Distributed Estimation of Mixture Models

Kamil Dedecius[1], Jan Reichl[1,2]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic. Prague, Czech Republic
dedecius@utia.cas.cz

[2] Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague. Prague, Czech Republic
reichja3@fjfi.cvut.cz

Abstract

The contribution deals with the dynamic decentralized distributed estimation of mixture models, exploiting the consistent Bayesian paradigm. A stochastic multimodal process of interest is observed and modelled by a network (i.e., directed or undirected graph) of nodes (vertices); the edges define the information-sharing possibilities. The additive noise, corrupting the process outcomes, is vertices-specific. Exploiting the shared knowledge of observations and/or posterior distributions, the aim is to arrive at the best estimates of the employed mixture model. Our contribution focuses on components' parameters and weights.

Keywords: mixture estimation, dynamic models, distributed estimation.

1 Introduction

The problem of the dynamic distributed estimation of various model types has attained a considerable attention in the last decade, particularly due to the rapid development of wireless ad-hoc networks, sensor networks and the emergence of the so-called big data phenomenon, e.g. [3] and many others. The underlying models comprise mostly the least-squares ones, for instance, the recursive least squares (RLS) [1], least mean squares (LMS) [7, 11] and Kalman filters [2, 8, 10]. The mixture models are covered, e.g., in [9] (EM algorithm) and [6] (Gaussian mixture Bernoulli filter).

Despite the great potential of the Bayesian paradigm in this field, its adoption is rather an exception than a rule. From the probabilistic viewpoint, the resulting "classical" (that is, non-Bayesian) algorithms often suffer statistical inconsistencies. For instance, the point estimators are often fused without reflecting their statistical properties, which may lead to statistically-absurd situations. The first author's work [4] aims to partially fill this gap. It proposes a fully Bayesian approach to decentralized distributed estimation with a fusion based on minimization of the Kullback-Leibler divergence. The present contribution extends the results to the case of mixture models.

2 Distributed Estimation

Let us assume a spatially distributed network, represented by an directed or undirected graph of N vertices (nodes, denoted i), where the edges define which vertices may share information. The vertices observe noisy outcomes $y_{i,k}$ of a single stochastic process; $k = 1, 2, \ldots$ stand for time indices. The observations $y_{i,k}$, possibly corrupted by some additive vertices-specific noise, generally obey a mixture distribution with K components, each having latent parameters $\theta_{i,k}, k = 1, \ldots, K$. That is, each node i exploits a model of the form

$$p_i(y_{i,k}|\phi_i, \theta_i) = \sum_{k=1}^{K} \phi_{i,k} p_{i,k}(y_{i,k}|\theta_{i,k}), \qquad k = 1, 2, \ldots, \quad i = 1, \ldots, N,$$

where p denote the components – the conditional probability density functions, and $\phi_{i,k}$ are weights (probabilities) of these components, taking values on a unit K-simplex. The goal is to collectively estimate the weights and the components' parameters from the dynamically incorporated observations $y_{i,k}$, exploiting the theory of Bayesian mixture estimation of Titterington [12] and Frühwirth–Schnatter [5]. The shared information relates either to the observations, the posterior distributions, or both. Furthermore, the imposed communication restrictions prohibits iterations among nodes: any information, once obtained, is directly incorporated.

From a vertice's viewpoint, a set of models and/or posterior distributions is available from the other (reachable) network nodes. The elements of this set have to be merged together, preferably resulting in a single model and/or posterior distribution, closest to the set elements in the statistical sense. That is, we search for such a probability distribution function p_i^*, whose Kullback-Leibler divergence D from all others is minimal,

$$\alpha_{ij} D(p_i^*||p_j) \rightarrow \min, \qquad \alpha_{ij} \in [0, 1] \quad \text{s.t.} \quad \sum_j \alpha_{ij} = 1.$$

Here, α_{ij} are weights assigned by the vertice i to its peers j, reflecting the degree of belief in their information.

This type of approximation can be used both for the models and the posterior distributions. Naturally, there is also the other order of the divergence arguments. It can be shown, that (i) the results coincide provided the (ideal but rather unexpected) statistical homogeneity of the vertices, (ii) the chosen order has a consistent interpretation of the Wang and Zidek's weighted likelihood [13] and (iii) is has appealing consequences for the exponential family distributions and their conjugate prior distributions.

Acknowledgement

The project is supported by the Czech Science Foundation, project number 14–06678P.

References

[1] Cattivelli, F.S., Lopes, C.G., Sayed, A.H. (2008) "Diffusion recursive least-squares for distributed estimation over adaptive networks." *IEEE Transactions on Signal Processing*, 56(5), 1865–1877.

[2] Cattivelli, F.S., Sayed, A.H. (2010) "Diffusion strategies for distributed Kalman filtering and smoothing." *IEEE Transactions on Automatic Control*, 55(9), 2069–2084.

[3] Chouvardas, S., Slavakis, K., Theodoridis, S. (2011) "Adaptive robust distributed learning in diffusion sensor networks." *IEEE Transactions on Signal Processing*, 59(10), 4692–4707.

[4] Dedecius, K., Sečkárová, V. (2013) "Dynamic diffusion estimation in exponential family models." *IEEE Signal Processing Letters*, 20(11), 1114–1117.

[5] Frühwirth-Schnatter, S. (2006) *Finite mixture and Markov switching models.* Springer.

[6] Guldogan, A.B. (2014) "Consensus Bernoulli filter for distributed detection and tracking using multi-static Doppler shifts." *IEEE Signal Processing Letters*, 21(6), 671–675.

[7] Lopes, C.G., Sayed, A.H. (2008) "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis." *IEEE Transactions on Signal Processing*, 56(7), 3122–3136.

[8] Olfati-Saber, R. (2007) "Distributed Kalman filtering for sensor networks." In *2007 46th IEEE Conference on Decision and Control*, 5492–5498. IEEE.

[9] Pereira, S.S., Lopez-Valcarce, R., Pages-Zamora, A. (2013) "A diffusion-based EM algorithm for distributed estimation in unreliable sensor networks." *IEEE Signal Processing Letters*, 20(6), 595–598.

[10] Ribeiro, A., Giannakis, G.B., Roumeliotis, S.I. (2006) "SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations." *IEEE Transactions on Signal Processing*, 54(12), 4782–4795,

[11] Schizas, I.D., Mateos, G., Giannakis, G.B. (2009) "Distributed LMS for consensus-based in-network adaptive processing." *IEEE Transactions on Signal Processing*, 57(6), 2365–2382.

[12] Titterington, D.M., Smith, A.F.M., Makov, U.E. (1985) *Statistical analysis of finite mixture distributions.* John Wiley, New York.

[13] Wang, X., Zidek, J. (2005) "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.