Neurocomputing 175 (2016) 466–472

Contents lists available at ScienceDirect

Neurocomputing

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

Brief Papers

Set-membership filtering for genetic regulatory networks with missing values ☆

Wu Wang [a,b,*], Xiaocheng Liu [a,b], Yurong Li [a,b], Yurong Liu [c,d]

[a] College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, Fujian 350116, China
[b] Fujian Key Lab of Medical Instrument and Pharmaceutical Technology, Fuzhou, Fujian 350002, China
[c] Department of Mathematics, Yangzhou University, Yangzhou 225002, China
[d] Communication Systems and Networks (CSN) Research Group, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Article history:
Received 17 July 2015
Received in revised form
27 September 2015
Accepted 18 October 2015
Communicated by Hongli Dong
Available online 31 October 2015

Keywords:
Missing values
Genetic regulatory networks
Set-membership filtering
Linear matrix inequality

ABSTRACT

The data missing is a common and important problem in the area of genetic regulatory networks (GRNs). A class of discrete-time GRNs with missing values, parameter uncertainties, time delays and molecular noise is considered in this paper. A set-membership filtering method is proposed to estimate the states of the underlying GRNs. Meanwhile, the corresponding problem of set-membership filtering is formulated as finding the set of estimations that belongs to an ellipsoid. The desired filter gains are characterized as the solution of a set of linear matrix inequalities. Finally, a numerical example is provided to illustrate the effectiveness of the proposed method, which shows that by using the proposed set-membership filtering algorithm, the concentrations of mRNA and protein could be estimated accurately.

1. Introduction

Genetic regulatory networks (GRNs) comprise the interactions of mRNAs and proteins in a biological cell. These mRNAs and proteins are the products of gene expression (a process including the transcription, RNA splicing, translation of a protein). In the GRNs, the levels of gene expression are regulated such that the function and structure of proteins (and or genes) are controlled and hence special functions of the cell are realized. Generally, gene expression levels can be measured on a genomic scale by DNA microarray technology. However, DNA microarray data often contain missing values, because of various reasons, such as dust or scratch on the slide, inappropriate thresholds in preprocessing, insufficient resolution of the microarray, experimental errors during the laboratory processes or image corruption, etc. [1]. These missing values severely affect the construction of GRNs, the performance of GRNs model and statistical analysis [2]. In addition, due to the slow transcription, the random births and deaths of individual molecules, and the environment fluctuations, the GRNs are of high complexity with time delays and molecular noise. Thus, the measured gene expression levels might be far different from the actual ones which leads that not all information about the gene expression levels are available in the network measurement. Hence, the consideration of filtering technique is very significant for subsequent analysis [3–7].

There are quite a few papers in the literature which are devoted to the study of the filtering problem of GRNs during the past decades. For instance, in [8], the adaptive filtering approach has been developed to estimate unknown delayed GRNs based on an adaptive synchronization setting. In [9], the intrinsic fluctuation was described as a state-dependent stochastic process, and the extrinsic noise was modeled as an arbitrary signal with bounded energy, then a class of Markov jump linear filter was designed to estimate the true concentrations of mRNA and protein. In [10], the authors investigated the robust filtering problem for a class of linear GRNs with parameter uncertainties, which was assumed to reside in a polytypic region. Furthermore, other typical filtering approaches, such as Kalman filtering [11] and H_∞ filtering [12–20] are also applied to the study of GRNs. However, the Kalman filtering approach requires the exact statistical knowledge of external noise, but the actual external noise may not be fully known. The H_∞ filtering approach does not take into account the estimation error bound. To overcome these drawbacks, the reference [2] concerned with the set-values filtering (or called set-membership filtering) for a class of discrete time-delay GRNs with time-varying parameters and nonstochastic noise (bounded external noise). Although these methods have been successfully

☆This work was partially supported by the Natural Science Foundation of Fujian Province, China (Grant no. 2013J01227) and the Foundation of Science and Technology of Fujian Province, China (Grant nos. 2013Y4003 and JK2014001)

* Corresponding author at: College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, Fujian 350116, China.
E-mail address: wangwu@fzu.edu.cn (W. Wang).

http://dx.doi.org/10.1016/j.neucom.2015.10.083
0925-2312/

implemented in the GRNs, all of them are mainly designed for GRNs without considering the missing values.

To deal with the missing values, there are mainly three kinds of methods up to date, i.e., repeating the experiment, ignoring objects containing missing value, and estimating the missing values. The first approach is expensive and time consuming. Using the second approach, too much useful information are lost and may bias the results if the remaining cases do not represent the entire sample appropriately. The third approach utilizes different algorithms to estimate the missing values, it is more suitable than the former two approaches [21,1,22]. In particular, in [21], a weighted K-nearest neighbor method is proposed to estimate the missing values. Based on the Bayesian principal component analysis (BPCA), the author introduces a new missing value as the simultaneous of probabilistic model and latent variables in [1]. To overcome the shortcoming of BPCA that is incapable of handling local structure of data, a local approach bicluster-based BPCA is developed to capture the local structure via biclustering in [23]. In this paper, we will introduce the set-membership filtering to deal with the missing values as the first attempt. The stability issue for GRNs has been intensively exploited [3,24,25], therefore it is not necessary to discuss the stability problem for GRNs in this paper.

In this paper, we consider a class of discrete-time GRNs model, including time-varying parameters, time delays, bounded external noise and missing values. We propose a set-membership filtering algorithm to compute the filter gains for the presented discrete-time GRNs with missing values. With the set-membership filtering [26,27], the actual estimation is a set in the state space, which contains the true state of the GRNs by assuming a hardbound on external noise. Hence, different from the Kalman filtering and H_∞ filtering, the set-membership filtering problem aims to find the small characterization of the feasible set of the state, rather than providing the most possible states under some optimality criteria [28]. The solution of the filter is obtained by solving a recursive linear matrix inequality (LMI). By using the set-membership filtering method, we provide a deterministic upper bound for the estimation error and minimize the quadratic estimation error at each time step. A numerical example is given to illustrate the effectiveness of our theoretical results.

The remainder of this paper is organized as follows. In Section 2, we consider a class of discrete-time GRNs, and propose the set-membership filtering framework. The set-membership filtering issue is investigated, and two critical lemmas are used to derive our main theoretical results in Section 3. In Section 4, a numerical example is given to illustrate the effectiveness of our theoretical results. Some concluding remarks are provided in Section 5.

2. Problem formulation

A class of discrete-time genetic regulatory networks (GRNs) can be described as [2,29,30]:

$$\begin{cases} \boldsymbol{M}_{k+1} = \boldsymbol{A}_k \boldsymbol{M}_k + \boldsymbol{B}_k f(\boldsymbol{N}_{k-\sigma}) + Z \\ \boldsymbol{N}_{k+1} = \boldsymbol{C}_k \boldsymbol{N}_k + \boldsymbol{D}_k \boldsymbol{M}_{k-\sigma} \end{cases} \tag{1}$$

where $\boldsymbol{M}_k \in \mathbb{R}^n$ and $\boldsymbol{N}_k \in \mathbb{R}^n$ are respectively the concentrations of mRNA and protein. $\boldsymbol{A}_k$ and $\boldsymbol{C}_k$ represent the degradation rates of mRNA and protein, respectively. $\boldsymbol{D}_k$ is the translation rate. σ denotes the translation and feedback regulation delay. $\boldsymbol{B}_k$ is the coupling coefficient of the genetic network. Z is the bounded constant and denotes the dimensionless transcriptional rate. In addition, $f(\cdot) \in \mathbb{R}$ is a monotonic function of the Hill form, which represents the feedback regulation of the protein on the transcription. Here, $f_i(x) = (x/\beta_i)^{H_i}/(1+(x/\beta_i)^{H_i})$, where H_i is the Hill coefficient and β_i is a positive constant. Clearly, it satisfies [29–31]:

$$|f_i(u) - f_i(v)| \le l_i |u - v| \tag{2}$$

$$\forall u, v \in R, i = 1, 2, \cdots, n \tag{3}$$

Denote $\boldsymbol{M}^*$ and $\boldsymbol{N}^*$ to be the equilibrium point of system (1). Let $\overline{\boldsymbol{M}}_k \triangleq \boldsymbol{M}_k - \boldsymbol{M}^*$ and $\overline{\boldsymbol{N}}_k \triangleq \boldsymbol{N}_k - \boldsymbol{N}^*$. Then, the intended equilibrium point $\boldsymbol{M}^*$ and $\boldsymbol{N}^*$ will shift to the origin. Thus, system (1) can be rewritten as

$$\begin{cases} \overline{\boldsymbol{M}}_{k+1} = \boldsymbol{A}_k \overline{\boldsymbol{M}}_k + \boldsymbol{B}_k g(\overline{\boldsymbol{N}}_{k-\sigma}) \\ \overline{\boldsymbol{N}}_{k+1} = \boldsymbol{C}_k \overline{\boldsymbol{N}}_k + \boldsymbol{D}_k \overline{\boldsymbol{M}}_{k-\sigma} \end{cases} \tag{4}$$

where $g(\boldsymbol{N}_k) \triangleq f(\boldsymbol{N}_k + \boldsymbol{N}^*) - f(\boldsymbol{N}^*)$. See [2] for more details about the model (4).

In practice, the actual GRNs might be influenced by the dynamic reaction of the networks, time delays, and molecular noise. Based on the above analysis, a more general discrete-time GRNs is considered in this paper:

$$\begin{cases} \boldsymbol{x}_{k+1} = \tilde{\boldsymbol{A}}_k \boldsymbol{x}_k + \tilde{\boldsymbol{D}}_k \boldsymbol{x}_{k-\sigma} + \tilde{\boldsymbol{B}}_k g(\boldsymbol{x}_{k-\sigma}) + \boldsymbol{W}_k \boldsymbol{v}_k \\ \boldsymbol{y}_k = \boldsymbol{E}_k \boldsymbol{x}_k + \boldsymbol{F}_k \boldsymbol{v}_k \end{cases} \tag{5}$$

where $\boldsymbol{x}_k \triangleq \left[\overline{\boldsymbol{M}}_k^T \overline{\boldsymbol{N}}_k^T\right]^T$, $\boldsymbol{y}_k \in \mathbb{R}^m$ is the sampled output, $\boldsymbol{v}_k$ is the external noise, and $\boldsymbol{W}_k$, $\boldsymbol{E}_k$ and $\boldsymbol{F}_k$ are constant matrices. Furthermore,

$$\tilde{\boldsymbol{A}}_k \triangleq \begin{bmatrix} \boldsymbol{A}_k & 0 \\ 0 & \boldsymbol{C}_k \end{bmatrix}, \quad \tilde{\boldsymbol{D}}_k \triangleq \begin{bmatrix} 0 & 0 \\ \boldsymbol{D}_k & 0 \end{bmatrix},$$

$$\tilde{\boldsymbol{B}}_k \triangleq \begin{bmatrix} 0 & \boldsymbol{B}_k \\ 0 & 0 \end{bmatrix},$$

and $\boldsymbol{A}_k, \boldsymbol{B}_k, \boldsymbol{C}_k, \boldsymbol{D}_k$ are time varying parameters indicating the influence of the dynamic biological response.

In this paper, we assume that the external noise $\boldsymbol{v}_k$ is unknown but bounded by

$$\boldsymbol{v}_k^T \boldsymbol{S}_k^T \boldsymbol{v}_k \le 1 \tag{6}$$

where $\boldsymbol{S}_k = \boldsymbol{S}_k^T > 0$ is a known matrix with compatible dimension.

Considering the missing values, we can rewrite the sampled output $\boldsymbol{y}_k$ as

$$\overline{\boldsymbol{y}}_k = \delta_k \boldsymbol{y}_k + (1 - \delta_k) \boldsymbol{y}_{k-1} \tag{7}$$

where δ_k is the indicator function defined by

$$\delta_k \triangleq \begin{cases} 1, & \text{if } \boldsymbol{y}_k \text{ is available,} \\ 0, & \text{if } \boldsymbol{y}_k \text{ is missing.} \end{cases} \tag{8}$$

Remark 1. It is the first attempt to consider the missing values problem in set-membership filtering. The missing values are described as a binary switching sequence which is viewed as a Bernoulli distributed white sequence taking on the values of 1 and 0. Although such the representation for missing values is simple, it is useful for many practical systems, such as the GRNs.

Our aim in this paper is to estimate the state of the system (5) with sampling data (7). For this, the following filter is considered:

$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{G}}_k \hat{\boldsymbol{x}}_k + \hat{\boldsymbol{L}}_k \overline{\boldsymbol{y}}_k \tag{9}$$

where $\hat{\boldsymbol{G}}_k$ and $\hat{\boldsymbol{L}}_k$ are the filter parameters to be determined.

The initial state $\boldsymbol{x}_0$ and the initial estimate state $\hat{\boldsymbol{x}}_0$ belong to a given ellipsoid

$$(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0)^T \boldsymbol{P}_0^{-1} (\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0) \le 1 \tag{10}$$

where $\boldsymbol{P}_0 = \boldsymbol{P}_0^T > 0$ is a known matrix.

However, the current state $\boldsymbol{x}_{k+1}$ does not in an optimal state estimation ellipsoid. Hence, we apply the convex optimization approach to determine an optimal ellipsoid. $\boldsymbol{P}_{k+1}$ can be obtained