国家自然科学基金面上项目（61472045，61573067）
内蒙古自治区高等学校科学技术研究项目（NJZY132）

布尔网络控制问题的研究

高博 著

北京邮电大学出版社
www.buptpress.com

国家自然科学基金面上项目(61472045,61573067)
内蒙古自治区高等学校科学技术研究项目(NJZY132)

布尔网络控制问题的研究

高 博 著

北京邮电大学出版社
www.buptpress.com

内 容 简 介

布尔网络是系统生物学的重要模型，它因为具有简单的结构和抽象的描述方法，引起了基因研究人员的广泛关注。目前布尔网络模型已经应用于生物学、博弈理论、信息科学等众多领域。本书介绍了布尔网络的概念、网络状态演化的规律和近年来以半张量积为工具在布尔网络控制领域取得的成果：首先介绍了系统生物学的发展历程和研究现状、基因调控网络的概念、半张量积的概念与数学性质和使用半张量积进行布尔网络状态空间描述的方法；其次，介绍了布尔网络吸引子的生物学意义，研究了吸引子单点、多点调控的方法，吸引子的鲁棒性评测的问题和具有脉冲扰动的布尔网络状态描述问题；最后，介绍了将布尔网络模型应用于序列密码器件研究所取得的成果。

图书在版编目(CIP)数据

布尔网络控制问题的研究/高博著. --北京 ：北京邮电大学出版社，2016.4
ISBN 978-7-5635-4736-4

Ⅰ.①布… Ⅱ.①高… Ⅲ.①生物模型－研究 Ⅳ.①Q141

中国版本图书馆 CIP 数据核字(2016)第 073434 号

书　　名：布尔网络控制问题的研究
著作责任者：高　博　著
责任编辑：徐振华　孙宏颖
出版发行：北京邮电大学出版社
社　　址：北京市海淀区西土城路 10 号(邮编：100876)
发 行 部：电话：010-62282185　传真：010-62283578
E-mail：publish@bupt.edu.cn
经　　销：各地新华书店
印　　刷：北京九州迅驰传媒文化有限公司
开　　本：720 mm×1 000 mm　1/16
印　　张：7.5
字　　数：150 千字
版　　次：2016 年 4 月第 1 版　2016 年 4 月第 1 次印刷

ISBN 978-7-5635-4736-4　　定　价：20.00 元

前　言

随着人类基因组测序技术的发展，产生了各种新兴的生物学分支，基因组学、蛋白质组学、代谢组学等新学科的出现表明了21世纪的生物学研究正在从分子生物学走向系统生物学。

在系统生物学研究中，研究者将实验数据进行整合，从而建立数学模型，然后通过实验不断地改进模型并调整参数，最终使模型具有实际意义并能够预测生物系统的表观形态变化。系统生物学是一门多学科交叉的新型学科，需要具备数学、物理学、化学、生物学、医学及控制理论和计算机科学等领域的知识。

在系统生物学的众多数学模型中，布尔网络因其简单而抽象的描述方法，引起了基因研究人员的广泛关注。本书介绍了布尔网络的概念、网络状态演化的一般规律，并介绍了近年来以半张量积为工具在布尔网络控制和状态描述等相关领域取得的成果。本书的内容分为以下5个部分。

第1章介绍了系统生物学的发展历程和研究现状、基因调控网络的概念和主要用途，并分析了基因调控网络的主要特征；然后介绍了几种常见的基因调控网络的数学模型；最后重点介绍了布尔网络模型和近年来以半张量积为工具在布尔网络研究方面取得的成果。

第2章介绍了半张量积的概念和数学性质，将半张量积作为工具将逻辑运算转变成矩阵的乘法运算的基本做法；然后介绍了布尔网络状态空间的描述方法和吸引子的计算方法；最后进行了实验仿真。

第3章介绍了布尔网络吸引子的生物学意义与相关研究现状，提出吸引子调控的方法，通过评估吸引子调控的可达性，然后求解控制序列的方式构建了吸引子调控的数学模型；在研究基因单点调控的基础上，研究了多基因点调控的问题；最后讨论了吸引子的鲁棒性评测的问题。

第4章介绍了具有脉冲扰动的布尔网络状态描述问题，首先给出具有脉冲扰

动的布尔网络的状态空间描述方法和有效扰动输入的判别方法；然后提出了未知脉冲序列的状态描述方法和具有周期特征的脉冲扰动的状态分析方法。

第5章介绍了序列密码的基本概念和序列密码设计中的常用器件；在此基础上，重点介绍NFSR器件的概念、分类和研究现状；然后介绍了将布尔网络模型应用于NFSR器件研究所取得的成果并进行了试验仿真。

目　录

第 1 章　布尔网络简介 …… 1

1.1　系统生物学 …… 1

1.1.1　系统生物学发展综述 …… 1

1.1.2　系统生物学的概念 …… 2

1.2　基因调控网络 …… 3

1.2.1　基因调控的概念 …… 3

1.2.2　基因调控网络的特性 …… 4

1.2.3　基因调控网络的数学模型 …… 5

1.3　布尔网络研究综述 …… 9

1.3.1　布尔网络的状态描述 …… 9

1.3.2　基于半张量积的布尔网络研究 …… 10

1.4　本书符号说明 …… 11

参考文献 …… 13

第 2 章　半张量积与布尔网络状态分析 …… 22

2.1　半张量积的概念 …… 22

2.2　半张量积的计算方法 …… 23

2.3　逻辑运算的矩阵表达 …… 25

2.4　布尔网络的状态描述 …… 26

2.4.1　L 矩阵的求解 …… 27

2.4.2　求解布尔网络的吸引子 …… 30

2.4.3　布尔网络状态空间描述 …… 31

2.5　实验仿真 …… 33

参考文献 …… 35

第 3 章　布尔网络的吸引子调控问题 …… 38

3.1　吸引子调控的应用背景 …… 38
3.2　单点调控问题 …… 39
3.2.1　有效输入的判别 …… 39
3.2.2　吸引子调控 …… 40
3.2.3　实验仿真 …… 42
3.3　多点调控问题 …… 48
3.4　布尔网络吸引子鲁棒性的评测 …… 50
3.4.1　研究背景 …… 50
3.4.2　模型与方法 …… 51
3.4.3　实验仿真 …… 53
参考文献 …… 55

第 4 章　具有扰动的布尔网络状态研究 …… 59

4.1　具有脉冲扰动的布尔网络的应用背景 …… 59
4.2　脉冲扰动的布尔网络的状态描述 …… 60
4.2.1　确定输入下状态分析 …… 60
4.2.2　未知扰动序列的状态分析 …… 61
4.2.3　实验仿真 …… 64
4.3　周期脉冲扰动的布尔网络的状态分析 …… 65
4.3.1　无效输入的判别 …… 66
4.3.2　周期输入的状态空间描述 …… 67
4.3.3　实验仿真 …… 69
参考文献 …… 76

第 5 章　布尔网络模型在序列密码研究中的应用 …… 79

5.1　密码学基础 …… 79
5.2　序列密码概述 …… 80
5.2.1　序列密码的设计 …… 81
5.2.2　序列密码的器件 …… 83
5.3　NFSR 简介 …… 84
5.4　基于布尔网络模型的 NFSR 周期研究 …… 86
5.4.1　NFSR 状态空间分析 …… 87
5.4.2　周期输入下状态空间的描述 …… 88

5.4.3　级联 NFSR 的周期研究 …… 91
5.4.4　NFSR 评测矩阵 …… 95
5.5　实验仿真 …… 98
5.5.1　Grain 型算法的周期分析 …… 98
5.5.2　NFSR 串联的周期分析 …… 102
参考文献 …… 105

后记 …… 111

第1章 布尔网络简介

1.1 系统生物学

系统生物学是现代科学的重要组成部分，也是人们认识生物现象的重要工具。系统生物学在整合已有实验数据的基础上，建立数学模型，并通过对模型的不断优化预测生物系统表观形态（或内部结构）的变化。本节首先介绍系统生物学的发展历史，然后介绍系统生物学的基本概念。

1.1.1 系统生物学发展综述

随着人类基因组测序技术的发展，产生了各种新兴的生物学分支——基因组学、蛋白质组学、代谢组学等，这些新学科的出现表明了21世纪的生物学研究正在从分子生物学走向系统生物学。许多生物学家都认为系统生物学将成为医学和生物学发展的核心驱动力。

在系统生物学研究中，研究者将实验数据进行整合，从而建立数学模型，然后通过实验不断地改进模型并调整参数，最终使模型具有实际意义并能够预测生物系统的表观形态变化（或者生物系统内部功能结构的变化）。系统生物学是一门多学科交叉的新型学科，需要具备数学、物理学、化学、生物学、医学及控制理论和计算机科学等领域的知识。系统生物学研究也是一门面向动态生物系统的，以假设驱动的，进行全局化、定量化研究的学科。

2000年，系统生物学的开创者Leroy Hood教授，在美国西雅图建立了世界上首个系统生物学研究所（Institute for Systems Biolgoy，ISB）。之后不久，日本的Hiroakitano在东京也建立了系统生物学研究所（SBI）。哈佛大学于2005年成立了系统生物学系，之后系统生物学在美国得到了迅速的发展，目前全美共有12个系统生物学研究中心。2003年3月，《科学》杂志出版了“系统生物学”专刊；之后，

自然出版集团和欧洲分子生物组织合作创立了《分子系统生物学》(Molecular Systems Biology),该杂志第一年的影响因子就达到 8,之后数年其影响因子不断增长,到 2008 年已经达到 12。

2003 年,上海交通大学与中国科学院联合成立了我国首个系统生物学研究所,此后国内的相关研究也陆续展开。多家科研院校相继成立系统生物学研究所,一些国外的论文和专著也相继在国内出版,国内生物领域的学术期刊也开始发表系统生物学的文章,其中,杨胜利院士发表的《系统生物学进展的综述》引起了广泛的关注。

1.1.2 系统生物学的概念

系统生物学是以实验数据与计算结果为基础来研究生物系统的学科,它对生物的系统性(生物的、遗传的、化学的)进行研究,检测基因、蛋白质以及信息通路的各项表征和反应,然后通过对数据的整合,最终建立数学模型,通过模型进一步描述生物系统的结构和它对各种扰动的反应。系统生物学的研究是生物系统内部所有的组成部分(这些组成部分包括基因、mRNA、蛋白质等)和在一定环境下这些组成部分之间的相互关系;以数学模型的方式定量地描述生物特征,并预测一定条件下生物体可能的表型、功能或行为。系统生物学从细胞、组织、器官和生物体等多个维度上,对生物系统的结构、功能进行研究,分析生物系统各部分之间的相互作用,通过计算的方式解释生物的功能、表型和行为,在此基础上,试图预测生物系统在功能、表型和行为方面可能的变化。

20 世纪 50 年代,DNA 双螺旋结构的提出标志着人类进入了分子生物学的时代,随着基因组学、蛋白质组学等各种新兴学科的出现和高通量检测工具的出现,人们逐渐认识到要从系统的视角去认识生物体。系统生物学的基础是将基因数字化,基因数字化信息分为两大类:第一类是编码蛋白质的基因;第二类是控制基因行为和细胞定向分化的调控网络。系统生物学研究的关键不是其生物系统各部分物质,而是这些物质之间的相互作用(或者各部分之间的相互关系)。这样,生物系统就被看作是一个信息流的过程,系统生物学就是要研究这种信息流运行的规律。

数据整合将每次实验的结果进行累加、分析,然后提一些假设,这些假设可以用来辨识生物系统的组成成分和结构。系统生物学要的整合是过层面的,从基因到细胞、到组织、到个体的各个层次都存在整合。数据整合是系统生物学的重要组成部分。简单的整合是把不同类型的数据资源存储到数据库里,进行数据归一化、表格级联、查询和统一的展示。进一步的书籍整合能够以生物学知识为指导,对来自于不同角度和不同层次的数据进行关联,采用模式识别和数据挖掘等方法,分析出有价值的相互关系、假设和趋势描述。

数学建模和计算机仿真是系统生物学研究的最终目标。通过建立接近生物系

统演化规律的数学模型，对生物体进行定量的描述，并预测生物的功能、表型和行为。通过数学模型的方式，以可控、可操作和可重复的方式描述系统各部分之间的相互作用。然后再把生物体实验得到的结果和模型仿真得到的结果进行对比，通过不断地与实验数据的对比，改进模型（或者相关参数），然后再实验，再对比，逐渐迭代，从而完成建模的工作。从短期来看，模型和仿真可以加快实验速度和减少实验成本；从长期来看，模拟和预测生物体对各种扰动的反应，能够帮助研究者选择最优的调控方案。计算机建模和分析已经在生物学和医学的许多领域有了应用，例如，细胞周期的分叉分析、代谢分析、生物振荡回路稳态的比较研究等。在新药的研发过程中，传统方法是基于动物细胞、器官或组织的药物筛选，这种方法虽然有效，但耗时长、花费大、药物作用的机制不明确。采用计算机建模和分析选定疾病的分子靶标后，根据靶标的结构特征和性质，利用计算机构建与受体活性部位能很好契合的模型，从药物吸收、分布、代谢、排泄和毒性等方面进行仿真研究，已成为新药研发的重要工具。

1.2　基因调控网络

本节介绍系统生物学的重要研究内容：基因调控网络。作为一种常用的分析模型，基因调控网络描述了调控因子调控基因表达的过程，基因调控网络可以用于描述 DNA、RNA、蛋白质和基因结构，并模拟各个基因之间的相互作用关系和演化趋势。本节首先介绍基因调控网络的概念，然后分析基因调控网络的主要特征，最后介绍几种常见的基因调控网络的数学模型。

1.2.1　基因调控的概念

基因的表达与调控问题是系统生物学研究的一个重要问题，基因的调控与表达主要指储存遗传信息的基因经过一系列转化，逐步表现其生物特征的整个过程。在整个转化的过程中，包括基因的活化、转录、加工、翻译及翻译的调控网络的形成等几个步骤，其中最关键的调节过程发生在基因转录的过程中。

基因调控网络描述了调控因子如何调控基因表达的过程，这个网络是由 DNA、mRNA、蛋白质、其他小分子，以及它们之间的相互作用关系所构成的。从网络中各个组成部分的相互作用来看，基因调控网络可以表示为一个由节点（调控元素）和边（调控作用）组成的有向图。

在基因调控网络中，整个网络的表达模式是局部调控反应的结果。在细胞环境中，基因调控的产物可以影响其他基因的表达和基因间的相互作用，这种制约的关系就构成了一种基因调控网络。基因之间调控关系的实质是遗传信息的交流、

能量的交换和物质的转移。这些交流、交换和转移使得生物体能够适应外界的环境，维持个体的生长、发育和繁殖。在生命活动中，基因调控表现为生物体内控制基因表达的机制，形成了核酸、蛋白质等物质的内部和它们之间相互的作用。

在实际研究中，有 3 种典型的基因调控网络：①代谢网络，这种网络体现了细胞内各种代谢底物与产物之间的化学反应链；②蛋白质网络，这种网络展示了细胞内各种蛋白质之间的相互作用关系，各种信号传导过程中蛋白复合物的形成；③基因转录调控网络，这种网络映射了所有基因之间一种抽象的相互作用关系，即一个基因的表达水平对其他基因表达的影响。

基因调控网络模型利用基因表达和调控所产生的数据，使用数据挖掘和模式识别等方法，反向分析和挖掘基因、蛋白质和大分子等各种物质之间的调控和被调控关系，采用复杂系统的观点，从基因（或者蛋白质）之间的相互关系的角度分析网络拓扑结构、作用机理和功能信息。基因调控网络需要研究大量生物基因表达的共同特征，是研究基因的共同表达、相互关联和相互作用的重要工具。

1.2.2 基因调控网络的特性

在一定条件下，某些基因的表达或停止（增强或抑制）是细胞完成基本生命活动的基础，也是细胞反应外界刺激的基础。从系统的观点看，基因表达（或调控）的过程是通过一个由多个基因所组成的逻辑网络所决定的，这种网络是由一系列节点（代表基因）和节点之间的连线（调控关系）组成的。

随着系统生物学研究的深入，人们对于基因的表达（或调控）的认识已经从单个基因的变化，扩展到以多基因和基因簇为基础的网络研究。需要强调的是，基因调控网络不同于一般的图形数据结构，它一定要具有生物学意义上的一些特征。这些特征主要表现在以下 4 个方面。

首先，随机性。基因调控在分子水平上呈现出多种因素复杂的交互作用，这些因素都能增强或抑制特定基因的表达。从一个因素发挥作用到下一个因素发挥作用，这个时间间隔，即使是同一个细胞，多次试验其结果也会有很大的不同。

其次，复杂性。高等生物中所具有的基因数量非常巨大，这些基因又分布在不同的染色体上，基因组的结构存在着分裂基因与重复序列的特点。因此染色质构象、结构基因上游各种各样的调节序列、反式作用因子、RNA 聚合酶活性等因素都决定着基因转录的复杂性。

再次，时空特异性。在高等生物的发育过程中，细胞逐步分化，形成各种功能的组织器官。细胞分化是一组特定基因表达与关闭的结果，这一表达形式受到严格的时（个体发育或细胞演化周期）、空（所在组织及邻近组织的关系）调控。由于组织器官高度分化，又有相对稳定的内环境，所以如果只有少部分组织细胞受影响，大部分组织仍维持正常功能。当细胞受到不同的外界刺激（或处于不同的发育

阶段)时,参与表达的基因是不同的,基因之间的调控结果也是不同的。细胞即使受到相同的外界刺激,其反应状态是有可能不同的。

最后,网络动态性。在基因调控网络中,反馈回路可以分为“正反馈”和“负反馈”。负反馈回路可能引起系统的一个稳定的振动行为,正反馈回路可以引起系统的多重稳定性。整个生物系统的稳定性在很大程度上可由调控元素的结构来决定,这体现了基因调控元素的结构与基因调控的动态性。

1.2.3　基因调控网络的数学模型

基因调控网络的数学模型根据不同的标准可以分为不同的类别。根据基因调控是否有方向,可以分为无向调控和有向调控;根据模型处理的数据是否连续,可以分为离散模型和连续模型;根据模型本身的变化程度,可以分为确定性模型和随机性模型。本节所介绍的模型可以分为两大类:第一类是确定性建模,其中包括布尔网络、信息论、概率图模型、贝叶斯网络、动态贝叶斯网络、线性/非线性微分方程和神经网络模型;第二类是随机性建模,其中包括随机模拟方法和随机微分方程模型。

1. 布尔网络模型

布尔网络模型(Boolean Network,BN)最初是由 Kauffman 等提出,它是一种以有向图为基础的离散系统。在布尔网络中,每一个基因只有两种状态:“开(on)”和“关(off)”,而在某个时刻每个基因点只能处于这两种状态中的某一种。“开(on)”和“关(off)”这两种状态可以理解为是基因表达和不表达,或者理解为基因的积极表达状态和消极表达状态,在数值上,“开(on)”可以用“1”表示,“关(off)”可以用“0”表示。布尔网络中,每个节点下一时刻的状态是由相邻节点的状态决定的,相邻节点的状态为输入,经过一系列的逻辑运算得到本节点的新状态。运算中使用的逻辑操作符包括:与(AND)、或(OR)、非(NOT)、异或(XOR)等。下列表达式是布尔网络的方程组形式:

$$\begin{cases} x_1(t+1)=f_1(x_1(t),\cdots,x_n(t)) \\ x_2(t+1)=f_2(x_1(t),\cdots,x_n(t)) \\ \qquad\qquad \vdots \\ x_n(t+1)=f_n(x_1(t),\cdots,x_n(t)) \end{cases} \tag{1.1}$$

其中 $x_i(t)$表示每个基因点在 t 时刻的状态,f_i,$i=1,2,\cdots,n$ 是逻辑运算的函数,也是每个基因点状态更新的计算规则。

布尔网络是离散时间系统,它将每个基因点的状态二值化,也将生物系统的状态进行了离散化的描述。然而在实际中,基因表达水平是一个连续值,所以要将基因表达的数据离散化、二值化,但是这样会造成信息的损失。同理,生物系统的状态演化也是连续的,对每个节点的离散化描述会造成整个系统信息的损失。而且

当布尔网络的规模增大时，网络的状态会以指数级别增长，模型的复杂程度会大大增加，模型对于生物系统描述的准确性也会大大下降。所以布尔网络模型是一个较为抽象的模型，它适合对系统进行宏观的描述，在对准确性要求不高的情况下，有较好的效果。Toussaint 使用布尔网络模型构建了和人类老化相关的基因调控网络。

由于生物系统的随机性，使得基因表达的过程中存在着大量噪声和扰动，布尔网络这种确定性的模型不能准确描述基因间的调控关系。为了在布尔网络模型中引入不确定的因素，概率布尔网络(Probabilistic Boolean Networks，PBN)应运而生，它在原来节点之间关系的基础上加入了概率模型。在概率布尔网络中，每一个节点的状态演化可能根据多个布尔函数，基因采用哪个布尔函数进行演化，需要根据一定的概率来决定。因为有了多个更新函数，所以系统的计算量也随之增大。

2. 信息论模型

信息论模型是使用相关系数或者互信息的方法来刻画基因之间的相互关系，从而构建基因调控网络。信息论模型首先要设定基因之间关系强弱的阈值，如果两个基因之间的相关系数达到或者超过这个阈值，就认为这两个基因之间存在一条无向边。这种方法能够识别稳态数据、时间序列数据和时间延迟数据等数据之间的依赖关系。

早期的信息论模型是相关网络(Relevance Network，RN)，它首先评估所有基因之间的互信息系数，然后根据阈值确定基因之间的连接关系。ARACN(Algorithm for the Reconstruction of Accurate Cellular Networks)模型与 RN 方法类似，它利用了数据处理不等式(Data Processing Inequality，DPI)来消除基因间连接关系较弱的边，但是 DPI 方法只能消除边，却不能增加边。CLR(Context Likelihood of Relatedness)方法采用自适应校正来删除错误的边，并把互信息转化为 Z-score来评价基因间的相互关系。C3NET 方法可以构建一个显著的最大互信息网络，它保留了网络的核心连接关系。这种方法要求只有两个基因的互信息高于其他基因和这两个基因的互信息这种情况下，才认为两个基因之间有连接的边。综上所述，信息论方法构建的基因关系图是无向网络，但是实际的生物基因网络是有向的，和样本的时间之间也存在依赖关系。

3. 概率图模型

概率图模型(Gaussian Graphical Models，GGM)，也称高斯图模型，是一种以无向图为基础的模型。在概率图模型中，假设基因的表达服从多维正态分布，首先计算基因之间的偏相关系数，然后通过统计检测偏相关系数矩阵中每一个元素是否满足相关的要求来建立网络的连接关系。在一个有 p 个基因点的网络中，基因表达水平为：$x_1,\cdots,x_p$，则这些基因之间的关系服从联合正态分布，偏相关系数可以表示为：$\rho_{ij}=\mathrm{Corr}(x_i,x_j\mid x_{-(i,j)})$，其中 $x_{-(i,j)}=\{x_k\mid k\geqslant 1 \text{ 且 } k\neq i,j\leqslant p\}$，当

$p_{i,j}\neq 0$时，则 i,j 之间存在一条边。所以概率图模型中的边，就表示了基因之间存着相互调控的关系。其偏相关系数可以表示为如下的形式：

$$\rho_{ij}=-\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \tag{1.2}$$

其中，σ_{ij} 是协方差矩阵逆的元素。为了简化求解协方差矩阵，研究者提出了一些基于低阶偏相关分析的方法。概率图模型可以消除许多基因之间的间接连接，方便进行分析，但是它构造的图是无向的，不能根据连接关系推断基因变化之间的因果关系，并且作为一种静态模型，概率图模型不能反映基因调控的动态演化。

4. 贝叶斯网络模型

贝叶斯网络模型(Bayesian Network，BN)是另一种基于图形数据结构的模型，这种模型是利用贝叶斯规则建立基因联合概率分布来构造基因调控网络的，构建的网络是有向无环的。贝叶斯网络中的每个节点代表一个基因的表达水平，每一条边表示基因之间的概率依赖关系。从 x_i 到 x_j 的边表示 x_i 是 x_j 的父节点。模型使用条件概率描述 x_j 和 x_i 之间的关系，其联合概率可以表示为以下的形式：

$$P(x_1,\cdots,x_p)=\prod_{i=1}^{p}P(x_i\mid \mathrm{Pa}(x_i)) \tag{1.3}$$

其中，P 是网络中基因的个数，$P(x_i|\mathrm{Pa}(x_i))$是条件概率。

贝叶斯网络模型以统计数据为基础，能够表示各种生物基因的关系，模型能够分析在基因表达数据中固有的噪声。模型可以使用离散和连续两种方式描述基因的状态，为基因调控网络提供了统一且简单的描述方式。但是，贝叶斯网络建立的调控网络是有向无环的，而且没有考虑基因之间动态的调控关系。

为了克服贝叶斯网络的上述缺点，动态贝叶斯网络模型被提了出来。动态贝叶斯网络(Dynamic Bayesian Network，DBN)既可以描述稳态数据，也可以描述基因表达时间序列，但构建动态贝叶斯网络需要花费大量时间来学习条件依赖关系。

5. 微分方程模型

微分方程是描述动态系统的常用数学工具，在生物学中微分方程常用来描述生物大分子随时间的演化过程。通过微分方程构建基因调控网络，要求微分方程描述标靶基因与调控基因之间的调控关系，方程中的相关参数决定了基因网络之间的相互关系，基因调控网络的常微分方程表示如下：

$$\frac{\mathrm{d}x_i}{\mathrm{d}t}=f_i(X_1,\cdots,X_p),i=1,\cdots,p \tag{1.4}$$

其中，p 是基因调控网络中基因的个数，X_i 表示基因 i 的表达水平，f_i 是 n 维空间的值函数。函数 f_i 的项数和对应的参数决定了基因调控网络相互影响关系和每一个因子对于标靶基因的调节强度。构造微分方程的过程就是确定 f_i 表达式并确定其参数的过程。一种简单的线性微分方程形式如下：

$$\frac{\mathrm{d}x_i}{\mathrm{d}t}=\sum_{k=1}^{p}w_{ik}x_k+\varepsilon_i(t)-\lambda_i x_i(t),i=1,\cdots,p \tag{1.5}$$

其中，λ_i 是自降解率，ε_i 是外界噪声函数，且服从正态分布。在生物中，细胞内的调控影响往往是非累积的，也就是说调控基因和其他对靶基因的有调控作用的基因，会由于二者的结合使得自身的调控力度增强或者削弱。这就可能导致微分方程模型构造的基因调控网络不准确，所以微分方程模型通常用来模拟真实系统中的大尺度网络。

6. 神经网络模型

神经网络模型(Neural Network，NN)用来描述复杂的线性和非线性的基因调控网络。在神经网络中，基因就是网络的节点，节点之间的边代表一个基因之间的调控关系，边的权值表示调控关系的强度和调控的类型。在各种神经网络模型中，最常用的是递归神经网络(Recurrent Neural Network，RNN)。

递归神经网络具有生物上的可信性和抗噪性，还兼顾了反馈回路，并在运行过程中考虑系统的内部状态。从而使网络在外部反馈缺乏的情况下振荡，所以使模型具有模拟振荡和周期活动的能力，而且可以分析系统随时间变化的动态行为。一个离散时间神经网络可以表示为以下形式：

$$g_i(t+\Delta t)=g_i(t)+\Delta t[a_iS(w_{ij}g_j(t)+b_i)-d_ig_i(t)] \tag{1.6}$$

其中，w，a，b 和 d 是模型的参数，S 是非线性函数。通过定义误差函数作为网络性能的评价标准，网络学习问题变成了全局最小化误差函数的参数优化问题。

7. 随机模型

生物过程中的随机性可以分为两类：一类是化学反应中的噪声，另一类是具有生物学意义的随机因素。基因调控过程中的随机性会影响细胞的功能，从而造就了个体在表现型上的差异。而这种差异又导致了生物群体中的优胜劣汰，从而造就了生物的进化史。从微观方面看，基因调控过程中的随机性可以增强生物系统的稳定性，帮助细胞应对复杂多变的环境，同时也会引导细胞的分化。

另一方面，基因的表达过程分为很多步骤：转录、mRNA 的剪切、翻译、蛋白质的折叠与修饰等，这些步骤又可以继续细分，在这些十分细小的过程中，都伴有随机性的出现。转录过程就是一个典型的例子，转录的起始会以一定的概率失败，转录的延伸过程中又伴随着 RNA 聚合酶的停止、转录提前结束等多种随机事件。这些随机性反映到模型中就会导致模型中状态变量的值有一定的随机性，时间参数具有滞后，而且时间滞后量也是随机的。

随机模型可以分为静态模型和动态模型。静态模型基本就是统计模型，常见的是回归分析和假设检验。静态模型虽然相对粗略，但却十分实用，因为这类模型可以应对数据十分稀少的情况，可以很好地配合生物学实验，而且模型也容易使用。

动态模型主要是随机模拟方法和随机微分方程模型。随机模拟方法最大的好

处是十分灵活，可以通过随机模拟生物学中的各种反应，来呈现生物系统中的各种过程。随机模拟的一大用途是通过模拟来发现生物过程的特征，如 H. Mcadams 和 A. Arkin 就通过模拟的结果指出基因的表达是猝发式的。

随机微分方程模型是从常微分方程模型扩展而来的，也是典型的反向工程模型。由于模型中含有随机积分，导致从系统的观测数据来估计内部参数十分困难，该问题一直没有较好的解决方案，已有的方案要么耗时过长，要么理论过于繁杂，难以实现。受制于这个因素，以随机微分方程模型为基础的反向工程方法在基因调控网络中的应用还比较有限，相当一部分研究是利用随机微分方程进行随机模拟，并分析系统的随机动力学特性。

1.3　布尔网络研究综述

在众多数学模型中，布尔网络因其简单的表达形式，引起了基因研究者的广泛关注。本节首先介绍布尔网络状态演化的一般规律、更新函数、吸引子和吸引域等相关概念，然后介绍近年来以半张量积为工具在布尔网络状态描述和相关领域取得的成果。

1.3.1　布尔网络的状态描述

上一节我们已经介绍了基因调控网络的主要数学模型，在众多模型中，布尔网络因其简单的表达形式，以及对实验结果的抽象化的描述，引起了学术界的广泛关注。布尔网络可以模拟封闭的基因调控网络，也可以描述有外界输入（或扰动）的基因调控网络。一般情况下，布尔网络的输入表示外界环境对生物系统的影响和扰动，其输出表示了生物系统的相关产物。

布尔网络是一个基于布尔运算的离散时间系统，网络中所有基因点的演化都在统一的时钟周期下进行，也就是说整个网络是同步更新的。每个基因节点在 $t+1$ 时刻的状态值，由和它相邻的基因节点在 t 时刻的状态值来决定。不同基因节点的更新函数，是不同的基于逻辑运算的函数，这种函数统称为布尔函数。而且每一个基因节点的值，只有“1”和“0”两种状态，在某个时刻每个基因点只能处于这两种状态中的某一种。“1”表示“真（on）”状态，说明是基因表达（或者基因的积极表达）；“0”表示“假（off）”状态，说明是基因不表达（或者基因的消极表达）。因为布尔网络的简化描述，所以它能够很好地模拟基因网络控制过程中出现的非线性动力学行为，并进一步揭示基因调控网络的演化规律。目前已经出现了很多使用布尔网络描述基因调控网络的成功案例，如酵母细胞周期网络、哺乳动物细胞周期网络、酵母转录网络、黑腹果蝇体内的基因表达网络等。