DISTRIBUTED LEARNING WITH KERNELS IN WIRELESS SENSOR NETWORKS FOR PHYSICAL PHENOMENA MODELING AND TRACKING

Cédric Richard [(1)], *Paul Honeine* [(2)], *Hichem Snoussi* [(2)], *André Ferrari* [(1)], *Céline Theys* [(1)]

[(1)] Laboratoire Fizeau (UMR CNRS 6525, OCA), Université de Nice Sophia-Antipolis, 06108 Nice, France

[(1)] Institut Charles Delaunay (FRE CNRS 2848), Université de technologie de Troyes,10010 Troyes, France

Wireless ad-hoc sensor networks have emerged as an interesting and important research area in the last few years. They rely on sensor devices deployed in an environment to support sensing and monitoring, including temperature, humidity, motion, acoustic, etc. Low cost and miniaturization of sensors involve limited computational resources, power and communication capacities. Consequently, wireless ad-hoc sensor networks require collaborative execution of a distributed task on a large set of sensors, with reduced communication and computation burden.

In this paper, we consider the problem of modeling physical phenomena such as temperature field distributions, and track their evolution. Many approaches have been proposed in the signal processing literature to address this issue with collaborative sensor networks. In the framework of distributed inference with parametric models, focus was clearly put on determining how the channel capacity limits the quality of estimates. See, e.g., [1] and references inside. Model-based techniques were also exploited to compress the data based on their temporal and spatial redundancy [2]. Since they do not depend on some arbitrary modeling assumptions, applications of wireless sensor networks provide a strong motivation for the use of nonparametric methods for decentralized inference. See [3] for a survey. Model-independent methods based on kernel machines have recently been investigated. In particular, a distributed learning strategy has been successfully applied to regression in wireless sensor networks [5]. In this paper, we take advantage of this framework to derive a new approach.

Kernel-based methods have gained wide popularity over the last decade. Initially derived for regression and classification with *support vector machines*, they include classical techniques such as least-squares methods and extend them to nonlinear functional approximation. We wish to determine a function $\psi^o(\boldsymbol{x})$ defined on the area of interest $\mathcal{X}$ that best models a distribution such as a temperature field. The latter is learned from the information coupling sensor locations and measurements. The information from the N sensors located at $\boldsymbol{x}_n \in \mathcal{X}$ which provide measurements $\boldsymbol{d}_n \in \mathbb{R}$, with $i = n, \ldots, N$, is combined in the set of pairs $\{(\boldsymbol{x}_1, \boldsymbol{d}_1), \ldots, (\boldsymbol{x}_N, \boldsymbol{d}_N)\}$. Consider a reproducing kernel $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Let us denote by $\mathcal{H}$ its reproducing kernel Hilbert space (RKHS) with inner product $\langle \cdot\,, \cdot \rangle_{\mathcal{H}}$. This means that every function ψ of $\mathcal{H}$ can be evaluated at any $\boldsymbol{x} \in \mathcal{X}$ by $\psi(\boldsymbol{x}) = \langle \psi, \kappa_{\boldsymbol{x}} \rangle_{\mathcal{H}}$, where $\kappa_{\boldsymbol{x}}$ denotes $\kappa(\cdot, \boldsymbol{x})$. This allows us to write $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \kappa_{\boldsymbol{x}_i}, \kappa_{\boldsymbol{x}_j} \rangle_{\mathcal{H}}$, which defines the reproducing property. One of the most widely used kernel is the Gaussian kernel $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \mathrm{e}^{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 / 2\sigma^2}$ with σ the kernel

bandwidth. We seek the optimal ψ^o that minimizes the following global cost function

$$J(\psi) = \sum_{n=1}^{N} E|\psi(\boldsymbol{x}_n) - \boldsymbol{d}_n|^2,$$

that is, $J(\psi) = \sum_{n=1}^{N} E|\langle \psi, \kappa_{\boldsymbol{x}_n}\rangle_{\mathcal{H}} - \boldsymbol{d}_n|^2$ using the reproducing property. The cost function $J(\psi)$ can be rewritten as follows, for any node $k \in \{1, \ldots, N\}$,

$$J(\psi) = J_k(\psi) + \sum_{\substack{n=1 \\ n\neq k}}^{N} J_n(\psi)$$

with $J_k(\psi) = \sum_{n\in\mathcal{N}_k} c_{n,k} E|\langle \psi, \kappa_{\boldsymbol{x}_n}\rangle_{\mathcal{H}} - \boldsymbol{d}_n|^2$ the local cost function, and $\mathcal{N}_k$ the set of neighbors of sensor k. Here, $c_{n,k}$ is the (n,k)-th component of C, which satisfies the following constraints: $c_{n,k} = 0$ if $n \notin \mathcal{N}_k$, $C\mathbb{1} = \mathbb{1}$ and $\mathbb{1}^\top C = \mathbb{1}^\top$. Let us write $\psi_n^o = \arg\min_{\psi\in\mathcal{H}} J_n(\psi)$. Finally, it can be shown that it is equivalent to minimize $J(\psi)$ or the following cost function

$$J_k^\ell(\psi) = \sum_{n\in\mathcal{N}_k} c_{n,k} E|\langle \psi, \kappa_{\boldsymbol{x}_n}\rangle_{\mathcal{H}} - \boldsymbol{d}_n|^2 + \sum_{\substack{n=1 \\ n\neq k}}^{N} \|\psi - \psi_n^o\|_{\mathcal{H}_n}^2$$

Minimizing $J_k^\ell(\psi)$ requires the nodes to have access to global information ψ_n^o. To facilitate distributed implementations, consider the relaxed local cost function

$$J_k^r(\psi) = \sum_{n\in\mathcal{N}_k} c_{n,k} E|\langle \psi, \kappa_{\boldsymbol{x}_n}\rangle_{\mathcal{H}} - \boldsymbol{d}_n|^2 + \sum_{n\in\mathcal{N}_k/\{k\}} b_{n,k}\|\psi - \psi_n^o\|_{\mathcal{H}}^2$$

where ψ_n^o is now the best estimate available at node n. The iterative steepest-descent approach for minimizing $J_k^r(\psi)$ can be expressed in the following form

$$\psi_{k,i} = \psi_{k,i-1} - \frac{\mu}{2}\nabla J_k^r(\psi_{k,i-1})$$

with

$$\nabla J_k^r(\psi_{k,i-1}) = \sum_{n\in\mathcal{N}_k} 2\, c_{n,k} E\left(\psi_{k,i-1}(\boldsymbol{x}_n) - \boldsymbol{d}_n\right)\kappa_{\boldsymbol{x}_n} + \sum_{n\in\mathcal{N}_k/\{k\}} 2\, b_{n,k}(\psi_{k,i-1} - \psi_n^o)$$

Incremental algorithms are useful for minimizing sums of convex functions. They consist of iterating sequentially over each sub-gradient, in some predefined order. This leads to the following possible distributed learning strategies.

1. Adapt-then-Combine kernel LMS

For each time instant i and each node k, repeat

$\phi_{k,i} = \psi_{k,i-1} - \mu_k \sum_{n\in\mathcal{N}_k} c_{n,k}\left(\psi_{k,i-1}(x_n) - d_{n,i}\right)\kappa_{x_n}$

$\psi_{k,i} = \sum_{n\in\mathcal{N}_k} b_{n,k}\,\phi_{k,i}$

2. Combine-then-Adapt kernel LMS

For each time instant i and each node k, repeat

$\phi_{k,i-1} = \sum_{n\in\mathcal{N}_k} b_{n,k}\,\psi_{k,i-1}$

$\psi_{k,i} = \phi_{k,i-1} - \mu_k \sum_{n\in\mathcal{N}_k} c_{n,k}\left(\phi_{k,i-1}(x_n) - d_{n,i}\right)\kappa_{x_n}$

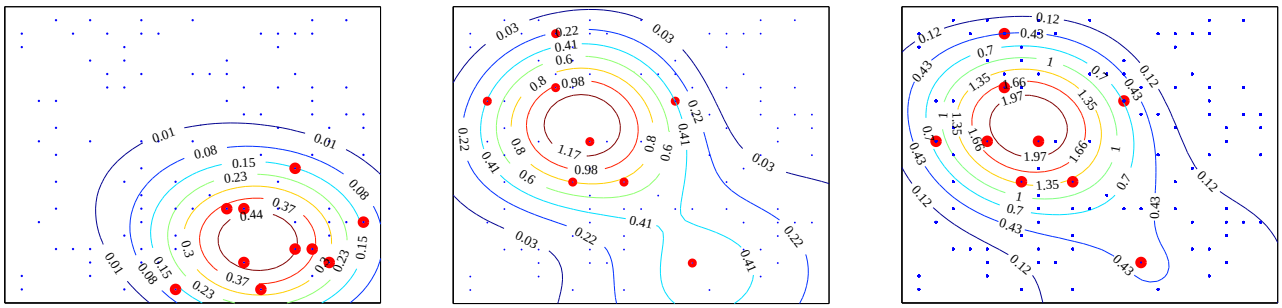

Fig. 1. Snapshots of the evolution of the estimated temperature at $t = 100$ (left), $t = 150$ (center) and $t = 200$ (right). Sensors are shown with small blue dots. Big red dots represent the sensors of interest according to a criterion that will be described in the camera-ready paper.

Both strategies consist of adapting local regressors based on measurements, and combining them if they are in the neighborhood of each other. Measurements and regressors are not exchanged between the nodes if $c_{n,k} = \delta_{n,k}$. In what follows, each $c_{n,k}$ and $b_{n,k}$ were set to $|\mathcal{N}_k|$. To illustrate the relevance of our technique, we considered a classical application of estimating a temperature field governed by the partial differential equation

$$\frac{\partial \Theta(\boldsymbol{x}, t)}{\partial t} - c\nabla^2_{\boldsymbol{x}}\Theta(\boldsymbol{x}, t) = Q(\boldsymbol{x}, t).$$

Here $\Theta(\boldsymbol{x}, t)$ denotes the temperature as a function of space and time, c is a medium-specific parameter, $\nabla^2_{\boldsymbol{x}}$ is the Laplace spatial operator, and $Q(\boldsymbol{x}, t)$ is the heat added. We studied the question of monitoring the evolution of the temperature field distribution in a square region with open boundaries and conductivity $c = 0.1$, using $N = 100$ sensors deployed randomly on a grid. Given some measurements $d_{n,i} = \Theta(x_n, t_i) + z_{n,i}$, with $z_{n,i}$ a i.i.d. Gaussian noise, the problem was to estimate the temperature $\Theta(x_n, t_i)$ via $\psi_i(x_n)$ based on the Gaussian kernel.

Two heat sources of intensity 200 W were placed within the region, the first one was activated from $t = 1$ to $t = 100$, and the second one from $t = 100$ to $t = 200$. The experimental setup will be described in the camera-ready paper. Fig. 1 illustrates the estimated temperature field distribution at different times by the Adapt-then-Combine strategy, for several values of $|\mathcal{N}_k|$. The convergence of the proposed algorithm is illustrated in Fig. 2 where we show the evolution over time of the normalized mean-square prediction error. The abrupt change in heat sources at $t = 100$ is clearly visible, and highlights the convergence behavior of the proposed algorithm. Comparison with an iterative distributed gradient algorithm clearly shows a faster convergence speed.

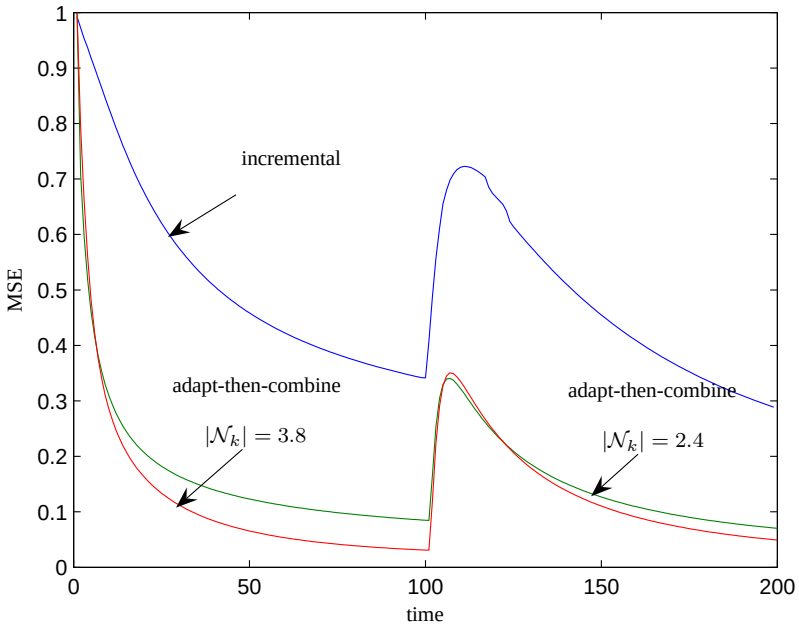

Fig. 2. Learning curve obtained from $t = 1$ to $t = 200$. Time $t = 100$ corresponds to a system modification.

1. REFERENCES

[1] A. Swami, Q. Zhao, and Y.-W. Hong, *Wireless Sensor Networks: Signal Processing and Communications*. Wiley, 2007.

[2] C. Guestrin, P. Bodi, R. Thibau, M. Paski, and S. Madde, "Distributed regression: an efficient framework for modeling sensor network data," in *Proc. third international symposium on information processing in sensor networks (IPSN)*. New York, NY, USA: ACM, 2004, pp. 1–10.

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

[4] ——, "Distributed regression in sensor networks: Training distributively with alternating projections," *CoRR*, vol. abs/cs/0507039, 2005, informal publication.

[5] ——, "Distributed kernel regression: An algorithm for training collaboratively," in *IEEE Proc. Information Theory Workshop*, 2006.