

肺 CT 图像的血管骨架化方法

陈 刚 吕 焯 王志成 陈宇飞

(同济大学 CAD 研究中心 上海 200092) (企业数字化技术教育部工程研究中心 上海 200092)

摘 要 在图像几何分析以及其他领域中,骨架化方法应用非常广泛。利用骨架表示图像可以保留图中的拓扑结构,减少冗余信息。在临床实践和治疗肺部疾病的过程中,如何有效地描绘并分析肺部血管骨架结构,对于计算机辅助诊断与辅助手术治疗是非常重要的。提出一种肺 CT 图像的血管骨架化方法:首先使用区域生长法将肺中的血管从胸腔其他组织中分离出来,接着使用数学形态学操作对分割结果进行处理,最后使用三维细化算法对血管进行骨架化。实验结果表明,该方法能快速、有效地实现肺部血管的骨架提取。

关键词 肺 CT 图像,区域生长,形态学运算,骨架化

中图分类号 TP391 **文献标识码** A

Vessel Skeletonization Method for Lung CT Images

CHEN Gang LV Xuan WANG Zhi-cheng CHEN Yu-fei

(CAD Research Center, Tongji University, Shanghai 200092, China)

(Engineering Research Center for Enterprise Digital Technology, Ministry of Education, Shanghai 200092, China)

Abstract The skeletonization method has been widely used in image analysis and other areas. Using skeleton to express image can preserve the topological structure and reduce redundant information. In the clinical practice of diagnosis and treatment of lung disease, how to effectively represent and analyze the vascular structure is very important for computer aided diagnosis and surgery. In this paper, a vessel skeletonization method for lung CT image was proposed. Firstly, vessels were segmented from the thoracic tissues by region growing algorithm. Secondly, the morphological operators were used to deal with the vessel segmentation results. Finally, skeleton of blood vessels was obtained by three-dimensional thinning algorithm. Experimental results show that the proposed method can accurately and efficiently extract vessel skeletons from lung CT image.

Keywords Lung CT image, Region growing, Morphology, Skeletonization

1 引言

肺癌是严重威胁人类生命与健康的最常见的恶性肿瘤之一,对肺癌的早期诊断和治疗是提高其治愈率的主要方法^[1]。CT 扫描一直以来都是诊断肺癌的重要手段,但 CT 图像不仅数量多、分析时间长,而且医生根据 CT 图像容易将肺癌误诊为假阳性。近些年随着计算机辅助诊断技术在临床上的不断使用,肺癌的识别和分析成为医学图像处理的重要课题,而准确地提取肺部血管的拓扑结构是其中一个重要的步骤^[2]。

骨架化理论和方法首先由 H. Blum 提出,其目的是降低冗余信息之后,能使用骨架来显示原图像,而且能保留原图像中目标物体的拓扑结构和形状特征。利用骨架表示原始图像,可以在保持图像重要拓扑特征的前提下,减少图像中的冗余信息^[3]。目前图像骨架提取方法主要分为以下 3 类:距离变换、V 氏图和细化。基于距离变换的方法首先将原始图像转换为特征和非特征元素,特征元素属于目标边界,接着针对

每个元素计算其与最近特征元素之间的近似欧氏距离,找到距离变化后图像中的全部局部极大值点,最后将这些局部极大值点连起来作为骨架化结果^[4]。该方法对噪声比较敏感,一般不能保证骨架的连通性和完整性,需要一些后续的处理方法。基于 V 氏图的算法是通过不断更新计算 V 氏图而最终得到骨架结果的,通过若干生成点的离散集将图像空间分割为若干单元,每个单元精确包含一个生成点和其所有最近像素点,当生成点的密度趋于无穷时,对应的 V 氏图就收敛为骨架^[5],该方法运行时间长,尤其不适用于处理大数据量的 CT 图像,而且在目标物体的弱连接处,不能保证骨架的连通性。基于细化的算法通过不断地从目标物体表面按顺序依次删除满足条件的点,直到没有点能够被删除为止,来获得骨架结果,该方法不会改变目标物体的拓扑结构^[6]。

上述 3 种方法有各自的适用范围,处理不同图像时应针对性地采用适当的方法才能产生良好的骨架化结果。由于基于细化的方法原理简单、容易实现,且处理速度快,能够保留

到稿日期:2012-08-22 返修日期:2012-11-25 本文受国家自然科学基金(61103070),中央高校基本科研业务费(0800219171),国家科技支撑计划课题(2012BAF10B12)资助。

陈 刚(1981—),男,博士生,主要研究方向为图像处理、计算机视觉,E-mail: HighFreeBird@gmail.com;吕 焯(1982—),男,博士生,主要研究方向为数字图像处理、计算机视觉。

骨架的拓扑结构和连通性,保证骨架的中间位置,因此本文采用该方法来提取肺部血管骨架。

本文第2节概述关于骨架的基本概念;第3节介绍如何使用区域生长法将肺部血管从胸腔CT图像中分割出来;第4节叙述基于细化的三维肺部血管骨架化方法;第5节给出实验结果和分析;最后给出结论。

2 基本概念介绍

骨架是指图像中目标物体的骨骼部分,是一种非常重要的图像几何特征。提取骨架的过程也就是对图像中的物体不断地进行细化的过程。细化的基本原理是按照顺序逐步剔除满足删除条件的点(这种点称之为简化点),直到没有可删除点为止^[6],但在三维数据空间内进行细化要比二维数据空间内困难得多,有3个主要原因:

- 1) 骨架应为单像素/体素宽^[7];
- 2) 骨架线要处于目标物体的中轴,并且要与原物体类似^[8];
- 3) 骨架应和原始图像拓扑等价,即具有相同数量的前景目标、背景目标和孔洞,目标物体不能被拆分或删除,空腔不能与背景或其他空腔合并。

三维空间中的体素分为两类:目标体素(像素值为1,也称黑点)、背景体素(像素值为0,也称白点)^[9]。由于在三维细化过程中保持连通性不变是一个至关重要的要求,因此需要构建一个 $3 \times 3 \times 3$ 的格子^[10],来检测目标体素的局部连通性。设点 $p(x, y, z)$ 为三维 Z^3 空间中的任意一点,如图1中的黑色体素,那么它的6邻域,如图1(b)中的红色体素所示,定义为:

$$N_6(p) = \{p(u, v, w) \in Z^3 \mid |x-u| + |y-v| + |z-w| \leq 1\} \quad (1)$$

它的26邻域,如图1(a)中的白色体素所示,定义为:

$$N_{26}(p) = \{p(u, v, w) \in Z^3 \mid \text{Max}[|x-u| + |y-v| + |z-w|] \leq 1\} \quad (2)$$

它的18邻域,如图1(c)中的蓝色体素所示,定义为:

$$N_{18}(p) = \{p(u, v, w) \in Z^3 \mid |x-u| + |y-v| + |z-w| \leq 2\} \cap N_{26}(p) \quad (3)$$

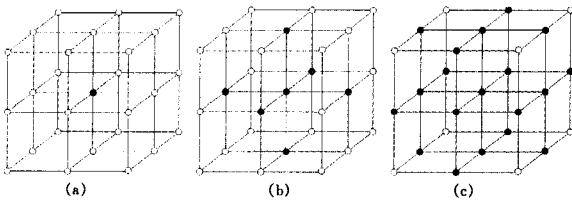


图1 体素的邻域系统表示

3 肺部血管的分割

胸腔CT图像中不仅只有肺部区域,还有整个胸腔、肌肉、气管、肋骨和胸腔外壁组织等(如图2(a)、(c)所示)。本文使用三维区域生长算法去除胸腔CT图像中肺区外的上述物体,将肺部血管分割出来。区域生长法是一种半自动的方法,它需要预先指定若干生长点和判断准则,根据图像中血管所占领区域内的性质,如灰度、纹理、色彩等,把血管所占领的区域寻找出来^[11]。

由一组设定的种子点开始,根据预先设置的生长准则把体素或子区域集合成较大区域,也就是将与种子点相似的邻近体素归纳到该种子的生长区域内,具体步骤为:

1) 由于肺部血管在灰度图像中为高密度,采用非最大值抑制法与人工标记共同产生的特征点作为种子点。

2) 对全部肺CT图像进行分析,根据经验确定全局阈值。一般情况下,肺部血管在CT图像中的灰度取值范围是 $[-1024, -500]$,判断体素与种子点之间的灰度差,决定体素是否进入生长区域:如果灰度差小于或者等于全局阈值,则认为该体素与种子点相似,归为该种子点的生长区域;否则,就不将它归入生长区域。

通过上述方法得到的肺部血管分割结果如图2所示,图2(a)为原始胸腔CT图像某切片的横断面图,图2(b)为肺部血管分割结果的三维重建图,图2(c)为胸腔与肺部血管的体重建结果,图2(d)为胸腔与血管某切片的矢状面重建结果。

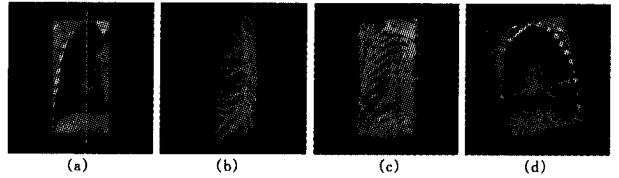


图2 胸腔CT肺部血管分割结果

4 肺部血管的骨架化

4.1 基于数学形态学的血管图像预处理

由于成像伪影、噪声数据集、有效空间的分辨率以及图像分割算法等因素的影响,分割出来的血管并不是光滑、准确的^[12],如果不对分割结果进行预处理,这些异常情况会极大地影响到血管骨架提取的准确性和有效性。在进行骨架化之前必须对血管图像进行预处理,本文采用数学形态学方法来消除血管分割结果中异常情况的影响。

数学形态学是用一组形态学的代数运算子度量、提取图像中的对应形态,从而实现对图像的分析 and 识别^[13]。最基本的形态学运算子有:腐蚀、膨胀、开和闭,用这些算子及其组合来进行图像形状和结构的分析及处理。其中图像 f 被结构元素 b 的开运算记为 $f \circ b$,表示 f 被 b 腐蚀后再用 b 来膨胀的结果:

$$f \circ b = (f \ominus b) \oplus b \quad (4)$$

图像 f 被结构元素 b 的闭运算记为 $f \cdot b$,表示 f 被 b 膨胀后再用 b 来腐蚀的结果:

$$f \cdot b = (f \oplus b) \ominus b \quad (5)$$

开运算完全删除了不能包含结构元素的对象区域,平滑了对象的轮廓,断开了狭窄的连接,去掉了细小的突出部分。闭运算也会平滑对象的轮廓,而且闭运算会将狭窄的缺口连接起来形成细长的湾口,并填充比结构元素小的洞。本文使用两组不同的结构元素: $D = \{D_1, D_2, D_3, D_4\}$ 和 $E = \{E_1, E_2, E_3, E_4\}$,来进行数学形态学计算。如图3所示,结构元素 D 用来去掉目标上、下、左、右4个方向的点,结构元素 E 用来去掉目标左上、左下、右上和右下方向上的点。图3中阴影部分的像素点表示中心点,1表示目标点,0表示背景点,X表示既可以是目标点也可以是背景点。

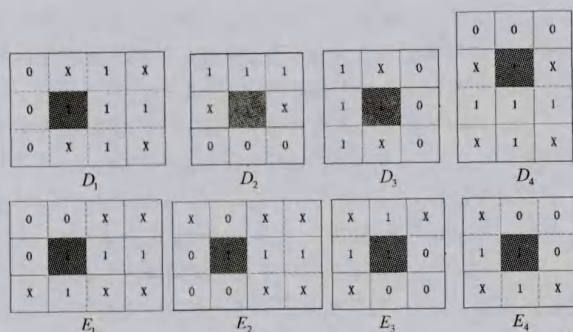


图3 结构元素模板

本文将开运算和闭运算结合使用,这样可以使图像前景和背景灰度拉伸,突显比背景暗的一些血管和细节,起到增强血管、消除噪声的作用。图4显示了相同数据下是否经过预处理的结果的比较。图4(a)是未经过预处理的血管切片图像,红色标记显示的是分割血管时产生的孔洞和湾。孔洞的存在对后续的骨架提取带来很大误差,而湾是不具有连通性的,会引起拓扑信息的丢失。为尽可能地减少孔洞带来的误差,保持湾原有的拓扑结构信息,采用了闭变换运算来填充孔洞和湾,结果如图4(b)所示。图4(c)中红色标记显示的是非主要小裂隙和小颗粒,这些噪声也会对后面的细化造成很大影响。使用 $f \otimes \{B\} = ((\dots((f \otimes B_1) \otimes B_2) \dots) \otimes B_n) (f \otimes B = f - (f \otimes B))$,其中结构元素序列为 $\{B\} = \{D_1, \dots, D_4, E_1, \dots, E_4\}$,从 D_1 到 E_4 依次计算、运行一遍,消除了小裂隙和小颗粒,结果如图4(d)所示。

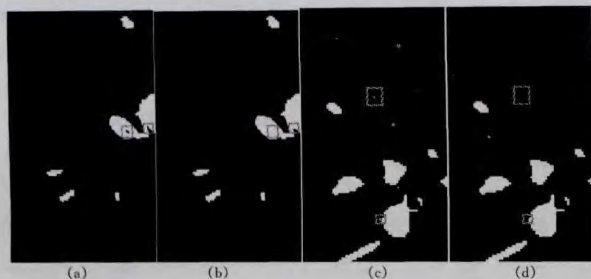


图4 血管图像预处理

4.2 基于细化的肺部血管骨架化

CT数据集产生的肺部血管图像是由许多二维切片组成的三维图像序列,图像数据量较大且在大多数情况下,血管图像中存在弱连接和非平滑管状结构,在血管网络的尾部管状结构变得非常细小,所以肺部血管骨架化方法需要保证骨架的连通性和完整性。为确保最终骨架位置的准确性和光滑性,同时也为了降低边界噪声的敏感度,本文对肺部CT血管图像进行细化时,对于简化点采用了与以往不同的定义方式,定义了4种简化点:边界体素、线性体素、欧拉不变性体素和简单体素。

定义1 如果某体素的6邻域中至少有一个体素值不为1,那么它就是边界体素。

定义2 如果某体素不位于线段的尾部,也就是说在它的26邻域中,不止有一个体素值为1,那么它就是线性体素。

定义3 如果删除某体素不改变欧拉特性,那么它就是欧拉不变体素。

定义4 如果删除某体素不改变26邻域连通性,那么它

就是简单体素。

判断一个体素是否为边界体素或线性体素,取决于其邻域系统的体素取值;判断它是否为欧拉不变体素就要涉及对整个连通物体、洞和腔数量的计算,其中欧拉特性函数为 $F = O - H + C$, O, H, C 分别表示相连物体、洞和腔的数量。判断欧拉不变体素与文献[6]中判断某个体素26邻域内任意两个目标点通过其他目标点可以连成一条26连通路,本质上是一样的;判断简单体素时涉及到的26邻域连通性是指删除该体素前后连通物体的数量不会发生变化,这与判断该体素6邻域内任意两个背景点通过18邻域中的背景点是否可以连成一条6连通的路径,效果也相同。因此这4种体素涵盖了需要删除的所有点,但以前在判断体素时先读入当前体素26个邻居中每个邻居26邻域内目标点的位置,并利用该位置信息来判断当前体素的邻居中目标点的连通情况。而采用欧拉不变体素和简单体素之后,在判断体素时是先读入当前体素的26邻居中的目标点,再循环判断两点之间的连通性,这样就可以将原来判断准则中的三重循环降低为两重循环,提高了计算速度。

图5(a)中间的黑点其6邻域内有2个目标体素、4个背景体素,它是一个边界体素。图5(b)中间的黑点其26邻域内有3个目标体素,它是一个线性体素。图5(c)中有1个相连物体、2个腔和1个洞, $F = 1 - 1 + 2 = 2$,如果把红色目标体素删除,相连物体数为1、腔的数量为2、洞的数量变成2, $F = 1 - 2 + 2 = 1$,就改变了整个欧拉特性,它不是欧拉不变体素;如果把绿色目标体素删除,相连物体数为1、腔的数量为2、洞的数量还是1, $F = 1 - 1 + 2 = 2$,那么它就是欧拉不变体素。如果把图5(d)中的红色目标体素删除,整个相连物体将由1个变成2个,那它就不是简单体素,而把绿色目标体素删除之后还是1个相连物体,所以绿点是简单体素。

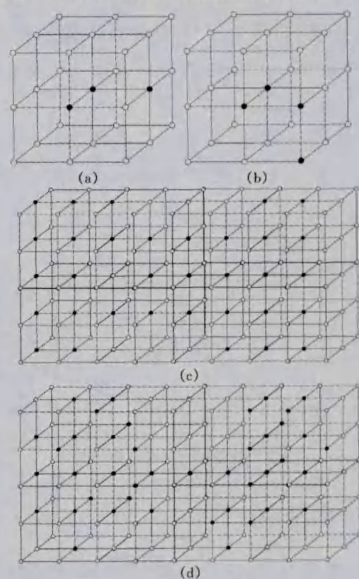


图5 待删除体素

本文采用的骨架化方法将不停地腐蚀上述定义的4种体素,直到无体素可被删除为止,总的来说,基于细化的肺部血管骨架化方法主要步骤如下所示。

Input: 胸腔CT图像切片集 $images \in R^{m \times n \times t}$

Initialize;

```

1) I=Segmentate(images)
% 区域生长法分割血管
2) I=Binary(I)
% 将分割出来的血管图像二值化
3) I=((I>b) & b)&imfill(I,'hole')
% 形态学运算预处理
Repeat
% 遍历整个图像,直到没有符合的体素能被删除为止
v=I(v)
if v∈Vdelete then delete(v);
% 如果该体素属于定义1-定义4中的体素,则删除
Until no voxels are deleted.
Output: 骨架结果

```

5 实验结果与分析

5.1 实验数据与环境

选取一系列肺部 CT 数据集进行肺部血管骨架化提取实验,CT 图像为 DICOM 格式数据,共有 322 个切片,每张切片大小为 92×150 像素,切片间隔为 2.5mm。用于测试的计算机为 Pentium(R) Dual-Core CPU E5200(2.50GHz),2GB 内存,操作系统为 Windows7,算法在 Matlab2009a 上实现,并封装到 MITK 平台中。

5.2 骨架化结果与计算复杂度分析

使用第 3 节的区域生长方法将肺部血管分割出来,再经过第 4.1 节形态学预处理得到的结果,分别采用文献[4]、文献[5]和本文的方法进行骨架化对比实验。实验结果表明,本文采用的血管骨架化方法是准确而有效的,如图 6、图 7 所示。为了保证实验的可靠性并对比不同方法的实验结果,在二维切片图像上进行方法对比实验。图 6(a)中所示为肺部血管的 5 张初始切片图像,图 6(b)~(d)分别为使用文献[4]、[5]和本文方法的骨架化结果。从图 6(b)可以直接看出,文献[4]的方法产生的骨架与原图目标物体的拓扑结构差距很大;在图 6(c)中,红色圆环标记区域表示文献[5]的方法在目标物体存在弱连接时,既不能保证骨架结果的连通性,也不能保证骨架结果的完整性。

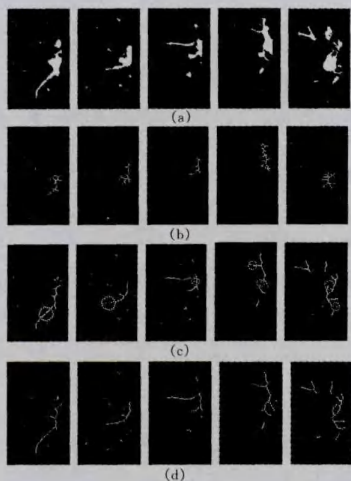


图 6 不同方法的骨架化对比实验结果

文献[4]中的方法首先将原始图像转换为特征元素和非特征元素,特征元素属于目标边界,接着对每个元素计算与其

最近特征元素之间的距离,并生成一个距离图,最后探测距离图中的屋脊点作为骨架点,整个过程存在二次两重循环,其执行时间复杂度为 $O(N^2)$ (N 为图像中像素的个数)。文献[5]中的方法在提取骨架之前要计算 Voronoi 图,虽然其骨架化过程的时间复杂度为 $O(n \log n)$ (n 为 Voronoi 图中顶点的数量),但构造 Voronoi 图需要花费较长时间,从而造成其效率不高。本文方法最多只存在一次两重循环,时间复杂度为 $O(N^2)$ (N 为图像中像素的个数),因此其效率较高。如表 1 所列,文献[5]的方法计算时间要远远大于其他两种方法,当处理数据量大的三维 CT 图像时性能最差。但是本文采用的方法能在较短时间内生成骨架,同时能保证骨架的拓扑结构和连通性,实验结果如图 6(d)所示。

表 1 不同方法的计算时间

图像	大小	计算时间(s)		
		文献[4]的方法	文献[5]的方法	本文的方法
图 6(a)左一	92 * 150	0.3588	0.5304	0.1872
图 6(a)左二		0.3744	0.5460	0.1404
图 6(a)中间		0.3900	0.3432	0.1248
图 6(a)右二		0.3432	0.4368	0.1543
图 6(a)右一		0.3276	0.5148	0.1308
图 7(a)	92 * 150 * 322	16.0548	内存溢出	7.2031

三维图像上使用文献[5]的方法会造成内存溢出,得不到其骨架化结果。图 7 显示了采用各种方法得到的肺部血管和三维骨架的可视化结果,图 7(a)是肺部血管分割结果的三维重建图,图 7(b)为文献[4]的方法提取的血管骨架化结果,图 7(c)显示使用本文方法提取的血管骨架化结果,图 7(d)为本文骨架与血管之间三维空间位置和拓扑结构的对比。从图中可以看出,文献[4]的方法得到的骨架与肺部血管的拓扑结构存在一定差距,本文所提方法得到的骨架基本位于血管的中心,同时保持了整个肺部血管的拓扑结构。

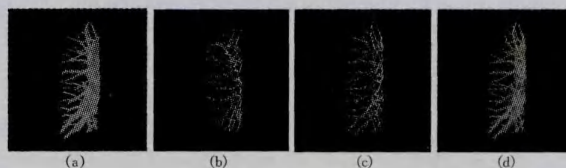


图 7 肺部血管和骨架的可视化结果

结束语 本文采用区域生长的方式分割出肺部血管,利用形态学操作对血管进行预处理,接着应用三维细化算法来获得肺部血管的骨架。实验结果表明,该方法可以准确地提取肺部血管拓扑结构,同时能去除噪声、病变的影响,对肺部 CT 图像处理具有实用价值。

参考文献

- [1] Ingrid S, Arnold S, Mathias P, et al. Computer Analysis of Computed Tomography Scans of the Lung: A Survey [J]. IEEE Transactions Medical Imaging, 2006, 25(4): 385-405
- [2] Agam G, Samuel G A. Vessel tree reconstruction in thoracic CT scans with application to Nodule Detection [J]. IEEE Transactions on Medical Imaging, 2005, 24(4): 486-499
- [3] 孙晓鹏, 张琪. 层次渐进的三维骨架算法[J]. 计算机科学, 2010, 12(37): 238-240
- [4] Couprie M, Coeurjolly D, Zrou R. Discrete bisector function and

Euclidean skeleton in 2D and 3D[J]. Image and Vision Computing, 2007, 25(10):1543-1556

- [5] Beristain A, Grana M. Pruning algorithm for voronoi skeletons[J]. Electronics Letters, 2010, 46(1):39-41
- [6] Couprie M, Bertrand G. New characterizations of simple points in 2D, 3D and 4D discrete spaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(4):637-648
- [7] Chen Y, Drechsler K, Zhao W, et al. A Thinning-based Liver Vessel Skeletonization Method[C]// International Conference on Internet Computing and Information Services, 2011:152-155
- [8] Palagyi K, Kuba A. Directional 3D Thinning Using 8-Subiterations[C]// Proceedings of Discrete Geometry for Computer Imagery. 1999, 1568:325-336

(上接第 252 页)

$$\text{sim}(E_1, E_2) = \max(\text{sim}(L_1, L_2), \text{sim}(L_1, L_2^1), \text{sim}(L_1, L_2^2), \dots, \text{sim}(L_1, L_2^{2^n-1})) \quad (3)$$

步骤 5 结束。

4 实验结果及分析

数学公式抄袭的手段主要有原样复制、修改变量名称和交换可交换运算符两边子式的位置等。为了验证算法的性能,本文实验从已公开发表的学术论文中选取 50 个含义和结构互不同的数学公式进行测试。根据抄袭手段的不同,人工对这些公式进行修改,具体的修改方式如表 2 所列。

表 2 修改方式的种类

修改种类	1	2	3	4	5
修改方式	未作修改	修改变量名称	交换可交换符号两边子式	修改变量名称和交换可交换符号两边子式	修改常量值

实验中,按表 2 的修改方式修改每个公式,并分别计算源公式与修改后的目标公式的相似度,每一种修改方式的平均相似度如表 3 所列。

表 3 各种修改方式的检测结果

修改方式	1	2	3	4	5
相似度(%)	100	100	100	100	78.41

从实验结果可以看出,未作修改的公式、修改变量名称的公式以及交换可交换运算符两边子式表达式的公式,其相似度都为 100%,这是因为对源公式和目标公式的树形结构作归一化处理后,源公式和目标公式的树形结构相同,虽然与源公式树形结构形同的目标公式二叉树可能有多种,但在对所有的先序遍历序列中变量名称作归一化处理后,目标公式的遍历序列中至少存在一个与源公式遍历序列完全相同的遍历序列,使修改后的目标公式恢复原形。修改常量值时的相似度较低,这是因为修改公式中的常量后,公式的含义在一定程度上与源表达式不同,即源表达式和目标表达式是两个不完全相同的表达式。例如, $\sum_{i=1}^{10} i^2$ 和 $\sum_{i=10}^{20} i^3$ 是两个含义不相同的数学表达式。另外,算法中用结构码标识树形结构,并且字符串匹配过程是以公式元素为单位,匹配速度较快。

结束语 基于数学公式的二叉树表示,本文提出了一种

- [9] 陈磊,王胜军,郑全录,等. 基于 CT 图像的三维拓扑细化算法及其在心脏 CAD 中的应用[J]. 计算机应用, 2007, 6(27):406-410
- [10] 滕奇志,康瑕,唐棠,等. 基于升序复核的并行三维图像骨架化算法[J]. 光学精密工程, 2009, 10(17):2528-2534
- [11] Friman O, Hindennach M, Khnel C, et al. Multiple hypothesis template tracking of small 3D vessel structures[J]. Medical Image Analysis, 2010, 14(2):160-171
- [12] 朱应礼,徐益明,崔秋梅. MSCT 肺血管成像对肺动脉栓塞的诊断价值[J]. 医学影像学杂志, 2008, 18(6):597-599
- [13] Lin K S, Tsai C L, Sofka M, et al. Vascular tree construction with anatomical realism for retinal images[C]// Proceedings of the 9th IEEE International Conference on BioInformatics and BioEngineering. 2009:313-318

LaTeX 格式的数学公式匹配算法。该方法对未作修改的公式、修改变量名称的公式和交换可交换运算符两边子式表达的公式进行的检测都十分精确,且具有较高的检测速度,有效地解决了论文抄袭检测中的数学公式检测问题。但有些问题还有待进一步研究,如具有分配属性的运算处理、矩阵处理以及复杂的多行表达式处理等。

参 考 文 献

- [1] Mander U, Baker B S. Deducing similarities in java sources from bytecode [C]// Usenix 1998 Annual Technical Conference. New Orleans: The Advanced Computing Systems Association, 1998: 179-190
- [2] 史彦军,滕弘飞,金博. 抄袭论文识别研究与发展[J]. 大连理工大学学报, 2005, 45(1):50-57
- [3] 鲍军鹏,沈钧毅,刘晓东,等. 自然语言文档复制检测研究综述[J]. 软件学报, 2003, 14(10):1753-1761
- [4] 金博,史彦军. 基于篇章结构相似度的复制检测算法[J]. 大连理工大学学报, 2007, 47(1):125-130
- [5] 秦新国. 基于句子相似度的文档复制检测算法研究[J]. 现代图书情报技术, 2007(11):63-66
- [6] Si A, Leong H V, Lau R W H. CHECK: A document plagiarism detection system[C]// Proceedings of the ACM Symposium for Applied Computing. 1997:70-77
- [7] 郭育生,黄磊,刘昌平. 基于多候选的数学公式识别系统[J]. 计算机研究与发展, 2007, 44(7):1144-1150
- [8] Chan K F, Yeung D Y. Mathematical Expression Recognition: A Survey [J]. International Journal on Document Analysis and Recognition, 2000, 3(1):3-15
- [9] 靳简明,江红英,王庆人. 数学公式图像处理综述[J]. 模式识别与人工智能, 2005, 18(4):429-440
- [10] Lee H-J, Wang J-S. Design of a Mathematical Expression Recognition System[J]. Pattern Recognition Letters, 1997, 18(8):289-298
- [11] 靳简明,江红英,王庆人. 数学公式识别系统: MatheReader[J]. 计算机学报, 2006, 29(11):2018-2026
- [12] 刘宏哲,须德. 基于本体的语义相似度和相关度计算研究综述[J]. 计算机科学, 2012, 39(2):8-13
- [13] 林学民,王炜. 集合和字符串的相似度查询[J]. 计算机学报, 2011, 34(10):1853-1862