Digital Signal Processing 78 (2018) 265–283

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Efficient linear fusion of partial estimators

David Luengo [a,*], Luca Martino [b], Víctor Elvira [c], Mónica Bugallo [d]

[a] Dep. of Signal Theory and Communications, Universidad Politécnica, Madrid, Spain
[b] Image Processing Lab., Universitat de València, Spain
[c] IMT Lille Douai & CRIStAL (UMR CNRS 9189), Villeneuve d'Ascq, France
[d] Dep. of Electrical and Computer Eng., Stony Brook University, NY, USA

ARTICLE INFO

Article history:
Available online 21 March 2018

Keywords:
Distributed estimation
Linear fusion
Constrained minimization
Big data
Bayesian estimation
Bernstein–von Mises theorem

ABSTRACT

Many signal processing applications require performing statistical inference on large datasets, where computational and/or memory restrictions become an issue. In this big data setting, computing an exact global centralized estimator is often either unfeasible or impractical. Hence, several authors have considered distributed inference approaches, where the data are divided among multiple workers (cores, machines or a combination of both). The computations are then performed in parallel and the resulting partial estimators are finally combined to approximate the intractable global estimator. In this paper, we focus on the scenario where no communication exists among the workers, deriving efficient linear fusion rules for the combination of the distributed estimators. Both a constrained optimization perspective and a Bayesian approach (based on the Bernstein–von Mises theorem and the asymptotic normality of the estimators) are provided for the derivation of the proposed linear fusion rules. We concentrate on finding the minimum mean squared error (MMSE) global estimator, but the developed framework is very general and can be used to combine any type of unbiased partial estimators (not necessarily MMSE partial estimators). Numerical results show the good performance of the algorithms developed, both in problems where analytical expressions can be obtained for the partial estimators, and in a wireless sensor network localization problem where Monte Carlo methods are used to approximate the partial estimators.

1. Introduction

Estimation theory addresses the problem of inferring a set of unknown variables of interest given a collection of available data [1,2]. This is a central problem in statistical signal processing, where a parametric model for the data is often assumed and its parameters have to be inferred from the observations [3–5]. Indeed, even non-parametric approaches typically have a reduced set of hyperparameters that need to be estimated from the data [6–8]. Unfortunately, determining the *global estimator* of these parameters using all the available information is often unfeasible or impractical for most real-world scenarios. Many current signal processing applications require performing statistical inference on large datasets, where the amount of data at hand imposes computational and/or storage constraints that impede the global estimation process [9]. Furthermore, even when approximate numerical solutions working directly on the whole dataset can be computed, they may not provide a satisfactory performance either. For example, Monte Carlo (MC) methods are often used to attain asymptotically exact estimators when closed-form analytical expressions cannot be obtained [10–12]. However, large datasets also pose a challenge for MC-based estimators, since the posterior density tends to be concentrated on a relatively small portion of the state space as the number of data increases [13]. Consequently, MC algorithms may have trouble locating this area (especially if the dimension of the state space is also large), and thus can lead to a poor performance in practice.

An alternative to *global estimation* consists of dividing the available data into groups of manageable information, and distributing them among multiple workers (cores, machines or a combination of both). The computations are then performed in parallel (with or without communication among the different workers) and *distributed* or *partial* estimators of the unknown parameters are obtained. In this setting, two extreme situations may arise, namely the multi-core and the multi-machine scenarios. On the one hand, in the *multi-core* case, the estimation is performed using several cores of a single machine (e.g., inside a graphics processing unit [GPU]) and communication among the cores can be considered costless [14,15]. This approach allows for communication

* Corresponding author.
E-mail addresses: david.luengo@upm.es (D. Luengo), luca.martino@uv.es (L. Martino), victor.elvira@telecom-lille.fr (V. Elvira), monica.bugallo@stonybrook.edu (M. Bugallo).

https://doi.org/10.1016/j.dsp.2018.03.006

among workers, can provide significant speed-ups (if synchronization issues are properly addressed), and solves the computational cost problem, but not the memory/disk storage bottleneck. On the other hand, in the *multi-machine* case, the estimation is distributed among several machines (typically lying inside a large cluster), and the cost of inter-machine communications cannot be ignored. This approach can alleviate all the issues associated to big data signal processing (i.e., both computational and memory/storage issues), but requires each machine to work independently without any communication among workers (which typically communicate only to the central node at the beginning and the end of their tasks) [16]. Finally, note that a combination of both scenarios often occurs in practice (i.e., a large cluster where each machine may have several cores), thus resulting in situations where a moderate amount of communications may be acceptable.

In this paper, we focus on the scenario where no communication exists among the workers, deriving efficient linear fusion rules for the combination of the partial estimators. Our main goal is finding an optimal combination of these partial estimators that allows us to achieve a performance which is as close as possible to the performance of the global estimator that has access to all the information. We concentrate on minimum mean squared error (MMSE) global estimators, but the developed framework is very general and can be used to combine any type of unbiased partial estimators. In the following, we review related works (Section 1.1), detail our main contributions (Section 1.2) and summarize the structure of the whole paper (Section 1.3).

1.1. Related works

The fusion of different models or estimators has been studied in many different areas including control, signal processing, econometrics and digital communications. The literature on the subject is rather vast, so here we only mention the most important results related to the problem addressed. On the one hand, a related field in the statistical literature is the combination of forecasts [17]. Indeed, the optimal linear combination for the single parameter case was already derived in [18,19], a Bayesian perspective was provided in [20], and a general procedure to combine estimators in the multiple parameter case has been proposed very recently in [21]. However, there are two important differences with respect to the scenario addressed here: (1) each forecaster is assumed to have access to the whole dataset; (2) the computational complexity issue is not addressed. Therefore, problems related to the scarcity of data per estimator (when the number of data is large but the ratio data/workers is not so large), such as the so-called *small sample bias* [22], the choice of the appropriate number of partial estimators or the feasibility of the optimal combination rules when the number of parameters to be estimated is also large, have never been investigated in this context as far as we know.

On the other hand, in wireless sensor networks the focus has been on distributed learning/estimation under communication constraints [23,24]. The optimal linear fusion rule for the multi-dimensional case has also been derived in this context [24,25], but the focus has been on developing optimal compression rules to restrict the amount of information being transmitted, rather than on obtaining efficient fusion schemes. Unfortunately, this compression is of limited use in the multi-machine learning scenario, since passing messages among multiple machines is expensive regardless of their size. Distributed fusion approaches, obtained by adapting methods developed for graphical models, have also been proposed [26], as well as many different consensus, gossip or diffusion algorithms [27–29]. However, all of these methods still require a non-negligible amount of communication that constitutes a burden for multi-machine signal processing.

Finally, there is currently a great deal of interest in parallel Bayesian computation using MC methods [30], and a few communication-free parallel Markov chain Monte Carlo (MCMC) algorithms working on disjoint partial datasets have been developed following the so-called *embarrassingly parallel* architecture [31]. In [32], four alternatives were proposed to combine the samples drawn from the partial posteriors using either a Gaussian approximation or importance resampling. Then, [33] derived the optimal linear combination of weights required to obtain samples approximately from the full posterior, noting that the approach is optimal when both the full and the partial posteriors are Gaussian. This was followed by [34], where three different approaches to approximate the full posterior from the partial posteriors were proposed: a simple parametric approach, a non-parametric estimator and a semi-parametric method. In [35], an improvement of the quality of the approximation to the full posterior was proposed by using the Weierstrass transform. A variational aggregation approach has been derived in [36], whereas an approach based on space partitioning and density aggregation has been introduced in [37]. However, none of these previous works addresses the potentially large dimension of the optimal combiners in practical problems, which demands the transmission of large matrices across the network. Furthermore, all of the aforementioned works focus on the generation of valid Monte Carlo samples at the fusion node (thus requiring the transmission of all the samples generated in the distributed nodes, which can be an excessive burden in some environments like wireless sensor networks), whereas we concentrate on the problem of obtaining a good global estimator using a single estimate from each distributed node.

1.2. Main contributions

The main contribution of this work is the derivation of two novel efficient linear schemes for the fusion of the partial estimators. Although we focus on minimum mean squared error (MMSE) partial estimators throughout the paper, the proposed fusion schemes are independent from the specific approach followed to obtain those partial estimators. Indeed, the only assumption is the unbiasedness of the partial estimators. Note that this requires both the total number of data and the number of data per partial estimator (i.e., N and N_ℓ) to be large. The case where N is large but N_ℓ is small requires either taking into account the bias in the formulation of the fusion rules (a very hard task which is still an open problem) or designing bias correction schemes, and falls outside of the scope of this paper.

Our motivation comes from the optimal linear combination, which involves the calculation of one weighting matrix per partial estimator and thus may be too computationally demanding for large dimensional systems (both in the number of unknowns and observations), as it requires as many weighting matrices (whose size depends quadratically on the number of unknowns) as partial estimators (whose number is typically a fraction of the number of observations). For instance, in a setting where the number of parameters to be estimated is D and the N observations available are equally distributed among L partial estimators, the optimal linear fusion approach requires computing one $D \times D$ matrix per partial estimator (L matrices and LD^2 parameters in total), which must be estimated from the partial dataset composed of N/L samples. In order to reduce the computational complexity, we propose two linear approaches: the single coefficient MMSE (SC-MMSE) fusion rule, that requires only a single weighting coefficient per partial estimator (i.e., L weights in total), and the independent linear MMSE (IL-MMSE) fusion, which requires one weighting coefficient per parameter and partial estimator (i.e., LD weights in total), respectively.