

医院智能导诊系统设计与实现^{*}

王嘉宁 苏 翀

任 瞳

(江苏科技大学电气与信息工程学院 张家港 215600)

(南京江北人民医院 南京 210000)

〔摘要〕 利用稀疏矩阵和杰卡德相似系数,结合医疗领域知识构建医院智能导诊系统,阐述系统需求、总体设计、关键技术和功能模块,包括智能导诊、综合信息查询、知识库管理及数据库设计,指出该系统能够在一定程度上克服医院现有分诊流程存在的弊端,缓解导诊服务压力。

〔关键词〕 智能导诊系统;稀疏矩阵;杰卡德相似系数

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.08.006

Design and Realization of the Intelligent Hospital Guide System WANG Jia-ning, SU Chong, School of Electrical and Information Engineering, Jiangsu University of Science and Technology, Zhangjiagang 215600, China; REN Tong, Jiangbei People's Hospital, Nanjing 210048, China

〔Abstract〕 By using sparse matrix and Jaccard similarity coefficient, the intelligent hospital guide system is built based on knowledge in the medical domain. The paper elaborates on the system requirements, overall design, key technologies and functional modules, including intelligent hospital guide, comprehensive information query, knowledge base management and database design, and points out that the system can overcome the disadvantages existing in the current hospital triage process to some extent and mitigate the pressure of hospital guide service.

〔Keywords〕 Intelligent guide system; Sparse matrix; Jaccard similarity coefficient

1 引言

随着生活水平的不断提高,人们对医疗服务的要求也越来越高。传统的医疗服务已经不能满足人们快节奏的生活需求,因此医院的信息化建设^[1]至关重要。现代化医院应该采取管理观念^[2],结合计算机技术,以提高患者的就诊体验和减轻医务人员

的工作量,使医院能够更加专注于医疗服务质量的提升,从而更好地服务于社会。为简化大型医院就诊流程、优化资源配置、提升服务质量,本研究充分利用现有医院信息平台^[3]研究并设计实现医院智能导诊系统。患者通过输入自身症状可以快速、准确地定位到应挂号的科室以及可能患有的疾病,在一定程度上减少挂错号等现象,提升患者就诊效率和就医感受,缓解医患矛盾。

〔收稿日期〕 2018-04-12

〔作者简介〕 王嘉宁,本科生;通讯作者:苏翀,博士,研究员。

〔基金项目〕 南京市医学科技发展一般性课题“基于随机森林算法的无创性乙肝代偿期肝硬化筛查模型的建立与研究”(项目编号:YKK16265)。

2 需求分析

2.1 背景

对于大型综合性医院而言,以专科为诊疗单元进行日常的诊疗工作。广大民众的医疗基本知识相

对匮乏，很多患者在就诊时往往不知道对应的挂号科室，继而导致医院挂号排队现象严重。虽然部分医院设置专门的导诊服务，但数量有限且工作负荷较大，催生出二次排队现象。这些在一定程度上导致医患关系不和谐，甚至发生冲突，为此有必要设计并实现智能导诊系统，以减少患者的窗口等待时间，缓解医院的“三长一短”现象。

2.2 知识库构建

系统需要知识库来存储所有发病部位、症状、对应的疾病名称以及门诊诊疗科室等信息。本研究中知识库的构建通过某医院信息系统进行整理，经过医学专家人工确认，范围仅涉及门诊常见疾病。为保证系统导诊功能的科学性和准确性，管理员可以对数据库中疾病部位表、症状表、疾病表以及科室表进行增删改操作，不断完善知识库。

2.3 智能导诊

系统需要友好的操作界面完成智能导诊流程。用户选择性别、年龄及患病部位和相关症状，系统就会显示用户可能患有的疾病和相关诊疗科室^[4]。为方便实际使用，系统会按照相似度降序排列，给出前几个疾病列表，本系统默认值为 3。导诊流程，见图 1。用户获得导诊结果后还可进一步查询疾病和相关科室的基本信息。

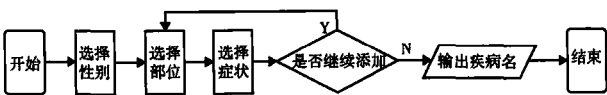


图 1 导诊流程

3 总体设计

系统主要由智能导诊和知识库管理两部分组成。智能导诊是面向用户，用户通过选择性别和年龄确定大致类别；随后根据用户选择的疾病部位和症状生成症状向量；再根据症状向量与知识库中疾病症状矩阵分别计算杰卡德相似度；最后按相似度降序排列，选取排名前 3 位的疾病作为最终推荐结果。此外支持用户查询相关疾病和诊疗科室信息。知识库管理部分面向系统管理员，管理员结合医学领域知识对疾病部位、疾病症状及对应的名称和诊

疗科室进行维护。系统总体框架，见图 2。

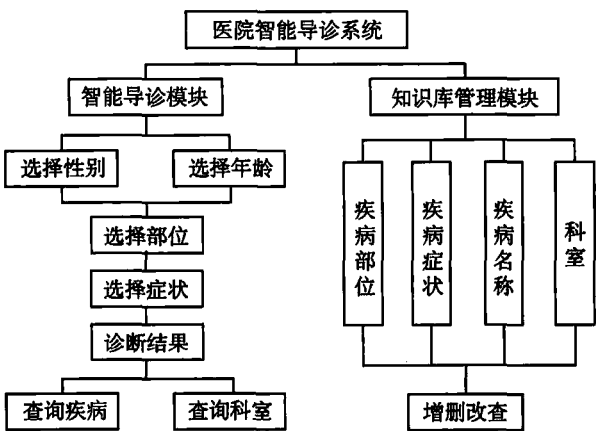


图 2 系统总体框架

4 关键技术

4.1 稀疏矩阵压缩存储

为将疾病症状存储到数据库中，系统设置数值 1 代表该疾病存在此症状，0 表示不存在该症状。所有疾病将以矩阵形式存储其表现的症状。考虑到绝大多数的疾病所表现出的症状往往只占少数，因此该疾病所对应的矩阵会存在大量的 0 元素，既浪费存储空间，又降低存取效率。为解决该问题本研究采用稀疏矩阵的数据结构来存储疾病与特征的对照信息^[5]。通过利用稀疏矩阵压缩存储技术将稀疏矩阵写成 3 元组表，将非 0 元素及其对应矩阵的行、列下标存入 3 元表中。

例 1：疾病症状空间有 6 个特征，分别为发烧、咳嗽、腹泻、皮肤、眼睛充血和失眠；对应的 5 种疾病分别为普通感冒、红眼病、皮炎、鼻炎和哮喘。红眼病一般仅存在眼睛充血症状，普通感冒往往具有发烧、咳嗽等症状。其存储的矩阵表示如下：

$$S = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

其中矩阵中列分别对应疾病的症状，行对应各种疾病。观测可知矩阵存在大量的 0 元素，占用 $5 \times 6 = 30$ 个存储空间，而利用 3 元表存储后如下：

$$S' = \begin{pmatrix} 1 & 1 & 5 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 3 & 4 \\ 1 & 4 & 2 \\ 1 & 6 & 6 \end{pmatrix}$$

每行代表矩阵中的非 0 元素，第 1 列代表该非 0 元素的值，第 2、3 列分别代表该元素所在矩阵中的行、列坐标。3 元表仅用 $3 \times 6 = 18$ 个单位空间，比原矩阵节省 12 个单位空间。因此采用稀疏矩阵压缩技术可以很好地节省存储空间，提高存取效率。

4.2 杰卡德相似系数

考虑到疾病与症状矩阵的稀疏特性，对于相比的两对症状向量均未出现的症状（即值为 0 的症状）个数显然很多，而这些症状对度量整体相似度贡献很小。因此采用更加适合稀疏数据的杰卡德相似系数度量症状向量之间的相似度。杰卡德相似系数（Jaccard）是衡量两个集合之间相似度的一种指标，常应用于网页去重、论文查重等^[6]。在本研究中假设 p 表示疾病 A 和用户输入的症状向量 B 中对应症状都为 1 的症状个数， q 表示疾病 A 对应的症状是 1，而用户输入的症状向量 B 对应的症状是 0 的症状个数， r 表示疾病 A 对应的症状是 0，而用户输入的症状向量 B 对应的症状是 1 的症状个数。那么 A 和 B 的杰卡德相似系数计算公式如下：

$$J(A,B) = \frac{p}{p+q+r}$$

可以看出分子是 A 和 B 出现相同症状的个数，分母是 A 和 B 中所出现症状的总个数。杰卡德相似系数很好地描述 A 和 B 具有的相同症状占 A 和 B 中所出现症状的比例。因此其能准确地描述 A 和 B 之间的相似程度。

例 2：假设所有症状包括如下：{发烧、咳嗽、头痛、眼干、胸痛、呼吸困难}。肺炎所表现的症状为发烧、咳嗽、胸痛、呼吸困难；而普通感冒所表现的症状为发烧和咳嗽。若用户提交的症状包括发烧、咳嗽和胸痛，则其对应的症状向量 $U = \{1,1,0,0,1,0\}$ ，对肺炎 A 和普通感冒 B 分别计算杰卡德相似系数：

$$A = \{1,1,0,0,1,1\}$$
$$B = \{1,1,0,0,0,0\}$$
$$J(U,A) = \frac{3}{3+1} = \frac{3}{4} = 0.75$$
$$J(U,A) = \frac{2}{2+2} = \frac{1}{2} = 0.5$$

可知 U 与 A 的相似度更大，该用户患有肺炎的可能性更大，因此系统会将肺炎及对应的科室反馈给用户。具体的智能导诊算法流程，见表 1。

表 1 具体的智能导诊算法流程

步骤	具体操作
输入	用户通过选项输入相应部位及出现症状 0-1 向量 U 、推荐的疾病及科室数目 k
Step1	获取用户输入的相应部位及出现症状向量 U 的特征空间集合 P
Step2	获取对照关系表中的疾病症状空间集合 Q
Step3	新的疾病症状空间集合 $R = P \cup Q$
Step4	按照新的疾病症状空间集合 R 重构 0-1 向量 U
Step5	将对照关系表转化为一个矩阵 M ，行代表疾病条目 s ，列数等于疾病症状空间集合 R 的基数 t
Step6	初始化一个具有 s 个元素的一维实型数组 a ，初值为 0
Step7	For $i=1$ to s
Step8	向量 $C =$ 向量 $U +$ 矩阵 M 中第 i 行行向量 L
Step9	two = 值为 2 的个数；one = 值为 1 的个数
Step10	$a[i] = \text{two} / (\text{two} + \text{one})$ ；//计算杰卡德相似系数
Step11	End for
Step12	对数组 a 按降序排序，返回前 k 个疾病及其对应诊疗科室

5 功能模块

5.1 智能导诊

在用户导诊界面中用户首先选择性别和年龄，然后选择患病部位，此时界面会提供属于该部位的相关症状，用户可以选择多个症状。当用户选择完毕并确认时，模块会将用户选择的患病部位及相关症状转化为症状向量，然后计算症状向量与知识库中疾病症状矩阵之间的杰卡德相似系数，进而给出用户可能患有的疾病列表，提供相关疾病和诊疗科室的信息^[7]。智能导诊数据流，见图 3。

5.2 综合信息查询

用户在获得导诊结果后可查询可能患有的疾病

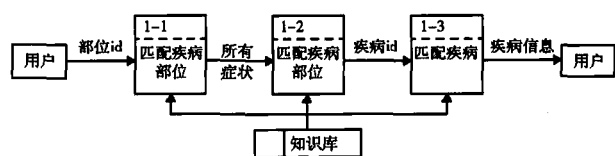


图 3 导诊模块数据流

和相关诊疗科室的信息；也可针对某种疾病和诊疗科室进行单独查询。查询结果主要包括疾病简要介绍、注意事项以及相关诊疗科室基本信息^[8]。

5.3 知识库管理

主要包括基本数据字典的管理以及科室、疾病、症状之间对照表的管理。基本数据字典的管理主要用于对部位、症状、疾病以及科室等进行单独维护。对照表的管理则相对比较复杂，除需要相应的医学领域知识外，还需要借助数据库的触发器技术保障数据之间的完整性和一致性^[9]。

5.4 数据库设计

根据需求分析，本系统包括疾病部位、疾病症状、疾病、科室以及对照关系表 5 个实体。其中疾病部位实体包括属性有部位名称和编号；疾病症状实体包括属性有症状名称、编号和所属部位；疾病实体包括属性有疾病名称、疾病编号、诊疗科室、所属部位、拥有症状、疾病介绍；科室实体包括属性有科室编号、名称、介绍以及地址。对照关系表主要体现上述 4 个实体之间的联系。具体实体联系，见图 4。

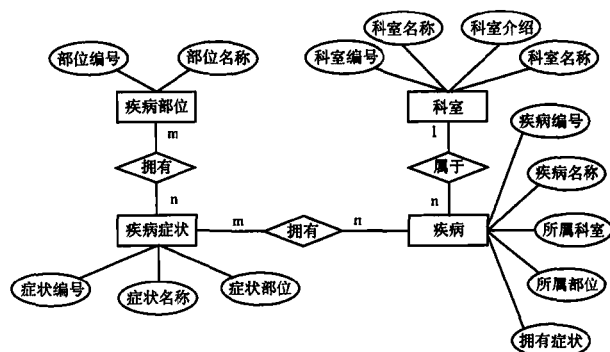


图 4 实体联系

例 3：对照关系表是存储疾病、部位、症状以及科室之间对应关系的表，其属性包括疾病编号、

部位、症状和科室编号。具体结构，见表 2。

表 2 对照关系

名称	类型	说明	备注
DiseaseID	int (10)	疾病编号	主键
PartID	int (10)	部位编号	主键
SymptomID	int (10)	症状编号	主键
DepartNO	int (10)	部门编号	主键

6 结语

本研究运用杰卡德相似系数实现智能导诊系统构建。用户登录系统选择相应症状后^[10]系统会给出用户可能患有的疾病和推荐的诊疗科室，从而满足用户的导诊需求，在一定程度上克服医院现有分诊流程存在的弊端，缓解医院导诊服务压力。将该系统移植到移动端时，鉴于内存的限制考虑采用矩阵分块技术实现对照关系表的内存载入。

参考文献

- 1 周越. 医院信息化建设对医院整体发展的意义 [J]. 电子技术与软件工程, 2018, 16 (1): 226.
- 2 李俊茹, 何晓俐, 邓爱萍. 超大型医院门诊分诊创新服务模式探讨 [J]. 四川医学, 2013, 34 (12): 1982 - 1983.
- 3 徐若然, 张毅, 周博雅, 等. 智慧医院建设中信息平台的构建与应用研究 [J]. 中国医院管理, 2018, 38 (3): 55 - 57.
- 4 Salton G. The SMART Retrieval System - Experiments in Automatic Document Processing [M]. Englewood Cliffs: Prentice Hall Inc, 1971.
- 5 王林, 张一帆. 基于节点相似度的有向网络社团检测算法 [J]. 微型机与应用, 2017, 36 (3): 19 - 22.
- 6 刘军娜, 刑琪, 赵卫东. 程序相似度检测算法 [J]. 计算机与数字工程, 2015, 43 (12): 2145 - 2149.
- 7 陈维. 医疗专家云诊断系统的设计与实现 [D]. 西安: 电子科技大学, 2016.
- 8 Zadeh L A. Fuzzy Sets as a Basic for a Theory of Possibility [J]. Fuzzy sets and Systems, 1978, 1 (7): 3 - 28.
- 9 韩如冰, 叶得学. 基于 VSM 的权重改进文档相似度算法研究 [J]. 软件, 2012, 33 (10): 103 - 105.
- 10 毛宇. 中医药症状的中文分词与句子相似度研究 [D]. 杭州: 浙江大学, 2017.