

基于 CBR 的原发性心脏恶性肿瘤诊断系统设计

武会苹 李 莉

(上海长征医院肿瘤科 上海 200070)

〔摘要〕 为对原发性心脏恶性肿瘤的早期诊断提供重要临床信息, 提出应用实例推理方法 (Case - based Reasoning, CBR) 模型建立原发性心脏恶性肿瘤诊断系统的构想, 并阐述 CBR 模型实例库设计和若干关键技术, 为原发性心脏恶性肿瘤的自动诊断提供新的思路。

〔关键词〕 原发性心脏恶性肿瘤诊断系统; 实例推理方法; 实例库

Primary Cardiac Malignant Tumor Diagnosis System Design Based on CBR WU Hui - ping, LI Li, Oncology Department, Shanghai Changzheng Hospital, Shanghai 200070, China

〔Abstract〕 In order to provide important clinical information for early diagnosis of primary cardiac malignant tumor, the paper proposes the idea of constructing primary cardiac malignant tumor diagnosis system using case - based reasoning (CBR) model. It elaborates the design of CBR model case database as well as several key technologies, provides a new idea for cardiac malignant tumor automatic diagnosis.

〔Keywords〕 Primary cardiac malignant tumor diagnosis system; Case - based reasoning (CBR); Case database

目前, 心脏肿瘤的术前诊断采用无创经胸或经食道超声心动显像、计算机断层显像 (CT) 以及磁共振显像代替心导管的检查方法^[1]。医学信息化的迅猛发展, 为原发性心脏恶性肿瘤的早期诊断提供重要临床信息, 给原发性心脏恶性肿瘤诊断系统的研究奠定了基础。本文将实例推理方法 (Case - based Reasoning, CBR) 引入原发性心脏恶性肿瘤的早期诊断中, 通过运用相似度量理论来实现相似案例快速检索, 为原发性心脏恶性肿瘤的诊断提供一条新的途径。

1 实例推理方法

实例推理方法最初来自耶鲁大学的 Roger Schank 等学者的研究工作, 是一种基于人的认知心理过程的机器自学习和推理方法^[2]。它强调人们在解决新问题的時候, 常常根据以往积累下来的类似情况处理, 并通过过去类似情况处理的适当修改来解决新问题。将 CBR 方法应用于原发性心脏恶性肿瘤的早期诊断中, 具体过程如下: 利用专家的知识 and 经验, 构建原发性心脏恶性肿瘤实例, 并按照一定的结果组成实例知识库; 当有新的病例出现时, 系统能够从实例库中检索出与当前病例相应的相似实例; 从检索出的结果中找出相似性最高的实例, 根据约束规则对其进行调整、修改和评定, 生

〔收稿日期〕 2010 - 07 - 14

〔作者简介〕 武会苹, 主要研究方向为肿瘤信息化; 通讯作者: 李莉。

成针对目标实例的诊断结果；根据给定的策略要求将新生成的实例加入到知识库中扩充，以备后用。从基于 CBR 的诊断过程来看，CBR 技术通过从一个旧实例到另一个新问题的类比推理，利用过去的经验来指导新问题的解决，这样可以有效降低从医经验对诊断结果的局限性影响，提高诊断结果的科学性和准确性。

2 基于 CBR 的原发性心脏恶性肿瘤诊断系统总体结构

在任何医疗部门，知名专家和资深医生的诊断都是十分重要的。为使大多数医生都能了解知名专家或资深医生的知识和技能，提高医疗水平，可以把实例推理技术引进医疗诊断过程中。基于 CBR 的原发性心脏恶性肿瘤诊断系统总体结构，见图 1。

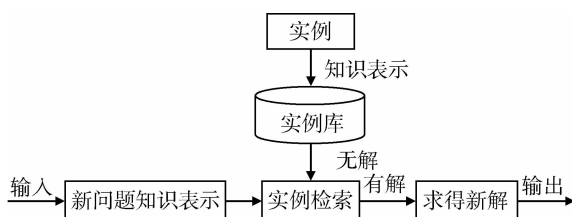


图 1 基于 CBR 的原发性心脏恶性肿瘤诊断系统总体结构

系统总体结构主要包括实例库和实例检索。实例库收集心脏恶性肿瘤的历史数据，结合专家的诊断经验对样本进行特征提取和知识表示，并存入数据库中；实例检索是在实例库中检索到与当前欲求解问题的“最相似”的实例，从而得到当前实例的诊断结果。反之，将新生成的实例加入到实例库中，以备后用。

3 关键技术

3.1 实例库的建立

存储在数据库中的原发性心脏恶性肿瘤实例，是其特征属性字段来表示的。基于 CBR 的诊断，其实例被描述为 $\{M_i = (T_i, S_i)\}$ 。其中 T_i 表示第 i 个实例的特征描述， S_i 表示相应的诊断结果。基

于临床诊断信息，可将特征归纳为以下几个方面：年龄、主要病史、肿瘤部位、肿瘤体积、肿瘤类型和诊断结果。不难发现，多数特征是基于文本匹配型的，因此匹配可分为两步进行。其一，语义匹配，采用词汇语义相似度检索相似分类^[3]。其二，数值匹配，根据属性值的特性，分精确值和区间值两种。而精确值可以看作是区间值的特殊情况，因此仅研究区间值之间的匹配。

3.2 相似度

基于案例推理的心脏恶性肿瘤诊断目的是要在实例库中检索到与当前欲求解问题“最相似”的实例，从而得到当前实例的诊断结果。两个实例对象 I_i 和 I_j 的相似度可定义为：

$$\text{Sim}(I_i, I_j) = \sum_{k=1}^n w_k \text{sim}(I_{ik}, I_{jk}) \quad (1)$$

其中 w_k 为各属性相对实例对象的重要程度，且 $\sum_{k=1}^n w_k = 1$ ， n 为实例对象的属性个数。 $\text{sim}(I_{ik}, I_{jk})$ 为实例对象 I_i 和 I_j 关于属性 k 的相似度。 $\text{sim}(I_{ik}, I_{jk})$ 的计算分两种情况：

当属性 k 为语义型时：

$$\text{sim}(I_{ik}, I_{jk}) = \begin{cases} 1, & I_{ik} \text{ 和 } I_{jk} \text{ 配对} \\ 0, & \text{其他} \end{cases} \quad (2)$$

当属性 k 为区间型时，假设 $I_{ik} = [I_{ik}, \bar{I}_{ik}]$ ， $I_{jk} = [I_{jk}, \bar{I}_{jk}]$ 为两区间值，采用相对长度法可计算两个区间属性值的相似度：

$$\text{sim}(I_{ik}, I_{jk}) = \frac{L(I_{ik} \cap I_{jk})}{L(I_{ik}) + L(I_{jk}) - L(I_{ik} \cap I_{jk})} \quad (3)$$

其中 L 表示相应区间的长度， $I_{ik} \cap I_{jk}$ 表示 I_{ik} 和 I_{jk} 的重叠区间。

3.3 相似匹配检索算法

检索是在心脏恶性肿瘤实例库中查找类似的实例，常用的方法是最近邻法、归纳索引法和知识索引法，其它测量相似性的算法还有神经网络、模糊逻辑和各种统计算法等^[3]。根据心脏恶性肿瘤诊断的特点及应用领域实际状况，本文采用基于相似度的最近邻法的实例检索算法。算法描述如下：预置诊断系统中查询实例各属性的权值为 $(w_1, w_2, L,$

w_n)。将求解问题各属性分为语义型和区间型,并表示为 $\{M_i = (T_i, S_i)\}$ 。选择权值最高的 N 个属性作为检索项,检查实例库中与查询实例类型相似的所有实例。根据求解问题的各属性对实例库中属性的权重,并用相似度计算公式 (2) 和 (3) 可得到该问题与实例库中现有的各个实例的相似度。若求解问题与现有实例的相似度大于预先设定的阈值,则为当前欲求解的新问题提供可能解。如果检索后的相似实例的解能够作为当前欲求问题的解,即当前欲求问题与实例库中的实例的相似度高于预先设置的阈值,则当前欲求解的问题实例不加入实例库中;反之,进行手工诊断,并且将当前欲求解的问题作为新实例加入实例库中。

原发性心脏恶性肿瘤诊断系统的设计包括过程模型设计、CBR 实例库设计和检索算法。它为原发

性心脏恶性肿瘤的早期诊断提供的重要临床信息,提高了心脏恶性肿瘤的自动诊断能力。随着医学信息化的不断发展,该系统设计为肿瘤诊断提供了新的思路,同时为病人的早期诊断预警提供指导。

参考文献

1 周维新,杨浣宜,张枝革,等. 超声心动显像评价心脏恶性肿瘤的手术、病例和尸检结果对照 [J]. 中国超声医学杂志. 2004, 20 (1): 38 - 42.

2 Watson, I. Applying Case - based Reasoning: techniques for enterprise systems [M]. San Francisco: Morgan Kaufmann Publishers, Inc. , 1997.

3 吴健,吴朝晖,李莹,等. 基于本体论和词汇语义相似度的 Web 服务发现 [J]. 计算机学报, 2005, 28 (4): 595 - 602.

(上接第 40 页)

GRNYCODE	GRNYNAME
001	旅店
001	文化娱乐场所
001	公共浴室
001	理发店,美容店
001	游泳池
001	体育场
001	图书馆,博物馆
001	商场,书店
001	餐饮
002	墙子街
002	王串场街
002	江都路街
002	大江路街
002	小树林街
002	小关街
002	光复道街
002	鸿顺里街
002	京津公路街
002	新开河街
002	东六街

图 1 公共场所卫生监测管理数据库的 GRNY 表

5 结语

目前此公共场所卫生监测管理系统已在天津市河北区疾病预防控制中心的业务工作中投入运行,使得程序不仅仅停留在设计阶段,更能在实际工作中发挥效能。系统运行结果表明,该系统可以满足公共场所卫生监测管理人员的工作需要,大大提高

了工作效率,使计算机技术在卫生工作中发挥作用。

参考文献

1 公共场所卫生监测技术规范 [S]. 1998.

2 室内环境质量及检测标准汇编 [M]. 北京: 中国标准出版社, 2004.

3 王立福. 软件工程 [M]. 第 4 版. 北京: 北京大学出版社, 2008.

4 潘锦平. 软件开发技术 [M]. 西安: 西安电子科技大学出版社, 2006.

5 龚沛曾,陆慰民,杨志强. Visual basic 程序设计教程 (6. 0 版) [M]. 北京: 高等教育出版社, 2006.

6 [美] Anthony Prince David Rinaldi 著. 王建华, 陈一飞, 吴柄林译. Visual Basic 6 开发人员指南. 北京: 机械工业出版社, 2007.

7 张俊玲. 数据库原理与应用 [M]. 北京: 清华大学出版社, 2005.

8 蔡翠平. Visual Basic 6. 0 中文版控件大全 [M]. 北京: 电子工业出版社, 2006.