



Study on Description and Prediction
for the Behavior of
China's Securities Markets

**中国证券市场行为
描述与预测**

**不确定时态数据挖掘方法
及其在证券市场中的应用**

Uncertainty Temporal Data Mining Method and
Its Application in the Securities Markets

谭华 谢赤 著

by

TAN Hua, XIE Chi



Study on Description and Prediction
for the Behavior of
China's Securities Markets

中国证券市场行为 描述与预测

责任编辑：胡捷晖
封面设计：殷健
刘卉

ISBN 978-7-5357-5192-8



9 787535 751928 >

共五本套价：300.00 元

不确定时态数据挖掘方法 及其在证券市场中的应用

Uncertainty Temporal Data Mining Method and
Its Application in the Securities Markets

谭华 谢赤 著
by
TAN Hua, XIE Chi



Study on Description and Prediction
for the Behavior of
China's Securities Markets

**中国证券市场行为
描述与预测**

图书在版编目 (CIP) 数据

不确定时态数据挖掘方法及其在证券市场中的应用 /
谭华, 谢赤主编. —长沙: 湖南科学技术出版社, 2009. 5
(中国证券市场行为描述与预测)
ISBN 978-7-5357-5192-8

I. 不… II. ①谭…②谢… III. 数据采集—应用—证券
交易—资本市场—中国 IV. F832. 51

中国版本图书馆 CIP 数据核字 (2009) 第 068498 号

中国证券市场行为描述与预测

不确定时态数据挖掘方法及其在证券市场中的应用

著 者: 谭 华 谢 赤

责任编辑: 胡捷晖 刘 卉

出版发行: 湖南科学技术出版社

社 址: 长沙市湘雅路 276 号

<http://www.hnstp.com>

印 刷: 湖南天闻新华印务邵阳有限公司

(印装质量问题请直接与本厂联系)

厂 址: 邵阳市双坡岭

邮 编: 422001

出版日期: 2009 年 5 月第 1 版第 1 次

开 本: 700mm×1000mm 1/16

印 张: 12.75

字 数: 205000

书 号: ISBN 978-7-5357-5192-8

共五本套价: 300.00 元

(版权所有 · 翻印必究)

摘 要

随着金融全球化与金融自由化的推进，金融行业的运行效率与水平在很大程度上决定了一个国家的经济竞争力，而信息化技术越来越成为影响金融行业创新能力与发展水平的重要因素。近年来，许多金融机构开始运用先进的信息技术与智能决策支持技术，对业务系统中积累的海量数据进行深入分析，以发现各种有价值的规律。数据挖掘技术作为一种新兴的智能决策支持技术，已经开始在金融行业的部分领域得到应用。在此背景下，研究如何从金融数据中挖掘出各种信息，更好地认识、掌握并利用其规律，无疑具有重要意义。

与此同时，金融市场的信息具有众多不确定性、非线性，其信息数据具有模糊性及非结构性等特点。金融市场中的不确定性还包含时间的不确定。这些问题都值得研究者进行深入的研究。

不确定性方法与数据挖掘技术有一定的重叠性，两者在单独使用时都具有一定局限性。数据挖掘中存在不确定性问题，金融时间序列分析中也存在不确定性问题，并且传统数理统计方法不适用于从大量的数据中主动发现各种潜在规则，而不确定性方法在单独进行预测时会遇到小数据量等问题。

本书根据具体选定的证券市场行情作为预测对象，将不确定性方法和数据挖掘技术的各自优势结合起来，得到一种基于不确定性方法和数据挖掘技术的不确定时态数据挖掘方法（UTDM）。该方法能更好地发挥不确定性方法和数据挖掘技术的优势，为证券市场的预测提供更好的技术分析方法，从而为投资决策者提供更为精确的定量分析结果。

在对不确定性方法和数据挖掘技术的相关理论、研究发展现状及不足进行深入讨论的基础上，本书选取不确定性方法和数据挖掘技术中几

种具有代表性的方法：在不确定性方法中选取模糊集方法、模糊相似关系下的模糊粗糙集及灰色理论；在数据挖掘中选取关联规则、神经网络等方法，用于构建证券市场的有效分析方法。在此基础上分别得到模糊相似关系下的模糊粗糙集挖掘预测方法、趋势特征挖掘预测方法、时间序列模糊关联规则挖掘预测方法及多灰色神经网络预测方法，用这些方法分别解决证券市场中短期的个股及股指的预测问题。研究的后续部分则以此为基础进行展开。

本书通过模糊相似关系下的模糊粗糙集和数据挖掘技术对股票价格进行预测研究，从证券市场的大量数据中得到强规则。利用模糊集和粗糙集方法将股票价格进行预分类，并按时间属性进行分组，通过给出的模糊相似关系下的模糊粗糙集计算每组的真值，利用数据挖掘技术获得候选属性，最终得到相应时间段内的有用规则，根据所得规则预测某一具体时间段内股票价格的变化趋势。将模糊粗糙集扩展到模糊相似关系下的模糊粗糙集，并应用到对股票价格的预测中，能较原模糊粗糙集方法得到更多的有用规则，准确率更高。

本书将股票中的时间序列转换为以价格变动率为变量的时间序列进行分析，并对趋势特征提取、聚类算法进行改进，将时间序列的预测问题转化为频繁和有效特征集来发现问题，进而对趋势特征模式进行挖掘预测，根据连续一段时间内的涨跌情况判断市场的发展趋势。

同时，本书将时间序列模糊关联规则应用于证券市场的交易规则抽取。选用聚类方法对模糊集属性进行离散化，构造模糊集和隶属函数，引入时间维度，提出适合股票交易规则抽取的时间序列模糊关联规则算法，对一定时间段内股票间及行业间的关联关系进行最大限度的挖掘预测。

此外，本书提出将三种灰色预测模型，即残差 $GM(1, 1)$ 、无偏 $GM(1, 1)$ 和 $pGM(1, 1)$ 与神经网络预测模型有机组合起来，建立一种新的多灰色神经网络组合预测方法，并通过对中国证券市场综合指数的模拟预测进行验证，对证券市场综合指数的预测及比较说明了该组合预测精度的有效性。

Abstract

With the development of financial globalization and liberalization, the efficiency of the financial sector determines the level of a country's economic competitiveness at a large extent, and information technology is one of the important factors to impact the financial industry innovation ability and development level. Many financial institutions use more advanced information technology and intelligent decision support technology to find out useful rules by analyzing massive data, which are accumulated in operational systems. As a new intelligent decision support technology, data mining technology has been used in some fields of financial sector. Based on these conditions, how to mining more useful information from financial data to understand, master, and use its own rules and undoubtedly has special significance.

The information of financial market has the characteristics of uncertainty, nonlinear, fuzzy nature of the information data and non-structure. The uncertainty in financial market, not only includes the uncertainty of time, but also the uncertainty of information and technology. All of these are worthy of deep research.

Uncertainty methods and data mining technology have similar usages to some extent, but they have some limitations when we use them alone. As we know, it has uncertainty in data mining, and the analysis of financial time series also has uncertainty problems. Furthermore, traditional mathematical statistics methods don't fit to find potential rules from large amounts of data.

In this book, we mainly study uncertain knowledge and data mining technology to propose a new method-uncertainty temporal data mining method (UTDM), from which we can obtain a series of new methods, and use the new methods to solve the problems we meet during the mining and forecasting processes. This book use these methods to study securities markets, primarily for the securities markets' trends, stock price forecasts, and stock index forecasts.

This book chooses some typical methods in uncertainty methods and data mining technology to setup useful methods to analyie securities markets: fuzzy sets, fuzzy rough sets based on fuzzy similar relationship, gray theory, association rules, networks method. This book summarize the related theories and methods of uncertain knowledge, data mining technology and financial time series, point out the limitations of the current studies, lay the foundation for uncertainty temporal data mining method, clarify the objections of this study, and give us important informations to make the further theoretical research.

This book uses fuzzy similar relationship through the rough sets and fuzzy data mining techniques to predict stock prices, obtain strong rules from the securities markets and economic data. This book uses fuzzy rough sets and data mining technology to forecast stock price at certain given time. At first, this book uses fuzzy sets and rough sets to make the stock price into some groups based on the attribute "time", and then compute each true value by the fuzzy rough sets, this book obtains the candidate properties by data mining method. In the end, this book gets the useful rules and forecast the change trading of stock price in a certain time. The result of the method this book pointed out is more exact than other methods.

This book shares the time series into a variable rate of change in prices of the time series analysis, improve the trend of feature extraction and clustering algorithm, transfore time series prediction into fre-

quent and effective feature sets to discover problems to mining forecast rules, judge the market trends according to the change in market during a period of time.

In order to assist stock investors to make reasonable decision, this book uses fuzzy association rules to mining securities markets' exchange rules. Firstly, the fuzzy sets' attributes are dispersed by clustering method. Secondly, fuzzy sets and their corresponding membership function of the quantitative attributes are generated by means of the methods. Finally, the algorithm is named as FARS. At the end of the 6th chapter, the fuzzy association rules are studies are combined is this book based on temporal forms.

Three gray models: residual GM(1, 1), unbiased GM(1, 1), p GM(1, 1) and neural network to propose a new combination forecasting model. And use it to make forecasts on Composite Stock Price Index of the securities markets in Shanghai, China. The results show that this model could gain optimized forecasting value and could be taken as an effective tool to predict Shares Price Composite Index. The model makes full use of grey prediction modeling information which requires less information and the neural network has strong ability and good nonlinear mapping and fault-tolerant, self-organizing and adaptive characteristics. The forecasts on Composite Stock Price Index of the securities markets in China shows that the combination of the effectiveness of prediction accuracy.

目 录

插图索引

附表索引

第 1 章 绪论	1
1.1 选题背景及课题来源	1
1.2 研究意义	3
1.3 主要工作及创新点	4
1.4 研究思路与内容	5
第 2 章 相关研究基础与文献综述	9
2.1 数据挖掘技术	9
2.1.1 数据挖掘模型的分类	10
2.1.2 数据挖掘的主要方法和算法	12
2.1.3 时态数据挖掘	17
2.2 金融时间序列分析方法研究	24
2.2.1 金融时间序列的传统分析方法	24
2.2.2 混合方法	26
2.2.3 传统时间序列分析挖掘方法的不足	28
2.3 不确定性及其方法	28
2.3.1 不确定性的定义	29
2.3.2 造成不确定性的原因	31
2.3.3 不确定性方法	32
2.4 本章小结	35
第 3 章 证券市场预测需求下新型数据挖掘方法的构建	36
3.1 金融市场可预测性	36

3.2 证券市场行情预测总体分析	38
3.2.1 证券市场预测的分类	41
3.2.2 研究对象的选择	42
3.2.3 证券市场科学预测的研究方法	43
3.3 不确定时态数据挖掘方法的构建	52
3.3.1 数据挖掘技术的应用领域	53
3.3.2 时间序列挖掘的研究领域	55
3.3.3 不确定时态数据挖掘方法的基本思路	59
3.4 证券市场行情预测研究方案的确定	60
3.5 本章小结	62
第4章 模糊相似关系下模糊粗糙集对股价的预测	63
4.1 模糊粗糙集理论基础	64
4.1.1 粗糙集理论基本概念	65
4.1.2 模糊集理论基本概念	68
4.1.3 模糊相似关系	70
4.2 模糊相似关系下的模糊粗糙集	72
4.2.1 模糊粗糙集近邻算法	73
4.2.2 基于 FRNN 算法的聚类分类算法	74
4.3 利用模糊粗糙集对证券市场进行预测	75
4.3.1 股票价格序列预定义	75
4.3.2 利用模糊粗糙集方法挖掘规则过程	77
4.3.3 模糊粗糙集规则挖掘	78
4.3.4 测试结果	80
4.4 本章小结	80
第5章 趋势特征挖掘对股价突变的预测	82
5.1 时间序列相似性模式挖掘研究	82
5.1.1 相似性度量	83
5.1.2 欧氏距离的定义	85
5.2 趋势特征挖掘预测方法构造过程	87
5.2.1 特征提取平滑处理	88

5.2.2 趋势特征抽取算法.....	88
5.2.3 趋势特征聚类算法选择.....	90
5.2.4 特征模式预测方法发现过程.....	91
5.3 趋势特征挖掘方法在证券市场中的应用.....	92
5.4 测试结果.....	94
5.5 本章小结.....	96
第6章 股票间时间序列模糊关联规则挖掘预测	97
6.1 模糊数据挖掘技术.....	97
6.2 时间序列关联规则挖掘方法.....	99
6.2.1 Apriori 算法	100
6.2.2 时间序列关联规则挖掘基本步骤	102
6.3 证券市场中时间序列模糊预处理过程	103
6.3.1 模糊集核与支集的定义	104
6.3.2 模糊集的隶属函数构造	105
6.3.3 聚类方法的选择	105
6.3.4 时间序列的模糊离散化	106
6.4 时间序列模糊关联规则挖掘算法构造	108
6.5 时间序列模糊关联规则在证券市场中的应用	109
6.5.1 时间序列数据预处理	110
6.5.2 规则抽取结果分析	112
6.6 本章小结	113
第7章 多灰色神经网络组合方法对股指的预测	115
7.1 灰色系统理论及其技术	116
7.1.1 灰色系统理论的形成	116
7.1.2 灰色预测分类	117
7.1.3 灰色理论在证券市场分析中的应用	118
7.1.4 灰色理论在证券市场分析应用中的不足	119
7.2 神经网络模型	120
7.2.1 基本人工神经元及其网络	121
7.2.2 神经网络的学习与训练	122

7.3 多灰色神经网络组合预测模型建模过程	123
7.3.1 带残差修正的灰色 GM(1, 1) 模型	125
7.3.2 无偏 GM(1, 1) 模型	127
7.3.3 p GM(1, 1) 模型	128
7.3.4 BP 神经网络	129
7.3.5 多灰色神经网络组合预测模型	134
7.4 组合预测方法实证分析	136
7.4.1 灰色模型进行预测	137
7.4.2 BP 神经网络模型进行预测	138
7.4.3 多灰色神经网络模型预测结果分析	138
7.5 本章小结	142
结 论	144
附录 相关算法部分程序源代码	147
参考文献	175

插图索引

图 1.1	研究结构图	6
图 2.1	不确定性的描述	30
图 3.1	不确定性方法、数据挖掘技术及金融时间序列的关系	59
图 3.2	证券市场行情预测的总体方案设计	61
图 5.1	影响序列相似性的常见因素	84
图 5.2	中牧股份股价变动时间序列	93
图 5.3	特征数量变化曲线	95
图 5.4	预测时段与预测精度曲线	95
图 7.1	一般前向神经网络结构	122
图 7.2	有指导学习原理图	122
图 7.3	人工神经元模型	129
图 7.4	组合预测模型结构图	135

附表索引

表 2.1	几种不确定性方法的比较	34
表 4.1	变化率的粗集分类	76
表 4.2	预分类及分类等级值	79
表 4.3	2006 年 5 月 10 日中牧股份预分类及分类等级值.....	79
表 4.4	13:30 时刻的候选属性.....	79
表 4.5	5 分钟数据测试结果.....	80
表 5.1	各时间段的相应趋势特征	93
表 6.1	股票行情的数据结构	110
表 6.2	中牧股份交易行情基本数据	110
表 6.3	部分预处理后的数据	111
表 6.4	部分模糊离散化结果	111
表 6.5	规则抽取部分实验结果	112
表 6.6	2007 年 11 月 5~25 日期间的部分具体规则	112
表 7.1	预测精度等级划分	127
表 7.2	上证综指月均指数各模型模拟预测值	138
表 7.3	深证成指月均指数各模型模拟预测值	140
表 7.4	上证综指月均指数模型检验与评价	141
表 7.5	深证成指月均指数模型检验与评价	141

第1章 绪 论

1.1 选题背景及课题来源

金融市场是国家经济运行系统的核心。探求金融市场的变化规律，进行有效的金融管理，提高金融投资效率，是各国政府与投资机构孜孜以求的主要目标。随着金融全球化与自由化的推进，金融行业的运行效率与水平在很大程度上决定了一个国家的经济竞争力，而信息化技术则成为影响金融行业创新能力与发展水平的一个重要因素。

进入 20 世纪 90 年代，国际金融业的竞争日趋激烈，政策分析与决策制定对数据的依赖性和敏感度也越来越高。在此背景下，许多金融机构开始运用更加先进的信息技术与智能决策支持技术，对业务系统中积累的海量数据进行深入分析，以发现各种有价值的规律。

作为一种新兴的智能决策支持手段，数据挖掘技术开始在金融行业的部分领域得到应用，如客户分析、欺诈行为检测、资产价格预测以及投资组合建议等^[1~3]。例如，法兰克福银行使用数据挖掘技术发现投资者的金融和商业兴趣，并对投资者进行分类信用评估与欺诈检测，从而确保利润并降低风险；美洲银行采用数据挖掘工具分析信用卡交易数据，从而全面掌握客户的消费情况和偏好，为其提供个性化的服务；由 JSA 公司开发的 EPOS 系统不仅具有良好的预测功能，而且能够对资本市场的数据进行分类，帮助用户构建合理的投资组合。

当前，中国经济与世界经济相互融合的程度越来越深，金融业面临着更大、更新的发展机遇和挑战。与此同时，全球金融创新活动日新月异，各国金融市场联动效应不断增强。因此，对金融市场本质规律的认识和把握，将直接关系到金融市场的稳定、高效与安全运行。

中国金融业经过十几年的高速发展，已经形成了相当规模的市场。金融机构的许多业务活动都越来越依赖于对大量历史数据的分析。国内投资者与金融机构也越来越清楚地认识到，分析金融数据，从中挖掘出有价值的信息，是其实现科学化管理决策的必要手段与基础工作。

金融市场是一个受多种因素影响的、庞大的系统，具有非常复杂的运行规律，而金融数据则是其综合表现形式。“本质决定现象，现象反映本质”，数据中必定蕴含了金融系统许多客观规律信息。以商业银行为例，它们已相继开发出比较完善的会计系统、储蓄系统、国际结算系统、资金系统、信贷系统、信用卡系统、办公自动化系统等，管理手段得到很大的提高。同时，它们充分利用现代科技的发展成果，追求利润和市场的最大化，向集中化处理、电子化处理和服务创新等方向发展。目前，许多数据库系统正逐步建立起来。基于这些条件，如何从金融数据中挖掘出各种信息，更好地认识、掌握并利用其规律，无疑具有重要的意义。

中国金融业在信息化建设上发展非常迅速。还是以银行业为例，20世纪80年代初开始采用计算机进行业务处理，90年代开始实现各银行内部的联网，进而实现各银行间、银行与商业机构间的数据交换。目前，已基本实现数据集中，即银行将全国各地各网点的数据统一集中到一起建立中心数据库。各个金融机构基本上都建立了数据中心，金融业已经成为一个数据高度密集的行业。随着数据大集中的逐步深入，结构化的数据和非结构化的数据搭建起来的统一数据平台，为在金融行业应用数据挖掘技术打下坚实的基础。在此基础之上，能够提供关于宏观政策环境、竞争对手、市场需求等信息，包括企业商业智能、客户关系管理、市场销售分析、财务、交易等相关战略的数据挖掘应用。

从金融时间序列中获取信息的传统做法是基于统计理论的金融时间序列分析方法，它是现代金融计量经济学理论中的重要内容，通过模型假设、参数估计、模型检验等手段和技术，获得描述金融时间序列规律的数学模型。现在的金融时间序列是由大量的数据组成的序列，已有的金融计量经济理论已经不能满足对金融时间序列数据的特征模式等的提取，而数据挖掘技术则正好符合了这一要求。