Feedback Identification Theory in Systems Biology

董朝轶◎著

系统生物学中的反馈辨识理论研究

北京理工大学出版社
BEIJING INSTITUTE OF TECHNOLOGY PRESS

Feedback Identification Theory in Systems Biology

Chao-Yi Dong
School of Electric Power
INNER MONGOLIA UNIVERSITY OF TECHNOLOGY

北京理工大学出版社
BEIJING INSTITUTE OF TECHNOLOGY PRESS

版权专有　侵权必究

图书在版编目(CIP)数据

系统生物学中的反馈辨识理论研究 = Feedback identification theory in systems biology:英文/董朝轶著. —北京:北京理工大学出版社,2013.3

ISBN 978-7-5640-7505-7

Ⅰ.①系… Ⅱ.①董… Ⅲ.①生物反馈控制-理论研究-英文 Ⅳ.①Q811.3

中国版本图书馆 CIP 数据核字(2013)第 046840 号

出版发行 / 北京理工大学出版社
社　　址 / 北京市海淀区中关村南大街 5 号
邮　　编 / 100081
电　　话 / (010)68914775(办公室)　68944990(批销中心)　68911084(读者服务部)
网　　址 / http://www.bitpress.com.cn
经　　销 / 全国各地新华书店
印　　刷 / 北京通州皇家印刷厂
开　　本 / 889 毫米×1194 毫米　1/32
印　　张 / 6
彩　　插 / 2
字　　数 / 140 千字
版　　次 / 2013 年 3 月第 1 版　2013 年 3 月第 1 次印刷　责任校对 / 周瑞红
定　　价 / 60.00 元　责任印制 / 王美丽

图书出现印装质量问题，本社负责调换

Preface

Fifty years ago, feedback identification theory was first proposed by Caines and Chan in 1975 to find the systematic structure of an industrial process. From then on, the theory was developed rapidly both in theoretical aspects and applicable aspects. Based on system identification theory, time series analysis, and stochastic theory, it now provides numerous applications in the field of engineering, economics, social science, and biology. In particular, it offers theoretically well founded means to find a crucial feedback mechanism underlying a complex network.

Despite these advantages, traditional feedback identification theories often suffer from the opinion that they usually address two-variate time-series data and are inappropriate for large-scale networks because of their

practical and theoretical limitations. Data acquisition is difficult or connective entanglements are fearing, which might hinder their applications to very large datasets, as occur more and more frequently nowadays. Now this is not the case, as many new experimental techniques, for example, real-time PCR, immunofluorescence, microarray, multi-electrode array and EEG, can now provide such time-series data in a cost efficient manner. Also a multi-variate time-series analysis theory has undergone a great development. The new theoretical contribution much helps to find the feedback loops in large-scale networks.

This book aims at tackling the problem of feedback loop identifications in multi-variate time series from large-scale biological networks, considering both the theoretical and practical points of view and bringing together contributions from various fields. The book gathers the up-to-date methods and algorithms that tackle this problem, showing that the proposed methods are a very powerful way to provide biologists with relevant results within reasonable time and memory. The book aims at providing researchers, master

此为试读,需要完整PDF请访问: www.ertongbook.com

students, engineers the state-of-the-art tools to address the new challenges of current applications. The readers will get a panorama of the existing methods, algorithms and applications devoted to biological feedback loop identifications. Besides, being the very first reference gathering feedback loop identification methods from various fields, this book contributes to bridging the gap between the research communities that are not always enough combined and mixed.

In this book, two effective methods, NIRCM (Non-causal Impulse Response Component Method) and MSGCM (Multi-step Granger Causality Method), are developed to identify the feedback loops embedded in biological neural networks, which use only time-series experimental data. The NIRCM, which is based on correlation identification and spectral factor analysis, provides a non-causal component criterion for the identification of feedback loops from a view point of a 2-variate system. The significant non-causal components of the impulse response sequences observed in the negative

time axis imply the existence of a feedback loop. It requires an assumption that the input noise and output noise are uncorrelated with each other. The MSGCM, stemming from a multivariate time-series analysis, employs a modified Wald test to infer the existence of Multi-step Granger Causality between a pair of network nodes. If the significant Multi-step Granger Causality exists in bi-directions between these two variates, it implies a feedback loop. The MSGCM can be viewed as a natural extension of NIRCM to a multivariate system by introducing a multivariate time-series analysis. One limitation of MSGCM is its high complexity of computation. This may deter the application of this method to a very large scaled network.

The two proposed identification methods were first applied to several 2-node SRM (Spike Response Model) networks and another more complicate 6-node SRM network. For these synthetic models, NIRCM and MSGCM correctly imply the existence of feedback loops and show their effectiveness of feedback loop identifications. Finally, NIRCM and MSGCM are applied to the time-series data

from in-vitro cultured neural networks (Micro-Electrode Array data). Some significant feedback loops can be found in those biological neural networks and they infer a close link between the synchronized oscillatory bursting behaviors and existence of DPPFLs (double-positive positive feedback loop).

The organization of the book is as follows. The book starts from one introductory chapter that provides general overviews on the Systems Biology, the definition of feedback loops, and the previous identification methods of biological feedback loops (Chapter 1). Chapter 2 is devoted to the Non-causal Impulse Response Component Method (NIRCM) and its computational complexity. Correlation Method and Spectral Factor Analysis are applied in this chapter to derive the non-causal components criterion for identifying the existence of a feedback loop. In Chapter 3, as a multiple time-series analysis method, the Multi-step Granger Causality Method (MSGCM) is introduced to identify the feedback loops in a large-scale network. In this chapter, we focus on a modified Wald statistics, which can help us to

infer the infinity-step causality between any pair of the two network nodes. To illustrate the effectiveness of the two proposed methods, synthetic pulsed neural network models are constructed in Chapter 4. Then the relations of their typical network behaviors and structures are also analyzed in the same chapter. In Chapter 5, we show that the two feedback loop identification methods can successfully identify the feedback loops embedded in synthetic neural networks with two-node network simulations and large scale random network simulations. Finally, the NIRCM and MSGCM are applied to the time-series data from in-vitro cultured neural networks in Chapter 6. Some significant feedback loops are found in those biological neural networks and they infer a close link between network dynamic and the existence of feedback loops.

The material presented in the book is self-contained. Using the text does not require much earlier knowledge on the control science. The presentation requires only the basic knowledge of linear algebra, differential equations and probability theory. I hope that the book can be useful for

students, researchers and all readers working in the field of Bioinformatics and Systems Biology. Anyway, the ambitious goal of the book is to aid biologists to investigate the underlying feedback circuit in biological systems and make some significant findings.

Many persons have contributed directly or indirectly to this book and I am very grateful to all of them. A number of colleagues have commented on parts of the manuscript and have been available for discussions on the topics covered. These comments and discussions have been very helpful for my own understanding of the subject and have resulted in improvements to the manuscript. Although the persons who have contributed to the project in some way or other are too numerous to be listed here, I wish to express my special gratitude to some of them. They include Professor Kwang-Hyun Cho (KAIST) who read and commented on a large part of the manuscript, Prof. Yoon-Key Nam (KAIST) who shared his many thought-provoking and beneficial discussions with me and kindly provided large amount of experimental data to this book, Dr. Dong-kwan Shin who

commented and helped to prepare the final versions of most figures. I deeply appreciate the helps of all these collaborators. Additionally, the publication of the monograph and parts of my recent researches in the book are supported by "The Project Sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry." I would like to express my thanks to the support from State Education Ministry, People's Republic of China.

此为试读,需要完整PDF请访问: www.ertongbook.com

Contents

1 Introduction 1

1.1 Systems Biology and Its Objective 1

1.2 Biological Feedback Loops 2

1.3 Identification Methods of Biological Feedback Loops 5

1.4 Outline of the Book 9

2 Non-causal Impulse Response Component Methods 11

2.1 Basic Concepts about the Stochastic Process 11

2.2 Correlation Identification Methods 19

2.3 Spectral Factor Analysis 24

2.4 Identification Algorithm of the NIRCM 30

2.5 Perturbation Methods 33

2.6 Factors Affecting the Identification Precision 37

3 Multi-step Granger Causality Methods ····· 51

3.1 Multivariate Time-Series Analysis ····· 51

3.2 Finite-order Vector Autoregressive Model and Its Corresponding Infinite-order Vector Moving Average Model ····· 53

3.3 Estimation of VAR Coefficients ····· 60

3.4 Granger Causality and Multi-step Causality ····· 64

3.4.1 Granger Causality Inference Between the 2-partitioned Variate Sets ····· 67

3.4.2 Granger Causality Between a Pair of Variate Sets ····· 69

3.4.3 Testing Multi-step Granger Causality Between a Pair of Variate Sets ····· 71

3.5 Identification Algorithm of the MSGCM ····· 75

4 Synthetic Spike Neural Networks and Their Dynamical Network Behaviors ····· 80

4.1 Spike Neural Networks ····· 80

4.2 Typical Network Behaviors ····· 85

4.3 Synchronized Bursting Behavior and Feedback Mechanism ····· 87

4.4 Feedback Motifs of Networks ····· 89

4.5 Dynamical Characteristics of Network Motifs ····· 91

5 Application of Feedback Loop Identification Methods to Synthetic Spike Neural Networks 98

5.1 Data Sampling Methods 98

5.2 Two-node Network Simulation 100

5.2.1 Scenario 1: Double-positive Positive Feedback Loops 103

5.2.2 Scenario 2: Negative Feedback Loops 104

5.2.3 Scenario 3: Double-negative Positive Feedback Loops 108

5.2.4 Scenario 4: Unidirectional Excitatory Connections 111

5.2.5 Scenario 5: Unidirectional Inhibitory Connections 114

5.2.6 Scenario 6: No Synaptic Connections 117

5.3 Large-scale Random Network Simulation 120

6 Feedback Loop Identifications for Biological Cultured Neural Networks 126

6.1 MEA Experiments and Its Protocols 126

6.2 Feedback Loop Investigation 128

7 Summary 133

7.1 Summary of the Book 133

7.2 Discussion and Future Research Topics 135

Appendix ····· 139

Proof of Theorem 2.1 ····· 139

Proof of Theorem 3.1 ····· 146

Proof of Theorem 3.2 ····· 149

Bibliography ····· 152

Introduction

1.1 Systems Biology and Its Objective

Systematic approaches have been applied not only in botany, zoology, and ecology, but also in the molecular biology of single cells for a long time. In the past, biologists thoroughly investigated how the individual parts of a cell work: They studied the biochemistry of small and large molecules, the structure of proteins, the structure of DNA and RNA, and the principles of DNA and RNA, and the principles of DNA replication as well as the transcription and translation and the structure and function of membranes (Hunter, 2000; Stryer, 1995). The next step in this line of research is a further effort towards a systematic investigation of cells, organs, cell divisions, homeostasis, and adaptation. This approach has been termed as systems biology.

Systems biology is a biology-based inter-disciplinary study field that focuses on the systematic study of complex interactions

in biological systems, thus using a new perspective to study them. Systems biology finds its roots in:

(i) The quantitative modelling of enzyme kinetics, a discipline that flourished between 1900 and 1970 (Cornish-Bowden, 2004).

(ii) The simulations developed to study neurophysiology (Hodgkin and Huxley, 1990; Le Novere, 2007).

(iii) Control theory and cybernetics (Di Bernardo, et al., 2005; Gardner, et al., 2003; Mesarovi, 1968; Wellstead, et al., 2008).

1.2 Biological Feedback Loops

Various complex control and regulation mechanisms appear within or between reacting biomolecular species that are coordinated in the context of cell functions such as growth, differentiation, apoptosis and communication. As a result, two of the central themes in current systems biology research are the identification of regulatory network motifs, and the development of a quantitative understanding of regulation, control and coordination in intracellular and intercellular networks.

The dynamic behavior of a biological network is often described as a set of nonlinear differential equations (Shinuk, et al.,

此为试读,需要完整PDF请访问: www.ertongbook.com