Towards Distributed Bayesian Estimation
A Short Note on Selected Aspects

K. Dedecius
Department of Adaptive Systems
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, Prague, 182 08, Czech Republic
dedecius@utia.cas.cz

V. Sečkárová
Department of Adaptive Systems
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, Prague, 182 08, Czech Republic
seckarov@utia.cas.cz

Abstract

The rapid development of ad-hoc wireless networks, sensor networks and similar calls for efficient estimation of common parameters of a linear or nonlinear model used to describe the operating environment. Therefore, the theory of collaborative distributed estimation has attained a very considerable focus in the past decade, however, mostly in the classical deterministic realm. We conjecture, that the consistent and versatile Bayesian decision making framework, whose applications range from the basic probability counting up to the nonlinear estimation theory, can significantly contribute to the distributed estimation theory.

The limited extent of the paper allows to address the considered problem only very superficially and shortly. Therefore, we are forced to leave the rigorous approach in favor of a short survey indicating the arising possibilities appealing to the non-Bayesian literature. First, we introduce the problem in a general Bayesian decision making domain and then narrow the scope to the estimation problem. In the ensuing parts, two mainstream approaches to common-objective distributed estimation are presented and the constraints imposed by the environment are studied.

1 Introduction

From the immense distributed decision making framework, we consider only the fully distributed decision problem (specifically parameter estimation) in which different agents (network nodes) obtain rather slightly different measurements from the environment. These measurements ideally obey the same distribution and differ only with respect to a realization of a noise variable. In this regard, the prescriptive methodology for designing agents proof of systematic error or bias is necessary. Furthermore, we assume that the agents have the same objective function. Our (practically unreachable) goal is to achieve the *general consensus on the decision*. In other words, if the decision is the evaluation of the posterior probability of some event, the goal is to achieve the state when all agents agree on it [22]. The solution has been proposed in [3]. Its time-dependent reformulation follows [22]:

Each time instant, the agents first communicate their own distributions among themselves and then update own distribution by reflecting the obtained distributions from the others. Following this procedure, the consensus is achieved as its limit case.

Since we deal with distributed estimation problem, it should be emphasized that our distributed decision problem differs from the team decision making [18],[19] and others. In our case, the *common consensus* on the estimate is the primary goal.

The theory of distributed estimation of an unknown common variable of interest has attained the prevailing focus in the last decade. The main cause was the increasing spatial complexity of large-scale ad-hoc wireless and sensor networks consisting of heterogeneous devices. Such an environment, more or less limited with respect to the energy, communication and processing resources, calls for efficient computational paradigms. The main tasks of interest in these networks, closely related to estimation, comprise in-network routing, signal processing, management, load balancing, sensor management, change point estimation etc.

From the Bayesian viewpoint, the theory of parameter estimation belongs to the decision theory. Suppose, that the considered network consists of $n \in \mathbb{N}$ nodes whose respective scalar or multidimensional measurements $y_1, \ldots, y_n$ are related to some unobservable quantity Θ, called parameter, some scalar or multidimensional input variables $u_1, \ldots, u_n$ and the task consists in estimation of Θ. This basically means, that the nodes seek the probability distribution of Θ given measurements and inputs, mostly in the form of a probability density function

$$f_\Theta\left(\Theta | y_0, y_1, \ldots, y_n, u_0, u_1, \ldots, u_n\right), \quad n \in \mathbb{N},$$

where y_0 and u_0 form the prior information, e.g. obtained from an expert, from past measurements or a noninformative prior is used.

The distributed systems can actively benefit from the higher number of participating sensors in the network and, potentially, from their technical heterogeneity, allowing to measure and compute with different performance according to the actual state of the observed reality.[1]

Some examples of distributed estimation problems comprise:

- collective estimation of a physical variable. This case is very important in sensor networks. Furthermore, it becomes popular in large physical experiments as well.
- fault tolerant systems with the voting circuit, in which several units collectively decide about a failure of a redundant device. Currently, the fuzzy approach dominates these solutions;
- classification networks, in which several nodes estimate the parameters of classifiers (represented, e.g., by beta-binomial or Dirichlet-multinomial models). This case is very important in bandwidth-limited networks;
- and many others.

From the communication and evaluation strategy, the estimation task in the distributed systems can be run in two different basic concepts, obviously influencing the network topology:

- The centralized approach in which the network nodes send their data to a dedicated unit responsible for computations;
- The decentralized approach in which all the network nodes posses and actively exploit own computing ability.

We will further describe them below.

[1] A particularly interesting case is the existence of the need of very precise measurements under several different conditions, preventing the user from using single measuring device.

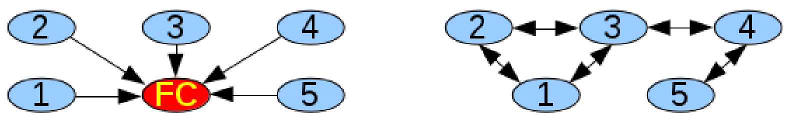

Figure 1: Principles of centralized and decentralized schemes. The centralized one (left) embodies a fusion center (FC) responsible for computation of estimates; in the decentralized concept (right) the computation task lies on the network nodes disseminating the available information, possibly partial. Remind, that the network topology of the decentralized scheme may differ from case to case.

2 Centralized and decentralized approaches

In this section, we describe the centralized and decentralized approaches to distributed parameter estimation, highlight the principal differences and mention several existing concepts. The purpose of this section is twofold: to induce contemplation on the pros and cons of the respective approaches and (maybe more importantly) to let the Bayesians take inspiration from the "disregarded deterministic world". Anyway, the good aspect is that in the Bayesian framework, the estimation mostly abstracts from the centralization or decentralization of information processing. In both cases, the basic methods can be the same. Some differences will arise with respect to the convergence properties of the estimators. Again, we stress that we focus only on the case of *distributed estimation of common parameter*.

3 Centralized approach

The centralized approach with a fusion center processing the measurements from the nodes in the network and potentially propagating the results back to them has appeared with the first occurrence of the distributed networks. This popular approach is widely used, e.g., in industrial applications, Internet services etc. However, it significantly suffers from high communication resources and high-availability (HA) demands to be able to transmit the data between the fusion center and the network nodes. The plethora of data is likely to saturate the (short term) memory and can lead to high system load. To some degree, these problems can be solved by data aggregation and quantization, e.g. [22] and references below.

On top of this, the fusion center represents a potential single point of failure (SPoF) requiring a special treatment, e.g., redundant hot and cold spare devices, data replications, graceful degradation systems etc., [28], which in turn increases the complexity of the whole system.

4 Decentralized approach

In the decentralized approach, the estimation is run directly in the nodes which share their information with other nodes. The decentralized approach can be further divided to incremental and diffusion approaches [4]. The former one is similar to the token-ring network topology, in which a closed cyclic path is performed, i.e., each node communicates with its neighbours within this path. In this setting, a failure of a link between two adjacent nodes can prevent the network from operation, hence the problem of SPoFs can be even worse. On the other hand, the diffusion approaches solve this issue by letting the nodes communicate within their closed neighbourhood, i.e., with their adjacent neighbours. In this case, the failure of a single link or node does not cause failure of the whole network, hence the estimation is preserved. It was shown, that the information propagating through the network leads in the limit case to the consensus [5], [4]. The diffusion approach possesses the adaptivity property that is very important for ad-hoc network, in which the topology of the network can dynamically change with time.

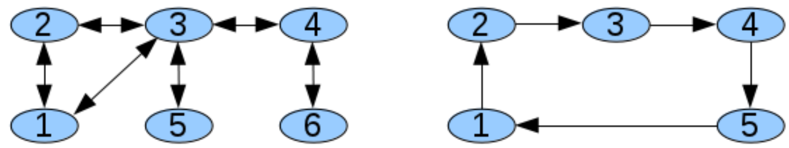

Figure 2: Two basic decentralized approaches. Left: diffusion approach, in which, e.g. node 3 communicates with nodes 1,2,4 and 5, while node 6 communicates only with node 4. Right: incremental approach (token-ring topology), the information sequentially circulates via all the network nodes which incorporate their contributions to it.

5 Communication constraints

The communication constraints still represent the most frequent restriction in distributed systems. In the centralized approach, the fusion center is often connected to other nodes via a backbone network, which enforces high bandwidth requirement. However, if the nodes are connected directly to the center via dedicated lines, the center must be able to effectively handle the incoming data traffic using a sort of an efficient switching protocol. This issue is far beyond the scope of the paper, we only reveal the need for as little traffic as possible.

In the decentralized estimation, the situation is much easier, since although the load of individual point-to-point communication lines remains the same, the high communication load typical for the fusion center is avoided. By using efficient network topology, the load can be decreased to a very low level, respecting the constraints of individual nodes.

Several possible communication strategies comprise:

- communication of all data, i.e., y_i, u_i and estimates $\hat{\Theta}$, possibly the parameters of distribution of the latter;;
- communication of sufficient statistics of respective distributions or, if applicable, non-sufficient statistics whose ancillary complements are naturally known to the network nodes;
- down to 1 bit communication strategy.

The communication of all data is a trivial case. The data quantization, i.e., the compression of data to be transmitted among the nodes in the network is an interesting option. Such a problem has been treated, for instance, in [8] and [1] for single node. For a decentralized distributed network, [13] restrict the local nodes to be data quantizers and develop the optimal design minimizing the estimation error. These results were further used by [2], [6], [9], [14], [16], [20], [27] and many others, dealing mostly with few or even one-bit messages. In this light, the assumption of possibility to communicate, for instance, sufficient statistics, so fundamental in the Bayesian framework, can simply fail. Therefore, it would be necessary to find out a way to fulfill the potential communication constraints.

6 Information fusion strategies

There exist several possible strategies for fusion of information obtained from nodes, irrespectively of the network scheme, topology or constraints. The Bayesian paradigm imposes constraints on entropy transform between the prior and posterior distribution, measured mostly by the information entropy (in terms of its maximization) or the Kullback-Leibler divergence and the cross-entropy (minimization). There are two main representation of the information from several information sources (i.e., network nodes), namely mixture, i.e. a convex combination of probability density functions [7], [10], [17], or a weighted likelihood [23], [24] and [25]. The latter has been proved suitable for distributed dynamic estimation in [21]. The mixture-based information treatment is a traditional and well-established way. The fusion of incompletely compatible probabilistic representations of information still represents a challenge. First steps towards this, exploiting the minimum cross-entropy principle, can be found, e.g. in [11].

7 Concluding remarks

We have outlined the possible future research trends towards the Bayesian distributed estimation of common parameter of interest using similar decision makers (estimators) with the same model. Unlike the traditional single problem oriented solutions employing mostly the non-stochastic methods, the Bayesian reasoning leads rather to a *methodology*, abstracting from a particular problem view. Being applicable to a large class of problems comprising, among others, dynamic estimation of least-squares problems, classification problems and many others, the distributed Bayesian framework only expects a suitable formulation of the problem. The application of the basic prescribed methods arising from the methodology can be done usually in a straightforward way.

The future research directions in the field of Bayesian distributed estimation will almost certainly deal with networks considerably constrained from specific aspects. This issue has been only recently dealt with from the non-Bayesian viewpoint, e.g., the energy-constrained networks in [12],[15],[26], the bandwidth constrained networks mentioned above etc.

Another very special issue is the design of intelligent nodes, able to agree on the form of information to be disseminated. However, this case is rather a part of the multi-agent systems (MAS) theory and definitely does not belong to the estimation theory.

Acknowledgement

This research has been supported by the Grant Agency of the Czech Republic, project number 102/08/0567.

References

[1] E. Ayanoglu. On optimal quantization of noisy sources. *Information Theory, IEEE Transactions on*, 36(6):1450–1452, November 1990.

[2] T. C. Aysal and K. E. Barner. Constrained decentralized estimation over noisy channels for sensor networks. *Signal Processing, IEEE Transactions on*, 56(4):1398–1410, April 2008.

[3] V. Borkar and P. Varaiya. Asymptotic agreement in distributed estimation. *Automatic Control, IEEE Transactions on*, 27(3):650–655, June 1982.

[4] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed. Diffusion recursive Least-Squares for distributed estimation over adaptive networks. *IEEE Transactions on Signal Processing*, 56(5):1865–1877, May 2008.

[5] F. S. Cattivelli and A. H. Sayed. Diffusion LMS strategies for distributed estimation. *IEEE Transactions on Signal Processing*, 58(3):1035–1048, March 2010.

[6] Hao Chen and P. K. Varshney. Nonparametric One-Bit quantizers for distributed estimation. *Signal Processing, IEEE Transactions on*, 58(7):3777–3787, July 2010.

[7] Morris H. Degroot. *Optimal Statistical Decisions (Wiley Classics Library)*. Wiley-Interscience, April 2004.

[8] Y. Ephraim and R. M. Gray. A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization. *Information Theory, IEEE Transactions on*, 34(4):826–834, July 1988.

[9] Jun Fang and Hongbin Li. Distributed adaptive quantization for wireless sensor networks: From delta modulation to maximum likelihood. *Signal Processing, IEEE Transactions on*, 56(10):5246–5257, October 2008.

[10] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař. *Optimized Bayesian Dynamic Advising: Theory and Algorithms*. Springer, London, 2006.

[11] M. Kárný and T.V. Guy. Sharing of knowledge and preferences among imperfect Bayesian participants. In *Proc. of the NIPS Workshop 'Decision Making with Multiple Imperfect Decision Makers'*, 2010.

[12] Alexey Krasnopeev, Jin J. Xiao, and Zhi Q. Luo. Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises. *EURASIP J. Wirel. Commun. Netw.*, 2005:473–482, September 2005.

[13] Wai-Man Lam and A. R. Reibman. Design of quantizers for decentralized estimation systems. *Communications, IEEE Transactions on*, 41(11):1602–1605, November 1993.

[14] Junlin Li and G. AlRegib. Rate-Constrained distributed estimation in wireless sensor networks. *Signal Processing, IEEE Transactions on*, 55(5):1634–1643, May 2007.

[15] Junlin Li and G. AlRegib. Distributed estimation in Energy-Constrained wireless sensor networks. *Signal Processing, IEEE Transactions on*, 57(10):3746–3758, October 2009.

[16] Zhi-Quan Luo. Universal decentralized estimation in a bandwidth constrained sensor network. *Information Theory, IEEE Transactions on*, 51(6):2210–2219, June 2005.

[17] G. J. McLachlan and D. Peel. *Finite mixture models*. Wiley series in probability and statistics: Applied probability and statistics. Wiley, 2000.

[18] Michael and Bacharach. Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics*, 53(2):117 – 147, 1999.

[19] R. Radner. Costly and bounded rationality in individual and team decision-making. *Industrial and Corporate Change*, 9(4):623, 2000.

[20] A. Ribeiro and G. B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks-part II: unknown probability density function. *Signal Processing, IEEE Transactions on*, 54(7):2784–2796, July 2006.

[21] V. Sečkárová and K. Dedecius. Distributed Bayesian Diffused Estimation. In *IFAC SYSID 2012, submitted.* IFAC, 2012.

[22] J. N. Tsitsiklis and M. Athans. Convergence and asymptotic agreement in distributed decision problems. 21:692–701, December 1982.

[23] X. Wang. Asymptotic properties of maximum weighted likelihood estimators. *Journal of Statistical Planning and Inference*, 119(1):37–54, January 2004.

[24] X. Wang. Approximating Bayesian inference by weighted likelihood. *Can J Statistics*, 34(2):279–298, 2006.

[25] Xiaogang Wang and James Zidek. Derivation of mixture distributions and weighted likelihood function as minimizers of KL-divergence subject to constraints. *Annals of the Institute of Statistical Mathematics*, 57:687–701, 2005.

[26] Jwo-Yuh Wu, Qian-Zhi Huang, and Ta-Sung Lee. Minimal energy decentralized estimation via exploiting the statistical knowledge of sensor noise variance. *Signal Processing, IEEE Transactions on*, 56(5):2171–2176, May 2008.

[27] J. J. Xiao and Z. Q. Luo. Decentralized estimation in an inhomogeneous sensing environment. *Information Theory, IEEE Transactions on*, 51(10):3564–3575, October 2005.

[28] B.Y. Zhao, J. Kubiatowicz, and A.D. Joseph. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. *Computer*, 74(11-20):46, 2001.