

基于孤立点分析的医学图像处理*

冯敏¹, 阴珊珊²

Medical image processing based on outlier analysis

Feng Min¹, Yin Shan-shan²

¹College of Information Engineering, Taishan Medical University, Taian 271016, Shandong Province, China; ²Taian Center for Disease Control and Prevention, Taian 271000, Shandong Province, China

Feng Min★, Master, Lecturer, College of Information Engineering, Taishan Medical University, Taian 271016, Shandong Province, China
fmxsc@126.com

Received: 2011-05-06
Accepted: 2011-06-19

Abstract

BACKGROUND: Due to the low resolution of naked eyes, many small details and texture changes in morphology are not easy to be found, it will affect the early judgement of diseases.

OBJECTIVE: To provide a new way of thinking and methods for the computer diagnosis of diseases in digital medical images, which helps doctors to detect and diagnose the early malignant lesions and improve diagnostic efficiency and accuracy.

METHODS: Outlier data mining technique was used to analyze large data sets, extract the hidden, unnoticed, and easily discarded, but very useful information, and find out the rules and patterns of medical diagnosis to assist doctors to diagnose disease.

RESULTS AND CONCLUSION: Experiments show that outlier analysis algorithm based on clustering of the medical image pixels is feasible for the brain lesions.

Feng M, Yin SS. Medical image processing based on outlier analysis. Zhongguo Zuzhi Gongcheng Yanjiu yu Linchuang Kangfu. 2011;15(39): 7340-7342. [http://www.crter.cn http://en.zglckf.com]

摘要

背景: 在进行临床诊断的时候, 医学影像中许多微小的纹理变化细节和形态特征不容易被发现, 会影响对病情的早期判断。
目的: 为数字医学图像中病变的计算机诊断提供一种新的思路和方法, 帮助医生及早发现和诊断恶性病变、提高诊断效率和准确性。

方法: 运用孤立点数据挖掘技术, 分析提取医学图像数据集中隐藏、不为人所注意、易被抛弃的但非常有用的信息, 找出其中的医学诊断规则和模式, 从而辅助医生进行疾病诊断。

结果与结论: 实验证明基于医学图像像素聚类的孤立点分析算法对于发现脑部病变是切实可行的。

关键词: 孤立点分析; 聚类; 图像挖掘; 医学图像; 数据挖掘

doi:10.3969/j.issn.1673-8225.2011.39.030

冯敏, 阴珊珊. 基于孤立点分析的医学图像处理[J]. 中国组织工程研究与临床康复, 2011, 15(39):7340-7342.

[http://www.crter.org http://cn.zglckf.com]

0 引言

在医学影像处理和医疗设备及软件的研究、开发领域, 许多机构都在进行通过计算机处理医学影像的尝试, 其中, 利用计算机将医学平面图片还原为三维数字模型的应用是目前进步明显、成就显著的一个方面, 这一类研究为医生提供了比以往更为直观、清晰的观察方式, 但还不能代替人工进行临床检查。在计算机辅助诊断方面, 人们进行了大量的探索和实践, 并由此发展出许多相关的计算机图像处理技术, 例如图形对中、图片比较、图像特征提取等等, 但是, 至今还没有一种可以完全代替人工进行临床医学影像检查的成熟技术出现。其制约因素, 主要有两个方面: 其一, 人体的组织结构极其复杂, 而人体的病变机制更是千差万别, 加之不同对象之间的个体差异以及人体正常生命活动造成的正常形变, 即使是同一个体同一身体部位的CT图像之间的差异也会非常大, 这就使得计算机区别健康个体图片与

带有病变信息图片的过程异常困难; 其二, 就研究内容和手段而言, 目前主要工作的特点是研究问题基本集中在关联规则和分类上, 研究的内容主要是医学图像数据库中每个对象的单一图像, 而且研究方法主要是先从图像上提取特征, 产生特征属性, 然后在特征属性上进行疾病检查。对图像特征的分析有利于区别身体不同结构的器官组织, 对关联规则的挖掘也能够揭示许多身体构造的内在规律, 但是对于检查对象的个体差异和生命活动造成的身体正常形变, 这些方法还难以处理。因此, 这些做法目前还局限在形变和个体差异很小的脑部及骨骼检查领域中, 对形变幅度较大的身体其他部位的处理效果还不理想。

本文针对目前各种图像数据库系统应用过程中积累的海量数据、但却难以提取出有价值信息这一难题, 运用孤立点数据挖掘技术^[1], 分析提取图像数据集中隐藏的, 不为人所注意的, 易被抛弃的但非常有用的信息, 找出其中的医学诊断规则和模式, 从而辅助医生进行疾病诊断。

¹泰山医学院信息工程学院, 山东省泰安市 271016; ²泰安市疾病预防控制中心, 山东省泰安市 271000

冯敏★, 男, 1980年生, 山东省泰安市人, 2009年山东师范大学毕业, 硕士, 讲师, 主要从事计算机辅助医学技术的研究。fmxsc@126.com

中图分类号: R318
文献标识码: B
文章编号: 1673-8225 (2011)39-07340-03

收稿日期: 2011-05-06
修回日期: 2011-06-19
(20110506012/M·W)

1 脑部医学图像聚类

1.1 图像数据采集 CT影像是医生在临床上进行诊断最常用的技术之一^[2-4]。CT是指从患者的头顶向下每隔一段距离(通常是几个毫米)生成一个图像,因此每例患者的CT就是一个图像的序列,这个序列中的图像之间存在着很强的位置关系,在这样的数据集合上利用孤立点分析来发现知识,辅助医生的诊断^[5-6]。本研究中,选取20例患者的脑部医学图像数据进行实验分析,给患者进行编号为: $IM_i=001, \dots, 020$, 每例患者的脑部图像依次编号为: $IM_{i,01}, IM_{i,02}, \dots, IM_{i,15}$, 然后在20例患者的有序图像数据集上进行实验比较分析。

1.2 图像数据预处理 为了减少冗余数据,提高算法的执行效率,需要对原始图像进行预处理,处理方法如下: **步骤1:** 利用图像处理中的剪切技术得到简化的脑部医学图像。**步骤2:** 将所有对象的有序图像集的像素灰度值与密度值的对应关系一致化^[7]。**步骤3:** 清除“噪声”数据^[8]。

1.3 基于像素的医学图像聚类 经过与医学影像专家的讨论,使用的算法是只对图像进行一次顺序扫描的算法,此算法的突出优点是:操作比较简单,计算过程快,很方便对大幅面的高分辨率的医学图像进行处理。

算法描述为:

输入: 一幅经过预处理之后的脑部医学CT图像的像素集 IM_{ij}, P ;

输出: 脑部医学CT图像的像素聚类结果。

说明: IM_{ij} 为第 i 位个体的第 j 张CT图像, 以下均是。

Sub Clustering()

{将脑部CT图片的像素点阵数据装入数组中; //减少I/O吞吐次数, 提高算法的运算速度;

Number置为零, 指针 $Po(x, y)$ 指向像素集中的初始位置;

WHILE (像素集中还存在没有被标记的像素)

{检查当前像素点与相邻各个聚类, 判断其是否可以加入其中;

if 像素 $p(i, j)$ 不属于任何聚类 then

{标记 $p(i, j)$ 为新聚类成员像素;

聚合所有与 $p(i, j)$ 密度可达的像素并予以标记编号;

Number=Number+1; //计数}

指针指向下一个点; }

}

2 基于孤立点分析的脑部医学图像挖掘

由前面的分析处理, 得到了每一幅脑部医学图像的像素聚类结果, 于是, 下面的工作就是利用这些聚类信

息, 用更加直观和可比性的数据资料, 从中发现异常数据, 帮助医生准确地进行疾病的分析和诊断, 为此, 人们引入了两个新的概念: 聚类参数矢量及聚类参数矢量组, 以帮助进行下一步的分析处理。

2.1 聚类参数矢量的构造 在前期的处理过程中, 脑部医学图像已经被使用基于医学图像像素的聚类方法分割为若干部分。从抽象的意义来说, 每一个部分表示了身体的一个特定的器官组织区域, 而对于计算机而言, 每一个部分都是若干像素点的聚集。需要从中提取有意义的、共性的特征值和参数, 从而实现将图像从以像素点为基本单元的阵列转化为以不同的器官组织为基本单元的抽象过程, 此处借鉴了现实医疗中医生通过自己的观察读片区分不同的组织区域的过程, 由此引入了聚类参数矢量这一工具, 借助它, 把所要处理的医学图像的一个个离散的像素点的灰度、坐标转化成了具备了密度、大小等具体参数特征的各种器官组织, 从而实现了数字化和抽象化。

一幅脑部图片的聚类参数矢量就是由经前期处理生成的每个聚类的编号、聚类的中心位置、每个聚类的像素点个数(即面积)、每个聚类的像素点灰度值(即密度值)这些描述信息所构成的一个数组, 每一幅图片的所有聚类参数矢量再构成该图片对象的聚类参数矢量组。通过这种方式, 就可以将一幅图片描述成数字化和抽象化的形式。

2.2 聚类参数矢量组的构造算法

输入: 一幅脑部医学CT图像的聚类结果集;

输出: 该幅脑部医学CT图像的聚类参数矢量组:

$IM_{ij}.Vector$ 。

Sub Clustering_Struct ()

{初始化聚类参数矢量组 $IM_{ij}.Vector$;

Number置为零; //计数

指针指向像素聚类结果集的初始位置;

While(还有聚类未被处理)

{计算当前聚类的中心位置Center;

计算当前聚类的像素点数/面积Area;

记录该聚类的灰度值/密度Density;

指针指向下一个聚类;

Number=Number+1;

}

}

2.3 定义孤立点 在实际医疗活动中, 医生是根据如下传统方式和依据来对医学CT图像进行分析和处理的: 当人体器官发生病变时, 会造成病变部位的结构、形态以及机制发生相应的变化, 反映到CT图像上就是该部位组织密度会发生相应的变化; 而CT图像恰巧建起了组织密度和图像像素灰度之间的映射关系, 因此医学CT图像可以直观地表现出身体部位的密度分布情况; 当组织病

变造成的组织密度变化发展到一定程度时, 就会被CT设备识别并反映到CT图像中, 从而导致了身体部位出现了密度异常, 这就是一个对医学诊断有辅助意义的孤立点^[9-10]。

2.4 孤立点挖掘算法 算法描述如下:

```

①首先设置阈值  $\lambda$  和  $\mu$ ;
②找出所有位于同一位置的图片中的相似聚类;
FOR j=1 TO y DO
  FOR i=1 TO x DO
    { k=1;
      WHILE IMij.Vecterk<>NULL DO
        { FOR w=2 TO x DO
          FOR V=1 TO z DO
            WHILE IMwj.VecterV<>NULL DO
              { IF IMij.Vecterk.Density=IMwj.VecterV.

```

Density AND

IMij.Vecterk.Area-IMwj.VecterV.Area|<= λ AND

|Distance(IMij.Vecterk.Center,IMwj.VecterV.Center)|<= μ

THEN 标记 IMij.Vecterk 和 IMwj.

VecterV为相似聚类;

(相似聚类矢量被加入到 IMij.

Vecterk所属的类)

}

k=k+1;

}

}

③没有相似聚类的象素区域即为所要寻找的孤立点, 即可能的病灶点。

说明: 此算法中的阈值 λ 和 μ 是随机设定的, 在本实验中, 当设置过大时, 会导致找不到孤立点, 也就失去了用计算机处理的意义, 当设置过小时, 会导致找到的孤立点太多, 无法克服脑部正常个体差异所造成的影响, 因此实验结果跟 λ 和 μ 的选择有较大关系。这也是今后需要改进的一个地方。

3 实验结果与分析

通过上面的算法处理之后, 孤立点就很方便的被找到了, 并且挖掘出来的孤立点与被处理的CT图像的特定区域存在一一对应关系, 孤立点分析结果的后期处理的任务就是还原被标记为孤立点的聚类参数矢量与相应医学图像区域之间的对应关系, 以比较直观的形式表现出来。采用了伪彩色处理技术^[11], 其处理算法比较简单, 具体是将灰度图像相应区域象素点的G、B值置零, R值置为255即用红色显示相关病变区域。

4 结论

本文以医学数据为背景, 研究了孤立点分析方法, 其实质就是将被处理的所有聚类参数矢量组中的孤立点予以标记, 或者说, 是将孤立点与正常聚类参数矢量分开来的过程。因此, 挖掘的结果, 就是一些被赋予了特殊标记的聚类参数矢量。它们的数量可能很多, 并且分布在多个不同的聚类参数矢量组中, 也可能极少甚至没有——这取决于前期象素聚类结果的粒度大小, 也就是阈值的大小。两个阈值的同时确定存在一定难度, 在今后的研究中需要在此加以改进, 另外, 对于实验结果, 也可以采用辅以文本报告的方式加以表示。

5 参考文献

- [1] Jiawei Han, Micheline Kamber. Beijing: China Machine Press. 2001. Jiawei Han, Micheline Kamber 等著. 范明, 孟小峰译. 数据挖掘: 概念与技术[M]. 北京: 机械工业出版社, 2001.
- [2] Tao YH, Miao T. Zhonghua Fangshexue Zazhi. 2000; 34(1): 66-67. 陶勇浩, 缪陶. 影像存档及通讯系统方案中医学数字成像及通讯标准遵从性的设计[J]. 中华放射学杂志, 2000, 34(1): 66-67.
- [3] Qiu F, Tian J, Cao Y, et al. Zhongguo Yixue Yingxiang Jishu. 2000; 16(1): 73-75. 邱峰, 田捷, 曹勇, 等. PACS系统综述[J]. 中国医学影像技术, 2000, 16(1): 73-75.
- [4] Liu JC, Tian J, Cao Y, et al. Zhongguo Yixue Yingxiang Jishu. 2000; 16(1): 76-78. 刘景春, 田捷, 曹勇, 等. PACS系统结构与实现[J]. 中国医学影像技术, 2000, 16(1): 76-78.
- [5] Shen HB, Wang ST, Wu XJ. Fuzzy Kernel Clustering with Outliers. Journal of Software. 2004; 15(7): 1021-1029.
- [6] He ZY, Xu XF, Huang JZ, et al. FP-Outlier: Frequent pattern based outlier detection. Computer Science and Information System. 2005; 2(1): 103-118.
- [7] Ruan QQ. Beijing: Publishing House of Electronics Industry. 2007: 192-194. 阮秋琦. 数字图像处理学[M]. 2版. 北京: 电子工业出版社, 2007: 192-194.
- [8] Cui L. Beijing: Higher Education Press. 2006: 57-60. 崔雷. 医学数据挖掘[M]. 北京: 高等教育出版社, 2006: 57-60.
- [9] Pan HW. Harbin: Heilongjiang People's Publishing House. 2007: 100-105. 潘海为. 医学图像数据挖掘关键技术的研究[M]. 哈尔滨: 黑龙江人民出版社, 2007: 100-105.
- [10] Shekhar S, Lu CT, Zhang P. A Unified Approach to Detecting Spatial Outliers. Geoinformatica. 2003; 7(2): 139-166.
- [11] Tian J, Bao SL, Zhou MQ. Beijing: Publishing House of Electronics Industry. 2004. 田捷, 包尚联, 周明全. 医学影像处理与分析[M]. 北京: 电子工业出版社, 2004.

来自本文课题的更多信息——

作者贡献: 实验设计、实施为冯敏, 实验评估、资料收集为冯敏、阴珊珊。冯敏成文、审校, 并对文章负责。

利益冲突: 课题未涉及任何厂家及相关雇主或其他经济组织直接或间接的经济或利益的赞助。

伦理批准: 所获得的实验数据均得到了患者本人的知情同意。

本文创新性: 以“孤立点分析、图像挖掘”为检索关键词, 检索中国知网数据库, 检索时间为2005年至今, 发现本研究存在一定的先进性。其创新点为: 将孤立点数据挖掘技术与医学图像挖掘相结合, 提出了一种基于医学图像象素聚类的孤立点挖掘算法, 并通过实验验证了本算法的有效性和先进性。