

# 多媒体数据集中的数据挖掘:系统框架和方法\*

武德峰, 李国辉, 来旭, 胡军涛

(国防科技大学 管理科学与工程系多媒体实验室, 湖南长沙 410073)

**摘要:** 分析了目前信息环境对多媒体挖掘技术提出的需求, 提出了一种适合多媒体挖掘的系统框架、一般结构和挖掘过程; 描述了各种媒体可用于挖掘的特征, 阐述相应的适合多媒体挖掘的基本方法; 讨论了多媒体挖掘技术今后发展可能遇到的问题和面临的挑战。

**关键词:** 多媒体挖掘; 智能多媒体; 多媒体信息管理; 多媒体数据库

中图法分类号: TP311.134

文献标识码: A

文章编号: 1001-3695(2005)02-0053-04

## Systematic Framework and Methods for Multimedia Mining

WU De-feng, LI Guo-hui, LAI Xu, HU Jun-tao

(Multimedia Lab., Dept. of Management Science & Engineering, National University of Defense Technology, Changsha Hunan 410073, China)

**Abstract:** Proposed a general framework and a model suitable for multimedia mining, including multi-dimensional data cube, mining engine, interactive mining interface. In this framework, heterogeneous-feature space, new mining methods, new form of patterns and knowledge applicable to multimedia data are crucial to be introduced and discussed. The preliminary experiment on video structure mining indicates the proposed framework is valid.

**Key words:** Multimedia Mining; Intelligent Multimedia; Multimedia Information Management; Multimedia Database

近年来, 数据挖掘技术一直是人们研究的热点, 在这个领域中已经取得了一些令人瞩目的成果<sup>[1]</sup>, 很多商用系统和研究原型已经推出。随着信息技术的发展, 人们日常生活中所接触的数据形式不断地丰富, 多媒体数据库的数量日益增多, 数据量日益增大, 出现了很多海量的多媒体数据库。

原有的数据库技术已满足不了应用的需要, 人们希望从这些媒体数据中得到一些高层的概念和模式, 找出蕴涵于其中的有价值的知识。随着多媒体数据库技术的日趋成熟、数据挖掘技术应用的成功, 使得人们将目光放到了在多媒体数据库中进行知识发现<sup>[2]</sup>。这种将数据挖掘技术和多媒体信息处理技术有机地结合起来形成的在多媒体数据中进行知识发现的信息处理方法就是多媒体挖掘<sup>[3]</sup>。多媒体挖掘就是从大量的多媒体数据集中, 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, 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。

多媒体挖掘是数据挖掘技术和多媒体技术的结合产物, 是知识发现、数据挖掘、人工智能、机器学习、数据库技术、多媒体技术等学科的交叉研究领域。它引起了相关学科研究者的高度重视, 已经成为一个引人瞩目的研究热点<sup>[4,5]</sup>。虽然上述几个学科和研究方向各自都已经有了较为深入的发展, 但目前多媒体数据挖掘处于研究的起步阶段。多媒体挖掘的概念及其系统框架、研究的范围以及使用的方法都没有一个系统的、完整的、统一的界定。另外, 对于各种多媒体挖掘任务, 也没有形成可重复的方法。

## 1 多媒体数据挖掘的系统框架

### 1.1 多媒体数据挖掘的概念

多媒体挖掘是一种智能多媒体技术, 使多媒体的处理和管理从信息存取上升到知识获取的层次。多媒体数据是非结构化或者是半结构化的<sup>[6]</sup>, 各媒体数据有着不同的特点, 有着各自表述信息的方式, 各媒体既可独立表示信息又可共同表示相同事件的不同特征, 共同描述事件的存在、发展和结果。因此, 多媒体数据集中必定存在关于信息主体的特征、属性以及它们之间的关系, 或者存在着某些人们从直观上无法得到的模式。

**定义** 多媒体数据挖掘。基于多媒体数据的内容特性  $C$ , 以及这些特性的相关语义, 从大型多媒体集  $M$  中 (不仅仅是多媒体数据库), 发现和分析出隐含的、有效的、有价值的、可理解的模式  $P$ 。可以抽象地把挖掘用多维特征空间的一种映射来表示:

$$P = (M|C) \quad (1)$$

### 1.2 多媒体挖掘的内容

多媒体数据包含着十分丰富的内容特性, 对于这些特性的分析、提取以及获得它们之间的关系和模式都属于多媒体数据挖掘的范畴。

(1) 图像挖掘。图像包含着丰富的视觉特性和空间特性。视觉特性表示颜色、纹理、形状、轮廓等特征; 空间特性包括图像中目标的模式、布局、空间关系等。运用图像处理技术, 如图像分割、边缘探测、边缘提取、模式识别等, 从图像 (视频帧) 中抽取能代表、区分该图像的结构化内容的特征, 同时收集用于图像处理和数据挖掘所需要的知识, 获得各种元数据及领域知识, 建立特征库和知识库。

当挖掘任务确定后,抽取与挖掘任务相关的特征向量,形成多维特征向量,在这些特征向量空间中比较、分析各向量之间的距离或相似关系,在知识库的指导下,完成对图像内容的分析、索引、摘要、分类、聚类、关联等操作,将所得到的结果进行解释和表示,从而获得高层的概念或模式。

(2) 视频挖掘。视频包括丰富的内容特性,除了图像具有的视觉特性和空间特性外,还具有时间特性、视频对象特性和运动特性等。运用视频处理技术,可以将视频按照各种属性(如场景、视频对象或运动特性)进行分割,然后进行分类、聚类等操作,得到视频的结构模式。也可以从视频中提取视频对象,跟踪其运动,结合时间特性分析其模式以及与其他对象之间的关联,从而发现高层次的事件摘要、概念或模式。

视频挖掘可以广泛地应用到新闻视频、监控视频、纪录片、数字图书馆等应用系统中,如从交通监视视频中,提取交通事故的模式、分析交通拥堵的原因、趋势;从连续的侦察图像中分析出战场情况的变化等。

(3) 音频挖掘。音频是听觉媒体,其主要特征有基音、音调、韵律或旋律等。音频挖掘通常有两种途径:运用语音识别技术将语音识别成文字,将音频挖掘转换成文本挖掘;直接从音频中提取声音特征,如音调、韵律等,运用聚类的方法分析声音模式。机器学习技术,包括粗糙集、人工神经网络和决策树技术能够用于分析音频的基频、能量分布及其他特征,从而获得音频事件和对象的结构,挖掘出隐含在音频流中的信息线索、规律和模式。如通过对海量语音数据库中语音特征的提取和学习,获得音调和韵律变化的模式,使得语音合成更加自然化和智能化。

(4) Web 挖掘。Web 是一个庞大的、动态的、超媒体结构的包含多媒体和数据元素的信息库。Web 挖掘是从 Web 文档、媒体、结构以及用户交互行为中抽取感兴趣的、潜在价值的模式和隐含信息的过程。Web 挖掘主要有三种形式: Web 内容挖掘。它是对 Web 页面进行挖掘,即对页面中的文本和各种媒体内容的挖掘,包括对页面进行摘要、分类、聚类和关联操作。 Web 结构挖掘。它是在 Web 链结构中发现知识,链结构能够反映 Web 空间中信息的流向和分布,能够反映 Web 元素的性质和特点,如一个站点的入链数目远大于出链数目,那么这个站点可能是服务器;如果一个页面经常被引用,那么可以反映该页面的流程度和重要性等。 Web 使用方式的挖掘。它主要是跟踪用户与 Web 的交互和操纵,包括对服务器方访问记录的挖掘,对访问路径的分析。挖掘的主要手段有路径分析、分类和聚类、关联和序列模式的发现等。对 Web 使用的挖掘可以获得用户的属性和使用 Web 的特点,如可以划分用户群体,获得访问模式和爱好,发现不合理的链等,从而帮助改进 Web 站点的设计和进行主动的信息分发。

Web 挖掘不仅涉及页面信息内容、站点分析和设计,在网上电子商务中也有着很大的应用价值。Web 挖掘是目前前沿的研究领域,在技术上是几个方向的交叉点,在应用上具有很大的潜在价值。

(5) 多媒体综合挖掘。多媒体概念与单媒体的区别在于,它是一个集成的系统概念,媒体之间有联系。多种媒体的有机复合效果(如信息量)大于单媒体效果的总和。上面提到的仅是单媒体的挖掘,多媒体的挖掘将是在挖掘过程的各个阶段,

综合利用多媒体的特性进行知识发现。

在数据准备阶段,多媒体的集成可以增强自动预处理分析的准确性,解决语义模糊性;在挖掘阶段,各种媒体的元数据可以互补使用,挖掘出意想不到的合理的信息线索、模式、趋势或关系;在结果表达和解释阶段,多媒体的同步和互补呈现使得知识可视化和交互接口更直观、更易于理解。

## 2 一般系统结构

### 2.1 系统结构

由于与常规数据的特性不同,针对多媒体数据的挖掘系统需要新的处理模块的支持。图 1 所示的一般系统结构表示在特征选取、多维内容空间构造、挖掘引擎和多媒体知识表达等方面需要新的理论、方法和模型。

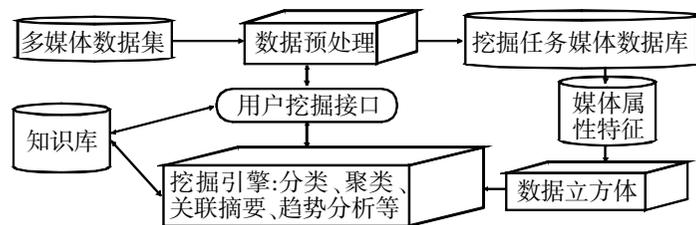


图 1 多媒体数据挖掘系统结构

(1) 预处理模块。它是对多媒体原始数据进行预处理,提取有效特征,可以是对多媒体数据的结构化处理,如图像对象分割、视频和音频对象分割、视频与音频逻辑和时空分段、视觉和听觉特征提取、运动特征提取、事件标记、叙事结构组织、语义关联等,它们以元数据的形式记录在元数据库中。

(2) 多媒体数据集。大型多媒体数据集可能包含几十万幅图片、几千小时的视频和音频,它们的媒体结构与元数据库中的描述关联,用于可视化表现和存取。元数据库是一种按照挖掘要求组织的多维、多层次、多媒体属性数据库,支持高效率的多媒体挖掘。

(3) 挖掘引擎。它包含一组快速的挖掘算法,如分类、聚类、关联、总结和摘要、趋势分析等,系统可以根据具体的应用选择一个或多个相应的挖掘算法,对元数据库进行挖掘。

(4) 用户接口。挖掘结果的可视化和解释界面,也可以为用户提供交互接口和扩展 SQL 挖掘语言。由于多媒体的视听和时空特性,挖掘出来的模式应该以新的表现方式呈现出来,如导航式知识展开和交互式问题求解过程,以及提供挖掘结果的可视化接口。

如果是实时获取的媒体数据或数据流,则需要实时处理媒体、实时挖掘、实时呈现趋势、关联和事件。

这一挖掘系统结构与多媒体挖掘模型一致。在图 2 所示的挖掘模型中,多媒体数据的信息建模是多媒体挖掘的基础,位于下层,它从大容量多媒体原始数据中建立多维、多层次的信息概括(抽象)模型;在此基础上,进行时空和视听模式的抽取;最高层是把抽取的模式用表示模型(或模板)可视化和解释出来。

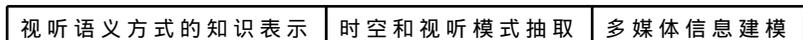


图 2 多媒体挖掘模型

### 2.2 挖掘过程

(1) 数据准备。集成处理大型的、异构的媒体资源,选择有效的和与任务相关的媒体数据集,提取多媒体内容特征和

属性,形成基础的元数据库。对于多媒体数据,需要利用内容处理技术获取多媒体对象和数据的时空、运动和视听特性,以及对于多媒体结构元素的语义,把这些特性有效地组织成为一个或多个多维元数据库(立方体),支持多媒体挖掘。

如何从大规模多媒体数据中提取有效的元数据,并把这些多维的(多维特征)、异构的(复杂结构的和各种类型的)元数据组织成有效的元数据库是一个关键问题。这里包括直接从压缩的多媒体数据中快速获取元数据,并按照一种适合的开采算法的结构组织起来。表达多媒体这些特性和语义的元数据不同于常规数据库中的表格数据,它们是多维异构的元数据集,有效地组织一种元数据库支持多媒体挖掘,是多媒体挖掘中关键的第一步。

(2) 挖掘。在挖掘过程中必须基于多媒体的内容特性,采用新的适合多媒体特性的挖掘方法和技术。因为多媒体与常规字符数字数据的特性在许多方面都不同,而且多媒体中的视频、图像和音频本身又各有其特有的性质,因此除了采用常规的数据挖掘中的统计和相关性分析等方法,从多媒体角度出发,对时间、空间、运动特性、视听特性及其语义进行分析之外,还需要采用适合多媒体挖掘中特有的方法,如概要、摘要、筛选、聚类、时间周期、位置分布、叙事结构等新的挖掘和分析方法。

目前一个主要的方法是通过构建多媒体数据立方体来对大型的多媒体数据库进行多维分析。多媒体数据立方体可以包含针对多媒体信息的维和度量,如颜色、纹理和形状维及其度量。多媒体数据立方体可以有好多维,如图像和视频的名称;图像的尺寸和视频的长短;建立时间;图像和视频的格式;关键字等。其中很多属性是集合值而不是单值,如一个图像可能包含一组关键字。设计多媒体数据立方体也十分复杂,维数过多,数据立方体计算复杂度过高;维数太少,模型过于粗糙,不利于数据的多维分析。

在多媒体数据立方体上可进行多种操作:上卷、下钻、切片和切块等。在此基础上进行基于内容的多维分析和多种知识的获得,包括摘要、汇总、比较、分类、聚类和关联等。

挖掘算法应该是高效快速的,并且能够切实地从元数据库中挖掘出有价值的知识。关键问题是如何根据多媒体特有的内容及其语义,研究出新的有效的挖掘算法,发现隐含的知识或验证用户的假设。例如,如何根据视频的运动特性分析出事件和事实的趋向和联系;如何挖掘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挖掘出事件和事实的空间分布等。

(3) 结果表达。多媒体是一种视觉和听觉媒体,挖掘的结果更有效的呈现方式应该是可视化的呈现。根据最终用户的任务要求和目的,把最有价值的信息区分出来,一目了然地以时间展开、空间分布和运动趋向等方式向用户呈现开采结果,可视化地呈现事实、事件的叙事和因果关系。另外一种可能的的方式是以多媒体特有的导航式知识呈现方式引导用户一步步发现感兴趣的知识。如果挖掘结果不够满意,可以调整参数,重复以上挖掘步骤,直到得出满意的结果。多媒体挖掘得出的知识可用概念、事件、规则、规律、模式、约束等形式可视化表示和解释出来。

### 3 多媒体数据挖掘的基本方法

多媒体数据在经过预处理之后成为结构化或半结构化的数据,并建立了元数据库和特征库。挖掘任务确定之后,抽取与之相关的特征矢量,建立多维特征矢量空间。在这个基础上,就可以对多媒体数据进行多维分析,发现有用的视听和结构模式。

#### 3.1 多媒体数据分类

分类是多媒体数据的一种分析方法。它是通过与给定的样板特征进行比较,将多媒体对象分成预定义类的过程,它可分两步进行:

(1) 建立一个模型用于描述预定义类集。分类通常是依据媒体数据的某一特性来确定的,该特性在元数据库中表现为视听描述子。通过分析元数据库中部分数据的该描述子值来构造模型,并把用于建立模型的媒体数据作为训练集。训练样本可以随机选取,并预先给出类标号。

(2) 利用模型进行分类。首先要评价分类模型分类的准确率,采用测试数据集来检测模型的准确率,然后用于对未知类型的数据进行分类。

分类是一个逐步细化的过程,对具有同一类标号的媒体数据集还可以按照上面的步骤继续分类。分类方法可以采用常用的判定树归纳、贝叶斯分类、神经网络算法等。

#### 3.2 多媒体数据聚类

多媒体数据聚类就是运用某种算法,使得多媒体对象集中具有相似特征的对象聚成组的过程。经过聚类,同一类的对象之间的相似性尽可能大,而不同类别的数据之间的相似性尽可能小。聚类不同于分类,分类事先要定义好类,并选择带有类标号的训练集来建立分类模型,属于有指导学习;而聚类不依赖预先定义的类和带标号的训练集,属于无指导学习。对象之间的相似性是根据对象特征向量之间的距离来计算的,常用的方法有加权欧氏距离、相异度矩阵等。

聚类分析可以发现全局的模式,获得数据的分布情况。观察每个簇的特点,对某些有价值的簇进行进一步分析,可以得出有价值的知识。例如在医学上对大量的CT照片进行聚类,找出某一疾病的特点,帮助诊断;在气象科学中,可以将大量的卫星云图进行聚类,找出同一类天气情况的特征,帮助提高天气预报和灾害预报的准确性。

#### 3.3 多媒体数据的关联规则挖掘

多媒体数据的关联规则挖掘是在相关的多媒体对象集中,找到一组关联规则,显示一组对象或特征的模式或相互关系的发生频率。一个典型的关联规则为

$$X \rightarrow Y [s\%, c\%] \quad (2)$$

其中,  $X$ 和  $Y$ 是一组特征描述的谓词,  $s\%$ 是规则的支持度( $X$ ,  $Y$ 共同出现的概率),  $c\%$ 是规则的可信度( $X$ 出现时  $Y$ 出现的概率)。下面以图像中相关联规则的挖掘为例来说明。

图像的关联规则是指图像对象或特性之间频繁出现的模式。设  $D$ 为图像集,那么关联规则可以表示为

$$P_1 P_2 \dots P_n Q_1 Q_2 \dots Q_m (c\%) \quad (3)$$

其中,  $P_1, P_2, \dots, P_n, Q_1, Q_2, \dots, Q_m$ 是  $D$ 中图像的描述谓词,可以是图像的大小、颜色、纹理、密度、对象、空间位置、文本描述

等;  $c\%$  表示该规则的可信度, 其含义为当  $P_1, P_2, \dots, P_n$  都发生时,  $Q_1, Q_2, \dots, Q_m$  也都发生的概率。

在挖掘的过程中, 我们首先建立描述集。描述集是一组谓词的集合, 如  $\{P_1, P_2, \dots, P_n\}$ 。对于每一个描述, 定义一个支持度来表示该描述在整个图像集中出现的概率, 记作  $(P_i/D)$ 。描述集的支持度表示集合内的所有描述同时发生的概率, 记作  $(P_1 P_2 \dots P_n/D)$ 。如果只考虑满足最小支持度的描述或描述集, 若一组描述集中有一个描述子集不满足最小支持度, 则忽略该描述集。满足最小值尺度的描述集被称为频繁描述集, 从频繁描述集中可以推导出规则。例如, 从频繁描述集  $(P_1, P_2, P_3, P_4)$  可以推出规则

$$P_1 P_2 P_3 P_4 \quad (4)$$

$$P_1 P_2 P_3 P_4 \quad (5)$$

$$P_1 P_2 P_3 P_4 \quad (6)$$

...

由频繁描述集推导出的规则不一定都有用, 只有满足最小可信度的规则才是我们所关心的规则, 即强关联规则。规则的可信度可以用条件概率公式求得, 如式(4)表示的第一个规则的可信度为

$$P(P_1 | P_2 P_3 P_4) = (P_1 P_2 P_3 P_4) / (P_1) \quad (7)$$

图像的关联规则的挖掘分两个步骤: 找出所有的频繁描述集; 从频繁描述集中推导出关联规则, 并确定强关联规则。

## 4 结束语

多媒体挖掘的方法和技术涉及到数据挖掘、多媒体内容管理、多媒体内容分析、知识发现、数据仓库等领域。多媒体挖掘包括集成媒体元素和部件的模型、媒体分析和元数据分析的有效算法、知识表示模式、可视化工具等。我们把本文的系统框

架和关联挖掘方法应用到视频结构挖掘当中去<sup>[9]</sup>, 获取到了视频镜头语法和情节模式等知识。多媒体挖掘的概念、内容和方法都没有一个统一的定论, 有很多问题有待于进一步研究, 归纳有三: 多媒体挖掘的体系结构和框架; 新的和增强的挖掘算法, 使之适合多媒体数据的内容特征; 新的挖掘任务和挖掘结果的表示、解释和可视化。由于多媒体数据挖掘涉及的技术面广、应用潜力大并且刚刚成为研究的热点, 在这一领域中的研究和开发一定具有广阔的前景。

参考文献:

- [1] M Fayyad, G Piatetsky-Shapiro, *et al.* *Advances in Knowledge Discovery and Data Mining* [M]. AAAI/MIT Press, 1996.
- [2] Simeon J Simoff, Osmar R Zaiane. Report on MDM/KDD2000 [R]. The 1st International Workshop on Multimedia Data Mining, SIGKDD Explorations, 2001. 103-105.
- [3] 李国辉, 张军, 汤大权, 等. 多媒体挖掘 [C]. 北京: 第十届全国多媒体技术学术会议, 2001.
- [4] Zajane O R, *et al.* *Multimediaminer: A System Prototype: For Multimedia Data Mining* [C]. Proc. of ACM SIGMOD Conf. on Management of Data, Seattle, 1998. 581-583.
- [5] Simeon J Simoff. *Variations on Multimedia Data Mining* [C]. Boston, USA: Proceedings of the International Workshop on Multimedia Data Mining, 2000.
- [6] 胡晓峰, 李国辉. 多媒体系统 [M]. 北京: 人民邮电出版社, 1997.
- [7] 李国辉, 等. 信息组织与检索 [M]. 北京: 科学出版社, 2003.
- [8] 胡军涛. 视频高层结构分析和挖掘 [D]. 长沙: 国防科技大学硕士论文, 2002.

作者简介:

武德峰, 男, 博士研究生, 主要研究方向为多媒体信息系统; 李国辉, 男, 教授, 博士生导师, 主要研究方向为多媒体信息系统与虚拟现实。

(上接第 52 页) 现, 该过程的目的是对来自客户端的请求参数, 即客户端经过编码器生成的 XML 文件的内容进行解析, 从中解析出各种操作并动态地组装成对应的 SQL 语句, 通过 ADO.NET 连接数据库执行相应的操作。该过程的重点在于对 XML 文件的解析, 该操作是客户端编码器的逆过程, 我们称之为解析器, 该解析器接收 XML 文件, 输出动态组装成的各种 SQL 语句。之后执行 SQL 语句, 则最终实现了将对客户端数据库的操作同步到服务器数据库上的目的。同步之后, 回复一个信号给客户端, 该信号是作为 Web 服务的返回值由 SOAP 消息带到客户端的, 目的在于告诉客户端同步已完成。

### 2.4 进一步分析

进一步考查整个过程, 可以看到, 我们的解决策略的关键思想在于将对本地数据库的操作序列化到 XML 文件中, 并依托 Web 服务将 XML 文件带到远程服务器。

这一过程的优势在于: Web 服务的引入使得位于公司 Intranet 中的数据库和位于 Internet 上的 Web 站点数据库同步成为可能。同时, 由于 Web 服务请求和响应是在 HTTP 之上进行的, 这就有效地解决了穿透防火墙的问题。

XML 的引入, 则充分利用了 XML 的自描述性、体积小和易于网络传输的优势。整个同步过程不再涉及连接远程数据库的操作, 这就有效地降低了网络流量, 减小了远程服务器资源的消耗。

## 3 局限

从上述的转换方式可以看出, 在 SQL 语句和 XML 节点之间进行转换的方法中, 所给出的 SQL 语句的形式都是非常简单的, 都是对单表的操作。我们已经实现的编码器和解析器尚不支持通用 SQL 语句同对应的 XML 节点形式的转换。

## 4 总结

随着应用系统向网络化和分布式的发展, 数据库的同步, 特别是异构数据库之间信息的同步问题是亟待解决的问题。本文从实际项目出发, 提出了一种基于 Web 服务和 XML 技术的数据库同步策略并详细说明了实现的方法和步骤。由于特定项目需求的限制, 我们只是实现了简单的 SQL 操作语句同 XML 节点描述之间转换的编码器和解析器, 为了能让该模型具有通用性, 还需要对编码器和解析器进行扩充, 以支持通用 SQL 操作语句。

参考文献:

- [1] 美] 威廉逊. XML 技术大全 [M]. 智慧东方工作室. 北京: 机械工业出版社, 2002.
- [2] 郑阿奇. SQL Server 实用教程 [M]. 北京: 电子工业出版社, 2002.
- [3] [美] 弗雷泽, 利文斯敦. C# XML 入门经典——C# 编程人员必备的 XML 技能 [M]. 毛尧飞, 崔伟. 北京: 清华大学出版社, 2003.

作者简介:

刘福顺 (1980-), 男, 在读硕士, 研究方向为软件工程与工具; 唐宁九 (1960-), 男, 教授, 研究方向为软件工程与工具。