

商务智能与大数据丛书

# 复杂时间序列 预测技术研究： 数据特征驱动分解集成方法论

汤 铃 余乐安 著  
李建平 汪寿阳



科学出版社

商务智能与大数据丛书

# 复杂时间序列预测技术研究： 数据特征驱动分解集成方法论

汤 铃 余乐安  
李建平 汪春明



科学出版社

北京

## 内 容 简 介

本书致力于时序预测技术创新,试图构建一个具有广泛适用性与高预测精度的预测方法论。针对预测模型具有各自的数据针对性与优劣势,本书创新性提出了“数据特征驱动”思想,旨在充分考虑研究样本的数据特征,相应设计与之相匹配的预测方法。在此基础上,本书将新思想与复杂系统前沿分析技术“先分解后集成”思想相结合,提出了一个新的复杂时序预测方法论——数据特征驱动分解集成方法论。数据特征驱动分解集成方法论以基于“先分解后集成”思想的分解集成模型为分析框架,包括时序分解、模态预测与集成预测3个主要步骤;并以“数据特征驱动”思想为建模依据,试图紧扣研究样本的数据特征,设计相应的分技术——时序分解技术、模态预测技术与集成预测技术。新方法论紧扣研究样本的数据特征,不限于特定的研究领域,具有广泛的适用性,为复杂数据分析特别是复杂时序预测提供了一个新的研究视角。

本书可作为理工类、经管类研究生学习数据分析与预测方法的教材,也可供科研机构、高等院校等单位从事相关研究的科研技术人员参考。

### 图书在版编目(CIP)数据

复杂时间序列预测技术研究:数据特征驱动分解集成方法论/汤铃等著. —北京:科学出版社,2016

(商务智能与大数据丛书)

ISBN 978-7-03-047636-4

I. ①复… II. ①汤… III. ①时间序列分析-预测技术-研究 IV. ①O211.61

中国版本图书馆CIP数据核字(2016)第045901号

责任编辑:马跃 王丹妮/责任校对:贾娜娜

责任印制:霍兵/封面设计:无极书装

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

http://www.sciencep.com

三河市骏立印刷有限公司印刷

科学出版社发行 各地新华书店经销

\*

2016年5月第一版 开本:720×1000 1/16

2016年5月第一次印刷 印张:12 1/2

字数:252 000

定价:68.00元

(如有印装质量问题,我社负责调换)

# 丛书编委会

## 顾问

- 汪寿阳 中国科学院数学与系统科学研究院, 研究员, 第三世界科学院院士  
James T. Tien 美国迈阿密大学工学院, 教授, 美国工程院院士  
黎建强 香港城市大学商学院, 讲座教授, 长江学者

## 主编

- 余乐安 北京化工大学经济管理学院, 教授, 国家杰出青年基金获得者

## 编委 (以姓氏笔划为序)

- 王帆 中山大学管理学院, 教授, 国家杰出青年基金获得者  
王红卫 华中科技大学自动化学院, 教授, 长江学者, 国家杰出青年基金获得者  
毛基业 中国人民大学商学院, 教授, 国家杰出青年基金获得者  
石勇 中国科学院大学管理学院, 教授, 国家杰出青年基金获得者  
叶强 哈尔滨工业大学管理学院, 教授, 国家杰出青年基金获得者  
吕金虎 中国科学院数学与系统科学研究院, 研究员, 国家杰出青年基金获得者  
李一军 国家自然科学基金委员会管理科学部, 教授, 长江学者  
李仲飞 中山大学管理学院, 教授, 长江学者、国家杰出青年基金获得者  
李建平 中国科学院科技政策与管理科学研究所, 研究员, 国家杰出青年基金获得者  
杨晓光 中国科学院数学与系统科学研究院, 研究员, 国家杰出青年基金获得者  
吴信东 合肥工业大学计算机学院, 教授, 国家“千人计划”, 国家杰出青年基金获得者  
余玉刚 中国科学技术大学管理学院, 教授, 国家杰出青年基金获得者  
邹国华 中国科学院数学与系统科学研究院, 研究员, 国家杰出青年基金获得者  
张顺明 中国人民大学财政金融学院, 教授, 国家杰出青年基金获得者  
张彦春 澳大利亚维多利亚大学, 教授, 国家“千人计划”  
徐泽水 四川大学工商管理学院, 教授, 长江学者、国家杰出青年基金获得者  
唐立新 东北大学信息学院, 教授, 长江学者、国家杰出青年基金获得者  
唐加福 东北财经大学管理科学与工程学院, 教授, 长江学者、国家杰出青年基金获得者  
寇刚 西南财经大学工商管理学院, 教授, 长江学者  
彭怡 电子科技大学经济管理学院, 教授, 长江学者、国家杰出青年基金获得者  
曾大军 中国科学院自动化研究所, 研究员, 国家杰出青年基金获得者  
樊治平 东北大学工商管理学院, 教授, 长江学者、国家杰出青年基金获得者

## 总 序

早在 1958 年，当 IBM 公司研究员 Hans Peter Luhn 开始研究如何利用具有强大运算能力的计算机辅助人类决策时，首次提出了商务智能或商业智能（business intelligence, BI）的概念。但这个概念到底是什么，具有什么样的内涵与外延，没有人能描述清楚，商务智能成为当时“高大上”的词汇。随着信息与通信技术（information and communication technology, ICT）的不断发展，Howard Dresner 在 1989 年正式将商务智能描述为“使用基于事实的决策支持系统来改善业务决策的一套理论与方法”。至此，商务智能如“天使”降临凡间，其概念经由 Howard Dresner 的通俗化而被人们广泛了解。

进入 20 世纪 90 年代，以互联网为代表的信息技术的蓬勃发展催生了对商务智能的强大需求。1996 年，全球最大的 IT 行业分析和咨询服务商——加特纳集团（Gartner Group）适时推出商务智能，并迅速成为时代的热点聚焦。通常，商务智能是由一系列技术与方法组成的解决方案，为各类组织迅速有效地收集、存储、分析、挖掘和利用数据，将这些数据转化为知识，帮助组织作出科学的决策。在某种意义上，商务智能就是一个典型的决策支持系统，是数据仓库（data warehouse）、联机分析处理（OLAP）工具和数据挖掘（data mining）等技术及其软件工具在组织决策中的综合运用。其中，数据仓库主要实现数据提取（extraction）、转换（transformation）、加载（load）和存储（storage）等多种数据管理功能；联机分析处理是数据仓库系统的主要应用，支持复杂的数据分析操作，满足在多维环境下对数据的处理与分析；而数据挖掘是商务智能的技术核心，用以发现数据之间的内在和外在关系，并作出相应的决策支持，包括分类、估计、预测、聚类、关联规则等多种分析方法。

随着物联网、云计算、移动互联网、社交网络等新一代信息技术的应用和推广，人类产生的数据成倍增长，数据种类繁多，数据在宽带网络中高速流动，数据的待开发价值越来越大，已经进入了大数据（big data）时代。通常，大数据是以海量数据的处理和分析，发现数据背后的本质，增加组织洞察力为目的而诞生的一类技术，这与商务智能的目的异曲同工。可以说，大数据是从数据挖掘、商务智能发展而来，是传统数据库、数据仓库、商务智能概念外延的扩展和手段的扩充。在这个意义上，大数据可理解为实现商务智能的一种新技术手段。

在大数据时代下，数据的典型特征是体量大、复杂性高、更新速度快，这给传统的商务智能理论与技术带来新的机遇与挑战。首先，如何有效处理与某一特

定现象相关的更多数据甚至是所有数据，这不仅是大数据时代带给基于数据驱动的商务智能的一个极好的机遇，同时也是对商务智能在数据存储、处理与分析挖掘功能上的技术挑战。在某种程度上，大数据可以看成是商务智能的一个新数据源。其次，面对如此之多的数据，商务智能的分析侧重点将从微观层面上的精确性转移至宏观层面上的方向把握性。最后，商务智能将不再热衷于挖掘数据之间的因果关系，而应寻找事物之间的相关关系，警示某现象或规则正在发生或将要发生，而非发生的具体因果原由。

在这样的背景下，数据技术从早期在单机上处理单一类型数据，发展到当前在计算机集群上处理多种类型数据，实现时间宽松的数据分析与挖掘应用。随着数据量发展到 PB、EB 级甚至更大，并且要求更快地处理分析时间，大数据专用计算机、异地分布式计算机集群、多来源数据融合、多类型数据处理、数据网络等复杂结构数据分析、秒级时间分析等通用技术以及各种面向领域的应用技术是大数据技术的发展趋势。在这个意义上，大数据技术可以理解为商务智能在数据量大规模爆发后的必然结果。实际上，传统的商务智能技术手段的确很难应付当前数据量的爆炸式增长，这也使得商务智能亟需新的理论与技术创新。然而，无论是“小数据”时代还是“大数据”时代，各类组织所面临的数据管理问题都是相同的，即挖掘数据之间的内在联系，催生新的商业价值。

在大数据发展如火如荼的当下，一大批研究学者在大数据环境下对商务智能已经进行了实质性的理论与方法创新工作。特别地，作为主要创始人，本人所创立的“商务智能与金融工程国际学术会议（International Conference of Business Intelligence and Financial Engineering）”在国际相关领域得到广泛关注和高度认可，被管理信息系统顶尖期刊 *MIS Quarterly* 列为商务智能与分析领域排名第一的国际会议。以该国际会议为学术交流平台，一大批学者就“商务智能与大数据”研究议题进行了热烈的学术交流与研讨。学术因论辩而产生共鸣，研究学者在不断研讨中碰撞出思想的火花，这极大地促进了商务智能与大数据研究工作不断取得新的进展与突破。

由科学出版社推出的这套“商务智能与大数据丛书”，将陆续展现商务智能在大数据背景下的理论与技术创新成果，记录商务智能在大数据时代下的变革与发展历程，集成商务智能与大数据两个新兴学科的交叉研究领域的探索性工作。在本套丛书的研究和出版过程中，得到了许多机构、同行与朋友的鼓励、支持与帮助。首先，“商务智能与大数据丛书”的出版计划得到了中国科学院数学与系统科学研究院汪寿阳院士和香港城市大学商学院黎建强教授的大力支持与鼓励。其次，感谢科学出版社的积极响应与全力支持，尤其是经管分社各位编辑对这套丛书的出版倾注了大量的时间与精力，从帮助策划到丛书宣传给予了许多具体的指导和帮助。同时，要特别感谢国家杰出青年科学基金（项目名称：基于商务智能的经

济预测与金融管理研究, 资助编号: 71025005)对这套丛书的出版所给予的经费支持。此外, 还要感谢中共中央组织部“万人计划”、中科院“百人计划”、国家自然科学基金重点项目、北京化工大学中央高校基本科研业务费等对我研究工作的支持。本丛书的部分专著就是所受资助的研究项目的部分成果的总结。

最后, 希望这套丛书的出版能够积极推动我国商务智能与大数据领域的研究和传播, 促进其在不同研究领域的实际应用, 推动我国相关学科领域的发展。

余乐安

北京化工大学经济管理学院院长  
北京化工大学大数据研究中心主任

2016年1月

# 前 言

时间序列（本书简称“时序”）预测一直是管理科学与大数据分析研究领域的重点与难点。准确把握主要变量在时间维度上的波动特性与趋势变化，是深入理解所研究系统的机制机理和演化规律的基础，为进一步的决策分析与政策制定工作提供强有力的理论依据与数据支持。作为学术界的研究热点，如何改进预测模型、提高预测精度，是时序预测研究领域的核心创新点。然而，不同的预测模型具有各自的数据针对性与优劣势，现有文献不足以证明某一个（类）方法在不同研究样本中均具有最高的预测准确性。这意味着，充分考虑研究样本的数据特征，并相应设计与之相匹配的预测方法，能为预测理论与方法研究提供一个新的创新视角。对此，本书的主要研究工作将集中于时序预测技术创新，试图构建一个具有广泛适用性与高预测精度的预测方法论。

在此背景下，本书紧扣研究样本的数据特征，创新性地提出了“数据特征驱动建模”思想。具体地，“数据特征驱动建模”思想可定义为，一个有效的预测模型必须与研究样本的各数据特征相匹配，并准确刻画其主要数据特征；相反，忽略数据特征而构建不恰当的数学模型，预测将会失效。在此基础上，本书将“数据特征驱动建模”思想引入复杂系统前沿分析技术“先分解后集成”思想，构建一个新的复杂时序预测方法论——数据特征驱动分解集成方法论。“先分解后集成”思想，即“分而治之”思想，旨在将复杂系统分解为一系列易于描述的简单子系统，以降低建模的困难性，提高系统分析研究的可行性与有效性。

具体地，本书主要的创新工作可概括为以下几个方面。

（1）提出一个新的复杂时序预测方法论——数据特征驱动分解集成方法论。本书创新性地提出了“数据特征驱动建模”思想，并将其与复杂系统前沿分析技术“先分解后集成”思想相结合，提出了一个新的时序预测方法论，即“数据特征驱动分解集成方法论”。该方法论以基于“先分解后集成”思想的分解集成模型为分析框架，包括时序分解、模态预测与集成预测 3 个主要步骤；并以“数据特征驱动”思想为建模依据，试图紧扣研究样本的数据特征，设计相应的分技术——时序分解技术、模态预测技术与集成预测技术。数据特征驱动分解集成方法论紧扣研究样本的数据特征，不限于特定的研究领域，具有广泛的适用性。



(2) 构建一套综合性的时序数据特征识别方案。本书对时序数据可能具有的数据特征及其相互关系进行了系统的梳理工作，并对相关特征检验方法进行了全面的分析工作。在此基础上，本书构建了一套综合性的时序数据特征识别方案，试图更加系统、全面地把握时序数据的本质特征与主要模式特征，是数据分析研究领域中的一项创新工作。

(3) 探讨数据特征驱动的时序分解。本书将“数据特征驱动建模”思想引入分解集成模型中的第 1 步，即根据研究样本的数据特征选择与之相匹配（即具有相同数据针对性）的分解方法。本书对现有 3 种主流分解方法（时域分解方法、时频分解方法与瞬频分解方法）的数据针对性进行深入的比较分析工作，初步探讨分解方法与数据特征的匹配关系。作为分解集成模型的关键性步骤，数据特征驱动的时序分解将确保数据分解的有效性，进而充分发挥“先分解后集成”思想在复杂系统分析中的优势——从不同时间尺度上识别数据动力系统的影响因素与潜在规律。

(4) 探讨数据特征驱动的模式预测。基于“数据特征驱动建模”思想，在分解集成模型第 2 步中，对各模态分量的本质特征进行全面检验，确定与之相兼容的预测方法，并对其主要模式特征进行识别，引入模式特征变量对其进行有效刻画。对此，本书对现有时序预测方法的数据针对性与优劣势进行充分的比较分析，从而提炼预测方法与数据特征的匹配关系，以及模式特征的有效刻画方法。数据特征驱动的模式预测紧扣各模态分量的数据特征，构建与之相匹配且有效刻画其主要特征的预测模型，保证了对各模态分量预测的准确性。

(5) 探讨了数据特征驱动的集成预测。在分解集成模型第 3 步中，紧扣研究样本的数据特征，确定与之相兼容（即具有相同数据针对性）的集成方法，以有效刻画原系统内部各因素（由模态体现）间的相互作用机制，形成最终预测结果。同样地，本书对现有主流的集成方法的数据针对性进行了相关分析，初步探索了集成方法与数据特征的匹配性。数据特征驱动的集成预测，通过识别原动力系统错综复杂的内部结构，确定与之相匹配的集成预测技术，确保最终预测结果的合理性。

在本书的研究和写作中，得到许多同行与朋友的鼓励、支持与帮助，包括：中国科学院科技政策所徐伟宣研究员、陈建明研究员、范英研究员、蔡晨研究员等。本书的研究得到了国家杰出青年科学基金、国家自然科学基金重点项目、国家自然科学基金青年科学基金、中组部“青年拔尖人才支持计划”、中国科学院“百人计划”、中央高校基本科研业务费专项资金等的资助，以及国家自然科学基金委员会、中国科学院和北京化工大学的大力支持。此外，要特别感谢科学出版社对本书的出版所提供的支持与帮助。

本书所提出的新思想与新方法仍处于初步形成阶段，加之作者水平有限，肯定存在很多缺陷与不足之处，需要进一步的检验、纠正、完善与扩展。恳请各位专家和广大读者批评指正。

作 者

2016年1月

# 目 录

前言

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	3
1.2.1 主流时序预测技术研究综述	3
1.2.2 清洁能源与国家风险预测技术研究综述	10
1.3 本书主要内容	15
1.3.1 研究思路与技术路线	15
1.3.2 章节结构安排	16
1.3.3 主要创新点	17
第 2 章 方法论理论框架	19
2.1 核心思想	19
2.1.1 “数据特征驱动建模”思想	19
2.1.2 “先分解后集成”思想	21
2.1.3 流程框架	23
2.2 实证设计	25
2.2.1 研究样本	25
2.2.2 评价标准	31
2.3 本章小结	32
第 3 章 数据特征识别	34
3.1 数据特征识别的重要性	34
3.2 数据特征及其相互关系	35
3.3 数据特征识别方案	39
3.3.1 本质特征确定	40
3.3.2 模式特征度量	46
3.3.3 主要步骤	50
3.4 数据特征识别方案的应用	53
3.4.1 研究样本的本质特征确定	53
3.4.2 研究样本的模式特征度量	57
3.5 本章小结	68

<b>第 4 章 分解集成模型</b> .....	70
4.1 分解集成模型的提出 .....	70
4.2 分解集成模型方法 .....	71
4.2.1 分解方法 .....	72
4.2.2 预测方法 .....	75
4.2.3 集成方法 .....	79
4.3 分解集成模型的应用 .....	79
4.3.1 准确性评价 .....	80
4.3.2 方向性评价 .....	87
4.3.3 统计检验 .....	91
4.3.4 结果小结 .....	93
4.4 本章小结 .....	93
<b>第 5 章 数据特征驱动的时序分解</b> .....	95
5.1 数据特征驱动时序分解的重要性 .....	95
5.2 数据特征驱动分解方法 .....	96
5.2.1 时域分析方法 .....	97
5.2.2 时频分析方法 .....	100
5.2.3 瞬频分析方法 .....	101
5.2.4 分解方法与数据特征 .....	103
5.3 数据特征驱动时序分解的应用 .....	104
5.3.1 传统分解方法下的预测结果 .....	106
5.3.2 自适性分解方法下的预测结果 .....	111
5.3.3 数据特征驱动分解的有效性 .....	116
5.4 本章小结 .....	122
<b>第 6 章 数据特征驱动的模式预测</b> .....	124
6.1 数据特征驱动模式预测的重要性 .....	124
6.2 数据特征驱动模式预测方法 .....	125
6.2.1 方法类别与本质特征 .....	127
6.2.2 模式变量与模式特征 .....	132
6.3 数据特征驱动模式预测的应用 .....	133
6.3.1 方法类别选择 .....	134
6.3.2 模式变量引入 .....	141
6.4 本章小结 .....	150
<b>第 7 章 数据特征驱动的集成预测</b> .....	152
7.1 数据特征驱动集成预测的重要性 .....	152

---

7.2 数据特征驱动集成预测方法	153
7.2.1 拟合回归方法	153
7.2.2 简单集成方法	154
7.2.3 集成方法与数据特征	155
7.3 数据特征驱动集成预测的应用	156
7.3.1 数据类型 1 的预测结果	157
7.3.2 数据类型 2 的预测结果	161
7.3.3 数据类型 3 的预测结果	162
7.4 本章小结	164
<b>第 8 章 总结与展望</b>	<b>165</b>
8.1 本书主要研究工作	165
8.2 本书主要研究成果	166
8.3 研究展望	168
<b>参考文献</b>	<b>169</b>

# 第1章 绪 论

本章拟阐述时序预测研究创新的重要意义，并对现有时序预测相关研究进行了梳理与总结，从而得出本书研究的创新思路与研究内容。1.1节论述本书的研究背景与意义；1.2节对现有时序预测相关研究进行全面、系统的梳理工作，总结提炼各类预测方法的特点及优劣势，并相应得出两个重要启示；在此基础上，1.3节给出本书的创新思路、研究内容与章节结构。

## 1.1 研究背景与意义

时序预测是管理科学与大数据分析领域重要的研究主题之一（Armstrong, 1988; Granger and Jeon, 2007）。不失一般性，时序主要指某统计指标的数值按时间先后顺序排列的序列，而时序预测指基于时序数据的波动历史，依照一定的规律对其未来走势进行的估计。时序预测研究一直是学术界的核心议题，涉及各种研究领域，如经济、社会、能源、环境、气候等研究领域（Pesaran et al., 2009; Miranda and Lima, 2011; Kurt and Oktay, 2010; Grell and Baklanov, 2011）。把握各主要变量在时间维度上的波动特性与趋势规律，是深入理解相关研究领域系统机制与发展动向的基础，能为信息处理、决策分析及政策制定等相关工作提供强有力的理论依据与方法支持（Gooijer and Hyndman, 2006）。

作为学术界的热点，时序预测性能的提高，包括水平与方向上的预测准确性，对相关研究领域的预警、管理、控制与决策方案制订具有重要的指导作用。例如，在清洁能源方面，提高对市场需求及价格等关键性因素的预测精度，将有利于政府相关部门及时了解能源发展动态，以制定相应的政策规划，确保相关战略规划目标的实现（如“中长期核电发展规划”与水电发展的“十二五规划”），促进清洁能源的快速健康发展（Wang et al., 2011; Tang et al., 2012）。同样在国际市场中，准确把握各种风险因素的波动与演化，能为跨国企业制定合理的对外经济贸易政策提供有力的决策支持，有效规避不确定性损失，最大化预期收益（Sun et al., 2011; Li et al., 2012）。

如何提高预测性能及预测准确性是时序预测研究领域最重要的难题之一。现有大部分研究均与提高预测精度相关，试图改进预测模型以提高预测性能指标。一方面，部分研究根据时序数据所具有的某些数据特征，如季节性、突变性、记忆性与非线性等数据特征，对已有预测方法进行扩展与修正，以提高模型的分析

与预测性能 (Box et al., 1976; Zivot and Andrews, 1992; Woods, 2009)。另一方面, 部分研究认为传统预测技术会在非平稳性、非线性及复杂性时序数据处理中无能为力, 而引入机器学习等其他手段, 通过让机器模拟人类智力或自然现象来解决复杂的工作, 可得到更快更准确的预测结果 (Han and Kamber, 2006; Witten and Frank, 2006)。

随着研究的不断深入, 人们逐渐意识到各主要时序变量均处于复杂系统中, 受多种内外部因素影响而呈现出多维数据特征; 此外, 各种突发事件也起着举足轻重的作用, 时序预测变得十分困难。因此, 单模型 (single model) 已经不能满足对复杂系统的分析与预测要求, 而集成两种以上模型方法的混合模型 (hybrid model), 即集成模型 (ensemble model) 在近年来的时序预测研究中受到广泛的关注。时序混合模型研究中的一个重要的工作是 Wang (2004) 针对复杂系统提出的 TEI@I 方法论, 该方法论以“先分解后集成 (decomposition and ensemble)”思想为核心, 将文本挖掘 (text mining)、传统计量 (econometrics)、人工智能 (intelligence) 及系统集成 (integration) 技术相融, 以发挥各模型优势并弥补其不足, 提高预测精度。此外, 作为时序混合模型中的一种主流方法, 分解集成模型 (多尺度模型) 基于“先分解后集成”思想, 将时序数据分解为易于刻画的简单分量, 以降低建模难度, 改善模型预测性能 (Yu et al., 2008; Zhang et al., 2008; Li et al., 2012)。

然而, 不同的预测模型具有各自的优势与缺陷, 现有文献不足以证明某一个 (类) 方法在各种时序样本中均具有最高的预测准确性。其中, 传统计量模型的函数形式较为固定与简化, 同时对样本数据施加较为严格的统计假设, 虽然对符合其假设条件的时序数据具有良好的预测性能, 但其假设与实际情况中绝大部分数据具有非平稳性、非线性及高复杂性等数据特征相悖。另外, 虽然已有大量研究证实人工智能 (artificial intelligence, AI) 算法在复杂系统研究中具有显著的高效分析优势, 但它们同样存在自身缺陷, 如参数敏感性、局部最优及过度拟合等。

相比单模型, 现有研究已充分证明混合模型具有较高的数据分析与预测性能 (Yu et al., 2008; Zhang et al., 2008; Li et al., 2012)。然而, 同样的问题仍然存在: 混合模型中的不同分技术 (如分解技术、预测技术与集成技术) 具有各自的数据针对性及优劣势, 没有充分的证据证明某一个特定的混合模型适用于各种样本数据, 并在预测性能上优于其他同类的混合模型。

上述分析意味着时序预测研究需要进一步的理论与方法创新, 而充分考虑研究样本的数据特征去构建具有广泛适用性与较高预测性能的预测方法, 将为预测研究提供一个新的创新思路 (Tang et al., 2014, 2015a)。在这一背景下, 本书研究工作将紧扣研究样本的数据特征 (Tang et al., 2013), 提出“数据特征驱动建

模 (data-characteristic-driven modeling)”思想,并在基于“先分解后集成”思想的分解集成模型基础上,开展时序预测理论与方法创新工作,提高模型的分析与预测性能。

## 1.2 国内外研究现状

时序预测一直是各研究领域重要的热点与难点问题,现有研究构建了各式各样的预测方法,试图提高对其研究对象的预测精度。文献综述将分别从两个角度展开讨论:首先,从主流预测方法出发,对广泛应用于各研究领域的主流时序预测技术进行全面的梳理工作;其次,以清洁能源与国家风险两个新兴研究领域为例,从特定研究对象出发,讨论其预测方法的演化过程。

### 1.2.1 主流时序预测技术研究综述

根据现有文献,时序预测研究从时间维度上,大致经历了3个阶段:基于传统计量模型的时序预测研究,基于人工智能算法的时序预测研究,以及近年来新兴的基于混合模型的时序预测研究。相应地,时序预测方法可大致分为3个类别,即传统计量模型、人工智能模型,以及集成两个或两个以上方法的混合模型(集成模型),它们在函数形式与模型估计等方面存在明显差异。值得注意的是,由于非参数模型存在预测外推的局限性(Iwata et al., 2003),本书所讨论的预测方法均为参数模型,通过引入有限个具有实际经济意义的实参数以体现输入输出变量间的相互关系,实现时序数据的分析与预测。

依次分析三类时序预测方法,包括传统计量模型、人工智能模型及混合模型(集成模型);在此基础上,总结各类预测方法的特点及优劣势,并相应得出一些重要启示,以支撑本书研究工作的开展。

#### 1. 传统计量模型

在时序预测研究中,传统计量模型是一类发展比较成熟的建模技术。传统计量模型,也称为经济计量模型、计量经济模型或计量统计模型,实质上是统计学在经济学中的应用,是经济学、数学和统计学的结合。具体地,传统计量模型以统计学知识为基础,用随机性的数学方程并引入实参数来描述时序内部各组成要素间的定量关系,也常被称为参数模型(Sims, 1980)。其中,传统计量模型常用的时序模型主要包括:指数平滑(exponential smoothing, ES)、自回归移动平均单整[auto regressive (AR) moving average (MA), ARMA]、自回



归移动平均单整 (auto regressive integrated moving average, ARIMA) 及状态空间结构等基础模型, 以及其扩展模型, 如季节 ARIMA (seasonal ARIMA, SARIMA)、X12-ARIMA 模型、自回归移动平均分整 (ARFIMA)、自回归条件异方差 (AR conditional heteroscedastic, ARCH) 及广义 ARCH (generalized ARCH, GARCH) 等模型。

### 1) ES 模型

ES 模型在统计学理论上的研究始于 20 世纪五六十年代, 最初以加法或乘法形式考虑时序数据的趋势与季节模式 (或噪声模式), 即简单指数平滑 (simple ES, SES) 模型, 实现数据简单且有效的外推 (Brown, 1963; Holt, 1957; Winters, 1960; Pegels, 1969)。随后, 通过对参数估计、初始条件及外推方式的调整与修订, 提出系列新的 ES 模型 (Ledolter and Abraham, 1984; McClain, 1988; Sweet and Wilson, 1988); 另外, Taylor (2003) 提出了一种新的 ES 模型, 即衰减的 ES 模型 (damped multiplicative model)。基于以上研究, 以 5 种不同形式的趋势 (无趋势、加法趋势、衰减加法趋势、乘法趋势、衰减乘法趋势) 与 3 种不同形式的季节模式 (无季节模式、加法季节模式与乘法季节模式) 可构建出目前常用的 15 种 ES 模型。其中, 应用最为广泛的 ES 模型有简单 ES 与二次 ES 模型 (无趋势无季节模式)、Holt-Winter 非季节模型 (加法趋势无季节模式)、Holt-Winter 季节乘积模型 (加法趋势乘法季节模式) 等 (Hyndman et al., 2005)。

关于 ES 模型的统计理论基础, 大量文献指出 ES 模型在统计意义上实为 ARIMA 模型中的一种特例 (Abraham and Ledolter, 1986), 也具有状态空间模型的统计特征 (Snyder, 1985)。

### 2) ARIMA 模型

在早期研究中, 时序数据被认为是由确定性的系统产生, 直到 Yule (1927) 提出随机性 (stochasticity) 概念, 并指出时序数据中存在不确定性因素。基于此概念, 研究构建了一系列新的时序模型, 而最著名的工作涉及 Sultsky 等提出的 AR 与 MA 概念 (Hyndman and Gooijer, 2006)。此后, 时序研究步入设计与求解线性方程的计量经济时代, 时序预测问题被相应地分解为线性函数设计、参数估计、模型检验与外推预测等统计理论在时序分析中的应用 (Kolmogorov, 1941; Newbold, 1983)。1970 年, Box 和 Jenkins 等对当时的时序研究进行了总汇, 出版了书籍《时序分析: 预测与控制》, 其中包括了 ARMA 模型及引入差分平稳概念的 ARIMA。从此, ARIMA 模型, 也被称为 Box-Jenkins 模型, 成为时序预测研究中的经典方法, 而随后诸多模型的提出与构建都是基于 ARMA 或 ARIMA 模型的修正与扩展工作。