

发现数据之美：数据分析原理与实践



[发现数据之美：数据分析原理与实践_下载链接1](#)

著者:彭鸿涛

出版者:电子工业出版社

出版时间:2014-8-1

装帧:

isbn:9787121235580

大数据时代已经来临，这将引起深刻的行业变革。但是，大数据的真意在于数据分析，即从繁多的数据中找出洞见，并将其应用于实际决策中，以产生更明智的决策。这是一

个看起来简单、做起来较难的事情。

本书从一个自底向上的角度，全面地阐述了数据分析所涉及的知识和技术，对于经典算法和工具的介绍也不止于泛泛而谈，而是加入了作者的经验 and 理解。所谓自底向上的角度，即从数据分析实践开始时所需要的数据准备、数据探查、数据再处理等，到经典的统计分析和数据挖掘算法及应用，还讲述了模型的部署，优化技术的引入，最终到决策自动化。

本书对企业管理者、数据分析从业者及高校的学生都有参考意义。管理者能看到一个较全面的数据分析的阐述，明确自身的需求；从业者能看到经验的总结及经典工具的使用；高校学生能看到数据分析所涉及的知识，对数据分析有一个全面的认识。

作者介绍:

目录: 第1章 业务分析是一个蓬勃发展的方向 1

1.1 业务分析是什么 2

1.2 业务分析的应用现状 3

1.3 如何应用业务分析 5

1.4 大数据与业务分析 8

1.5 我们还在等什么 9

第2章 开始我们的旅程——从数据谈起 10

2.1 我们讨论的数据结构 11

2.1.1 行 (Row) 是什么 12

2.1.2 列 (Column) 是什么 13

2.1.3 多少行数据才合适 15

2.1.4 我们需要什么样的列 16

2.2 Statistics 和 Modeler 的基本知识 18

2.3 数据导入 (Loading Data) 24

2.4 数据探查 (Data Exploring) 27

2.4.1 正态分布 (Normal Distribution) 28

2.4.2 数据探查的常见统计量 30

2.4.3 数据可视化 35

2.5 本章小结 47

第3章 在分析之前，还需要数据预处理 48

3.1 数据的问题 49

3.2 数据校验 50

3.2.1 验证规则 50

3.2.2 验证数据 53

3.2.3 数据审计 (Data Audit) 57

3.2.4 识别异常数据 60

3.3 数据集成 (Data Integration) 65

3.3.1 在 Statistics 中进行数据集成 66

3.3.2 在 Modeler 中进行数据集成 68

3.4 数据转换 (Data Transformation) 73

3.4.1 分箱 (Binning) 73

3.4.2 数据调整 (Data Rescale) 78

3.4.3 数据重新编码 (Recode) 79

3.5 自动数据准备 83

3.5.1 Statistics 中的自动数据准备 83

3.5.2 Modeler 中的自动数据准备 88

3.6 本章小结 89

第4章 经典分析——统计学的魅力 91

4.1 随机变量及分布	92
4.2 数理统计导引	94
4.3 参数估计	96
4.3.1 点估计	96
4.3.2 区间估计	97
4.4 假设检验	98
4.4.1 正态分布检验和t 检验	101
4.4.2 非参数检验	108
4.5 相关分析	111
4.6 方差分析	113
4.7 回归分析	114
4.7.1 线性回归分析	114
4.7.2 自动化线性回归分析	120
4.7.3 广义线性模型	122
4.7.4 广义线性混合模型 (Generalized Linear Mixed Mode, GLMM)	128
4.8 本章小结	135
第5章 我想预测未来	136
5.1 数据挖掘的技术分类	136
5.1.1 有监督的建模技术	137
5.1.2 无监督的建模技术	138
5.1.3 Feature Selection 对于分类的意义	139
5.1.4 查看建模的结果	139
5.2 决策树	140
5.2.1 C5.0 算法	141
5.2.2 分类和回归树	145
5.2.3 卡方自动交互检测法 (CHAID)	147
5.2.4 快速、无偏、高效的统计树 (QUEST)	148
5.2.5 交互式的决策树构建方式	149
5.3 决策表	150
5.3.1 决策表算法的设置	151
5.3.2 交互式决策表的生成方式	153
5.4 贝叶斯网络	154
5.4.1 一些基本概念	154
5.4.2 IBM SPSS 的做法	156
5.5 神经网络 (Neural Networks)	158
5.5.1 神经网络是什么	158
5.5.2 SPSS 神经网络算法	160
5.6 支持向量机 (Support Vector Machine)	162
5.6.1 什么是线性分类器	162
5.6.2 Modeler 中的支持向量机	163
5.7 最近相邻 (Nearest Neighbor)	165
5.8 我该选用哪种算法	167
5.9 如何评价预测结果	170
5.9.1 基本指标	170
5.9.2 Gains	171
5.9.3 Lift	173
5.9.4 Response	175
5.9.5 Profit	175
5.9.6 ROI	177
5.10 本章小结	177
第6章 我想发现聚类 (Cluster)	179
6.1 聚类技术	180
6.2 分层聚类	181
6.3 K-means	184

6.4 TwoStep	188
6.4.1 预聚类	189
6.4.2 离群值处理	189
6.4.3 聚类	189
6.4.4 TwoStep 的使用	190
6.5 Kohonen network	192
6.6 我怎么知道聚类结果是好的	194
6.6.1 考察聚类的数量和每个聚类中的记录数	194
6.6.2 考察聚类内的特征	195
6.6.3 考察聚类间的特征	195
6.6.4 一个综合的考察指标Silhouette	196
6.7 自动聚类	197
6.8 理解聚类的结果	198
6.9 一个聚类分析应用的例子	201
6.10 本章小结	202
第7章 周而复始的规律——时间序列分析	203
7.1 时间序列	204
7.1.1 时间序列的类型	204
7.1.2 时间序列的特征	205
7.2 指数平滑模型	206
7.2.1 简单指数平滑法	206
7.2.2 带有趋势调整的指数平滑法（霍尔特指数平滑法）	208
7.2.3 带有阻尼趋势的指数平滑法	208
7.2.4 简单季节指数平滑法	209
7.2.5 带有趋势和季节调整的指数平滑法（温特斯指数平滑法）	209
7.2.6 指数平滑法的初始化	210
7.2.7 去除时间序列的趋势和季节性因素	211
7.3 自回归模型	212
7.3.1 自回归模型	212
7.3.2 移动平均模型	213
7.3.3 自回归移动平均模型（ARMA）	213
7.3.4 差分自回归移动平均模型	214
7.4 SPSS 产品中的时间序列模型	214
7.4.1 Statistics 中的时间序列模型	214
7.4.2 Modeler 中的时间序列模型	235
7.5 时间序列分析的评价	238
7.6 本章小结	239
第8章 你的行为完全可能被猜中——关联规则分析	240
8.1 基本概念	241
8.2 Apriori 算法	245
8.2.1 Apriori 算法工作步骤	245
8.2.2 Apriori 算法的评估方法	246
8.2.3 Apriori 节点	247
8.3 CARMA 算法	249
8.3.1 CARMA 算法的工作步骤	249
8.3.2 CARMA 节点	251
8.4 序列算法	252
8.5 关联规则的评价	255
8.6 典型应用案例	256
第9章 我们还需要优化技术的帮忙	257
9.1 什么是优化技术	258
9.2 优化问题的分类	259
9.2.1 线性规划	260
9.2.2 整数规划	261

9.2.3 多目标规划	262
9.2.4 动态规划	262
9.3 IBM ILOG Optimization 介绍	263
9.4 本章小结	265
第10章 有关方法论的问题	266
10.1 为什么我们要讨论方法论	267
10.2 CRISP-DM	267
10.2.1 CRISP-DM 方法学	268
10.2.2 CRISP-DM 参考模型	270
10.3 IBM SPSS CaDS	273
10.3.1 Repository	273
10.3.2 Job	274
10.3.3 Model Refresh and Champion Challenger	274
10.3.4 Scoring	274
10.4 模型的部署不是终点	275
第11章 一个时髦的领域——决策管理	276
11.1 决策管理系统	276
11.1.1 什么是决策	277
11.1.2 什么是决策管理系统	279
11.1.3 决策支持与决策管理的比较	281
11.2 构建决策管理系统	282
11.2.1 构建决策支持系统的原则	282
11.2.2 合适的决策	283
11.2.3 如何找到合适的决策	285
11.2.4 怎样在决策管理系统中定义决策	287
11.2.5 决策管理系统中的优化技术	292
11.2.6 决策影响的评估	294
11.2.7 监控决策	297
11.2.8 决策的持续改进	298
11.2.9 构建和部署决策服务	299
11.2.10 实施决策管理的一些要求	300
11.3 IBM ADM	301
11.3.1 ADM 是SPSS 数据分析能力的窗口	301
11.3.2 ADM 的着眼点是将数据分析结果转化为决策	302
11.3.3 ADM 是一个可以配置的决策服务平台	303
11.3.4 ADM 的工作步骤	306
11.4 本章小结	308
后记 为未来做好准备	309
• • • • • (收起)	

[发现数据之美：数据分析原理与实践 下载链接1](#)

标签

数据分析

数据挖掘

spss

大数据

统计

科学

机器学习

数据科学

评论

从书名和简介你绝对看不出这就是本IBM的SPSS使用指南，坑爹

见过。不过是spss。

介绍了一个工具，随意看看还可以，走心就算了

SPSS新手指南……不过简单介绍了下现在流行的大数据算法和模型，对于科普来说是不错的

结合spss讲解的一本书。

看推荐借来看的，实际真的是spss软件的使用入门介绍，而且很多方法和数学方法菜鸟表示看不懂，强撑看完，受到了一万点伤害～

书评

就是一本介绍SPSS操作的书，题目取得太大了。卖软件的，不太适合搞数据分析的人。比如想数据分析，但是不想弄懂那些模型具体内涵，可以试试SPSS。。。。。。。