第 34 卷 第 20 期 计 算 机 工 程 2008 年 10 月
Vol.34 No.20 Computer Engineering October 2008
·开发研究与设计技术· 文章编号：1000—3428(2008)20—0283—03 文献标识码：A 中图分类号：N945.12

基于 Web 的通用疾病计算机辅助诊断平台

曹雨田[1]，崔杜武[1]，崔颖安[1]，强永乾[2]
(1. 西安理工大学计算机科学与工程学院，西安 710048；2. 西安交通大学第一附属医院影像中心，西安 710061)

摘 要：针对传统医疗专家系统设计的复杂性和局限性，提出一种基于 Web 的通用疾病辅助诊断平台。阐述系统的整体架构和工作流程，讨论该解决方案的设计思想及实现方法。介绍数据规范化、病历模板、病历采集、知识获取、在线诊断功能的实现技术。实际测试表明，该系统在功能和性能上都是可行的。

关键词：计算机辅助诊断；疾病属性；模板；知识挖掘；规则匹配

General Platform for Disease Computer Aided Diagnosis Based on Web

CAO Yu-tian[1], CUI Du-wu[1], CUI Ying-an[1], QIANG Yong-qian[2]
(1. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048;
2. Imaging Center, the 1st Affiliated Hospital, Xi'an Jiaotong University, Xi'an 710061)

【Abstract】 A resolve scheme of a general platform for disease computer-aided diagnosis based on Web is proposed, which aims at complexity and limitations of traditional medical experts system. This system and its work flow are illustrated, and the design thought and implementation method of this reslove scheme are discussed. This paper also presents the data normalization, form board of case, case collection, knowledge acquisition and implementation of diagnosis on line. Practical test results show that this system is feasible on both function and performance.

【Key words】 Computer Aided Diagnosis(CAD); disease attribute; template; knowledge mining; regulation matching

1 概述

计算机技术的迅猛发展使得各行业的信息化得以普及。在医疗信息化建设中，计算机技术更是发挥了重要作用。计算机网络、人工智能、图形图像、多媒体等技术已经普遍应用于其中。人们越来越多地想借助计算机强大的处理能力和智能性来辅助医生的日常工作，而疾病的辅助诊断就是人们所关注的问题之一。

计算机辅助诊断(Computer Aided Diagnosis, CAD)给医生提供了诊断参考和建议，不仅减轻了医生的工作负担和压力，也在一定程度上避免了医生由于疏忽或劳累而造成的漏诊和误诊情况。

计算机辅助诊断系统是专家系统的一个典型应用。专业的人工智能专家系统(experts system)的开发需要专门的人工智能开发语言和工具(如：PROLOG, LISP)。这些系统往往比较复杂(如：早期用于细菌感染疾病诊疗和咨询的MYCIN系统[1])。此外，传统的专家系统主要面向人与单机的交互，开放性和灵活性不强，影响其应用的广泛性。

随着计算机网络、远程通信、分布式技术和多媒体技术的发展，构建基于 Web 的专家系统已成为可能。本文提出一种基于 Web 的疾病辅助诊断的通用平台解决方案，在数据标准化及疾病属性规范化的基础上，提出一种动态可配置的病历模板技术，利用产生式规则对知识进行匹配，从而得出诊断结论。该系统可以随着疾病属性集和知识数据库的不断完善和丰富而不断地扩大所诊断疾病的种类，因此，具有较强的通用性。

2 基于 Web 的辅助诊断系统

2.1 系统架构

专家、知识工程师和普通用户可以通过浏览器访问应用服务器，将问题传递给推理机，推理机对数据库和知识库进行存取，推断出结论，然后将这些结论反馈给用户。基于 Web 的辅助诊断系统结构如图 1 所示，主要分为 3 个层次：浏览器，应用逻辑层和数据库层。

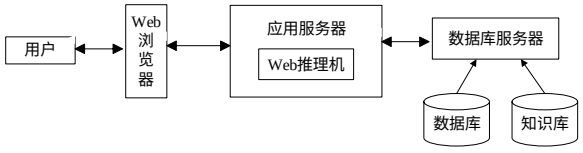


图 1 系统架构

2.2 系统流程

该系统主要有 3 类角色：专家，医生和管理员。(1)专家可以对模板进行创建、审核和维护。制定好的模板可以动态生成病历表。专家还可以制定诊断知识，也可以通过病历库挖掘来获得知识。(2)医生通过浏览器的诊断界面，录入病人

基金项目：国家自然科学基金资助项目(60501006)；陕西省教育厅科学技术研究基金资助项目(07JK327)
作者简介：曹雨田(1981－)，男，硕士研究生，主研方向：计算机医学应用，数据库技术；崔杜武，教授、博士生导师；崔颖安，讲师；强永乾，副教授
收稿日期：2008-01-20 **E-mail：**ladanyu@126.com

的病历数据，提交给系统进行辅助诊断，所得到的确诊结果及数据可作为病历存储到确诊病历库。(3)系统管理员主要负责各类数据，包括标准编码、疾病属性集的维护。图 2 描述了系统的主线流程。

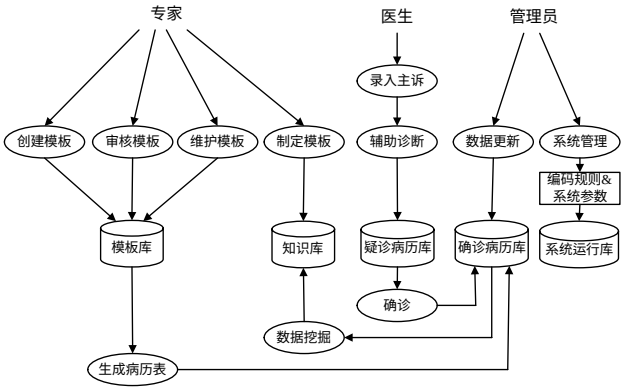


图 2　系统主线流程

2.3　系统特点

本系统具有以下特点：采用国际标准，基于疾病属性构建基础数据；基于产生式规则构建诊断知识；系统的通用性较强，适合各类疾病的在线诊断；系统功能便于扩展；系统人机友好，操作方便。

3　数据组织

数据组织是系统实施的前提和关键，本文所涉及的数据主要包括疾病属性、病历模板、病历、解剖学身体部位、疾病名称等。

3.1　数据规范化

为了利用和组织数据，需要对数据进行规范。数据规范化能简化医院之间、医院各系统之间的数据交流和共享，且便于系统版本的更新。

该系统所采用的标准包括：国际疾病标准编码ICD-10，影像学分类的病理编码(ACR Key, ACR Index for Radiological Diagnosis)，医学数字成像和通信标准(DICOM3.0, Digital Imaging and Communications in Medicine)，国际医疗电子信息交换标准(HL7, Health Level Seven)等[2]。

例如，疾病属性按身体部位分类时，采用 ACR Key 编码的前 4 位表示身体的解剖学部位，疾病名称采用 ICD-10 国际疾病标准编码，如 C34 表示支气管和肺恶性肿瘤。

疾病属性是描述疾病的症状词，包括临床属性和影像属性两部分。它是各种与疾病诊断有关的属性全集。规范化就是要规范统一所有属性名称和属性值。属性值有中心、周围、散列，它还具有层次关系，如图 3 所示。

<table>
<tr><td rowspan="11">边缘</td><td colspan="3">光滑</td><td></td></tr>
<tr><td rowspan="10">不光滑</td><td rowspan="3">毛刺</td><td colspan="2">无</td></tr>
<tr><td rowspan="2">有</td><td>长毛刺</td></tr>
<tr><td>短毛刺</td></tr>
<tr><td rowspan="4">分叶</td><td colspan="2">无</td></tr>
<tr><td rowspan="3">有</td><td>浅分叶</td></tr>
<tr><td>中分叶</td></tr>
<tr><td>深分叶</td></tr>
<tr><td rowspan="2">脐样切迹</td><td colspan="2">无</td></tr>
<tr><td colspan="2">有</td></tr>
<tr><td rowspan="2">晕征</td><td colspan="2">无</td></tr>
<tr><td colspan="2">有</td></tr>
</table>

图 3　属性层次关系

作为一个实例，西安交通大学第一附属医院做好了周围性肺癌的属性规范化工作。随着工作的不断深入，疾病属性库不断扩充，直至覆盖大部分疾病的属性信息。

3.2　病历模板

病历模板是针对某类疾病建立的疾病属性集合。疾病属性的规范化是实现系统通用性的基础，而病历模板则是通用性的体现。考虑到粒度的粗细及其耦合性，将模板定位到 ICD-10 编码的二级目录(代表一类疾病)。这样 300 个左右的病历模板将覆盖所有疾病(目前 ICD-10 中有 2 万余种疾病)。定义模板可以减轻医生录入病历数据的工作量，数据录入采用选择形式，既快捷又避免了错误的发生，而且便于数据维护，增强系统的灵活性。属性与模板的关系如图 4 所示。

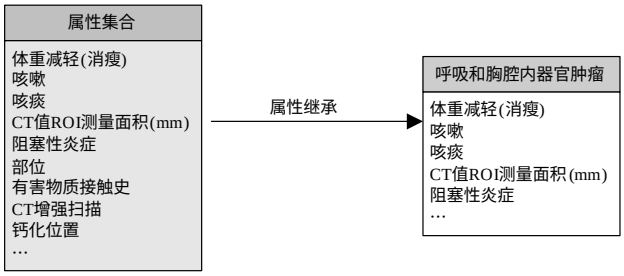


图 4　属性与模板的继承关系

为便于管理，模板中还包含其他说明信息，如申请人、申请时间、创建人、创建时间、上次修改时间、状态标志。系统还提供疾病模板的审核机制，保证了模板从提出申请到最终投入使用的整个流程的科学性和规范性。

3.3　病历

系统的病历库表由模板动态生成，是模板的实例，包括：头信息，属性分项和图像信息 3 个部分。其中，属性分项来自模板中的属性列表集，字段名是属性名，值为对应的属性值；头信息是病历的基本信息，包括：病历编号，病人姓氏，性别，年龄，民族，居住地以及病历的来源单位、确诊疾病；图像信息是指属于该病历的 CT 影像数据，采集病历时上传的 CT 影像为 DICOM3.0 格式，系统同时兼容如 Bmp, JPEG, TIFF 等其他格式的影像数据。

4　知识获取

系统运用归纳推理技术自动从确诊病历表中抽取专家知识，同时可以借助专家经验编辑疾病诊断知识。

4.1　知识结构

系统的诊断知识使用产生式规则来表示，具体表述形式如下：

<规则>=(<规则基本信息>,<前提>,<结论>,<其他信息>)

其中，<规则基本信息>包括来源、制定者、制定时间、修改者、修改时间、疾病系统、身体部位、标志位等；<前提>表示为 $A1*a1*q1/A2*a2*q2/\cdots/An*an*qn$，$Ai$ 表示属性名称，ai 表示属性值，qi 表示权重；<结论>即诊断结果疾病的编码；<其他信息>包含可信度(Certainty_Factor, CF)、支持度(Support_degree, SD)、使用次数、正确次数等。

例如，有一条诊断规则为

规则的前提　咳嗽*有*25.3%/毛刺*短毛刺*23.5%/边缘*不光滑*23.5%/乏力*有*15.9%/咳痰*粘液性*11.8%

规则的结论　支气管和肺恶性肿瘤(C34)

其他信息　(24.1% ,56.4%,10,8)

它表明如果病人有咳嗽症状、其肺 CT 片有肿块边缘不光滑呈短毛刺状，且有乏力症状，咳痰有粘液，则该病人得

支气管和肺恶性肿瘤病的可信度为 24.1%，其支持度为 56.4%。

4.2 数据挖掘模块

诊断规则挖掘系统相对独立于Web辅助诊断系统，通过数据连接模块来连接数据源，支持多种算法，包括关联规则(association rules)[3]、粗糙集(rough set)[4]、神经网络(Neural Network, NN)[5]等算法，具有自学习以及规则优化评估的特性。通过该模块可生成诊断规则。

5 辅助诊断

5.1 在线诊断

医生通过系统的前台界面录入病人的基本信息及主要症状，提交给 Web 诊断推理机，系统诊断出结果(支持度最大的 5 条结论按降序排列)，医生结合自己的经验知识选择若干个诊断结论保存至疑似病历库中，这样就完成了一次在线诊断过程。

5.2 规则匹配算法

输入 症状编码 Symptom

输出 支持度最大的前 5 个诊断结论

算法如下：

```
初始化
Get:Symptom
规则匹配
FOR(i=0;i<Rule_Count;i++){
Get:Rule[i]
//求匹配序列 Rule_Flag
FOR(j=0;j<Rule[i].Size;j++){
 IF(Symptom.Sx[j] ∈ Rule[i]){
    //比较症状与规则的属性值
    IF(Symptom.Sxz[j]==Rule[i].Sxz[j])
        Rule_Flag[i][j]=1;
    ELSE
        Rule_Flag[i][j]=0;
  }
 }
}
//计算可信度 CF
FOR(i=0;i<Rule_Count;i++){
 FOR(j=0;j<Rule[i].Size;j++){
 total=total+Rule_Flag[i][j]*Rule[i].Qz[j];
 }
 CF[i]=total;
}
输出支持度最大的 5 个诊断结论，按降序排列；
```

其中，症状 Symptom 形如：咳嗽*有/毛刺*短毛刺/边缘*不光滑/乏力*有/咳痰*粘液性；在求匹配序列前，先检查规则中是否有录入的症状信息项；可信度计算方法是将匹配序列的每一项乘以该项对应的权重值，将所有的乘项相加。

6 系统实现

系统使用的开发平台为 JBuilderX，后台数据库系统为 Oracle9i，采用 MVC 架构思想，将系统分为表现层、业务逻辑层、数据层加以实现。其中，病历的动态生成及显示采用了面向对象、ORM、改进 Validator 等技术，扩展了系统的灵活性，使模板的动态配置和表现成为可能。另外，部署在 Weblogic 上的 Web 推理机实现了辅助诊断的核心算法。

图 5 是病历数据录入的前台界面，图 6 是经在线辅助诊断后，系统给出的诊断报告。

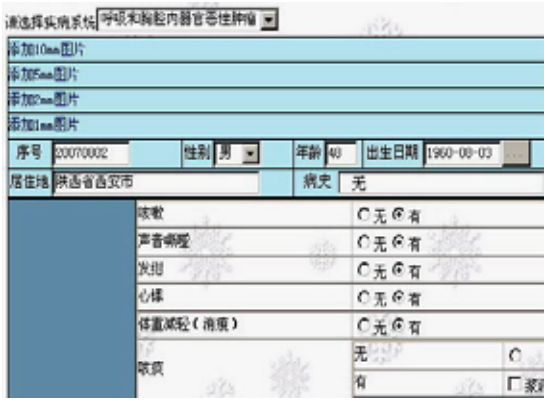


图 5 系统诊断症状录入界面



图 6 诊断报告

7 结束语

该项目由西安理工大学计算机学院与西安交通大学第一附属医院合作开发。目前系统中的数据以周围性肺癌为主，经过近半年的试运行，说明该系统在功能和性能上是可行的。今后将进一步提高系统诊断的准确率。

参考文献

[1] Shortliffe E H. Computer-based Medical Consultation: MYCIN[M]. New York, USA: American Elsevier, 1976.

[2] 电子病历研究中心. NEMA PS 3(1-16)-2004 DICOM3.0[S]. 2004.

[3] 蔡伟杰, 张晓辉, 朱建秋, 等. 关联规则挖掘综述[J]. 计算机工程, 2001, 27(5): 31-33.

[4] 辛旭明, 孙福良. 粗集在知识发现中的应用[J]. 东华大学学报, 2003, 29(5): 131-134.

[5] 周志华, 陈世福. 神经网络规则抽取[J]. 计算机研究与发展, 2002, 39(4): 398-405.