

<<SAS编程与数据挖掘商业案例>>

图书基本信息

书名：<<SAS编程与数据挖掘商业案例>>

13位ISBN编号：9787111305354

10位ISBN编号：7111305353

出版时间：2010-5

出版时间：机械工业出版社

作者：姚志勇

页数：345

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<SAS编程与数据挖掘商业案例>>

前言

当前国内的诸多数据挖掘书籍几乎都是基于理论说明，很少深入介绍数据挖掘实践，涉及SAS开发的更是少见。

因此，从商业应用出发，基于实践而不是基于理论的数据挖掘书籍呼之欲出。

本书作者从商业需求出发，以商业人士的眼光来看待企业数据挖掘，并给出大量的商业实践案例。

把主流的数据挖掘技术用真实案例来实现是本书出版的初衷，同时为了满足初学者需求，作者也给出了数据挖掘必备的基础编程知识模块。

全书共分两部分。

第一部分是SAS编程：第1章和第2章主要介绍SAS系统和编程基础，同时介绍SAS数据处理最核心的内容——数据指针和PDV流程。

该核心内容贯穿第一部分，是已出版的其他SAS图书没有的。

第3~9章主要介绍SAS的数据处理技术，也是第一部分的主要内容，包括数据集处理、变量处理和观测处理等多种数据处理技术，同时也介绍了循环控制等稍难的内容，重要的是给出了诸多实际案例及商业应用。

尽管第3~9章从表面上看和诸多已经出版的SAS图书没有什么大的不同，但是这些章节最大的亮点是作者对每一个示例和案例从数据指针和PDV流程的角度给予了最详细的程序解读，让读者真正读懂程序，而不是停留在程序的表面。

第10章是第一部分的难点。

作者还是站在商业实践的角度逐一介绍宏最常用的部分，同时也给出了非常详细的程序解读。

第11章介绍SQL过程。

有关内容在国内同类书中都出现过，但是作者独辟蹊径，融合了项目实践中诸多真正有用的语句，同时也给出了诸多开发建议和应注意的问题。

第12章介绍数据处理实践。

该章共包括四个方面的内容，几乎都是目前国内没有出现过的，如HASH对象及商业应用、正则表达式等。

随机抽样也是数据处理经常面临的问题，这里作者开发了在SAS系统中如何处理分层不等比例抽样的代码，这也是目前国内其他SAS图书没有介绍过的。

第二部分是数据挖掘商业案例：第13章主要介绍数据挖掘概念和流程。

数据挖掘流程尤其是商业流程是本章的重点。

该流程告诉读者一个真正的商业数据挖掘流程在商业环境中是如何实施的。

第14章重点介绍响应模型。

响应模型是商业实践中最常用的预测模型，基于第13章的流程规范给出了一个具体的商业案例研究。

第15章是客户行为分析。

该章有目前全球最流行的行为分析，包括“行为年龄”和“行为性别”（注意完全不同于具有自然属性特征的“真实年龄”和“真实性别”），作者运用NaiveBayesian技术开发出一整套模型，并对该模型拥有完全自主知识产权。

第16章介绍文本挖掘。

该章首先介绍了文本挖掘的流程，然后开发出基于NaiveBayesian文本分类算法和EM迭代思想的大型代码，并成功应用于商业实践。

<<SAS编程与数据挖掘商业案例>>

内容概要

从PDV角度详尽剖析Base SAS常用语句代码及应用，数据挖掘理论和商业应用紧密结合，原创相互贝叶斯文本分类和EM迭代算法代码，三个典型的数据挖掘商业案例分析。

本书是作者多年来在企业实践工作中的经验总结，详细讲解了使用SAS进行商业数据挖掘的方法，其中包含了目前公开出版的诸多SAS教材没有的大量实战内容。

本书内容全面、新颖独创、综合性强，适合企业人员使用，也可作为数学、统计学、金融、电子商务、医药等专业的本科生、硕士生学习SAS编程和数据挖掘的参考资料。

<<SAS编程与数据挖掘商业案例>>

作者简介

姚志勇，南开大学企业管理硕士，进修统计学硕士。
现供职于全球500强商业公司，从事SAS数据挖掘工作8年，有丰富的项目实践经验。
管理并自主开发多个大型数据挖掘算法源代码，成功应用于商业实践。

<<SAS编程与数据挖掘商业案例>>

书籍目录

出版说明前言第1章 SAS系统简介1.1 系统简介1.1.1 SAS系统与商务智能系统1.1.2 SAS系统与其他数据库的数据交换1.1.3 SAS语言与SAS系统1.1.4 SAS9浏览窗口简介1.2 一个简单的编程实例1.2.1 编写一个SAS程序1.2.2 提交一个SAS程序1.2.3 保存和打开一个SAS程序1.3 DATA步的数据指针和PDV流程1.3.1 数据指针和PDV流程1.3.2 DATA步执行次数第2章 SAS编程基础2.1 SAS逻辑库2.1.1 创建SAS逻辑库2.1.2 删除SAS逻辑库2.1.3 永久逻辑库和临时逻辑库2.2 SAS数据集2.2.1 SAS数据集命名规则2.2.2 永久SAS数据集和临时SAS数据集2.2.3 SAS数据集结构2.2.4 SAS数据集形式2.3 SAS索引2.3.1 创建索引2.3.2 删除索引2.4 SAS目录2.5 数据字典2.6 SAS变量2.6.1 变量属性2.6.2 变量列表2.6.3 自动变量第3章 数据获取与数据集操作3.1 数据获取3.1.1 LIBNAME方式3.1.2 PASSTHROUGH方式3.1.3 IMPORT方式3.1.4 INPUT方式3.2 SET语句3.2.1 语法说明3.2.2 实例详解3.2.3 商业实践3.3 BY语句3.3.1 语法说明3.3.2 实例详解3.4 MERGE语句3.4.1 语法说明3.4.2 实例详解3.5 UPDATE语句3.5.1 语法说明3.5.2 实例详解3.6 MODIFY语句3.6.1 语法说明3.6.2 实例详解3.6.3 商业实践3.7 PUT语句3.7.1 语法说明3.7.2 实例详解3.7.3 商业实践3.8 FILE语句3.8.1 语法说明3.8.2 实例详解3.8.3 商业实践3.9 INFILE语句3.9.1 语法说明3.9.2 实例详解3.9.3 商业实践第4章 SAS变量操作4.1 赋值语句和累加语句4.1.1 赋值语句4.1.2 累加语句4.2 KEEP语句和DROP语句4.2.1 KEEP语句4.2.2 DROP语句4.3 IRETAIN语句4.3.1 语法说明4.3.2 实例详解4.3.3 商业实践4.4 ARRAY语句4.4.1 语法说明4.4.2 实例详解4.4.3 商业实践4.5 其他语句4.5.1 RENAME语句4.5.2 LENGTH语句4.5.3 LABEL语句第5章 SAS观测值操作5.1 OUTPUT语句5.1.1 语法说明5.1.2 实例详解5.2 子集IF语句5.2.1 语法说明5.2.2 实例详解5.2.3 子集IF与OUTPUT语句比较5.3 WHERE语句5.3.1 语法说明5.3.2 实例详解5.3.3 子集IF与WHERE语句比较5.4 REPLACE语句和REMOVE语句5.4.1 REPLACE语句5.4.2 REMOVE语句5.4.3 REPLACE、REMOVE与OUTPUT应用5.5 DELETE语句与STOP语句5.5.1 DELETE语句5.5.2 STOP语句第6章 SAS数据集管理6.1 APPEND过程6.1.1 语法说明6.1.2 实例详解6.2 SORT过程6.2.1 语法说明6.2.2 实例详解6.2.3 商业实践6.3 TRANSPOSE过程6.3.1 语法说明6.3.2 实例详解6.4 CONTENTS过程6.4.1 语法说明6.4.2 实例详解6.5 DATASETS过程6.5.1 语法说明6.5.2 实例详解第7章 DATA步循环与控制7.1 IF.THEN / ELSE语句与SELECT语句7.1.1 IF.THEN / ELSE语句7.1.2 SELECT语句7.2 DO语句7.2.1 DO组语句7.2.2 DO循环语句7.2.3 DOWHILE语句7.2.4 DOUNTIL语句7.2.5 DOOVERRIDE语句7.2.6 商业实践7.3 各种控制语句7.3.1 GOTO语句7.3.2 CONTINUE语句与LEAVE语句7.3.3 RETURN语句第8章 常用全程语句8.1 COMMENT语句8.2 X语句8.3 FILENAME语句8.4 %INCLUDE语句8.5 TITLE语句8.6 FOOTNOTE语句第9章 输出控制9.1 LOG窗口输出控制9.2 OUTPUT窗口输出控制9.3 常用ODS输出控制9.3.1 ODSLSTING9.3.2 ODSRESULTS9.3.3 ODSTRACE9.3.4 ODSOUTPUT9.3.5 ODSHTML9.3.6 ODSCSVALL9.3.7 ODSSELECT9.3.8 ODSEXCLUDE第10章 SAS宏变量10.1 宏运行的内在机制10.2 宏变量10.2.1 定义宏变量10.2.2 显示宏变量10.2.3 引用宏变量10.3 宏程序10.3.1 定义宏10.3.2 调用宏10.3.3 宏内宏10.3.4 宏存储10.4 宏参数10.4.1 创建参数10.4.2 参数赋值10.5 宏函数10.5.1 通配函数10.5.2 计算函数10.5.3 字符函数10.5.4 引用函数10.6 宏语句10.6.1 %IF.%THEN / %ELSE语句10.6.2 %DO组语句10.6.3 %DO循环语句10.6.4 %DO%WHILE循环语句10.6.5 %DO%UNTIL循环语句10.7 宏应用10.7.1 创建宏变量的八种方法10.7.2 宏程序一般应用10.7.3 宏程序高级应用第11章 SQL过程11.1 单表操作11.2 多表操作11.2.1 多表关联11.2.2 子查询11.2.3 合并查询11.2.4 MERGE与SQL比较11.3 创建、更新与删除表操作11.3.1 创建表11.3.2 行操作11.3.3 列操作11.3.4.删除表11.4 使用SQL注意的几个问题第12章 数据处理实践12.1 随机抽样12.1.1 简单无重复随机抽样12.1.2 分层等比例随机抽样12.1.3 分层不等比例随机抽样12.1.4 随机抽样MACRO12.2 HASH对象12.2.1 HASH对象的引例12.2.2 HASH对象的语法12.2.3 HITER对象的引例12.2.4 HITER对象的语法12.2.5 商业实践12.3 FORMAT综述12.3.1 PROC步创建12.3.2 DATA步创建12.3.3 永久存储及调用12.4 正则表达式12.4.1 语法说明12.4.2 常用函数12.4.3 实例详解12.5 宏在SAS与Excel转换中的应用12.5.1 SAS数据集转换成Excel12.5.2 Excel转换成SAS数据集第13章 数据挖掘概念、任务和流程13.1 数据挖掘概念13.2 数据挖掘任务13.3 数据挖掘流程13.3.1 定义商业目标13.3.2 编制需求文档13.3.3 选择数据源13.3.4 建模流程图13.4 LOGISTIC建模及结果详解13.4.1 数学模型13.4.2 参数估计13.4.3 模型评价指标13.4.4 回归系数13.4.5 变量筛选方法13.4.6 应用举例及输出结果详解13.4.7 多

<<SAS编程与数据挖掘商业案例>>

值LOGISTIC模型第14章 响应模型：定位新客户14.1 前期准备14.1.1 商业需求14.1.2 定义目标14.1.3 选择变量14.2 数据获取与数据处理14.2.1 创建建模数据集14.2.2 变量首次筛选14.2.3 数据探索14.2.4 数据清洗14.2.5 变量二次筛选14.2.6 变量三次筛选14.2.7 字符变量压缩14.3 模型开发14.3.1 全模型法选择所有候选模型14.3.2 逐步回归法筛选候选模型14.3.3 创建两个重要数据集14.3.4 创建LIFT图14.3.5 创建评分卡文件14.4 模型验证14.4.1 评分卡文件导入14.4.2 LIFT图比较14.4.3 模型确认14.5 模型实施与监控14.5.1 模型实施14.5.2 模型监控14.6 小结第15章 行为建模：客户行为属性分析15.1 前期准备15.1.1 商业需求15.1.2 定义目标15.1.3 选择建模方法15.2 数据获取与处理15.3 模型开发15.4 模型验证15.5 模型打分15.6 模型预测15.7 模型实施15.8 小结第16章 文本挖掘：Web文本分析16.1 文本挖掘概念与流程16.1.1 文本挖掘概念16.1.2 文本挖掘流程16.2 商业案例16.2.1 商业需求16.2.2 建模框架设计16.2.3 结合朴素贝叶斯文本分类的EM迭代16.2.4 数据获取与数据预处理.....参考文献

章节摘录

插图：6.流失在商业领域，流失意味着利润的减少甚至账户关闭，主要发生在电信行业和银行业。由于竞争对手的存在，流失会经常发生。

如果能够通过建立一个比较健壮的流失模型，能够对那些可能在未来的几个月内流失的客户做出准确的预测，则从营销角度就可以提前做好一些准备，如通过一些营销手段来挽留这些客户。

从技术上，流失模型本质上和风险模型是一样的，只是目标定义可能会有所不同。

一种常见的流失定义是：在过去的6个月内账户余额呈现减少趋势，并在第6个月余额低于公司规定的阈值或比例。

注意：响应模型是基于时间点的模型，而风险模型和流失模型是基于时间段的模型。

7.提升销售和交叉销售提升销售是指预测客户购买更多同样产品的可能性。

交叉销售是指预测客户购买公司不同产品的可能性。

提升销售和交叉销售对于纵向挖掘一个客户的潜在利润是非常重要的。

从技术上来说，关联规则也许能够帮助公司发现客户的特征，著名的“啤酒和尿布”就是多数数据挖掘图书“言必称希腊”的经典案例，但是在作者看来，“啤酒和尿布”如同“尼斯湖怪兽”一样并非值得绝对信赖。

在提升和交叉销售领域，更多的关注应该是客户消费对象本身的关联性以及客户的消费心理，这是更偏向于定性分析的技术，而不是定量分析的方法。

事实上，以上列举的商业需求只是千千万万个商业需求中的部分代表，读者能够从中得到这样的启发：所有的数据挖掘技术都是商业目标的一个实现，或简单，或复杂。

而商业应用的最高原则就是“效率、效果”。

13.3.2.编制需求文档在明确了商业目标之后，接下来就需要分析师编制需求文档。

需求文档是商业目标的细化。

完整的需求文档应包括以下几个部分：项目计划文档（PPT）。

方法论设计文档（Word）。

变量需求文档（Excel）。

这三个项目涵盖了商业客户需求、团队头脑风暴成果、项目数据收集指标三个重要内容。

1.项目计划文档项目计划文档是指实现客户需求而制定的需求框架、计划内容、路线图和资源。

由于该部分内容主要呈现给商业客户，因此一般以PPT形式，幻灯片尽量控制在10张以下。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>