

Applied Time Series Analysis and Innovative Computing

Lecture Notes in Electrical Engineering
Volume 59

For other titles published in this series, go to
www.springer.com/series/7818

Sio-Iong Ao

Applied Time Series Analysis and Innovative Computing

 Springer

Sio-Long Ao
International Association of Engineers
Unit 1, 1/F, 37–39 Hung To Road
Hong Kong
publication@iaeng.org

ISBN 978-90-481-8767-6 e-ISBN 978-90-481-8768-3
DOI 10.1007/978-90-481-8768-3
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010924436

© Springer Science+Business Media B.V. 2010

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To My Lovely Wife Choi, Wai-Ming

Preface

There are many reasons to analyze the time series data, for example, to understand the underlying generating mechanism better, to achieve optimal control of the system, or to obtain better forecasting of future values. Applied time series analysis consists of empirical models for analyzing time series in order to extract meaningful statistics and other properties of the time series data. With the advances in computer technology, nowadays huge amounts of time series data are stored in data warehouses. Different innovative computing techniques are needed to extract information from these datasets. Innovative computing paradigms can extract the patterns from the time series and to present it in such a way that can better our understanding of the structure, relation, and function of the subjects. The purpose of this book is to illustrate how to apply innovative computing paradigms for the applied time series analysis, with frontier application studies of the time series problems based on the recent works at the Oxford University Computing Laboratory, University of Oxford, the University of Hong Kong, and the Chinese University of Hong Kong. The monograph was drafted when the author was a post-doctoral fellow in Harvard School of Engineering and Applied Sciences, Harvard University. It provides a systematic introduction to the use of innovative computing paradigms as an investigative tool for applications in time series analysis.

In this book, innovative computing algorithms have been illustrated for solving some frontier problems in time series analysis. The book is organized as follows. In Chap. 1, it is the brief introduction to the applied time series analysis and the advances in innovative computing paradigms. The three real-word applications of innovative computing paradigms for time series problems and the contributions of these algorithms to the time series analysis are described briefly. In Chap. 2, we describe about the applied time series analysis generally. Time series analysis models including time domain models and frequency domain models are covered. In Chap. 3, we describe about the recent advances in innovative computing paradigms. Topics like computing algorithms and databases, integration of hardware, systems and networks, Internet and grid computing, and visualization, design and communication, will be covered. The advances of innovative computing for time series problems are also discussed, and an example of building of an innovative computing algorithm for some simulated time series is illustrated. In Chap. 4, we present the real-word application of innovative computing paradigms for time series problems.

The interdisciplinary innovative computing techniques are applied to understand, model and design systems for business forecasting. In Chap. 5, the second real-word application is for the analysis of the biological time series. Recurrent Elman neural networks and support vector machines have been outlined for temporal modeling of microarray continuous time series data sets. In Chap. 6, we present the last real-word application for the astronomical time series. Some innovative computing algorithms are described about how to classify the light curves of the quasars against light curves of the other stars.

In business organizations, forecasting is one of the most important activities that form the basis for strategic and operational decision. Traditionally, business time series forecasting has been dominated by linear methods. However, the linear models have serious limitation with problems of nonlinear relationships. It may be unsatisfactory to approximate the linear models for these nonlinear relationships. The importance and complexity of the business time series forecasting problem paves way for the importance of innovative computing paradigms. In Chap. 4, the interdisciplinary innovative computing techniques are applied to understand, model and design systems for business forecasting. Two business case studies are described to demonstrate the advantages of our proposed system. In global business, the interactions between different markets cause collective lead-lag behavior having special statistical properties which reflect the underlying dynamics. The internal structure of a complex system can manifest itself with correlations among its components. An innovative computing system of combining the vector autoregression and genetic algorithm with neural network is outlined how to take advantage of the lead-lag dynamics, to make the neural network forecasting process more transparent and to improve the prediction capability of the neural network.

The temporal patterns exhibited by biological time series are often complex, nonlinear or chaotic, nevertheless, the analysis are emerging as important means for many biological studies. To address the nonlinear properties of the biological time series, innovative computing algorithms are becoming important tools to study these systems. In Chap. 5, recurrent Elman neural networks and support vector machines are outlined for temporal modeling of microarray continuous time series data sets. In a single microarray experiment, the expression levels of as many as thousands of genes can be measured simultaneously. Thus, it can enable the genome-wide measurement of gene expression, and the construction of genetic network from gene expression time series with innovative computing approaches. An ensemble of the ENN and SVM models are described to better improve the transparency and robustness of the recurrent neural networks, and to further improve the prediction accuracy of the individual models. In order to provide the neural networks with explanation capabilities, a pedagogical rule extraction technique is considered for inferring the output of our proposed ensemble system. The gene regulatory network can be reconstructed satisfactorily with this hybrid innovative computing methodology. The proposed rule extraction technique provides the network inference of the ensemble. The reconstructed gene regulatory network enables us to better understand the dynamics of the underlying biological processes.

With the advances of the technologies for the sky surveys, massive amount of survey data become available. It is very helpful for the automatic and semi-automatic innovative computing methods in the classifications and detections of the astrophysical objects. In Chap. 6, the first section describes about the existing automatic and semi-automatic innovative computing methods for the comprehensive search of quasars. Quasars are interesting astrophysical objects that have been recently discovered more comprehensively from the sky surveys. Secondly, some innovative computing algorithms are described about how to classify the light curves of the quasars against light curves of the other stars.

Topics covered in the monograph include Frequency Domain, Correlation, Smoothing, Periodogram, Autoregression, ARIMA Models, Discrimination Analysis, Clustering Analysis, Factor Analysis, Dynamic Fourier Analysis, Random Coefficient Regression, Discrete Fourier Transform, Innovative Computing Algorithms, Knowledge Extraction, Large Complex Databases, Modeling and Simulations, Integration of Hardware, Systems and Networks, Grid Computing, Visualization, Design and Communication, Business Time Series Applications, Biological Time Series Applications, and Astronomical Time Series Applications. The book offers the state of art of tremendous advances in applied time series analysis and innovative computing paradigms and also serves as an excellent reference work for researchers and graduate students working on applied time series analysis and innovative computing paradigms.

The author is grateful for the supports of Dr. Vasile Palade throughout the author's research in Oxford University Computing Laboratory, University of Oxford, UK, and the supports of colleagues in Harvard throughout the author's research in Harvard School of Engineering and Applied Sciences, Harvard University, USA.

May 2009

Sio-Long Ao, Ph.D.
International Association of Engineers
Unit 1, 1/F, 37-39 Hung To Road
Hong Kong

Contents

1	Introduction	1
1.1	Applied Time Series Analysis	2
1.1.1	Basic Definitions	2
1.1.2	Basic Applied Time Series Models	2
1.1.3	Frequency Domain Models	2
1.2	Advances in Innovative Computing Paradigms	3
1.2.1	Computing Algorithms and Databases	3
1.2.2	Integration of Hardware, Systems and Networks	3
1.2.3	Internet, Web and Grid Computing	4
1.2.4	Visualization, Design and Communication	4
1.3	Real-World Applications: Innovative Computing Paradigms for Time Series Problems	5
1.3.1	Developing Innovative Computing Algorithms for Business Time Series	5
1.3.2	Developing Innovative Computing Algorithms for Biological Time Series	6
1.3.3	Developing Innovative Computing Algorithms for Astronomical Time Series	6
2	Applied Time Series Analysis	9
2.1	Basic Characteristics of Time Series	10
2.1.1	Estimation of Correlation	10
2.1.2	Stationary Time Series	12
2.1.3	Smoothing of the Time Series	12
2.1.4	Periodogram Analysis	13
2.2	Autoregression and ARIMA Models	14
2.2.1	Time Series Regression	14
2.2.2	Autoregressive Moving Average Models	14
2.2.3	Building ARIMA Models	15
2.2.4	Forecasting and Evaluation	16
2.2.5	Causality of the Time Series	16
2.3	Mathematical Models in the Frequency Domain	17
2.3.1	Introduction	17

2.3.2	Discrimination Analysis.....	18
2.3.3	Clustering Analysis.....	19
2.3.4	Principal Components and Factor Analysis.....	21
2.3.5	Dynamic Fourier Analysis.....	22
2.3.6	Random Coefficient Regression.....	23
2.3.7	Discrete Fourier Transform.....	24
3	Advances in Innovative Computing Paradigms.....	25
3.1	Research Advances in Computing Algorithms and Databases.....	25
3.1.1	Knowledge Extraction Methods.....	25
3.1.2	Exploiting Large Complex Databases.....	26
3.1.3	Neural Computing Algorithms.....	26
3.1.4	Fuzzy Computing Algorithms.....	27
3.1.5	Evolutionary Computing Algorithms.....	27
3.1.6	Quantum Computing Algorithms.....	28
3.1.7	Swarm-Based Computing Algorithms.....	28
3.1.8	DNA Computing Algorithms.....	29
3.1.9	Theoretical Modeling and Simulations.....	29
3.2	Research Advances in Integration of Hardware, Systems and Networks.....	29
3.2.1	Innovative Experimental Hardware System.....	29
3.2.2	Data-Acquisition Devices.....	30
3.2.3	Interaction Devices for Visual Exploration.....	30
3.2.4	Graphics Processing Units and Co-Processors for Innovative Computing.....	31
3.2.5	Networking and Interoperability.....	31
3.2.6	Code Optimization and Integration.....	32
3.3	Research Advances in Internet, Web and Grid Computing.....	32
3.3.1	Distributed Computation and Data Sharing.....	32
3.3.2	Large-Scale Collaborations over the Internet.....	32
3.3.3	Grid Computing.....	33
3.3.4	Pooling of Remote Computer Resources.....	33
3.3.5	Integration of Knowledge Metadata Systems.....	33
3.4	Research Advances in Visualization, Design and Communication.....	34
3.4.1	Novel Solutions to Visualization and Communication Challenges.....	34
3.4.2	Displaying of Complex Information.....	34
3.4.3	Escaping Flatland.....	35
3.4.4	Systems Integration for High Performance Image Processing.....	35
3.4.5	Representation of Uncertainties.....	36
3.4.6	Informative Graphics for Scientific Communication.....	36
3.5	Advances and Applications for Time Series Problems.....	37
3.5.1	Efficient Retrieval of Similar Time Series.....	37

3.5.2	Automatic Classification of Time Series Sequences	37
3.5.3	Time Warping Algorithms	38
3.5.4	Time Frequency Clustering of Time Series Datasets	40
3.5.5	Enhanced Representation for Complex Time Series	40
3.5.6	Automatic Monitoring of Large and Complex Time Series	41
3.6	An Illustrative Example of Building an Innovative Computing Algorithm for Simulated Time Series	41
3.6.1	Description of the Simulated Time Series Problem	41
3.6.2	Background of the Methodology	42
3.6.3	Building the Innovative Regression Model	44
3.6.4	Experimental Results with the Simulated Time Series	48
3.6.5	Discussions and Further Works	50
4	Real-Word Application I: Developing Innovative Computing Algorithms for Business Time Series	51
4.1	Business Time Series	51
4.2	Advances in Business Forecasting	52
4.2.1	Basic Econometrics Models	52
4.2.2	Neural Computing Models	52
4.2.3	Evolutionary Computing Models	53
4.2.4	Hybrid Intelligent Models	53
4.3	Developing a Hybrid Intelligent Econometrics Model for Business Forecasting	54
4.3.1	Vector Autoregression	54
4.3.2	Neural Network	55
4.3.3	Genetic Algorithm	58
4.3.4	A Cybernetic Framework of Hybrid Vector Autoregression, Neural Network and Genetic Algorithm	60
4.4	Application for Tourism Demand Forecasting	61
4.4.1	Quantifying Cross-Market Dynamics	62
4.4.2	Experimental Results	62
4.5	Application for Cross-Market Financial Forecasting	63
4.5.1	Quantifying the Cybernetic Lead-Lag Dynamics across Different Markets	64
4.5.2	Benchmark Stand-Alone Neural Network	64
4.5.3	Hybrid Innovative System and Results Comparison	65
4.6	Discussions and Further Works	66
5	Real-Word Application II: Developing Innovative Computing Algorithms for Biological Time Series	67
5.1	Biological Time Series	67
5.2	Advances in Experimental Designs for Microarray Time Series	68

5.2.1	Microarray Experiments	68
5.2.2	Microarray Time Series and Applications	69
5.3	Reverse Engineering of Biological Networks.....	70
5.3.1	Introduction.....	70
5.3.2	Materials and Methods.....	71
5.4	Models for Biological Network Inference	78
5.4.1	Biological Time Series Datasets	78
5.4.2	Analysis with Simulated Non-stationary Datasets.....	79
5.4.3	Analysis with Real Biological Datasets	79
5.4.4	Rule Extraction for Reverse Engineering of Biological Networks	80
5.5	Discussions and Further Works.....	81
6	Real-Word Application III: Developing Innovative Computing Algorithms for Astronomical Time Series	83
6.1	Astronomical Time Series.....	83
6.2	Advances and Applications of Innovative Computing Paradigms.....	84
6.2.1	Classification of Astronomical Time Series	84
6.2.2	Clustering of Astronomical Time Series.....	84
6.2.3	Semi-Supervised Learning for Astronomical Time Series	85
6.2.4	Anomaly Detection of Astronomical Time Series.....	86
6.3	Motivations for Investigating the Quasar Time Series with Innovative Approaches.....	86
6.4	Advances in Emerging Methods for Quasar Studies	87
6.4.1	Variability Properties of the Quasar Light Curves.....	87
6.4.2	Algorithms Based on Variability and Proper Motion for Quasar Classification.....	89
6.4.3	Bayesian Classification for Efficient Photometric Selection of Quasars	93
6.4.4	Machine Learning Paradigms for Quasar Selection	97
	Bibliography	99