

多媒体开采初探^{*}

陈剑贇 老松扬 吴玲达

(国防科技大学多媒体研究开发中心 长沙410073)

Multimedia Mining

CHEN Jian-Yun LAO Song-Yang WU Ling-Da

(Multimedia Technology Research Center of National University of Defence and Technology, Changsha 410073)

Abstract This paper is a primary study of multimedia mining. It first explains the basic concepts of multimedia mining and classifies multimedia mining. Then it outlines two kinds of frameworks of multimedia mining function-driven one and information-driven one. At last it introduces the process and methods of multimedia mining.

Keywords Multimedia mining, Knowledge

1 引言

对于多媒体的管理,早期采用文件的管理方式。70到80年代期间,数据库学派牵头,采用关键词的描述方法建立媒体数据的索引以达到管理的目的。90年代以后,人们转向研究面向网络环境下的支持基于内容检索的大规模的多媒体数据库。基于内容的检索在一定程度上解决了信息搜索和资源发现的问题,但是基于内容的检索只能获取用户要求的相关信息,而不能从大量多媒体数据中分析出隐含的有价值的模式和知识。另一方面,研究人员在90年代开展了大量的针对关系数据库中的结构化数据进行数据开采的研究工作。而日常生活中还包含大量的非结构化的图像、视频和音频等多媒体的数据,这使得数据开采的研究正向非结构化数据的方向发展。从用户需求的角度上讲,人们不再满足于多媒体信息存取这个层次,而需要信息的高层组织和知识发现。这些都为新的研究领域的出现提出了要求。这种新的研究领域需要采用比多媒体信息检索更高层次的新的理论和方法,从大量的多媒体数据中分析获取蕴含的但对用户来说有意义的模式和知识,这就是多媒体开采(Multimedia Mining),或者称为多媒体数据开采。

本文对多媒体开采这门新兴学科进行初探。首先简单阐述了多媒体开采的定义,澄清了一些概念理解上的误区;并且在概念的基础上从开采的媒体和开采的任务两种角度对多媒体开采进行了分类。然后从功能和信息两个角度描述了目前多媒体开采系统所采用的两种框架结构,介绍多媒体开采的一般过程和常用方法。最后是总结。

2 多媒体开采的定义以及分类

2.1 多媒体开采的基本概念

所谓多媒体开采,就是从大量多媒体数据集中,通过综合分析视听特征和语义,发现隐含的、有效的、有价值的、可以理解的模式(知识),得出事件的关联和趋向,为用户提供问题求解层次的决策支持能力。

多媒体开采是一门多学科交叉的新兴学科,包括计算机

视觉、图像处理、模式识别、人工智能等等。但是多媒体开采与这些相关学科的侧重点不同。计算机视觉和图像处理的焦点是从单个图像中分析和提取特定的特征,而多媒体开采的焦点是从大量的多媒体数据集中抽取一般的模式。对于多媒体开采和基于内容的信息检索,看起来两者的研究有重叠的部分(例如都是处理大量的媒体数据),但是多媒体开采的研究范围远远超出了基于内容信息检索的研究范围,多媒体开采的目标是从大量媒体数据中发现有意义的一般模式或者知识。

从多媒体开采的产生历史来看,另一个理解的误区是多媒体开采是现有的数据开采的算法对于多媒体数据的一个应用和扩展,而事实绝非如此。其中一个重要原因就是传统的数据开采处理的是关系数据库的结构化数据,而多媒体开采处理对象是非结构化的多媒体数据,两者存在有很大的差别。多媒体数据的结构非常复杂,多媒体对象的描述也不同于普通的数据对象,多媒体对象的特征描述是多维的。多媒体开采绝对不仅仅是数据开采技术在多媒体中的一个扩展:如何从多媒体的内容着手,利用多媒体的时间、空间、视觉特性、视听对象、运动特征,适当地结合数据开采的基本理论和方法,开采出有价值的隐含的模式和知识,是多媒体开采面临的重要挑战。多媒体开采也必然产生新的相关研究方向,例如:

·特征的选择:多媒体特征是多维的,如何选择多媒体数据的特征或者如何融合多种特征以利于多媒体开采将是多媒体开采研究的重要内容。

·模式的表示:如何在一种丰富的视觉环境下表示多媒体数据的模式将是多媒体开采的一项研究内容。

2.2 多媒体开采的分类

多媒体结构复杂,内容丰富;数据开采的方法也很丰富,那么多媒体开采如何分类呢?以下从开采的媒体和开采的任务两种角度对多媒体开采进行了分类。

2.2.1 从处理的媒体对多媒体开采进行分类 根据所处理的媒体类型,多媒体开采分为:

·Web 开采:Web 是一个庞大的、动态的、超媒体结构的包含媒体和数据元素的信息库。对 Web 进行有效的资源和知

^{*}基金项目:国防预研项目资助。陈剑贇 博士研究生,目前主要研究方向为基于内容检索。老松扬 博士,副教授,主要从事多媒体数据库的研究。吴玲达 教授,博士生导师,主要从事多媒体信息系统和虚拟现实技术的研究。

识发现具有极大的挑战性,因为 Web 是一个动态性极强的信息源,Web 网页的复杂性高于任何传统的文本文档,而 Web 面向的又是一个广泛的形形色色的用户群体,用户往往只关心 Web 中很少的一部分信息。Web 开采就是要从 Web 文档和 Web 活动中抽取感兴趣的潜在的有用模式和隐藏的信息。Web 开采可以分为三个方面:Web 内容开采(Web Content Mining)、Web 结构开采(Web Structure Mining)和 Web 使用开采(Web Usage Mining)。Web 内容的开采是从文档内容或者其描述中抽取知识的过程,目前主要研究从 HTML 文档内容进行开采。Web 结构的开采是对 WWW 的组织结构和链接关系中推导知识,利用文档之间的互联关系对页面进行排序,找出重要的网页。Web 使用开采是追随用户对 Web 的交互和操纵,从中开采出有价值的存取模式;目前主要研究通过 Web 日志记录来发现用户访问 Web 页面的模式,例如通过分析和探究 Web 日志记录的规律识别电子商务的潜在用户,改进 Web 服务器的性能和结构等。

•图像开采:图像含有丰富的视觉和空间特性。图像开采的目的就是从图像中自动抽取有意义的语义信息(知识)。目前图像开采的根本问题在于底层特征如何关联转换为高层对象以及语义概念^[1]。

•视频开采:视频不仅包含有图像的视觉和空间特性,还具有时间特性、视频对象特性、运动特性等丰富的内容线索。视频开采技术可以广泛应用到新闻视频、监控视频、记录影片和数字图书馆等应用系统中,例如从交通监视视频中分析出交通堵塞的趋势,从新闻中分析出不同事件之间的因果关系等。

•音频开采:音频开采应用非常广泛,目前主要研究的是语音和音乐的开采。重点是针对听觉特性中的基音、音调、旋律、音频事件和对象的结构,开采出隐含在音频流中的信息线索、规律和特性。

以上提到的是单一媒体的开采,多媒体是一个集成的系统概念,如何在多媒体开采的各个阶段,综合利用多种媒体的特性进行知识发现也将是研究的重点。

2.2.2 从开采的任务对多媒体开采进行分类 多媒体开采有不同层次和类型的任务,按照这种不同,可以把多媒体开采分为关联规则的开采、聚类的开采、分类的开采、偏差分析、趋势分析等。

•关联规则的开采:如果多个数据项或者特征的取值之间重复关联出现并且频率很高,就意味着它们之间存在着某种联系,这就是关联规则。例如文[2]对国旗的颜色特征和地理分布进行了分析,作者发现一般带有黄色和绿色的国旗的国家位于南美洲,这就是多媒体数据中颜色特征与地理位置的关联。作者通过关联分析的方法,计算出支持度和可信度以筛选出这条有价值的关联规则。

•聚类的开采:聚类是指在没有任何辅助和监督的情况下,将数据库中的数据划分为一系列有意义的子集,即类别。在同一类别中,个体之间的距离较小,而不同类别上的个体之间的距离较大。聚类是一种自动(unsupervised)的分类。文[3]是一篇开采多媒体数据空间关系的文章,作者采用聚类的方法,将空间上零散杂乱的点按照一定的衡量原则划分为有意义的子集,以辅助用户的决策。

•分类的开采:分类在多媒体开采中用得很多。分类是找到一个类别的概念描述,它代表了这类数据的整体信息,即该类的内涵描述。分类的模式能把数据集中的元素映射到给定

类别中的某一类中去。与聚类不同,分类的开采是有一定辅助(supervised)的过程。在文[4]中,分类开采应用到乳腺 X 照片图像(mammography)的自动判定中。作者首先对图像进行预处理,去噪平滑,用直方图均衡化加强图像的对比度;然后结合神经网络的方法将图像分为正常和非正常两类,非正常类还细分为六类;当一旦有新的图片,即可把新图片归到正常或者非正常类中,从而自动判断是否病变。

•偏差分析的开采:所谓偏差分析就是通过数据分析的方法发现数据的异常情况,例如:分类中的反常实例、模式的例外等。偏差分析常用的方法是寻找观察结果和参照之间的差别,参照一般是给定模型的预测或者外界提供的标准,而观察结果一般是一个特征的取值或者融合多个特征的汇总。

•趋势分析:趋势分析属于高层的多媒体数据开采的范畴,它是利用大量的数据找出变化规律,建立模型,并用此模型来预测未来的发展趋势。在文[5]中,作者通过分析图像序列中对象的时空关系来挖掘交通堵塞的趋势。文章首先进行对象的分割和识别,例如小汽车或者大型的公共汽车对象,然后进行对象跟踪,接着分析对象之间的时空关系,最终开采出知识,例如堵塞或者有汽车等待着进入交叉路口等。

3 多媒体开采系统的框架

一个好的多媒体开采系统的框架将有利于多媒体数据的开采。好的多媒体数据开采系统不仅要提供有效的数据存储、处理和检索,还要利于模式和知识的产生和发现。目前主要有功能驱动和信息驱动两种多媒体数据开采系统的框架。前者强调组织不同的功能模块以实现多媒体数据开采,后者为一个包含不同信息层次需求的结构。以下简单介绍这两种系统框架。

3.1 功能驱动的多媒体开采系统的框架

目前大部分多媒体开采系统多为这种框架,这种框架是由功能模块来组织的。文[6]开发的是一个基于卫星图像的开采系统,其体系框架如图1所示。从图中可见,这个开采系统包含有多个功能模块,概括来讲主要分为两类:一是卫星图像的分析处理和检索模块,负责卫星图像数据的获取、预处理、存档以及图像特征的提取和检索;二是图像的开采模块,它是建立在图像分析处理和检索模块的基础上的,它能使用户发现图像的意义,探测相关的事件,得到对用户有意义的知识。

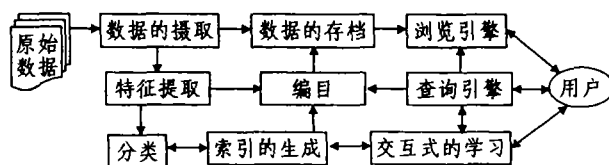


图1 功能驱动的多媒体开采系统的框架

在文[7]中,加拿大 Simon Fraser 大学的 Osmar R. Z 博士开发的多媒体开采的原型系统也是功能驱动的框架结构。

功能驱动的框架区分了多媒体开采系统中不同的任务和角色,但它没有考虑信息的不同层次的需求和表示,因而产生了信息驱动的多媒体开采系统的框架,下面简单介绍这种框架。

3.2 信息驱动的多媒体开采系统的框架

在文[8]中,作者提出了一种信息驱动的图像开采系统的框架,如图2所示。这个框架着重于不同层次的信息的表示和处理,文中区别了四层信息:最底层的是像素级的,包括源图

像信息和底层的物理特性(例如颜色、纹理、形状等);第二层是对像级的,主要分析处理基于底层特性的对象或者区域;第三层是语义级的,主要考虑从被鉴别的对象或者区域中得到的高层语义概念或者领域知识;最高层是模式和知识级的,主要指把语义概念和相应的文字结合起来获得隐含的模式和知

识。概括来讲,以上四层可以分为两个大层:一是较低层次的像素级和对象级,主要包括原始图像和抽取的图像信息,相关领域包括图像分析、处理和识别。二是较高层次的语义概念级和模式知识级,主要是指高层的图像操作,例如知识发现等。较高层次比较低层次含有更多的语义信息。

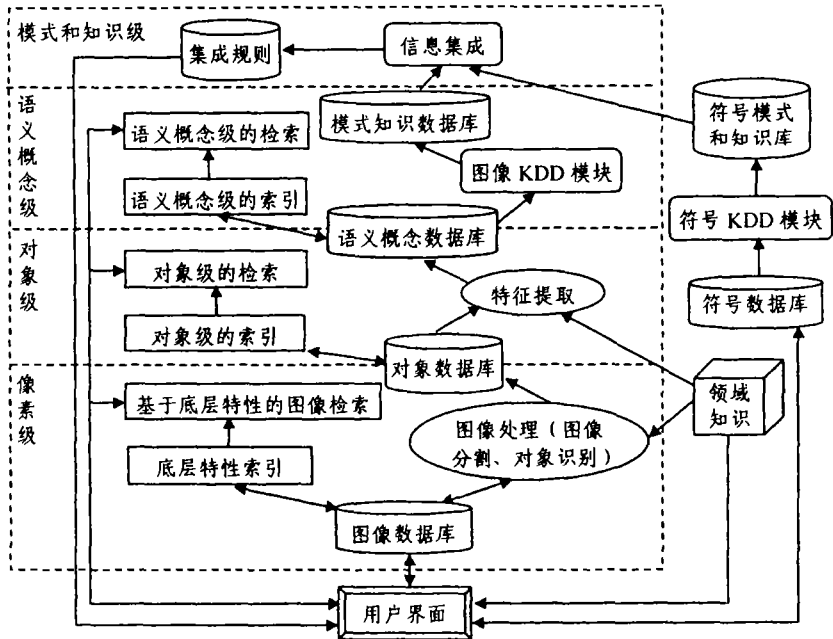


图2 信息驱动的多媒体开采系统的框架

以上谈到的是图像开采系统的信息层次结构。类似地,视频也可以考虑低、中、高三层开采结构。视频是非结构化的数据,从信息驱动的角度,可以把视频分为低、中、高三层,相应的多媒体开采可以在这三层中展开。低层处理的是原始的视音频信号,在分析过程中利用物理特征的客观知识,主要采用信号处理技术,得到的结果是视频的基本语义单元(如镜头);而中层处理的对象是附带有低层特征的视音频信息,结果具有更高的颗粒度和抽象度,例如故事单元等;高层一般与相关领域关联,适合于特定应用,结果往往是符合特定要求的内容单元和语义抽象。

4 多媒体开采的过程和方法

前面讨论的是多媒体开采系统的一般框架,以下简要介绍一下多媒体开采的具体过程和常用的方法。

4.1 多媒体开采的过程

与传统的数据开采的过程基本相似,多媒体开采的一般过程如图3所示。

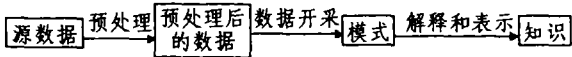


图3 多媒体数据开采的基本过程

数据的预处理就是从大量的甚至异构的媒体数据中选取与任务有关的媒体数据。多媒体数据的预处理应该从多媒体内容着手,获取多媒体对象的时空、运动和视听特征,使得处理后的数据适合于接下来的开采过程。

开采是多媒体开采整个过程中的核心部分。一方面,现有的数据开采的技术和方法可以应用于多媒体开采中;另一方面因为多媒体数据库不同于传统的关系数据库,从多媒体时

间、空间、运动特征、视听特性等角度,必然需要新的多媒体开采的方法来实现多媒体开采,例如,如何开采出视频和音频的内容结构(包括叙事结构),如何从多媒体语义中开采出事件和事实的空间分布,如何根据视频的运动特性分析出事件的联系和趋向等。

模式可以理解知识的雏形,所以需要通过解释和表示以形成知识。多媒体强调视听效果,如何根据用户的要求,以空间分布、运动趋向等可视化的方式向用户呈现最有价值的开采结果是多媒体开采研究的重点内容。

4.2 多媒体开采的方法

那么,在多媒体开采过程中有哪些常用的方法呢?数据开采的方法由人工智能、机器学习发展而来,并且结合了传统的统计分析、模糊数学等方法,目前常用的多媒体开采方法有:

1) 统计分析的方法

·相关分析:即通过大量的数据度量变量间的相关程度,例如关联规则的开采就可以采用这种方法分析大量的历史数据,找出数据项之间的联系。

·回归分析:求回归方程来表示变量间的数量关系。

·聚类分析:聚类是统计分析中的一种重要的方法。如前所述,聚类是将数据库中的数据划分为一系列有意义的子集(即类),在同一类别中,个体之间的距离较小,而不同类别的个体之间的距离较大。聚类增强了人们对于客观现实的认识,通过聚类可以建立事物的宏观概念。

2) 仿生物技术的方法

·神经网络:从结构上模仿生物神经网络,是一种通过训练来学习的非线性预测模型,是数据开采的一种重要的方法。

·遗传算法:这是一种模拟生物进化过程的算法,由繁殖(选择)、交叉(重组)和变异(突变)三个基本算子组成。遗传算

(下转第79页)

集合的相似度量对其应用有着重要的作用^[2-4],对 Vague 集这里我们给出一种新的相似度量方法。

设 X 中有元素 $x_i (i=1, 2, \dots)$, 对 Vague 集合记:

$$\Sigma_{count}(A) = \Sigma_i (t_A(x_i) + f_A(x_i))$$

对 Vague 集中任意两集合 A, B 定义其相似度量 $S(A, B)$ 为:

$$S(A, B) =$$

$$1 - \frac{\Sigma_i \max(0, t_A(x_i) - t_B(x_i)) + \max(0, t_B(x_i) - t_A(x_i))}{\Sigma_{count}(A \cup B)}$$

$$- \frac{\Sigma_i \max(0, f_A(x_i) - f_B(x_i)) + \max(0, f_B(x_i) - f_A(x_i))}{\Sigma_{count}(A \cup B)}$$

$$= \frac{\Sigma_{count}(A \cap B)}{\Sigma_{count}(A \cup B)}$$

对于相似度量 S 可以验证具有以下性质:

性质1 $0 \leq S(A, B) \leq 1$, 当且仅当 $A=B$ 时, $S(A, B) = 1$ 。

性质2 $S(A, B) = S(B, A)$ 。

性质3 $S(A \cup B, A \cap B) = S(B, A)$ 。

性质4 $S(A \cup B, C) \leq S(A, C) + S(B, C)$ 。

结论 通过对 Vague 集的探讨本文给出了 Vague 集中的分解定理以及相似性度量公式。近来 Vague 集作为一种新的工具被广泛研究和应用,但处于起步阶段。本文的结果是为发展其理论所做的探讨。

参考文献

- 1 Gau W L, Buehrer D J. Vague sets. IEEE Trans. Systems Man Cybernet, 1993, 23(2): 610~614
- 2 Chen S M. Measures of Similarity Between Vague Sets. Fuzzy Sets and Systems, 1995, 74: 217~223
- 3 Hong D H, Kim C. A Note on Similarity Measures Between Vague Sets and Between Elements. Information Sciences, 1999, 115(1): 83~96
- 4 闫德勤. 新的模糊度量与模糊似然函数. 模式识别与人工智能, 2001, 14(1): 23~26
- 5 郭宗祥, 杨鸿铨. 模糊信息处理基础. 成都: 成都电讯工程学院出版社, 1989. 22~29

(上接第59页)

法是一种新的优化技术,它基于生物进化的概念设计这一系列的过程来达到优化的目的。在多媒体开采中,为了应用遗传算法,需要把多媒体开采任务表达为一种搜索问题而发挥遗传算法的优化搜索能力。

3) 归纳学习的方法: 归纳学习的方法是传统的数据开采方法的重要研究方向。主要包括决策树、粗集(Rough Set)等方法。

4) 模糊数学的方法: 利用模糊集合理论对实际问题进行模糊评判、模糊决策、模糊模式识别等处理。

5) 基于本体论(Ontology)和多主体的方法: 在文[9]中,作者提出了一种利用智能主体和本体论理论设计的多媒体信息检索和开采系统,其利用 Ontology 对多媒体数据进行领域分类,同时对用户的查询信息进行规范,较好地完成多媒体开采和检索的任务。

除了以上谈到的方法外,还有象 Bayesian 分类和 Bayesian 学习算法等方法。

以上谈到的这些开采方法基本上是传统的数据开采的方法在多媒体开采中的应用和扩展。如果从多媒体时间、空间、运动特征、视听特性等角度考虑,必然有新的多媒体开采的方法的出现。例如: Web 结构开采是分析页面的结构和 Web 的结构,在此基础上对页面进行分类和聚类从而找到权威的页面。Brin 等人从 Web 的结构出发就提出了不同于传统数据开采方法的 Page-rank 方法^[10]。

总结 本文对新一代多媒体内容管理中的关键技术——多媒体开采的概念和分类、一般的系统框架、开采的过程和方法作了初步的探讨。多媒体开采是多媒体内容分析、多媒体内容管理、数据仓库和数据开采等多领域的一个交叉学科,也是一门新兴的学科。关于多媒体开采的方法、综合多种媒体的多媒体开采、多媒体开采的交互和表现等诸多问题都需要深入的研究。

参考文献

- 1 Zhang Ji, Hsu W, Lee M L. IMAGE MINING: ISSUES, FRAMEWORKS AND TECHNIQUES. In: the 2nd Intl. Workshop on MDM, San Francisco, California, USA, Aug. 2001
- 2 Uehara Y, et al. A Computer-aided Visual Exploration System for Knowledge Discovery from Images. In: the 2nd Intl. Workshop on MDM, San Francisco, California, USA, Aug. 2001
- 3 Tung A K H, Ng R T, Lakshmanan L V S, Han Jiawei. Geo-spatial Clustering with User-Specified Constraints. In: the 1st Intl. Workshop on MDM, Boston MA, USA, July 2000
- 4 Antonie M L, Zaiane O R. Application of Data Mining Techniques for Medical Image Classification. In: the 2nd Intl. Workshop on MDM, San Francisco, California, USA, Aug. 2001
- 5 Chen Shu-Ching, Shyu Mei-Ling, Zhang Chengcui, Strickrott J. Multimedia Data Mining for Traffic Video Sequences. In: the 2nd Intl. Workshop on Multimedia Data Mining. San Francisco, California, USA, Aug. 2001
- 6 Datcu M, Seidel K. Image information mining: exploration of image content in large archives. IEEE Conference on Aerospace, Vol. 3, 2000
- 7 Osmar R Z, Han Jiawei, et al. MultiMediaMiner: A System Prototype for MultiMedia Data Mining, 1999
- 8 Zhang J, Hsu W, Lee M L. An Information-driven Framework for Image Mining. In: Proc of 12th Intl. Conf. on Database and Expert Systems Applications, Munich, Germany, Sep. 2001
- 9 Simoff S J, Maher M L. Ontology-based multimedia data mining for design information retrieval
- 10 Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. In: Proc of 7th World Wide Web Conf. Brisbane, Austria, 1998
- 11 熊华. 视频结构化技术的研究与实现. [国防科技大学博士论文]. 2001