Digital Signal Processing 58 (2016) 10–19

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Diffusion maximum correntropy criterion algorithms for robust distributed estimation

Wentao Ma [a], Badong Chen [b,*], Jiandong Duan [a], Haiquan Zhao [c]

[a] School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, China
[b] School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China
[c] School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China

ARTICLE INFO

Article history:
Available online 19 July 2016

Keywords:
Correntropy
Maximum correntropy criterion
Diffusion
Robust distributed adaptive estimation

ABSTRACT

Robust diffusion adaptive estimation algorithms based on the maximum correntropy criterion (MCC), including adapt then combine MCC and combine then adapt MCC, are developed to deal with the distributed estimation over network in impulsive (long-tailed) noise environments. The cost functions used in distributed estimation are in general based on the mean square error (MSE) criterion, which is desirable when the measurement noise is Gaussian. In non-Gaussian situations, especially for the impulsive-noise case, MCC based methods may achieve much better performance than the MSE methods as they take into account higher order statistics of error distribution. The proposed methods can also outperform the robust diffusion least mean p-power (DLMP) and diffusion minimum error entropy (DMEE) algorithms. The mean and mean square convergence analysis of the new algorithms are also carried out.

1. Introduction

As an important issue in the field of distributed network, the distributed estimation over network plays a key role in many applications, including environment monitoring, disaster relief management, source localization, and so on [1–4], which aims to estimate some parameters of interest from noisy measurements through cooperation between nodes. Much progress has been made in the past few years. In particular, the diffusion mode of cooperation for distributed network estimation (DNE) has aroused more and more concern among researchers, which keeps the nodes exchange their estimates with neighbors and fuses the collected estimates via linear combination. So far a number of diffusion mode algorithms have been developed by researchers, such as the diffusion least mean square (DLMS) [5–8], diffusion recursive least square (DRLS) [9] and their variants [10–13]. These algorithms are derived under the popular mean square error (MSE) criterion, of which the optimizations are well understood and efficient. It is well-known that the optimality of MSE relies heavily on the Gaussian and linear assumptions. In practice, however, the data distributions are usually non-Gaussian, and in these situations, the MSE is possibly no longer an appropriate one especially in the presence of heavy-tailed non-Gaussian noise [14]. In distributed networks, some impulsive noises are usually unavoidable.

Recently, some researchers focus on improving robustness of DNE methods. The efforts are mainly directed at searching for a more robust cost function to replace the MSE cost (which is sensitive to large outliers due to the square operator). To address this problem, the diffusion least mean p-power (DLMP) based on p-norm error criterion was proposed to estimate the parameters of the wireless sensor networks [15]. For non-Gaussian cases, *Information Theoretic Learning* (ITL) [16] provides a more general framework and can also achieve desirable performance. The diffusion minimum error entropy (DMEE) was proposed in [17]. Under the MEE criterion, the entropy of a batch of N most recent error samples is used as a cost function to be minimized to adapt the weights. The evaluation of the error entropy involves a double sum over the samples, which is computationally expensive especially when the window length L is large.

In recent years, the correntropy as a nonlinear similarity measure in ITL, has been successfully used as a robust and efficient cost function for non-Gaussian signal processing [18]. The adaptive algorithms under the maximum correntropy criterion (MCC) are shown to be very robust with respect to impulsive noises, since correntropy is a measure of local similarity and is insensitive to outliers [19]. Moreover, MCC based algorithms are, in general, computationally much simpler than the MEE based algorithms. Research results on dimensionality reduction [20], feature selection

* Corresponding author.
E-mail address: chenbd@mail.xjtu.edu.cn (B. Chen).

http://dx.doi.org/10.1016/j.dsp.2016.07.009

[21], robust face recognition [22], robust principal component analysis [23], robust sparse representation [24], robust regression [25] and adaptive filtering [26–31] have demonstrated the effectiveness of MCC when dealing with occlusion and corruption problems. Recently, a robust diffusion adaptive filtering algorithm under the MCC has been proposed in [32], however, the detailed derivation, convergence performance analysis and comprehensive comparison study were not presented in that work.

Motivated by the desirable features of correntropy, we further develop in this work a novel diffusion algorithm, called *diffusion MCC* (DMCC), for robust distributed network estimation in impulsive noise environments. The main contributions of the paper are three-folds: (i) a correntropy-based diffusion scheme is proposed to solve the distributed estimation over networks; (ii) two MCC based diffusion algorithms, namely *adapt then combine* (ATC) and *combine then adapt* (CTA) diffusion algorithms are developed, which can combat impulsive noises effectively; (iii) the mean and mean square performances have been analyzed. Moreover, simulations are conducted to illustrate the performance of the proposed methods under impulsive noise disturbances.

The remainder of the paper is organized as follows. In Section 2, we give a brief review of MCC. In Section 3, we propose the DMCC method and present two adaptive combination versions. The mean and mean square analysis are performed in Section 4. Simulation results are then presented in Section 5 Finally, the paper is concluded in Section 6.

2. Maximum correntropy criterion

The correntropy between two random variables X and Y is defined by

$$V(X,Y) = E\big[\kappa(X,Y)\big] = \int \kappa(x,y)dF_{XY}(x,y) \tag{1}$$

where $E[\cdot]$ denotes the expectation operator, $\kappa(\cdot,\cdot)$ is a shift-invariant Mercer kernel, and $F_{XY}(x,y)$ denotes the joint distribution function of X and Y. In practice, only a finite number of samples $\{x(i), y(i)\}_{i=1}^{N}$ are available, and the joint distribution is usually unknown. In this case, the correntropy can be estimated by the sample mean as

$$\hat{V}(X,Y) = E\big[\kappa(X,Y)\big] \approx \frac{1}{N}\sum_{i=1}^{N}\kappa\big(x(i), y(i)\big) \tag{2}$$

The most popular kernel used in correntropy is the Gaussian kernel:

$$\kappa_\sigma(x,y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{e^2}{2\sigma^2}\right) \tag{3}$$

where $e = x - y$, and σ denotes the kernel size. With Gaussian kernel, the instantaneous MCC cost is [18]

$$J_{MCC}(i) = G_\sigma^{MCC}\big(e(i)\big) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{e^2(i)}{2\sigma^2}\right) \tag{4}$$

where i denotes the time instant (or iteration number), and $e(i) = x(i) - y(i)$.

Important properties of correntropy can be found in [18]. In recent years, correntropy has been successfully applied in robust pattern recognition [24,25] and non-Gaussian signal processing [26–31]. More recently, in [33], a generalized correntropy is proposed as a novel cost for robust adaptive filtering. Compared with other similarity measures, such as the mean square error (MSE), correntropy (with Gaussian kernel) has some desirable properties [19]: 1) it is always bounded for any distribution; 2) it contains all even-order moments, and the weights of the higher-order moments are determined by the kernel size; 3) it is a local similarity measure and is robust to outliers. As shown in Fig. 1, the correntropy cost is sensitive to the residual error only within an "observation window" (whose range is controlled by the kernel size). Thus, the MCC criterion is robust to large outliers.

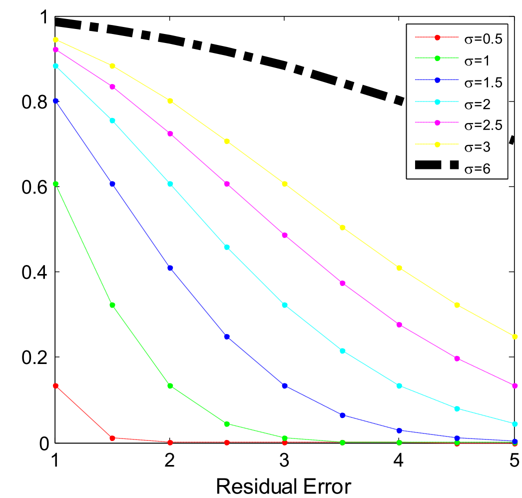

Fig. 1. Illustration of the correntropy cost with different values of kernel size.

3. Diffusion MCC algorithms

3.1. General diffusion MCC

Consider a network composed of N nodes distributed over a geographic area to estimate an unknown vector w_o of size $(M \times 1)$ from measurements collected at N nodes. At each time instant i $(i = 1, 2, \cdots, I)$, each node k has access to the realization of a scalar measurement d_k and a regression vector u_k of size $(M \times 1)$, related as

$$d_k(i) = w_o^T u_k(i) + v_k(i) \tag{5}$$

where $v_k(i)$ denotes the measurement noise, and T denotes transposition.

Given the above model, for each node k, the DMCC seeks to estimate w_o by maximizing a linear combination of the local correntropy within the node k's neighbor N_k. The cost function of the DMCC for each node can be therefore expressed as

$$\begin{aligned} J_k^{local}(w_k) &= \sum_{l\in N_k}\alpha_{l,k}G_\sigma^{MCC}\big(e_{l,k}(i)\big) \\ &= \sum_{l\in N_k}\alpha_{l,k}G_\sigma^{MCC}\big(d_l(i) - w_k^T u_l(i)\big) \end{aligned} \tag{6}$$

where w_k is the estimate of w_o, $e_l(i) = d_l(i) - w_k^T u_l(i)$, $\{\alpha_{l,k}\}$ are some non-negative combination coefficients satisfying $\sum_{l\in N_k}\alpha_{l,k} = 1$, and $\alpha_{l,k} = 0$ if $l \notin N_k$, and

$$\begin{aligned} G_\sigma^{MCC}\big(e_{l,k}(i)\big) &= \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}\big(e_l(i)\big)^2\right) \\ &= \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}\big(d_l(i) - w_k^T u_l(i)\big)^2\right) \end{aligned} \tag{7}$$

Taking the derivative of (6) yields

$$\begin{aligned} \nabla J_k^{local}(w_k) &= \sum_{l\in N_k}\alpha_{l,k}\frac{\partial G_\sigma^{MCC}(e_l(i))}{\partial w} \\ &= \frac{1}{\sigma^2}\sum_{l\in N_k}\alpha_{l,k}G_\sigma^{MCC}\big(e_l(i)\big)e_l(i)u_l(i) \end{aligned} \tag{8}$$