• 博士论文 •

文章编号: 1000-3428(2008)01-0053-03

文献标识码: A

中图分类号: TP391

# 基于形状保持主动轮廓模型长直条的检测

李小毛 1,2 , 王智峰 1,2 , 唐延东 1

(1. 中国科学院沈阳自动化研究所,沈阳 110016; 2. 中国科学院研究生院,北京 100039)

摘 要:通过形状约束方程(组)与一般主动轮廓模型结合,将目标形状与主动轮廓模型融合到统一能量泛函模型中,提出一种形状保持主动轮廓模型。模型通过参数化水平集函数的零水平集表示某一类特定形状,不仅达到了分割即目标的目的,而且能够给出特定目标的定量描述。根据形状保持主动轮廓模型,建立一个用于长直条状目标检测的统一能量泛函模型,导出相应的Euler-Lagrange常微分方程并用水平集方法实现了长直条状区域的检测。此形状保持模型的一种特殊情况可以用于直线状地平(海天)线提取。实验结果表明,该模型不仅能够准确地检测出给定图像中的长直条状区域而且有很强的抗噪、抗变形及遮挡性能。

**关键词:**形状保持; MS 模型; 主动轮廓; 长直条状; 地平线

# Long Straight Bar Detection Based on Shape Preserving Active Contour Model

LI Xiao-mao<sup>1,2</sup>, WANG Zhi-feng<sup>1,2</sup>, TANG Yan-dong<sup>1</sup>

(1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016;

2. Graduate School, Chinese Academy of Sciences, Beijing 100039)

[Abstract] This paper integrates the detected object shape represented by shape restraint equation and general active contour model into an unified energy functional generating the shape preserving active contour model. In this model, a specific class contour shape is represented as the zero level line of some certain parametric level set function. This model can not only detect the given object correctly but also characterize the object shape quantificationally via these parameters. In addition, it builds an energy functional for long straight bar detection using the proposed model. It deduces the corresponding Euler-Lagrange ODE(s) of its shape parameters and implements them using level sets method. It also deduces that a special case of the long straight bar detection model can be used for the linear skyline (sea level) detection. Numerical experiments present this model not only can detect the long straight bar correctly but also is robust to noise, deformity and occlusion.

[Key words] shape preserving; Mumford-Shah model; active contour; long straight bar; skyline

# 1 概述

近年来国内外有很多人在基于先验形状约束的图像分割 方面做了大量的研究工作。如何将更多高层先验信息,如特 定形状目标融合于一个统一分割框架中,准确地分割出所期 望的目标是当前主动轮廓模型研究的热点和难点。1995年, COOTES 等人基于 Snake 方法[1]提出了 ASM 模型[2], 将物体 的形状变化分为刚性变化和弹性变化。刚性变化是通过姿态 空间的参数描述,而弹性变化是通过统计的方法将物体的弹 性变化限制在一定的弹性空间。2000 年 Leventon 等人[3]基于 统计模型和水平集方法[4],将目标的先验形状知识与测地线 动轮廓模型<sup>[5]</sup>结合起来。2002 年 Chen 等人将目标的先验形 状信息与测地主动轮廓模型融合到统一的变分能量模型 中[6],并确定了能量函数极小值解的存在性。2001年山世光 等人提出了一种基于特征模板参数形状约束的能量函数项用 于面部特征提取[7]。2003年上海交大的黄福珍等人将人脸先 验形状椭圆约束与 CV<sup>[8]</sup>模型结合用于人脸分割<sup>[9]</sup>。2005 年肖 秀春和邹北骥根据眼睛和嘴巴的形状特征,基于 Snake 方法, 提出了一种基于抛物线形状约束能量项用于人脸特征提 取[10]。2002 年, Cremers 等人在文献[11]中将先验形状统计 模型与 MS 模型结合在统一的变分框架中。 随后,在文献[12] 中将符号距离函数表示的先验形状与 CV<sup>[8]</sup>模型结合 ,并引入 动态标记函数的概念用以标记形状约束起作用的区域。 Cremers 在文献[13]中引入多个动态标记函数 ,并且在形状约束能量项中加入了姿态参数。 Tony 与 Zhu 在文献[14]中也做了类似的工作。不过他们变分能量模型中的形状约束项是关于水平集函数表征的先验形状对应的 Heavisied 函数的积分。 2006年 Hassan 在文献[15]中提出了一种用水平集表示形状的新方法并与 CV 模型结合。这种新的形状表示方法的优势就是能够更好地控制形状训练阶段的校准过程而且能够有效地推广到多维空间。

以上这些模型都是基于形状约束的主动轮廓模型。本文基于一般主动轮廓模型与形状约束方程(组),将曲线的显式与隐式表示方法优点结合在一起,提出了一种形状保持主动轮廓模型。此模型的思想与 Han 提出的拓扑保持主动轮廓模型思想很相似<sup>[16]</sup>。在已知待分割目标先验形状情况下,将主动轮廓模型能量泛函用一类零水平集具有固定形状的水平集

基金项目:中国科学院创新基金资助项目(CX01-04-02)

作者简介: 李小毛(1981 - ), 男, 博士研究生, 主研方向: 基于变分

方法的图像处理;王智峰,博士研究生;唐延东,研究员 **收稿日期:**2007-02-05 **E-mail:**lixiaomao@sia.cn 函数表示。形状保持主动轮廓模型还具有如下一些特点:形状描述简单、方便;无须正则项因子<sup>[14]</sup>;水平集函数由方程(组)中的参数唯一确定无须重复初始化;由极小能量泛函推导出的 Euler-Lagrange 方程是关于目标形状参数的常微分方程。

# 2 形状保持主动轮廓模型

设一个目标  $O \subset \mathbb{R}^2$  闭合有界 则目标 O 与一种特殊的水平集函数,即式(1)中的符号距离函数存在一一对应关系 $^{[14]}$ 。

$$\begin{aligned} |\nabla \phi| &= 1 \\ \phi &= 0 & X \in O \setminus \partial O \\ &= 0 & X \in \partial O \\ &< 0 & X \in R^2 \setminus O \end{aligned}$$
 (1)

形状保持主动轮廓模型在本质上是在已知待分割目标形状先验知识下,将主动轮廓曲线约束为具有固定形状的曲线,从而实现分割即目标的目的。形状约束的主动轮廓模型写为更一般的如下形式:

$$\min_{\phi} \int_{\Omega} F(\phi, \nabla \phi|, u, \nabla u) dx$$

$$subject to \underbrace{\left|\phi - T_s(\phi_0)\right|}_{\text{形状约束}} \varepsilon$$
(2)

其中,F 为一般主动轮廓能量泛函模型; $\phi$  表示水平集函数;u 为原图像; $\phi$ 。表示初始轮廓函数并具有待分割目标的先验形状特征; $T_s$  为同一类形状变换算子,其变换参数由约束条件和求取泛函的极小值自适应确定; $\varepsilon$  为形状变化的容许误差,当 $\varepsilon=0$  时就变成形状保持主动轮廓模型。根据不同形状约束条件,可以建立新的算法模型。

# 3 长直条状区域检测模型

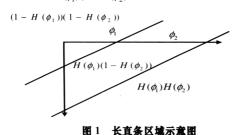
CV 模型本质上是一种基于水平集方法简化的 MS 模型 $^{[17]}$ ,它将图像分割成两块分段常数区域。假设  $\Omega$  为  $R^2$  空间的一个有界开子集, $\omega$  为  $\Omega$  中的一个有界开子集,其边界表示为  $\partial \omega$  。对于一幅给定的图像 u 定义为  $u:\Omega \to R$  。假设  $\phi:\Omega \to R$  为水平集函数,其零水平集为  $\partial \omega$  。H 为阶跃函数,则 CV 模型可以表示为

$$\begin{split} E(c_1,c_2,\phi) &= \mu \int_{\Omega} \left| \nabla H\left(\phi(x,y)\right) \right| \mathrm{d}x \mathrm{d}y + \nu \int_{\Omega} H\left(\phi(x,y)\right) \mathrm{d}x \mathrm{d}y + \\ \lambda_1 \int_{\Omega} \left| u - c_1 \right|^2 H\left(\phi(x,y)\right) \mathrm{d}x \mathrm{d}y + \\ \lambda_2 \int_{\Omega} \left| u - c_2 \right|^2 (1 - H\left(\phi(x,y)\right)) \mathrm{d}x \mathrm{d}y \end{split} \tag{3}$$

其中, $\mu$ , $\nu$ , $\lambda$ 1 和  $\lambda$ 2 为非负常数; $c_1$  和  $c_2$  为区域  $\omega$  和  $\Omega \setminus \varpi$  的灰度平均值;第 1 项代表  $\partial \omega$  的长度即正则项;第 2 项为区域的面积项;最后两项表示分块常数函数与原图像的逼近程度;统计特征量  $c_1$  和  $c_2$  可以用其它特征量来代替,如曲率、方向及分形维数等。

设  $\phi_1$  和  $\phi_2$  分别为式(5)、式(6)表示的水平集函数,其零水平集分别表示两条相互平行的直线。这两个水平集函数的零水平集将整个图像区域分为

 $(1-H(\phi_1))(1-H(\phi_2)), H(\phi_1)(1-H(\phi_2), H(\phi_1)H(\phi_2)$  3 部分。其中, $H(\phi_1)(1-H(\phi_2))$  为长直条状区域,如图 1 所示。



因此,长直条状区域检测的能量泛函模型为

$$\begin{split} E(c_1,c_2,c_3,\phi_1,\phi_2) &= \lambda_1 \int\limits_{\Omega} \left| u - c_1 \right|^2 (1 - H(\phi_1(x,y))) (1 - H(\phi_2(x,y))) \mathrm{d}x \mathrm{d}y + \\ \lambda_2 \int\limits_{\Omega} \left| u - c_2 \right|^2 H(\phi_1(x,y)) (1 - H(\phi_2(x,y))) \mathrm{d}x \mathrm{d}y + \\ \lambda_3 \int\limits_{\Omega} \left| u - c_3 \right|^2 H(\phi_1(x,y)) H(\phi_2(x,y)) \mathrm{d}x \mathrm{d}y \end{split}$$
 (4)

Subject to:

$$\phi_1 = x \cos \theta + y \sin \theta - \rho_1 \tag{5}$$

$$\phi_2 = x \cos \theta + y \sin \theta - \rho_2 \tag{6}$$

其中, $\rho_2 > \rho_1$ , $\theta \in [0^\circ, 180^\circ)$ 。 u 为原图像, $\lambda_1$ , $\lambda_2$  和  $\lambda_3$  为非负常数。首先固定  $\phi_1$  和  $\phi_2$  ,分别关于  $c_1$ , $c_2$  和  $c_3$  极小化式 (4),得到它们关于水平集函数的表达式如下:

$$c_1 = \frac{\int u(1 - H(\phi_1))(1 - H(\phi_2)) dx dy}{\int (1 - H(\phi_1))(1 - H(\phi_2)) dx dy}$$
(7)

$$c_2 = \frac{\int_{\Omega} u H(\phi_1) (1 - H(\phi_2)) dx dy}{\int_{\Omega} H(\phi_1) (1 - H(\phi_2)) dx dy}$$
(8)

$$c_{3} = \frac{\int_{\Omega} uH(\phi_{1})H(\phi_{2})dxdy}{\int_{\Omega} H(\phi_{1})H(\phi_{2})dxdy}$$
(9)

实际上  $c_1$  ,  $c_2$  和  $c_3$  分别表示图像在  $(1-H(\phi_1))(1-H(\phi_2))$ ,  $H(\phi_1)(1-H(\phi_2))$ ,  $H(\phi_1)(1-H(\phi_2))$ ,  $H(\phi_1)(1-H(\phi_2))$  个区域的灰度平均值。为了得到未知形状参数  $\rho_1$  ,  $\rho_2$  和  $\theta$  相应的 Euler-Lagrange 方程 ,用文献[8]中的 H 和  $\delta$  的正则项形式  $H_\varepsilon$  和  $\delta_\varepsilon$  来代替 H 和  $\delta$  。当  $\varepsilon \to 0$  ,  $H_\varepsilon \to H$  ,  $\delta_\varepsilon = H_\varepsilon'$  。引入虚拟时间变量 t ,得到关于形状参数的常微分发展方程分别如下:

$$\frac{\mathrm{d}\rho_1}{\mathrm{d}t} = \int_{\Omega} \left[ -\lambda_1 (u - c_1)^2 (1 - H_{\varepsilon}(\phi_2)) + \lambda_2 (u - c_2)^2 (1 - H_{\varepsilon}(\phi_2)) + \lambda_3 (u - c_3)^2 H_{\varepsilon}(\phi_2) \right] \delta_{\varepsilon}(\phi_1) \mathrm{d}x \mathrm{d}y \tag{10}$$

$$\begin{split} \frac{\mathrm{d}\rho_2}{\mathrm{d}t} &= \int_{\Omega} \left[ -\lambda_1 (u - c_1)^2 (1 - H_{\varepsilon}(\phi_1)) - \right. \\ \left. \lambda_2 (u - c_2)^2 H_{\varepsilon}(\phi_1) + \lambda_3 (u - c_3)^2 H_{\varepsilon}(\phi_1) \right] \delta_{\varepsilon}(\phi_2) \mathrm{d}x \mathrm{d}y \end{split} \tag{11}$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \int_{\Omega} [\lambda_1(u - c_1)^2 (\delta_{\varepsilon}(\phi_1)(1 - H_{\varepsilon}(\phi_2)) + (1 - H_{\varepsilon}(\phi_1))\delta_{\varepsilon}(\phi_2)) + \\
\lambda_2(u - c_2)^2 (-\delta_{\varepsilon}(\phi_1)(1 - H_{\varepsilon}(\phi_2)) + H_{\varepsilon}(\phi_1)\delta_{\varepsilon}(\phi_2)) + \\
\lambda_3(u - c_3)^2 (-\delta_{\varepsilon}(\phi_1)H_{\varepsilon}(\phi_2) - \\
H_{\varepsilon}(\phi_1)\delta_{\varepsilon}(\phi_2))](y \cos\theta - x \sin\theta) \mathrm{d}x \mathrm{d}y$$
(12)

因为最后导出的方程是关于参数演化的常微分方程,所以在数值算法实现的时候无须重新初始化水平集函数,而且最后的结果是关于目标形状的定量描述。根据式(5)和式(6)的形状保持模型,再来考虑另外两种情况。当两条直线不平行时,在式(5)和式(6)中的角度  $\theta$  分别用  $\theta_1$  和  $\theta_2$  代替,相应的形状约束方程组如下:

$$\phi_1 = x \cos \theta_1 + y \sin \theta_1 - \rho_1 \tag{13}$$

$$\phi_2 = x \cos \theta_2 + y \sin \theta_2 - \rho_2 \tag{14}$$

 $ho_1$ 和  $ho_2$ 的常微分发展方程与式(10)和式(11)相同,而关于角度  $heta_1$ 和  $heta_2$ 的演化方程如下:

$$\frac{\mathrm{d}\theta_{\mathrm{i}}}{\mathrm{d}t} = \int_{\Omega} \left[ \lambda_{\mathrm{i}} (u - c_{\mathrm{i}})^{2} (1 - H_{\varepsilon}(\phi_{2})) - \lambda_{\mathrm{2}} (u - c_{\mathrm{2}})^{2} (1 - H_{\varepsilon}(\phi_{2})) - \lambda_{\mathrm{3}} (u - c_{\mathrm{3}})^{2} H_{\varepsilon}(\phi_{\mathrm{2}}) \right] \delta_{\varepsilon}(\phi_{\mathrm{i}}) (y \cos \theta_{\mathrm{i}} - x \sin \theta_{\mathrm{i}}) \mathrm{d}x \mathrm{d}y \tag{15}$$

$$\frac{\mathrm{d}\theta_{2}}{\mathrm{d}t} = \int_{\Omega} \left[ \lambda_{1} (u - c_{1})^{2} (1 - H_{\varepsilon}(\phi_{1})) + \lambda_{2} (u - c_{2})^{2} H_{\varepsilon}(\phi_{1}) - \lambda_{3} (u - c_{3})^{2} H_{\varepsilon}(\phi_{1}) \right] \delta_{\varepsilon}(\phi_{2}) (y \cos \theta_{2} - x \sin \theta_{2}) dx dy \tag{16}$$

第 2 种情况是当  $\phi_1$  和  $\phi_2$  始终相同时 ,用下面的方程表示目标形状。

$$\phi = x\cos\theta + y\sin\theta - \rho \tag{17}$$

得到直线形状保持的能量函数模型:

$$E(c_1, c_2, \phi) = \lambda_1 \int_{\Omega} |u - c_1|^2 (1 - H(\phi)) dx dy + \lambda_2 \int_{\Omega} |u - c_2|^2 (1 - H(\phi)) dx dy$$
(18)

关于形状参数  $\rho$  和  $\theta$  的演化方程如下:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \int_{\Omega} \left[ -\lambda_1 (u - c_1)^2 + \lambda_2 (u - c_2)^2 \right] \delta_{\varepsilon}(\phi) \mathrm{d}x \mathrm{d}y \tag{19}$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\int_{\Omega} \left[ -\lambda_1 (u - c_1)^2 + \lambda_2 (u - c_2)^2 \right]$$

$$\delta_2(\phi) (y \cos \theta - x \sin \theta) \mathrm{d}x \mathrm{d}y \tag{20}$$

这个模型不仅可以用于直线状海天线的检测,而且还可用 于直线状目标将图像分成两块具有明显不同特征区域的检测。

为简单起见,这里只给出式(19)和式(20)的数值差分求 解格式,其它公式以此类推。由于是常微分方程,可以用显 式数值差分格式求解。设h和 $\Delta t$ 分别表示空域和时域步长,  $(x_i, y_i) = (ih, jh)$  为离散点,则

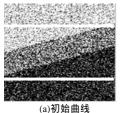
 $\rho^n = \rho(n\Delta t), \theta^n = \theta(n\Delta t), \phi^n_{i,j} = \phi(n\Delta t, x_i, y_j) \quad 1 \quad i \quad M,$ 1 j N。则式(19)与式(20)的离散化方程如下:

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} = \sum_{i=1}^{M} \sum_{j=1}^{N} \left[ -\lambda_1 (u_{i,j} - c_1(\rho^n, \theta^n))^2 + \lambda_2 (u_{i,j} - c_2(\rho^n, \theta^n))^2 \right] \delta_{\varepsilon}(\rho^n, \theta^n)$$
(21)

$$\frac{\theta^{n+1} - \theta^n}{\Delta t} = -\sum_{i=1}^M \sum_{j=1}^N \left[ -\lambda_1 (u_{i,j} - c_1(\rho^n, \theta^n))^2 + \lambda_2 (u_{i,j} - c_2(\rho^n, \theta^n))^2 \right] 
\delta_{\varepsilon}(\rho^n, \theta^n) (j \cos \theta^n - i \sin \theta^n)$$
(22)

# 4 试验结果

下面试验结果图像的灰度值范围为[0,1],参数的选择如 下:空间域的离散步长 h=1;参数  $\rho$  的时间步长  $\Delta t=0.1$ ; 参数  $\theta$  的时间步长  $\Delta t = 0.05$ ;  $\lambda_1 = 1$ ,  $\lambda_2 = 1.1$ ,  $\lambda_3 = 1$ .



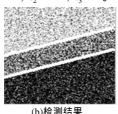
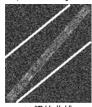


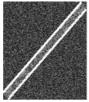
图 2 长直条区域示例图

因为图 2(a)最上面和最下面的区域的灰度值相差很大。 若将这两部分归为一个类不合理而且也得不到准确的检测结 果 ,所以有必要将  $H(\phi_1)(1-H(\phi_2))$  (即长直条状目标)之外的区 域分成两块不同类别,用统计量 $c_1$ 和 $c_3$ 表征。初始形状参数:  $\rho_1 = 30, \, \rho_2 = 100, \, \theta = 90$ ; 最后结果:  $\rho_1 = 61.8, \, \rho_2 = 89.4, \, \theta =$ 74.3

第2个试验是对一个机场跑道的俯视图检测(见图3)。初 始形状参数: $\rho_1 = 80$ ,  $\rho_2 = 160$ ,  $\theta = 50$ ; 最后结果: $\rho_1 = 116.5$ ,  $\rho_2$  $=132.1, \theta =40.3$ 







(a)初始曲线

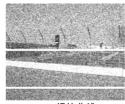
(b)CV 模型检测结果

(c)本文模型结果

图 3 机场跑道的俯视图检测

从图 3 中可以看出,机场部分已经模糊不清,而且还有 很多地方被遮挡了。图 3(b)是基于 CV 模型的检测结果。从

此结果可以看出,由于 CV 模型中缺少目标的先验形状知识, 因此最后的结果是很多零碎的块。然而形状保持的主动轮廓 模型不仅能够准确地检测出目标而且能够给出目标的定量描 述(图 3(c))。还给出了另外一个试验结果,用于平行长直条状 区域的检测(见图 4)。初始形状参数:  $\rho_1 = 10$ ,  $\rho_2 = 180$ ,  $\theta = 90$ ; 最后结果:  $\rho_1 = 120.3$ ,  $\rho_2 = 135.3$ ,  $\theta = 93.8$ 。



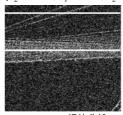


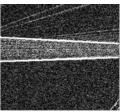
(a)初始曲线

(b)检测结果

图 4 平行长直条状区域的检测

两条不平行直线的长直条状区域检测结果见图 5。初始 形状参数:  $\rho_1 = 20$ ,  $\rho_2 = 110$ ,  $\theta_1 = 90$ ,  $\theta_2 = 90$ ; 最后结果:  $\rho_1 =$ 82.0,  $\rho_2 = 139.0$ ,  $\theta_1 = 93.1$ ,  $\theta_2 = 87.7$ 





(a)初始曲线

(b)检测结果

图 5 两条不平行直线的长直条状区域检测

关于直线状地平线一个检测结果见图 6。初始形状参数:  $\rho = 150, \theta = 45$ ; 最后结果:  $\rho = 93.2, \theta = 91.1$ 。





(a)初始曲线

(b)检测结果

图 6 直线状地平线的检测

从以上实验结果可看出,基于形状保持的主动轮廓模型 无须前处理及后处理过程,将分割与识别过程融合为一体, 达到分割即目标,同时给出目标定量描述。

#### 5 结束语

本文提出了一种用于特定目标形状检测的形状保持主 动轮廓模型。此模型最后导出的发展微分方程是关于目标形 状参数的常微分方程。因为形状保持主动轮廓模型中特定目 标形状是通过参数化水平集函数的零水平集表示,所以轮廓 在演化过程中完全不会自交,而且能够通过参数控制及描述 特定目标形状,将曲线的显式(定量描述)与隐式(曲线不会自 交)表示的优点融合在一起。根据这种模型建立了长直条状区 域及直线状目标检测能量函数模型,将目标的分割与识别统 一于一个完整的能量泛函模型中,达到了分割即目标,同时 给出目标定量描述,这是基于形状约束的主动轮廓模型不具 有的。形状保持主动轮廓模型的局限性,在于最后的检测结 果和初始轮廓有一定关系。形状保持主动轮廓模型中的约束 条件也可以用其他形状约束方程(组)代替,达到关于其他特 定目标的检测目的。

(下转第58页)

**—**55—

#### 3.4 模式评估

对挖掘结果进行验证和评价,这是一个反复实验的过程,运用独立于训练集的测试集进行验证,直到得到满意的结果为止。

#### 3.5 其他问题

由于部分测试工具具有一定的针对性,为了有效地利用 这些工具,要适当地编写一些应用接口将测试结果转换为可 存储的形式以降低操作的复杂度。

## 4 模型应用

将该模型应用到数据库管理系统的分类中,测试时记录如下信息作为输入属性:所支持的最大用户数,所支持的存储数量级(GB级,TB级),是否支持多CPU、运行时的平均内存量,平均CPU占用量等,得到3个类别:个人版(PsnEd),标准版(StdEd)和企业安全版(EnterSecEd)等。

将测试数据经过集成后生成 XML 文档,其中,性能、功能中的各字段作为分类属性,所有属性在处理过程中经过预排序生成属性表。在决策树生长过程中,采用预先设置的阈值防止决策树的过度生长:当叶结点的样本数量小于阈值后,停止分裂该结点,并用多数表决法给该结点作类标记。生成的决策树以线性方式存储到数据库中,因为 SLIQ 算法生成的是二叉树,所以只要对结点进行编号,就可以知道一个结点的父结点,规定左子结点的编号为 2×(父结点编号),右子结点的编号为 2×(父结点编号)+1。为了方便使用,在根结点和中间结点中记录左、右子结点的编号。

决策树建好后,可以很方便地转换为分类规则。下面是 挖掘得到的3条简单的规则:

规则 1 IF(Num of Users ≤5)THEN(class=PsnEd)

规则 2 IF(5<Num of Users≤24 AND Order of Mag =GB

AND MultiCpu=0)THEN(class=StdEd)

规则 3 IF(24< Num of Users ≤1000 AND Order of Mag=GB AND MultiCpu=1)THEN(class=EnterSecEd)

根据这些规则,可以对测试数据集进行分类,然后与实际的类型进行比较。如果分类错误率在可接受的范围内,就 将这些规则存储到规则库中。

# 5 结束语

本文对将决策树分类方法引入到软件的细粒度分类中进行了研究,并在数据库管理系统的分类中应用,验证了方案的可行性。主要选取的是 SLIQ 分类算法以及 MDL 剪枝算法,关于分类和剪枝算法的改进工作近年来均有很大进展。下一步工作将针对其他决策树算法如 SPRINT 算法<sup>[5]</sup>进行探讨,此外,在其他领域的软件的细粒度分类中运用该模型也是一个主要方面。

# 参考文献

- [1] Han Jiawei, Kamber M. 数据挖掘: 概念与技术[M]. 范 明, 孟 小峰, 译. 北京: 机械工业出版社, 2001.
- [2] Mehta M, Agrawal R, Rissancn J. SLIQ: A Fast Scalable Classifier for Data Mining[C]//Proc. of the 5th Int'l Conf. on Extending Database Technology. Avignon, France: [s. n.], 1996.
- [3] Clinkenbeard R A, Xin Feng. An Unsupervised Learning and Fuzzy Approach for Software Category Identification and Capacity Planning[J]. IEEE Neural Networks, 1992, 3(7): 358.
- [4] 郭志懋, 周傲英. 数据质量和数据清理研究综述[J]. 软件学报, 2002, 23(11): 76.
- [5] Shafer J, Agrawal R, Mehta M. SPRINT: A Scalable Parallel Classifier for Data Mining[C]//Proc of the 22nd VLDB Conference. Mumbai, India: [s. n.], 1996: 544-555.

(上接第55页)

## 参考文献

- Kass M, Witkin A, Terzopoulos D. Snakes: Active Contour models[J]. International Journal of Computer Vision, 1987, 1(4): 321-331.
- [2] Cootes T, Cooper D, Taylor C, et al. Active Shape Models——Their Training and Application[J]. Computer Vision and Image Understanding, 1995, 61(1): 38-59.
- [3] Leventon M E, Grimson W E L, Faugeras O. Statistical Shape Influence in Geodesic Active Contours[C]//Proceedings of the IEEE CVPR, Hilton Head Island. South Carolina, USA: [s.n.], 1988: 316-323.
- [4] Osher S, Sethian J A. Fronts Propagating with Curvature Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations[J]. Journal of Computational Physics, 1988, 79(1): 12-49.
- [5] Casselles V, Catte F, Coll T, et al. A Geometric Model for Active Contours in Image Processing[J]. Numerische Mathematik, 1993, 66(1): 1-31.
- [6] Chen Y, Tagare H D, Thiruvenkadam S, et al. Using Prior Shapes in Geometric Active Contours in a Variational Framework[J]. International Journal of Computer Vision, 2002, 50(3): 315-328.
- [7] 山世光, 高 文. 基于纹理分布和变形模板的面部特征提取[J]. 软件学报, 2001, 12(4): 570-577.
- [8] Chan T, Vese L A. Active Contours Without Edges[J]. IEEE Transactions on Image Processing, 2001, 10(2): 266-277.
- [9] 黄福珍, 苏剑波, 席裕庚. 基于几何活动轮廓模型的人脸轮廓提取方法[J]. 中国图象图形学报, 2003, 8(5): 546-550.

- [10] 肖秀春, 邹北骥. 一种复杂背景下的人脸特征提取算法[J]. 湛 江海洋大学学报, 2005, 25(4): 72-75.
- [11] Cremers D, Tischhauser F, Weickert J, et al. Diffusion Snakes: Introducing Statistical Shape Knowledge into the Mumford-Shah Functional[J]. International Journal of Computer Vision, 2002, 50(3): 295-313.
- [12] Cremers D, Sochen N, Schnorr C. Towards Recognition-based Variational Segmentation Using Shape Priors and Dynamic Labeling[C]//Proc. of Conf. on Scale Space Theories in Computer Vision. [S. 1.]: IEEE Press, 2003: 388-400.
- [13] Cremers D, Sochen N. Multiphase Dynamic LabelingModel for Variational Recognition-driven Image Segmentation[C]//Proc. of European Conf. on Computer Vision. [S. 1.]: IEEE Press, 2004: 74.
- [14] Chan T, Zhu W. Level Set Based Shape Prior Segmentation[R]. Los Angeles: University of California, Technical Report: 03-66, 2003.
- [15] Hassan H, Farag A A. A Shape-based Segmentation Approach: An Improved Technique Using Level Sets[C]//Proc. of 10th the IEEE International Conference on Computer Vision. [S. 1.]: IEEE Press, 2005: 930-935.
- [16] Han X, Xu C Y, Prince J L. A Topology Preserving Level Set Method for Geometric Deformable Models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(6): 755-768.
- [17] Mumford D, Shah J. Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problem[J]. Communications on Pure and Applied Mathematics, 1989, 42(4): 577-685.