

量化投资：数据挖掘技术与实践（MATLAB版）



[量化投资：数据挖掘技术与实践（MATLAB版）_下载链接1_](#)

著者:卓金武

出版者:电子工业出版社

出版时间:2015-5

装帧:平装

isbn:9787121259265

全书内容分为三篇。第一篇（基础篇）主要介绍数据挖掘与量化投资的关系，以及数据挖掘的概念、实现过程、主要内容、主要工具等内容。第二篇（技术篇）系统介绍了数据挖掘的相关技术及这些技术在量化投资中的应用，主要包括数据的准备、数据的探索

、关联规则方法、数据回规方法、分类方法、聚类方法、预测方法、诊断方法、时间序列方法、智能优化方法等内容。第三篇（实践篇）主要介绍数据挖掘技术在量化投资中的综合应用实例，包括统计套利策略的挖掘与优化、配对交易策略的挖掘与实现、数据挖掘在股票程序化交易中的综合应用，以及基于数据挖掘技术的量化交易系统的构建。

《量化投资：数据挖掘技术与实践（MATLAB版）》的读者对象为从事投资、数据挖掘、数据分析、数据管理工作的专业人士；金融、经济、管理、统计等专业的教师和学生；希望学习MATLAB的广大科研人员、学者和工程技术人员。

作者介绍:

卓金武，中国量化投资学会专家委员会成员，MathWorks中国区数据挖掘和量化投资总监，主要职责是向中国区MATLAB正版用户提供数据挖掘和量化投资解决方案。曾2次获全国大学生数学建模竞赛一等奖（2003，2004），1次获全国研究生数学建模竞赛一等奖（2007）；已主编专著两部：《MATLAB在数学建模中的应用》（第一版和第二版）。

周英，现就职于某知名搜索引擎公司，主要从事移动搜索引擎的研发，研究方向为互联网数据挖掘和互联网金融。曾获美国大学生数学建模竞赛二等奖一项，全国研究生数学建模竞赛二等奖一项。著有专著《大数据挖掘的技术与实践》。

目录: 第一篇基础篇

第1章绪论 2

1.1 量化投资与数据挖掘的关系 2

1.1.1 什么是量化投资 2

1.1.2 量化投资的特点 3

1.1.3 量化投资的核心——量化模型 5

1.1.4 量化模型的主要产生方法——数据挖掘 7

1.2 数据挖掘的概念和原理 8

1.2.1 什么是数据挖掘 8

1.2.2 数据挖掘的原理 10

1.3 数据挖掘在量化投资中的应用 11

1.3.1 宏观经济分析 11

1.3.2 估价 13

1.3.3 量化选股 14

1.3.4 量化择时 14

1.3.5 算法交易 15

1.4 本章小结 16

参考文献 16

第2章数据挖掘的内容、过程及工具 17

2.1 数据挖掘的内容 17

2.1.1 关联 17

2.1.2 回归 19

2.1.3 分类 20

2.1.4 聚类 21

2.1.5 预测 22

2.1.6 诊断 24

2.2 数据挖掘过程 25

2.2.1 数据挖掘过程概述 25

2.2.2 挖掘目标的定义 26

2.2.3 数据的准备 26

2.2.4 数据的探索 28

2.2.5 模型的建立	30
2.2.6 模型的评估	34
2.2.7 模型的部署	35
2.3 数据挖掘工具	36
2.3.1 MATLAB	36
2.3.2 SAS	37
2.3.3 SPSS	38
2.3.4 WEKA	40
2.3.5 R	41
2.3.6 工具的比较与选择	42
2.4 本章小结	43
参考文献	44
第二篇技术篇	
第3章数据的准备	47
3.1 数据的收集	47
3.1.1 认识数据	47
3.1.2 数据挖掘的数据源	49
3.1.3 数据抽样	50
3.1.4 量化投资的数据源	51
3.1.5 从雅虎获取交易数据	53
3.1.6 从大智慧获取财务数据	56
3.1.7 从Wind获取高质量数据	57
3.2 数据质量分析	59
3.2.1 数据质量分析的必要性	59
3.2.2 数据质量分析的目的	60
3.2.3 数据质量分析的内容	60
3.2.4 数据质量分析的方法	61
3.2.5 数据质量分析的结果及应用	66
3.3 数据预处理	67
3.3.1 为什么需要数据预处理	67
3.3.2 数据预处理的主要任务	68
3.3.3 数据清洗	69
3.3.4 数据集成	73
3.3.5 数据归约	74
3.3.6 数据变换	74
3.4 本章小结	77
参考文献	77
第4章数据的探索	78
4.1 衍生变量	79
4.1.1 衍生变量的定义	79
4.1.2 变量衍生的原则和方法	80
4.1.3 常用的股票衍生变量	80
4.1.4 评价型衍生变量	85
4.1.5 衍生变量数据收集与集成	87
4.2 数据的统计	88
4.2.1 基本描述性统计	89
4.2.2 分布描述性统计	90
4.3 数据可视化	90
4.3.1 基本可视化方法	91
4.3.2 数据分布形状可视化	92
4.3.3 数据关联情况可视化	94
4.3.4 数据分组可视化	95
4.4 样本选择	97
4.4.1 样本选择的方法	97

4.4.2 样本选择应用实例	98
4.5 数据降维	100
4.5.1 主成分分析 (PCA) 基本原理	100
4.5.2 PCA应用案例：企业综合实力排序	103
4.5.3 相关系数降维	106
4.6 本章小结	107
参考文献	108
第5章关联规则方法	109
5.1 关联规则概要	109
5.1.1 关联规则提出背景	109
5.1.2 关联规则的基本概念	110
5.1.3 关联规则的分类	112
5.1.4 关联规则挖掘常用算法	113
5.2 Apriori算法	113
5.2.1 Apriori算法的基本思想	113
5.2.2 Apriori算法的步骤	114
5.2.3 Apriori算法的实例	114
5.2.4 Apriori算法的程序实现	117
5.2.5 Apriori算法的优缺点	120
5.3 FP-Growth算法	121
5.3.1 FP-Growth算法步骤	121
5.3.2 FP-Growth算法实例	122
5.3.3 FP-Growth算法的优缺点	124
5.4 应用实例：行业关联选股法	124
5.5 本章小结	126
参考文献	127
第6章数据回归方法	128
6.1 一元回归	129
6.1.1 一元线性回归	129
6.1.2 一元非线性回归	133
6.1.3 一元多项式回归	138
6.2 多元回归	138
6.2.1 多元线性回归	138
6.2.2 多元多项式回归	142
6.3 逐步回归	145
6.3.1 逐步回归的基本思想	145
6.3.2 逐步回归步骤	146
6.3.3 逐步回归的MATLAB方法	147
6.4 Logistic回归	149
6.4.1 Logistic模型	149
6.4.2 Logistic回归实例	150
6.5 应用实例：多因子选股模型的实现	153
6.5.1 多因子模型的基本思想	153
6.5.2 多因子模型的实现	154
6.6 本章小结	157
参考文献	157
第7章分类方法	158
7.1 分类方法概要	158
7.1.1 分类的概念	158
7.1.2 分类的原理	159
7.1.3 常用的分类方法	160
7.2 K-近邻 (KNN)	161
7.2.1 K-近邻原理	161
7.2.2 K-近邻实例	163

7.2.3 K-近邻特点	166
7.3 贝叶斯分类	167
7.3.1 贝叶斯分类原理	167
7.3.2 朴素贝叶斯分类原理	167
7.3.3 朴素贝叶斯分类实例	170
7.3.4 朴素贝叶斯特点	170
7.4 神经网络	171
7.4.1 神经网络的原理	171
7.4.2 神经网络的实例	173
7.4.3 神经网络的特点	174
7.5 逻辑斯蒂 (Logistic)	175
7.5.1 逻辑斯蒂的原理	175
7.5.2 逻辑斯蒂的实例	175
7.5.3 逻辑斯蒂的特点	175
7.6 判别分析	176
7.6.1 判别分析的原理	176
7.6.2 判别分析的实例	177
7.6.3 判别分析的特点	177
7.7 支持向量机 (SVM)	178
7.7.1 SVM的基本思想	178
7.7.2 理论基础	179
7.7.3 支持向量机的实例	182
7.7.4 支持向量机的特点	182
7.8 决策树	183
7.8.1 决策树的基本概念	183
7.8.2 决策树的建构的步骤	184
7.8.3 决策树的实例	187
7.8.4 决策树的特点	188
7.9 分类的评判	188
7.9.1 正确率	188
7.9.2 ROC曲线	191
7.10 应用实例：分类选股法	193
7.10.1 案例背景	193
7.10.2 实现方法	194
7.11 延伸阅读：其他分类方法	197
7.12 本章小结	197
参考文献	198
第8章 聚类方法	199
8.1 聚类方法概要	200
8.1.1 聚类的概念	200
8.1.2 类的度量方法	201
8.1.3 聚类方法的应用场景	203
8.1.4 聚类方法的分类	204
8.2 K-means方法	205
8.2.1 K-means的原理和步骤	205
8.2.2 K-means实例1：自主编程	206
8.2.3 K-means实例2：集成函数	208
8.2.4 K-means的特点	212
8.3 层次聚类	212
8.3.1 层次聚类的原理和步骤	212
8.3.2 层次聚类的实例	214
8.3.3 层次聚类的特点	217
8.4 神经网络聚类	217
8.4.1 神经网络聚类的原理和步骤	217

8.4.2 神经网络聚类的实例	218
8.4.3 神经网络聚类的特点	219
8.5 模糊C-均值 (FCM) 方法	219
8.5.1 FCM的原理和步骤	219
8.5.2 FCM的应用实例	220
8.5.3 FCM算法的特点	221
8.6 高斯混合聚类方法	222
8.6.1 高斯混合聚类的原理和步骤	222
8.6.2 高斯聚类的实例	224
8.6.3 高斯聚类的特点	225
8.7 类别数的确定方法	225
8.7.1 类别的原理	225
8.7.2 类别的实例	227
8.8 应用实例：股票聚类分池	229
8.8.1 聚类目标和数据描述	229
8.8.2 实现过程	229
8.8.3 结果及分析	231
8.9 延伸阅读	233
8.9.1 目前聚类分析研究的主要内容	233
8.9.2 SOM智能聚类算法	234
8.10 本章小结	235
参考文献	235
第9章预测方法	236
9.1 预测方法概要	236
9.1.1 预测的概念	236
9.1.2 预测的基本原理	237
9.1.3 量化投资中预测的主要内容	238
9.1.4 预测的准确度评价及影响因素	239
9.1.5 常用的预测方法	240
9.2 灰色预测	241
9.2.1 灰色预测原理	241
9.2.2 灰色预测的实例	243
9.3 马尔科夫预测	246
9.3.1 马尔科夫预测的原理	246
9.3.2 马尔科夫过程的特性	247
9.3.3 马尔科夫预测的实例	248
9.4 应用实例：大盘走势预测	252
9.4.1 数据的选取及模型的建立	252
9.4.2 预测过程	253
9.4.3 预测结果与分析	254
9.5 本章小结	255
参考文献	256
第10章诊断方法	257
10.1 离群点诊断概要	257
10.1.1 离群点诊断的定义	257
10.1.2 离群点诊断的作用	258
10.1.3 离群点诊断方法分类	260
10.2 基于统计的离群点诊断	260
10.2.1 理论基础	260
10.2.2 应用实例	262
10.2.3 优点与缺点	264
10.3 基于距离的离群点诊断	264
10.3.1 理论基础	264
10.3.2 应用实例	265

10.3.3 优点与缺点	267
10.4 基于密度的离群点挖掘	267
10.4.1 理论基础	267
10.4.2 应用实例	268
10.4.3 优点与缺点	270
10.5 基于聚类的离群点挖掘	270
10.5.1 理论基础	270
10.5.2 应用实例	271
10.5.3 优点与缺点	273
10.6 应用实例：离群点诊断量化择时	273
10.7 延伸阅读：新兴的离群点挖掘方法	275
10.7.1 基于关联的离群点挖掘	275
10.7.2 基于粗糙集的离群点挖掘	276
10.7.3 基于人工神经网络的离群点挖掘	276
10.8 本章小结	277
参考文献	277
第11章时间序列方法	279
11.1 时间序列的基本概念	279
11.1.1 时间序列的定义	279
11.1.2 时间序列的组成因素	280
11.1.3 时间序列的分类	281
11.1.4 时间序列分析方法	282
11.2 平稳时间序列分析方法	283
11.2.1 移动平均法	283
11.2.2 指数平滑法	284
11.3 季节指数预测法	285
11.3.1 季节性水平模型	285
11.3.2 季节性趋势模型	286
11.4 时间序列模型	286
11.4.1 ARMA模型	286
11.4.2 ARIMA模型	287
11.4.3 ARCH模型	288
11.4.4 GARCH模型	289
11.5 应用实例：基于时间序列的股票预测	289
11.6 本章小结	293
参考文献	293
第12章智能优化方法	294
12.1 智能优化方法概要	295
12.1.1 智能优化方法的概念	295
12.1.2 在量化投资中的作用	295
12.1.3 常用的智能优化方法	295
12.2 遗传算法	297
12.2.1 遗传算法的原理	297
12.2.2 遗传算法的步骤	298
12.2.3 遗传算法实例	306
12.2.4 遗传算法的特点	307
12.3 模拟退火算法	309
12.3.1 模拟退火算法的原理	309
12.3.2 模拟退火算法步骤	310
12.3.3 模拟退火算法实例	313
12.3.4 模拟退火算法的特点	319
12.4 应用实例：组合投资优化	320
12.4.1 问题描述	320
12.4.2 求解过程	320

12.5 延伸阅读：其他智能方法	321
12.5.1 粒子群算法	321
12.5.2 蚁群算法	323
12.6 本章小结	325
参考文献	325
第三篇实践篇	
第13章统计套利策略的挖掘与优化	327
13.1 统计套利策略概述	327
13.1.1 统计套利的定义	327
13.1.2 统计套利策略的基本思想	327
13.1.3 统计套利策略挖掘的方法	328
13.2 基本策略的挖掘	329
13.2.1 准备数据	329
13.2.2 探索交易策略	329
13.2.3 验证交易策略	330
13.2.4 选择最佳的参数	331
13.2.5 参数扫描法	334
13.2.6 考虑交易费	335
13.3 高频交易策略及优化	337
13.3.1 高频交易的基本思想	337
13.3.2 高频交易的实现	339
13.4 多交易信号策略的组合及优化	341
13.4.1 多交易信号策略	341
13.4.2 交易信号的组合优化机理	343
13.4.3 交易信号的组合优化实现	344
13.5 本章小结	347
参考文献	348
第14章配对交易策略的挖掘与实现	349
14.1 配对交易概述	350
14.1.1 配对交易的定义	350
14.1.2 配对交易的特点	350
14.1.3 配对选取步骤	351
14.2 协整检验的理论基础	352
14.2.1 协整关系的定义	352
14.2.2 EG两步协整检验法	353
14.2.3 Johansen协整检验法	353
14.3 配对交易的实现	355
14.3.1 协整检验的实现	355
14.3.2 配对交易函数	356
14.3.3 协整配对中的参数优化	359
14.4 延伸阅读：配对交易的三要素	360
14.4.1 配对交易的前提	360
14.4.2 配对交易的关键	360
14.4.3 配对交易的假设	360
14.5 本章小结	361
参考文献	361
第15章数据挖掘在股票程序化交易中的综合应用	362
15.1 程序化交易概述	362
15.1.1 程序化交易的定义	362
15.1.2 程序化交易的实现过程	363
15.1.3 程序化交易的分类	365
15.2 数据的处理及探索	366
15.2.1 获取股票日交易数据	366

15.2.2 计算指标	369
15.2.3 数据标准化	375
15.2.4 变量筛选	377
15.3 模型的建立及评估	379
15.3.1 股票预测的基本思想	379
15.3.2 模型的训练及评价	379
15.4 组合投资的优化	381
15.4.1 组合投资的理论基础	381
15.4.2 组合投资的实现	385
15.5 程序化交易的实施	389
15.6 本章小结	389
参考文献	390
第16章基于数据挖掘技术的量化交易系统	392
16.1 交易系统概述	393
16.1.1 交易系统的定义	393
16.1.2 交易系统的作用	393
16.2 DM交易系统总体设计	394
16.2.1 系统目标	394
16.2.2 相关约定	395
16.2.3 系统结构	395
16.3 短期交易子系统	396
16.3.1 子系统功能描述	396
16.3.2 数据预处理模块	396
16.3.3 量化选股模块	397
16.3.4 策略回测模块	397
16.4 中长期交易子系统	398
16.4.1 子系统功能描述	398
16.4.2 导入数据模块	398
16.4.3 投资组合优化模块	399
16.5 系统的拓展与展望	401
16.6 本章小结	401
参考文献	402
• • • • •	(收起)

[量化投资：数据挖掘技术与实践（MATLAB版）](#) [下载链接1](#)

标签

量化投资

投资

量化

数据挖掘

数学

金融

股票

经济

评论

真的是相当科普向……不过配合光盘可以当练习用

工具在次，思想最重要

翻了翻，感觉还不错

内容较多，但也有中学教科书式的罗嗦。例子太过简略，部分例子甚至不是投资方面的。适于了解些业内概念、算法，想要实用的量化算法，尚需很多在书本外的努力。

作为科普还是挺不错的…

[量化投资：数据挖掘技术与实践（MATLAB版）_下载链接1](#)

书评

抱着试一试的态度，去浏览了一遍，彻头彻尾的粗制滥造产品。
整本书大部分都是在粘贴定义，然后举的例子，几乎是不相关的例子，真的只是在说这个例子，没有策略分析之类的干货。
没什么代码，举的例子中只要稍微好一点的例子，都是没有代码的，这有个毛线参照意义。顺便再说一...

[量化投资：数据挖掘技术与实践（MATLAB版）_下载链接1](#)