



21世纪高等院校示范性实验系列教材

数据挖掘与商务智能实验教程

SHUJU WAJUE YU SHANGWU ZHINENG SHIYAN JIAOCHENG

主编 张大斌



教育部直属师范大学
华中师范大学出版社

数据挖掘与商务智能 实验教程

主编 张大斌

华中师范大学出版社
2015年·武汉

内 容 提 要

数据挖掘与商务智能是高等院校电子商务、信息管理与信息系统、计算机科学与技术等相关专业的核心课程,也是近年来企业信息化的热点内容。本实验教程通过一系列数据挖掘与商务智能工具的实验练习,将商务智能的概念和理论知识融入实践中,从而加深读者对数据挖掘与商务智能原理和算法的理解。实验内容涉及商务智能的终端用户查询和报告、联机分析处理、数据挖掘、数据仓库等软件工具,主要包括 SPSS Modeler、SQL Server Analysis Services、WEKA 和 SAP Business Object。全书分为17 个小实验和 2 个综合设计实验,每个实验都包含背景知识、实验目的、所需要的工具/准备工作、实验内容及步骤、实验分析与扩展练习等,以帮助读者掌握一些主流数据挖掘与商务智能工具的基本使用方法,通过规范的实验训练,培养其综合解决实际管理问题的能力。

本书可作为高等院校相关专业的“数据挖掘”、“数据仓库”、“商务智能”等课程的实验辅助教材,也可作为单独开设“数据挖掘与商务智能实验”课程的主教材。

图书在版编目(CIP)数据

数据挖掘与商务智能实验教程 / 张大斌主编. —武汉:华中师范大学出版社,2015. 1

(21 世纪高等院校示范性实验系列教材)

ISBN 978-7-5622-6880-2

I. ①数… II. ①张… III. ①数据采集—应用—电子商务—高等学校—教材 IV. ①TP274 ②F713. 36

中国版本图书馆 CIP 数据核字(2014)第 290592 号

数据挖掘与商务智能实验教程

主编 张大斌◎

责任编辑:王 炜

责任校对:王 胜

封面设计:罗明波

编辑室:第二编辑室

电话:027-67867362

出版发行:华中师范大学出版社有限责任公司

社址:湖北省武汉市珞喻路 152 号

电话:027-67863426/67863280(发行部) 027-67861321(邮购)

传真:027-67863291

网址:<http://www.ccnupress.com>

电子信箱:hscbs@public.wh.hb.cn

印刷:虎彩印艺股份有限公司

督印:章光琼

字数:420 千字

印张:14.5

开本:889mm×1194mm 1/16

版次:2015 年 1 月第 1 版

印次:2015 年 1 月第 1 次印刷

定价:28.00 元

欢迎上网查询、购书

前 言

以云计算、物联网、移动互联网、智慧城市等为代表的新一代数据中心的迅猛发展和“大数据”引领的智慧科技时代的来临,开启了信息领域的重大转型。人们的生产、经营、决策、服务等活动日益依赖于数据分析和挖掘,由此带来技术、管理、思维方式、商业模式的巨变。因此,如何快速、高效地从海量数据中获取有价值的信息,进而分析、判断、抉择、预测,已经成为业界最关注和亟待优先解决的重点问题,与数据收集、分析挖掘和综合处理相关的技术、专利、人才等也日益变得炙手可热。

为了适应这一市场需求,高等院校的电子商务、信息管理与信息系统及计算机科学与技术等相关专业纷纷开设了“数据挖掘”、“数据仓库”和“商务智能”等课程,以培养学生智能分析数据的能力。目前,这类课程的教学主要以具体理论及算法分析为主,通过例题或习题方式来讲授,缺少针对企业实际案例设计的相应的实践教学内容。为此,我们以数据挖掘与商务智能涉及的主要方法为对象,通过市场上的主流工具,并结合企业的实际管理数据,来设计实践内容并编写了本实验教程。

本实验教程的主要指导思想是以电子商务、信息管理与信息系统、计算机科学与技术等相关专业的本科人才培养方案中设计的专业课程体系为基础,利用市场上主流或开源的软件工具资源,充分集成智能分析课程群的实验项目,高层次培养学生进行信息智能分析的实践能力。

本实验教程涉及电子商务、信息管理与信息系统、计算机科学与技术等相关专业的“商务智能原理与方法”、“数据仓库”、“数据挖掘”等课程。这几门课程各有自己的研究对象,但同时又是一个不可分割的整体,可以用已有的商业软件将这些知识全部联系在一起。“商务智能原理与方法”课程的重点在于业务绩效管理、可视化技术,其中可视化技术包括报表和查询、用户界面,用户界面核心是仪表盘和其他信息广播工具。“数据仓库”课程主要介绍数据预处理,即数据的 ETL(数据提取、转换和加载)过程以及 OLTP(联机事务处理)与 OLAP(联机分析处理)等。而“数据挖掘”课程主要介绍一些智能方法和工具,对数据库或数据仓库中的数据进行深度挖掘。不过也有文献从广义上认为商务智能包括数据仓库和数据挖掘。但不同企业设计的软件会将这些功能区分开来,比如企业管理系列软件(SAP)的商务智能软件 Business Object 就只包括“数据仓库”和“商务智能原理与方法”课程中涉及的内容,没有数据挖掘模块。而 SPSS 有专门的数据挖掘软件 Clementine,现在命名为 Modeler。为了全面掌握数据挖掘与商务智能的知识内容,本教程综合集成了这三门课程的实验内容,使之形成一个完备的体系。

为了便于教师根据本校的实验条件以及讲授课程内容的不同,灵活选择对应的实验项目,本实验教程共分为四个部分:实验环境工具篇、数据挖掘实验篇、商务智能实验篇和综合实验篇。实验的主要工具分为 SPSS Modeler、SQL Server Analysis Services、WEKA 和 SAP Business Object,并用这些工具分别设计了对应的实验项目。本教程共设计了 19 个实验,包括 17 个小实验和 2 个综合设计实验,每个实验都包含背景知识介绍、实验目的、实验工具/准备工作、实验内容及步骤、实验分析与扩展练习等,以帮助读者加深对原理与方法的理解,熟练掌握市场上主流的数据挖掘与商务智能工具。

本书既可以作为高等学校电子商务、信息管理与信息系统、计算机科学与技术等相关专业的高年级本科生和管理科学与工程学科的“数据挖掘与商务智能”课程群的实验辅助教材,也可作为单独开设“数据挖掘与商务智能实验”课程的主教材,同时可作为企业、事业单位信息化的培训教材以及相关工程与管理决策人员的参考书。

本书的推出要感谢华中师范大学出版社、教务处和实验室与设备管理处共同立项的“21世纪高等院校示范性实验系列教材”出版项目的资助,另外,本书的编写得到了上海数聚公司和博易智讯公司的支持和帮助,感谢他们提供了大量技术资料和部分软件工具。

参加本书编写的有刘雯(实验1、实验6~11、实验18~19)、周志刚(实验4、实验12~15)、邵鹏(实验2、实验6~11)、江华(实验5、实验16~17)、王畅(实验3),全书大纲由张大斌教授拟定并由其统稿。由于编者水平有限,书中难免存在错漏之处,敬请广大读者批评指正。

编者

2015年1月

目 录

第一部分 实验环境工具篇	1
实验1 数据挖掘软件 SPSS Modeler 的使用	3
1.1 背景知识	3
1.2 实验目的	5
1.3 工具/准备工作	5
1.4 实验内容及步骤	5
1.5 实验分析与扩展练习	8
实验2 数据挖掘套件 SQL Server Analysis Services 的使用	9
2.1 背景知识	9
2.2 实验目的	10
2.3 工具/准备工作	10
2.4 实验内容及步骤	10
2.5 实验分析与扩展练习	12
实验3 开源数据挖掘软件 WEKA 的使用	13
3.1 背景知识	13
3.2 实验目的	19
3.3 工具/准备工作	20
3.4 实验内容及步骤	20
3.5 实验分析与扩展练习	23
实验4 商务智能软件 SAP Business Object 的使用	25
4.1 背景知识	25
4.2 实验目的	30
4.3 工具/准备工作	30
4.4 实验内容及步骤	31
4.5 实验分析与扩展练习	36
实验5 商务智能可视化工具 Dashboard 和 Crystal Reports 的使用	37
5.1 背景知识	37
5.2 实验目的	42
5.3 工具/准备工作	42
5.4 实验内容及步骤	42
5.5 实验分析与扩展练习	48
第二部分 数据挖掘实验篇	49
实验6 数据挖掘的基本数据分析	51
6.1 背景知识	51
6.2 实验目的	51

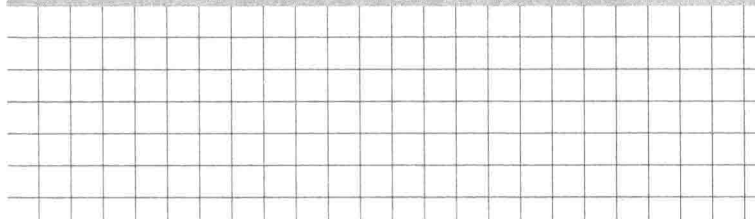
6.3	工具/准备工作	51
6.4	实验内容及步骤	51
6.5	实验分析与扩展练习	55
实验7	统计分析:逻辑回归	56
7.1	背景知识	56
7.2	实验目的	58
7.3	工具/准备工作	58
7.4	实验内容及步骤	58
7.5	实验分析与扩展练习	63
实验8	关联分析:关联规则	64
8.1	背景知识	64
8.2	实验目的	66
8.3	工具/准备工作	67
8.4	实验内容及步骤	67
8.5	实验分析与扩展练习	72
实验9	分类分析:决策树	74
9.1	背景知识	74
9.2	实验目的	78
9.3	工具/准备工作	79
9.4	实验内容及步骤	79
9.5	实验分析与扩展练习	83
实验10	聚类分析:K-Means	85
10.1	背景知识	85
10.2	实验目的	88
10.3	工具/准备工作	88
10.4	实验内容及步骤	89
10.5	实验分析与扩展练习	94
实验11	预测分析:人工神经网络	95
11.1	背景知识	95
11.2	实验目的	99
11.3	工具/准备工作	99
11.4	实验内容及步骤	100
11.5	实验分析与扩展练习	104
第三部分	商务智能实验篇	105
实验12	数据仓库的创建	107
12.1	背景知识	107
12.2	实验目的	111
12.3	工具/准备工作	111
12.4	实验内容及步骤	111
12.5	实验分析与扩展练习	117
实验13	数据抽取、转换、装载(ETL)	118
13.1	背景知识	118

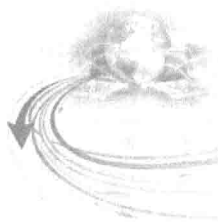
13.2	实验目的	120
13.3	工具/准备工作	120
13.4	实验内容及步骤	120
13.5	实验分析与扩展练习	129
实验14	联机分析处理(OLAP)	130
14.1	背景知识	130
14.2	实验目的	134
14.3	工具/准备工作	134
14.4	实验内容及步骤	134
14.5	实验分析与扩展练习	147
实验15	“水晶易表”的仪表盘制作	148
15.1	背景知识	148
15.2	实验目的	148
15.3	工具/准备工作	148
15.4	实验内容及步骤	148
15.5	实验分析与扩展练习	154
实验16	“水晶易表”的动态波士顿矩阵制作	155
16.1	背景知识	155
16.2	实验目的	155
16.3	工具/准备工作	156
16.4	实验内容及步骤	156
16.5	实验分析与扩展练习	160
实验17	“水晶报表”的制作——建立一个“高校学生成绩水晶报表”	161
17.1	背景知识	161
17.2	实验目的	161
17.3	工具/准备工作	161
17.4	实验内容及步骤	162
17.5	实验分析与扩展练习	170
第四部分	综合实验篇	171
实验18	保险业的欺诈行为挖掘	173
18.1	背景知识	173
18.2	实验目的	173
18.3	工具/准备工作	173
18.4	实验内容及步骤	173
18.5	实验分析与扩展练习	205
实验19	电信业客户流失分析	207
19.1	背景知识	207
19.2	实验目的	207
19.3	工具/准备工作	207
19.4	实验内容及步骤	207
19.5	实验分析与扩展练习	223
参考文献		224



第一部分 实验环境工具篇

SHIYAN HUANJING GONGJU PIAN





实验 1

数据挖掘软件 SPSS Modeler 的使用

1.1 背景知识

1.1.1 数据挖掘的产生背景

20 世纪 80 年代以来,随着计算机数据库技术和产品的日益成熟以及计算机应用的普及深化,各行业的数据采集能力得到了前所未有的提高,通过各自内部的业务处理系统、管理信息系统以及外部网络系统,获得并积累了浩如烟海的数据。

在严酷的市场竞争压力下,为更客观地把握自身和市场状况,提升内部管理和决策水平,企业管理者们面对如此丰富的海量数据,分析需求越来越强烈。他们希望利用有效的数据分析工具,更多地挖掘出隐藏在数据中的、有价值的、能够辅助管理和决策的信息。然而,问题接踵而至,原先管理者们得不到想要的信息,是因为数据库中没有充足的数据,但现在他们似乎仍然无法快捷地得到想要的信息,其原因是数据库里的数据资源成为使用者的负担,组织中的管理决策者无奈地感慨:基层业务人员尚且能够通过业务处理系统快速访问一定范围内的业务数据,而高层决策者却似乎缺少有效的工具,从数据库中获得利于决策制定的有价值的信息。于是,所谓的“信息爆炸”、“数据多但知识少”成为一种很普遍的怪现象。针对这一现象,数据挖掘作为 20 世纪 90 年代中后期兴起的技术,因其注重减少数据分析方法对数据的限制性和约束性,注重与计算机技术结合以实现数据的管理性及分析的易操作性,已成为数据分析中获取知识的重要手段。同时,随着数据挖掘方法的不断成熟及其应用的日益普及化,数据挖掘软件的研发也取得了可喜的成果。

1.1.2 数据挖掘的概念

数据挖掘(Data Mining, DM),就是从大量数据(包括文本)中挖掘隐含的、未知的、对决策有潜在价值的关系、模式和趋势,并且这些知识和规划建立用于决策支持的模型,提供预测性决策支持的方法、工具和过程;是利用各种分析工具在海量数据中发现模型和数据之间关系的过程。这些模型和关系可以被企业用来分析风险、进行预测。

1.1.3 数据挖掘的功能

数据挖掘的功能是现有商务智能平台的核心组成部分,图 1-1 中的结构图能更加直观地帮助企业定性、定量地了解各种市场活动和企业内部运作可能带来的利益,从而不断发现新的收益增长点。具体的商务智能分析中的数据挖掘模型如图 1-1 所示。

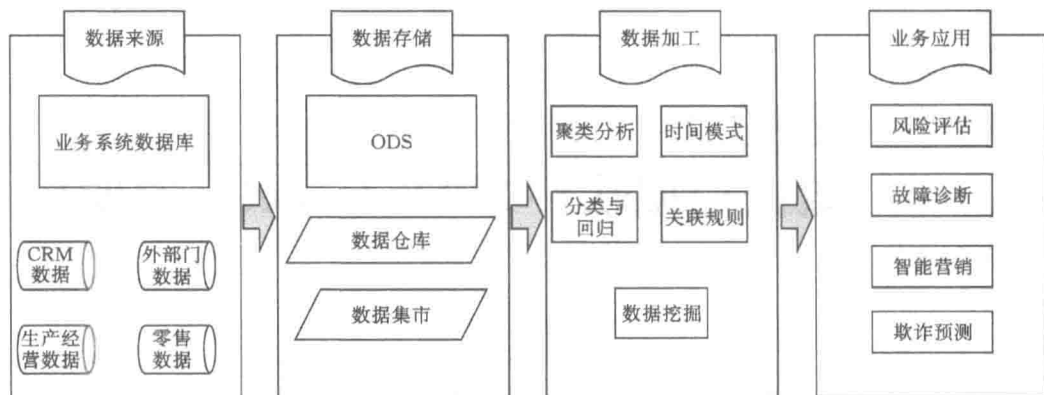


图 1-1 商务智能分析中的数据挖掘模型

数据挖掘的主要功能：可以完成数据的总结、分类、关联、聚类等若干主要任务。

(1) 数据总结。

数据总结是对数据的基本特征进行概括。通过数据总结，不仅能够实现对数据多维度多层次的汇总，还能够得到数据分布特征的精确概括。

例如，为制订不同种类的商品在不同城市 and 不同季节的销售情况，首先可对现有销售数据进行汇总。如果数据为月度数据，那么可以按照季节，汇总出不同种类商品在不同城市各个季度的销售量；也可以按地理区域，汇总出不同种类商品在不同月份各个区域的销售量，形成各种统计报表，等等。这种多角度的汇总能够直观地反映销售的情况，是对原始数据的提炼和总结。

可利用数据仓库的联机分析处理(OLAP)技术进行数据的多维查询汇总，也可以通过基本统计方法计算测度数据分布的集中趋势、离散程度以及分布对称性和陡缓程度。

(2) 分类。

分类的主要目的是通过向数据“学习”，分析数据不同属性之间的联系，得到一种能够正确区分数据所属组别的规律，即通过“学习”建立一种包含分类规律的分类模型，且该模型能够对新数据所属的组别进行自动预测。

例如，一份超市的会员购买行为的数据，其中包括顾客的性别、职业、收入、年龄以及相关的消费记录。如果希望分析顾客的消费行为跟性别、职业、收入、年龄等属性特征是否相关，可以将曾经购买的顾客分为一类，另一类则为未购买的顾客，由此通过“学习”来找到特征属性与消费行为之间的联系规律，从而实现对新客户是否购买进行分类预测。

常见的分类方法有机器学习中的决策树、神经网络，以及统计学的 Logistic 回归、判别分析等。评价模型的优劣性的重要方面是分类预测的准确性。

(3) 关联。

关联就是通过数据分析来发现事物之间的相互联系，包括简单关联规则和时序关联规则。

例如，很著名的“尿布与啤酒”的故事。在美国沃尔玛连锁超市中，沃尔玛对其顾客的购物行为进行购物篮分析，想弄清楚顾客经常一起购买什么，发现一些年轻的父亲在下班后经常要到超市去买婴儿尿布，而他们中有 30%~40% 的人同时也为自己买了一些啤酒。

当然在关联分析中，通常不知道关联性是否确实存在，因此关联分析所产生的规则是带有一定置信度的，所谓的置信度就是如果 A 发生，则 B 有百分之 C 的可能方法，其中 C 称为关联规则的置信度(Confidence)。

常用的关联规则是机器学习中的相关规则等。

(4) 聚类。

聚类是在没有给定分类的情况下，根据信息的相似度进行的信息聚类的一种方法，因此聚类又称为无指导的学习。

例如,根据顾客的消费行为的数据,企业可以做聚类分析,在不指定任何分类标准下,根据数据全面客观地进行客户群划分,不同群组的客户消费行为和客户特征总体上相近。这样,可以不同客户群采用不同的营销策略。

采用的聚类方法有层次聚类、K-Means 聚类以及两步聚类、Kohonen 聚类等。

1.1.4 数据挖掘的工具

目前市场上的数据挖掘软件主要有 SPSS Modeler、Statsoft Statistics、SAS Enterprise Miner、Oracle DM、MATLAB、Angoss,以及开源数据挖掘软件 WEKA 等。由于 SPSS Modeler 拥有丰富的数据挖掘算法,支持与数据库之间的数据和模型交换,同时,具有可视化操作界面,简单易用,分析的结果直观易懂,图形功能强大等特点,SPSS Modeler 产品一直居市场占有率首位。

1.2 实验目的

- (1)熟悉 SPSS Modeler 的窗口基本构成。
- (2)基本掌握 SPSS Modeler 中工具的相关操作。

1.3 工具/准备工作

- (1)熟悉相关智能算法的基本原理。
- (2)一台装有 SPSS Modeler 软件的电脑。

1.4 实验内容及步骤

本次实验所采用的数据来自一份关于药物研究的数据。本数据中相关的变量 x_1-x_7 依次是药物 Drug (DrugA, DrugB, DrugC, DrugX, DrugY), 血压 BP (High、Normal、Low), 胆固醇 Cholesterol (Normal、High), 唾液元素含量 (Na、K), 以及病人的年纪 Age, 性别 Sex (M、F)。本次实验共有 200 条记录。通过数据分析发现以往处方适用的规律,给出不同临床特征病人更加适合服用哪种药物的建议,为未来医生填写处方提供参考。

步骤 1 构建数据流

- (1)启动桌面/开始菜单中的 SPSS Modeler 图标后会出现 SPSS Modeler 主窗口。如图 1-2 所示。



图 1-2 SPSS Modeler 主窗口

(2)在节点工具箱窗口—“源”下,选择合适的节点,拖入数据流编辑区,本例选择“可变文本”,在文件选项中指定的文件 DRUG.txt 中读入数据。

(3)在“输出”选项卡下,选择“Table”节点,连在数据流中,运行后得到相应的输出结果。“输出”选项卡如图 1-3 所示。



图 1-3 “输出”选项卡相关节点

(4)在“图形”选项卡下,选择“图”节点,连在数据流中。

(5)根据上述的方法,按照图 1-4 依次连入相应的节点。

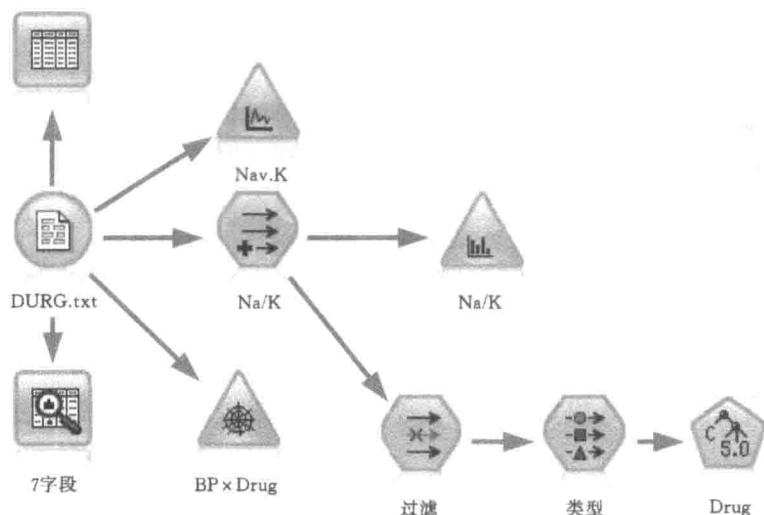


图 1-4 数据流的构建

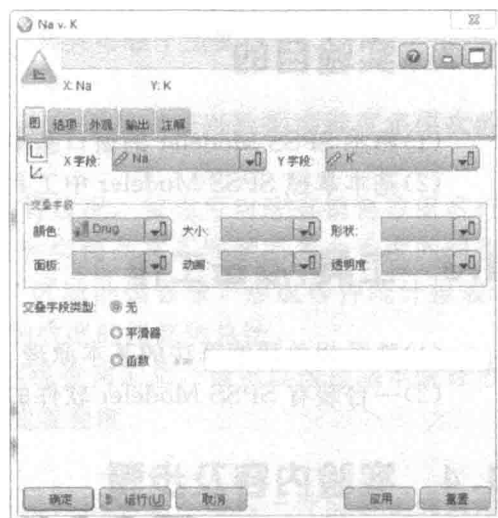


图 1-5 “图”节点参数的设置

步骤 2 相关参数设置

(1)在“图”节点处,双击点开编辑框,弹出的编辑框中 X 字段为 Na, Y 字段为 K;颜色设置为 Drug。如图 1-5 所示。

(2)在“导出”节点处,双击点开编辑框,在弹出的窗口中,导出字段为 Na/K;公式为 Na/K。如图 1-6 所示。

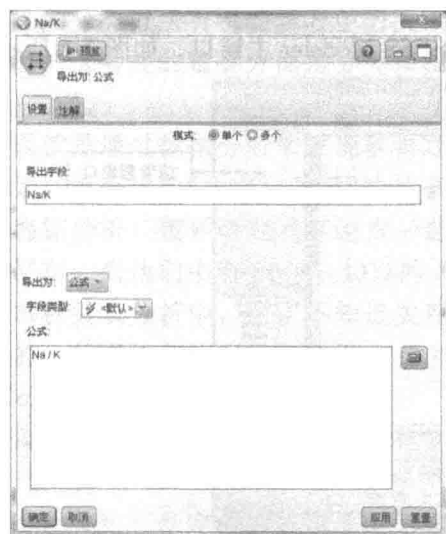


图 1-6 “导出”节点的参数设置

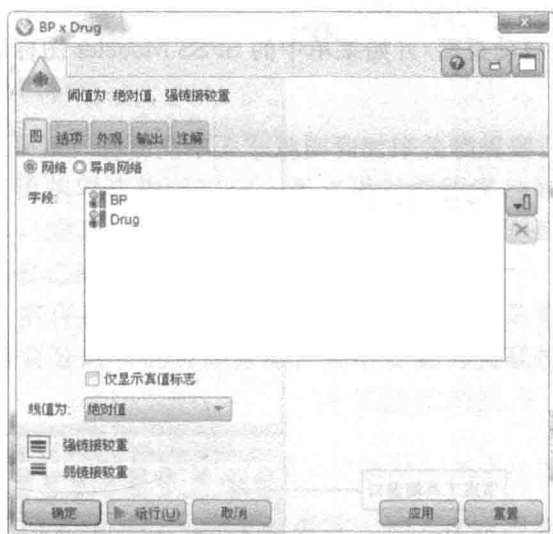


图 1-7 “网络”节点的参数设置

(3)在“网络”节点处，双击，在弹出的编辑框中，字段处选入 BP 和 Drug；线性为绝对值。如图 1-7 所示。

(4)在“直方图”节点处，双击，弹出编辑窗口，设置字段为 Na/K；颜色为 Drug。如图 1-8 所示。

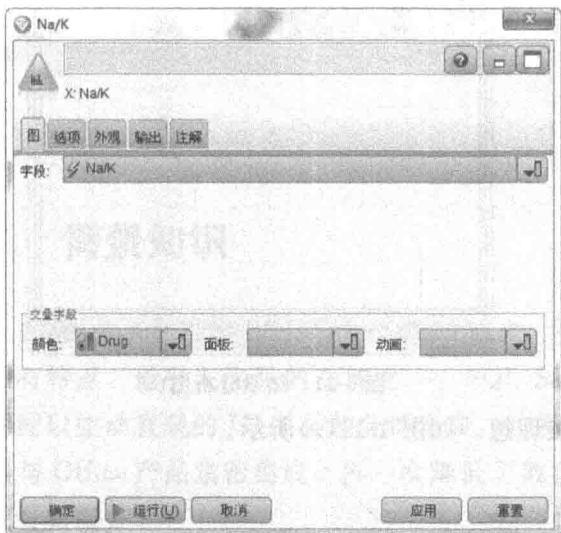


图 1-8 “直方图”参数的设置

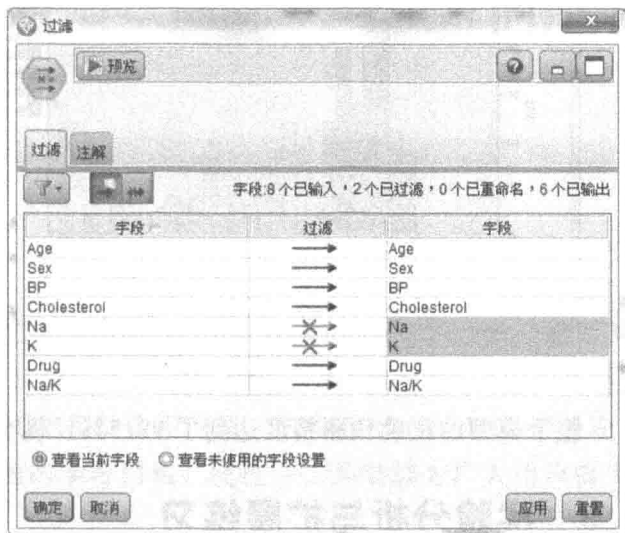


图 1-9 “过滤”的参数设置

(5)在“过滤”节点处，双击，弹出编辑窗口，过滤的字段为 Na、K。如图 1-9 所示。

(6)在“类型”节点处，双击，弹出编辑窗口，设置字段 Drug 的角色为目标。如图 1-10 所示。

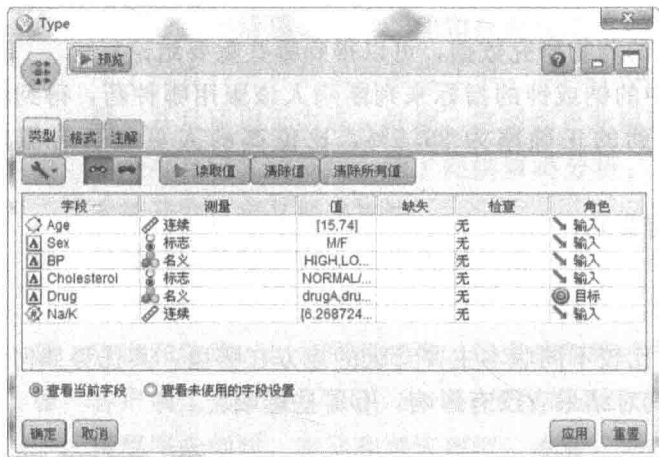


图 1-10 “类型”的参数设置

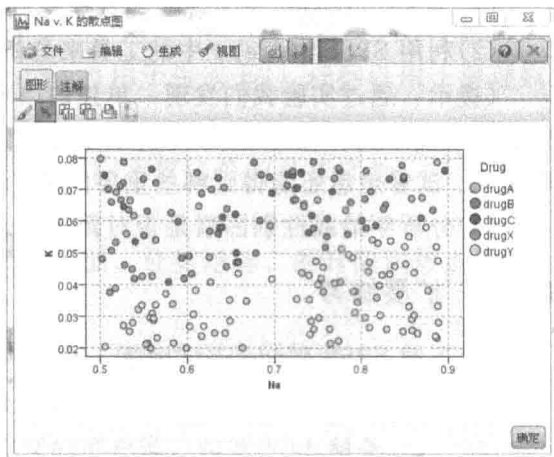


图 1-11 结果分析图(a)

步骤 3 数据流运行

单击“运行”按钮，得到相应的部分结果如图 1-11 所示。

在图 1-11(a)中可以看出，服用 DrugY 的病人，其唾液中钾的含量明显低于其他类病人，但钠的含量有的高有的低，单纯地看，钾含量低的人服用 DrugY 比较理想。

单纯从钠、钾的含量观察是不全面的，应观察钠和钾的浓度比值指标，它能更好地反映病人的肾上腺的功能情况。针对 Na/K 值处在高水平的病人服用 DrugY 应是理想的选择。如图 1-11(b)所示。

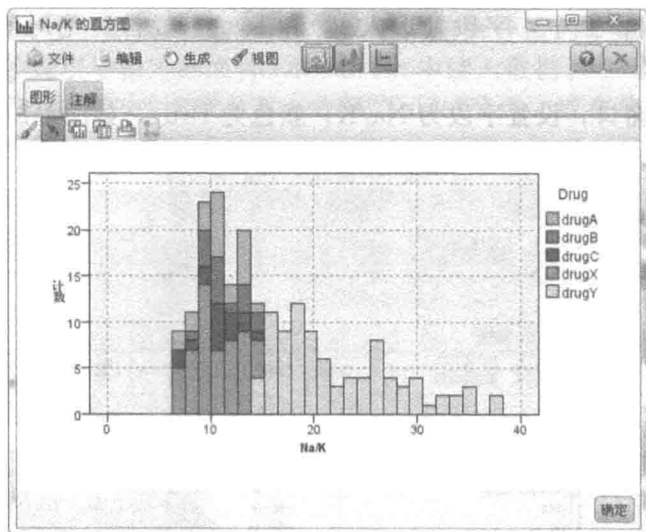


图 1-11 结果分析图(b)

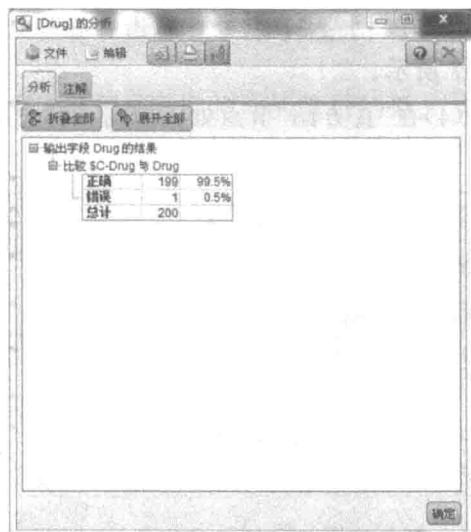


图 1-11 结果分析图(c)

整个模型的正确预测精度达到了 99.5%，模型比较理想。如图 1-11(c)所示。

1.5 实验分析与扩展练习

1.5.1 实验分析

本实验通过一份关于药物研究数据的分析，基本掌握 SPSS Modeler 的使用方法。请总结分析下面问题：

- (1) SPSS Modeler 能实现哪些数据挖掘功能？
- (2) 利用 SPSS Modeler 中的工具来分析上述的药物研究数据，可以得到哪些重要结论？

(提示：通过实验我们发现，单独的以唾液中的钠或钾的指标来判断病人该服用哪种药，得到结果的精准度不高。而用钠与钾的含量比值，得到的正确率为 99.5%，比值高的人更加适合服用 DrugY。试着想想还能得到哪些重要的结论。)

- (3) 分析年龄或性别因素是否对药物的选择有影响。

1.5.2 扩展练习

- (1) 下载 SPSS Modeler 试用版，通过练习，比较不同版本主要升级的地方在哪里，其优势如何？
- (2) 改变上述实验中的相关参数设置，分析其对结果有没有影响，有哪些影响？



实验 2

数据挖掘套件 SQL Server Analysis Services 的使用

2.1 背景知识

自 Microsoft 推出 SQL Server 2000 以来，数据挖掘 SSAS(SQL Server Analysis Services)成为其中最特点、最引人注目的功能之一。SQL Server 开发组将多种经典统计算法置于 SSAS 产品中，同时配以生动直观的 UI 工具和分析图表，扩展了数据挖掘的应用范围。在 SQL Server 2008 产品中，更是与 Office 产品紧密集成，再一次降低了数据挖掘的学习门槛，使这一工具继续为广大用户提供深层分析服务。

针对集成的数据挖掘解决方案，SQL 提供了以下的功能支持：

- 多个数据源：无需创建数据仓库或 OLAP 多维数据集就可以执行数据挖掘。可以使用来自外部提供程序、电子表格甚至文本文件的表格格式数据。此外，还可以轻松地挖掘在 Analysis Services 中创建的 OLAP 多维数据集。但是，不能使用在内存中存放的数据库中的数据。
- 集成的数据清理、数据管理和数据提取、转换和加载(ETL)：Data Quality Services 提供了用于探查和清理数据的高级工具。Integration Services 可用于生成 ETL 进程以用于清理数据，并且还用于生成、处理、定型和更新模型。
- 多个可自定义的算法：除了提供聚类分析、神经网络和决策树之类的算法之外，该平台还支持开发自定义插件算法。
- 模型测试基础结构：使用重要的统计工具(例如交叉验证、分类矩阵、提升图和散点图)测试模型和数据集。轻松创建和管理测试和定型集。
- 查询和钻取：创建预测查询、检索模型模式和统计信息以及钻取到事例数据。
- 客户端工具：除了 SQL Server 提供的开发和设计工具之外，还可以使用 Excel 数据挖掘外接程序来创建、查询和浏览模型。或者，创建自定义的客户端，包括 Web 服务。
- 脚本语言支持和托管 API：所有数据挖掘对象都是完全可编程的。可通过用于 Analysis Services 的 MDX、XMLA 或 Power Shell 扩展插件来撰写脚本。使用数据挖掘扩展插件(DMX)语言来进行快速查询和脚本撰写。
- 安全性和部署：通过 Analysis Services 提供基于角色的安全性，包括用于钻取到模型和结构数据的单独的权限，轻松地将模型部署到其他服务器，以便用户可以访问模式或执行预测。

SSAS 数据挖掘功能在算法模型构建方面，难度不是很大，真正具挑战性的一面是理解算法的真正功能，然后在创建的挖掘结构中包含最符合具体业务需求的算法。另一项重要考虑就是如何向最终用户呈现结果。