NEAREST-NEIGHBOR ESTIMATION IN SENSOR NETWORKS

Stefano Marano, Vincenzo Matta

DIEM, Univ. of Salerno, Fisciano (SA), Italy
{marano,vmatta}@unisa.it

Peter Willett

ECE, Univ. of Connecticut, Storrs, CT, USA
willett@engr.uconn.edu

(Invited Paper)

ABSTRACT

This contribution reviews some recent advances in the field of nearest-neighbor (NN) nonparametric estimation in sensor networks. Upon observing X_0, the problem is to estimate the corresponding response variable Y_0 by using the knowledge contained in a training set $\{(X_i, Y_i)\}_{i=1}^n$, made of n independent copies of (X_0, Y_0). In the distributed version of the problem, a network made of spatially distributed sensors and a common fusion center (FC) is considered. As X_0 is made available at the FC, it is broadcast to all the sensors. Relying upon the locally available pair (X_i, Y_i) and upon X_0, sensor i sends a message containing Y_i to the FC, or stays silent: only the few most informative response variables $\{Y_i\}$ should be sent, but no inter-sensor coordination is allowed.

The analysis is asymptotic in the limit of large network size n and we show that, by means of a suitable ordered transmission policy, only a vanishing fraction of NN messages can be selected, yet preserving the consistency of the estimation even under communication constraints.

***Index Terms*—** Nearest Neighbor, Nonparametric Regression, Ordered Transmissions, Sensor Networks.

1. INTRODUCTION

An influential paper by Predd, Kulkarni and Poor addressed the issue of distributed learning in wireless sensor networks, with specific focus on the so-called naive kernel [1]. The other major regression approach –that based on the nearest-neighbor (NN) paradigm– was not considered for difficulties of implementation without inter-sensor coordination. This paper, which provides an overview of recent advances by our group [2–5], addresses this issue.

We refer to a network consisting of n cheap remote sensors equipped by very simple hardware/software devices and powered by low-capacity batteries, each connected to a common fusion center (FC) with powerful computational and communication capabilities. As it is usually assumed, the link FC between and sensors is asymmetric: communication from the FC to sensors is essentially unconstrained, while the reverse link is subject to severe communication constraints [1, 6, 7].

Two main approaches to model such constraints have been conceived [2–5]. The former is in terms of number of channel accesses (i.e., number of messages) and it has been exploited in [2] to design a distributed single-transmission NN estimation scheme [8]. The latter –assumed here– is based on more conventional measures of communication costs, e.g., energy, bandwidth, etc.

The classical formulation of supervised nonparametric regression in the distributed setting described above is as follows. We want to estimate at the FC the value of the random variable $Y_0 \in \mathbb{R}$ (response), upon observing $X_0 \in \mathbb{R}^d$ (observation). The statistical distribution of the pair (X_0, Y_0) is completely unknown. Beforehand, a training set $\mathcal{T}_n = \{(X_i, Y_i)\}_{i=1}^n$, made of i.i.d. (independent, identically distributed) variables is disseminated through the network, and the statistical distribution of (X_i, Y_i) is exactly the same of that of (X_0, Y_0) [9]. For simplicity we assume that the i-th sensor owns just one single pair (X_i, Y_i) (thus, n represents both the network size and the cardinality of $\mathcal{T}_n$), and that the Y_i's are scalar.

When the FC collects the measurement X_0, this is broadcast to all the nodes of the network by the powerful (i.e., noiseless) link FC $\rightarrow$ sensors. The i-th node, without inter-sensor coordination, compares X_0 with the locally available X_i, and prepares a message containing Y_i, to be sent to the FC to provide information about the unknown Y_0. Whether or not this message is actually delivered to the FC depends upon a transmission policy, which will be detailed soon.

Let $r_n : \mathbb{R}^d \longrightarrow \mathbb{R}$ the estimator computed by the FC. From the orthogonality principle

$$\mathbb{E}\{[r_n(X_0) - Y_0]^2 | \mathcal{T}_n\} = \mathbb{E}\{[r_n(X_0) - r^*(X_0)]^2 | \mathcal{T}_n\} + \text{MMSE},$$

where $r^*(x_0) = \mathbb{E}\{Y_0 | X_0 = x_0\}$ is referred to as the (optimal) regression function, which is the estimator achieving the minimum mean square error MMSE $= \mathbb{E}\{[r^*(X_0) - Y_0]^2\}$. This implies that the goodness of any estimator $r_n(x_0)$ is measured in terms of the L_2 error with respect to the optimal regression function. We are interested in weakly universally consistent estimators, for which [9]

P. Willett was supported by ONR under contract N000014-13-1-0231.

$\lim_{n\to\infty} \mathbb{E}\left\{[r_n(X_0) - r^*(X_0)]^2\right\} = 0$, for all distributions of (X, Y) with $\mathbb{E}\{Y^2\} < \infty$.

Given $x_0 \in \mathbb{R}^d$, let:

$$\left(X_{(1,n)}(x_0), Y_{(1,n)}(x_0)\right), \ldots, \left(X_{(n,n)}(x_0), Y_{(n,n)}(x_0)\right),$$

be the sequence of pairs ordered according to $\|X_{(1,n)}(x_0) - x_0\| \leq \cdots \leq \|X_{(n,n)}(x_0) - x_0\|$, where $\|\cdot\|$ denotes the standard Euclidean norm in $\mathbb{R}^d$. The second sub-index in the notation $X_{(i,n)}$ emphasizes the dependence on the training set size n.

To understand how the FC can estimate Y_0 according to the NN rules, let us first assume that the FC may receive from the sensors the whole training set $\mathcal{T}_n$, which would be the case if even the reverse channel from the sensors to the FC was communication-unconstrained. In this case the FC would compute the weighted combination of the Y_i:

$$r_n^{NN}(x_0) = \sum_{i=1}^{n} W_{ni}(x_0)\, Y_i = \frac{1}{k_n} \sum_{i=1}^{k_n} Y_{(i,n)}(x_0) \tag{1}$$

with n-dependent weights

$$W_{ni}(x_0) = \begin{cases} \dfrac{1}{k_n}, & \text{if } X_i \text{ is one of the } k_n\text{-NN of } x_0, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where k_n is the number of neighbors to be accounted for. It is known that in the regime $n, k_n \to \infty$ with $k_n/n \to 0$, the k_n-NN regression rule in (1) is weakly universally consistent [9].

In the presence of communication constraints from sensors to the FC, the key point is how to select the k_n nearest neighbors without inter-sensor coordination. This problem is solved in Sect. 2, which enables us to elaborate on the distributed k_n-NN regression in Sects. 3 and 4 for continuous and quantized data, respectively.

2. UNIVERSAL CHANNEL ACCESS

The approach that makes it possible to design the distributed inference system has been pioneered by Blum and Sadler in [10], and since then has been exploited in a series of papers, e.g., [11–13]. The idea is that of *ordered transmissions*: sensors owning more informative labels have access priority with respect to those with less informative data.

Applied to our setting, the ordered transmission scheme works as follows. The i-th sensor computes the distance $\|X_i - X_0\|$ between its observation X_i and the measurement X_0. The smaller this distance, the smaller is the time instant T_i at which the i-th sensor attempts to access the channel. More precisely:

$$T_i := \tau_{\min} + \alpha_n \|X_i - X_0\|^d, \tag{3}$$

where $\tau_{\min}$ is an offset value introduced for avoiding the collapse of the first transmission epoch to zero when n grows, and the factor α_n, to be suitable designed, depends upon the network size n. To ensure synchronization, it is crucial that the time origin is the same for all the sensors. This can be guaranteed by setting to zero the sensors' clock when they receive a broadcast message from the FC (e.g., when they receive the broadcast message that conveys X_0). The assumption is made that differences in the transmission delays between different sensors and the FC can be neglected. Once that k_n messages have been received, further sensor transmissions can be avoided by means of a broadcast halting message sent by the FC. At this point the k_n "most informative" labels $Y_{(i,n)}(x_0)$, $i = 1, \ldots, k_n$ are available to the FC for the inference task, either as they are or suitably quantized and/or possibly distorted by channel impairments.

From a detailed analysis of the access protocol [2–5], two key issues arise in the regime of increasingly large n:

1. The *message crowding*: the time instants of the message deliveries might become too close to each other.
2. The total time required for making the inference task might be prohibitively large.

To elaborate briefly on these, one expects that the problem could be avoided by making α_n in (3) to grow with n sufficiently fast. On the other hand, it is clear that if α_n grows too fast, this makes the total time required for making the inference task too large. Does a compromise solution exist? The answer is not obvious, especially because we are looking for a scaling law for α_n that is valid irrespectively of the data distribution, i.e., a *universal* scaling law valid for all the distributions of interest.

To show how this compromise can be found, let $T_{(i)}$ denote the i-th order statistics of the sequence $\{T_i\}_{i=1}^n$. The transmission *spacing* Δ_i and the *minimum* spacing M_n are defined as follows:

$$\Delta_i := T_{(i+1)} - T_{(i)}, \quad i = 1, \ldots, n-1, \tag{4}$$

$$M_n := \min_{i=1,\ldots,k_n} \Delta_i. \tag{5}$$

Technical assumption: It is assumed that the pdf of the random variables $\{\|X_i - x_0\|^d\}_{i=1}^n$ is bounded in a neighborhood of x_0, for all x_0 except sets of zero probability.

THEOREM 1 *(Feasibility of ordered data delivering) In the limit for $n \to \infty$, let $k_n \to \infty$, with $k_n/n \to 0$, and let ξ_n be a sequence such that $\xi_n \to \infty$. Set:*

$$\alpha_n := n\, k_n\, \xi_n. \tag{6}$$

Then, under the technical assumption stated above, $\forall \delta > 0$ and $\tau > 0$, we have

$$\mathbb{P}\{M_n \leq \delta\} \to 0 \tag{7}$$

$$\mathbb{P}\{T_{(k_n)} > n\tau\} \to 0. \tag{8}$$

□

Ideas behind the proof. A classical problem in statistics is that of characterizing the distance between consecutive order statistics from a set of n i.i.d. random variables with known probability distribution. In particular, the *spacing theory* pioneered by [14, 15] tells us how to find a stabilizing multiplicative factor α_n such that the spacings multiplied by α_n converge in distribution. To elaborate, let us think of n points placed uniformly at random in the interval $[0,1]$: the average spacing is in the order of $1/n$, meaning that choosing $\alpha_n \sim n$ would make the spacing stable, i.e., the spacing multiplied α_n would converge in the limit $n \to \infty$. If one is interested in stabilizing the *minumum* spacing, a simple application of the union bound reveals that the appropriate factor is $\sim n^2$. In our case we deal with the minimum spacing among the first k_n ordered statistics from an ensemble of n, and we are able to prove that the appropriate multiplicative factor is $k_n\, n$: any scaling law faster than $k_n\, n$ makes the crowding probability vanish. This leads to (7), while result (8) is achieved with similar tools; the details can be found in [2]. •

Some comments are in order. Result (7) ensures that the probability that the minumum spacing between two delivered messages is less than δ be vanishingly small. Since this involves the minumum spacing, condition (7) is rather strong; weaker conditions could be conceived involving, e.g., the average spacing. Note also that including in the definition of M_n the spacing $T_{(k_n+1)} - T_{(k_n)}$ ensures that the (k_n+1)-th transmission attempt is sufficiently spaced apart from the k_n-th one. This is desirable, since it ensures that the k_n-th transmission is not impaired by the subsequent access attempts, and the FC has enough time to send its broadcast halting message. Finally, it can be shown that the scaling law for α_n is essentially the best: for any normalizing sequence α_n growing any slower than nk_n, eq. (7) does not hold true universally.

As to $T_{(k_n)}$, which is a measure of the overall channel occupation, i.e., a *global* resource employed for the inference task, it makes sense to ensure that it would not grow proportionally to network size n, or more precisely, that the per-sensor resource $T_{(k_n)}/n$ remains close to zero for large n. This is exactly what is ensured by (8).

3. UNCODED COMMUNICATION

The paradigm of *uncoded* communication has attracted increasing interest, since the pioneering work in [16], especially in connection with networks deployed for estimation/detection purposes [17, 18]. In this section, we assume that the sensor labels are sent over the common AWGN channel with coherent receiver and with random attenuation (consistency results for noncoherent channels are provided in [2]). Let $\mathcal{V}_n{:=}[-V_n, V_n]$, with some range limit V_n, and let $I_{\mathcal{V}_n}(y)$ be the indicator function of the set $\mathcal{V}_n$. The message delivered by the i-th sensor (if not inhibited by the FC) is $(Y_i/V_n)\, I_{\mathcal{V}_n}(Y_i)$, where an energy peak constraint has been imposed.

The channel access policy discussed in Sect. 2 allows us to recover the different labels at the receiver: the FC obtains (for those i for which the communication is not inhibited)

$$R_i = A_i\, \frac{Y_i}{V_n}\, I_{\mathcal{V}_n}(Y_i) + N_i, \tag{9}$$

where $\{N_i\}_{i=1,2,\dots}$ are i.i.d. zero-mean Gaussian random variables with variance σ^2, and $\{A_i\}_{i=1,2,\dots}$ are i.i.d. nonegative random variables modeling the received signal strengths. These two random sequences are assumed to be independent of all other random mechanisms present in the system, and independent of each other.

From the received messages, the FC builds the following regression function:

$$r_n(x_0) = \frac{V_n}{\mathbb{E}\{A\}} \sum_{i=1}^{n} W_{ni}(x_0) R_i. \tag{10}$$

Note that $i)$ no knowledge at all is required about the characteristics of the Gaussian noise; $ii)$ the *received* signal strengths $\{A_i\}_{i=1,2,\dots}$ are different from link to link, and no power control is necessary to compensate for them; $iii)$ the only required knowledge about the links connecting the sensors to the FC is the expectation $\mathbb{E}\{A\}$. The following theorem is proved in [2].

THEOREM 2 *(Uncoded k_n-NN over AWGN) Let $n \to \infty$. For any $k_n \to \infty$, with $k_n/n \to 0$, the regression function (10) is weakly universally consistent, provided that $V_n \to \infty$, $V_n^2/k_n \to 0$, and that $\mathbb{E}\left\{A^2\right\} < \infty$.* □

4. QUANTIZED SCHEME

4.1. Quantization

Let $\mathcal{Q}_n(y)$ be a quantized version of y to b bits. In this scheme of quantized communication, we assume that the i-th sensor, upon observing X_i, delivers the quantized version $\mathcal{Q}_n(Y_i)$ of the corresponding label Y_i. The goal is to obtain a consistent regression and, to this end, the structure of the quantizers is key; in particular it is necessary to ensure that the quantized versions of the labels are unbiased.

The random quantization rule proposed in [19, 20] provides the appropriate tool. Let us start by assuming that Y is a bounded random variable, i.e., $|Y| \le V$. Let $\mathcal{V}{:=}[-V, V]$. The range $\mathcal{V}$ of the random variable Y is divided into $2^b - 1$ equal bins of length $\delta_s{:=}2V/(2^b-1)$ whose endpoints (quantizer thresholds) are $s_i = -V + i\,\delta_s$, $i = 0, 1, \dots, 2^b - 1$. To quantize the label y, we first find the bin to which y belongs to, and then use as reconstruction level one of the two endpoints of that bin. The choice between the two endpoints is made at random according to the following rule: if $y \in [s_i, s_{i+1})$, then $\mathbb{P}\{\mathcal{Q}_n(y) = s_{i+1}\} = p = 1 - \mathbb{P}\{\mathcal{Q}_n(y) = s_i\}$, where

$$p = \frac{y - s_i}{\delta_s}. \tag{11}$$

The choice of the probability p in (11) ensures unbiasedness of the quantized label:

$$\mathbb{E}\{\mathcal{Q}_n(y)\} = p\, s_{i+1} + (1-p)s_i = p\,\delta_s + s_i = y. \tag{12}$$

In addition, the mean square quantization error can be written as

$$\begin{aligned}\mathbb{E}\{[\mathcal{Q}_n(y)-y]^2\} &= \text{VAR}[\mathcal{Q}_n(y)] = \delta_s^2 p(1-p)\\ &\leq \frac{\delta_s^2}{4} = \frac{V^2}{(2^b-1)^2}.\end{aligned} \qquad (13)$$

The extension to the case of unbounded Y can be obtained as done in [7]: Set $\mathcal{V}_n = [-V_n, V_n]$, with V_n arbitrarily chosen and depending on n. The random variable is quantized as described above if the realization $y \in \mathcal{V}_n$; conversely, if the value taken by Y falls outside the interval $\mathcal{V}_n$, then $\mathcal{Q}_n(y) = \pm V_n$, each with probability $1/2$, yielding $\mathbb{E}\left[\mathcal{Q}_n(y)\,|\,Y \notin \mathcal{V}_n\right] = 0$, and $\text{VAR}\left[\mathcal{Q}_n(y)\,|\,Y \notin \mathcal{V}_n\right] = V_n^2$. The unbiasedness of the quantizer is lost, but we obtain:

$$\mathbb{E}\{\mathcal{Q}_n(y)\} = y\, I_{\mathcal{V}_n}(y), \qquad \text{VAR}\{\mathcal{Q}_n(y)\} \leq V_n^2 \qquad (14)$$

(recall that $I_{\mathcal{V}_n}(y)$ is the indicator function of the set $\mathcal{V}_n$). These properties are sufficient to design an asymptotically consistent universal estimator, using the trick of letting $V_n \to \infty$, under the constraint $V_n^2/k_n \to 0$, as stated below.

4.2. Estimation

The sensors deliver the quantized versions of their labels so that, at the FC, the regression function is obtained by modifying eq. (1) as follows:

$$r_n(x_0) = \sum_{i=1}^{n} W_{ni}(x_0)\, \mathcal{Q}_n(Y_i). \qquad (15)$$

THEOREM 3 *(Quantized k_n-NN) Let $n \to \infty$. For any $k_n \to \infty$, with $k_n/n \to 0$, the regression function (15) is weakly universally consistent, provided that $V_n \to \infty$, and that $V_n^2/k_n \to 0$.* □

Ideas behind the proof. Let us define the expectation of $r_n(x_0)$ conditioned on $\mathcal{T}_n$ and X_0: $\bar{r}_n(x_0) := \mathbb{E}\{r_n(x_0)|\mathcal{T}_n, X_0 = x_0\}$. The residual randomness is in the quantizers' output, whose average is given by the first of eqs. (14), yielding

$$\begin{aligned}\bar{r}_n(x_0) &= \sum_{i=1}^{n} W_{ni}(x_0) Y_i\, I_{\mathcal{V}_n}(Y_i)\\ &= r_n^{NN}(x_0) - \sum_{i=1}^{n} W_{ni}(x_0) Y_i\, I_{\mathcal{V}_n^c}(Y_i),\end{aligned} \quad (16)$$

where $\mathcal{V}_n^c$ is the complement of set $\mathcal{V}_n$ with respect to $\mathbb{R}$, and $r_n^{NN}(x_0)$ is the standard NN regression function, i.e., that with unquantized data, given in (1). Thus, $\bar{r}_n(x_0)$ is the sum of the desired term $r_n^{NN}(x_0)$, and of an additional term, due to the overload region of the quantizers. The technical steps of the proof (see [2]) show that the last term can be made asymptotically negligible if $V_n \to \infty$, with $V_n^2/k_n \to 0$. •

It can be shown that the universal consistency can be retained even if the link between the sensors and the FC is noisy.

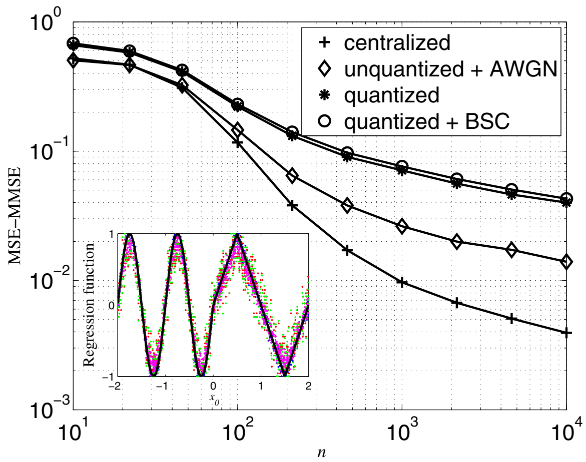

Fig. 1. In the inset box, the optimal regression function $r^*(x_0)$ (bold line) is displayed, along with samples (circles), corresponding to different Monte Carlo runs, of the regression functions of the proposed strategies. In the main plot, the L_2 errors between these regression functions and the optimal $r^*(x_0)$ is shown.

In particular, for binary quantization, it has been shown that knowledge of the crossover probability p_e of the binary symmetric channel (BSC) connecting the i-th sensor and the FC is sufficient to build at the FC a universally consistent estimator in the following form (see [2]):

$$r_n(x_0) = \frac{1}{1-2p_e} \sum_{i=1}^{n} W_{ni}(x_0)\, Z_i, \qquad (17)$$

where $Z_i := \mathcal{Q}_n(Y_i)(1-2E_i)$ is the received binary quantized label ($E_i = 1$ if a channel error occurred, $E_i = 0$ otherwise): the effect of the errors is simply accounted for by the corrective factor $(1-2p_e)^{-1}$.

5. COMPUTER EXPERIMENTS

The effect of the finiteness of the training set size n is now investigated by means of standard Monte Carlo counting procedure. Additional details and examples can be found in [2]. Let us refer to the following observation model:

$$X \text{ uniform in } [-2, 2], \qquad (18)$$

$$Y = r^*(X) + \frac{E}{\sqrt{\text{VAR}\{E\}}}\sqrt{\text{MMSE}}, \qquad (19)$$

where the optimal regression function $r^*(x_0)$ is the concatenation of a sinusoidal and a triangular wave, see Fig. 1, while the pdf $f_E(x)$ of the error term E is chosen as follows:

$$f_E(x) = \frac{e^{-x^2/(2a_1^2)}}{2\sqrt{2\pi a_1^2}} + \frac{a_3}{4a_2^{a_3}}|x|^{a_3-1} e^{-(|x|/a_2)^{a_3}}, \qquad (20)$$

that is, a balanced mixture of a Gaussian and a symmetrized Weibull distribution. We assume: $a_1 = 0.1, a_2 = 1, a_3 = 0.9$, and MMSE in (19) is $1/4$.

The theory developed in Sects. 3 and 4 ensures asymptotic consistency of the estimator. Nothing is claimed about the speed of convergence of the estimator towards the optimum $r_n^*(x_0)$, and therefore little can be said about the MSE of the universal estimators for finite values of n. Figure 1, based on 10^4 Monte Carlo runs and in which $k_n = 3\,n^{1/3}$ and $V_n = k_n^{1/10}$, provides this. The curve labeled as "centralized" refers to the case of unconstrained reverse link, i.e. to (1). For the quantized transmission we also account for the presence of channel errors, i.e., the links connecting sensors and FC is a BSC with crossover probability $p_e \approx 0.02$.

We see that in this example, for moderate crossover probability, the presence of channel errors has little impact on the estimation performance: the two curves showing the performance of the quantized systems are quite close to each other. We also see that the unquantized communication over the AWGN channel with random attenuation provides improved performance with respect to the binary quantized system over BSCs. Here we pay attention to making the comparison meaningful by using compatible parameters for the two scenarios, i.e., the crossover probability of the BSC correspond to the probability of error of a coherent AWGN channel with Rayleigh distributed attenuations.

More important, we see that the curves shown in Fig. 1, seem to decay as $n^{-\nu}$, with one and the same $\nu \in (0, 1)$. This is remarkable, read in the light of known results about the rates of convergence of k_n-NN estimates, see [9].

6. SUMMARY

Nearest-neighbor distributed regression in a nonparametric framework is studied. The key enabler for distributed implementation of the estimation scheme is the ordered transmission policy, in which more informative sensors have delivering priority. Taking also into account a noisy channel, two estimators are investigated: one in which data are delivered without qantization, and the other that assumes quantized messages. For both cases asymptotic consistency is proved, thus enriching the class of asymptotically consistent distributed nonparametric estimators with the important case of the nearest-neighbor regression.

REFERENCES

[1] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 56–69, Jul. 2006.

[2] S. Marano, V. Matta, and P. Willett, "Nearest-neighbor distributed learning by ordered transmissions," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5217–5230, Nov. 2013.

[3] ——, "How many bits from how many sensors? A trade-off in distributed nearest-neighbor learning," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, Florence, Italy, May 4-9 2014.

[4] ——, "Decentralized nearest-neighbor learning over noisy channels: the uncoded way," in *Proc. of the 16th International Conference on Information Fusion (FUSION 2013)*, Istanbul, Turkey, July, 9-12 2013, pp. 426–431.

[5] ——, "Nearest-neighbor distributed learning under communication constraints," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, Vancouver, Canada, May 26-31 2013, pp. 3278–3282.

[6] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "A collaborative training algorithm for distributed learning," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1856–1871, Apr. 2009.

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

[8] T. M. Cover, "Estimation by the nearest neighbor rule," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 50–55, Jan. 1968.

[9] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*. New York: Springer, 2002.

[10] R. S. Blum and B. M. Sadler, "Energy efficient signal detection in sensor networks using ordered transmissions," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3229–3235, Jul. 2008.

[11] R. S. Blum, "Ordering for estimation and optimization in energy efficient sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2847–2856, Jun. 2011.

[12] P. Braca, S. Marano, and V. Matta, "Single-transmission distributed detection via order statistics," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 2042–2048, Apr. 2012.

[13] ——, "Asymptotic consistent one-bit detection in large sensor networks," in *Proc. of 19th European Signal Processing Conference (EUSIPCO 2011)*, Barcelona, Spain, Aug. 29- Sept. 2 2011, pp. 1035–1039.

[14] P. Lévy, "Sur la division d'un segment par des points choisis au hasard," *Comptes Rendus Acad. Sci. Paris*, vol. 208, pp. 147–149, 1939.

[15] R. Pyke, "Spacings," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 27, no. 3, pp. 395–449, 1965.

[16] M. Gastpar, M. Vetterli, and P. L. Dragotti, "Sensing reality and communicating bits: a dangerous liaison," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 70–83, Jul. 2006.

[17] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 613–626, Feb. 2006.

[18] S. Marano, V. Matta, L. Tong, and P. Willett, "A likelihood based multiple access for estimation in sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5155–5166, Nov. 2007.

[19] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, Jun. 2005.

[20] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 413–422, Feb. 2006.