ARTICLE IN PRESS

Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Information reliability in complex multitask networks

Sadaf Monajemi [a,*], Saeid Sanei [b], Sim-Heng Ong [c]

[a] NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore
[b] Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK
[c] Electrical and Computer Engineering, Faculty of Engineering, National University of Singapore, Singapore

HIGHLIGHTS

- The multitask network is studied where sharing genuine information is subject to a cost.
- An information credibility model to obtain the probability of sharing genuine information.
- Adaptive reputation protocol to select the most reputable neighbors for sharing true information.

ARTICLE INFO

Article history:
Received 31 October 2016
Received in revised form 18 May 2017
Accepted 10 July 2017
Available online xxxx

Keywords:
Adaptive distributed networks
Information and communication security
Reputation protocol
Information credibility
Diffusion adaptation

ABSTRACT

In this paper, we study the performance of multitask distributed networks where sharing genuine information is costly. We formulate an information credibility model which links the probability of sharing genuine information at each time instant to the cost. Each agent then shares its true information with only a subset of its neighbors while sending fabricated data to the rest according to this probability. This behavior can affect the performance of the whole network in an adverse manner especially in cases where the cost is high. To overcome this problem, we propose an adaptive reputation protocol which enables the agents to evaluate the behavior of their neighbors over time and select the most reputable subset of neighbors to share genuine information with. We provide an extensive simulation-based analysis to compare the performance of the proposed method with several other distributed learning strategies. The results show that the proposed method outperforms the other learning strategies and enables the network to have a superior performance especially when the cost of sharing genuine information is high.

1. Introduction

Decentralized systems have attracted much interest in several situations as the information is processed in a distributed and collaborative manner. Following these strategies, a set of spatially distributed agents who are linked to each other form an adaptive network. These nodes or agents of the network have local interactions with other agents which enable them to have cooperation. The performance of these self-organized networks depends on the learning abilities and the localized cooperation of the interconnected nodes. In these types of networks, each agent can communicate and share information with its neighboring nodes. As a result of this cooperation and information sharing, the agents can solve particular tasks or optimization problems (such as estimating unknown parameters, tracking objects, etc.). Several strategies have been proposed for distributed information processing over networks, such as incremental [1–5], consensus [6–9] and diffusion strategies [10–15]. In the incremental strategy, a Hamiltonian cycle has to be determined across the nodes of the network, which is generally an NP-hard task [16]. Therefore, topology changes in the network over time presents a considerable obstacle for incremental methods. On the contrary, it has been shown that among these strategies, diffusion algorithm is robust, scalable, and capable of real-time adaptation and learning. Diffusion strategies have also superior performance and stability compared to consensus methods in data processing applications [14,17].

The diffusion strategies and their performances have been studied extensively in several scenarios [10,14,18]. In most of the prior studies, it is assumed that the agents are genuine, trustworthy and obey certain distributed information sharing protocols which may be a strong limitation in real-world applications [10,17,19,20]. Although sharing genuine information and being cooperative is an essential element for the efficiency of these networks, there are many types of cooperative networks (such as online social networks, websites, and social forums) where some of the agents might disobey the protocols, stop cooperating, or feed others with

* Corresponding author.
E-mail address: sadaf.monajemi@u.nus.edu (S. Monajemi).

http://dx.doi.org/10.1016/j.future.2017.07.023
0167-739X/© 2017 Elsevier B.V. All rights reserved.

Nomenclature

$\mathcal{N}_k$	Neighbors of node k including node k
$d_k(i)$	Scalar measurement of node k at time instant i
$\boldsymbol{x}_k(i)$	Regression vector of node k at time instant i
$\boldsymbol{\omega}_k^o$	Optimum parameter vector of node k at time instant i
$\mathcal{C}_q$	Set of nodes belonging to cluster q
$\mathcal{C}(k)$	The cluster that node k belongs to
$a_{\ell k}(i)$	Weight assigned by node k to the information of node ℓ
$\mathcal{N}_{\mathbb{G}}^{\ell}(i)$	The subset that node ℓ shares its genuine information with
c_k	Cost of sharing genuine information
$p_{k,\mathbb{G}}^i$	Probability of sharing genuine information for agent k
$R_{\ell k}(i)$	Reputation score assigned to node ℓ by node k

malicious and misleading information [21–23]. One of the reasons causing this behavior is the fact that sharing genuine information can be sensitive due to privacy preservation in several situations [24–26]. In some other scenarios, selfish agents might prefer to disobey the rules and share misleading information to minimize their own costs as sharing genuine information can be expensive [27–29]. Moreover, some agents might share misleading information with others in order to obtain an unfair advantage or bias due to their own superior performance. Therefore, it is crucially important to consider the freedom of agents for sharing genuine information in order to analyze the behavior and performance of these networks.

Several research articles in this area are dedicated to incentive mechanisms that encourage the agents to cooperate with their neighbors. Some of these mechanisms are based on pricing strategies where payments are used to reward or punish the agents for their behaviors [30,31]. These mechanisms might be adequate for some settings, but are not appropriate for several cases where the information is free such as online social networks. Moreover, these systems usually require complex monitoring and accounting infrastructures that result in significant computation costs. In other studies, differential services are considered to reward and punish the agents based on their actions. These services can be provided by a network operator in the case of centralized systems [23,32]. However, as there is no central operator in decentralized systems, differential services are provided by the other agents based on their interactions. In reputation mechanisms, a reputation score is assigned to an agent based on its past behavior with the other agents. However, in most of the studies on reputation mechanisms the focus is mainly on practical implementation details [33,34]. In [21], the authors studied a network where the agents provide different services to their neighbors and designed incentive-compatible rating systems. In another work [22], the authors studied the case where the agents are self-interested and try to minimize their own cost and estimation error over a single task network. In this work, the agents are randomly paired with each other and each agent can share information with only one agent at a time. In [23], a centralized rating protocol is formulated where all the agents have the incentive to follow the recommended strategy. In [27], the authors studied energy expenditure of agents in a cooperative network. They proposed a game-theoretic approach to help the agents decide about their activation based on the trade-off between their contribution and energy expenditure over a single task network.

This work differs from the existing methods in the literature in several aspects. First, we study multitask networks where there are several connected clusters of nodes with different objectives. To make it more realistic, we remove any presumption about prior clustering information. Specifically, the agents have no former knowledge regarding the cluster that they or their neighbors belong to. Moreover, we do not assume that the objectives of the clusters are necessarily related. Second, we study the case where sharing genuine information is subject to a cost and each node can decide whether to share genuine or misleading information. To consider this problem in a more generalized way, we allow the nodes to make this decision in a pairwise manner rather than holistic approaches where genuine or false information is sent to all the neighbors. This means that each node can share genuine information with a subset of its neighbors while sending misleading information to the remaining neighbors at each time instant. Third, we define the necessary utility functions for sharing genuine and misleading information to obtain a credibility model for the cooperative network. Credibility equilibrium can be found using the model which determines the probability of sharing genuine information for each agent. It is obvious that a low probability of genuine information sharing results in a degraded global benefit for the cooperative network. Lastly, we propose a reputation approach which allows the agents to evaluate the importance of their neighbors for performing their own estimation task. This is the first time that a reputation protocol has been incorporated in the multitask diffusion strategy. Considering the spontaneous behavior of the nodes and using the reputation scores, each node can select the subset of neighbors to share genuine information with according to the credibility equilibrium. With the help of this dynamic and adaptive protocol, each node would be able to select its most important and trustworthy neighbors to share information with while taking into account its own privacy and cost budgets. Table 1 highlights the differences of our method from the existing works.

The rest of this paper is organized as follows. In Section 2 we introduce our system model and the multitask diffusion adaptation strategy for information processing over the network. In Section 3, we introduce the utility functions and derive the credibility equilibrium. We propose the adaptive reputation protocol for sharing information in Section 4 and provide the simulation results in Section 5. The paper is concluded in Section 6.

2. System model

First we provide a summary of some of the main symbols and notations that are used in this article. Other symbols are defined in the context where they are used:

2.1. Network model

Consider a connected network of N nodes as shown in Fig. 1. As can be seen in this figure, each node k is connected to a number of neighboring nodes represented by $\mathcal{N}_k$. Each agent k of the network has access to the scalar measurements $d_k(i)$ and an $M \times 1$ regression vector $\boldsymbol{x}_k(i)$ with covariance matrix $\boldsymbol{R}_{x,k} = \mathbb{E}\boldsymbol{x}_k(i)\boldsymbol{x}_k^*(i) > 0$ at every time instant i and M represents the dimension of the problem at hand. It is assumed that each node is interested to estimate an $M \times 1$ unknown parameter vector $\boldsymbol{\omega}_k^o$ that is related to the data $\{d_k(i), \boldsymbol{x}_k(i)\}$ via a linear regression model:

$$d_k(i) = \boldsymbol{x}_k^T(i)\boldsymbol{\omega}_k^o + n_k(i), \tag{1}$$

where $n_k(i)$ is the measurement noise of node k at time instant i. To better understand this linear regression model, we present a physical example from [35], where a network of agents are spread over a geographical area observing realizations of an auto-regressive (AR) random process $d_k(i)$ of order M. The AR process observed by agent k satisfies the model:

$$d_k(i) = \sum_{m=1}^{M} \alpha_m d_k(i-m) + n_k(i), \quad k = 1, 2, \ldots N \tag{2}$$