Received December 18, 2019, accepted January 10, 2020, date of publication January 14, 2020, date of current version January 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2966508

Distributed Adaptive Clustering Based on Maximum Correntropy Criterion Over Dynamic Multi-Task Networks

QING SHI[1,2,3,4], FULIANG HE[2], JIAGUI WU[2,3,4], (Member, IEEE), AND FENG CHEN[1,2,3,4], (Member, IEEE)

[1]College of Artificial Intelligence, Southwest University, Chongqing 400715, China
[2]College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
[3]Brain-Inspired Computing and Intelligent Control Key Laboratory, Southwest University, Chongqing 400715, China
[4]Chongqing Collaborative Innovation Center for Brain Science, Chongqing 400715, China

Corresponding author: Feng Chen (fengchen.uestc@gmail.com)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant XDJK2020B034, in part by the National Key Research and Development Program of China under Grant 2018YFB1306600, in part by the National Natural Science Foundation of China under Grant 61875168, and in part by the Chongqing Research Program of Basic Research and Frontier Technology under Grant cstc2017jcyjAX0265.

ABSTRACT This paper focuses on the problem of distributed adaptive estimation over dynamic multi-task networks, where a set of nodes is required to collectively estimate some parameters of interest from noisy measurements. Besides, since nodes in the network are constrained by communication power consumption and external interference in a non-stationary environment, the objective pursued by the node is prone to change or abnormality. The problem is worth considering in several contexts including multi-target tracking, multi-model classification and heterogeneous network segmentation. We propose a distributed adaptive clustering strategy, which is mainly composed of two procedures: normal task adaptation and the same task cluster. The task anomaly detection based on non-cooperative least-mean-squares (NC-LMS) algorithm and task switching detection based on diffusion maximum correntropy criterion (D-MCC) algorithm are provided. A series of scenarios, such as dynamic network, time-varying tasks and non-stationary (Gaussian and pulse interference) are simulated. We also discuss optimization schemes to design the NC-LMS and D-MCC weights and examine the estimate performance and clustering effects of the proposed algorithm by simulation results.

INDEX TERMS Adaptive clustering, distributed estimation, multi-task, maximum correntropy criterion.

I. INTRODUCTION

Distributed estimation for adaptation, learning, modeling, and optimization through cooperation between nodes plays an key role in reinforcement learning, signal processing, and online supervised learning and many other application areas, which aims to estimate a single parameter vector collaboratively. However, in reality, there are many parameters of interest happening to be multitask-oriented. In other words, there are multiple optimum parameter vectors that are simultaneously inferred in a collaborative manner. Multi-task problems have been studied in many important applications, such as multi-task clustering [1]–[4], multi-target tracking [5]–[7], and multi-model classification [8]–[10]. In our work, we consider the situation where there are connected clusters of nodes, and each cluster has a parameter vector to estimate.

In recent years, several useful distributed strategies have been proposed in the literature, including incremental strategies [11]–[13], consensus strategies [14], [15] and diffusion strategies [21]–[23], [48]. In particular, many researchers are attracted by diffusion strategies because of their scalability and reliability. It is worth noting that [22] has proved that the diffusion strategy in data processing on adaptive networks has better stability and robust performance than consensus-based strategies. Accordingly, the diffusion adaptive learning algorithm is mainly considered in our work. Adaptive networks are well-suited for decentralized inference, filtering and

The associate editor coordinating the review of this manuscript and approving it for publication was Theofanis P. Raptis.

TABLE 1. Possible motivations.

Reference	Adaptive clustering	Correntropy	Multi-task Networks
[8]– [10]	×	×	✓
[29]– [33]	×	✓	×
[1]– [4], [38]	✓	×	✓
This work	✓	✓	✓

clustering tasks. However, previous work on topology design and tuning techniques that was not dynamic [18]–[20], and in the sense that they cannot track changes in the network. Motivated by the problem, we develop an adaptive clustering algorithm over dynamic multi-task network in this paper, which can reduce the impact of weak links on network estimation by selecting data subsets from neighbour nodes with normal tasks.

In many previous works, distributed algorithms based on diffusion strategy have been proposed under the background of diffusion LMS [24]–[26]. An inspection of the existing articles on the above algorithms shows that most works are based on the mean-square error (MSE) cost function because it has attractive characteristics such as smoothness, convenience, low computational burden and optimality under Gaussian assumptions. If the signal is Gaussian, then MSE is desirable. However, in the case of non-Gaussian, its performance may be significantly reduced. In these cases, the non-secondary cost is usually better than MSE [27]. The kernel function in entropy is usually a Gaussian kernel due to its smoothness and strict positive determination. These properties showed the effectiveness of maximum correntropy criterion (MCC) for occlusion and corruption problems [28]–[32], [46], [47]. In particular, MCC is suited for dealing with impulsive noises.

Motivated by the desirable features of correntropy and others(see Table 1), we propose in this work a novel distributed clustering strategy based on diffusion maximum correntropy criterion (D-MMC), for robust distributed multi-task network estimation in a nonstationary environment. Moreover, we consider a general situation where there are connected clusters of nodes, and each cluster has a parameter vector to be estimated.

The main contributions of the paper are three-folds: (i) In non-stationary multi-task networks, an adaptive clustering strategy is derived, which can make the nodes in a network correctly clustered and improve clustering accuracy through enhanced intra-cluster cooperation. (ii) Normal task adaptation based on non-cooperative least-mean-squares (NC-LMS) algorithm is developed, which can discriminate the abnormality of the task to combat interference effectively. (iii) The same task clustering based on adaptive D-MCC algorithm is provided to solve the distributed estimation over multi-task networks. In addition, simulations are conducted to illustrate the performance of the proposed methods under mixed noise (Gaussian and impulsive) disturbances.

This work is organized as follows. Section II describes the system model. Section III describes the problem statement and presents a solution. In Section IV, we motivate and derive a family of diffusion LMS algorithms under the MCC for distributed clustering estimation. In Section V, we simulate different choices of weighting rules for the diffusion algorithms over MCC, and discuss the effect of kernel size on the MSD performance. In addition, we conclude our work in Section IV.

Notation: We use $(\cdot)^T$ and $|\cdot|$ represent the transpose and represents the absolute value respectively, $\mathbb{E}$ for expectation and $\|\cdot\|$ for Euclidean norm. In addition, we use the symbol $\mathbf{1}_N$ denote the vector with unit entries, and other used symbols are defined in the context of the article.

II. SYSTEM MODEL

In the section, we describe the system model and give a brief review of adaptive diffusion strategy based on minimum mean-square-error (MMSE) criterion.

A. NETWORK MODEL

We focus mainly on a set $V = \{1, 2, \cdots, N\}$ of N nodes with limited processing capabilities, which are divided into Q clusters. Particularly, the nodes are distributed over a given geographical area, and the position of them changes over time. The nodes are connected directly by an edge if they are neighbors, and note that nodes from different clusters may be connected, which forms a partially connected network. The neighborhood of an arbitrary node k is denoted as $\mathcal{N}_k$, and its size as $|\mathcal{N}_k|$ by the notation n_k. Fig. 1 provides a graphical representation of a connected network with $N = 9$ nodes, and Fig. 1 illustrates the neighborhood of node 6, which consists of nodes $\{3, 4, 8, 9\}$. Accordingly, node 6 has degree 5, which is the size of its neighborhood.

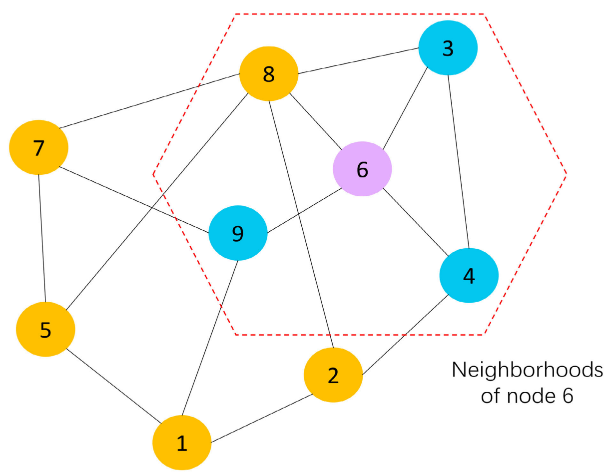


FIGURE 1. A network consists of a collection of cooperating nodes.

In our work, we consider the situation where each node k in the connected network observes random measurements $\{d_{k,i}, u_{k,i}\}$, where $d_{k,i}$ is a scalar data and $u_{k,i}$ is $1 \times L$

regression vector data, which are assumed to be related to some unknown $L \times 1$ parameter vector $w^o_{k,i}$ by a linear regression model of the form:

$$d_{k,i}=u_{k,i}w^o_{k,i}+v_{k,i}, \tag{1}$$

where $u_{k,i}$, at time instant i, is temporally white and independent over space with covariance matrix $R_{u,k} = \mathbb{E}u_{k,i}u^T_{k,i} > 0$ and zero means. $v_{k,i}$ is an additive temporally and spatially independent zero-mean noise process with a time-independent variance $\sigma^2_{v,k}$, and it is independent of every other signal over space. It is considered that nodes of different clusters track different objectives (which also call tasks) [26], and there is one task per cluster, namely,

$$w^o_k = w^o_{\mathcal{C}_q}, \quad for\ \forall\ k \in \mathcal{C}_q. \tag{2}$$

In wireless sensor networks, sensors are responsible for information collection, and different ocean buoys carry different sensors. When the sensor is accelerating or decelerating, the sensor chooses to change the target due to the constraints of communication distance and power consumption to ensure the integrity and accuracy of information collection. To understand this dynamic process of each node's tasks more clearly, we introduce the task time-varying model as follows:

$$w^o_{k,i} = s_{k,i}w^o_{k,i-1} + (1 - s_{k,i})w^o_{\mathcal{C}_p} + z_{k,i-1}, \tag{3}$$

where $z_{k,i-1}$ is the process noise for node k at time instant $i-1$, which is independent of measurement noise $v_{k,i}$ and regression vector $u_{k,i}$. When $s_{k,i} = 1$, then the task $w^o_{k,i}$ pursued by the node $k$at time i will switch from $w^o_{k,i-1}$ to $w^o_{\mathcal{C}_p}$.

In our work, we assume that the cost of one information communication between adjacent nodes is equal to c_0, and the potential communication cost of each node at the time i is $c_{k,i} = c_0 n_{k,i}$, where $n_{k,i}$ represents the number of neighbor nodes of node k at time i. When the communication cost $c_{k,i-1}$ of each node k exceeds the threshold c_r tolerated by the node k, then $s_{k,i} = 1$, otherwise $s_{k,i} = 0$. Therefore, we introduce the task switching condition:

$$s_{k,i} = \begin{cases} 1, & if\ c_{k,i-1} > c_r \\ 0, & otherwise \end{cases} \tag{4}$$

Although the cost c_0 may vary depending on the distance of the network nodes, we assume that c_0 is the same for all nodes.

B. ADAPTIVE DIFFUSION STRATEGY

Adaptive diffusion strategy based on MMSE criterion has been studied in previous studies [21], [22], [34], [39] for single-task problems, which seek the optimal linear estimator w^o that minimizes the following global cost function:

$$J^{glob}(w) = \sum_{k=1}^{N} J_k(w), \tag{5}$$

where $J_k(w)$ is the local cost function that is developed based on MSE by

$$J_k(w) = \mathbb{E}\left|d_{k,i} - u_{k,i}w_{k,i-1}\right|^2. \tag{6}$$

Adaptive diffusion strategy based on MMSE criterion contains adapt-then-combine (ATC) diffusion strategy and combine-then-adapt (CTA) diffusion strategy. The ATC algorithm has the same processing and communication complexity as the CTA algorithm while the former outperforms the latter. Therefore, in what follows, we focus on the ATC strategy to illustrate the main results. Each node k of the network tries to learn its optimum vector w^o_k from collected data $\{d_{k,i}, u_{k,i}\}$ through the ATC diffusion strategy based on MMSE, including adaptation and combination steps:

$$\begin{cases} \psi_{k,i} = w_{k,i-1} - \bar{\mu}_k \sum\limits_{\ell \in \mathcal{N}_k} a_{\ell k} \nabla J_\ell(w_{k,i-1})(\text{adaptation}) \\ w_{k,i} = \sum\limits_{\ell \in \mathcal{N}_k} c_{\ell k}\psi_{\ell,i}(\text{combination}), \end{cases} \tag{7}$$

where $\bar{\mu}_k$ is a positive constant step-size for the process repeated continuously, and the combination coefficients $\{a_{\ell k}, c_{\ell k}\}$ are used to share the local data between connected nodes. The matrices A and C are collections of coefficients $\{a_{\ell k}\}$ and $\{c_{\ell k}\}$, respectively, which are required to satisfy

$$A^T\mathbf{1}_N = \mathbf{1}_N, \quad a_{\ell k} \geq 0,\ a_{\ell k} = 0\ if\ \ell \notin \mathcal{N}_k, \tag{8}$$
$$C^T\mathbf{1}_N = \mathbf{1}_N, \quad c_{\ell k} \geq 0,\ c_{\ell k} = 0\ if\ \ell \notin \mathcal{N}_k. \tag{9}$$

By using instantaneous approximation, the approximate gradient vector can be given by

$$\nabla J_\ell(w_{k,i-1}) \approx -2(d_{\ell,i} - u_{\ell,i}w_{k,i-1})u^T_{\ell,i}. \tag{10}$$

Substituting the approximate gradient vector $\nabla J_\ell(w_{k,i-1})$ into the steepest descent strategy in (7), we can rewrite ATC diffusion strategy as

$$\begin{cases} \psi_{k,i} = w_{k,i-1} + \mu_k \sum\limits_{\ell \in \mathcal{N}_k} a_{\ell k}(d_{\ell,i} - u_{\ell,i}w_{k,i-1})u^T_{\ell,i} \\ w_{k,i} = \sum\limits_{\ell \in \mathcal{N}_k} c_{\ell k}\psi_{\ell,i}, \end{cases} \tag{11}$$

where $\mu_k = 2\bar{\mu}_k$. Indiscriminate cooperation between nodes that belong to different clusters pursuing different task, may bring undesired results and even worse results than non-cooperative approach. Accordingly, what deserves our consideration is that solving clustering problems over multi-task network through ATC diffusion strategy in (11) is challenging.

III. PROBLEM FORMULATION AND SOLUTION

A. PROBLEM FORMULATION

We consider that every node k in the network has individual cost function $J_k(w) : \mathbb{R}^{1\times L} \to \mathbb{R}$ for a vector parameter w with a unique minimized point w^o_k. Since nodes belonging to different clusters share different minimizers in the network, then the goal of nodes in the network is to deal with the clustered multi-task problem through seeking the unique minimizer of the aggregate cost function $J^{glob}(w)$, defined as

$$\min_{\{w_q\}_{q=1}^{Q}} \ J^{glob}(w_1, w_2, \cdots, w_Q) \triangleq \sum_{q=1}^{Q} \sum_{k \in \mathcal{C}_q} J_k(w_q). \tag{12}$$

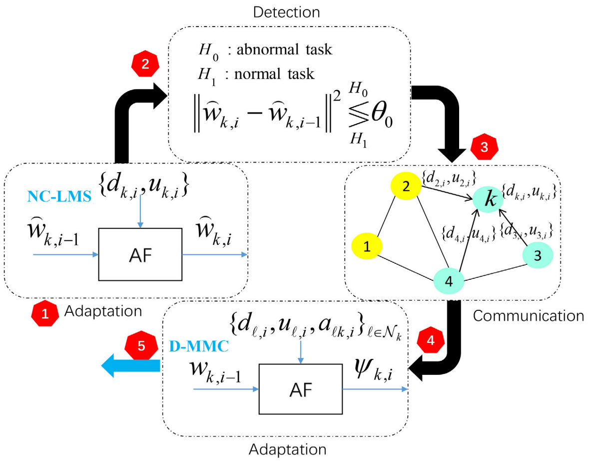


FIGURE 2. Normal task adaptation process.

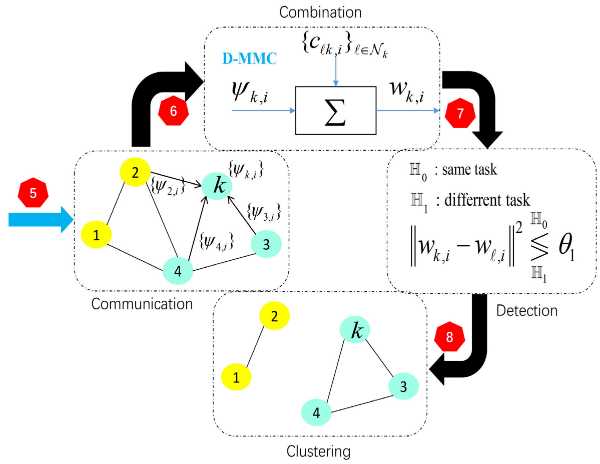


FIGURE 3. The same task clustering process.

If the cluster information $\mathcal{C}_q$ is available to nodes in the network, the clustered multi-task problem in (4) above can be broken down into separate optimization problems through subnetworks related to the clusters, namely,

$$\min_{w} J_q^{\mathcal{C}}(w) \triangleq \sum_{k\in\mathcal{C}_q} J_k(w), \tag{13}$$

for $q = 1, 2, \cdots, Q$. It is assumed that all cluster topologies are as connected as possible, and each cluster uses a diffusion strategy to find the corresponding minimizer w_q^o. The above process means that the networks connected by different cluster nodes are decomposed into subnetworks connected just to the same cluster nodes.

In our work, we consider the challenging scenario:

1) the cluster information is entirely unavailable;
2) the network changes dynamically in a non-stationary environment;
3) the task of nodes changes over time due to noise, power consumption and communication constraints.

In this case, the problem we need to solve is how to realize the decomposition of the network connected by different clusters into subnetworks connected just for the same cluster nodes.

B. SOLUTION SCHEME

To solve the problem in the previous section, we propose a distributed clustering scheme based on the maximum entropy criterion estimation algorithm, which mainly includes the two processes of normal task adaptation and the same task clustering. The former is used to provide a normal task intermediate estimate for the fusion step in the latter, which eliminates interference of abnormal tasks on parameter estimation. Whereas the latter is used to achieve distributed clustering estimation over the same task detection.

From the viewpoint of implementation, normal task adaptation process mainly includes four steps: adaptation, detection, communication, and adaptation, see the block diagram given in Fig. 2.

(1) *Adaptation:* each node k updates the estimate based on its individual measurements $\{d_{k,i}, u_{k,i}\}$;

(2) *Detection:* a threshold test is made to detect the abnormality of the task;

(3) *Communication:* each node k sends its local measurements for normal task to its neighbors and also receives the measurements $\{d_{\ell,i}, u_{\ell,i}\}$ from its neighbors;

(4) *Adaptation:* each node k updates the intermediate estimate $\psi_{k,i}$ based on local measurements of neighbors $\{d_{\ell,i}, u_{\ell,i}\}$ for normal task.

The same task clustering process mainly includes four steps: communication, combination, detection and clustering, see the block diagram given in Fig. 3.

(5) *Communication:* each node k sends its intermediate estimate to its neighbors and also receives the intermediate estimate $\psi_{\ell,i}$ from its neighbors;

(6) *Combination:* the intermediate estimates are combined based on an adaptive fusion weight $c_{\ell k,i}$;

(7) *Detection:* the same task is detected via another threshold test;

(8) *Clustering:* adjacent links within the same cluster are kept active, and adjacent links from different clusters are dropped.

The distributed clustering algorithm involved in this scheme will be elaborated in the next section.

IV. DISTRIBUTED CLUSTERING ALGORITHM

In this section, a distributed clustering algorithm is proposed for achieving reliable distributed estimation in a non-stationary environment. Each node k in the network can updates the estimate by using the non-cooperative least-mean-squares (NC-LMS) learning strategy, namely,

$$\widehat{w}_{k,i} = \widehat{w}_{k,i-1} + \mu_k(d_{k,i} - u_{k,i}\widehat{w}_{k,i-1})u_{k,i}^T. \tag{14}$$

Through the distributed diffusion strategy, we can see that each node applies adaptive gain to the measurement, during the estimation update process, to limit the impact of abnormal and damaged measurements. Since the task anomaly can be detected by the fluctuation of the adaptive gain, Thus, a hypothesis test judge whether the task is normal or not based on the updated estimate $\widehat{w}_{k,i}$ is developed to ascertain whether the task of node k is abnormal, namely,

$$\left\| \widehat{w}_{k,i} - \widehat{w}_{k,i-1} \right\|^{2} \underset{H_1}{\overset{H_0}{\lessgtr}} \theta_0, \tag{15}$$

where the H_0 hypothesis denotes the task of node k is normal, and node k sends data $\{d_{k,i}, u_{k,i}\}$ to neighbors ℓ. Conversely, the hypothesis H_1 denotes the task of node k is abnormal, and node k does not send data $\{d_{k,i}, u_{k,i}\}$ to neighbors ℓ. The threshold θ_0 is predefined. Besides, no exchange of data for abnormal task is needed during the adaptation, which makes the communication cost relatively low.

The correntropy between two random variables x and y is associated with a generalized correlation function [29], which scales the similarity of x and y via

$$V_\beta(x, y) = \mathbb{E}\left[\frac{1}{\beta\sqrt{2\pi}}\exp(-\frac{(x-y)^2}{2\beta^2})\right], \tag{16}$$

where β is the Gaussian kernel size. With Gaussian kernel and local error $e_{k,i} = d_{k,i} - u_{k,i}w_{k,i-1}$, the instantaneous correntropy cost function $J_k^{MCC}(w_{k,i-1})$ is

$$J_k^{MCC}(w_{k,i-1}) \triangleq G_\beta^{MCC}(e_{k,i}) = \mathbb{E}\left[\frac{1}{\beta\sqrt{2\pi}}\exp(-\frac{e_{k,i}^2}{2\beta^2})\right]. \tag{17}$$

To reduce the impact of noise on the estimate, each node k in the network can updates the intermediate estimate through the diffusion learning strategy over MCC. The ATC usually outperforms the CTA. Since the ATC-DMCC algorithm tends to outperform the CTA-DMCC [31], we consider ATC-DMCC in this work, namely,

$$\begin{cases} \psi_{k,i} = w_{k,i-1} + \mu_k \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \nabla J_\ell^{MCC}(w_{k,i-1}) \\ w_{k,i} = \sum_{\ell \in \mathcal{N}_k} c_{\ell k} \psi_{\ell,i}. \end{cases} \tag{18}$$

By using instantaneous approximation for (17), an approximation of the gradient vector is obtained

$$\nabla J_k^{MCC}(w_{k,i-1}) \approx \frac{1}{\beta^2} G_\beta^{MCC}(e_{k,i}) e_{k,i} u_{k,i}^T. \tag{19}$$

Based on the correntropy cost function in (18), a stochastic gradient for adaptive algorithm, the diffusion algorithm based on MCC (D-MCC), can be simply derived as

$$\psi_{k,i} = w_{k,i-1} + \frac{\mu_k}{\beta^2} \sum_{\ell \in \mathcal{N}_k} a_{\ell k} G_\beta^{MCC}(e_{\ell,i}) e_{\ell,i} u_{\ell,i}^T, \tag{20}$$

where $G_\beta^{MCC}(e_{\ell,i})$ is a Gaussian kernel, as kernel size $\beta \to \infty$, then $G_\beta^{MCC}(e_{\ell,i}) \to 1$. Note that the kernel size for correntropy function is quite important, and the Gaussian

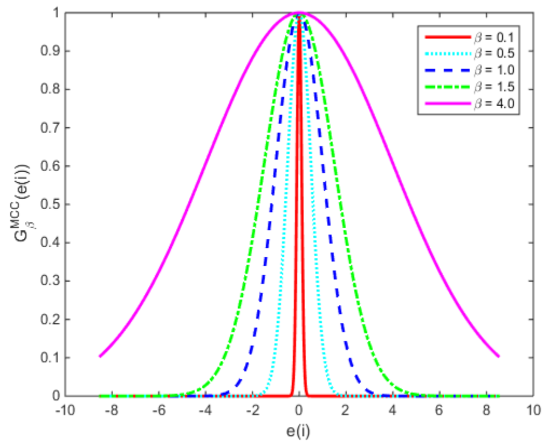


FIGURE 4. The Gaussian kernel $G_\beta^{MCC}(e(i))$ versus local error $e(i)$ for different values of size β.

kernel $G_\beta^{MCC}(e(i))$ versus local error $e(i)$ for different values of size (We choose in this work $\beta = \{0.1, 0.5, 1.0, 1.5, 4.0\}$) as shown Fig. 4.

Since the network is considered to be dynamically changing and the adaptive step in (18) only fuses the measurement data $\{d_{\ell,i}, u_{\ell,i}\}$ for the normal task. Thus, the neighborhood set $\mathcal{N}_k$ will be time-dependent and expressed as $\mathcal{N}_{k,i}$, and combination coefficient $a_{\ell k,i}$ is improved as follows:

$$a_{\ell k,i}^{+} = \begin{cases} a_{\ell k}, & if \left\| \widehat{w}_{\ell,i} - \widehat{w}_{\ell,i-1} \right\|^2 < \theta_0 \\ 0, & otherwise \end{cases} \tag{21}$$

where $a_{\ell k} \geq 0, \sum_{\ell}^{N} a_{\ell k} = 1, a_{\ell k} = 0 \, if \, \ell \notin \mathcal{N}_k$ for $k = 1, 2, \ldots, N$. Therefore, the algorithm D-MCC will be rewritten as

$$\psi_{k,i} = w_{k,i-1} + \eta_k \sum_{\ell \in \mathcal{N}_{k,i}} a_{\ell k,i}^{+} G_\beta^{MCC}(e_{\ell,i}) e_{\ell,i} u_{\ell,i}^T, \tag{22}$$

where $\eta_k = \frac{\mu_k}{\beta^2}$ is the step size. It is worth mentioning that the selection of combination coefficients $c_{\ell k,i}$ has a significant impact on the performance of multi-task networks. As mentioned earlier, the task of nodes may switch due to power consumption and communication constraints. Blindly blending intermediate estimates will reduce the accuracy of parameter estimation and affect clustering accuracy. Hence, an adaptive combination weights rule should be formulated to help the nodes ignoring this misleading task. Note that the weights can be used to minimize the instantaneous mean-square deviation (MSD) of the network:

$$\min_{\{c_{\ell k,i}\}} MSD(i) \triangleq \frac{1}{N} \sum_{k=1}^{N} \mathbb{E} \left\| w_k^o - w_{k,i} \right\|. \tag{23}$$

The combination coefficients $c_{\ell k,i}$ can be obtained through (24)(see Appendix I), which can be approximated by

$$c_{\ell k,i}=\begin{cases}\dfrac{\left\|\psi_{k,i}-w_{k,i-1}\right\|^{-2}}{\sum\limits_{j\in\mathcal{N}_{k,i}^{-}}\left\|\psi_{j,i}-w_{k,i-1}\right\|^{-2}}, & if\ \ell\in\mathcal{N}_{k,i}^{-}\\ 1-\sum\limits_{m\in\mathcal{N}_{k,i}^{-}}c_{\ell m,i}, & if\ \ell=k\\ 0, & otherwise,\end{cases}\tag{24}$$

where $\mathcal{N}_{k,i}^{-}\triangleq\mathcal{N}_k\backslash\{k\}$. From the combination rule in (23) we can find that the closeness of the local estimate to the neighboring intermediate estimates is adaptive to adjust the combination weights. In other word, the combination rule gives larger weights to neighbors with common cluster and smaller weights to neighbors that come from different clusters. Then, the combination step in (18) is rewritten by

$$w_{k,i}=\sum_{\ell\in\mathcal{N}_{k,i}}c_{\ell k,i}\psi_{k,i}.\tag{25}$$

Because $w_{k,i}$ and $w_{\ell,i}$ can be accessed through local interaction in the community, and the estimated error values $\left|w_{k,i}-w_{\ell,i}\right|$ of neighbors can reflect the similarity of the two tasks, the larger the error value, the smaller the similarity of their tasks. Therefore, we introduce another hypothesis test by using these dynamically-evolving estimates. The hypothesis test is based on the updated estimate $w_{k,i}$, and it is developed to ascertain whether the tasks of node k and node ℓ are the same at time i,namely,

$$\left\|w_{k,i}-w_{\ell,i}\right\|^2\underset{\mathbb{H}_1}{\overset{\mathbb{H}_0}{\lessgtr}}\theta_1,\tag{26}$$

where θ_1 is a predefined threshold. In general, task anomalies have a much greater impact than task similarity, so the threshold θ_0 is greater than θ_1. The $\mathbb{H}_0$ hypothesis denotes the tasks of node k and node ℓ are the same, and the link between node k with neighbor ℓ are active. Conversely, the hypothesis $\mathbb{H}_1$ denotes the tasks of node k and node ℓ are different, and the links between node k with neighbor ℓ are dropped. Then, the cluster connection coefficient $l_{k\ell,i}^{c}$ is given by

$$l_{k\ell,i}^{c}=l_{\ell k,i}^{c}=\begin{cases}1, & if\ \mathbb{H}_0\ success\\ 0, & otherwise.\end{cases}\tag{27}$$

The proposed diffusion clustering algorithm over dynamic multi-task network with adaptive combination rules is presented (see Table 2).

V. NUMERICAL EXPERIMENTS

In this section, some numerical simulations are performed to evaluate the estimate performance and clustering effects of the proposed method in a variety of different scenarios, i.e., dynamic network, time-varying tasks, non-stationary (Gaussian interference or pulse interference) and all nodes have no prior knowledge about the clusters. In addition, we compare it with those of several other diffusion clustering strategies.

TABLE 2. Summary of the proposed algorithm.

Algorithm: Diffusion clustering algorithm
Initialize: Start with $\{w_{k,-1}=0\}$,for each node k. Let $0<T_s<T$ Set values of β, μ_k, and $0<\theta_0<\theta_1<1$.   **Algorithm steps:** For each node k of the network at time $i=1:T$:   **Stage 1 (Adaptation for NC-LMS):** $\widehat{w}_{k,i}=\widehat{w}_{k,i-1}+\mu_k(d_{k,i}-u_{k,i}\widehat{w}_{k,i-1})u_{k,i}^T$ **Stage 2 (Detection for abnormal tasks):** $\left\|\widehat{w}_{k,i}-\widehat{w}_{k,i-1}\right\|^2\underset{H_1}{\overset{H_0}{\lessgtr}}\theta_0$ **Stage 3 (Adaption for D-MCC based on normal tasks):** $\psi_{k,i}=w_{k,i-1}+\eta_k\sum_{\ell\in\mathcal{N}_k}a_{\ell k}^{+}G_\beta^{MCC}(e_{\ell,i})e_{\ell,i}u_{\ell,i}^T$ with the adaptive combination coefficient $a_{\ell k,i}^{+}$: $a_{\ell k,i}^{+}=\begin{cases}a_{\ell k}, if\left\|\widehat{w}_{\ell,i}-\widehat{w}_{\ell,i-1}\right\|^2<\theta_0\\ 0, otherwise\end{cases}$ **Stage 4 (Combine weights):** $w_{k,i}=\sum_{\ell\in\mathcal{N}_{k,i}}c_{\ell k,i}\psi_{k,i}$ with the adaptive combination coefficient $c_{\ell k,i}$: $c_{\ell k,i}=\begin{cases}\frac{\left\|\psi_{k,i}-w_{k,i-1}\right\|^{-2}}{\sum_{j\in\mathcal{N}_{k,i}^{-}}\left\|\psi_{j,i}-w_{k,i-1}\right\|^{-2}}, & if\ \ell\in\mathcal{N}_{k,i}^{-}\\ 1-\sum_{m\in\mathcal{N}_{k,i}^{-}}c_{\ell m,i}, & if\ \ell=k\\ 0, & otherwise\end{cases}$ **Stage 5 (Detection for the same tasks and clustering):** $l_{k\ell,i}^{c}=l_{\ell k,i}^{c}=\begin{cases}1, if\left\|w_{k,i}-w_{\ell,i}\right\|^2<\theta_1\\ 0, otherwise\end{cases}$

A. MODEL VALIDATION

The topology of the network consisting of $N=20$ nodes divided into $Q=3$ clusters, i.e., $\mathcal{C}_1=\{1-5\}$, $\mathcal{C}_2=\{7-14\}$, and $\mathcal{C}_3=\{15-20\}$, with connection is generated as a random geometric graph model as shown in Fig. 5. The location coordinates $(x_{k,i},y_{k,i})$ of each node k in the square region $[0,110]\times[0,110]$. In the time-varying scenario, they vary according to the first-order Markov vector process:

$$\begin{aligned}x_k(i)&=bx_k(i-1)+h(i),\\ y_k(i)&=by_k(i-1)+h(i).\end{aligned}\tag{28}$$

where $b=0.98$ and $h(i)$ is an independent zero-mean Gaussian vector process with variance

$$\sigma_h^2=\begin{cases}0.01\rightarrow\text{Small interference}\\ 1\rightarrow\text{Big interference.}\end{cases}\tag{29}$$

The input regression data with size $L=4$ and the zero-mean Gaussian noise are independent in time and space with statistical profiles shown in Fig. 6. The parameters of the algorithms are fixed to $\mu_k=0.04$, $\beta=3$, $\theta_0=0.3$, and $\theta_1=0.04$. The loading factors for the three clusters, w_1^o, w_2^o, and w_3^o are randomly generated set to $\frac{randn(L,1)}{\sqrt{L}}$, where $L=4$ and $randn(\cdot)$ is the function of generating Gaussian random. In the time-varying scenario, they vary as mentioned early (4)-(5) and $c_r=0.03$. The corresponding probability density

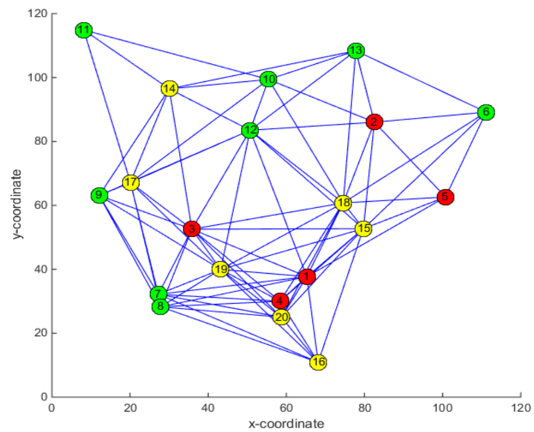


FIGURE 5. Network initial topology.

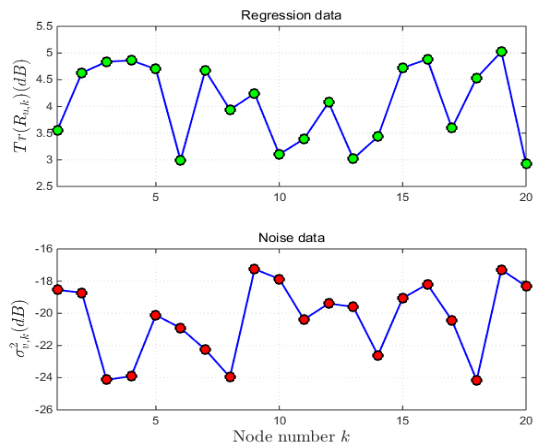


FIGURE 6. Network data statistical profiles.

function of the process noise z_{i-1} is attained by.

$$p_n(z_i)=\frac{1-\kappa_i}{\sqrt{2\pi\sigma_{z0}^2}}\exp\left(-\frac{z_i}{2\sigma_{z0}^2}\right)+\frac{\kappa_i}{\sqrt{2\pi\sigma_z^2}}\exp\left(-\frac{z_i}{2\sigma_z^2}\right), \tag{30}$$

where $\sigma_z^2 = \sigma_{z0}^2 + \sigma_{zi}^2$. Increasing κ_i leads to more frequent impulses:

$$\begin{aligned} &if\ \kappa_i = 0 \rightarrow \text{Gaussian} \\ &if\ \kappa_i \neq 0 \rightarrow \text{Impulsive}. \end{aligned} \tag{31}$$

B. ILLUSTRATIVE NUMERICAL SIMULATION

In this subsection, to illustrate the adaptive network performance, we provide some simulation examples in Figs. 7-9, and we initialize all nodes with the parameter vectors $w_{k,-1} = 0$. All simulation curves are obtained by an average of 50 runs, and the number of repetitions for per simulation is set to 500. In addition, all results are obtained by taking the overall average of the network MSDs in 300 independent Monte Carlo runs.

1) THE EFFECT OF KERNEL SIZE ON THE MSD PERFORMANCE

In this subsubsection, to investigate the effect of kernel size on the MSD performance, we show the convergence curves of the proposed algorithm with different kernel size in Fig. 7. Specifically, four kernel sizes $\beta = \{0.1, 0.5, 1.0, 1.5\}$ are chosen, and all the other parameters remain unchanged to look into the performance of the proposed algorithm. According to the Fig. 7, the best MSD performance of the proposed algorithm obtain by using a large kernel size. Thus, we conclude that a larger size is able to efficiently aggregate dynamic nodes of different tasks on the network.

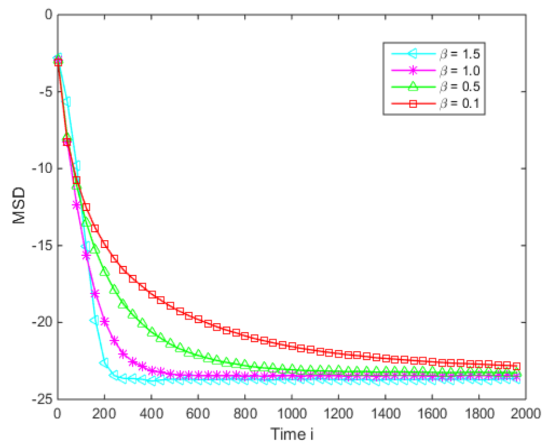


FIGURE 7. The MSD(dB) performance versus different kernel sizes $\beta = \{0.1, 0.5, 1.0, 1.5\}$.

2) THE EFFECT OF COMBINATION RULE ON THE MSD PERFORMANCE

For comparison aims, the properties for uniform, metropolis, laplacian combination rule and the proposed in this work are presented (see Table 3). In addition, the MSD performances for diffusion strategy with the above five combination rules are depicted in Fig. 8. It can be found that since uniform, metropolis and laplacian cooperation rules simply deal with the estimates from the neighbors without reasonable error penalty, these cooperation rules algorithms introduce biases that cause them to provide lower performance. In comparison, since in our work nodes with the different task are prevented to cooperate with each other in non-stationary environments, the proposed algorithm with cooperation policy based on an adaptive MCC combination weight attains a superior MSD learning performance.

3) THE MSD PERFORMANCE FOR DIFFERENT CLUSTERING ALGORITHMS

In order to investigate the estimated performance of our proposed algorithm, we compare it with the diffusion LMS introduced in previous study [26] by their MSD learning curves, over a dynamic multi-task network. In more detail, the algorithms that need to be compared are as follows: diffusion

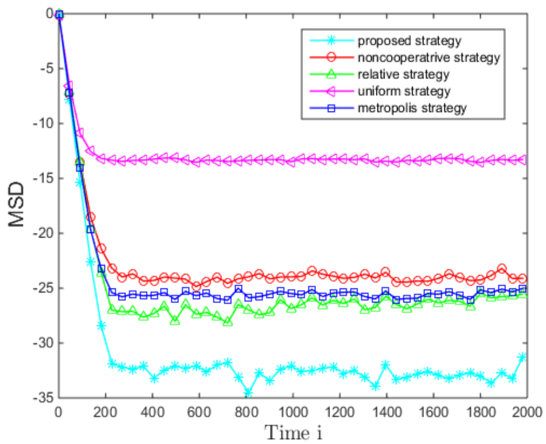


FIGURE 8. The MSD(dB) performance versus different combination rule.

TABLE 3. Possible combination rules properties.

Name	Rule $\ell \in \mathcal{N}_k, \ell \neq k$	Reference
Uniform	$c_{\ell k} = 1/n_k$	[40]
Laplacian	$c_{\ell k} = 1/n_{\max}$	[41]
Metropolis	$c_{\ell k} = 1/\max\{n_k, n_\ell\}$	[42]
Relative degree	$c_{\ell k} = n_\ell/(\sum_{m\in\mathcal{N}_k} n_m)$	[39]
Relative degree-variance	$c_{\ell k} = \sigma_\ell^{-2}/(\sum_{m\in\mathcal{N}_k} \sigma_m^{-2})$	This work

Note that in all cases, $c_{\ell k} = 0, if\ \ell \notin \mathcal{N}_k$, and for all node k, c_{kk} is chosen such that $\sum_{\ell=1}^{N} c_{\ell k} = 1$

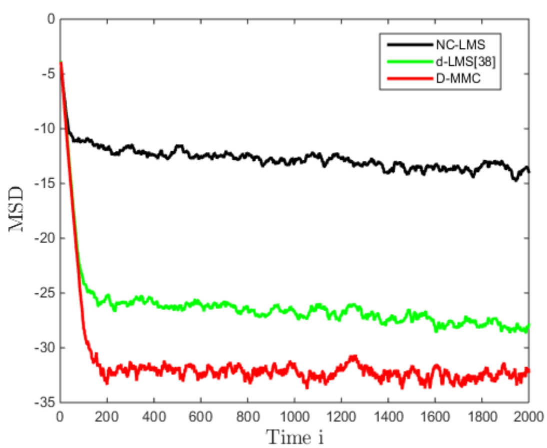


FIGURE 9. The average MSD(dB) of the network for different algorithms.

LMS (d-LMS) with the cooperation policy over multi-task networks [26], non-cooperative LMS (NC-LMS), the proposed diffusion algorithm over MCC (D-MCC). The MSD learning curves of the above three algorithms are depicted in Fig. 9. From Fig. 9, we can find that the proposed algorithm has a superior performance in comparison with the other two algorithms. In another word, the proposed algorithm can distinguish the tasks well for clustering, resulting in a better MSD learning performance in comparison with diffusion LMS in previous work.

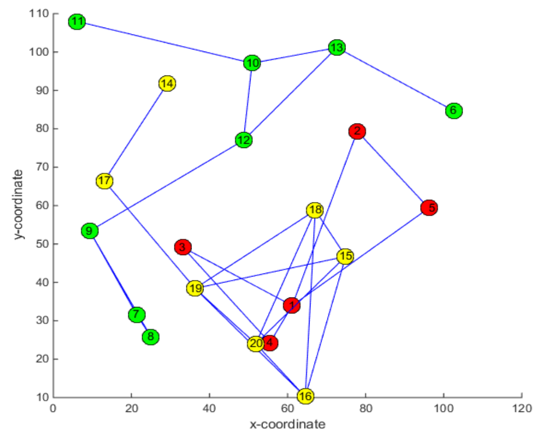


FIGURE 10. The resulting topology of the subnetwork over Scenario 1.

C. ADAPTIVE CLUSTERING OVER TIME-VARYING MODEL

In the subsection, under the assumption that all nodes have no prior knowledge about the clusters, clustering effects of the proposed method in the following two scenarios are illustrated. Scenario 1: The location coordinates of nodes in network fluctuate is gentle (the network structure changes slightly, i.e., $\sigma_h^2 = 0.01$), then for zero-mean Gaussian interference with $\kappa_i = 0$ and $\sigma_{z_i}^2 = 0.5$, the task of nodes does not switch and cannot evolve to exceptions. Scenario 2: The location coordinates of nodes in network fluctuate is dramatic (the network structure changes wildly, i.e., $\sigma_h^2 = 1$), then for Impulse interference with and $\sigma_{z_i}^2 = 10^3$, the task of nodes is switched and abnormal.

1) SCENARIO 1 (NO TASK SWITCHING OR EXCEPTION)

In the first scenario, the network node position fluctuates slightly, and it suffers from constrained Gaussian interference. After approximate 400 iterations of MSD curves, The clustering decision of the proposed algorithm does not change with time. The neighboring links within the same cluster are active whereas the neighboring links, which come from different clusters, are dropped. Fig. 10 illustrates the resulting topology when the network is in steady-state. It can be seen that the underlying topology as shown in Fig. 5 is pruned and divided into three disjoint subnetworks. This result implies that the proposed clustering strategy can suppress the interference between clusters. From the simulation results, we find that there is no task switching or exception when the nodes are under Gaussian interference with small constraint. The MSD learning curves for the proposed clustering algorithm consisting of the recursions NC-LMS and D-MCC are plotted in Fig. 11. It is obvious that three clusters take MCC cooperation clustering policy to improve their MSD performance on average.

2) SCENARIO 2 (TASK SWITCHING AND EXCEPTION)

In the second scenario, the network node position fluctuates wildly, and it suffers from Impulse interference.

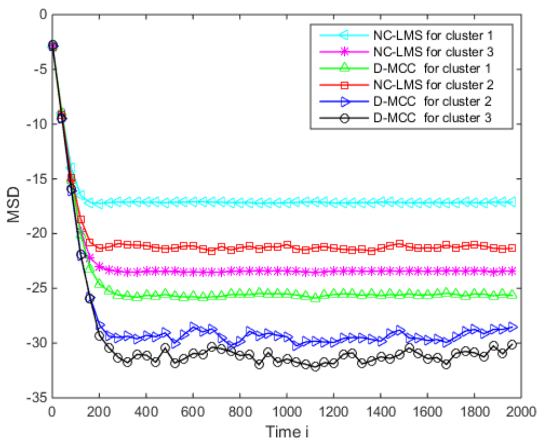


FIGURE 11. The subnetwork average MSDs for NC-LMS and D-MCC over Scenario 1.

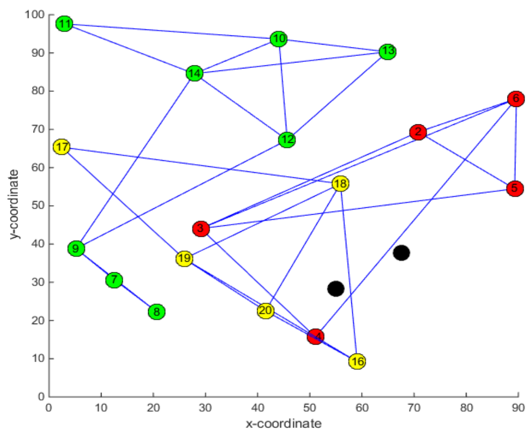


FIGURE 12. The resulting topology of the subnetwork over Scenario 2.

After approximate 400 iterations of MSD curves, similarly, the clustering decision of the proposed algorithm does not change with time. In addition, the neighboring links within the same cluster are active, whereas the neighboring links that come from different clusters, are dropped. Fig. 12 illustrates that the three subnetworks are themselves connected when network is at steady-state. There is a similar implied result with scenario 1 that the proposed clustering strategy can suppress the interference between clusters. From Fig. 12, we can see that there are task switching and exceptions when the nodes are Impulse interference. The MSD learning curves for the proposed clustering algorithm consisting of the recursions NC-LMS and D-MCC are plotted in Fig. 13. Obviously, the proposed algorithm for recursion D-MCC has a superior performance in comparison with recursion NC-LMS. Additionally, comparing Fig. 11 and Fig. 13, we can find that the MSD performance of the proposed algorithm decreases with increasing noise.

In our work, we mainly focus on realistic wireless sensor networks, and then consider dynamic multi-task networks. When the network topology changes randomly, the neighbor domain of the network node changes, and at the same time,

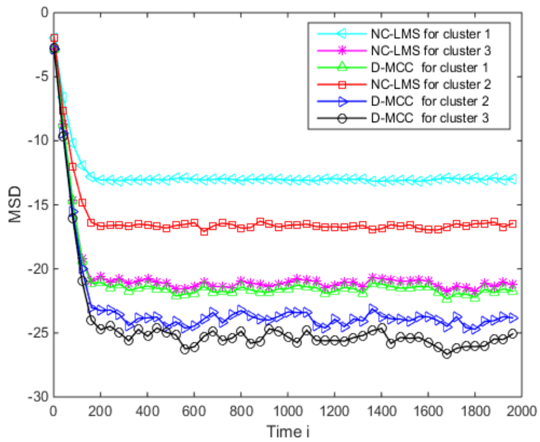


FIGURE 13. The subnetwork average MSD(dB) for NC-LMS and D-MCC over Scenario 2.

the communication power consumption of the network node gradually decreases. In order to complete the network communication, the node switches to a task with relatively low consumption. Therefore, the network we have considered in this article is general and has certain applicability in other more complex networks, such as multi-agent networks, complex networks, and wireless sensor networks, etc. For details, see references [40], [43], [44]

VI. CONCLUSION

In this paper, we consider a dynamic non-stationary multi-task network where nodes are constrained by communication power consumption and external interference. To solve the multi-task problems, we develop a distributed adaptive clustering strategy based on MCC, which can enable nodes that get a similar normal task to have a collaboration with each other, and nodes that have discommoned or abnormal tasks are prevented from collaborating. The clustering approach can be used not only to segment heterogeneous networks to enhance intra-cluster collaboration and suppress cross-cluster interference, but also to prevent intrusions or interference by isolating abnormal tasks from normal tasks for homogeneous networks. In addition, we simulate a variety of scenarios and examine the estimated performance and clustering effect of the proposed algorithm.

APPENDIX

Let $\vartheta_{k,i} = w_k^{\circ} - w_{k,i}$, At each instant i, following the expression (23), we can give the instantaneous MSD at node k by

$$E\left\{\left\|\vartheta_{k,i}\right\|^2\right\} = E\left\{\left\|w_k^{\circ} - \sum_{\ell\in\mathcal{N}_k} c_{\ell k}\psi_{\ell,i}\right\|^2\right\}. \tag{32}$$

Because the matrix C is assumed left-stochastic in (9), the expression (32) can be rewritten as

$$E\left\{\left\|\vartheta_{k,i}\right\|^2\right\} = \sum_{\ell\in\mathcal{N}_k}\sum_{p\in\mathcal{N}_k} c_{\ell k}c_{pk}E\left\{\left[w_k^{\circ}-\psi_{\ell,i}\right]^T\left[w_k^{\circ}-\psi_{p,i}\right]\right\}. \tag{33}$$

Let ψ_k be the matrix at each node k with (ℓ, p)-th entry defined as

$$[\psi_k]_{\ell p} = \begin{cases} E\left\{\left[w_k^\circ - \psi_{\ell,i}\right]^T \left[w_k^\circ - \psi_{p,i}\right]\right\}, & \ell, p \in \mathcal{N}_k \\ 0, & otherwise. \end{cases} \tag{34}$$

Let $c_k = [c_{1k}, \ldots, c_{Nk}]^T$. Minimizing (33) for node k at time i, subject to left-stochasticity of C and $c_{\ell k} = 0$ for $\ell \notin \mathcal{N}_k$, can be formulated as follows:

$$\begin{aligned} &\min c_k^T \psi_k c_k \\ &subject\ to\ \mathbf{1}_N^T c_k = 1, \quad c_{\ell k} \geq 0, \\ &\qquad c_{\ell k} = 0 \quad if\ \ell \notin \mathcal{N}_k. \end{aligned} \tag{35}$$

Generally, since ψ_k and w_k° are unknown, it is impossible to solve (35) at each node k. Thus, we use an approximation for w_k° to approximate matrix ψ_k, and to drop its off-diagonal entries in order to make the problem tractable. The resulting problem is as follows:

$$\begin{aligned} &\min \sum_{\ell=1}^{N} c_{\ell k}^2 \left\| w_{k,i-1} - \psi_{\ell,i} \right\|^2 \\ &subject\ to\ \mathbf{1}_N^T c_k = 1, \quad c_{\ell k} \geq 0, \\ &\qquad c_{\ell k} = 0 \quad if\ \ell \notin \mathcal{N}_k. \end{aligned} \tag{36}$$

with $w_{k,i-1}$some approximation for w_k°. The objective function shown above has the natural interpretation of penalizing the combination weight $c_{\ell k}$ assigned by node ℓ to node k if the local estimate at node ℓ is far from the objective at node k. The solution to this problem is given by

$$c_{\ell k,i} = \frac{\left\| w_{k,i-1} - \psi_{\ell,i} \right\|^{-2}}{\sum_{j \in \mathcal{N}_k} \left\| w_{k,i-1} - \psi_{j,i} \right\|^{-2}}, \quad for\ \ell \in \mathcal{N}_k. \tag{37}$$

REFERENCES

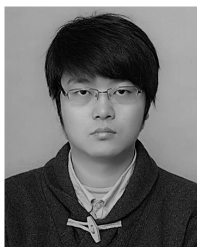
[1] X. Zhao and A. H. Sayed, "Clustering via diffusion adaptation over networks," in *Proc. 3rd Int. Workshop Cogn. Inf. Process. (CIP)*, Baiona, Spain, May 2012, pp. 1–6.
[2] A. Sayed, "Adaptation, learning, and optimization over networks," *Found. Trends Mach. Learn.*, vol. 7, nos. 4–5, pp. 311–801, 2014.
[3] L. Jacob, F. Bach, and J.-P. Vert, "Clustered multitask learning: Aconvex formulation," in *Proc. Neural Inf. Process. Syst. (NIPS)*, Vancouver, BC, Canada, Dec. 2008, pp. 1–8.
[4] X. Zhao and A. H. Sayed, "Distributed clustering and learning over networks," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3285–3300, Jul. 2015.
[5] J. Liu, M. Chu, and J. Reich, "Multitarget tracking in distributed sensor networks," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 36–46, May 2007.
[6] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control.*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
[7] X. Zhang, "Adaptive control and reconfiguration of mobile wireless sensor networks for dynamic multi-target tracking," *IEEE Trans. Autom. Control.*, vol. 56, no. 10, pp. 2429–2444, Oct. 2011.
[8] I. Francis and S. Chatterjee, "Classification and estimation of several multiple regressions," *Ann. Statist.*, vol. 2, no. 3, pp. 558–561, May 1974.
[9] X.-R. Li and Y. Bar-Shalom, "Multiple-model estimation with variable structure," *IEEE Trans. Autom. Control.*, vol. 41, no. 4, pp. 478–493, Apr. 1996.
[10] V. Cherkassky and Y. Ma, "Multiple model regression estimation," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 785–798, Jul. 2005.
[11] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM J. Optim.*, vol. 12, no. 1, pp. 109–138, Jan. 2001.
[12] M. Rabbat and R. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 798–808, Apr. 2005.
[13] C. Lopes and A. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4064–4077, Aug. 2007.
[14] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.
[15] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.
[16] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.
[17] P. Di Lorenzo, S. Barbarossa, and A. H. Sayed, "Sparse diffusion LMS for distributed adaptive estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3281–3284.
[18] C. G. Lopes and A. H. Sayed, "Diffusion adaptive networks with changing topologies," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Las Vegas, NV, USA, Mar. 2008.
[19] B. H. Fadlallah and J. C. Príncipe, "Diffusion least-mean squares over adaptive networks with dynamic topologies," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Dallas, TX, USA, Aug. 2013.
[20] T. Wimalajeewa and S. K. Jayaweera, "Distributed node selection for sequential estimation over noisy communication channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2290–2301, Jul. 2010.
[21] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.
[22] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.
[23] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4289–4305, Aug. 2012.
[24] S.-Y. Tu and A. H. Sayed, "Distributed decision-making over adaptive networks," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1054–1069, Mar. 2014.
[25] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4129–4144, Aug. 2014.
[26] J. Chen, C. Richard, and A. H. Sayed, "Diffusion LMS over multitask networks," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2733–2748, Jun. 2015.
[27] A. Singh and J. C. Príncipe, "Using correntropy as a cost function in linear adaptive filters," in *Proc. Int. Joint Conf. Neural Netw.*, Jun. 2009.
[28] B. Chen and J. C. Príncipe, "Maximum correntropy estimation is a smoothed MAP estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 491–494, Aug. 2012.
[29] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Príncipe, "Generalized correntropy for robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.
[30] B. Chen, L. Xing, B. Xu, H. Zhao, N. Zheng, and J. C. Príncipe, "Kernel risk-sensitive loss: Definition, properties and application to robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2888–2901, Jun. 2017.
[31] W. Ma, B. Chen, J. Duan, and H. Zhao, "Diffusion maximum correntropy criterion algorithms for robust distributed estimation," *Digit. Signal Process.*, vol. 58, pp. 10–19, Nov. 2016.
[32] W. Liu, P. P. Pokharel, and J. C. Príncipe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[33] B. Chen, Y. Zhu, J. Hu, and J. C. Príncipe, *System Parameter Identification: Information Criteria and Algorithms*. New York, NY, USA: Newnes, 2013.
[34] F. Cattivelli and A. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.
[35] F. Cattivelli, C. Lopes, and A. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.

[36] V. Blondel, J. Hendrickx, A. Olshevsky, and J. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *Proc. 44th IEEE Conf. Decision Control*, Seville, Spain, Dec. 2006, pp. 2996–3000.
[37] D. S. Scherber and H. C. Papadopoulos, "Locally constructed algorithms for distributed computations in ad-hoc networks," in *Proc. 3rd Int. Symp. Inf. Process. Sensor Netw. (IPSN)*, Berkeley, CA, USA, Apr. 2004, pp. 11–19.
[38] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, Sep. 2004.
[39] W. Ma, J. Duan, W. Man, J. Liang, and B. Chen, "General mixed-norm-based diffusion adaptive filtering algorithm for distributed estimation over network," *IEEE Access*, vol. 5, pp. 1090–1102, 2017.
[40] Y. Liu, C. Li, W. K. Tang, and Z. Zhang, "Distributed estimation over complex networks," *Inf. Sci.*, vol. 197, pp. 91–104, Aug. 2012.
[41] F. Chen and X. Shao, "Broken-motifs diffusion LMS algorithm for reducing communication load," *Signal Process.*, vol. 133, pp. 213–218, Apr. 2017.
[42] Y. Wang, H. Liu, W. Zheng, Y. Xia, Y. Li, P. Chen, K. Guo, and H. Xie, "Multi-objective workflow scheduling with deep-Q-network-based multi-agent reinforcement learning," *IEEE Access*, vol. 7, pp. 39974–39982, 2019.
[43] X. Li, Q. Shi, S. Xiao, S. Duan, and F. Chen, "A robust diffusion minimum kernel risk-sensitive loss algorithm over multitask sensor networks," *Sensors*, vol. 19, no. 10, p. 2339, May 2019.
[44] F. Chen, X. Li, S. Duan, L. Wang, and J. Wu, "Diffusion generalized maximum correntropy criterion algorithm for distributed estimation over multitask network," *Digit. Signal Process.*, vol. 81, pp. 16–25, Oct. 2018.
[45] W. Li, K. Liao, Q. He, and Y. Xia, "Performance-aware cost-effective resource provisioning for future grid IoT-cloud system," *J. Energy Eng.*, vol. 145, no. 5, Oct. 2019, Art. no. 04019016.
[46] L. Dang, B. Chen, S. Wang, Y. Gu, and J. C. Príncipe, "Kernel Kalman filtering with conditional embedding and maximum correntropy criterion," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 11, pp. 4265–4277, Nov. 2019.
[47] B. Chen, L. Dang, Y. Gu, N. Zheng, and J. C. Príncipe, "Minimum error entropy Kalman filter," *IEEE Trans. Syst., Man, Cybern. Syst.*, to be published.
[48] F. Chen, T. Shi, S. Duan, L. Wang, and J. Wu, "Diffusion least logarithmic absolute difference algorithm for distributed estimation," *Signal Process.*, vol. 142, pp. 423–430, Jan. 2018.

QING SHI received the B.S. degree in communication engineering from Hubei Normal University, in 2017. She is currently pursuing the master's degree with the College of Electronic and Information Engineering, Southwest University, Chongqing, China. Her research interests include machine learning, multitask networks, and adaptive signal processing.

FULIANG HE received the Ph.D. degree from Chongqing University, in 2018. He is currently a Lecturer with the College of Electronic and Information Engineering, Southwest University, Chongqing, China. He has published over ten articles in various journals. His research interests include artificial neural networks, infrared image processing, and computer vision.

JIAGUI WU (Member, IEEE) was born in Sichuan, China, in 1981. He received the B.Sc. degree in physics and the M.Sc. degree in optics from Southwest University, Chongqing, China, in 2003 and 2006, respectively, and the Ph.D. degree in non-linear dynamics from the Sichuan University, Chengdu, China. He is currently a Professor with the College of Electronic and Information Engineering, Southwest University, and a Visiting Scholar with the Electrical Engineering and Computer Science Department, University of California, Los Angeles, USA. He has authored or coauthored over 60 publications, including about 40 journal articles. His current research interests include the information security, nonlinear dynamics of lasers, micro-nanophotonic devices, signal and information processing, chaos generation and synchronization, and secure communication with improved privacy.



FENG CHEN (Member, IEEE) received the Ph.D. degree from the University of Electronic Science and Technology of China, in 2013. Since 2008, he visited in Prof. Emery Brown's research group at the Massachusetts Institute of Technology, and also worked in Harvard University, Massachusetts General Hospital. Since 2013, he has been an Associate Professor with the College of Electronic and Information Engineering, Southwest University, Chongqing, China. He has published over 20 articles in various journals and conference proceedings. His research interests include machine learning, wireless sensors networks, and signal processing.