<<数据挖掘与应用>>

图书基本信息

书名：<<数据挖掘与应用>>

13位ISBN编号：9787301152393

10位ISBN编号：7301152396

出版时间：2009-6

出版时间：北京大学出版社

作者：张俊妮

页数：185

前言

教材建设是大学人才培养和知识传授的重要组成部分。
对管理教育而言，教材建设尤为重要，一流的商学院不仅要有一流的师资力量、一流的生源、一流的教学管理水平，而且必须使用一流的教科书。
一流的管理类教科书必须满足以下标准：第一，能把所在领域的基础知识以全面、系统的方式和与读者友好的语言呈献给读者；第二，必须有时代感，能把学科前沿的研究成果囊括进去；第三，必须做到理论和实务（包括案例分析）相结合，有很强的实用性；第四，能够启发学生思考现实的管理问题，培养他们分析问题和解决问题的能力；第五，可以作为研究人员和管理人士的工具书。
中国的管理教育是伴随改革开放而产生的。
真正意义上的管理教育在中国不过十多年的历史，但巨大的市场需求使得管理教育成为中国高等教育各学科中发展最快的领域，管理类教科书市场异常繁荣。
但总体而言，目前国内市场上管理类教科书的水平仍不能令人满意。
国内教科书作者大多数在所涉及领域并没有真正的原创性研究和学术贡献，所撰写的教科书普遍停留在对国外教科书的内容进行中国式排列组合的水平上；国外引进的原版教科书虽然具有学术上的先进性，但由于其写作背景是外国的管理实践和制度安排，案例也都是取自于西方发达国家，对中国读者而言，总有一种隔靴搔痒的感觉。
如何写出一流的中国版的管理类教材，是中国管理教育发展面临的重要任务。
北京大学光华管理学院一直重视教材建设工作。
1999年夏，我们曾与经济科学出版社签约，以每本20万元的稿酬，向全国征集MBA教科书作者。
这个计划公布之后，我们收到了十几本教科书的写作方案。

<<数据挖掘与应用>>

内容概要

本书全面地介绍了数据挖掘的相关主题，包括数据理解与数据准备、关联规则挖掘、多元统计中的降维方法、聚类分析、神经网络、决策树方法、模型评估等内容。
全书体系完整，文字精炼，注重对数据挖掘方法的直觉理解及其应用；同时，保持了一定的严谨性，为学生理解和运用这些方法提供了坚实的基础。
本书实例丰富，并附有相应SAS程序，以便于学生尽快理解相关内容并用以解决实际问题。
本书配有教辅，可以免费提供给任课教师使用。
如需要，欢迎填写书后的“ 教师反馈及课件申请表 ”索取。

作者简介

张俊妮，美国哈佛大学统计学博士，现为北京大学光华管理学院商务统计及经济计量系副教授。研究领域包括因果推断、贝叶斯分析、蒙特卡洛方法、数据挖掘。

<<数据挖掘与应用>>

书籍目录

章节摘录

插图：对于定序自变量，最常用的一种转换是按各类别的序号直接将该变量转换为数值自变量。
对于名义自变量，最常用的转换是将该变量转换为哑变量。
例如，对于性别而言，可以生成一个二元哑变量，取值1表示“女”，0表示“男”。
对于有多个取值的名义自变量，可以生成一系列二元哑变量。
例如，中国内地有31个省、自治区和直辖市，可以据此生成30个哑变量。
但是，如果一个名义自变量取值过多，生成过多的哑变量容易造成过度拟合。
一个简单而有效的方法是只针对包含观测比较多的类别生成哑变量，而将剩余的类别都归于“其他”这个大类别。
还有一种方法是利用领域知识，将各类别归为几个大类之后再生成哑变量，例如，将中国内地31个省、自治区和直辖市归为华北、华中、华东、华南、西北、东北、西南等地区，再生成地区的哑变量。
五、处理时间变量时间变量无法直接进入建模数据集，因为时间是无限增长的，在历史数据中出现的时间肯定不同于将来模型所需应用的数据集中出现的时间，所以直接使用历史数据的时间建立的模型就无法应用于将来的数据集。
如果要在建模过程中考虑时间变量，就必须对其进行转换。
常用的转换有如下几种：1.转换为距某一基准时间的时间长短，例如，“距离××年××月××日的天数”、“距离下一次春节的周数”等。
2.转换为季节性信息，例如，一年中第几季度或第几个月，每个季度或月对应于一个二元哑变量。
很多情形下可以考虑对时间进行多种转换，把所有可能影响因变量的时间信息都放人建模过程中。
例如，对于某些食品的购买量而言，不仅存在节日效应，也存在季节性效应，这时就需要同时使用上述两种转换。
六、异常值自变量的异常值对一些模型会产生很大影响。
在图2.2 的示例中，大部分数据点的，值都分布在-2.2 和2.4 之间，但有一个数据点的x值为8，它对拟合的回归线会有很大的影响；如果它落在点0或点6，拟合出的回归线分别为线a和线b，它们的差别颇大。
因变量的异常值同样可能对模型有很大影响，在这里不赘述。
第五章将介绍的聚类算法可以用来发现异常值，如果少数几个观测自成一类，它们很有可能是异常值。
发现异常值后需要查看它们为什么异常。

<<数据挖掘与应用>>

编辑推荐

《数据挖掘与应用》是张俊妮编写的，由北京大学出版社出版。

<<数据挖掘与应用>>

版权说明