

新闻视频挖掘技术研究^{*}

栾悉道 谢毓湘 韩智广 吴玲达

(国防科技大学多媒体研究开发中心 长沙 410073)

摘要 新闻视频挖掘是一个新兴的研究领域,也是多媒体数据挖掘的典型代表。本文对新闻视频挖掘技术进行了全面深入的讨论,首先从概念上对新闻视频挖掘进行了界定,提出了新闻视频挖掘的层次框架和技术框架,指出新闻视频挖掘包括低层视频挖掘和高层视频挖掘两个层次。其中,低层视频挖掘是利用数据挖掘的方法对视频内容进行分析的过程,而高层数据挖掘则是在低层挖掘的基础上进一步发现视频中的知识的过程。新闻视频挖掘的技术框架则对挖掘所涉及到的具体技术进行了分析。最后,对新闻视频挖掘中的结构挖掘、语义内容挖掘、视频摘要、趋势挖掘、关联挖掘等任务进行了详细的阐述,并对各种任务举出了具体的示例加以说明。

关键词 新闻视频挖掘,视频内容分析,视频摘要,结构挖掘,语义内容挖掘

Research on News Video Mining Techniques

LUAN Xi-Dao XIE Yu-Xiang HAN Zhi-Guang WU Ling-Da

(Multimedia R&D Center, National University of Defense Technology, Changsha 410073)

Abstract News video mining is a rising field that is the sub field of multimedia data mining. A survey of news video mining techniques is discussed. First, the definition of news video mining and the relationships of news video mining and other fields are analyzed. Then the layered and technical frameworks of news video mining are designed. In the layered framework, news video mining is divided into two layers. One is the low-level video mining, which is the process of analyzing video content by the methods of data mining. The other is the high-level video mining, which is the process of discovering knowledge on the basis of low-level video mining. The tasks of news video mining that include structure mining, semantic content mining, video summarization, tendency mining, association rule mining, classification, clustering, etc. are also described in detail.

Keywords News video mining, Video content analysis, Video summarization, Structure mining, Semantic content mining

1 引言

人们每天都会接触到大量的新闻视频,这些新闻视频能够为我们提供极为丰富的信息。由于新闻视频大多具有及时、准确的特点,因此时常被人们当作一个有用的工具来指导决策。举个简单的例子,许多股民就是通过收看电视新闻节目来了解股市的规律,从而决定手中股票的抛售与否。一些重大的新闻事件往往也会对人们的日常生活产生极大的影响,如已经过去的伊拉克战争对世界石油价格产生了极大的冲击;前段时间爆发的 SARS 疫情,对世界特别是中国的旅游业以及其它行业产生了消极的影响,甚至影响到了人们日常的生活起居以及心态。所有这些都暗示着我们新闻视频不仅仅只是一个简单的一次性消费品,而是一个极富价值的资料库。通过对大量的新闻视频节目进行分析和挖掘,完全可以从中发现有价值的信息。

新闻视频挖掘是一个新兴的研究领域,它隶属于多媒体数据挖掘的范畴。多媒体数据挖掘^[1]是近年来国外研究的一个新热点,它从多媒体数据库、数据挖掘、信息系统等领域发展而来,是一个交叉性的学科。Osmar 是较早从事多媒体数据挖掘的研究者之一。他设计的 MultimediaMiner^[2]原型系统概括性地说明了图像与视频数据的挖掘过程,但在具体实现上他更侧重于图像对象空间关系的关联挖掘,而没有涉及到视频挖掘的内容。George Mason 大学的 Duminda^[3]设计

了电影挖掘的系统,综合了视音频处理技术,目的是要挖掘出视频的分层结构、感兴趣的事件、电影和元数据的关联以及进行电影分类。美国德州大学计算机科学与工程系的 JungHwan^[4]提出了一个视频挖掘的系统框架,主要是针对没有编辑的实时视频流,如交通视频、监控视频等进行挖掘。Zhu^[5]提出了一个挖掘医学视频结构的工具 ClassMiner,目的在于挖掘出视频的分层结构。另外他还利用视音频分析技术来发现节目中的特殊事件,如医生给病人诊断的片断、临床操作的片断等。Kim^[6]提出了一种综合视频、音频分析技术和领域知识来分析新闻视频的方法,以发现新闻节目的叙事结构,并将得出的结构信息用于新闻视频的摘要和浏览。

国内有关多媒体数据挖掘的研究还很初步。比较有代表性的研究包括:武汉大学曹加恒^[7]提出的多媒体数据挖掘原型系统 MDMP。该系统以多媒体数据库为平台,根据用户请求,利用基于内容检索和相关数据收集,建立媒体数据特征立方体,挖掘出隐含规则,并通过图形化用户界面向用户解释获取的知识。国防科大李国辉^[9]教授对多媒体数据挖掘框架及技术进行了探索性的研究。

从上述研究来看,多媒体数据挖掘已成为了近年来国内外研究的热点之一。但还存在以下一些问题需要解决:一是多媒体数据挖掘系统体系框架的设计问题。它需要反映出多媒体数据自身的一些特征。二是概念界定问题。现有的研究很少对多媒体数据挖掘、多媒体信息检索以及数据挖掘等概

^{*} 本文得到国家自然科学基金项目资助(60473117)。栾悉道 博士;谢毓湘 博士,讲师;主要研究方向为多媒体信息系统,视频分析与挖掘;吴玲达 教授,博导,主要研究领域为多媒体技术。

念进行界定,导致许多错误的观点认为多媒体数据挖掘只是传统的数据挖掘在多媒体方面的扩展,或者认为多媒体数据挖掘只是多媒体信息检索在概念上的一个提升。如何理清这些概念之间的关系是需要深入考虑的一个问题。第三,多媒体数据挖掘技术的应用问题。不少研究都已经给出了多媒体数据挖掘的原型系统框架,但真正实用的系统还有待进一步的开发。

考虑到新闻视频作为公开情报的重要来源之一以及它在辅助情报分析方面所特有的优势,我们将选择新闻视频挖掘作为本文研究的重点。本文的组织如下:第2节将对新闻视频挖掘的概念进行界定,第3节提出新闻视频挖掘的框架,第4节对新闻视频挖掘的任务加以阐述,最后是总结与展望。

2 新闻视频挖掘的概念

新闻视频挖掘是多媒体数据挖掘中有代表性的一个研究方向。多媒体数据挖掘就是从大量多媒体数据集中,通过综合分析视听特性和语义,发现隐含的、有效的、有价值的、可理解的模式,进而发现知识,得出事件的趋向和关联,为用户提供问题求解层次的决策支持能力^[9]。多媒体数据挖掘是数据挖掘、多媒体数据库、基于内容检索等技术发展的共同产物,与计算机视觉、模式识别等学科也有着千丝万缕的联系。新闻视频挖掘是多媒体数据挖掘的典型代表,其定义与多媒体数据挖掘类似,所不同的是它所针对的数据集是新闻视频数据集。

在数据库界,普遍存在这样一种错误的观点,即认为多媒体数据挖掘是数据挖掘的一个子集,是数据挖掘技术在多媒体领域的应用。从多媒体数据挖掘的发展进程来看,它的研究最早开始于20世纪90年代末,比数据挖掘技术的研究晚了整整十年。而且较早从事多媒体数据挖掘研究的人员大多是从数据挖掘的研究中分离出来的。从这种角度来说,数据挖掘技术是多媒体数据挖掘技术产生的主要源泉。但这并不能说明多媒体数据挖掘技术就是数据挖掘技术的子集。事实上,多媒体数据有着传统数据所不具备的诸多特点,它的大容量、多维、非结构化的数据特征使得传统的数据挖掘方法在处理多媒体数据时显得无能为力。因此,迫切需要研究新的适合多媒体数据挖掘的框架和方法,以解决大容量多媒体数据集中的知识发现问题。

与数据库界相对应,在信息系统界,普遍存在着另外一种错误的观点,即认为多媒体数据挖掘其实质就是多媒体信息检索,它本身并没有任何新意,仅是在概念上对多媒体信息检索加以提升而已。所谓多媒体信息检索,就是通过对多媒体数据进行分析,从中提取出特定的信息线索,然后根据这些线索在多媒体数据库中进行查找,检索出具有相似特征的媒体数据^[10]。我们承认多媒体数据挖掘需要用到多媒体信息检索的一些技术来解决多媒体数据的特征提取等问题,但这些方法并不能直接帮助我们解决知识发现问题。没有新的多媒体数据挖掘技术做支撑,多媒体信息检索将永远停留在信息的浏览与检索层次,而不是知识层次。应该说,多媒体信息检索技术为多媒体数据挖掘奠定了一个良好的数据分析基础,是多媒体数据挖掘的重要组成部分。反过来,多媒体数据挖掘能够在很大程度上改善多媒体信息检索。

多媒体数据挖掘与模式识别之间的区别在于:首先,模式识别关注具体的模式,其目的在于识别出已知的特定目标。而多媒体数据挖掘在没有先验条件的情况下去发现多媒体集合中所有可能的有意义的模式。另外,挖掘出来的模式和知识可以用于模式识别。其次,模式识别只关心模式的产生和

对模式的分析,而多媒体数据挖掘关注多媒体数据库的各个方面的,包括多媒体数据的存储、索引的建立以及多媒体数据的检索、表示等。第三,模式识别中的模式主要是指分类模式,而多媒体数据挖掘的模式可以是分类、描述、空间、关联等多种模式。

与计算机视觉之间的区别在于:计算机视觉关注的是从单张的图像中抽取或理解特定的特征,而多媒体数据挖掘是从大量的多媒体数据集中抽取模式。

与专家系统的区别在于:专家系统是从领域专家或学习案例里获取知识,形成规则和模型,是一个问题求解系统。而多媒体数据挖掘是从大量的多媒体数据中抽取知识。

与机器学习的区别在于:机器学习是将一些已知的并被成功解决的问题当作范例输入计算机,机器通过学习这些范例总结并生成相应的规则。它输入的数据集一般比多媒体数据挖掘的数据集小,关注于小样本的学习模式。而多媒体数据挖掘面对的是现实生活中大量的、不完全的、有噪声的、模糊的、随机的大数据样本。

综上所述,多媒体数据挖掘不是数据挖掘在多媒体领域的扩展,也不是简单的多媒体信息检索在概念上的提升,而是一个新的学科方向,它建立在数据挖掘、多媒体数据库、多媒体信息检索、计算机视觉、模式识别、人工智能、机器学习等多个学科的基础之上,具有广阔的发展前景。

3 新闻视频挖掘的框架

如前所述,已经有大量的研究致力于多媒体数据挖掘框架的设计^[2~4,7]。人们在从事多媒体数据挖掘研究的过程中发现,传统的数据挖掘方法并不能直接作用于多媒体数据。主要存在以下一些困难:

一是多媒体语义信息的提取问题。多媒体数据含有丰富的视频、音频和图像信息,从这些信息中提取出有价值的语义信息一直困扰着人们对多媒体数据的充分利用,而这个问题在传统的数据挖掘中并不突出。

二是对语义信息的理解问题。即使可以抽取多媒体数据中隐含的人脸、字幕、运动、对象等类型的语义信息,将这些信息转换成人们所能理解的形式依然存在极大的困难。导致这一困难的原因很多:首先,从视音频数据中识别语义信息本身就是一个非常复杂的任务;其次,高层的语义信息往往并不单纯地存在于某种媒体中,而是隐含在多种类型的媒体中。因此,需要采用信息融合的方法来解决语义信息的抽取问题。第三,现有的数据挖掘的概念和技术是否可以应用于复杂的、融合的多媒体数据是需要解决的问题。只有当前两个问题得以有效解决时,这个问题才会有比较完美的一个答案。多媒体信息检索技术与数据挖掘技术的有机结合,让我们看到了解决上述问题的曙光。

本节将以新闻视频为例,来说明多媒体数据挖掘的框架。

3.1 新闻视频挖掘的层次框架

新闻视频挖掘的层次框架如图1所示。图1反映的是一个完整的从新闻视频数据中发现知识的过程,包括数据准备、低层视频挖掘、高层视频挖掘以及解释评估等阶段。图1说明了以下几个问题:

(1)与传统的知识发现过程相比,图1在视频挖掘过程中引入了层次的概念,即认为视频挖掘可分为低层视频挖掘和高层视频挖掘两层。低层的视频挖掘包括数据的多维分析、结构挖掘和内容挖掘,它主要对应着多媒体信息检索领域所涉及的一些技术,如视频内容结构化分析技术。但是这一阶段的视频内容结构化分析已经发生了质的变化,它在分析的

过程中大量运用了传统的数据挖掘技术,如分类、聚类等。因此,我们称这一阶段的挖掘为低层的视频挖掘。低层视频挖掘的主要目标是完成视频句法分段和语义标注的任务,为高层的视频挖掘提供特征数据和元数据。我们认为低层的视频挖掘是视频挖掘区别于传统数据挖掘的重要特征之一。

是一个非常复杂的任务。在实际的挖掘过程中,往往根据具体的应用需求,选择相应的特征维数据进行分析,从而大大降低数据挖掘的复杂度。

(5) 视频的结构挖掘是指对视频的分层结构、叙事结构等结构特征进行挖掘,对应着传统视频内容结构化分析技术中的句法分段;视频的内容挖掘是指对视频中的字幕、人脸、运动、对象、暴力、感人等内容进行挖掘,对应着视频内容结构化分析技术中的语义标注。两者之间是一个互补的关系,视频的内容挖掘可以辅助视频的结构挖掘,例如新闻视频中口播帧的探测可以辅助新闻视频故事单元的切分;视频的结构挖掘则为视频的内容挖掘奠定了一个良好的基础。

(6) 视频摘要视频高层挖掘的主要方法之一,也是视频挖掘区别于传统数据挖掘的又一重要特征。视频摘要是视频内容的浓缩和精华,它建立在低层视频结构和内容挖掘的基础之上。其它的高层视频挖掘技术包括聚类、分类、比较、统计等等。

(7) 挖掘出来的模式并非都是知识,只有经过解释和评估了的模式才能成为知识。在模式的解释和评估阶段,需要引入人和知识库的判断。有趣的模式将为决策提供有力的支持。

(8) 可视化技术贯穿了整个知识发现的全过程。在视频数据准备阶段,用户可能需要使用直方图等统计可视化技术来显示有关数据,以期对数据有一个初步的了解;在视频挖掘阶段,需要使用与具体领域相关的可视化工具;特别是在挖掘结果的表达与解释评估阶段,更离不开可视化技术的帮助。

(9) 视频的知识发现过程是一个循环的过程。在这个过程的每一个阶段,如果发现某个阶段产生的结果与预想有出入,则需要用户重复前面阶段的某个过程。

综上所述,新闻视频挖掘是一个分层的、循环的过程。低层的视频挖掘为高层视频挖掘提供基础和保障,高层视频挖掘为决策提供支持。挖掘出来的知识反过来又可以用于视频挖掘的各个层次。它们之间互为促进,互为补充。

3.2 新闻视频挖掘的技术框架

新闻视频挖掘的层次框架将新闻视频挖掘分为了低层新闻视频挖掘和高层新闻视频挖掘两个层次,本节将在此基础上对新闻视频挖掘的技术框架进行设计,以进一步细化各个层次的挖掘内容。如图2所示。

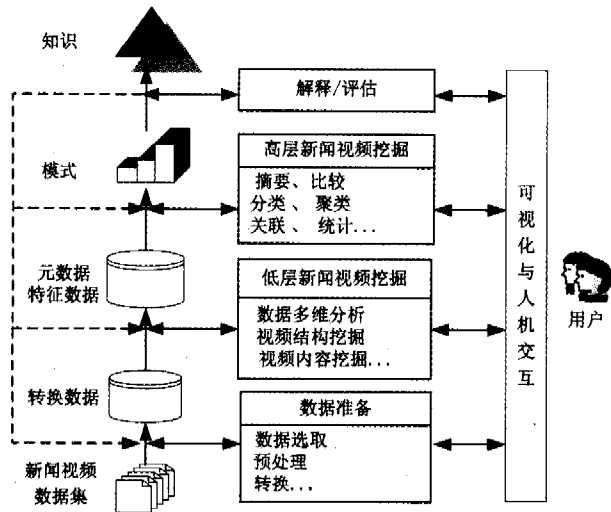


图1 新闻视频挖掘的层次框架

(2) 高层的视频挖掘包括摘要、分类、聚类、比较、关联、统计等任务。这一阶段的任务似乎更接近于传统的数据挖掘,但也蕴含视频数据挖掘自身的一些特征。比如摘要,就是高层的视频挖掘所特有的任务之一。

(3) 从挖掘的数据源来看,视频挖掘的数据源是视频数据集,这个集合可大可小,既可以是单个的视频文件,也可以是视频文件的集合,或者视频数据库以及视频数据仓库。这是由视频数据自身的特点所决定的。由于视频数据的数据量非常庞大,即使是单个的视频文件,其数据量也可以达到上兆字节。从数据量的角度而言,不论是单个的视频文件,还是庞大的视频数据仓库,都可以成为视频挖掘的数据源。

(4) 视频数据不仅数据量大,而且内容丰富,具有多维的特征。其多维特征包括:视频数据的大小、视频数据的拍摄或编辑日期、视频主题等描述性特征;也包括视频帧的颜色、纹理、形状等物理特征;视频的故事、场景、镜头、帧等结构特征,等等。要对所有的视频特征进行多维分析并建立数据立方体

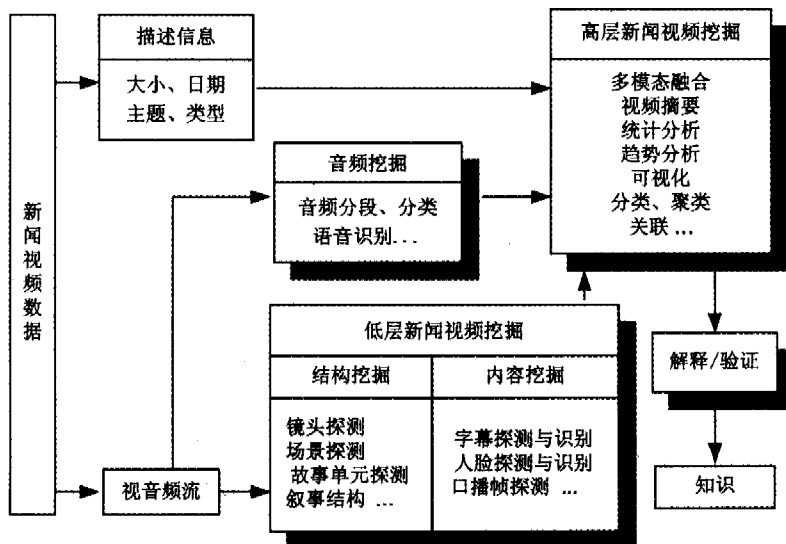


图2 新闻视频挖掘的技术框架

从图2我们可以看出:

(1)新闻视频的挖掘需要用到描述信息、视频信息、音频信息等多元信息,因此,整个挖掘是一个多源、多层次的过程。高层新闻视频挖掘阶段的多模态融合技术是新闻视频挖掘技术中的一个亮点,它能够有效地解决多源信息的融合问题。

(2)新闻视频挖掘中的音频挖掘技术主要是指音频内容的分段与分类技术,用于识别音频流中的语音、音乐、静音、环境音等声音。语音识别技术主要是在音频分段与分类的基础上,对语音进行识别,将语音转换成文字,用以辅助高层的新闻视频挖掘。

(3)低层的新闻视频挖掘主要是指充分利用多媒体内容分析技术以及传统的数据挖掘技术(如聚类、分类等)来进行新闻视频的结构和内容挖掘。其中,新闻视频的结构挖掘技术包括镜头探测、场景探测、故事单元探测、关键帧提取等分层结构挖掘技术以及叙事结构挖掘技术;新闻视频的内容挖掘技术包括字幕探测与识别、人脸探测与识别、口播帧探测等语义内容的标注技术。这两种挖掘技术之间存在互补关系。

(4)高层的新闻视频挖掘技术是在低层新闻视频挖掘技术、音频挖掘技术以及描述信息的基础上,运用类似传统数据挖掘的方法来提供用户解决问题的能力。这里涉及到的关键技术包括多模态融合、视频摘要、统计分析、可视化、分类、聚类、关联、趋势分析等等,它们的称呼虽然与传统的数据挖掘相似,但这些技术因为具体应用领域的不同将被分别赋予不同的操作内涵。例如视频摘要技术,在传统的数据挖掘技术中被更多地称为“总结”或“概要”。而多模态融合技术在传统的数据挖掘技术中并不存在,是多媒体数据挖掘技术所特有的。

4 新闻视频挖掘的任务

JiaWei Han 在其著作《数据挖掘:概念与技术》一书^[1]中曾经提出过这样几个问题,即是不是所有的模式都是有趣的?数据挖掘是不是能够挖掘出所有有趣的模式?他的答案是否定的。这个回答至少说明了两个问题:首先,挖掘是有目的的。妄想做一个大而全的能够挖掘出所有有趣模式的数据挖掘系统不太现实,其效率也不能得到保证。我们需要根据自己特定的目的和任务来对搜索进行聚焦。其次,挖掘出来的模式未必都是有趣的。一个模式是有趣的,必须满足以下条件^[1]:它易于被人理解;是有效的;是潜在有用的;是新颖的。有趣的模式才能成为知识。在许多情况下,有时候系统挖掘出来的结果可能与用户最初的兴趣存在差异,但这一结果将可能改变用户的初衷,指导用户进行下一步更深层次的挖掘。

既然挖掘带有一定目的性和引导性,那么在实施具体的新闻视频挖掘之前,首先需要解决这样一个问题,即挖掘什么?我们认为新闻视频的挖掘应包括以下任务:

4.1 结构的挖掘

即通过对新闻视频的分析,挖掘出与新闻视频结构有关的知识。

4.1.1 分层结构挖掘

同其它类型的视频类似,新闻视频也存在着层次化的结构,其最高层是整个视频流,按照颗粒度从大到小的次序依次以故事单元、场景、镜头、关键帧来排列。新闻视频分层结构的挖掘是在镜头分割的基础上,利用镜头间的相似性,采用数据挖掘的一些方法获取结构方面的信息,将视频组织成更高层次的结构的过程。其间需要用到镜头探测、场景探测、故事单元探测、代表帧提取等关键技术。

4.1.2 叙事结构挖掘

新闻视频一般都以播音员镜头(口播帧)作为新闻故事的开始,以新的口播帧的出现作为上一个新闻故事的结束。也就是说,每条新闻的内部多半以口播帧、新闻故事的结构排列。除此之外,各个电视台新闻节目的编排结构也存在很大差异。例如,中央台每天的新闻联播存在这样的结构:即新闻提要、国内新闻、国际新闻、广告、天气预报;凤凰卫视中文台的台北直播室栏目采用这样的结构编排:即今日要闻、头条话题、财经报道、重点扫描以及专题报道。像这类新闻视频结构的挖掘需要借助视频分析的方法以及数据挖掘技术中的分类等方法来完成,挖掘出来的结构将有助于获得不同电视台节目的编排风格,以及发现不同电视台舆论的倾向性。同时,也可以为其它的新闻视频挖掘任务奠定基础。

有关这方面的工作最早可以追溯到 Yeo 提出的场景转移图^[11](HSTG 图),当时设计 HSTG 图的初衷是为了更好地浏览与检索视频。JiaYu Pan 提出的 VideoGraph^[12]是一种利用图的方式来挖掘视频结构的方法。该方法在镜头分割的基础上对镜头进行分组,进而建立图的方式反映视频在结构上的特征。他特别分析了新闻访谈节目的结构以及广告节目的结构,发现广告节目中 VideoGraph 图的结点较为孤立,而新闻节目则多半存在重复的镜头交叉。这种方法的思路与早期的 HSTG 图不谋而合,但 VideoGraph 图的意义远远超出了 HSTG 仅用于浏览与检索的局限性,它可以广泛地运用于视频的分类(如新闻与广告的分类)上,也可以用于提取关键帧。除此之外,Marcin^[13]采用模糊决策树的方法来对视频关键帧的组成进行分析,用于发现新闻视频中特殊的结构,例如发现 inlays in video frames, errors in shot detection 以及 host3 类结构。

4.2 语义内容的挖掘

在新闻视频中,一些事件的出现往往蕴涵着丰富的语义内容,我们称之为语义事件。例如新闻视频中出现的标题事件,对于人们理解新闻故事的语义内容具有非常重要的作用;又如新闻视频中出现的人物特写,很容易引导人们产生各种联想。例如新闻中出现江主席和美国总统布什的特写,即使没有标题字幕的辅助,人们基本上也可以肯定这大概是一条与中美关系有关的新闻。当然,这样的关联是建立在一定的知识辅助的基础之上的。

有关这方面的研究,比较有代表性的是美国 Purden 大学 Zhu^[14]提出的医学视频挖掘方法,该方法通过对视频的结构进行分析,对 dialog、presentation、clinical operation 三类事件进行了挖掘,并在此基础上生成视频摘要。

4.3 视频摘要

视频摘要是以自动或半自动的方式对视频的结构和内容进行分析,从原视频中提取出有意义的部分,并将它们以某种方式进行组合,形成简洁的、能够充分表现视频语义内容的视频概要。它是对视频内容概括性特征的总结,是对长数据的浓缩,也是一个从低层次抽象到高层次的过程。

4.4 趋势挖掘

趋势挖掘是自动在大型的数据集中寻找预测性信息,根据时间序列性数据,由历史的和当前的数据推测未来的数据。新闻视频中趋势的挖掘主要是通过通过对某个新闻事件的跟踪,发现它的发生、发展的趋势,并对未来可能的趋势进行预测。例如,通过对连续一段时期新闻事件的跟踪,发现在前段时期内有关 SARS 的新闻报道非常频繁,从而可以得出这样的结

论:前段时期 SARS 疫情比较严重,从而引起了媒体界的普遍关注。近段时期,有关 SARS 的报道呈减弱趋势,于是人们可能会有这样的预测:即 SARS 疫情已经逐步缓解,可望在未来的一个月里解除警戒。

4.5 关联挖掘

在传统的数据挖掘领域,有关关联挖掘的典型事例就是购物篮分析^[1]。在新闻视频挖掘中,同样存在关联规则的挖掘问题,主要用来发现不同事件或特征之间的内在联系。如果两项或多项属性之间存在关联,那么其中一项属性值就可以依赖其它属性值进行预测。关联挖掘通常都带有不确定性,因此关联分析可用可信度和支持度来度量。新闻视频中的关联挖掘主要从以下几个方面展开:

4.5.1 低层特征和高层语义之间的关联

传统的多媒体信息检索技术没有得到进一步推广和应用的很大一部分原因就是未能有效地解决低层的物理特征和高层语义之间的关联问题,从而造成了低层的特征与高层的语义之间存在鸿沟。如何发现低层特征和高层语义之间的关联是多媒体数据挖掘需要解决的一个问题。

从新闻视频的角度去考虑,低层特征与高层语义之间的关联如图 3 所示。低层的物理特征指的是视频数据本身所固有的一些特征,如镜头代表帧的颜色、纹理、形状、边缘、背景等等;中层的逻辑特征在物理特征的基础上略带了一层语义,例如播音员对象,它表现在物理特征上就是镜头的运动量小、背景比较固定。字幕对象在物理特征上表现为边缘变化较为剧烈,一般会在固定区域周期性出现。这说明视频低层的物理特征与中层的逻辑特征之间存在某种隐含的关联。同样地,在中层的逻辑特征与高层的语义特征之间也存在隐含的关联。如果一个视频中出现了多个播音员镜头,且在整个视频的分布中具有一定的周期性,那么这个视频是播报新闻的可能性为 80%。用规则可以表示成如下的形式:

(Anchor Shot, Periodicity) ⇒ News Report

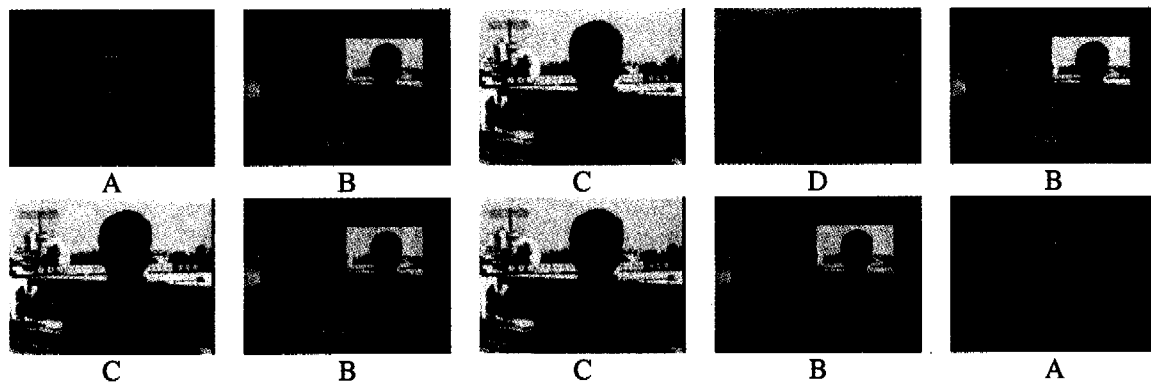


图 4 新闻视频中镜头的关联挖掘

图 4 所示的是播音员贺红梅通过电话采访水均益的新闻,讲述的是伊拉克加紧备战这样一则新闻故事。通过对新闻视频中镜头的排列顺序进行分析,我们可以发现电话采访类视频编辑的风格以及句法结构,同时对视频场景的构造以及基于关联的视频摘要生成都有着积极的影响。有关视频镜头关联挖掘的研究有代表性的是 Zhu^[16]提出的利用镜头关联挖掘生成摘要的方法以及 Yuya^[17]提出的从视频流中挖掘镜头编辑规则的方法。

4.6 兴趣点挖掘

在新闻视频的挖掘中,不同的决策人员所关注的话题是

[support=30%, confidence=80%]

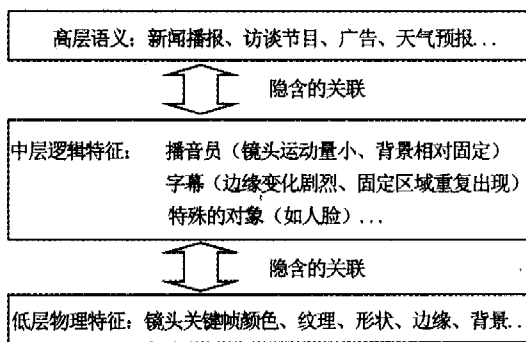


图 3 低层特征和高层语义之间的关联

4.5.2 新闻视频事件之间关联

新闻视频中的事件可分为两类:一类是从视频编辑角度提出的事件,例如镜头边界事件、故事边界事件、人脸事件、字幕事件等等,这些事件的出现之间可能存在某种关联和规律性。通过对这类事件之间的关系进行分析,可以得出视频的编辑句法。另一类事件指的是新闻视频故事中所讲述的含有语义内容的事件,例如美伊战争、非典事件、恐怖主义活动等等。这些事件之间也可能存在某种关联,例如美伊战争对世界石油价格的影响、对世界格局的影响;SARS 对旅游业的影响、对经济的影响;以及这些事件之间存在着怎样的关联等等。JiaYu Pan^[15]提出了一种 GeoPlot 方法,该方法用以挖掘不同事件之间的关联。例如:地震与龙卷风的关系、洪水和飓风的关系,等等,并通过可视化的方式来表现事件之间的关联。

4.5.3 视频镜头关联挖掘

视频镜头之间的关联挖掘是指对镜头的排列次序进行分析,发现其中隐含的关联模式。例如,对于访谈类新闻节目,可能会存在如下的镜头排列顺序:ABCDBCBCBA...,如图 4 所示。

不同的。为了更好地辅助决策人员的决策,有必要对新闻视频中的新闻事件进行分类,建立一个决策人员所关注话题的数据超立方体。例如,决策人员关心与伊拉克战争相关的话题,那么伊拉克战争问题就可以看作一个大的数据超立方体。与伊拉克战争有关的其它事件,如战争期间伊拉克人民的生活问题,不同国家对伊拉克战争的不同立场,伊拉克战争给世界政治、经济带来的影响以及伊拉克战后重建等问题都可以看作这个大数据库立方体的各个面,从而实现用数据超立方体来描述决策人员兴趣点的目的,这个超数据立方体的构建可以通过新闻事件的分类来完成。

4.7 分类

分类用于反映同类事物共同性质的特征性知识和不同事物之间的差异性知识。分类的导出模型一般用神经网络、决策树、if-then 规则来表示。

对新闻视频的分类研究最多的就是新闻视频中新闻与广告的划分。有关这方面的研究包括 JiaYu Pan^[18] 提出的 VideoCube 方法。该方法通过对视频流的时域和空域信息进行建模,得到 VideoCube 作为元数据,并对其进行 ICA 特征提取,得到 VideoBases。接下来对 VideoBases 进行训练,得到某类的基本函数,利用基本函数来确定视频类别。这种方法对视频特征都比较有效,而且符合人的感知,可以达到 81% 的分类准确率。

另外,根据具体应用的不同,可将新闻视频划分为不同的类型。例如,根据制作风格的不同,可将新闻视频分为播报类新闻(如央视的新闻联播)和访谈类新闻(如中央电视台的焦点访谈节目);根据节目内容的不同可分为国内新闻、国际新闻、体育新闻等,或进一步划分为政治类、经济类、军事类、科技类新闻等等。

4.8 聚类

所谓聚类,是指对一组个体按照相似性聚集成类,使得类间的距离尽可能的大,而类内的距离尽可能的小。与分类有指导性的学习不同,聚类是一个无监督的过程。聚类作为视频挖掘的一种方法,被广泛地运用到了各个领域。很早就有人利用聚类方法来构造视频的场景或抽取关键帧,如 Zhuang^[19] 等提出的利用聚类抽取关键帧, Yueng^[20] 等人利用聚类来分割视频并构造场景, Hanjalic^[21] 利用聚类的方法来生成视频摘要等等。相比其它的挖掘任务,聚类技术的研究最早,应用得也最广泛,几乎涉及到了视频所能应用的各个领域。

4.9 偏差分析

偏差分析是对差异或极端特例的描述,用于揭示事物偏离常规的异常现象。偏差分析一般表现在与数据的一般行为或模型表现不一致,通常应用于特定的领域。比较常见的例子是信用卡诈骗。在情报分析领域中的预警,也可以看作是偏差分析的一个应用。如果发现某个事件发生非常频繁,超出了预先所能承受的范围,则认为该事件预示着异常出现,需要引起有关部门的高度重视。

结束语 新闻视频挖掘技术是一个新兴的、富有挑战性的前沿性研究领域,它受数据挖掘、多媒体信息检索、多媒体数据库、计算机视觉、模式识别以及人工智能等多个学科领域的影响,作为多媒体数据挖掘的典型代表必将对其产生巨大的推动作用。新闻视频挖掘需要进一步解决的问题包括:如何将数据挖掘的基本理论与方法和新闻视频的特征有机的结合起来?如何从新闻视频的内容入手,利用其丰富的时空特性以及视觉特性来解决知识发现问题?如何设计新的适合新闻视频挖掘的方法,如何充分利用多种媒体特征之间的互补性和增强性,如何利用视频直观的特点可视化地表达知识,以及如何开发更加高效的视频挖掘方法等等。有关这方面的研究在国内外还刚刚起步,有许多新的理论、方法和应用等待我们去探索和发现。

本文首先从概念上对新闻视频挖掘进行了界定,并对它与其它学科之间的关系进行了分析。在此基础上提出了新闻视频挖掘的层次框架和技术框架,指出新闻视频挖掘是分层次的,包括低层视频挖掘和高层视频挖掘。其中,低层视频挖掘是利用数据挖掘的方法对视频内容进行分析的过程,而高

层数据挖掘则是在低层挖掘的基础上进一步发现视频中的知识的过程。新闻视频挖掘的技术框架则对挖掘所涉及到的具体技术进行了分析。最后,在框架的基础上对新闻视频挖掘的任务进行了详细的阐述。指出新闻视频挖掘的任务包括结构挖掘、语义内容挖掘、视频摘要、趋势挖掘、关联挖掘、兴趣点挖掘、分类、聚类、偏差分析等等,并对各种任务举出了具体的示例加以说明。本文从总体的角度对新闻视频挖掘的概念、框架、技术和任务等各方面都进行了积极深入的思考,我们相信它的深入研究将在战略决策、情报分析、企业规划等领域发挥积极的作用。

参考文献

- Han Jiawei, Kamber M. 数据挖掘概念与技术. 范明, 孟小峰, 等译. 北京:机械工业出版社, 2001
- Zaiane O R, Han JiaWei, Li Ze-Lian, et al. MultimediaMiner: A system prototype for multimedia data mining. In: Proc. 1998 ACM-SIGMOD Conf on Management Of Data (SIGMOD'98), Seattle, WA, June 1998. 581~583
- Wijesekera D, Babara D. Mining cinematic knowledge: Work in progress-An extended abstract. In: Proc. of the 1st International Workshop on Multimedia Data mining (MDM/KDD'2000), Boston, MA, USA, August 2000. 98~103
- Oh JungHwan, Bandi B. Multimedia data mining framework for raw video sequences. In: Proc. of the 3rd International Workshop on Multimedia Data Mining (MDM/KDD'2002), Edmonton, Alberta, Canada, July 2002. 1~10
- Zhu Xingquan, Fan Jianping, Aref W G, et al. ClassMiner: Mining medical video content structure and events towards efficient access and scalable skimming. In: Proc. of ACM SIGMOD Workshop on Data Mining and Knowledge Discovery (DMKD 2002), Madison, WI, June 2002. 9~19
- Shearer K, Dorai C, Venkatas S. Incorporating domain knowledge with video and voice data analysis in news broadcasts. In: Proc. of the 1st International Workshop on Multimedia Data Mining (MDM/KDD'2000), Boston, MA, USA, August 2000. 46~53
- 曹加恒, 舒风笛, 张凯, 等. 基于多媒体数据库的数据挖掘原型系统. 武汉大学学报(自然科学版), 2000, 46(5): 567~570
- 史忠植. 知识发现. 北京:清华大学出版社, 2002
- 李国辉, 张军, 汤义. 多媒体挖掘技术. 计算机世界报, 2002-07-22
- 吴玲达, 老松杨, 王辰, 等. 多媒体信息系统. 北京:电子工业出版社, 2002
- Yeo B L, Yeung M. Classification, simplification and dynamic visualization of scene transition graphs for video browsing. In: Proc. SPIE'98, Storage and Retrieval of Image and Video Databases VI, San Jose CA, Feb. 1998
- Pan Jia-Yu, Faloutsos C. Video Graph: A new tool for video mining and classification. In: JCDL'01, Roanoke, Virginia, USA, June 2001
- Detyniecki M. Discovering indexing rules for video-news. In: Proc. of the European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems - EUNITE'2002, Algarve, Portugal, September 2002. 44~46
- Zhu Xing-Quan, Aref W G, Fan Jianping, et al. Medical video mining for efficient database indexing, management and access. In: Proc. of 19th ICDE Conference, Mar. 2003
- Pan Jia-Yu, Faloutsos C. GeoPlot: Spatial data mining on video libraries. CIKM'02, McLean, Virginia, USA, Nov. 2002
- Zhu Xing-Quan, Wu Xindong. Sequential association mining for video summarization. In: Proc. of IEEE Int Conf on Multimedia & Expo (ICME 2003), Vol3, Baltimore, MD, July 2003. 333~336
- Matsuo Y, Amano M, Uehara K. Mining video editing rules in video streams. In: Proc. of ACM Multimedia'02, Juan-les-Pins, France, Dec. 2002
- Pan Jia-Yu, Faloutsos C. VideoCube: A novel tool for video mining and classification, 2002
- Zhuang Y, Rui Y, Huang T S, et al. Adaptive key frame extraction using unsupervised clustering. In: Proc. of ICIIP'98, 1998
- Yeung M, Yeo B L, Liu B. Segmentation of video by clustering and graph analysis. Computer Vision and Image Understanding, 1998, 71(1)
- Hanjalic A, Zhang H J. An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis. IEEE Trans on Circuits and Systems for Video Technology, 1999, 9(8): 1280~1289