

基于人工智能的 多媒体数据挖掘和 应用实例

张鸿 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

基于人工智能的 多媒体数据挖掘和 应用实例

张鸿 著

常州大学图书馆
藏书章



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

基于人工智能的多媒体数据挖掘和应用实例/张鸿著. —武汉:武汉大学出版社, 2018. 1

ISBN 978-7-307-19956-9

I. 基… II. 张… III. 人工智能—数据采集 IV. ①TP18
②TP274

中国版本图书馆 CIP 数据核字(2018)第 322035 号

责任编辑:王智梅

责任校对:汪欣怡

版式设计:马 佳

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:cbs22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:湖北金海印务有限公司

开本:720 × 1000 1/16 印张:15.75 字数:227 千字 插页:1

版次:2018 年 1 月第 1 版 2018 年 1 月第 1 次印刷

ISBN 978-7-307-19956-9 定价:42.00 元

版权所有,不得翻印;凡购买我社的图书,如有质量问题,请与当地图书销售部门联系调换。

致 谢

本书得到国家自然科学基金资助项目(No.61373109),以及武汉科技大学研究生教材专项基金资助,特此致谢。

前 言

近年来,随着人工智能技术和多媒体技术的飞速发展,人们周围环绕着各种类型的多媒体数据,如:文字、图像、音频、视频等。人工智能技术已经嵌入人们生活的方方面面,尤其是在信息检索领域,传统基于文字的查询方式,已经逐渐扩充为各种类型数据相融合的综合检索模式。多媒体数据挖掘技术的不断发展使得人们可以不受地域和时间的限制,看得更远、听得更多。信息世界正变得更加丰富和立体,更加唾手可得。

多媒体数据类型多样,可以从音、形、意等不同角度表达出丰富的语义信息。然而,从底层数据结构上来看,很多多媒体数据是非结构化和半结构化的,且普遍缺乏统一索引和有效标注。这就使得多媒体数据的科学管理和有效利用变得异常困难。同时,近年来随着电子信息技术的蓬勃发展,多媒体数据在数据量上飞速膨胀,形成异构的多媒体数据海洋。为了从这片数据汪洋中快速、准确地挖掘出有用的多媒体信息,大量的研究者关注了多媒体数据的存储管理和挖掘等研究工作,其中涉及人工智能、多媒体内容分析和检索、机器学习和统计分析等多个交叉领域的知识和技术。多媒体数据挖掘研究一方面使得多媒体数据汪洋变得更加井然有序,另一方面也实现了从数据到信息、从信息到知识的有效提取和科学利用。

计算机需要综合处理图像、音频等不同类型的多媒体数据,犹如人脑对外界事物的认知需要跨越视觉、听觉等不同感官传递的信息,以做出综合判断。那么,用计算机来处理图像、音频、视频等多媒体信息,就好比用人脑来处理视觉、听觉等多种感官传递的信号,只不过目前的人工智能技术还远远不及人脑的智能。这方面的研究工作也显得极为重要和意义深远,已有的研究成果还较为有限,仍需要有

大量的、创新性的研究工作来实现进一步的突破。

作者长期从事多媒体数据的特征分析和语义挖掘工作,尤其关注了不同类型多媒体数据的综合语义理解研究。该项研究一直是计算机视觉和人工智能界国内外研究中的热点,许多权威的国际期刊、高级别的国际会议,都陆续发表了多篇相关论文,我国自然科学基金委、科技部、教育部也常年资助一些与之相关的科研项目。而且,从实际应用方面来看,越来越多的软件开发项目将人工智能技术应用到各种多媒体数据场景中,如:各种基于手机客户端的多媒体数据浏览和检索软件。

本书从多媒体数据的底层特征分析到高层语义理解、从单一类型多媒体数据的知识挖掘到不同类型多媒体数据的综合理解等,多个角度详细介绍了基于人工智能的多媒体数据挖掘传统方法和前沿知识。更进一步的,结合目前的一些典型领域,列举了一些实际应用场景和支撑技术。力求理论与实际相结合,既有理论意义上的前沿算法,又有能够解决实际应用问题的具体案例。

本书共分为四个部分:

第一部分概述了多媒体数据挖掘的基本知识,包括从数据到语义的理解与挖掘,以及典型的图像语义挖掘系统——基于内容的图像检索(CBIR)。纯文本信息检索是最早被提出、也是应用最为广泛的一种文本数据挖掘方式,目前取得了广泛的应用和很好的效果;包括CBIR在内的基于内容的多媒体检索是一种结合了人工智能、机器学习、统计分析等多学科综合性的多媒体数据分析技术,针对特征高维、异构的多媒体数据,而提出的一种数据管理和信息挖掘的典型模式。

第二部分介绍了多媒体数据的底层特征分析技术。包括图像、音频、视频三种类型的多媒体信息的结构化表达方法,以及面向多模态数据的综合分析方法,该部分是本书的理论基础。数据表达是数据分析和检索的前提和基础,而多媒体数据通常是非结构化的,这一部分主要介绍了如何通过对底层特征的分析,计算各种类型的特征模型,实现数据结构化。

第三部分主要从跨媒体检索这一具体的应用角度,介绍了不同

类型多媒体信息的融合分析和跨媒体检索方法。这一部分是对前两部分的理论深入和综合应用,介绍了跨媒体检索的基本概念、算法流程、仿真实例,并结合 Web 应用环境介绍了一种基于跨媒体关联图的检索方法。

第四部分进一步扩充到了复杂数据环境下的多媒体语义挖掘。介绍了多种数据类型环境下的非线性数据分析,探讨了海量多媒体数据的存储问题。随着多媒体数据的不断扩充,海量数据的存储和复杂数据关系的挖掘问题已经日益显现。针对上述问题,第四部分做了初步的介绍和探索。

本书得到国家自然科学基金资助项目 (No.61373109),以及武汉科技大学研究生教材专项基金资助。

由于作者水平有限,时间紧迫,再加上基于人工智能的多媒体数据挖掘是当前的技术前沿,发展迅速,书中遗漏之处,敬请读者不吝指正,以便本书日后再版时予以更正。

张鸿

2017 年于武汉

目 录

第一部分 多媒体数据挖掘概述

第一章 从数据到语义的理解与挖掘.....	3
1.1 相关背景	3
1.2 信息检索的基本概念	4
1.3 纯文本信息的检索	5
1.3.1 技术背景	5
1.3.2 经典模型	6
1.4 基于关键字的多媒体检索	8
1.5 本章小结	11
第二章 图像数据挖掘概述——以 CBIR 为例	12
2.1 基本概念	12
2.2 主要技术流程和经典算法	14
2.2.1 特征建模	15
2.2.2 特征降维与索引	16
2.2.3 语义相似度度量	17
2.2.4 基于相关反馈的结果优化	20
2.3 CBIR 的相关应用	22
2.3.1 国外 CBIR 开发简介	22
2.3.2 数字图书馆中基于内容的多媒体检索	23
2.4 本章小结	27

第二部分 多媒体数据的底层特征分析技术

第三章 图像的视觉特征分析方法	31
3.1 基本概念	31
3.1.1 数字图像的组成和存储	31
3.1.2 图像的特征类型和表达方式	32
3.2 经典的视觉特征及其计算方法	33
3.2.1 颜色特征	34
3.2.2 纹理特征	39
3.2.3 形状特征	42
3.3 特征降维的关键技术	44
3.3.1 线性降维方法	44
3.3.2 非线性降维方法	50
3.4 基于深度学习的图像特征表示	51
3.5 本章小结	51
第四章 视频多模态底层特征的综合理解	53
4.1 视频时序性特征的结构化	53
4.2 视频关键帧的计算方法	56
4.3 视频镜头的自动分割	58
4.4 视频特征的融合分析和深度挖掘	61
4.5 视频数据挖掘的相关应用	62
4.6 本章小结	64
第五章 音频时序性特征的结构化方法	65
5.1 基本概念	65
5.2 听觉特征的形式化表达和计算	66
5.2.1 时域特征	67
5.2.2 频域特征	68
5.2.3 压缩域特征	69
5.2.4 特征计算的基本单位	70

5.3 音频数据挖掘的相关应用	72
5.3.1 基于内容的音频检索	72
5.3.2 音乐信号分析	73
5.3.3 音频流的自动分割	74
5.4 本章小结	76
 第六章 多模态特征的综合分析	77
6.1 底层异构特征间的内容鸿沟	78
6.2 异构特征间的统计相关性挖掘	80
6.2.1 数学模型	80
6.2.2 多模态特征综合分析实例:图像和音频特征间的典型 相关性分析	81
6.2.3 技术特点	83
6.3 实验测试和结果分析	83
6.3.1 数据收集和特征提取	84
6.3.2 实验环境和参数设置	85
6.3.3 扩展实验	85
6.4 本章小结	89
 第三部分 多模态数据挖掘实例:跨媒体检索	
 第七章 跨媒体检索基础知识	95
7.1 跨媒体的基本概念	96
7.1.1 人脑对多模态信息的认知特性	96
7.1.2 跨媒体数据挖掘的主要范畴	98
7.1.3 跨媒体检索的重要意义	99
7.2 跨媒体检索的相关研究与应用	100
7.2.1 多特征的融合分析与挖掘	100
7.2.2 跨媒体潜在关联挖掘	102
7.2.3 跨语言检索	103
7.2.4 视频中的说话人脸检测	105
7.2.5 多媒体交叉索引技术	106

7.3 本章小结	106
第八章 跨媒体检索系统的仿真实例	108
8.1 基于子空间映射的多模态数据表达	108
8.2 跨媒体距离的综合度量	110
8.3 基于相关反馈的跨媒体检索结果优化	111
8.3.1 算法描述	112
8.3.2 算法分析	114
8.4 新数据的引入	115
8.5 实验测试和结果分析	116
8.5.1 子空间维数的选取	117
8.5.2 跨媒体检索结果	118
8.5.3 相关反馈的实验结果	119
8.5.4 扩展实验	120
8.6 本章小结	123
第九章 基于隐性语义索引的跨媒体检索	124
9.1 跨媒体的特征共生矩阵	124
9.1.1 隐性语义索引	125
9.1.2 视觉和听觉特征的共生估计	126
9.2 相关性融合与优化算法	127
9.2.1 形式化描述	127
9.2.2 算法分析	128
9.3 主动学习策略	129
9.3.1 主动学习的概念	130
9.3.2 候选集计算	131
9.3.3 条件概率计算	132
9.3.4 主动学习规则	132
9.4 实验测试和结果分析	133
9.4.1 矩阵秩的选取	134
9.4.2 主动学习策略对检索性能的影响	135

9.4.3 LSI-Active 和 CCA-Passive 两种线性方法的性能对比	137
9.5 本章小结	138
第十章 Web 环境中的跨媒体数据挖掘	140
10.1 基于 Web 的信息检索	140
10.2 跨媒体关联图的建模过程	142
10.2.1 预处理过程	143
10.2.2 链接分析	143
10.2.3 图模型的定义	144
10.3 基于图模型的全局相关性推理	145
10.4 跨媒体关联图的更新策略	147
10.5 本章小结	148
第四部分 复杂数据环境下的多媒体语义挖掘	
第十一章 复杂数据关系的非线性分析	153
11.1 相关概念	153
11.1.1 复杂数据关系	153
11.1.2 非线性模型	154
11.2 流形学习模型	155
11.2.1 流形与流形学习	155
11.2.2 经典的流形学习方法	157
11.3 复杂数据关系的非线性流形建模	161
11.3.1 多特征观测空间	162
11.3.2 流形学习之:构造邻接图	163
11.3.3 流形学习之:计算测地线距离和子空间坐标	165
11.4 短期修正和长期修正策略	166
11.5 增量学习能力探讨	168
11.5.1 几何方法	168
11.5.2 交互方法	169
11.6 应用实例分析	170
11.6.1 多模态检索	171

11.6.2 图像和音频之间的跨媒体检索	172
11.6.3 新数据的引入	173
11.7 本章小结	175
第十二章 海量多媒体资源的网格化存储	177
12.1 海量多媒体资源的存储问题	178
12.2 网格相关知识	179
12.2.1 定义和特征	179
12.2.2 体系结构	180
12.2.3 网格的类型	180
12.2.4 网格资源管理	181
12.2.5 相关应用项目	182
12.3 仿真环境下的网格设计:以数字图书馆应用为例	183
12.3.1 系统构架	183
12.3.2 虚拟多媒体资源空间	186
12.3.3 检索算法设计	187
12.3.4 网格服务的发布和使用	189
12.4 本章小结	192
附录	193
参考文献	229
后记	240

第一部分

多媒体数据挖掘概述

第一章 从数据到语义的理解与挖掘

1.1 相关背景

多媒体数据包含文本、图像、视频、音频、3D 模型等不同类型的。近年来,随着人类社会的飞速发展,各种类型的信息资源急剧增长,要从茫茫的信息海洋中快速、有效地找到所需的资源,变得越来越困难。由于多媒体数据往往是无结构的、缺乏语义的,需要经过结构化处理和语义挖掘才能有效地管理多媒体大数据,从数据中提取语义,进而挖掘信息,形成有用的知识。

多媒体数据挖掘和检索在人们的日常生活中变得越来越普及和重要,并且在这些年中不断地变化和更新。该技术最早是在 1949 年由美国学者 C.Mooers 提出并使用,目前已经涉及了自然语言处理、人工智能、机器学习、多媒体检索技术等多个学科的知识。信息检索不仅是学术界一直以来的热门话题,而且已经成功地移植到许多商业应用领域中,如:Google、百度等 Web 搜索引擎。

因此,多媒体数据挖掘和检索是一种典型的需要利用人工智能方法实现从原始数据到高层语义挖掘和理解的技术领域。本书在介绍多媒体数据挖掘和应用实例的内容中,也会以不同类型多媒体数据的底层特征分析和建模,以及以高层检索应用为主题,展开深入分析。本章将对信息检索的基本概念、纯文本信息的检索和基于文本的多媒体检索进行介绍。

1.2 信息检索的基本概念

从广义上讲,信息检索是一种有目的性和组织性的信息存取活动,包括信息的“存储”和“检索”两个部分。其中,前者主要研究如何将各种异构、海量、无序的信息进行有效的组织和存储,存储的内容可以是文献的书目信息、文摘或全文等文本信息,也可以是图像、音频或视频等多媒体信息;后者则是面对用户提出的各种检索需求,快速、准确地查找到相关信息,相应地,检索方式可以是文献的作者、提名、关键词等,也可以是图像的颜色和形状、一段乐曲、一个关键帧,等等。

基于文本的信息检索在存储和检索的过程中都包括了文本信息,是信息检索领域的一个重要组成部分,主要分为纯文本信息检索和基于文本的多媒体检索两个方面,分别介绍如下:

(1) 纯文本信息检索较早被提出,从早期的结构化书目信息检索,到当前的无结构或半结构化的自由文本检索,从关键词检索,到基于概念的语义检索,一直都是较为热门研究方向。

(2) 基于文本的多媒体检索(也称为基于关键字的多媒体检索)是指:用户提交一个或多个查询关键字,就可以检索到与关键字在语义上相关的图像、音频、视频等多媒体信息。例如:提交关键字“爆炸”,可以找到“爆炸”的图像、音频和视频。从本质上讲,基于关键字的多媒体检索是对存储的多媒体信息建立文字索引,在检索的过程中,根据用户提交的关键字与文字索引之间的匹配结果,查找相应的多媒体信息,因此,也属于基于文本的信息检索范畴。

不论是何种类型的信息检索,在系统实现时主要包括四个关键步骤,即:预处理、建立索引、查询处理、搜索算法。分别介绍如下:

(1) 由于数据的种类和来源各不相同,如:结构化的书目信息、半结构化的网页数据、非结构化的多媒体数据,等等,因此,在建立数据库时需要进行预处理。预处理的主要任务是提取出结构化特征、统一编码转换等。例如:从网页数据集中提取正文、链接信息,从图像数据集中提取颜色、纹理、形状等特征,并形成结构化的视觉特征