

信息科学技术学术著作丛书



CRC Press
Taylor & Francis Group

多媒体数据挖掘 系统引论

Zhongfei Zhang Ruofei Zhang 著
王兴起 张仲非 译

Multimedia Data Mining:
A Systematic Introduction to Concepts and Theory



科学出版社

信息科学技术学术著作丛书

多媒体数据挖掘系统引论

Zhongfei Zhang Ruofei Zhang 著

王兴起 张仲非 译



科学出版社

北京

图字:01-2011-3735

内 容 简 介

本书全面系统地介绍了多媒体数据挖掘的基本概念与经典体系结构,多媒体数据挖掘的基本理论(包括多媒体数据特征和知识表示、统计数据挖掘理论和基于软计算的理论),多媒体数据挖掘理论解决实际多媒体数据挖掘问题的具体实例(包括图像检索和挖掘、图像语义标注、视频检索,以及音频分类)等。同时,介绍了多媒体数据挖掘的最新研究成果和应用前景。

本书可作为计算机科学、人工智能等相关专业高年级本科生和研究生的参考教材,也可作为从事多媒体数据挖掘及其相关领域研究的专业人员和系统开发工程师的参考用书。

Multimedia Data Mining: A Systematic Introduction to Concepts and Theory, by Zhongfei Zhang, Ruofei Zhang.

Copyright© 2010 by Taylor & Francis Group LLC. All Rights Reserved.

Authorized translation from English language edition published by CRC Press, part of Taylor & Francis Group LLC.

本书封面贴有 Taylor & Francis 集团防伪标签,未贴防伪标签属未获授权的非法行为。

图书在版编目(CIP)数据

多媒体数据挖掘系统引论/(美)张仲非(Zhongfei Zhang),张若非著;王兴起,张仲非译. —北京:科学出版社,2018. 5

(信息科学技术学术著作丛书)

书名原文:Multimedia Data Mining: A Systematic Introduction to Concepts and Theory

ISBN 978-7-03-057186-1

I. ①多… II. ①张…②张…③王… III. ①多媒体-数据采集 IV. ①TP274

中国版本图书馆 CIP 数据核字(2018)第 081445 号

责任编辑:魏英杰 / 责任校对:郭瑞芝

责任印制:师艳茹 / 封面设计:陈 敬

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

艺堂印刷(天津)有限公司 印刷

科学出版社发行 各地新华书店经销

*

2018 年 5 月第 一 版 开本:720×1000 1/16

2018 年 5 月第一次印刷 印张:15

字数:298 000

定价:90.00 元

(如有印装质量问题,我社负责调换)

关于作者

Zhongfei (Mark) Zhang: 纽约州立大学宾汉姆顿分校计算机科学系终身教授,多媒体实验室主任。他在浙江大学获得电子工程学士学位和信息科学硕士学位,马萨诸塞大学阿姆斯特分校获得计算机科学博士学位。Zhongfei (Mark) Zhang 曾任教于纽约州立大学布法罗分校计算机科学与工程系和文本分析与识别研究中心,研究领域包括多媒体信息索引与检索、数据挖掘和知识发现、计算机视觉和图像理解、模式识别、生物信息学等。他是国际学术界中最早从事多模态数据挖掘和多模态信息检索的研究者之一,主持过多项受美国联邦政府、纽约州政府和企业等资助的相关领域科研项目,拥有多项发明专利,是多个会议和期刊的审稿人和程序委员会成员。2000 年以来,一直是美国联邦政府基金机构(如美国国家科学基金会、美国国家航空航天局等)、纽约州政府基金机构和一些民间基金机构评审专家。他是多个期刊的编委,受聘于国际多所著名高等院校和科研机构,作为技术顾问服务于多个企业和政府机构,也是多个著名奖项的获得者。

Ruofei Zhang: 微软全球合伙人,人工智能与研究院高级总监。Ruofei Zhang 的研究领域包括机器学习、数据挖掘、大数据、计算广告学、个性化推荐、算法优化、计算机视觉和多媒体信息检索,以及这些技术在互联网及智能系统的应用。他发表高水平论文 40 余篇,获得美国专利 11 项,是美国国家科学基金会智能系统委员会项目评委,IEEE 和 ACM 高级会员。

《信息科学技术学术著作丛书》序

21 世纪是信息科学技术发生深刻变革的时代,一场以网络科学、高性能计算和仿真、智能科学、计算思维为特征的信息科学革命正在兴起。信息科学技术正在逐步融入各个应用领域并与生物、纳米、认知等交织在一起,悄然改变着我们的生活方式。信息科学技术已经成为人类社会进步过程中发展最快、交叉渗透性最强、应用面最广的关键技术。

如何进一步推动我国信息科学技术的研究与发展;如何将信息技术发展的新理论、新方法与研究成果转化为社会发展的新动力;如何抓住信息技术深刻发展变革的机遇,提升我国自主创新和可持续发展的能力? 这些问题的解答都离不开我国科技工作者和工程技术人员的求索和艰辛付出。为这些科技工作者和工程技术人员提供一个良好的出版环境和平台,将这些科技成就迅速转化为智力成果,将对我国信息科学技术的发展起到重要的推动作用。

《信息科学技术学术著作丛书》是科学出版社在广泛征求专家意见的基础上,经过长期考察、反复论证之后组织出版的。这套丛书旨在传播网络科学和未来网络技术,微电子、光电子和量子信息技术、超级计算机、软件和信息存储技术,数据知识化和基于知识处理的未来信息服务业,低成本信息化和用信息技术提升传统产业,智能与认知科学、生物信息学、社会信息学等前沿交叉科学,信息科学基础理论,信息安全等几个未来信息科学技术重点发展领域的优秀科研成果。丛书力争起点高、内容新、导向性强,具有一定的原创性;体现出科学出版社“高层次、高质量、高水平”的特色和“严肃、严密、严格”的优良作风。

希望这套丛书的出版,能为我国信息科学技术的发展、创新和突破带来一些启迪和帮助。同时,欢迎广大读者提出好的建议,以促进和完善丛书的出版工作。

中国工程院院士

原中国科学院计算技术研究所所长



原 书 序

能向大家介绍这本多媒体数据挖掘方面的重要著作,我感到非常荣幸。得知该书作者是这一领域中最活跃的两位年轻学者时,我倍感欣慰。这本书在这一领域发展的初期就得以出版,而该领域比其他领域更加需要这样的一本书。在大多数新兴的研究领域,图书在将一个领域带向成熟的过程中起着非常重要的作用。研究领域是随着研究成果及发布而发展的。然而,研究论文所提供的研究领域的内容、应用前景、已经提出的方法,以及需要解决的技术毕竟有限,这本书正好给我们提供了这样一个机会。我非常喜欢写关于某一领域的一本书这样的想法,因为它通过将研究论文中已经讨论的、不易发现和理解的不同主题内容整合在了一起。在我看到这本书之前,就一直支持这本书的出版,出版这本书是这两位思维活跃的年轻学者美好和勇敢的愿望,现在这本书的内容就在我的电脑屏幕上,这更坚定了我的想法。

多媒体数据挖掘作为一个研究领域逐渐得到大家的认可是在 20 世纪 90 年代,多媒体处理、存储、通信、获取和显示等相关技术已经取得长足发展,使得科研技术人员能够通过构造一些方法,将音频、图像、视频和文本等的多种形式的信息组合在一起。多媒体计算和通信技术已经意识到多种信息源间信息的相关性,以及任一单一信息源内信息的不足,希望通过正确地选择信息源来提供彼此间的互补信息,通过仅仅使用不同信息源的部分信息便能够实现对环境整体的描述。这样的系统有些类似于人类的感知系统。

数据挖掘是数据存储能力和处理速度发展的必然结果,当大规模数据存储及不同的统计计算用于分析所有可能,甚至是不可能的数据间的关联成为可能时,数据挖掘这一领域就诞生了。数据挖掘允许人们对数据间的关系做假设,然后探索对这些假设的支持,这一技术已经成功应用到许多领域中,并得到越来越多的应用。实际上,许多新领域的产生是由数据挖掘引发的,而且数据挖掘可能成为继许多新兴的自然和社会科学后的一个有力的计算工具。

考虑多媒体数据的数据量,以及构建跨越语义鸿沟的机器感知系统的难度,很自然地,多媒体与数据挖掘便结合在一起,可以应用到一些最富有挑战性的问题上。数据挖掘最困难的挑战都是由多媒体系统造成的。类似地,潜在的、最富价值的数据挖掘的应用也来源于多媒体数据。

自然而然,在一个领域发展的早期,人们常常只是对现有算法的改进,多媒体数据挖掘也不例外,许多早期的工具是用来处理图像这类的单一模态数据,这是一

一个好的开端,但真正的挑战还是处理多媒体数据,解决那些使用单一模态不能解决的问题。通过这种方式,多媒体数据挖掘可以成为通过各种方法相互影响使得分析取得进展的一个领域,我非常希望一些年轻的研究人员能够受到启发,关注这一值得研究的领域。

该书列出了多媒体数据挖掘领域的最新资料,具有三个显著特点。

第一,将多媒体数据挖掘的文献整合在一起,给出了相关定义,同时将该领域与其他具有较好理论基础的研究领域比较。

第二,涵盖最新的有关多媒体数据的基础理论,包括特征抽取和表示、知识表示、统计学习理论、软计算理论等。作者花费了大量的精力确保该书包含的理论和技术的反映的是最新研究成果,虽然不能穷尽,但是给出了多媒体数据挖掘理论基础的全面系统的介绍。

第三,为了向读者展示多媒体数据挖掘研究成果在实际领域的潜在应用,该书给出了多媒体数据挖掘理论用于解决实际多媒体数据挖掘问题的具体实例,包括图像检索和挖掘、图像语义标注、视频检索和挖掘和音频分类等。

多媒体数据挖掘的发展才刚刚起步,目前还处于飞速发展阶段,希望该书的出版可以引领和促进多媒体数据挖掘在学术领域、政府、企事业单位的研究和社会各行各业的应用。

Ramesh Jain

加州大学欧文分校

前 言

多媒体数据挖掘是一个多学科交叉的研究领域,由多媒体和数据挖掘两个学科发展而来。多媒体和数据挖掘都是最近十几年才发展起来的新兴学科,多媒体数据挖掘在最近几年才得到真正发展。本书是第一本有关多媒体数据挖掘的学术著作,全书自成体系,包括多媒体数据挖掘的相关基础、方法和应用,系统地阐述了相关的理论和概念,同时给出了具体应用,这些应用说明该领域研究的相关技术所具备的巨大潜力和影响。

我们多年来一直活跃于多媒体数据挖掘领域,书中内容是我们在该领域多年来长期研究工作的总结。本书既可以作为相关领域研究人员用书,也可以作为工程技术人员的参考书,同时还可以作为研究生高级研讨班之教材。此外,本书也可以用作研究生或高年级本科生的入门教材。书中提供的参考文献可供读者进一步阅读和参考。

由于多媒体数据挖掘是一个多领域的交叉学科,而且近几年来得到飞速发展,因此在一本书中包含该领域的全部内容是不可能的,我们尽量囊括相关研究主题的最新进展。对于从事多媒体数据挖掘或对该领域有一定了解的读者,本书可以作为对该领域知识的完整梳理,而对于刚刚进入该研究领域的读者,本书可以作为对该领域系统、形式化的介绍。

本书的完成离不开众多人员和机构的大力支持,这里特别感谢出版商——泰勒-弗朗西斯出版集团 CRC 出版社给了我们这个机会,使得本书能够作为由明尼苏达大学 Vipin Kumar 教授主编的“数据挖掘与知识发现系列丛书”中的一本与读者见面。同时,我们要感谢本书的编辑 Randi Cohen,感谢她热情、耐心的支持与宝贵的建议;感谢 Judith M. Simon 和众多的校对人员,他们对初稿中的错误进行的一丝不苟的努力。感谢国际排版公司 Shashi Kumar 给予的技术支持。感谢加州大学欧文分校的 Ramesh Jain 教授对本书的支持,并欣然为本书作序。感谢西北大学的 Ying Wu 教授、法国里尔理工大学的 Chabane Djeraba 教授和其他热心的读者,感谢他们阅读本书,并提出很多有价值的修改建议。本书的部分内容得益于作者和同事的工作,在此感谢下列同事对本书做出的贡献,他们是 Jyh-Herng Chow、Wei Dai、Alberto del Bimbo、Christos Faloutsos、Zhen Guo、Ramesh Jain、Mingjing Li、Weiyang Ma、Florent Masseglia、Jia-Yu (Tim) Pan、Ramesh Sarrukkai、Eric P. Xing 和 Hongjiang Zhang。本书的编写得到 Maria Zemankova 博士

主持的美国国家科学基金(IIS-0535162)的支持。书中任何结论、意见和建议都是作者本人的,不代表美国国家科学基金会的任何观点。

最后,还要感谢我们的家人,他们的爱与支持是完成这本书的基础。

目 录

《信息科学技术学术著作丛书》序

原书序

前言

第一部分 引 论

第 1 章 简介	3
1.1 多媒体数据挖掘定义	3
1.2 多媒体数据挖掘系统经典体系结构	6
1.3 本书内容与组织	7
1.4 本书受众	8
1.5 进一步读物	8

第二部分 理论和技术

第 2 章 多媒体数据特征与知识表示	13
2.1 引言	13
2.2 基本概念	14
2.2.1 数字采样	14
2.2.2 媒体数据类型	16
2.3 特征表示	18
2.3.1 统计特征	19
2.3.2 几何特征	23
2.3.3 元特征	26
2.4 知识表示	26
2.4.1 逻辑表示	26
2.4.2 语义网络	28
2.4.3 框架	29
2.4.4 约束	31
2.4.5 不确定性表示	33
2.5 小结	36

第3章 统计数据挖掘理论与技术	37
3.1 引言	37
3.2 贝叶斯学习	38
3.2.1 贝叶斯定理	38
3.2.2 贝叶斯最优分类器	40
3.2.3 Gibbs 抽样算法	41
3.2.4 朴素贝叶斯分类器	41
3.2.5 贝叶斯信念网络	42
3.3 概率潜在语义分析	45
3.3.1 潜在语义分析	46
3.3.2 潜在语义分析概率扩展	47
3.3.3 基于期望最大化的模型拟合	48
3.3.4 潜在概率空间与概率潜在语义分析	49
3.3.5 模型过拟合与强化的期望最大化算法	50
3.4 用于离散数据分析的隐含狄利克雷分配模型	51
3.4.1 隐含狄利克雷分配模型	52
3.4.2 与其他隐变量模型关系	54
3.4.3 隐含狄利克雷分配模型推理	56
3.4.4 隐含狄利克雷分配模型参数估计	58
3.5 层次狄利克雷过程	58
3.6 多媒体数据挖掘中的应用	60
3.7 支持向量机	60
3.8 面向结构化输出空间的间隔学习	65
3.9 Boosting	70
3.10 多示例学习	72
3.10.1 构建语义词空间与图像视觉代表对象空间映射	73
3.10.2 词到图像的查询	76
3.10.3 图像到图像的查询	76
3.10.4 图像到单词的查询	76
3.10.5 多模态查询	77
3.10.6 可扩展性分析	77
3.10.7 适应性分析	77
3.11 半监督学习	80
3.11.1 监督学习	83
3.11.2 半监督学习	84

3.11.3 半参数正则化最小二乘	87
3.11.4 半参数正则化支持向量机	88
3.11.5 半参数正则化算法	90
3.11.6 直推方法与半监督学习	91
3.11.7 与其他方法的比较	91
3.12 小结	92
第4章 基于软计算的理论与技术	93
4.1 引言	93
4.2 软计算方法特点	94
4.3 模糊集理论	95
4.3.1 模糊集基本概念和性质	95
4.3.2 模糊逻辑和模糊推理规则	97
4.3.3 模糊集在多媒体数据挖掘中的应用	98
4.4 人工神经网络	99
4.4.1 神经网络基本结构	99
4.4.2 神经网络中的监督学习	102
4.4.3 神经网络中的强化学习	106
4.5 遗传算法	109
4.5.1 遗传算法简述	109
4.5.2 遗传算法极值搜索与传统极值搜索方法比较	112
4.6 小结	116
 第三部分 多媒体数据挖掘应用实例	
第5章 图像数据库建模——语义库训练	119
5.1 引言	119
5.2 研究背景	119
5.3 相关工作	120
5.4 图像特征和视觉词典	122
5.4.1 图像特征	122
5.4.2 视觉词典	123
5.5 α -语义图与语义库模糊模型	125
5.5.1 α -语义图	126
5.5.2 语义库模糊模型	128
5.6 基于分类的检索算法	129

5.7 实验结果	131
5.7.1 给定数据库上的分类性能	132
5.7.2 基于分类的检索结果	133
5.8 小结	138
第6章 图像数据库建模——潜在语义概念发现	139
6.1 引言	139
6.2 研究背景和相关工作	139
6.3 基于区域的图像表示	141
6.3.1 图像分割	142
6.3.2 视觉符号目录	144
6.4 概率潜在语义模型	147
6.4.1 概率数据库模型	147
6.4.2 使用期望最大化构建模型	148
6.4.3 概念数估计	149
6.5 基于后验概率的图像挖掘与检索	150
6.6 算法分析	152
6.7 实验结果	153
6.8 小结	160
第7章 图像数据挖掘和概念发现的多模态方法	161
7.1 引言	161
7.2 研究背景	161
7.3 相关工作	162
7.4 概率语义模型	163
7.4.1 概率语义标注图像模型	164
7.4.2 基于期望最大化的模型拟合过程	165
7.4.3 概念数估计	166
7.5 基于模型的图像语义标注与多模态图像挖掘和检索	167
7.5.1 图像语义标注与图像到文本查询	167
7.5.2 文本到图像查询	168
7.6 实验	169
7.6.1 数据库与特征集合	169
7.6.2 评估度量	170
7.6.3 图像自动语义标注结果	171
7.6.4 单个文本到图像的查询结果	173
7.6.5 图像到图像的查询结果	174

7.6.6 与纯文本查询方法的性能比较结果	175
7.7 小结	176
第8章 视频数据库概念发现与挖掘	177
8.1 引言	177
8.2 研究背景	177
8.3 相关工作	178
8.4 视频分类	180
8.4.1 朴素贝叶斯分类器	180
8.4.2 最大熵分类器	182
8.4.3 支持向量机分类器	183
8.4.4 基于元数据与基于内容的分类器组合	184
8.5 查询分类	185
8.6 实验	186
8.6.1 数据集	186
8.6.2 视频分类结果	188
8.6.3 查询分类结果	193
8.6.4 查找相关性结果	193
8.7 小结	194
第9章 音频数据库概念发现与挖掘	196
9.1 引言	196
9.2 研究背景与相关工作	196
9.3 特征抽取	199
9.4 分类方法	201
9.5 实验结果	202
9.6 小结	205
参考文献	206

引 论

第1章 简介

1.1 多媒体数据挖掘定义

多媒体数据挖掘,顾名思义就是两个新兴领域——多媒体和数据挖掘的结合。然而,多媒体数据挖掘并不是多媒体和数据挖掘两个研究领域的简单结合。实际上,多媒体数据挖掘倾向于多媒体和数据挖掘两个领域融合的相关主题的研究,以提高和促进对建立在多媒体数据上的知识发现的理解和发展,因此多媒体数据挖掘作为一个完全独立的、不同的研究领域在一定程度上依赖于多媒体和数据挖掘两个领域的最新研究成果,同时从根本上又区别于多媒体和数据挖掘及其简单组合。

多媒体和数据挖掘都是多学科交叉研究领域,相较于其他比较成熟的计算机研究领域,如操作系统、程序设计语言、人工智能等,多媒体和数据挖掘发展历史较短,研究开始于20世纪90年代初期,属于比较年轻的研究领域。此外,随着大量的应用需求,最近几年两个研究领域都各自经历了快速发展的过程。

多媒体是一个不断发展的多学科交叉研究领域。“多媒体”一词指的是多种媒体形式的结合。20世纪90年代初期,计算机和数字技术飞速发展,多媒体作为一个新兴研究领域应运而生^[87,197],其研究内容是面向具体应用的高效多媒体系统的研发。从这一点来讲,多媒体技术的研究包含十分广泛的研究课题,如多媒体索引与检索、多媒体数据库、多媒体网络、多媒体信息表示、多媒体服务质量、多媒体使用与用户研究等,甚至多媒体标准,而这仅仅是列举了其中的一小部分。

多媒体技术是一个研究广阔的领域,与多媒体数据挖掘相关的研究内容主要包括多媒体索引与检索、多媒体数据库、多媒体信息表示^[72,113,198]等。如今,在众多应用中,多媒体信息技术已经成为一项无处不在、被广泛使用的技术,这使得多媒体数据库变得异常庞大,目前已有一些工具用于管理和检索这样庞大的数据库,但是在许多决策支持系统应用领域,从庞大的多媒体数据库中抽取隐含的、有用的知识的工具显得格外迫切。例如,我们目前极其需要有这样的一些工具,包括发现图像中的对象或区域之间的关系、根据图像的内容对其进行分类、识别声音中的模式、对语音和音乐进行归类、识别和跟踪视频中的对象等。

在寻找改进多媒体信息索引和检索方法的过程中,多媒体信息系统领域的研究人员正在寻求用于发现索引信息的新方法。各种技术和方法,包括机器学习、统计学、数据库、知识获取、数据可视化、图像分析、高性能计算和基于知识