

使用模型检测解决概率布尔网络优化控制

郭宗豪 魏 欧

(南京航空航天大学计算机科学与技术学院 南京 210016)

摘 要 系统生物学期望对复杂生物系统建立一个真实的、可计算的模型,以便于以系统的角度去理解生物系统的演变过程。在系统生物学中,一个重要的主题是通过外部的干预控制发展关于基因调控网络的控制理论,以作为未来基因治疗技术。目前,布尔网络及其扩展的概率布尔网络已经被广泛用于对基因调控网络进行建模。在控制问题的研究中,概率布尔控制网络的状态迁移本质上构成一条有限状态空间的离散时间马尔科夫决策过程。依据马尔科夫决策过程的理论,通过概率模型检测方法解决网络中有限范围优化控制问题和无限范围优化控制问题。针对带有随机干扰且上下文相关的概率布尔控制网络,使用概率模型检测器 PRISM 对其进行形式化建模,然后将两类优化控制问题描述为相应的时序逻辑公式,最后通过模型检测寻找出最优解。实验结果表明,提出的方法可以有效地用于生物网络的分析和优化控制。

关键词 基因调控网络,概率布尔网络,优化控制,概率模型检测

中图分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.05.035

Optimal Control of Probabilistic Boolean Networks Using Model Checking

GUO Zong-hao WEI Ou

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract Systems biology expected to construct a realistic and computational model of complex biology systems that aims at system-level understanding of biological systems. One of the significant topics in the field of system biology is that the control theory of gene regulatory networks (GRNs) is developed by applying external intervention control for gene theory technologies in the future. At present, Boolean networks and extended probabilistic Boolean networks have been used as the model of GRNs widely. In the research of control problem, the state transition of probabilistic Boolean control networks essentially forms a finite-state and discrete-time Markov decision processes (MDP). According to MDP theory, finite-horizon optimal control problem and infinite-horizon optimal control problem can be solved by using probabilistic model checking. For content-sensitive probabilistic Boolean control networks with random perturbation, probabilistic model checker PRISM is used to model formally. Then two kinds of optimal control problems are expressed by temporal logic. Finally, the optimal solution is found via model checking. The results indicate that this proposed approach can be used for analysis and optimal control of biology networks effectively.

Keywords Gene regulatory networks, Probabilistic Boolean network, Optimal control, Probabilistic model checking

1 引言

系统生物学的目标是对复杂生物系统构建一个可计算的数学模型,以用于从系统层面上理解分子间的相互作用。随着系统生物学的发展,如今各种形式体系已经被用于对生物网络进行建模和分析,如动态贝叶斯网络^[1]、Petri 网^[2]、布尔网络(Boolean Network, BN)^[3]和概率布尔网络(Probabilistic Boolean Network, PBN)^[4]等离散状态模型以及常微分或偏微分方程和线性模型等连续状态模型。不同的模型对生物网络特征的刻画具有不同的精准度。然而,具有高非线性

和高维度的模型对于网络的分析和控制是复杂的。BN 是一种离散、确定的模型,由一些二进制值的结点和表示结点之间动态交互关系的布尔函数构成。虽然布尔网络的结构比较简单,但是它不仅能够很好地模拟网络控制中出现的非线性动力学行为,而且能够反映网络运行过程中出现的复杂动态行为。因此,布尔网络模型成为研究基因调控网络的一种重要模型并且经常被用于控制问题的研究。随后,考虑在生物系统建模的过程中存在不确定因素,PBN 被提出。概率布尔网络是一种离散、随机的模型,在布尔网络的基础上加以概率变化形成。本质上,它表示多个布尔网络的集合,每个时刻根据

到稿日期:2016-04-05 返修日期:2016-07-11 本文受国家自然科学基金项目(61170043),国家重点基础研究发展计划(973 计划)(2014CB744904)资助。

郭宗豪(1992—),男,硕士生,主要研究领域为概率模型检测,E-mail:guozhwork@163.com;魏 欧(1974—),男,博士,副教授,主要研究领域为形式化方法、软件自动验证,E-mail:owei@nuaa.edu.cn。

给定的概率分布来决定哪个布尔网络将会被选择作为当前网络的演变规则。

一种生物网络的控制方式是通过外部的控制输入,改变网络的动态行为从而避免网络演变到一些不期望的状态,这些状态可能与疾病相关,如癌症。在癌症治疗的案例中,控制输入可能是辐射(radiation)、化疗(chemo-therapy)等。控制的目的在于防止恶性细胞(malignant cell)的增长。现实中由于药物治疗的成本比较昂贵,因此考虑如何在较少的成本下改善病情是很重要的。当考虑成本因素时,人们更加关注优化控制问题。因此,无论是BN还是PBN,优化控制问题作为一个关键的问题都早已经被提出。

关于优化控制问题的复杂性,布尔网络的控制和概率布尔网络的控制都属于NP-hard问题。根据控制时间的不同,优化控制问题可被分为有限范围的优化控制^[5]和无限范围的优化控制^[6]。对于优化控制问题的解决,已有一些传统的解决方案,例如利用矩阵半张量积、概率动态规划方法。概率模型检测(Probabilistic Model Checking)是一种基于模型的自动化验证技术。首先,使用离散时间的马尔科夫链(Discrete-Time Markov Chains, DTMC)、马尔科夫决策过程(Markov Decision Processes, MDP)等概率模型表示现实系统。另外,利用概率时序逻辑公式规约系统各种行为属性。模型检测算法将在概率模型上验证概率时序逻辑规约的属性并且得到定性或者定量的结果。

本文首先使用概率模型检测器PRISM提供的建模语言对概率布尔控制网络(Probabilistic Boolean Control Network, PBCN)进行描述;然后利用概率计算树逻辑(Probabilistic Computation Tree Logic, PCTL)^[7]公式分别表达概率布尔网络中有限范围的优化控制问题和无限范围的优化控制问题。PRISM将构建的模型和时序逻辑公式作为输入,并通过模型检测算法自动地计算出公式的结果。

本文第2节介绍布尔网络和概率布尔网络的相关内容,同时也解释了概率模型检测的相关原理;第3节从DTMC的角度分析概率布尔网络的动态行为以及从MDP的角度探讨概率布尔网络的控制;第4节描述有限范围和无限范围的优化控制问题,并且使用概率模型检测作为其解决方式;第5节和第6节以细胞凋亡网络为例,说明构建其概率布尔控制网络模型的过程,并且通过实验对两类优化控制问题进行求解和分析;第7节讨论了相关工作;最后总结全文。

2 预备知识

2.1 布尔网络

布尔网络^[3](BN)是一种重要的模型,用于研究基因调控网络的动态行为。假设基因的状态可被划分为两种:“0”和“1”,其中,“0”代表基因未被表达,“1”代表基因被表达。因此,布尔网络中每个结点可代表一个基因的状态。结点间的交互规则通过布尔函数进行表示。布尔网络可以表示成 $B=(G, F)$ 。 $G=(V, E)$ 是一个有向图,其中 $V=\{x_1, x_2, x_3, \dots, x_n\}$ 表示网络中结点的集合, E 表示结点间的关联关系。每一个结点 $x_i \in \{0, 1\}$ 代表基因 i 的布尔状态值。 $F=\{f_1, f_2, f_3, \dots, f_n\}$ 是一个布尔函数集合,表示所有相互关联的

基因之间调控交互的规则。每一个布尔函数 $f_i: \{0, 1\}^{k(i)} \rightarrow \{0, 1\} (1 \leq i \leq n)$ 实现一个状态的映射,决定基因 i 下一时刻的状态。 $k(i)$ 是每个布尔函数 f_i 输入结点的个数。若用 $x_i(t)$ 表示基因 i 在 t 时刻的布尔状态值,则 $x_i(t+1)=f_i(x_{i_1}(t), x_{i_2}(t), \dots, x_{i_{k(i)}}(t))$ 。另外,如果使用向量 $X(t)=\{x_1(t), x_2(t), \dots, x_n(t)\}$ 表示所有基因在 t 时刻的布尔状态值,同时定义了网络在 t 时刻的状态,那么网络下一时刻的状态 $X(t+1)=F(x_1(t), x_2(t), \dots, x_n(t))$ 。

根据所有基因状态同步和异步的更新模式,布尔网络又可以分为同步布尔网络^[8]和异步布尔网络^[9]。

2.2 概率布尔网络

概率布尔网络^[4]是在布尔网络的基础上为每个基因提供多个布尔函数,用于处理来自数据和模型选择方面的不确定性。与布尔网络相比,布尔函数向量 $F=\{F_1, F_2, F_3, \dots, F_n\}$ 中的每一个 $F_i=\{f_i^{(1)}, f_i^{(2)}, \dots, f_i^{(l(i))}\}$ 代表一个由 $l(i)$ 个布尔函数组成的集合。每一个 $f_i^{(j)}$ 决定基因 i 下一个时刻可能的状态,并且被选择的概率为 $c_i^{(j)}$ 。

当每个基因 i 对候选布尔函数的选择相互独立时,概率布尔网络本质上可以看作是由 N 个布尔网络构成。每一个构成的布尔网络可以分别用网络迁移函数 $f_1, f_2, f_3, \dots, f_N$ 表示。在每个离散时刻,概率布尔网络按照确定的概率分布随机地选择其中一个构成的布尔网络 $f_k=(f_k^{(1)}, f_k^{(2)}, \dots, f_k^{(n)})$ 作为网络的演变规则,其中, $1 \leq k \leq N, 1 \leq k_i \leq l(i)$ 。每个布尔网络 f_k 被选择的概率为:

$$P_k = \sum_{i=1}^n c_{k_i}^{(i)} \quad (1)$$

根据所有构成的布尔网络之间不同的切换频率,概率布尔网络可以分为瞬时随机的概率布尔网络(Instantaneously Random PBN)^[4]和上下文相关的概率布尔网络(Context-Sensitive PBN, CS-PBN)^[10]。在瞬时随机的概率布尔网络中,所有构成的布尔网络在每个时刻按照概率分布可以相互地切换。相对而言,为了更好地表示生物网络的稳定性,CS-PBN以一定的概率控制所有构成的布尔网络之间相互切换的发生。在CS-PBN中,一个二进制的转换开关 s 和转换概率 q 被添加到瞬时随机的概率布尔网络中。转换概率 q 是一个系统参数。当转换开关 s 以概率 q 处于状态1时,PBN按照确定的概率分布随机地选择一个布尔网络作为演变规则;否则,保持当前布尔网络不变。

在基因调控网络的演变过程中,每个基因的状态还会受到外界的随机干扰(perturbations)^[11],即状态的值不受布尔函数的控制,非确定地跳转到0或者1。通常情况下,每个基因受到干扰是相互独立的并且干扰发生的概率非常小。使用二进制的变量 $peri$ 代表基因 i 的干扰事件, p 代表干扰发生的概率。当 $peri=1$ 时,表示干扰发生。

2.3 概率模型检测

概率模型检测^[12]是一种自动的形式化验证技术,能够对具有随机行为的系统进行定性和定量的分析。概率模型检测需要两个输入:一个对系统进行描述的形式化建模语言和一个对系统的行为属性进行规约的时序逻辑。所描述出的系统通常被一些概率模型所表示,如DTMC和MDP等。关于系

如果固定决策者在所有状态下对控制输入的选择,即每个状态仅有一个 action,那么 PBCN 的状态迁移将转变为 DTMC。

4 优化控制问题

概率布尔网络的一个核心控制问题是优化控制问题。优化控制问题^[14]是指在特定的时间区间内选择优化的控制输入策略,使网络尽可能处于期望状态,同时最小化总的控制成本。因此,可以通过最小化期望总控制成本来解决优化控制问题。在生物领域中,特定的时间区间代表着一个治疗干预阶段,控制输入策略意味着一个治疗干预策略,控制成本反映了在不同治疗时刻所需付出的代价,如人力、设施等,具体成本值的设定应取决于领域专家。

4.1 有限范围和无限范围的优化控制

根据优化控制问题的定义^[14],假设 PBCN 有 n 个基因和 m 个控制输入且控制时长为 M 。 $t=0,1,2,\dots,M-1$ 表示 PBCN 处于不同的控制阶段。 $X(t)=(x_1(t),x_2(t),\dots,x_n(t))$ 描述 PBCN 所有基因的状态, $u(t)=(u_1(t),u_2(t),\dots,u_m(t))$ 表示 t 时刻所有控制输入的状态。因此,控制输入策略 $\pi=\{u(0),u(1),\dots,u(M-1)\}$,其中 $u(t):X(t)\rightarrow\{0,1\}^m$ 。另外,控制成本函数 $C(X(t),u(t))$ 表示在 t 时刻,当网络状态为 $X(t)$ 且控制输入为 $u(t)$ 时所需付出的成本。 $C(X(M))$ 是与终端状态 $X(M)$ 相关联的终端成本。根据以上定义,网络从初始状态 $X(0)$ 出发,在控制时长 M 内的期望总控制成本为:

$$E\left[\sum_{t=0}^{M-1} C(X(t),u(t))+C(X(M))\mid X(0)\right] \quad (4)$$

为了寻找最小的期望总控制成本,使用 $V(X(0))$ 表示网络在控制时长 M 内最终获得的优化控制成本。

$$V(X(0))=\min E\left[\sum_{t=0}^{M-1} C(X(t),u(t))+C(X(M))\right] \quad (5)$$

在优化控制成本公式中,当控制时长 M 为一个确定的整数值时,代表有限范围的优化控制;当控制时长 M 的值趋于无穷大或者不受限制时,代表无限范围的优化控制。

对于有限范围优化控制成本的问题,通常可以使用向后推导的概率动态规划方法^[14]进行求解。通过对终端状态 $X(M)$ 设置不同的终端成本值 $C(X(M))$,可以指出哪些状态是期望的状态,哪些状态是不期望的状态。终端成本值较小的状态表示越期望到达。对于控制成本函数而言,如果状态 $X(t)$ 的期望程度越高,那么 $C(X(t),u(t))$ 的值越小。因此,使期望总控制成本最小的优化控制输入策略,同时也意味着使得网络在演变过程中尽可能地处于期望的状态。

在无限范围的优化控制问题中,由于控制范围是无限的,因此没有终端状态。所以,无法利用向后推导的概率动态规划法计算出优化控制成本,否则将导致控制成本无限大。对于无限范围控制成本的计算,需要设定网络的初始状态 $X(0)$ 和目标状态 $X(M)$ 。因此,无限范围的优化控制问题是指选择优化控制输入策略,使网络从初始状态出发,沿着每条可行路径行进,当满足目标状态时,最小化期望总控制成本。所以对于无限范围的优化控制,最终获得的优化控制成本为:

$$V(X(0))=\min \lim_{M\rightarrow\infty} E\left[\sum_{t=0}^{M-1} C(X(t),u(t))\right] \quad (6)$$

4.2 使用时序逻辑说明优化控制问题

为了使用概率模型检测方法解决以上两类优化控制问题,需要将 PBCN 的状态演变转换为一个相应的 MDP 模型。两类控制问题分别对应着有限范围马尔科夫决策过程和无限范围马尔科夫决策过程。根据优化控制问题的定义,首先使用 reward 结构对 MDP 模型进行扩展来指定状态回报值,用于表示控制成本函数 $C(X(t),u(t))$ 和终端成本 $C(X(M))$ 。然后,使用基于 reward 扩展的 PCTL 表示对有限范围的优化控制成本和无限范围的优化控制成本的计算。最后,通过模型检测算法进行自动求解。

对于解决有限范围的优化控制问题,使用时序逻辑公式 " $R_{\min}=?[C\leq T+1]$ " 表示在 MDP 上计算最小累积回报,目的是得到在 T 时间范围内,在控制输入策略的作用下网络演变所产生的优化控制成本。公式中使用 " $T+1$ " 而不是 " T " 的原因是,模型检测算法在累积回报的计算过程中不包含对终端时刻状态回报的累加。根据累计回报计算过程的定义^[13],计算结果是结合迁移概率对每个时刻状态回报值和每个状态迁移回报值进行累加。然而,由于仅仅指定了状态回报用于表示控制成本函数和终端成本,因此公式 " $R_{\min}=?[C\leq T+1]$ " 的计算过程意味着是在有限控制时长内对每个时刻的控制成本函数值以及终端成本的累加,满足有限范围优化控制成本的计算要求。

对于解决无限范围的优化控制问题,使用时序逻辑公式 " $R_{\min}=?[F \text{ prop}]$ " 表示在 MDP 上计算最小可达性回报,目的是得到从初始状态出发,在控制输入策略的作用下,沿着每一条可执行的路径,当到达目标状态 prop 时,网络演变所产生的优化控制成本。根据可达性回报计算过程的定义^[13],计算结果是结合迁移概率对每个时刻状态回报值和每个状态迁移回报值进行累加,直到到达满足 prop 的状态。同样,由于仅仅指定了状态回报值用于表示控制成本函数和终端成本,因此公式 " $R_{\min}=?[F \text{ prop}]$ " 的计算过程意味着是在网络到达目标状态之前对每个时刻的控制成本函数值的累加,满足无限范围优化控制成本的计算要求。

5 案例研究:细胞凋亡网络的优化控制

5.1 细胞凋亡网络模型

本节选取一个细胞凋亡网络(apoptosis network)^[15]作为优化控制问题的研究对象。首先在 PRISM 中对简化的细胞凋亡网络建立一个带有干扰的且上下文相关的概率布尔控制网络模型。细胞凋亡网络的概率布尔控制网络模型的动态方程如式(7)所示:

$$\begin{aligned} X_1(t+1) &= \begin{cases} u(t) \wedge (\neg X_2(t)), & \text{with prob } 0.5 \\ X_1(t), & \text{with prob } 0.5 \end{cases} \\ X_2(t+1) &= \begin{cases} (\neg X_1(t)) \wedge X_3(t), & \text{with prob } 0.5 \\ X_2(t), & \text{with prob } 0.5 \end{cases} \\ X_3(t+1) &= \begin{cases} u(t) \vee X_2(t), & \text{with prob } 0.5 \\ X_3(t), & \text{with prob } 0.5 \end{cases} \end{aligned} \quad (7)$$

在模型中, X_1 代表 IAP 的浓度水平, X_2 代表 C3a 的浓度水平, X_3 代表 C8a 的浓度水平, u 代表 TNF 的浓度水平。 $X_i=1$ 表示高浓度水平, $X_i=0$ 表示低浓度水平。通过概率

扩展,每个基因的状态都具有同步更新和异步更新两种方式。另外,设置所有结点的候选布尔函数的被选择概率为 0.5。

5.2 在 PRISM 中建模

在式(7)所示的概率布尔控制网络(PBCN)的基础上,为了考虑网络的稳定性和外界随机干扰的影响,将转换开关 s 和干扰变量 $peri$ 添加到概率布尔控制网络中,最终对细胞凋亡网络建立一个带有随机干扰且上下文相关的概率布尔控制网络模型。该网络模型^[16]在 PRISM 中的代码描述如下。

```
1. mdp
2. constint c=0;
3. formula f11= $u * (1-x_2)$ ;
4. formula f12= $x_1$ ;
5. formula f21= $(1-x_1) * x_3$ ;
6. formula f22= $x_2$ ;
7. formula f31= $u+x_2-u * x_2$ ;
8. formula f32= $x_3$ ;
9. module switch
10.  $s:[0..1]$  init 1;
11.  $[csPBN]c=0 \rightarrow 0.3:(s'=1)+0.7:(s'=0)$ ;
12. endmodule
13. module pert1
14.  $per1:[0..1]$  init 0;
15.  $[csPBN]c=0 \rightarrow 0.9:(per1'=0)+0.1:(per1'=1)$ ;
16. endmodule
17. module BN1
18.  $x1:[0..1]$  init 1;
19.  $d1:[0..1]$  init 0;
20.  $[csPBN]c=0 \& s=1 \& per1=0 \rightarrow 0.5:(x1'=f11) \& (d1'=0) + 0.5:(x1'=f12) \& (d1'=1)$ ;
21.  $[csPBN]c=0 \& s=0 \& per1=0 \& d1=0 \rightarrow (x1'=f11)$ ;
22.  $[csPBN]c=0 \& s=0 \& per1=0 \& d1=1 \rightarrow (x1'=f12)$ ;
23.  $[csPBN]c=0 \& per1=1 \rightarrow (x1'=0)$ ;
24.  $[csPBN]c=0 \& per1=1 \rightarrow (x1'=1)$ ;
25. endmodule
26. module input
27.  $u:[0..1]$  init 0;
28.  $[csPBN]c=0 \rightarrow (u'=0)$ ;
29.  $[csPBN]c=0 \rightarrow (u'=1)$ ;
30. endmodule
```

以上仅给出了与转换开关 s 、布尔结点 x_1 、随机干扰变量 $per1$ 以及布尔控制结点 u 有关的 PRISM 建模代码,与布尔结点 x_2 和 x_3 以及随机干扰变量 $per2$ 和 $per3$ 的相关代码与之类似。在建模代码中,从第 3—8 行,每一个 formula 定义了结点的一个布尔函数。第 10—11 行代码说明转换开关 s 可以取二进制的值并且设定转换概率 $q=0.3$ 。第 14—15 行代码说明干扰变量 $per1$ 可以取二进制值并且设定干扰概率 $p=0.1$ 。第 23—24 行代码说明当干扰事件发生($per1=1$)时,布尔结点 x_1 下一个时刻非确定地变为 1 或 0。第 28—29 行的代码说明布尔控制结点 u 的取值是非确定的。

除此以外,使用 PRISM 提供的 reward 结构对网络的状态设置不同的状态回报值,用于说明控制成本函数和终端成本。Reward 结构的代码描述如下。

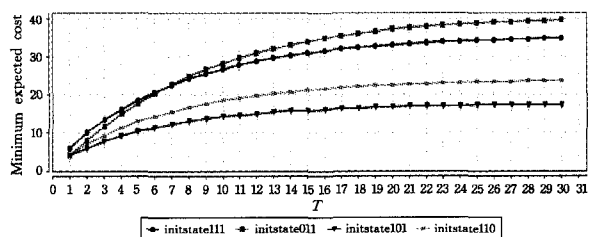
```
1. rewards "cost"
2.  $x_1=0 \& x_2=0 \& x_3=0 \& u=0:0$ ;
3.  $x_1=0 \& x_2=0 \& x_3=0 \& u=1:1$ ;
4.  $x_1+x_2+x_3=1 \& u=0:2$ ;
5.  $x_1+x_2+x_3=1 \& u=1:3$ ;
6.  $x_1+x_2+x_3=2 \& u=0:4$ ;
7.  $x_1+x_2+x_3=2 \& u=1:5$ ;
8.  $x_1=1 \& x_2=1 \& x_3=1 \& u=0:6$ ;
9.  $x_1=1 \& x_2=1 \& x_3=1 \& u=1:7$ ;
10. endrewards
```

在以上 PRISM 代码中,对网络所有状态设置了不同的状态回报值。网络状态的变化范围为 $[0,0,0]$ 到 $[1,1,1]$ 。第 2—9 行的代码说明了当网络处于所有不同的状态时,布尔控制结点 $u=0$ 和 $u=1$ 分别所对应的控制成本函数值。同时,第 2,4,6,8 行的代码也说明了所有不同状态的终端成本。在实际中,状态回报值必须要能够捕捉到干预的成本以及对不同状态的相对偏好。根据控制成本函数值的设置,可以看出当 $u=1$ 即应用干预时,与 $u=0$ 相比,需要付出额外的干预成本,本文设置了干预成本为 1。通过比较每个状态的终端成本,表明了状态 $[0,0,0]$ 是最期望到达的状态,状态 $[1,1,1]$ 是最不期望到达的状态。通过不同状态的期望程度,表明了我们对细胞凋亡网络控制的目的是通过外部的干预抑制每个分子的浓度水平,从而使得每个分子的浓度水平变为 0。

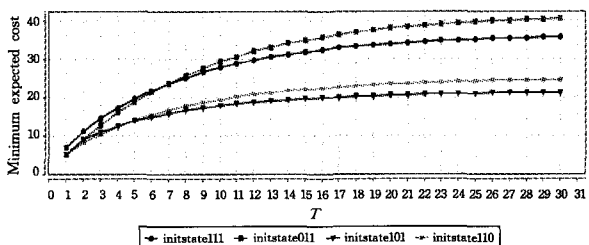
6 实验结果和分析

6.1 实验结果

在上文构建的模型中,对于两类优化控制成本的计算,分别使用 PCTL 公式 $R\{“cost”\} \min = ?[C \leq T+1]$ 和 $R\{“cost”\} \min = ?[F x_1=0 \& x_2=0 \& x_3=0]$ 表示计算有限范围内的和无限范围内的优化控制成本。在有限范围的优化控制中, T 为一个整数变量, C 代表累计时间。设定 T 变化范围为 1~30,且增加的步长为 1。在无限范围的优化控制中,指定网络的目标状态是 $[0,0,0]$ 。实验环境是 Intel(R) core (TM) 2quad CPUi5-4200U@1.60GHz 2.30GHz,操作系统是 64bits-ubuntu12.04。实验运行结果如图 3 和表 1 所示。



(a) 布尔控制结点初始值 $u=0$



(b) 布尔控制结点初始值 $u=1$

图 3 公式 $R\{“cost”\} \min = ?[C \leq T+1]$ 的计算结果

表 1 公式 $R\{\text{"cost"}\} \min = ? [F_{x_1=0} \ \& \ x_2=0 \ \& \ x_3=0]$ 的计算结果

	u 初始值	初始状态($x_1x_2x_3$)			
		[1,1,1]	[0,1,1]	[1,0,1]	[1,1,0]
优化	0	35.43	40.58	17.67	24.09
控制成本	1	36.43	41.58	21.45	25.09
计算时间/ s	0	1.75	1.80	1.83	1.78
	1	1.74	1.76	1.79	1.74
消耗内存/ MB	0	83.63	82.05	80.08	80.69
	1	79	79.23	79.56	80.09
目标状态($x_1x_2x_3$)		[0,0,0]			

6.2 分析实验结果

图 5 的结果展示了在布尔控制结点初始值 $u=0$ 和 $u=1$ 的情况下,网络分别从初始状态[1,1,1],[0,1,1],[1,0,1]以及[1,1,0]出发,在 $T=30$ 的最大控制时长内,有限范围优化控制成本随时间增加的变化情况。在相同的控制时间内,当网络初始状态为[1,0,1]时,优化控制成本始终是最小的。该结果表明,与其他 3 种初始状态相比,网络从状态[1,0,1]出发,在有限范围控制策略的作用下,能够更容易地向较期望的状态演变,即意味着浓度水平的变化有更高的概率是被抑制的。随着控制时长的增加,优化控制成本的增长趋势都会逐渐减缓并且趋于水平,这表明经过较长时间的控制后,网络能够以较大的概率保持在最期望的状态[0,0,0],即意味着所有分子浓度水平都保持被抑制的可能性很高。

表 1 列出了在布尔控制结点初始值 $u=0$ 和 $u=1$ 的情况下,网络分别从 4 种初始状态出发,在无限范围控制策略的作用下,达到目标状态[0,0,0]时所需要的无限范围优化控制成本、CPU 计算时间以及内存消耗。无限范围优化控制成本的计算结果不会无限大,是由于存在随机干扰使得网络状态具有遍历性。从优化控制成本的比较可以得出,当网络初始状态为[1,0,1]且布尔控制结点初始值 $u=0$ 时,优化控制成本最小。该结果表明,与其他 3 种初始状态相比,网络从状态[1,0,1]出发,在无限范围优化控制策略作用下,能够更容易地到达目标状态,即意味着在无限的控制时长内,分子浓度水平的变化有更高的概率是被抑制的。另外,实验数据表明,使用 PRISM 计算无限范围内的优化控制成本比较快速,且消耗的内存大小在可接受的范围内。

关于优化控制策略,目前在 MDP 模型上对可达性回报进行计算时,对应的无限范围优化控制策略将可以在 PRISM 中被获取。但是在 MDP 模型上对累积回报进行计算时,PRISM 无法显示地获取对应的有限范围优化控制策略。

7 相关工作

对于传统的布尔网络的研究,许多领域的不同方法已经被使用。在数学领域中,矩阵半张量积(STP)^[17]的方法作为一种有力的工具可被用于对布尔网络中的大部分问题进行分析求解。该方法主要是通过矩阵表示布尔变量和布尔函数,将逻辑动态系统转化为标准的离散动态系统,然后通过矩阵的张量积运算求解相关问题。但是该方法在对问题的求解过程中,计算比较复杂。在计算机领域中,符号化模型检测器 NUSMV 也已经被用于对布尔网络和布尔控制网络进行建模和分析^[18],并且使用计算树逻辑(CTL)公式寻找出有效的控制策略。但是 NUSMV 不能够描述 PBN。

在本文的案例研究中,最终对实际的生物网络建立的是

一个带有干扰的且上下文相关的概率布尔控制网络模型,该模型考虑了真实基因调控网络中存在的各种因素。与 Kobayashi, Hiraishi^[19]所建立的模型相比,所提模型通过添加随机干扰的影响而具有更好的完备性。在文献[5]中,对于优化控制问题的研究,Ching 等设定控制次数为一种硬性的约束条件,并且将最小化终端状态和期望状态之间的距离作为优化控制的目标。代替最小化控制成本,Liu 等^[20]研究如何选择控制输入使得在特定的时间内以最大的概率到达期望的网络状态。

结束语 本文详细地描述了如何使用概率模型检测方法解决概率布尔网络中有限范围和无限范围的优化控制问题。概率布尔网络的控制可以看作一个马尔科夫决策过程。原本用于求解优化控制成本的数学公式可以转换为基于 reward 结构扩展的 PCTL 公式。因此,优化控制成本的求解等同于在 MDP 模型上对回报属性的测量,控制输入策略等同于在 MDP 模型上对每个时刻不同状态上的 action 选择。在 PRISM 中,输入对 PBN 进行描述的形式化建模语言和待求解问题的时序逻辑公式,结果将会通过概率模型检测算法自动地被计算出来。概率模型检测方法的优势是可以作为一个简单的、方便的工具用于生物网络的分析和控制,并且在生物网络模型中很容易考虑一些复杂的情况,如外界的随机干扰或者两个以上的布尔控制结点等。

参 考 文 献

[1] LI P,ZHANG C,PERKINS E J,et al. Comparison of probabilistic Boolean network and dynamic Bayesian network approaches for inferring gene regulatory networks[J]. BMC Bioinformatics, 2007,8(1):S13.

[2] STEGGLES L J,BANKS R,SHAW O,et al. Qualitatively modelling and analysing genetic regulatory networks: a Petri net approach[J]. Bioinformatics,2007,23(3):336-343.

[3] KOBAYASHI K,HIRAIISHI K. Optimal control of gene regulatory networks with effectiveness of multiple drugs: a Boolean network approach [J]. Biomed Research International, 2013, 2013(23):10233-10240.

[4] SHMULEVICH I,DOUGHERTY E R,KIM S,et al. Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks[J]. Bioinformatics,2002,18(2):261-274.

[5] CHING W K,ZHANG S Q,JIAO Y,et al. Optimal finite-horizon control for probabilistic Boolean networks with hard constraints[C]//International Symposium on Optimization & Systems Biology. Beijing,China,2007:21-28.

[6] PAL R,DATTA A,DOUGHERTY E R. Optimal infinite-horizon control for probabilistic Boolean networks[J]. IEEE Transactions on Signal Processing,2006,54(6):2375-2387.

[7] CIESINSK F,GRÖßER M. On Probabilistic Computation Tree Logic[M] // Validation of Stochastic Systems. Berlin Heidelberg:Springer,2004:147-188.

[8] ZHANG H,WANG X,LIN X. Synchronization of Boolean Networks with Different Update Schemes[J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2014,11(5): 965-972.

同的标题根据相似度聚类为话题,然后利用结构信息建立话题层次的贝叶斯网络,最后通过 Chu-Liu/Edmonds 算法计算最优树,获得最终的层次话题结构。实验结果证明了本文方法的有效性。在今后的研究中,将进一步研究如何提高聚类准确度以及尝试将跨语言的在线百科合并到已有的层次话题结构中。

参考文献

- [1] TED P, SIDDHARTH P, JASON M. Wordnet: Similarity-measuring the relatedness of concepts[C]// HLT-NAACL 2004. Association for Computational Linguistics, 2004: 38-41.
- [2] WU F, WELD D S. Automatically refining the wikipedia infobox ontology[C]// Proceedings of the 17th International Conference on World Wide Web. ACM, 2008: 635-644.
- [3] WU F, HOFFMANN V, WELD D S. Information extraction from wikipedia: Moving down the long tail[C]// Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2008: 731-739.
- [4] LI R, BAO S H, YU Y, et al. Towards effective browsing of large scale social annotations[C]// Proceedings of the 16th International Conference on World Wide Web. ACM, 2007: 943-952.
- [5] NASTASE V, STRUBE M. Decoding wikipedia categories for knowledge acquisition[C]// AAAI 2008: 1219-1224.
- [6] DMBTL G, MIJJB T. Hierarchical topic models and the nested chinese restaurant process[J]. Advances in Neural Information Processing Systems, 2004, 16: 17.
- [7] MIMNO D, LI W, MCCALLUM A. Mixtures of hierarchical topics with pachinko allocation[C]// Proceedings of the 24th ICML. ACM, 2007: 633-640.
- [8] ZAVITSANON E, PALIOURAS G, VOUIROS G A. Non-parametric estimation of topic hierarchies from texts with hierarchical dirichlet processes[J]. The Journal of Machine Learning Research, 2011, 12: 2749-2775.
- [9] CHUANG S L, CHIEN L F. A practical web-based approach to generating topic hierarchy for text segments[C]// Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management. ACM, 2004: 127-136.
- [10] TANG J, LEUNG F, LUO Q, et al. Towards ontology learning from folksonomies[C]// IJCAI 2009: 2089-2094.
- [11] ZHU X W, MING Z Y, ZHU X Y, et al. Topic hierarchy construction for the organization of multi-source user generated contents[C]// Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2013: 233-242.
- [12] NAVIGLI R, VELARDI P, FARALLI S. A graph-based algorithm for inducing lexical taxonomies from scratch[C]// IJCAI 2011: 1872-1877.
- [13] MONGE A E, ELKAN C, et al. The field matching problem: algorithms and applications[C]// Proceedings of the 2nd ACM SIGKDD. 1996: 267-270.
- [14] CHU Y J, LIU T H. On shortest arborescence of a directed graph[J]. Scientia Sinica, 1965, 14(10): 1396.
- [15] VINH N X, EPPS J, BAILEY J. Information theoretic measures for clusterings comparison: is a correction for chance necessary? [C]// Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009: 1073-1080.
- [16] LIU X, SONG Y, LIU S, et al. Automatic taxonomy construction from keywords[C]// KDD. 2012: 1433-1441.
- [17] WANG C, DANILEVSKY M, DESAI N, et al. A phrase mining framework for recursive construction of a topical hierarchy[C]// KDD. New York, NY, USA, ACM, 2013: 437-445.
- [18] TOURNIER L, CHAVES M. Uncovering operational interactions in genetic networks using asynchronous boolean dynamics [J]. Journal of Theoretical Biology, 2009, 260(2): 196-209.
- [19] KOBAYASHI K, HIRAISHI K. Symbolic approach to verification and control of deterministic/probabilistic Boolean networks [J]. IET Systems Biology, 2012, 6(6): 215-222.
- [20] FENG J, YAO J, PENG C. Singular Boolean networks: Semi-tensor product approach[J]. Science China (Information Sciences), 2013, 56(11): 1-14.
- [21] LANGMEAD C J, JHA S K. Symbolic approaches for finding control strategies in Boolean Networks[J]. Journal of Bioinformatics & Computational Biology, 2009, 7(2): 307-319.
- [22] KOBAYASHI K, HIRAISHI K. Verification and Optimal Control of Context-Sensitive Probabilistic Boolean Networks Using Model Checking and Polynomial Optimization[J]. The Scientific World Journal, 2014, 2014(2): 295-318.
- [23] LIU Q, GUO X, ZHOU T. Optimal control for probabilistic Boolean networks[J]. IET Systems Biology, 2010, 4(2): 99-107.
- [24] SHMULEVICH I, DOUGHERTY E R, ZHANG W. Gene perturbation and intervention in probabilistic Boolean networks[J]. Bioinformatics, 2002, 18(10): 1319-1331.
- [25] PARKER D. Implementation of Symbolic Model Checking for Probabilistic Systems[D]. University of Birmingham, 2002.
- [26] KWIATKOWSKA M, NORMAN G, PARKER D. Stochastic Model Checking[M]. REMKE A, STOELINGA M, eds. Berlin Heidelberg: Springer, 2014.
- [27] DATTA A, CHOUDHARY A, BITTNER M L, et al. External Control in Markovian Genetic Regulatory Networks[J]. Journal of Chemical Physics, 2003, 119(11): 4569-4575.

(上接第198页)