

智能信息管理研究

Research on Intelligent Information Management

主 编 刘念祖

副主编 王汉兴 王双成



立信会计出版社

LIXIN ACCOUNTING PUBLISHING HOUSE

智能信息管理研究

Research on Intelligent Information Management



立信会计出版社

LIXIN ACCOUNTING PUBLISHING HOUSE

图书在版编目(CIP)数据

智能信息管理研究 / 刘念祖主编. —上海: 立信
会计出版社, 2010.12

ISBN 978-7-5429-2688-3

I. ①智… II. ①刘… III. ①智能控制-管理信息系统
-研究 IV. ①TP273②C931.6

中国版本图书馆 CIP 数据核字(2010)第 233572 号

责任编辑 陈 旻

封面设计 周崇文

智能信息管理研究

出版发行 立信会计出版社

地 址 上海市中山西路 2230 号 邮政编码 200235

电 话 (021)64411389 传 真 (021)64411325

网 址 www.lixinaph.com E-mail lxaph@sh163.net

网上书店 www.shlx.net Tel: (021)64411071

经 销 各地新华书店

印 刷 上海申松立信印刷有限责任公司

开 本 787 毫米×1092 毫米 1/16

印 张 11 插 页 4

字 数 214 千字

版 次 2010 年 12 月 第 1 版

印 次 2010 年 12 月 第 1 次

书 号 ISBN 978 - 7 - 5429 - 2688 - 3/C

定 价 50.00 元

如有印订差错, 请与本社联系调换

《智能信息管理研究》编委会

编委会主任 刘念祖

编委会副主任 王汉兴 王双成

编委会成员 (按姓氏笔画排序)

王小玲 王双成 王汉兴 方茂中

刘念祖 那丽春 李 荣 张 明

单 薇 赵斯泓 曹顺良 傅云斌

前 言

智能信息管理学科建设获得上海市市本级财政资助。智能信息管理学科是综合了管理科学、计算机科学与数学科学,而形成的一门交叉学科。按照信息科学(理论)、信息技术(技术)和信息管理(应用)三个方面,我们确定了智能信息管理学科的三个研究方向,即智能信息处理(Intelligent Information Processing)、网格信息技术(Grid Information Technology)和风险信息管理(Risk Information Management)。

信息作为人类社会的三大资源之一,与物质和能量相比,它具有一些很独特的性质。比如信息使用不会有损耗;信息可以大量复制;信息可以脱离所反映的事物被保存和传播;信息可以不断增长;所有事物都有信息。正是由于信息具有以上这些特征,人类的文明、社会的发展,甚至人类大脑的进化,都与信息有着非常密切的关系。智能信息管理学科将是运用数学的理论、计算机的技术、管理的思想进行信息的获取、信息的传输、信息的处理和信息的应用。

本书由刘念祖负责主编。第一章基于贝叶斯网络的数据挖掘研究由王双成执笔,内容包括数据挖掘简介;贝叶斯方法、贝叶斯公式、全概率公式和乘法公式以及贝叶斯概率;贝叶斯网络的概念、贝叶斯网络的类型、贝叶斯网络的构造与特点、贝叶斯网络的学习和贝叶斯网络的应用。第二章基于支持向量机的财务风险研究由张明执笔,内容包括国内外财务风险研究的情况;财务危机预兆分析、财务危机特征分析、财务危机原因分析以及财务危机预警系统;分类方法及其应用;支持向量机的分类方法及其相应问题分析;支持向量机在财务信用评价预测中的应用。第三章网格技术与应用研究由刘念祖、那丽春、王小玲执笔,内容包括基于 Web 服务的会计网格特点、系统结构与技术实现;基于 Web 服务的

图像检索网格的特点、运行架构与图像特征自学习代理服务组件;基于并发 workflow(网格)的财务监管数据汇聚模型结构、财务信息汇总树的进程部件描述、工作过程及其算法。第四章保险中的若干风险模型研究由王汉兴、赵斯泓、傅云斌执笔,内容包括经典风险模型;马氏风险模型及破产概率;离散的三项分布风险模型及破产概率;依生灭过程索赔的两险种风险模型及破产概率。第五章智能技术在统计数据分析中的应用由单薇、方茂中执笔,内容包括科技投融资生态系统特征和评价原则、评价指标体系的设计及界定;评价指标的模糊统计及优化设计、指标体系的模糊聚类综合评价;社会经济统计数据智能分析与预测,比如2010年上海世博会参观人数的预测,上海文化产业对区域经济拉动作用的实证研究。第六章智能审计模型研究及应用由曹顺良、李荣执笔,内容包括智能审计国内外研究现状;智能审计系统的应用体系架构;数据挖掘子系统与专家审计子系统的分析与设计;LixinAudit——一个智能审计系统;总结和展望。

总之,研究高效率、高精度的智能信息处理算法,基于现代网格计算技术,从而实现智能风险管理信息系统。在此基础上,将逐步交叉渗透到学校会计与治理、开放经济与风险管理两大学科群之中。在理论、方法和技术上提供了有力的支持;从定性到定量,建立相关数学模型,提升重点项目的研究水平。

本书的出版得到上海立信会计学院领导、相关职能部门和数学与信息学院的老师以及立信会计出版社编辑的大力支持,在此一并表示衷心的感谢。

刘念祖

2010年12月于松江大学园区

目 录

第 1 章 基于贝叶斯网络的数据挖掘研究	1
1.1 数据挖掘简介	1
1.2 贝叶斯概率	2
1.2.1 贝叶斯方法	2
1.2.2 贝叶斯公式	3
1.2.3 全概率公式和乘法公式	3
1.2.4 贝叶斯概率	4
1.3 贝叶斯网络	6
1.3.1 贝叶斯网络的概念	6
1.3.2 贝叶斯网络的类型	7
1.3.3 贝叶斯网络的构造与特点	7
1.3.4 贝叶斯网络的学习	8
1.3.5 贝叶斯网络的应用	13
参考文献	16
 第 2 章 基于支持向量机的财务风险研究	 17
2.1 研究综述	17
2.1.1 国外研究综述	18
2.1.2 国内研究综述	19
2.2 财务危机分析	21
2.2.1 预兆分析	21
2.2.2 特征分析	22
2.2.3 原因分析	23
2.2.4 财务危机预警系统	24

2.3 分类方法及其应用	26
2.3.1 线性判别分析法	26
2.3.2 Logistic 回归模型法	27
2.3.3 其他分类方法	29
2.4 支持向量机分类方法及其相应问题分析	29
2.4.1 支持向量机简介	30
2.4.2 函数集的学习性能和 VC 维	34
2.4.3 结构风险最小化	36
2.4.4 高斯核函数	37
2.4.5 SVM 多分类问题	39
2.5 支持向量机在财务信用评价预测中的应用	42
2.5.1 基于支持向量机方法的模型建立及其过程	42
2.5.2 传统方法与支持向量机方法的比较和分析	42
参考文献	44
 第 3 章 网络技术与应用研究	48
3.1 基于 Web 服务的会计网络研究与实现	48
3.1.1 引言	48
3.1.2 会计网络的特点与系统结构	49
3.1.3 会计网络的技术实现	50
3.2 基于 Web 服务的图像检索网络研究与实现	53
3.2.1 引言	53
3.2.2 图像检索网络的特点与运行架构	54
3.2.3 图像特征自学习代理服务组件	55
3.3 基于并发 workflow(网络)的财务监管数据汇聚模型	56
3.3.1 引言	56
3.3.2 GMFD 模型结构	57
3.3.3 财务信息汇总树的进程部件描述	59
3.3.4 GMFD 模型工作过程及其算法	59
3.3.5 分析与实验	62
参考文献	64

第 4 章 保险中的若干风险模型研究	67
4.1 风险模型综述	67
4.1.1 经典风险模型	67
4.1.2 风险模型研究中的两种经典方法	69
4.1.3 风险模型研究中的有代表性方向	74
4.2 马氏风险模型及破产概率	76
4.2.1 马氏风险模型	76
4.2.2 破产概率的估计	77
4.2.3 指数分布情形下的破产概率	81
4.2.4 混合指数分布情形下的破产概率	87
4.3 离散的两项分布风险模型及破产概率	90
4.3.1 离散的两项分布风险模型	90
4.3.2 破产概率的解析解	92
4.3.3 破产概率的渐进解	98
4.4 依生灭过程索赔的两险种风险模型及破产概率	104
4.4.1 生灭过程简介	104
4.4.2 依生灭过程索赔的两险种风险模型	108
4.4.3 破产概率方程	109
4.4.4 破产概率收敛速率的上界	113
参考文献	117
 第 5 章 智能技术在统计数据分析中的应用	121
5.1 统计数据分析指标的优化设计与模糊智能评价	121
5.1.1 引言	121
5.1.2 科技投融资生态系统特征和评价原则	122
5.1.3 科技投融资生态系统评价指标体系的设计及界定	124
5.1.4 评价指标的模糊统计及优化设计	131
5.1.5 指标体系的模糊聚类综合评价	135
5.2 社会经济统计数据智能分析与预测	141
5.2.1 2010 年上海世博会参观人数的预测	141
5.2.2 上海文化产业对区域经济拉动作用的实证研究	146

参考文献	152
第 6 章 智能审计模型研究及应用	154
6.1 引言	154
6.2 国内外研究现状	154
6.3 智能审计系统的应用体系架构	156
6.4 几个关键子系统的分析与设计	157
6.4.1 数据挖掘子系统的分析与设计	157
6.4.2 专家审计子系统的分析与设计	158
6.5 LixinAudit —— 一个智能审计系统	160
6.6 总结和展望	161
参考文献	162

Contents

Chapter 1 Research of Data Mining Based on Bayesian Networks	1
1.1 Brief Introduction to Data Mining	1
1.2 Bayesian Probability	2
1.2.1 Bayesian Method	2
1.2.2 Bayesian Formula	3
1.2.3 Full-Probability Formula and Multiplication Formula	3
1.2.4 Bayesian Probability	4
1.3 Bayesian Networks	6
1.3.1 Concepts of Bayesian Networks	6
1.3.2 Types of Bayesian Networks	7
1.3.3 Structure and Characters of Bayesian Networks	7
1.3.4 Learning of Bayesian Networks	8
1.3.5 Applications of Bayesian Networks	13
References	16
 Chapter 2 Research of Financial Risk Based on Support Vector Machine	 17
2.1 Research Summarization	17
2.1.1 Overseas Research Summarization	18
2.1.2 Domestic Research Summarization	19
2.2 Financial Crisis Analysis	21
2.2.1 Omen Analysis	21
2.2.2 Characteristic Analysis	22
2.2.3 Reason Analysis	23

2.2.4 Financial Crisis Prediction System	24
2.3 Classification Method and Applications	26
2.3.1 Linear Discriminate Analysis	26
2.3.2 Logistic Regression Model	27
2.3.3 Other Classification Methods	29
2.4 SVM Classification Method and Corresponding Problem Analysis	29
2.4.1 Brief Introduction to SVM	30
2.4.2 Study Performance of Function Set and VC Dimension	34
2.4.3 Structural Risk Minimization	36
2.4.4 Gaussian Kernel Function	37
2.4.5 SVM Multi-Classification Problem	39
2.5 Application of SVM on Finance Credit Scoring	42
2.5.1 Model Building and Process Based on SVM Method	42
2.5.2 Comparison and Analysis of Traditional Method and SVM Method	42
References	44

Chapter 3 Research of Technique and Application on Grid 48

3.1 Research and Implementation Based on Web Service for Accounting Grid	48
3.1.1 Introduction	48
3.1.2 Characteristics and System Architecture of Accounting Grid	49
3.1.3 Technique Implementation of Accounting Grid	50
3.2 Research and Implementation Based on Web Service for Image Retrieval Grid	53
3.2.1 Introduction	53
3.2.2 Characteristics and Running Architecture of Image Retrieval Grid	54
3.2.3 Proxy Service Module of Image Features Auto-Learning	55
3.3 Gather Model Based on Concurrent Workflow (Grid) for Finance	

Monitoring Data	56
3. 3. 1 Introduction	56
3. 3. 2 GMFD Model Architecture	57
3. 3. 3 Description of Process Unit on FT	59
3. 3. 4 Working Routine and Algorithms of GMFD Model	59
3. 3. 5 Analysis and Experiment	62
References	64
Chapter 4 Research of Some Risk Models on Insurance	67
4. 1 Summarization to Risk Model	67
4. 1. 1 Classical Risk Model	67
4. 1. 2 Two Classical Methods on Research of Risk Model	69
4. 1. 3 Representative Direction on Research of Risk Model	74
4. 2 Markov Risk Model and Ruin Probability	76
4. 2. 1 Markov Risk Model	76
4. 2. 2 Approximation of Ruin Probability	77
4. 2. 3 Ruin Probability in Case Exponential Distribution	81
4. 2. 4 Ruin Probability in Case Mixed Exponential Distribution	87
4. 3 Trinomial Distribution Risk Model in Discrete Setting and Ruin Probability	90
4. 3. 1 Trinomial Distribution Risk Model in Discrete Setting	90
4. 3. 2 Recursive Formulae and Analytical Solutions for Ruin Probability	92
4. 3. 3 Asymptotic Formulae for Ruin Probability	98
4. 4 Bi-Type Insurance Risk Model with Claims by Birth-Death Process and Ruin Probability	104
4. 4. 1 Introduction to Birth-Death Process	104
4. 4. 2 Bi-Type Insurance Risk Model with Claims by Birth-Death Process	108
4. 4. 3 Equation for Ruin Probability	109

4.4.4 Upper Bound of Convergence Rate in Ruin Probability	113
References	117

Chapter 5 Intelligent Technology Application in Statistics Data

Analysis	121
5.1 Optimization of Statistics Data Analysis Criteria and Evaluation of Fuzzy Intelligent	121
5.1.1 Introduction	121
5.1.2 Characteristics and Evaluation Principle of Sciences Investment and Financing Ecological System	122
5.1.3 Criteria System Design and Definition of Sciences Investment and Financing Ecological System	124
5.1.4 Fuzzy Statistics and Optimization of Evaluation Criteria	131
5.1.5 Fuzzy Cluster Comprehensive Evaluation of Criteria System	135
5.2 Data Intelligent Analysis and Prediction of Society Economics Statistics	141
5.2.1 Number of Audience Prediction on Expo 2010 Shanghai	141
5.2.2 Empirical Study of Shanghai Cultural Industry Stimulating Effect on Regional Economy	146
References	152

Chapter 6 Research and Application of Intelligent Audit Model

6.1 Introduction	154
6.2 Current Research at Home and Abroad	154
6.3 Application Architecture of Intelligent Audit System	156
6.4 Analysis and Design of Some Key Subsystems	157
6.4.1 Analysis and Design of Data Mining System	157
6.4.2 Analysis and Design of Expert System	158
6.5 Lixin Audit — A Intelligent Audit System	160
6.6 Summarize and Prospect	161
References	162

第 1 章 基于贝叶斯网络的数据挖掘研究

1.1 数据挖掘简介

数据是没有经过加工的原材料,一般不适合于直接使用,但在数据中往往蕴含着有用的知识,具有潜在的应用价值。知识是人类在认识和改造客观世界的过程中所积累的经验的结晶,而来源于数据的知识,则是数据的抽象和概括。在现实中,数据的获得或积累往往比较容易,但数据的直接利用价值很有限,从数据中获得所需要的有用知识,是人们追求的目标。知识发现是研究如何从数据中发现有用知识的一个研究领域,从数据中发现知识是一个需要多个环节的复杂过程,数据挖掘是这个链条上的一个关键环节。多阶段的数据库中知识发现的处理过程模型,如图 1.1.1 所示。

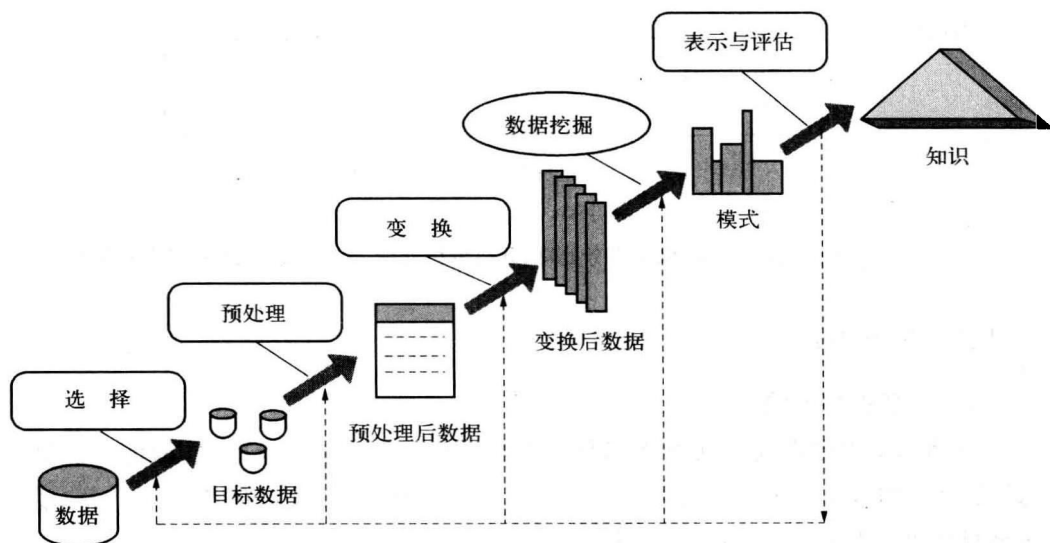


图 1.1.1 多阶段的数据库中知识发现的处理过程模型

数据库知识发现(KDD, Knowledge Discovery in Database)的处理过程模型共分九个处理阶段。这九个处理阶段分别是数据准备、数据选择、数据预处理、数据缩减、KDD目标确定、挖掘算法确定、数据挖掘(DM, Data Mining)、模式解释和知识评价,其中数据挖掘是核心。

联机分析处理主要用于决策支持,但人们逐渐发现数据挖掘能够为决策者提供更强有力的支撑。数据挖掘技术从数据库中获取决策者所需要的有用知识,实现人机结合,可提高决策的科学性,在传统的联机分析处理决策支持系统中融入数据挖掘技术的新决策支持框架。数据库、数据仓库、联机事务处理、联机分析处理、数据挖掘和决策支持关系图,如图 1.1.2 所示。

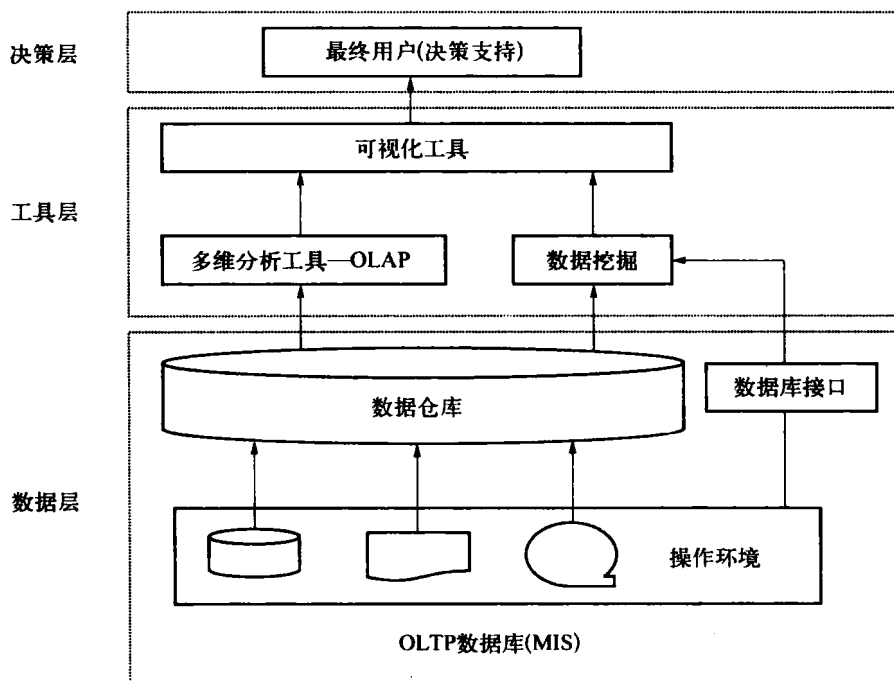


图 1.1.2 数据库、数据仓库、联机事务处理、联机分析处理、数据挖掘和决策支持关系图

1.2 贝叶斯概率

1.2.1 贝叶斯方法

(1) 将未知参数看成随机变量(或随机向量),记为 θ [$\theta = p(x)$ 在实验前的情况],于是当 θ 已知时, $x_1, \dots, x_n$ 的联合分布密度 $p(x_1, \dots, x_n; \theta)$ 就看成是 $x_1, \dots, x_n$ 对 θ 的条件密度,记为 $p(x_1, \dots, x_n | \theta)$ 或简写为 $p(x | \theta)$ 。

(2) 设法确定先验分布 $\pi(\theta)$ 。这是根据以往对参数 θ 的知识来确定的,是贝叶斯方

法中易引起争议的一步。

(3) 利用条件分布密度 $p(x_1, \dots, x_n | \theta)$ (似然函数) 和先验分布 $\pi(\theta)$, 可以求出 $x_1, \dots, x_n$ 与 θ 的联合分布和样本 $x_1, \dots, x_n$ 的分布(边界似然), 于是就可利用它们求得 θ 对 $x_1, \dots, x_n$ 的条件分布密度, 也就是用贝叶斯公式求得后验分布密度 $p(\theta | x_1, \dots, x_n)$ 。

(4) 利用后验分布密度 $p(\theta | x_1, \dots, x_n)$ 作出对 θ 的推断(估计或对 θ 作检验)。

$$p(\theta | x_1, \dots, x_n) = \frac{\pi(\theta)p(x_1, \dots, x_n | \theta)}{p(x_1, \dots, x_n)}$$

$$p(x_1, \dots, x_n) = \int \pi(\theta)p(x_1, \dots, x_n | \theta)d\theta$$

$$p(x | x_1, \dots, x_n) = \int p(x | \theta)p(\theta | x_1, \dots, x_n)d\theta =$$

$$\int \theta p(\theta | x_1, \dots, x_n)d\theta =$$

$$E_{p(\theta | x_1, \dots, x_n)}(\theta)$$

即实验后的概率是实验前的概率 $[p(\theta | x_1, \dots, x_n)]$ 的加权平均。

1.2.2 贝叶斯公式

(1) 贝叶斯公式的事件形式:

$$p(A_i | B) = \frac{p(A_i)p(B | A_i)}{\sum_{i=1}^k p(A_i)p(B | A_i)}$$

(2) 贝叶斯公式的随机变量形式:

$$g_{\xi|\eta}(x | y) = \frac{p(x)f_{\eta\xi}(y | x)}{\int_{-\infty}^{\infty} p_\xi(x)f_{\eta\xi}(y | x)dx}$$

$$f_{\eta\xi}(y | x) = \frac{q(y)g_{\xi\eta}(y | x)}{\int_{-\infty}^{\infty} q_\eta(x)g_{\xi\eta}(x | y)dy}$$

1.2.3 全概率公式和乘法公式

1. 全概率公式

(1) 全概率公式的事件形式:

$$p(A) = \sum_{i=1}^n p(B_i)p(A | B_i)$$