

图像纹理特征数据挖掘软件原型的设计与实现

吴涛^{1,2}, 秦昆¹, 尹宁¹, 肖启芝¹

(1. 武汉大学 遥感信息工程学院, 湖北 武汉 430079;

2. 湛江师范学院 信息科学与技术学院, 广东 湛江 524048)

摘要: 为了实现图像处理与分析的自动化、智能化, 把数据挖掘引入到图像处理与分析领域, 针对图像纹理数据的特点, 在深入研究数据挖掘软件的基础上, 提出了一种图像纹理特征数据挖掘的理想系统结构; 介绍了基于图像(遥感图像)数据挖掘的软件原型系统 RSIImageMiner 中图像纹理特征数据挖掘模块的各子功能模块的设计与实现; 并以图像纹理分割为例, 给出实验结果及分析。实验表明, 该软件原型能得到较满意的图像纹理特征, 并能有效地指导图像处理与分析。

关键词: 图像数据挖掘; 纹理分析; RSIImageMiner; 关联规则

中图分类号: TP311.52

文献标识码: A

文章编号: 1673-629X(2007)08-0019-04

Design and Development of Image Texture Feature Data Mining Software Prototype System

WU Tao^{1,2}, QIN Kun¹, YIN Ning¹, XIAO Qi-zhi¹

(1. School of Remote Sensing Info. Eng., Wuhan Univ., Wuhan 430079, China;

2. School of Info. Sci. and Tech., Zhanjiang Normal Univ., Zhanjiang 524048, China)

Abstract: For the intelligence and automatization of image processing and analyzing, data mining is introduced into image processing and analyzing. Aimed at the characteristic of image texture data, based on the in-depth studies for data mining software, a novel ideal model of image texture feature data mining system was designed; each subsystem of image texture data mining module in RSIImageMiner was designed and developed; then with examples of image segmentation, the result and some analysis of experiment are listed. And it suggests that preferable image texture features could be produced, and these features could guide image processing and analyzing effectively.

Key words: image data mining; texture analysis; RSIImageMiner; association rule

0 引言

随着信息技术的不断发展, 卫星遥感影像、生物学影像、数码照片等不断涌现, 人们被这些海量的图像数据所淹没, 人们迫切需要摆脱这种困境, 渴望能主动地从这些海量数据中获取丰富的有用信息, 图像数据挖掘就应运而生。随着 1993 年 Fayyad 博士首次提出挖掘多媒体数据的思想^[1]、2000 年首届多媒体数据挖掘年会的召开^[2], 图像数据挖掘技术有了长足的发展。图像纹理特征数据挖掘是图像数据挖掘的一个重要分支, 它把数据挖掘的方法应用到图像处理与分析中来。

将图像的纹理理解成是在像元亮度(灰度值)空间上变化的函数, 从大量的、不完全的、有噪声的、模糊的、随机的图像纹理样本数据中提取隐含其中的, 人们事先不知道的, 但又是潜在有用的知识, 发现图像纹理中所蕴含的有趣的、规律性模式, 并利用这些模式来描述图像纹理特征, 指导图像的分类、分割、检索, 为推进图像处理与分析的自动化、智能化服务, 以上是文中给出的图像纹理特征数据挖掘的解释。目前, 图像纹理特征数据挖掘的理论与方法尚不成熟, 其开发方式也以软件原型系统的开发为主^[3-6]。文中研究图像纹理特征数据挖掘软件的结构, 通过软件原型的开发, 进一步明确该类软件系统的需求, 探索图像纹理特征数据挖掘方法, 完善图像数据挖掘的理论和方法体系。

1 RSIImageMiner 简介

RSIImageMiner 是笔者参与设计开发的一个基于图像(遥感图像)数据挖掘的软件原型系统, 该软件已

收稿日期: 2006-11-15

基金项目: 973 计划资助项目(2006CB701305); 国家自然科学基金资助项目(40523005, 50379048); 地理空间信息工程国家测绘局重点实验室开放基金资助项目

作者简介: 吴涛(1980-), 男, 湖北汉川人, 硕士, 研究方向为数据挖掘、图像处理; 秦昆, 博士后, 研究方向为数据挖掘、空间分析、图像理解。

经在中华人民共和国国家版权局进行了软件登记,已经获得“计算机软件著作权登记证书”。该系统包括图像数据管理模块、光谱(颜色)特征数据挖掘模块、纹理特征数据挖掘模块、形状特征数据挖掘模块、空间分布规律挖掘模块、图像知识的存储与管理模块,以及图像知识的应用模块,即基于知识的图像分类模块、基于知识的图像检索模块、基于知识的目标识别模块等。该系统的特色在于:将图像数据挖掘理解成一个动态的概念,看成是一个由数据到知识的过程,用商空间^[7]的思想来描述该过程中的概念粒度,结合概念格理论^[8],建立了一个概念驱动的知识发现体系。

2 图像纹理特征数据挖掘软件原型的设计

图像纹理特征数据挖掘模块(Image Texture Data Mining,简称 ITDM)是 RSIImageMiner 的重要组成部分,下面详细介绍该模块。

2.1 理想结构

在前人的研究^[9,10]基础上,结合文中对图像纹理特征数据挖掘的研究,针对图像纹理数据的特点,提出了一个图像纹理特征数据挖掘的理想系统总体结构,如图 1 所示。

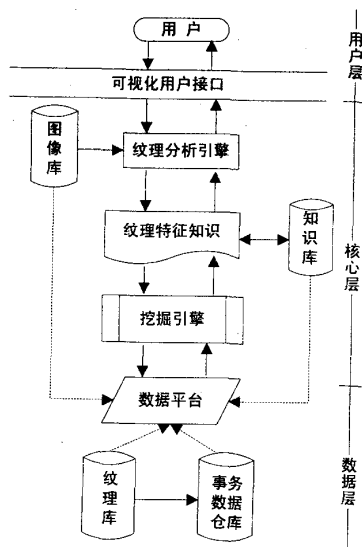


图 1 理想的图像纹理特征数据挖掘系统结构

该结构主要包含数据平台、挖掘引擎、纹理特征知识评估与表示、纹理分析引擎、可视化用户接口等子功能模块。

其中,数据平台实现对所有输入、输出数据进行存储、管理、维护,如图像数据、纹理数据、特征知识等,对事务数据仓库的维护是为了系统能够极其方便地实现增量式纹理特征数据挖掘。挖掘引擎是整个模块的核心,通过各种算法实现图像纹理数据挖掘,挖掘出来的

纹理特征知识以两种方式存储:一种是文本文件方式,方便用户针对挖掘结果进行分析,以便向系统提交反馈信息,如修改关联规则支持度、置信度或者分类等;另一种就是关系数据库方式,方便纹理特征知识评估,同时能够使数据平台对其实实现统一存储、管理、维护。纹理特征知识评估与表示主要是通过各种算法实现对规则的选择,获取最优的规则子集,并将其表示成适合用户理解的形式。纹理分析引擎是系统的最终目的,利用纹理特征关联知识,实现纹理分析,如纹理检索、纹理分割、纹理分类与聚类。一个好的软件系统,交互式的用户接口是必不可少的,图像纹理特征数据挖掘系统同样也需要一个高效的可视化用户接口。

按照数据流的处理过程,文中将图像纹理特征数据挖掘模块分为三个层次:数据层、核心层、用户层。数据层主要负责处理输入、输出数据,核心层是若干算法的集中地,是系统的“大脑”,负责输入/输出之间的转换,隐含着概念由底层数据到高层语义的提升。用户层通过交互性的可视化接口与核心层进行信息交流,实现对核心层的结果反馈,使整个模块成为一个完整的、运行良好的反馈系统。

2.2 ITDM 模块的结构

一般来说,由于各种限制,实际的图像纹理特征数据挖掘系统并不一定包含了上述理想结构中的所有方面。根据 RSIImageMiner 的实际需要,ITDM 模块重点研究了图像纹理特征关联规则挖掘,设计了 ITDM 模块的结构图,如图 2 所示。

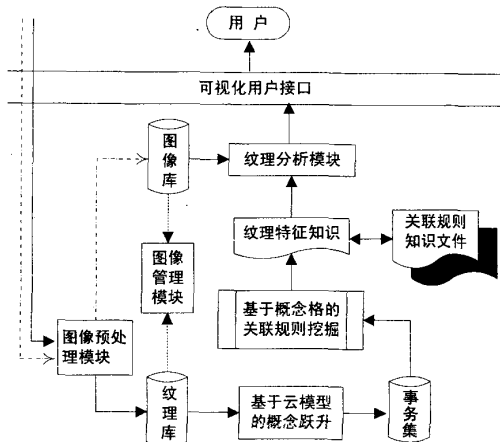


图 2 ITDM 模块的结构图

该模块结构的特色在于:以概念为主线,用云模型对原始数据进行不确定性处理,然后用基于概念格的关联规则挖掘方法,发现图像纹理关联规则,选取若干规则作为纹理特征知识存储到文本文件中。在进行纹理分析时,可以方便地调用这些特征知识指导纹理分析与处理。

3 ITDM 模块的实现

基于上述结构,文中用 VC6.0 作为系统开发工具,以原型系统的开发方式,实现了 RSImageMiner 的图像纹理特征数据挖掘模块 ITDM 的开发。其主要功能是对图像纹理特征进行挖掘,用关联规则知识描述纹理特征,进而指导纹理分析。具体包括以下五大子模块:图像数据管理模块、图像预处理模块、基于云模型的概念跃升模块、基于概念格的关联规则挖掘模块、纹理分析模块。下面分别对各个模块加以展开。

图像数据管理模块主要负责图像与纹理样本的存储、管理、维护,实际上是一个小型的图像数据库管理系统。文中选用小型关系数据库 Access 实现图像、纹理样本的一体化存储,其中图像数据存储存储在 BLOB 字段(长二进制对象),详细的表结构如图 3 所示。该模块实现了对数据的浏览、添加、删除、修改等功能。

字段名称	数据类型	说明
image_ID	自动编号	图像序号
width	数字	宽(以像素为单位)
height	数字	高(以像素为单位)
image	OLE 对象	图像数据
type	文本	图像类型(图像/纹理样本)
inDate	日期/时间	入库日期
description	文本	备注

图 3 表结构

图像预处理模块主要负责图像的一些基本预处理功能,如图像灰度直方图均衡化、亮度或对比度调整、颜色模型(RGB,CMYK,HIS,HLS)的转换、颜色位数的调整、彩色图像灰度化。该模块属于一般的图像预处理,是后续图像纹理特征数据挖掘的基础性模块。

基于云模型^[1]的概念跃升模块主要完成从定量的图像数据到定性的纹理亮度概念的不确定性转换,是一种数据预处理方法,属于系统的数据层。首先对原始图像进行样区切割,切割成一些小的子图像样本,并保存。ITDM 模块默认的子样区大小为 30×30,可以通过“设定范围”子菜单进行自定义。获取子样本图像的直方图特征后,再以该特征作为输入,进行不确定性概念粒度转换,可以交互地查看转换后的效果,将云概念用正向云发生器生成云滴,返回给用户一个可视化的结果,如图 4 所示,某图像的原始直方图与 7 个正态云图的拟合效果对比。最后,建立数据库连接,将上述概念特征存入数据集。

其中基于云模型的不确定概念粒度转换实际上是对纹理亮度(灰度值)空间进行划分的过程,是一个从亮度空间的灰度值分布中抽取出现念的过程,是从定量的灰度值表示到定性的亮度概念描述的转换,是一个概念规约的过程,由定量的灰度值数据不断地进行概念粒度提升得到满意的定性的图像灰度概念描述。

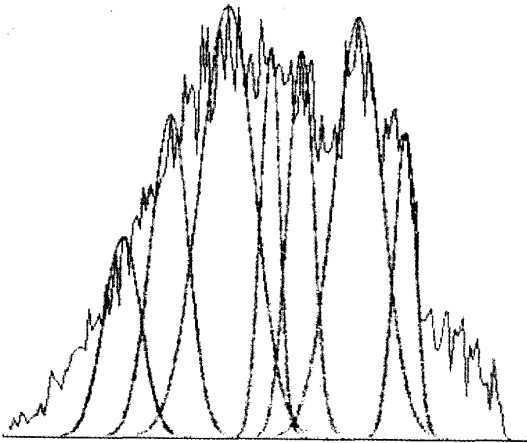


图 4 一个可视化的效果

基于概念格的关联规则挖掘模块属于系统的核心层。可以通过参数设置菜单设置关联规则挖掘的支持度阈值、置信度阈值,规则文件名(文中将挖掘出的纹理特征关联规则存储在文本文件、二进制文件中)等。接下来就是概念格的建立、Hasse 图的生成、规则的生成。

概念格的每一个节点是一个概念,包含内涵和外延两个部分,通过 Hasse 图可以很方便地描述这些概念之间的泛化和例化关系。基于概念格的关联规则挖掘在概念格构建的过程中直接产生频繁封闭项集,有效地避免了大量冗余规则的生成。另外,云模型和概念格都对概念进行形式化处理,所以用概念格表示由上一环节(基于云模型的概念跃升)得到的概念,无疑在两个环节之间建立了一种以概念为主线索的联系,这就为 RSImageMiner 后续研究概念驱动的数据挖掘与知识发现体系提供了便利。

纹理分析模块主要是应用图像纹理特征进行纹理分析,目前只完成了纹理分割,其它子功能与纹理分割模块有类似之处,有待于进一步的开发。

文中以图 5 (a)为例,进行分割实验。



(a) 原始图像 (b) 概念提升结果 (c) 最终分割结果

图 5 图像纹理分割效果

(1)对于两类样本,采用前述方法分别挖掘出纹理特征关联规则,其支持度与置信度就构成了一个模式向量 $R_1 \sim R_3$,如表 1 所示。其中参数设置如下:样本大小为 30×30,邻域大小设置为 3×3,概念个数为 2 个,支持度阈值为 0.35,置信度阈值为 0.85。

表 1 样本纹理特征

规则		R_1	R_2	R_3
样本		$\{1,7\} \Rightarrow \{4\}$	$\{2,8\} \Rightarrow \{5\}$	$\{3,9\} \Rightarrow \{6\}$
	支持度	0.422194	0.419643	0.400510
	置信度	0.887399	0.889189	0.889518
	支持度	0.441327	0.447704	0.437500
	置信度	0.895954	0.897436	0.903790

这里对规则的含义解释如下:例如规则 $R_1: \{1,7\} \Rightarrow \{4\}$, 表示在 3×3 邻域(如图 6 所示)中的 1 号、7 号位置的灰度概念为“亮”, 可以推断出 4 号位置的灰度概念为“亮”, 其中在样本的所有根像元(阴影所示部分)中满足该规则的百分比为 42.2194%, 该推断的置信度为 88.7399%。

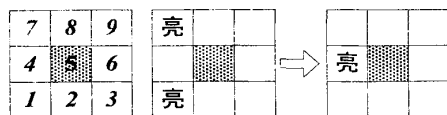


图 6 规则的解释

(2) 将待分割的图像采用同样的方法进行亮度概念提升, 结果如图 5(b) 所示。

(3) 对于图 5(a) 的每个样本大小的区域, 分析其对应每条规则的置信度 S_i 与支持度 C_i , 比较该结果与模式向量的距离 D , 按照极大判定原则对根像元作分割标记; $D = \sqrt{(S_i - S_{ij})^2 + (C_i - C_{ij})^2}$ 。

(4) 按标记对图像进行分割, 结果如图 5(c) 所示。

4 结束语

RSImageMiner 是一个基于(遥感)图像数据挖掘的空间数据挖掘工具, 图像纹理特征数据挖掘模块 ITDM 是其重要模块, 采用概念驱动的方式, 将云模型与概念格结合起来, 对图像纹理概念的粒度进行不确定性提升, 从中提取有价值的语义特征, 指导纹理分析与处理。实验表明, 该软件原型能得到较满意的结果。

但是, 文中实现的模块只是一个软件原型, 需要进一步地修改完善, 使其能更好地融合 RSImageMiner, 使 RSImageMiner 成为一个具有自主知识产权的、实用的空间数据挖掘软件。

参考文献:

- [1] Fayyad U M, Weir N, Djorgovski S G. SKICAT: A Machine Learning System for Automated Cataloging of Large Scale Sky Surveys[C]//International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers, 1993: 112 - 119.
- [2] Zaiane O R. The First International Workshop on Multimedia Data Mining (MDM/KDD 2000)[EB/OL]. 2006 - 10 - 20. http://www.cs.ualberta.ca/~zaiane/mdm_kdd2000/.
- [3] Rushing J A. Image segmentation using association rule features[J]. IEEE Transaction on Image Processing, 2002, 11 (5): 558 - 567.
- [4] Zaiane O R, Han Jiawei. MultiMediaMiner: A system Prototype for MultiMedia Data Mining[C]//Proceedings of ACM SIGMOD International Conference on Management of Data. Seattle: ACM Press, 1998: 581 - 583.
- [5] 秦 昆. 基于形式概念分析的图像数据挖掘研究[D]. 武汉: 武汉大学, 2004.
- [6] Dateu M, Seidel K. An Innovative Concept for Image Information Mining[C]//MDM/KDD 2002: International Workshop on Multimedia Data Mining (with ACM SIGKDD 2002). Edmonton: University of Alberta, 2002: 11 - 18.
- [7] 张 钹, 张 铃. 问题求解理论及应用[M]. 北京: 清华大学出版社, 1990.
- [8] Ganter B, Wille R. Formal Concept Analysis - Mathematical Foundations[M]. Berlin: Springer Verlag, 1999.
- [9] 朱 明. 数据挖掘[M]. 合肥: 中国科技大学出版社, 2002.
- [10] Han Jiawei, Kamber M. 数据挖掘概念与技术[M]. 范 明, 孟小峰等译. 北京: 机械工业出版社, 2001.
- [11] 李德毅, 杜 鹂. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005.

(上接第 4 页)

of the IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP - KE 2005). Beijing: China Scientific Book Service Co. Ltd, 2005: 632 - 640.

- [11] Burger J. Issues, tasks and program structures to roadmap research in Question Answering[C]//In Proceedings of The Ninth Text Retrieval Conference (TERC 2000). TERC - 9. Gaithersburg, Maryland: NIST Special Publication, 2000: 500 - 521.
- [12] Cristianini N, Shawe - Taylor J. An Introduction to Support Vector Machines and Other Kernel - based Learning Meth-

ods[M]. 北京: 机械工业出版社, 2005.

- [13] Vapnik V. The Nature of Statistical Learning Theory[M]. [s.l.]: Springer - Verlag, 1995.
- [14] Amari S, Wu S. Improving support machine classifier by modifying kernel Function[J]. Neural Networks, 1999, 12: 783 - 789.
- [15] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121 - 167.
- [16] 张运良, 张 全. 基于 HNC 理论的语义相关度计算方法[J]. 计算机工程与应用, 2005(34): 1 - 3.