

信息管理专业前沿论丛

学科交叉知识发现及 可视化

● 魏建香 著

南京大学出版社

国家社科基金项目“基于聚类分析的学科交叉研究”（09CTQ022）

信息管理专业前沿论丛

学科交叉知识发现及 可视化

魏建香 著

南京大学出版社

图书在版编目(CIP)数据

学科交叉知识发现及可视化 / 魏建香著. — 南京:
南京大学出版社, 2011. 12

ISBN 978 - 7 - 305 - 09535 - 1

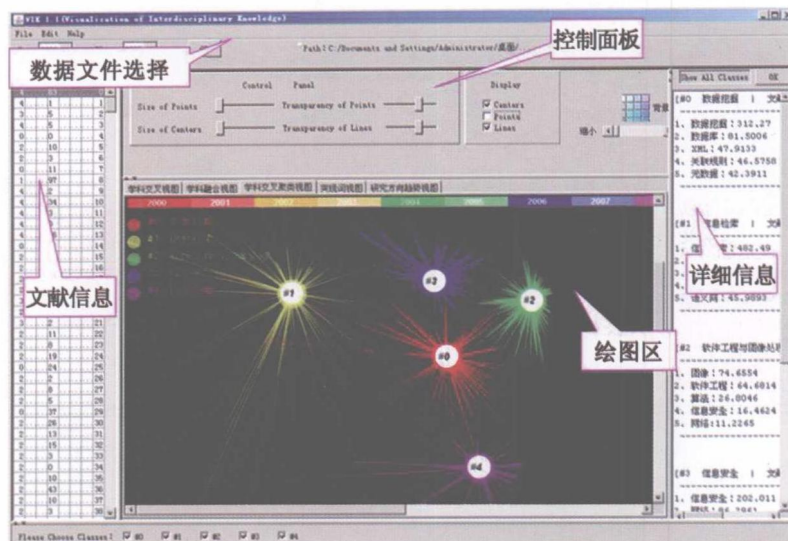
I. ①学… II. ①魏… III. ①交叉科学—研究 IV.
①G301

中国版本图书馆 CIP 数据核字(2012)第 000316 号

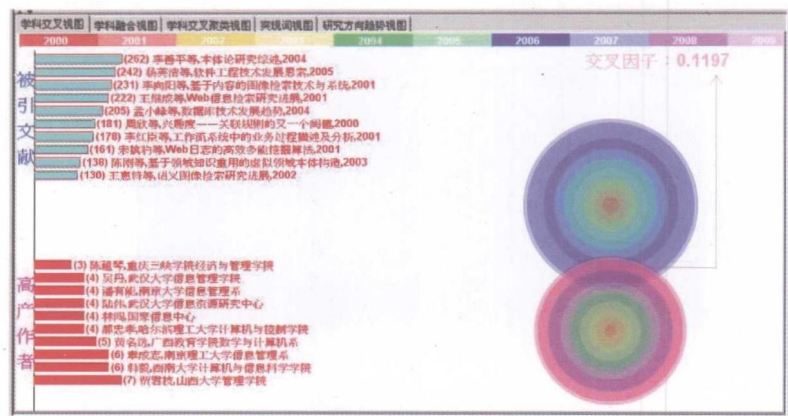
出版发行 南京大学出版社
社 址 南京市汉口路 22 号 邮 编 210093
网 址 <http://www.NjupCo.com>
出 版 人 左 健
书 名 学科交叉知识发现及可视化
著 者 魏建香
责任编辑 王 年 编辑热线 025 - 83686531
照 排 南京南琳图文制作有限公司
印 刷 南京京新印刷厂
开 本 880×1230 1/32 印张 6.375 字数 160 千
版 次 2011 年 12 月第 1 版 2011 年 12 月第 1 次印刷
ISBN 978 - 7 - 305 - 09535 - 1
定 价 30.00 元
发行热线 025 - 83594756 83686452
电子邮箱 Press@NjupCo.com
Sales@NjupCo.com(市场部)

* 版权所有,侵权必究

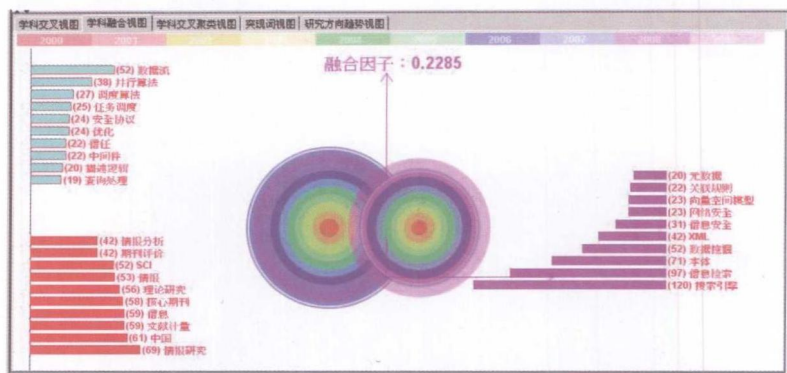
* 凡购买南大版图书,如有印装质量问题,请与所购
图书销售部门联系调换



VIK 主界面



学科交叉视图



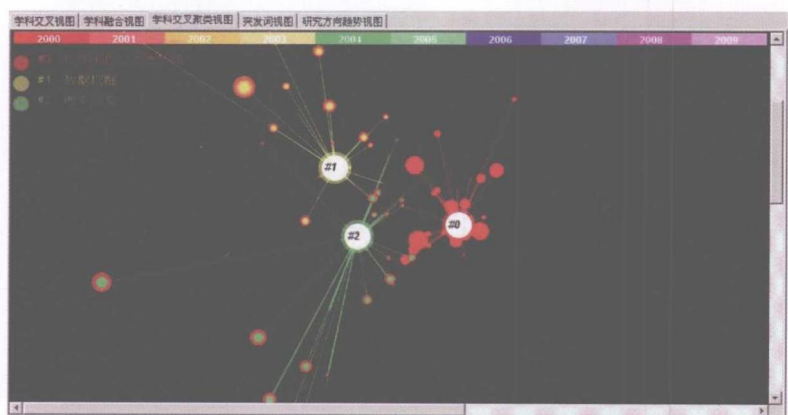
学科交叉共词视图



学科交叉聚类视图

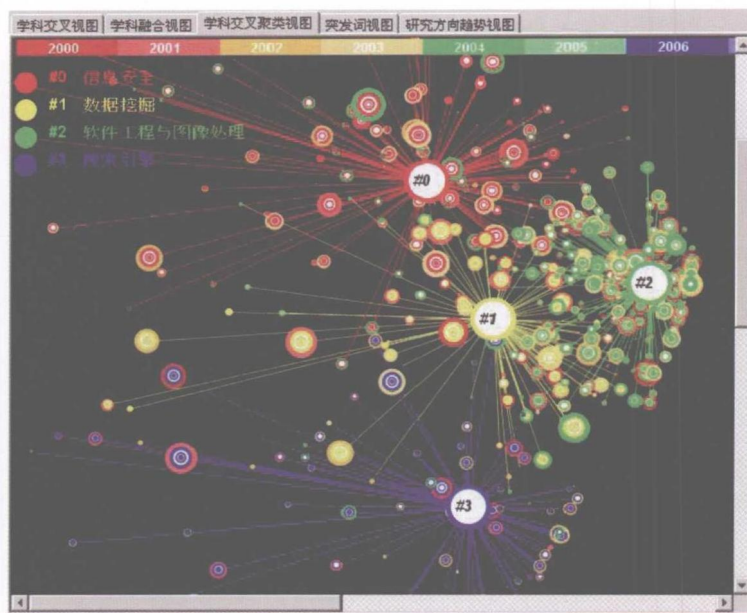


低透明模式下文献圆

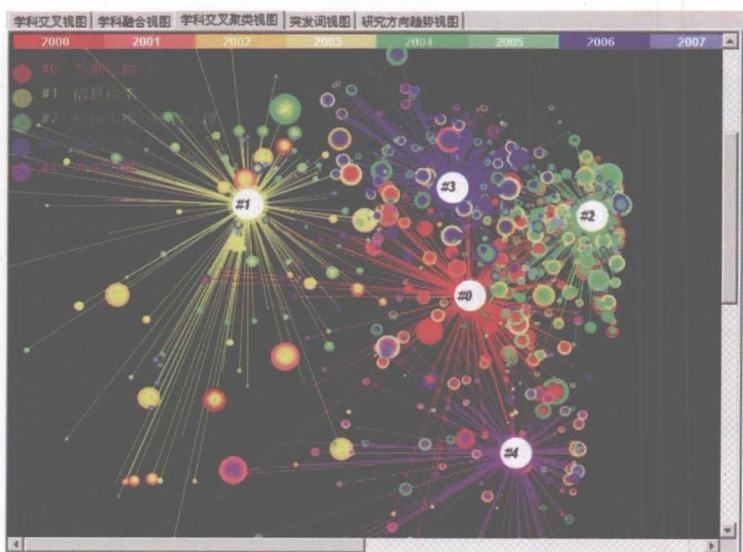


3个研究方向视图

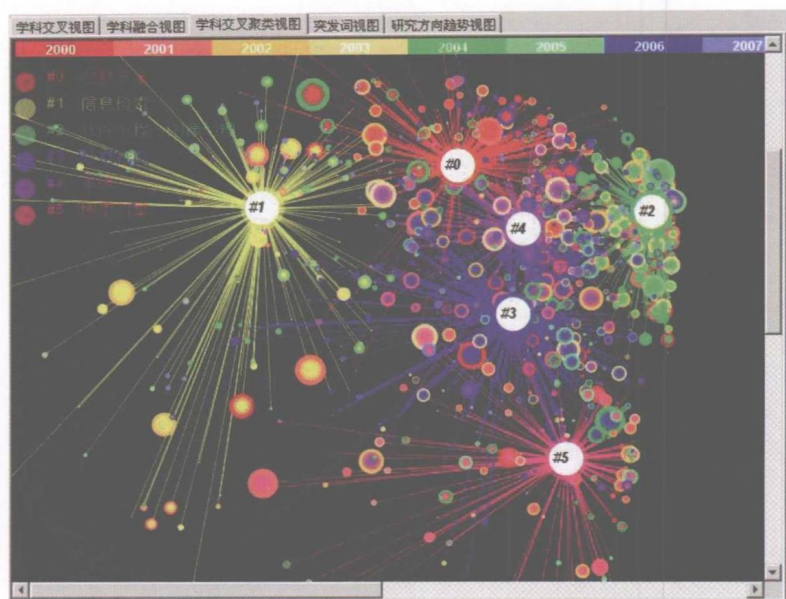
4 学科交叉知识发现及可视化



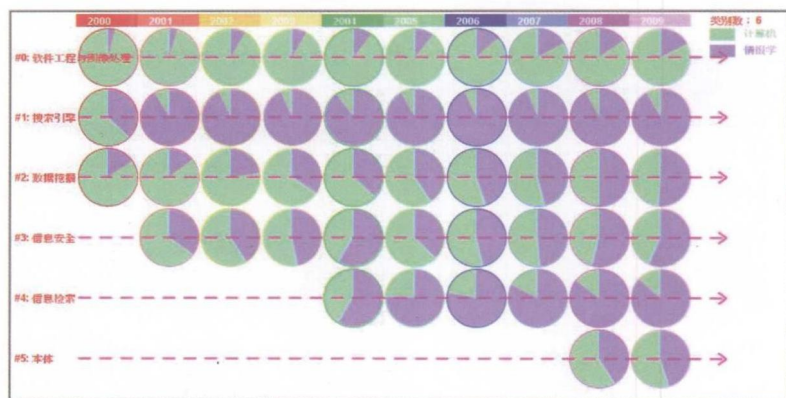
4 个研究方向视图



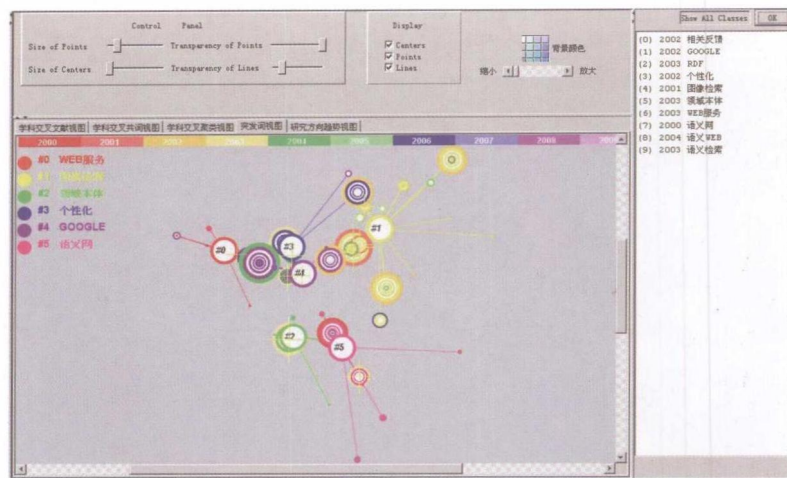
5 个研究方向视图



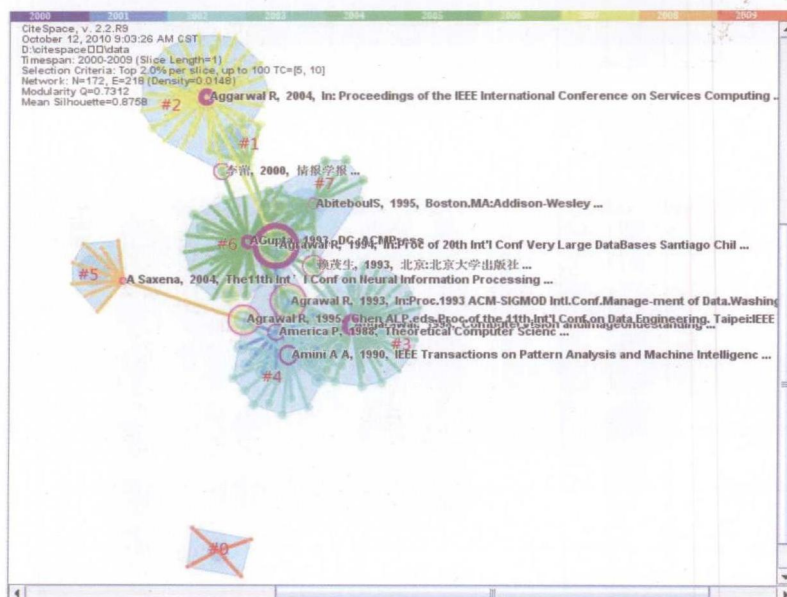
6个研究方向视图



学科交叉研究方向发展趋势



突现词聚类视图



基于 CiteSpace 的引文文献聚类视图

自序

这几年,我一直在从事文档聚类研究.在研究的过程中,经常会出现聚类的效果不尽人意的现象.不管采用什么聚类算法,科技文献的分类结果总是不稳定.究其原因,发现是某些文献的学科特征的模糊性和不确定性所导致,它们属于多个学科研究的交叉领域.我从而产生了利用文献聚类研究学科交叉的想法.

大量的资料表明,目前的许多最新研究成果都属于交叉学科.多学科交叉研究已经成为知识创新的一个主流.随着科技进步和社会发展,许多复杂问题单纯依赖单个学科已经无法解决,需要综合多个学科的知识 and 能力,这一需求促进了学科之间的交叉与融合.在国家层面上,国家宏观科技政策、项目立项等都大力提倡和鼓励多学科交叉研究;一个科研单位在人才培养与引进、科研团队建设、科研经费投入时也需要充分考虑学科之间的平衡与学科融合问题;作为科学研究者自身在从事某一领域研究的同时,也不能“一叶障目,不见森林”,要注重从其他学科吸收知识和营养,拓展自己的研究视野.学科交叉研究成果日益丰富,如何从海量文献中有效地挖掘学科之间的交叉知识成为本书要解决的主要问题.

我所学习的专业包括:数学、计算机和情报学,自认为在数学建模、算法设计和文献信息处理等方面有着优越的专业背景,为学科交叉的研究奠定了一些基础.但一本好书需要知识的科学性、系统性、全面性,自己的能力和水平与此要求还有相当的距离,只能尽自己最大的努力缩小这个差距.

本书的研究内容主要包括聚类算法、知识发现与信息可视化技术,因此更偏重技术研究,需要读者对一些相关技术有前期的了解.在写作的过程中,我尽可能地以通俗、详细的方式将相关技术知识呈现给大家,但由于自身水平以及文献资料和数据的准备等方面存在的问题,不一定能使所有读者都满意.

本书从起草到完稿得到我的导师——南京大学信息管理系的苏新宁教授的大力指导,在此表示衷心的感谢.也要感谢我的学生杜振东、李晓迪、夏真友等和同事朱云霞、薛景等,他们在数据处理、实验仿真、书稿校对中做了大量卓有成效的工作.感谢家人对我的关心和支持,同时要感谢南京大学出版社的大力支持.

本书的部分研究成果已发表或录用,详见参考文献[131]—[136].

魏建香

2011年7月于南京人口管理干部学院

目 录

1 绪论	1
1.1 学科交叉研究进展	3
1.2 主要内容	7
1.3 研究思路与技术路线	8
1.3.1 研究思路	8
1.3.2 技术路线	10
2 文本挖掘关键技术研究	12
2.1 文本挖掘概述	12
2.1.1 向量空间模型	14
2.1.2 相似度测度	16
2.1.3 评价函数	17
2.2 文本特征提取	19
2.2.1 互信息方法	20
2.2.2 特征频度与文档频度	20
2.2.3 χ^2 统计量方法	21
2.3 文本分类	22
2.4 文本聚类	25
2.5 一种改进的文档相似度模型	28
2.5.1 相似度计算模型	29
2.5.2 最大距离聚类算法	31
2.5.3 实验仿真与结果分析	33

2.6	最佳聚类数目确定方法	36
2.6.1	基于熵的评估方法	37
2.6.2	基于混合 F 统计量的评估方法	37
2.7	本章小结	39
3	聚类算法研究	40
3.1	模糊 C -均值聚类	40
3.1.1	FCM 基本原理	41
3.1.2	FCM 聚类算法	42
3.1.3	实验仿真	43
3.2	基于遗传算法的聚类	45
3.2.1	遗传算法	45
3.2.2	目标函数	46
3.2.3	算法实现	47
3.2.4	实验仿真	49
3.3	基于粒子群优化算法的聚类	50
3.3.1	粒子群算法简介	50
3.3.2	算法特点	52
3.3.3	文献聚类数学模型	53
3.3.4	文档聚类算法设计	54
3.3.5	算法仿真与分析	55
3.4	基于粒子对称分布多样性的 PSO 算法	59
3.4.1	多样性函数的定义及调节方法	61
3.4.2	算法设计	63
3.4.3	算法仿真与分析	64
3.5	基于免疫选择的粒子群优化算法	71
3.5.1	人工免疫算法	73
3.5.2	相关定义	74
3.5.3	算法设计	77

3.5.4	实验仿真与分析	79
3.6	基于 IS-PSO 与 FCM 的聚类融合算法	83
3.6.1	聚类融合算法的基本思想	83
3.6.2	聚类融合算法设计	84
3.6.3	实验仿真与分析	85
3.7	本章小结	87
4	知识可视化及降维技术研究	88
4.1	可视化概述	88
4.2	知识可视化	90
4.3	知识可视化软件	92
4.3.1	CiteSpace	93
4.3.2	Ucinet	95
4.4	降维技术	96
4.4.1	MDS	97
4.4.2	ISOMap	98
4.4.3	PCA	99
4.5	三种降维方法的比较	102
4.5.1	降维质量评价函数	102
4.5.2	降维效果比较	103
4.6	基于 PCA 可视化实例	105
5	学科交叉知识发现与可视化模型研究	110
5.1	学科交叉文献发现模型	110
5.2	学科交叉知识发掘模型	112
5.2.1	基于共词聚类的学科交叉点挖掘模型	113
5.2.2	基于互信息的研究方向描述	114
5.2.3	学科交叉突现词检测方法	114
5.2.4	基于发文比例的学科方向演变描述模型	116
5.3	学科交叉知识可视化模型	117

5.3.1	学科交叉度与融合度定义	117
5.3.2	学科交叉可视化模板	118
5.3.3	学科融合可视化模板	119
5.3.4	学科交叉聚类可视化模板	120
5.3.5	研究方向趋势可视化模板	122
5.3.6	突现词聚类可视化模板	123
5.4	本章小结	124
6	学科交叉知识发现及可视化实证研究	125
6.1	数据的采集与预处理	125
6.2	学科交叉文献的选择方法	126
6.2.1	学科交叉特征词选择与分析	126
6.2.2	学科交叉文献选择策略	129
6.3	基于融合算法的文档聚类	131
6.3.1	基于主题词表的特征约简	131
6.3.2	基于 F 统计量的最佳类别数确定	134
6.3.3	基于融合算法文档聚类的实现	138
6.3.4	基于互信息的类别特征描述	142
6.3.5	学科交叉因子与融合因子分析	147
6.3.6	突现词检测及聚类分析	149
6.4	基于 VIK 的学科交叉知识可视化	151
6.4.1	VIK 可视化软件简介	152
6.4.2	学科交叉视图	154
6.4.3	学科融合视图	156
6.4.4	学科交叉聚类视图	158
6.4.5	研究方向发展趋势视图	165
6.4.6	突现词聚类视图	167
6.5	基于 CiteSpace 的知识可视化分析	168
6.6	基于作者引用关系的社会复杂网络分析	172

6.7 本章小结	173
7 结语	174
7.1 主要工作	174
7.2 存在的不足	178
7.3 进一步的工作	179
主要参考文献	181

1 绪 论

科学是人类对客观世界认识的结晶,学科是对科学的细分,是人类认识客观世界实践经验的概括和总结.科学的基本使命是认识客观世界,是以概念、范畴、原理、定律等理论形态,按照周围世界的本来面貌反映周围世界,正确地揭示客观世界所有现象和过程的本质的、必然的规律.为了研究上的方便,人们按不同的标准将这些概念、范畴、原理、定律等分为不同的类别,构成我们今天的学科^[1].随着社会不断进步,学科正在朝着两个方向发展:一方面,随着研究的深入,一个学科正在逐渐细分为多个子学科,向更加专业化、精细化的方向发展;另一方面,由于许多现实的问题需要综合运用多个学科领域的知识,这种需求促进了学科之间在理论、方法及技术上的交叉融合,形成新的交叉学科.交叉学科是对原有学科分类体系的补充和发展,学科交叉是形成交叉学科的前提和基础.

对于学科交叉的概念,许多学者都提出了自己的看法和见解.中科院院士路甬祥^[2]认为:学科交叉是“学科际”或“跨学科”研究活动,其结果导致的知识体系构成了交叉科学.刘仲林^[3]在《现代交叉科学》一书中指出:所谓学科交叉,指跨出已有学科边界,实

[1] 贾剑平,郭凤英,王建坤.从“杂交优势”看学科交叉与交叉学科发展[J].西安电子科技大学学报(社会科学版),2005,15(4):116~120

[2] 路甬祥.学科交叉与交叉科学的意义[J].中国科学院院刊,2005,1:58~60

[3] 刘仲林.现代交叉科学[M].杭州:浙江教育出版社,1998