Nonparametric Decentralized Detection Based on Weighted Count Kernel

Jiayao Hu and Yingbin Liang
Dept of Electrical Engineering and Computer Science
Syracuse University
Syracuse, NY 13244 USA
Email: {hjiayao,yliang06}@syr.edu

Eric P. Xing
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15203 USA
epxing@cs.cmu.edu

Abstract—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. **The nonparametric decentralized detection problem is investigated, in which the joint distribution of the environmental event and the sensors' observations are not known and only a set of training samples are available. The system features rate constraints, i.e., integer bit constraints on sensors' transmissions, different qualities of observations, additional observations to the fusion center, and multi-level tree-structured network. Our study adopts the kernel-based nonparametric approach proposed by Nguyen, Wainwright, and Jordan with the following generalization. A weighted count kernel is introduced so that the corresponding reproducing kernel Hilbert space (RKHS) (over which the fusion center's decision rule is optimized) allows the fusion center's decision rule to count information from sensors and its own observations differently. In order to find the optimal decision rules, our optimization is solved by alternatively and recursively conducting three optimization steps: finding the optimal weight parameters in the weighted count kernel for selecting the best associated RKHS, finding the best optimal decision rule for the fusion center over the identified RKHS, and finding the local decision rules for sensors. Generalization to multilevel tree-structured network is also discussed. Finally numerical results are provided to demonstrate the performance based on the proposed weighted count kernel.**

I. INTRODUCTION

As a classical decision-making problem, decentralized detection has been extensively studied in the literature, e.g., [1]–[3] and references therein. Most of previous work on this topic used parametric approaches, which assumed the joint distribution of the environmental event and the sensors' observations are known in advance.

Nonparametric (de)centralized detection was studied previously in, e.g., [4], [5], which employed detectors that perform well for certain statistical environments. A learning-based nonparametric linear regression problem was studied in [6], [7]. More recently, a kernel-based classification approach was proposed in [8], which is more generally applicable with mathematical guarantee on the performance. The basic idea is to introduce a kernel function that determines a reproducing kernel Hilbert space (RKHS), over which the decision rule of the fusion center is searched to optimize a given risk function. It has been shown by numerical examples in [8] that the kernel-based approach yields better performances than other approaches based on estimating joint distributions. Furthermore, compared to parametric approaches, such a kernel-based nonparametric approach is also applicable for the case with correlated observations, in which the correlation is implicitly embedded in training data and their influence on the decision rules are automatically incorporated by optimizing empirical risk functions determined by the training data. In our previous work [9], we generalized the kernel-based approach in [8] to tree-structured sensor networks, and proposed a distributed protocol which achieves an efficient implementation for finding the optimal decision rules over a tree structure.

In this paper, we study more realistic sensor networks, which generalizes the models studied in [8], [9] to include several new features: (1) sensors' observations can have different qualities and hence different alphabet sizes due to their different locations in capturing the environmental event; (2) the fusion center can receive observations of the environmental event directly; (3) sensors' transmissions to the fusion center are subject to certain rate bit constraints and hence sensors' quantization levels can be different; and (4) sensors may be networked in a multilevel tree structure toward the fusion center. Our goal is to characterize the impact of these practical features on the decision rules of the fusion center and sensors in the nonparametric decentralized detection.

Our study adopts the nonparametric kernel-based approach proposed by Nguyen, Wainwright, and Jordan in [8] with the following generalization. We introduce a weighted count kernel so that the corresponding Hilbert space, i.e., the RKHS, (over which the fusion center's decision rule is optimized) allows the fusion center's decision rule to count information from sensors and its own observations differently based on the quality of these information sources. In this way, by introducing the weight parameters in the weighted count kernel into the risk minimization framework, the best RKHS associated with the weighted count kernel is selected jointly with the decision rules for the fusion center and sensors. Thus, the impact of the network features including the quality of sensors' observations, fusion center's direct observations, and rate constraints on sensors' transmissions are naturally incorporated into the fusion center's decision rules via selecting the RKHS that these decision rules lie in.

We solve the risk minimization for finding the decision rules by recursively and alternatively conducting three optimization steps: finding the optimal weight parameters for selecting the

best associated RKHS associated with the weighted count kernel, finding the optimal decision rule for the fusion center over the identified RKHS, and finding the local decision (i.e., quantization) rules for sensors. For each step, risk functions are typically convex, but not differentiable everywhere. By adopting the approach in [8] based on conjugate dual arguments, we analytically characterize the optimal decision rule for the fusion center and some component in subdifferential for optimizing weight parameters and decision rules for the sensors for efficient implementation of the optimization algorithm. We further discuss the generalization to multilevel tree-structured network, in which the impact of the network structure on the decision rules is also captured by the selection of the optimal weighted count kernel.

We also derive an upper bound on the true risk function based on the approximate empirical risk function, whose asymptotic behavior suggests that additional optimization over RKHSs associated with the weighted count kernel does not require more training samples for the approximate empirical risk function to be close to the true risk function from the above. We finally provide numerical results to demonstrate the impact of the rate constraints of sensors' transmissions to the fusion center and direct observations of the fusion center on the detection error probability based on our weighted count kernel approach.

The rest of the paper is organized as follows. In section II, we provide the necessary background on learning by kernels. In section III, we describe our system model and problem formulation. In section IV, we provide our main results in finding the optimal decision rules. In section V, we discuss about the generalization of our study to multilevel tree-structured sensor networks. In section VI, we provide the simulation results, and finally in section VII, we conclude our paper with a few remarks.

II. BACKGROUND ON KERNELS

In this section, we introduce the basic concepts and definitions on learning by kernels, which is the basic technique applied in this paper.

Definition 1. *A function k on $\mathcal{X}^2 \to \mathcal{R}$ is called a kernel if for all positive integer m and all $x_1, \cdots, x_m \in \mathcal{X}$, the $m \times m$ matrix K with elements $K_{ij} = k(x_i, x_j)$ for $i, j = 1, \ldots, m$ is positive semidefinite.*

Definition 2. *A Hilbert space $\mathcal{H}$ containing functions $f : \mathcal{X} \to R$ is called a reproducing kernel Hilbert space (RKHS) if there exists a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathcal{R}$ with the following properties:*

- *k has the reproducing property:*

  $$\langle f, k(x, \cdot) \rangle = f(x) \quad \text{for all } f \in \mathcal{H},$$

  in particular, $\langle k(x, \cdot), k(x', \cdot) \rangle = k(x, x')$, and
- *k spans $\mathcal{H}$.*

Given a kernel k, we define a feature mapping $\Phi : x \in \mathcal{X} \to k(\cdot, x)$, which maps an element $x \in \mathcal{X}$ to a function.
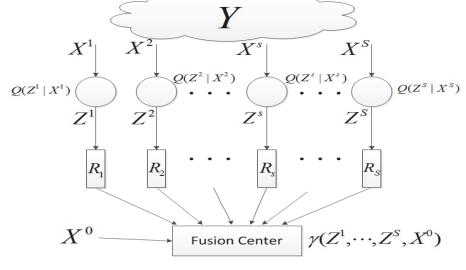

Fig. 1.   A sensor network for decentralized detection

We then define a vector space containing

$$f(\cdot) = \sum_{i=1}^{m} \alpha_i k(\cdot, x_i)$$

where m is any positive integer, $\alpha_i \in \mathcal{R}$, and $x_1, \cdots, x_m \in \mathcal{X}$ are arbitrary. For this vector space, we define an inner product between f and another function $g(\cdot) = \sum_{j=1}^{m'} \beta_j k(\cdot, x'_j)$ as

$$\langle f, g \rangle = \sum_{i=1}^{m} \sum_{j=1}^{m'} \alpha_i \beta_j k(x_i, x'_j).$$

It can be shown that after completing this vector space, we obtain a RKHS associated with the kernel k.

III. MODEL AND PROBLEM FORMULATION

We study the decentralized detection problem over a sensor network (see Fig. 1), in which sensors receive observations about an environmental event, quantize their observations based on their own local decision rules (i.e., quantization rules), and then forward their quantized information to a fusion center, which will make the decision about the state of the environmental event. We use Y to denote the environmental event, which can take binary values $+1$ and -1. We assume there are S sensors in the network. We use $X^s \in \mathcal{X}$ to denote the observation received by sensor s for $s = 1, \ldots, S$, and use Z^s to denote the quantized value of sensor s. The sensor's observation X^s can have different alphabet sizes, which may possibly due to nonuniform noise corruption of signals received by these sensors.

The decision rule of a sensor can be characterized by a probability distribution $Q^s_{Z^s|X^s}(z^s|x^s)$ mapping from its input variable X^s to an output variable Z^s, thus allowing a random decision rule. In particular, we assume that there is a bit constraint R_s (which is assumed to be an integer) on each sensor's transmission to the fusion center, and hence each Z^s has an alphabet size 2^{R_s}. Consequently, Z^s may also have different alphabet size due to different rate constraints. We also assume that the fusion center receives not only quantized information from all sensors but also observations directly from the environment denoted by X^0, and hence the fusion center's decision rule can be written as a function $\gamma(Z^1, \ldots, Z^S, X^0)$.

In our problem, we assume that the joint probability distribution of the event and the observations for all sensors and the fusion center, i.e., $P(Y, X^0, \ldots, X^S)$, is unknown. Instead,

a set of training data are available, i.e., $(x_i^0, \ldots, x_i^S, y_i)$ for $i = 1, \ldots, N$. We adopt the framework of empirical risk minimization for decentralized detection in [8] to find the jointly optimal decision rule γ for the fusion center and decision rules Q^s for all sensors that minimize a given risk function $\phi(\cdot)$ which is properly chosen as the system performance measure.

We consider decision rule for the fusion center that lies in the RKHS $\mathcal{H}$ determined by a kernel function $k(\cdot, \cdot) : (\mathcal{Z} \times \mathcal{X}) \times (\mathcal{Z} \times \mathcal{X}) \to \mathcal{R}$. We note that the domain of the kernel has one more space compared to that in [8] to take into account the observations of the fusion center. As such, we can express the fusion center's decision rule as:

$$\gamma(\underline{z}, x^0) = \langle w, \Phi(\underline{z}, x^0) \rangle_{\mathcal{H}}$$

where $w \in \mathcal{H}$ and $\Phi(\underline{z}, x^0) = k(\cdot, (\underline{z}, x^0))$.

Our problem is then formulated as the following optimization problem:

$$\min_{w \in \mathcal{H}, Q \in \mathcal{Q}} \sum_{\underline{z}} \sum_{i=1}^{N} \phi(y_i \langle w, \Phi(\underline{z}, x^0) \rangle_{\mathcal{H}}) Q(\underline{z}|\underline{x}_i) + \frac{\lambda}{2} \|w\|_{\mathcal{H}}^2 \quad (1)$$

where $\mathcal{Q}$ is the set which includes all possible conditional probabilities for every sensor. In particular, $Q(\underline{z}|\underline{x}_i)$ can be decomposed as $Q(\underline{z}|\underline{x}_i) = \prod_{i=1}^{s} Q(z^s|x^s)$, because sensors follow independent decision rules.

Since it is computationally complex to solve the above optimization problem in general. Similarly to [8], by applying Jensen's inequality, we obtain a lower bound for (1) as a relaxation.

$$\min_{w \in \mathcal{H}, Q \in \mathcal{Q}} \phi(y_i \langle w, \Phi'(\underline{x}_i, x_i^0) \rangle_{\mathcal{H}}) + \frac{\lambda}{2} \|w\|_{\mathcal{H}}^2 \quad (2)$$

where $\Phi'(\underline{x}_i, x_i^0) = \sum_{\underline{z}} \Phi(\underline{z}, x_i^0) Q(\underline{z}|\underline{x}_i) \in \mathcal{H}$. It can be shown that the approximate empirical risk function approaches to the true risk function from the above as the number of training data becomes large as in section IV-D.

IV. MAIN RESULTS

In this section, we introduce a weighted count kernel, which thus defines a Hilbert space of the RKHS that enables to count contributions from sensors differently based on quality of observations, transmission constraints, additional observations of the fusion center, and the impact network structure for designing decision rules of the fusion center and sensors. We then optimize the risk function over the weight parameters in the weighted count kernel so that to select the optimal Hilbert space that the decision rules lie in jointly with the decision rules for the fusion center and sensors.

We introduce the following weighted count kernel, which can be shown to satisfy the definition of kernel given in Definition 1.

$$k_\beta((\underline{z}, x^0), (\underline{z}', x'^0)) = \sum_{s=1}^{S} \beta_s \mathbb{I}[z^s = z'^s] + \beta_0 \mathbb{I}[x^0 = x'^0] \quad (3)$$

where $\mathbb{I}[\cdot]$ is an indicator function, and $\beta_s \geq 0$ for $s = 0, 1, \ldots, S$. Here, each weight parameter, say β_s, represents the contribution of sensor s in the decision rule of the fusion

center. In particular, β_0 represents the contribution of the direct observations of the fusion center. Thus, the Hilbert space $\mathcal{H}_\beta$ over which the decision rule of the fusion center is chosen is spanned by the weighted count kernel $k_\beta(\cdot, \cdot)$.

Remark 1. *Although our study focuses on the weighted count kernel, the idea of introducing weights to kernels for counting information differently may be applicable for more general types of kernels.*

By introducing the optimization of the weight parameters into the risk minimization problem, our optimization problem now becomes:

$$\min_{\substack{\beta \in \mathcal{R}^{S+1}, w \in \mathcal{H}_\beta, Q \in \mathcal{Q} \\ \beta_s \geq 0 \text{ for } s = 0, 1, \ldots, S}} \phi(y_i \langle w, \Phi'_\beta(\underline{x}_i, x_i^0) \rangle_{\mathcal{H}_\beta}) + \frac{\lambda}{2} \|w\|_{\mathcal{H}_\beta}^2 \quad (4)$$

where $\Phi'_\beta(\underline{x}_i, x_i^0) = \sum_{\underline{z}} \Phi_\beta(\underline{z}, x_i^0) Q(\underline{z}|\underline{x}_i)$

We note that without loss of generality, we set one dimension of β_s to be fixed as a reference value. Due to possibly large number of training samples and sensors, dimensions of the parameters to be optimized can be very large. Hence, optimizing over all parameters simultaneously is very complex. Furthermore, the risk function $\phi(\cdot)$ such as the hinge loss function is not differentiable everywhere, which adds more complexity to the problem. Thus, we adopt the coordinate gradient algorithm to recursively and alternatively optimize over these three types of parameters, i.e., β, w and Q. This approach is justified because the objective function is convex over any type of parameters given two other types of parameters.

A. Optimization over β

We optimize the objective function over the weight parameters $\beta_s, s = 0, 1, \ldots, S$ one after another with w and all $Q(\cdot|\cdot)$s fixed. Since risk functions are typically nondifferentiable everywhere (e.g. the hinge loss function we use for our numerical simulation), we characterize an analytical expression for an element in the subdifferential of the objective function with respect to each component β_s in order for implementing the gradient algorithm, which can be complex otherwise.

Proposition 1. *Consider β_0, which corresponds to the contribution of direct observations of the fusion center. An element in the subdifferential of the objective function with respect to the weight parameter β_0 with w, all Qs and other $\beta_s, s \neq 0$ being fixed is given by*

$$\partial_{\beta_0} G = \sum_{i,j} -\lambda \alpha_i \alpha_j y_i y_j \mathbb{I}[x_i^0 = x_j^0], \quad (5)$$

where G denotes the objective function.
Consider β_s for $s = 1, \ldots, S$, which corresponds to the contribution of sensor s. An element in the subdifferential of the objective function with respect to the weight parameter β_s with w, all Qs and all other $\beta_{s'}, s' \neq s$ being fixed is given by

$$\partial_{\beta_s} G = \sum_{i,j,z^s} -\lambda \alpha_i \alpha_j y_i y_j Q(z^s|x_i^s) Q(z^s|x_j^s) \quad (6)$$

The above proposition can be proved by conjugate dual arguments, which is omitted here due to the space limitations.

B. Optimization over w

Given the optimal RKHS associated with k_β, we now find the optimal decision rule for the fusion center by optimizing $G(\beta; w; Q)$ with respect to w. Due to the arguments similar to the proof of the kernel representor theorem [10], the optimal $w \in \mathcal{H}_\beta$ can be expressed as $w = \sum_{i=1}^{N} \alpha_i y_i \Phi'_\beta(\underline{x}_i, x_i^0)$. Following the arguments in [8], the coefficients α_i for $i = 1, \ldots, N$ in w should solve the following maximization problem.

$$\max_{0 \le \alpha_i \le \frac{1}{\lambda}} \left\{ \sum_i^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \langle \Phi'_\beta(\underline{x}_i, x_i^0), \Phi'_\beta(\underline{x}_j, x_j^0) \rangle_{\mathcal{H}_\beta} \right\}.$$

The above problem is a quadratic optimization problem and can hence be solved easily by alternatively updating the value of each α_i. We note that the difference here from [8] is that the inner product $\langle \Phi'_\beta(\underline{x}_i, x_i^0), \Phi'_\beta(\underline{x}_j, x_j^0) \rangle_{\mathcal{H}_\beta}$ is taken over the Hilbert space RKHS which allows the fusion center's decision rule to count contributions from sensors and its own observations differently based on the quality of sensors' observations, the quality of the fusion center's observations, and the rate constraints on sensors' transmissions.

C. Optimization over $Q(\cdot|\cdot)$

In this subsection, we find the optimal decentralized decision rules for sensors. Similar to the optimization over β, the major step here is to find an element in the subdifferential of the objective function $G(\beta; w; Q)$ with respect to each $Q^s(\cdot|\cdot)$.

Proposition 2. *Consider sensor s for $s = 1, \ldots, S$. An element in the subdifferential of the objective function with respect to the local decision rule of sensor s with w, β and all other $Q(\cdot|\cdot)s$ being fixed is given by*

$$-\lambda \sum_{i,j,\underline{z},\underline{z}'} \alpha_i \alpha_j Q(\underline{z}'|\underline{x}_j) \frac{Q(\underline{z}|\underline{x}_i)}{Q(x^s|x_i^s)} k_\beta((\underline{z}, x_i^0), (\underline{z}, x_j^0)) \tag{7}$$
$$\times \mathbb{I}[x_i^s = \bar{x}_i^s] \mathbb{I}[z^s = \bar{z}^s]$$

The proof here is based on conjugate dual arguments and is omitted due to the space limitations.

D. Bounds on True Risk Function

In this section, we derive bounds on the true risk function $\mathbb{E}\phi(Y\gamma(\underline{Z}, X^0))$ based on the approximate empirical risk function $\inf_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N \phi(y_i \langle w, \Phi'_\beta(\underline{x}_i, x_i^0) \rangle_{\mathcal{H}_\beta})$, where $\gamma(\underline{Z}, X^0) = \langle w, \Phi_\beta(\underline{Z}, X^0) \rangle_{\mathcal{H}_\beta}$, $w \in \mathcal{H}_\beta$, $Q \in \mathcal{Q}$, $\beta \in \mathcal{R}^{S+1}$, and $\mathcal{F}$ corresponds to the set $\mathcal{H}_\beta \times \mathcal{Q} \times \mathcal{R}^{S+1}$.

Following the arguments similar to [8], we show that with a probability at least $1 - 2\delta$ the true ϕ-risk $\mathbb{E}\phi(Y\gamma(\underline{Z}, X^0))$

is bounded by the approximate empirical ϕ-risk as follows:

$$\inf_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N \phi(y_i \langle w, \Phi'_\beta(\underline{x}_i, x_i^0) \rangle_{\mathcal{H}_\beta}) - 2l R_N(\mathcal{F}) - \sqrt{\frac{\ln(\frac{2}{N})}{2N}}$$
$$\le \inf_{f \in \mathcal{F}} \mathbb{E}\phi(Y\gamma(\underline{Z}, X^0)) \le$$
$$\inf_{f \in \mathcal{F}_0} \frac{1}{N} \sum_{i=1}^N \phi(y_i \langle w, \Phi'_\beta(\underline{x}_i, x_i^0) \rangle_{\mathcal{H}_\beta}) + 2l R_N(\mathcal{F}_0) + \sqrt{\frac{\ln(\frac{2}{N})}{2N}}$$

where ϕ is Lipschitz with constant l, $\mathcal{F}_0$ includes only deterministic rules for sensors, and $R_N(\mathcal{F})$ is the Rademacher complexity of the function class $\mathcal{F}$ given in [8]. It is clear that the bounds on the Rademacher complexity characterize how close the approximate empirical risk function that the designed decision rules achieve is to the true risk function. In the following, we provide an bound on the Rademacher complexity for our problem.

Proposition 3. *Let the alphabet sizes of the observations be bounded by C_x and the alphabet sizes of the quantized variables by all sensors be bounded by C_z, i.e. the rate constraints from the sensors to the fusion center be bounded by $\log C_z$. An upper bound on the Rademacher complexity is given by*

$$R_N(\mathcal{F}_0) \le \frac{2B\sqrt{D}}{N} \left[\mathbb{E} \sup_{Q \in \mathcal{Q}_0} \sum_{i=1}^N \langle \Phi'(\underline{x}_i, x_i^0), \Phi'(\underline{x}_i, x_i^0) \rangle_{\mathcal{H}_\beta} + \right.$$
$$\left. 2(N-1)\sqrt{\frac{N}{2}} \sup_{\underline{z}, \underline{z}'} k'((\underline{z}, x^0), (\underline{z}', x'^0)) \sqrt{2 C_x(S+1) \log C_z} \right]^{\frac{1}{2}}$$

Where $k'((\underline{z}, x^0), (\underline{z}', x'^0)) = \sum_{s=1}^S \mathbb{I}[z^s = z'^s] + \mathbb{I}[x^0 = x'^0]$, $\Phi'(\underline{x}_i, x_i^0) = \sum_{\underline{z}} k'((\underline{z}, x_i^0), \cdot) Q(\underline{z}|\underline{x}_i)$, B is the upper bound on $\|w\|_{\mathcal{H}_\beta}$, and D is the upper bound on β_s.

We note that the above bound approaches zero as the number N of training samples approaches infinity, thus providing a tightest upper bound. It can be seen from the above that the number of samples needed for $R_N(\mathcal{F}_0)$ to approach zero is the same as in [8] suggesting that optimization over weight parameters in the weighted count kernel does not require more number of samples for the approximate empirical risk function to be close to the true risk function.

V. GENERALIZATION TO TREE-STRUCTURED NETWORKS

In this section, we briefly describe how to generalize our study to a tree-structured sensor network. In particular, we now consider a sensor network with sensors configured in a tree structure with the fusion center being the root of the tree. All sensors and the fusion center can receive observations of an environmental event. Sequentially from the leave sensors, each sensor node quantizes its observation and its received information from all its children in the tree and then forwards the quantized information to its parent. Finally, the root, which is the fusion center, makes a decision about the event state based on its own observations and the information received from all its children sensors. There are rate constraints (i.e.,

		$S=1$	$S=2$	$S=3$	$S=4$	$S=5$
Case 1	Error	0.0452	0.0205	0.0061	0.0053	0.0016
	Error	0.0154	0.0077	0.0034	0.0020	0.0007
Case 2	β_0	3.4712	2.8214	3.0838	3.3351	2.4300

integer bit constraints) for all transmission links in the network so that the quantized alphabet size of each sensor is determined by the rate constraint on this sensor's transmission to its parent.

The optimization problem for finding the decision rules for the fusion center and sensors can be formulated and solved in the same fashion as for the one-level sensor network studied in Section IV. The difference lies in that now the optimization over the weights in the weighted count kernel also takes into account the impact of the network structure on the fusion center's decision rule.

VI. NUMERICAL RESULTS

We first study the impact of the fusion center's observations on the performance of the system. In particular, we compare the decision error probabilities for two cases with the fusion center having or not having observations, respectively. For both cases, we assume that the sensors have independent observations. We generate 300 data samples for training, and 10000 data samples for testing. For each sample, we generate y_i with uniform probability on $+1$ and -1. We then generate the noise variable n^s for each sensor and the fusion center independently, where $P(n^s = 0) = 0.6, P(n^s = +1) = 0.2$ and $P(n^s = -1) = 0.2$. Each observation equals $x^s = y + n^s$ for $s = 0, 1, \ldots, S$. In Table I, we compare the error probabilities for cases 1 and 2 with the fusion center respectively receiving or not receiving observations. It can be seen that case 2 always has smaller error probabilities than case 1 due to the additional observations at the fusion center. Such improvement gets smaller as the number of sensors increases, because the observations by the fusion center should have less effect on the performance of the system. It can also be seen that the weight β_0 of the fusion center's direct observations is bigger than the weights of sensors which are set to be one, suggesting that the fusion center count on direct observations more than quantized information.

We now study the impact of the rate constraints on the performance of system. We study a four-sensor network. Training samples are generated in the same fashion as the first experiment. We study the following five cases. For the first case, all sensors' transmission rates to the fusion center are limited to 1 bit. For the second case, one sensor's transmission rate increases to 2 bits, and all other sensors' rates are still 1 bit. For each of the remaining three cases, we allow one more sensor's transmission rate to increase to 2 bits.

In Table II, we provide the optimal weights corresponding to all sensors and the resulting detection error probability for the five cases. It is clear that the detection error decreases as more sensors are allowed to transmit at 2 bits. Furthermore, for each case, the weight parameters corresponding to the sensors that can transmit at 2 bits to the fusion center are higher than those

	β_1	β_2	β_3	β_4	Error prob.
4 sensors: 1 bit	1	0.8502	0.9495	0.9634	0.0059
3 sensors: 1 bit 1 sensor: 2 bits	1	0.9024	1.1541	2.2776	0.0050
2 sensors: 1 bit 2 sensors: 2 bits	1	1.0254	2.3541	2.3016	0.0043
1 sensors: 1 bit 3 sensors: 2 bits	1	2.4425	3.0014	2.9601	0.0035
4 sensors 2 bits	1	0.8972	0.9961	0.8875	0.0029

that can transmit only at 1 bit. This is reasonable because the fusion center counts more on the less quantized information transmitted from sensors.

VII. SUMMARY AND CONCLUSIONS

In this paper, we have proposed a weighted count kernel, and introduced these weight parameters into the risk minimization formulation for finding decision rules for the fusion center and sensors. Consequently, these decision rules can take into account the quality of sensors' observations, the quality of the fusion center's observations, and the rate constraints on sensors' transmissions. We have also exploited the properties of the optimization problem for simplifying the optimization algorithm. We have further discussed the generalization to multilevel tree-structured network based on our previous work [9]. Moreover, we have demonstrated the performance of the weighted count kernel via numerical results.

REFERENCES

[1] J. N. Tsitsiklis, "Decentralized detection," *in Advances in Signal Processing, edited by H. V. Poor and J. B. Thomas, JAI Press*, vol. 2, pp. 297–344, 1993.

[2] 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.

[3] J. F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Proc.*, vol. 51, no. 2, pp. 407–416, 2003.

[4] M. M. AI-Ibrahim and P. K. Varshney, "Nonparametric sequential detection based on multisensor data," in *Proc. 23rd Annu. Conf. Information Science Systems*, Mar. 1989, pp. 157–162.

[5] A. Nasipuri and S. Tantaratana, "Nonparametric distributed detection using wilconxin statistics," *Signal Processing*, vol. 57, no. 2, pp. 139–146, 1997.

[6] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "Distributed learning in wireless sensor networks: Application issues and the problem of distributed inference," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 56–69, Jul. 2006.

[7] ——, "Consistency in models for distributed learning under communication constraints," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 52–63, Jan. 2006.

[8] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "Nonparametric decentralized detection using kernel methods," *IEEE Trans. Signal Proc.*, vol. 53, no. 11, pp. 4053–4066, Nov. 2005.

[9] J. Hu, Y. Liang, and E. P. Xing, "Nonparametric decision making based on tree-structured information aggregation," in *Proc. Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep. 2011.

[10] B. Scholkopf and A. Smola, *Learning with Kernels*. MA: MIT press, 2002.