probabilities P_{C_i} and $P_{M_{i,j}}$. According to (1) and (2), the statistic $T_i^{(k)}$ is Gaussian distributed with

$$E[T_i^{(k)}] = \begin{cases} (h_k - \frac{1}{2})\varepsilon_i, & \mathcal{H}_i \\ h_k \sum_{n=0}^{N-1} s_j(n) s_i(n) - \frac{1}{2}\varepsilon_i, & \mathcal{H}_j \end{cases} \tag{9}$$

Denote $\boldsymbol{\mu_{ij}}$ by the mean of $\boldsymbol{T}_i$ under hypothesis $\mathcal{H}_j$. The variance of the k-th agent

$$Var\left(T_i^{(k)}\right) = \left(\sum_{n=0}^{N-1} s_i^2(n)\right) (\boldsymbol{\Sigma_v})_{kk} \tag{10}$$

is all the same under different hypotheses, where $(\boldsymbol{\Sigma_v})_{kk}$ is the kk-th element of the noise covariance matrix $\boldsymbol{\Sigma_v}$. Also, the covariance between $T_i^{(k)}$ and $T_i^{(m)}$

$$\begin{aligned} & Cov\left(T_i^{(k)}, T_i^{(m)}\right) \\ & = E\left[\left(\sum_{n=0}^{N-1} v_k(n) s_i(n)\right)\left(\sum_{n=0}^{N-1} v_m(n) s_i(n)\right)\right] \\ & = \left(\sum_{n=0}^{N-1} s_i[n]\right)^2 (\boldsymbol{\Sigma_v})_{km} \end{aligned} \tag{11}$$

is the same under different hypotheses.

From (3), it is obvious that $G_i(\boldsymbol{x})$ is also Gaussian distributed. Thus, if $\mathcal{H}_j$ is true we have

$$E\left(G_i|\mathcal{H}_j\right) = \boldsymbol{w}_i^T E\left(\boldsymbol{T}_i|\mathcal{H}_j\right) = \boldsymbol{w}_i^T \boldsymbol{\mu_{ij}} \tag{12}$$

and

$$\begin{aligned} Var\left(G_i|\mathcal{H}_j\right) &= E\left[(G_i - \boldsymbol{\mu_{ij}})^2|\mathcal{H}_j\right] \\ &= \boldsymbol{w_i}^T E\left[(\boldsymbol{T}_i - \boldsymbol{\mu_{ij}})(\boldsymbol{T}_i - \boldsymbol{\mu_{ij}})^T|\mathcal{H}_j\right] \boldsymbol{w_i} \\ &= \boldsymbol{w_i}^T \boldsymbol{\Sigma_{ij}} \boldsymbol{w_i} \end{aligned} \tag{13}$$

where the kk-th element of $\boldsymbol{\Sigma_{ij}}$ is given by (10), and the kj-th element is given by (11). As a matter of fact, $\boldsymbol{\Sigma_{ij}}$ is not relevant to hypothesis $\mathcal{H}_j$, such that we can represent it as $\boldsymbol{\Sigma_i}$. Therefore, if $\mathcal{H}_j$ is true, we have $G_i(\boldsymbol{x}) \sim N(\boldsymbol{w_i}^T\boldsymbol{\mu_{ij}}, \boldsymbol{w_i}^T\boldsymbol{\Sigma_i}\boldsymbol{w_i})$.

Moreover,

$$P_{C_i} = Q\left(\frac{\tau_i - \boldsymbol{\mu_{ii}}^T\boldsymbol{w_i}}{\sqrt{\boldsymbol{w_i}^T\boldsymbol{\Sigma_i}\boldsymbol{w_i}}}\right) \tag{14}$$

and

$$P_{M_{i,j}} = Q\left(\frac{\tau_i - \boldsymbol{\mu_{ij}}^T\boldsymbol{w_i}}{\sqrt{\boldsymbol{w_i}^T\boldsymbol{\Sigma_i}\boldsymbol{w_i}}}\right) \tag{15}$$

where $Q(.)$ is the tail probability of a zero mean unit variance Gaussian variable.

Since $Q(.)$ is a monotonic decreasing function, problem (8) is equivalent to

$$\begin{aligned} \min_{(\boldsymbol{w}_i,\tau_i)} \quad & \frac{\tau_i - \boldsymbol{\mu_{ii}}^T\boldsymbol{w_i}}{\sqrt{\boldsymbol{w_i}^T\boldsymbol{\Sigma_i}\boldsymbol{w_i}}} \\ \text{st.} \quad & \frac{\tau_i - \boldsymbol{\mu_{ij}}^T\boldsymbol{w_i}}{\sqrt{\boldsymbol{w_i}^T\boldsymbol{\Sigma_i}\boldsymbol{w_i}}} \geq Q^{-1}(\epsilon_{i,j}) \\ & \forall j \in \{0, 1, ..., M-1\}, j \neq i \end{aligned} \tag{16}$$

where the scalar τ_i and vector $\boldsymbol{w}_i$ are the variables that we need to optimize. For simplicity, we can write them into one vector, that is,

$$\tau_i - \boldsymbol{\mu_{ij}}^T\boldsymbol{w_i} = \begin{pmatrix} 1 & -\boldsymbol{\mu_{ij}}^T \end{pmatrix} \begin{pmatrix} \tau_i \\ \boldsymbol{w_i} \end{pmatrix} \tag{17}$$

and

$$\boldsymbol{w_i}^T\boldsymbol{\Sigma_i}\boldsymbol{w_i} = (\tau_i\ \boldsymbol{w_i}^T)\hat{\boldsymbol{\Sigma}}_i(\tau_i\ \boldsymbol{w_i}^T)^T \tag{18}$$

where

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}} = \begin{pmatrix} 0 & \boldsymbol{0}_{1\times(K-1)} \\ \boldsymbol{0}_{(K-1)\times 1} & \boldsymbol{\Sigma_i} \end{pmatrix}. \tag{19}$$

By introducing new variables

$$\boldsymbol{g_{ij}} = \begin{pmatrix} 1 \\ -\boldsymbol{\mu_{ij}} \end{pmatrix} \tag{20}$$

and

$$\boldsymbol{u_i} = \frac{(\tau_i\ \boldsymbol{w_i}^T)^T}{\sqrt{(\tau_i\ \boldsymbol{w_i}^T)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}(\tau_i\ \boldsymbol{w_i}^T)^T}}, \tag{21}$$

where $\boldsymbol{u_i}$ is constrained by

$$\boldsymbol{u_i}^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}\boldsymbol{u_i} = 1, \tag{22}$$

we can transform (16) into an equivalent form

$$\begin{aligned} \min_{\boldsymbol{u}_i} \quad & \boldsymbol{g_{ii}}^T\boldsymbol{u_i} \\ \text{st.} \quad & \boldsymbol{u_i}^T\hat{\boldsymbol{\Sigma}}_i\boldsymbol{u_i} = 1 \\ & \boldsymbol{g_{ij}}^T\boldsymbol{u_i} \geq Q^{-1}(\epsilon_{i,j}), \\ & \qquad \forall j \in \{0, 1, ..., M-1\}, j \neq i. \end{aligned} \tag{23}$$

The objective is a linear function constrained by $M-1$ linear and one quadratic constraints. As the quadratic constraint is non-convex, the optimization problem is non-convex and cannot be solved efficiently using the existing methods.

Fortunately, we can relax the quadratic equality constraint to a quadratic inequality constraint that is convex, i.e.,

$$\boldsymbol{u_i}^T\hat{\boldsymbol{\Sigma}}_i\boldsymbol{u_i} \leq 1. \tag{24}$$

We later show that the optimal solution of (23) always occurs on the edge of the ellipsoid (24).

Recall from (2) that the minimum distance receiver exploited the correlation between the received signal and the hypothesis under consideration. Consequently, $E\left(G_i\right)$ should be greater if the received signal is from hypothesis $\mathcal{H}_i$ than from other hypotheses. Otherwise, $\mathcal{H}_i$ cannot be distinguishable from other hypotheses using the decision rule (4). Therefore, it is always true throughout this paper that $\boldsymbol{\mu_{ii}} \succ \boldsymbol{\mu_{ij}}$ and $\boldsymbol{g_{ij}} \succeq \boldsymbol{g_{ii}}$, where the notation $\succeq$ means component-wise inequality for vectors and $\succ$ is strictly component-wise inequality.

Consider the case where $M = 3$ (i.e., $\mathcal{H}_0$, $\mathcal{H}_1$, and $\mathcal{H}_2$) and $K = 1$. For $\mathcal{H}_2$, the second element of $\boldsymbol{g}_{22}$ is $\boldsymbol{g}_{22}(1) = -\mu_{22}$. Please note that $\boldsymbol{g}_{ij}(0)$ is always one from (20). If $\boldsymbol{g}_{22}(1) \geq 0$, we have

$$0 \leq \boldsymbol{g}_{22}(1) < \boldsymbol{g}_{20}(1) < \boldsymbol{g}_{21}(1). \tag{25}$$

As illustrated in Fig.2, the feasible set of (24) is split into the sector unshadowed. Because the objective function $\boldsymbol{g}_{22}^T\boldsymbol{u}_2$ is a line with slope less than $\boldsymbol{g}_{20}^T\boldsymbol{u}_2$ and $\boldsymbol{g}_{21}^T\boldsymbol{u}_2$, the optimal solution can be obtained by moving the line $\boldsymbol{g}_{22}^T\boldsymbol{u}_2$ over the sector towards the origin. In this example, the optimum occurs at A where $\boldsymbol{u_i}^T\hat{\boldsymbol{\Sigma}}_i\boldsymbol{u_i} = 1$ and $\boldsymbol{g}_{20}^T\boldsymbol{u}_2$ intersect.

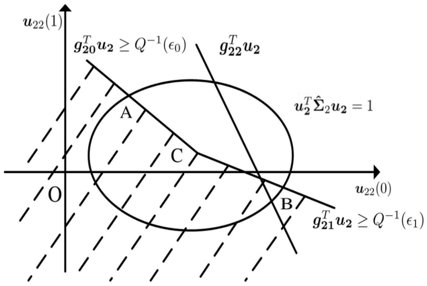


Fig. 2. A geometric illustration of the non-convex optimization problem (23).

Similarly, we can also show that the optimum occurs on the edge of the ellipsoid $\boldsymbol{u_i}^T \hat{\boldsymbol{\Sigma}}_i \boldsymbol{u_i} \leq 1$ for $\boldsymbol{g}_{22}(1) < 0$.

As a result, the non-convex problem (23) is equivalent to the following convex problem from relaxation

$$\begin{aligned} \min_{\boldsymbol{u}} \quad & \boldsymbol{g_{ii}}^T \boldsymbol{u_i} \\ \text{st.} \quad & \boldsymbol{u_i}^T \hat{\boldsymbol{\Sigma}}_i \boldsymbol{u_i} \leq 1 \\ & \boldsymbol{g_{ij}}^T \boldsymbol{u_i} \geq Q^{-1}(\epsilon_{i,j}) \\ & \qquad \forall j \in \{0, 1, ..., M\}, j \neq i \end{aligned} \tag{26}$$

The problem can be readily solved using a standard toolbox such as CVX [12]. Once the optimal solution of (26) is obtained, denoted by $\hat{\boldsymbol{u}}_i$, we can calculate the optimal weight vector $\hat{\boldsymbol{w}}_i$ using

$$\hat{\boldsymbol{w}}_i = \frac{\hat{\boldsymbol{u}}_{\boldsymbol{i}}(2:M)}{\|\hat{\boldsymbol{u}}_{\boldsymbol{i}}(2:M)\|}. \tag{27}$$

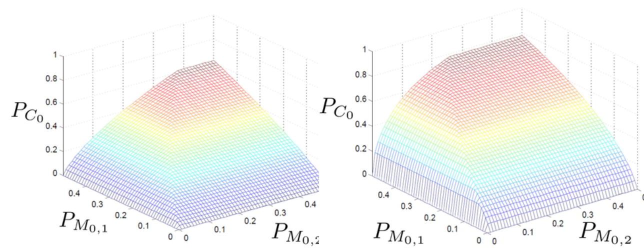
Please note that the result of (8) is not relevant to $\|\boldsymbol{w}_i\|$.

5. SIMULATION RESULT

In this section, we assess the performance of the proposed linear corporation scheme for the distributed signal classification system. Consider the case where we have $M = 3$ hypotheses, i.e., $\boldsymbol{s}_0 = [1, 0, 0.5, 0, 1, 0, 0.5, 0]$, $\boldsymbol{s}_1 = [1, 0, 0, 0.5, 1, 0, 0, 0.5]$ and $\boldsymbol{s}_2 = [1, 0, 0.5, 0, 1, 0, 0, 0.5]$. We would like to compare the classification performance of a single agent (A) against that of a network of two agents (B and C). In the case of a single agent, the channel gain $h_0 = 3$ and $\boldsymbol{\Sigma_v} = 1$. In the case of two agents, the channel gains are $h_0 = 2$ and $h_1 = 2.2$, which are worse than the case of a single agent, with $\boldsymbol{\Sigma_v} = [1, 0; 0, 1]$. That is said, the observations of B and C have lower SNRs than that of A. To visualize the performance of signal classification, we plot the correct classification probability P_{C_0} versus the misclassification probabilities $P_{M_{0,1}}$ and $P_{M_{0,2}}$ in a 3D figure.

For $\mathcal{H}_0$, we can obtain $\boldsymbol{w_0}$ by solving (8). As shown in Fig.3 (a) and Fig.3 (b), a network of two agents with cooperation outperforms a single agent, although the two agents (B and C) have signal quality disadvantages compared with the single agent (A).

Also we can see from Fig.3 (a) and Fig.3 (b) that P_{C_0} is mainly constrained by $P_{M_{0,2}} \leq \epsilon_{0,2}$, i.e.,, $\boldsymbol{g}_{0,2}^T \boldsymbol{u}_0 \geq Q^{-1}(\epsilon_{0,2})$. This is caused by the fact that $\boldsymbol{g}_{0,2} \preceq \boldsymbol{g}_{0,1}$. When $Q^{-1}(\epsilon_{0,2}) = Q^{-1}(\epsilon_{0,1})$, the feasible set of $\boldsymbol{u}_0$ constrained by $\boldsymbol{g}_{0,2}^T \boldsymbol{u}_0 \geq Q^{-1}(\epsilon_{0,2})$ is a subset of the feasible set constrained by $\boldsymbol{g}_{0,1}^T \boldsymbol{u}_0 \geq Q^{-1}(\epsilon_{0,1})$. In this case, the constraint $\boldsymbol{g}_{0,2}^T \boldsymbol{u}_0 \geq Q^{-1}(\epsilon_{0,2})$ dominates the problem. Moreover, P_{C_0} increases faster as $P_{M_{0,1}}$ increases since the distance from $\boldsymbol{s}_0$ to $\boldsymbol{s}_1$ is greater than that from $\boldsymbol{s}_0$ to $\boldsymbol{s}_2$ in the signal space, i.e., $\|\boldsymbol{s}_0 - \boldsymbol{s}_1\| > \|\boldsymbol{s}_0 - \boldsymbol{s}_2\|$. In other words, it is less likely to misclassify $\boldsymbol{s}_1$ as $\boldsymbol{s}_0$.



(a) Single agent. (b) Two agents with linear operation.

Fig. 3. The correct classification probability versus the misclassification probabilities for $M = 3$.

The total classification probability is given by $P_C = \frac{1}{3}(P_{C_0} + P_{C_1} + P_{C_2})$ as the hypotheses have equal prior probabilities. The receiver operating characteristic (ROC) curve plotted in Fig.4 also shows the benefit of cooperation among agents.

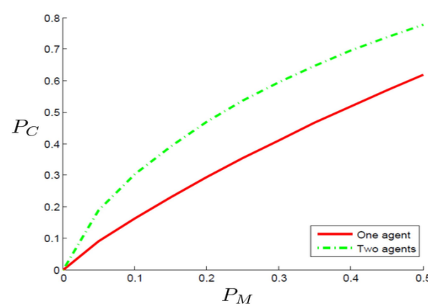


Fig. 4. The ROC curve with $P_M = P_{M_{0,1}} = P_{M_{0,2}} = P_{M_{1,0}} = P_{M_{1,2}} = P_{M_{2,0}} = P_{M_{2,1}}$.

6. CONCLUSION

In this paper, we study a distributed system that is able to identify a signal from multiple hypotheses. We have proposed a simple but effective linear cooperation scheme for a network of agents to jointly perform the task of signal classification. The design of such a distributed system can be formulated into a convex program that minimizes the error probability.

7. REFERENCES

[1] A. Fehske, J. D. Gaeddert, and J. H. Reed, "A new approach to signal classification using spectral correlation and neural networks," in *Proc. IEEE DySPAN*, 2005, pp. 144–150.

[2] B. Lerner, W. F. Clocksin, S. Dhanjal, M. A. Hulten, and C. M. Bishop, "Feature representation and signal classification in fluorescence in-situ hybridization image analysis," *IEEE Trans. Systems, Man, and Cybernetics, Part A*, vol. 31, no. 6, pp. 655–665, 2001.

[3] G. J. Miao and M. A. Clements, *Digital Signal Processing and Statistical Classification*, Artech House, 2002.

[4] S. Boyd and L. Vandenberghe, *P. K. Varshney, Distributed Detection and Data Fusion*, New York:Springer-Verlag, 1997.

[5] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part ii–advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.

[6] Z. Quan, W.-K. Ma, S. Cui, and A. H. Sayed, "Optimal linear fusion for distributed detection via semidefinite programming," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2431–2436, Apr. 2010.

[7] B. Chen and P. K. Willett, "On the optimality of the likelihood-ratiotest for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 693–699, Feb. 2005.

[8] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 28–40, Feb. 2008.

[9] Bernard C. Levy, *Principles of Signal Detection and Parameter Estimation*, pp. 16–25, Springer, 2008.

[10] S. M. Kay, *Fundamentals of statistical signal processing: Detection theory*, vol. II, pp. 119–122, Prentice Hall, 1993.

[11] Z. Quan, W. Zhang, S. J. Shellhammer, and A. H. Sayed, "Optimal spectral feature detection for spectrum sensing at very low snr," *IEEE Trans. Wireless Commun.*, vol. 59, no. 1, pp. 201–212, 2011.

[12] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambrige, U.K.:Cambridge Univ. Press, 2003.