Automatica 47 (2011) 702–710

Contents lists available at ScienceDirect

Automatica

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

Identification of Boolean control networks[☆]

Daizhan Cheng[*], Yin Zhao

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, PR China

ARTICLE INFO

Article history:
Available online 26 February 2011

Keywords:
Boolean control network
Identification
Controllability
Observability
Network graph

ABSTRACT

In this paper the identification of Boolean control networks is addressed. First, necessary and sufficient conditions are obtained for the identification of state equation from input-state data. Then a necessary and sufficient condition for a controllable Boolean network to be observable is presented. Based on these two results, a necessary and sufficient condition for the identification from input–output data is achieved. To practically identify the model, a numerical algorithm is proposed. Two particular cases: (i) identification of systems with a known network graph; (ii) identification of a higher order Boolean network, are also investigated. Finally, the approximate identification for large size networks is explored.

1. Introduction

The Boolean network was firstly proposed by Kauffman in his initial work (Kauffman, 1969). Then it becomes a hot topic in Systems Biology, Physics, as well as Systems Science, because investigations showed that it is a proper tool for modeling complex and nonlinear biological systems (Albert & Othmer, 2003; Aldana, 2003; Drossel, Mihaljev, & Greil, 2005; Huang, 2002; Huang & Ingber, 2000; Kauffman, 1993, 1995). When part of the network, such as genetic regularity network in cellular network, acts as controls, the Boolean control network becomes a proper model for analyzing the behavior of the network (Akutsu, Hayashida, Ching, & Ng, 2007; Data, Choudhary, Bittner, & Dougherty, 2004; Pal, Datta, Bittner, & Dougherty, 2006). Some practically useful control techniques have also been developed (Bansal, Belcastro, Ambesi-Impiombato, & Bernardo, 2007; Cantone et al., 2009).

Recently, a new technique, called the semi-tensor product of matrices and the matrix expression of logic (Cheng, 2007), has been used to the study of Boolean (control) networks, including the analysis of Boolean networks (Cheng, 2009; Cheng & Qi, 2010) and the control design of Boolean control networks (Cheng, Li, & Qi, 2010; Cheng & Qi, 2009). The basic idea is to convert the logical dynamic model of a Boolean (control) network into a discrete time dynamic system by using the vector expression of logic.

Identification is an important topic for Boolean (control) networks. For instance, though a real cellular network is huge, we may be interested in a particular function of the network. Hence we may have some input–output data related to this particular function, and we can use these data to identify a network, which can be used to describe this particular function of the original complicated network. Identification of the network transition mappings has been considered in Akutsu, Miyano, and Kuhara (2000); Liang, Fuhrman, and Somogyi (1998) and Nam, Seo, and Kim (2006). Using the semi-tensor product approach, a new method for identifying Boolean networks (without control) via observed data was proposed in Cheng, Qi, and Li (submitted for publication). To identify a logical control system directly seems difficult because it can hardly be parameterized. Fortunately, in our new approach, the dynamics of a Boolean (control) network can be converted into a discrete time linear (correspondingly, bilinear) systems, and hence they can be described via transition matrix. Therefore, the system identification problem becomes the problem of identifying the related matrices, which makes the identification realizable.

The rest of this paper is organized as follows: Section 2 provides a framework for the problem. Section 3 considers the identification of the state equations via input-state data. In Section 4 we first prove a necessary and sufficient condition for the observability of Boolean control network. Using this result, a necessary and sufficient condition is obtained for the identifiability of a Boolean control network. Section 5 gives a numerical algorithm for the identification using input–output data. Two particular cases: (i) when the input-state transfer graph is known; (ii) higher order Boolean control networks, are also investigated. Approximate identification of large size Boolean networks is considered in Section 6. Finally, a brief summary is given in Section 7.

2. Problem formulation

Denote by $\mathcal{D} := \{0, 1\}$ the domain of a logical variable. A logical variable $x(t)$ is assumed to evolve over it. That is, $x(t) \in \mathcal{D}$,

☆ This work is supported partly by SIC07010201 and NNSF 60736022, 60221301 and 61074114 of China. This paper was not presented at any IFAC meeting. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +86 10 62651445; fax: +86 10 62587343.
E-mail addresses: dcheng@iss.ac.cn (D. Cheng), zhaoyin@amss.ac.cn (Y. Zhao).

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.01.083

$t = 0, 1, \ldots$. The dynamics of a Boolean control network is described as

$$\begin{cases} x_1(t+1) = f_1(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t)) \\ x_2(t+1) = f_2(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t)) \\ \vdots \\ x_n(t+1) = f_n(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t)) \end{cases} \quad (1)$$
$$y_j(t) = h_j(x_1(t), \ldots, x_n(t)), \quad j = 1, \ldots, p,$$

where $x_i(t), u_k(t), y_j(t) \in \mathcal{D}, i = 1, \ldots, n, k = 1, \ldots, m, j = 1, \ldots, p$, are states, inputs (controls), and outputs respectively, and $f_i : \mathcal{D}^{n+m} \to \mathcal{D}^n$, $h_j : \mathcal{D}^n \to \mathcal{D}^p$ are logical functions.

The identification problem is stated as follows.

Definition 1. Assume that we have a Boolean control network with dynamic structure (1). The identification problem involves finding the functions $f_i, i = 1, \ldots, n$, and $h_j, j = 1, \ldots, p$, via certain input–output data $\{U(0), U(1), \ldots\}$, $\{Y(0), Y(1), \ldots\}$. The identification problem is said to be solvable if f_i and h_j can be uniquely determined by using designed inputs $\{U(0), U(1), \ldots\}$.

Note that here we use the following notation:

$$X(t) := (x_1(t), x_2(t), \ldots, x_n(t)),$$
$$Y(t) := (y_1(t), y_2(t), \ldots, y_p(t)),$$
$$U(t) := (u_1(t), u_2(t), \ldots, u_m(t)).$$

We need some other notations:

- δ_n^k is the kth column of the identity matrix I_n.
- $\Delta_n := \{\delta_n^1, \ldots, \delta_n^n\}$. For compactness, $\Delta := \Delta_2$.
- $\mathrm{Col}_i(A)$ is the ith column of a matrix A, the set of all the columns of A is denoted by $\mathrm{Col}(A)$.
- $\mathrm{Blk}_i(A)$ is the ith $n \times n$ block of an $n \times mn$ matrix A.
- A matrix $L \in M_{n\times m}$ is called a logical matrix if its columns, $\mathrm{Col}(M) \subset \Delta_n$.
- The set of $n \times m$ logical functions is denoted by $\mathcal{L}_{n\times m}$.
- Let $L \in \mathcal{L}_{n\times m}$. Then

  $L = [\delta_n^{i_1}, \delta_n^{i_2}, \ldots, \delta_n^{i_m}]$.

  It is briefly denoted as

  $L = \delta_n[i_1, i_2, \ldots, i_m]$.
- Let $A = (a_{ij}), B = (b_{ij}) \in M_{m\times n}$ be real matrices. (i) $A \geq 0$ ($A > 0$) means $a_{ij} \geq 0$ ($a_{ij} > 0$), $\forall i, j$; (ii) $A \geq B$ ($A > B$) means $a_{ij} \geq b_{ij}$ ($a_{ij} > b_{ij}$) $\forall i, j$.
- Let $A = (a_{ij}) \in M_{m\times n}$, if $a_{ij} \in \mathcal{D}$, $\forall i, j$, then A is called a Boolean matrix. The set of $m \times n$ Boolean matrices is denoted by $\mathcal{B}_{m\times n}$.
- $W_{[n,m]}$ is a swap matrix. We refer to Cheng (2007) for its definition and properties.

Throughout this paper the matrix product is assumed to be semi-tensor product as $A \ltimes B$, and the symbol "$\ltimes$" is omitted in most places. We refer to Cheng (2007) for details.

By identifying $1 \sim \delta_2^1$ and $0 \sim \delta_2^2$, refer to Cheng (2009), the algebraic form of (1) is

$$\begin{cases} x(t+1) = Lu(t)x(t) \\ y(t) = Hx(t), \end{cases} \quad (2)$$

where $x = \ltimes_{i=1}^n x_i$, $u = \ltimes_{k=1}^m u_k$, and $y = \ltimes_{j=1}^p y_j$, $L \in \mathcal{L}_{2^n\times 2^{m+n}}$, and $H \in \mathcal{L}_{2^p\times 2^n}$.

Note that $X = (x_1, \ldots, x_n)$ and $x = \ltimes_{i=1}^n x_i$ are in one-to-one correspondence and can be easily converted from one form to the other easily. Similarly, Y and y (U and u) are in one-to-one correspondence. Therefore, (1) and (2) are equivalent, and hence identifying $f_i, i = 1, \ldots, n$ and $h_j, j = 1, \ldots, m$ is equivalent to identifying L and H.

Remark 2. Assume that we have a coordinate transformation $z = Tx$, where $T \in \mathcal{L}_{2^n\times 2^n}$. Then the algebraic form (2) becomes

$$\begin{cases} z(t+1) = \tilde{L}u(t)z(t) \\ y(t) = \tilde{H}z(t), \end{cases} \quad (3)$$

where

$$\tilde{L} = TL(I_{2^m} \otimes T^T); \qquad \tilde{H} = HT^T. \quad (4)$$

It is obvious that (L, H) and $(\tilde{L}, \tilde{H})$ are not distinguishable by any input–output data. So precisely speaking, we should say the pair (L, H) is identifiable up to a coordinate transformation.

Keep this in mind, what we are going to identify is the equivalence class, but a particular (L, H).

3. Identification via input-state data

In this section, we assume that the state is measurable. Alternatively, we may assume that

A1 $p = n$ and $y_i(t) = x_i(t), i = 1, \ldots, n$.

We recall the following definition.

Definition 3 (*Cheng & Qi, 2009*)**.** System (1) is controllable, if for any initial state $X_0 = (x_1(0), \ldots, x_n(0)) \in \mathcal{D}^n$ and destination state X_d, there is a sequence of controls $U_0, U_1, \ldots$, where $U_t = (u_1(t), \ldots, u_m(t))$, such that the trajectory $X(t, X_0, U)$ satisfies $X(0, X_0, U) = X_0$, and $X(s, X_0, U) = X_d$, for some $s > 0$.

For the identifiability from input-state data we have the following result.

Theorem 4. *System (1) is input-state identifiable, iff the system is controllable.*

Proof. (Sufficiency) Since the system is controllable, for any $x_d = \delta_{2^n}^i$, we can find a set of controls such that at time s $x(s) = \delta_{2^n}^i$. Now to identify $\mathrm{Col}_k(L)$, we can find a unique pair (i, j) such that

$$\delta_{2^m}^j \delta_{2^n}^i = \delta_{2^{n+m}}^k.$$

In fact, $i = k\%2^m$ and $j = \frac{k-i}{2^m} + 1$. Hence, we can first choose control $u_0, u_1, \ldots$ to drive the system to $x(s) = \delta_{2^n}^i$, at certain moment $s > 0$, and then choose $u(s) = \delta_{2^m}^j$. It follows that

$$\mathrm{Col}_k(L) = x(s+1).$$

(Necessity) Split $\tilde{L} = LW_{[2^n,2^m]}$ into 2^n equal blocks as $\tilde{L} = [\tilde{L}_1, \tilde{L}_2, \ldots, \tilde{L}_{2^n}]$. Now assume that $\delta_{2^n}^i$ is not reachable, then the columns of $\mathrm{Blk}_i(\tilde{L}) = \tilde{L}_i$ can never be shown in the state $x(t), t = 1, 2, \ldots$. So $\mathrm{Blk}_i(\tilde{L})$ is not identifiable. □

One sees now the controllability is the key for identifiability. We refer to Cheng and Qi (2009) for the necessary and sufficient condition for controllability. A recent alternative necessary and sufficient condition might be more convenient. We introduce it as follows.

Define a $2^{n+m} \times 2^{n+m}$ matrix J as

$$J = \left.\begin{bmatrix} L \\ \vdots \\ L \end{bmatrix}\right\} 2^m,$$

which is called the incidence matrix of the input-state transfer graph of system (1), where L comes from (2). Denote the top