

统一医疗系统的设计与实现

童刚,刘焕,姜宁

(青岛科技大学 信息科学技术学院 山东 青岛 266061)

摘要:充分利用各个医疗系统的相关医疗数据资源优势,创建可视化、可操作化、可挖掘化的统一医疗系统。通过云计算技术对各个医疗系统的数据实现共享,数据仓库的创建,数据挖掘算法的选择与嵌入和用户端的设计,进而完成整个系统架构的设计。创建多系统集合、多数据类型、可数据处理的适用性更强的统一医疗系统。统一医疗系统的设计与实现,充分利用了云计算、数据仓库及数据挖掘等技术,将医院的各个系统的数据资源整合到一起,避免了医疗数据资源的浪费,为挖掘阶段提供了强有力的数据支持,并为从大规模医疗临床数据中发现和提炼进而形成临床有用知识提供了技术支持,能大大推动医学知识创新的开展以及推动医学领域的进步。

关键词:数据共享;数据仓库;算法选择;决策树;统一医疗系统

中图分类号: TP399

文献标识码: A

文章编号: 1673-629X(2019)02-0190-05

doi: 10.3969/j.issn.1673-629X.2019.02.040

Design and Implementation of Unified Medical System

TONG Gang, LIU Huan, JIANG Ning

(School of Information Science and Technology, Qingdao University of Science and Technology,
Qingdao 266061, China)

Abstract: Making full use of the advantages of relevant medical data resources in each medical system, we create a unified medical system which is visual, operable and tapped. Through cloud computing technology, realizing data sharing of each medical system, creation of data warehouse, selection and embedding of data mining algorithm, and design of user end, the whole system architecture is completed. A unified medical system with multiple system sets, multiple data types and stronger applicability of data processing is created. The research and design of unified medical system makes full use of cloud computing, data warehouse and data mining technology to integrate data resources in each medical system, avoiding the waste of medical data resources, providing a powerful support for the data mining and a technical support for discovering, extracting and then producing clinical useful knowledge from the large-scale medical and clinical data, which can greatly promote the development of medical knowledge innovation and progress in the field of medicine.

Key words: data sharing; data warehouse; algorithm selection; decision tree; unified medical system

0 引言

近年来,医疗信息化的发展已经推动了整个医疗事业的进步,医院的医疗系统可以看作一个有机的大整体,包括医院管理系统(HIS)、体检管理系统(PEIS)、检验信息系统(LIS)、影像信息系统(PACS/RIS)、电子病历系统(EMR)、移动护理系统等等。目前的现象是各个医疗系统各自为战,浪费了很多不同类型的医疗数据资源,因此创建一个统一的医疗系统就显得意义重大。将医院的各个信息系统进行有机结合,实现信息共享,进而结合数据挖掘技术,为其提供

强有力的数据支持进而协助医生辅助诊断及发现一些潜在的规则联系,其潜在的应用价值将推动整个医疗领域的信息化发展^[1]。

1 医疗系统的架构设计

系统架构大致分为四层:源数据层、医院综合数据仓库层、挖掘内核层及可视化界面层。

源数据层就是将医院内包括 HIS、EMI、LIS、PACS、CIS 及一些其他的财务后勤、病案等数据资源进行共享作为源数据,根据需要对其进行初步的数据

收稿日期: 2018-03-07

修回日期: 2018-07-10

网络出版时间: 2018-11-15

基金项目: 国家自然科学基金(61572268)

作者简介: 童刚(1962-),男,副教授,研究方向为工业自动化、信息管理、智能交通;刘焕(1992-),女,硕士,CCF 会员(74967G),研究方向为数据挖掘。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20181115.1047.034.html>

预处理,包括数据抽取、数据清洗、数据整合与加载,为构成医院综合数据仓库层提供数据源^[2]。

医院综合数据仓库层是为医院量身定制的企业级的数据仓库,该数据仓库将成为既支持医院运营管理的管理业务数据中心,又能整合病人的各项临床数据,成为未来支持各类临床科研的临床数据中心,从而实现对医院各业务系统数据和科研数据的集中治理并为挖掘层提供了强有力的数据支持^[3]。

挖掘内核层是整个挖掘系统的核心,该层以医院综合数据仓库层作为数据支撑,从中结合数据挖掘技术,筛选出适用性最强的算法进行嵌入,进而实现辅助医疗诊断及潜在规则的挖掘功能。

可视化界面层即用户层,便于用户根据自己的需求输入并获得对应的挖掘结果。

统一医疗系统的整体架构设计如图1所示。

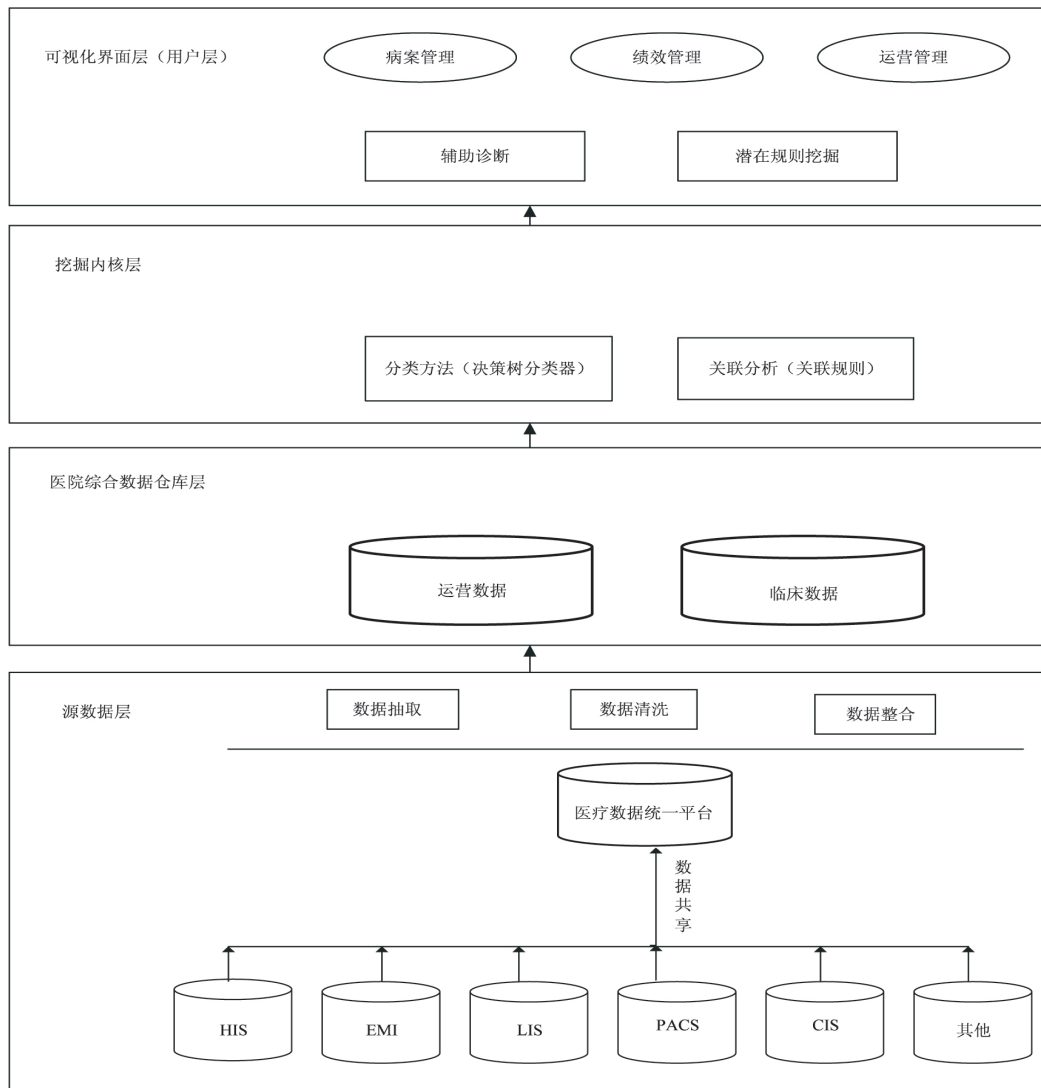


图1 统一医疗系统的架构

在整个系统搭建过程中,存在三个创新架构方面:数据共享的实现;医院综合数据仓库的建设;挖掘内核层中算法的筛选。

2 医疗系统的架构创新

2.1 数据共享

云计算是基于互联网的服务的增加、使用和交付模式,通常涉及通过互联网来提供动态易扩展且经常是虚拟化的资源^[4]。近年来,随着医疗信息化的发展以及医疗信息系统的普及,大部分医院均已配备了属

于自己的信息系统。由于这些系统深深地受地理位置、区域经济以及人才结构等的影响,所以导致各个医院的信息系统的规模及功能存在较大的差异。为了缩小这种差异以及更好地满足病患对医疗服务质量日益增长的要求,将云计算技术应用于医疗信息系统中,这样既可以将不同区域的医疗数据整合到一起,也可以建立标准化的信息共享平台,进而从根本上消除“信息孤岛”。医院的数据共享云平台由三部分组成,即患者信息索引、数据集成网关及医疗云服务器^[5]。

患者信息索引:云存储中的索引文件主要用来存

储就医者的医疗档案信息,它记录了每位患者在不同医院的就诊数据并对其进行统一编码,从而使各医院的医疗数据得以共享进而实现分工协作。其中索引数据主要由能够唯一标识患者身份及就诊记录的相关信息组成,如:患者身份证号码、就诊时间、医疗机构名称等^[6]。医院相关工作人员便可以利用该索引对每个患者的医疗信息进行检索,最终以可视化的结构展现出来。

数据集成网关:目前各医院均存在相当数量的异构平台和信息系统,这些系统并没有统一的数据交换准则,如何为不同的医疗信息系统提供一个既方便又灵活的信息交换准则成为了一项重要的研究课题。这里采用多标准兼容医疗网关。该网关是一种较为机敏的系统集成机制,具有多标准的特色同时也可以实现协议兼容。

医疗云服务器:云服务器实现了数据的提供及存储功能,它不仅提供很大的存储空间而且扩容也相当简便,因此可以存储共享后的海量医疗数据并对其进行备份。其中的云数据库既可以实现融合各医院的数据库,也可在融合的基础上实现数据的增删改查操作。

将云计算嵌入到医疗领域当中,不仅大大地提升了医疗信息的利用率,同时也为医疗信息系统的维护带来了便利。云计算在医疗数据共享中的应用能够使患者享受到更便捷更优质的服务,同时共享后的海量医疗数据能够为数据挖掘阶段提供强有力的数据支持,这也是实现整个统一医疗系统的主要课题之一^[7]。

2.2 医院综合数据仓库的建立

数据仓库的作用分为以下三点:分析和预测、支持多维分析、利用最宝贵的业务数据做出最明智的商业决策。具体功能内容如下:

(1) 数据仓库支持多维分析。一个实体由若干个维度属性共同描述,这样便增强了数据的丰富度,也增强了数据仓库的信息处理能力。

(2) 数据仓库中的数据来自于企业内部各不同的源业务系统中,这些数据是经过业务分析并且根据一定的业务逻辑经过了清洗转换之后而集成得到的。因此数据仓库可以为数据挖掘技术提供更加丰富且质量更高的数据。

(3) 数据挖掘技术是在数据仓库已有数据的基础上,挖掘出数据的隐藏价值,给整个企业的发展和未来前景做出较为完整、合理、准确的分析和预测。

目前医院各系统现阶段面临的问题包括:业务的快速发展、数据的急速增长以及随着政策的放开,医疗业务的信息化发展等多方面影响,使得医疗数据急速增长,为了保障医院系统正常运营,各大医疗机构也不断地采取措施,如增加在信息化方面的投入、购买服务

器等来满足医院数据的急速增长。除海量数据之外,“信息孤岛”也是一个亟待解决的问题,它并非一个普遍的问题,也并非中国信息化所特有的问题,而是全球企业信息化发展必经的时期。此外,医院有众多的信息系统,但是各信息系统之间并没有进行很好的数据共享及数据交换,还只是停留在完成业务的基础上,而且医院内各个信息系统水平参差不齐,数据质量急待加强。基于以上调研发现的问题,可以看出,医院现阶段构建集成平台数据仓库已经迫在眉睫。

医院综合数据仓库的建立是一个周而复始的历程,主要包括数据的选择、变换、建模、评估、解释模型、运用和巩固模型等几个阶段^[8]。

确定主题:医院机构通过对要研究的主题进行确定进而组织医院综合数据仓库的数据,所以确定主题对医院综合数据仓库的建设尤为重要,不仅要确定主题,而且要确定相应的子主题、维度及数据来源。

数据准备:由于数据挖掘的结果受数据质量的影响极大,因此数据准备阶段显得尤为重要。该阶段主要包括数据的选取(尽可能选取关联度强的数据)、探索(充分了解数据的分布情况及异常数据等)、修正(对缺失值进行填充及对错误数据进行纠错)、变换(将离散值变换成连续值以及数据项间的计算合并等)。

建立模型:它是建立医院综合数据仓库中最为重要的一步,首先选择挖掘工具提供的某种算法,将选取出来的算法应用在准备好的数据中,设定参数,最后完成整个模型的建设。

评估、解释模型:该步骤主要是对以上建立的模型不断地进行评估比较,进而生成一个相比较而言最好的模型,接着用言语对“最好的模型”进行解释说明,若不存在问题,就可以对其进行试验性的应用。反之,继续进行上面的操作直至生成较为满意的模型。

运用、巩固模型:将评估之后选出来的“最优模型”投入到实际应用中并对其监测,若在此过程中存在不好的情况,应对其进行考察并反映其业务规律的变化。

医院综合数据仓库是通过数据库引擎或者应用软件将来自不同系统不同架构的数据库中的数据按照相应的规则进行转换、抽取形成的^[9]。医院综合数据仓库中既包含关系型数据库,也包含面向不同主题的多维数据模型,像门诊量分析的模型就有时间、病人身份信息、疾病等若干个分析维度。医院的管理决策指标体系中就包含这些维度模型。

通过数据仓库的构建,最终提供了对现有以及历史数据的分析,并且为医疗系统的挖掘内核层提供强有力的数据支撑打下了可靠的数据基础^[10]。

2.3 算法的选择

2.3.1 辅助诊断功能的算法筛选

在挖掘内核层最关键的技术是算法的嵌入,即针对医疗数据自身的独特性(冗余性、多样性、不完整性)及需求筛选出合适的算法进而实现辅助诊断及潜在规则的挖掘。这个步骤是整个挖掘过程的精髓所在^[11]。

辅助诊断功能体现在分类方法上,使用大型三甲医院的医疗数据库,利用这些数据分别采用决策树、神经网络、支持向量机、贝叶斯这四种分类挖掘算法进行挖掘实验对比分析,从中选取一个综合性能最高的算法作为分类器嵌入到挖掘内核层中,原始数据样例部分截图如图2所示。

211327	14761	41943	静配配药	三升袋	注射用复方维生素3【(爱尔维泰)注射用复方维	Y000000016982	CZ	长期医嘱	0	1	20 mg	QD
211328	14761	41943	静配配药	三升袋	代甲【代甲(门冬氨酸钾注射液)】	Y000000016238	CZ	长期医嘱	0	1	20 ml	QD
211329	14761	41943	静配配药	三升袋	【胰岛素注射液】	Y000000002517	CZ	长期医嘱	0	1	28 IU	QD
211330	14761	41943	静配配药	三升袋	佳乐同怡【佳乐同怡(多种微量元素II注射液)】	Y000000001018	CZ	长期医嘱	0	1	10 ml	QD
211331	14761	41943	静配配药	三升袋	洛安命【(洛安命)5%复方氨基酸18AA注射液】	Y000000016113	CZ	长期医嘱	0	1	12.5 g	QD
211332	14761	41943	静配配药	三升袋	澳宜林【(澳宜林)复方氨基酸15双肽2液】	Y000000001562	CZ	长期医嘱	0	1	67 g	QD
211333	14761	41943	静配配药	三升袋	欣坤畅【欣坤畅(注射用丙氨酸谷氨酰胺)】	Y000000002393	CZ	长期医嘱	0	1	20 g	QD
211334	14761	41943	静配配药	三升袋	【软包0.9%氯化钠注射液】	Y000000001807	CZ	长期医嘱	0	1	500 ml	QD
211335	14761	41943	静配配药	三升袋	【50%葡萄糖注射液】	Y000000000088	CZ	长期医嘱	0	1	80 ml	QD
211336	14761	41943	静配配药	三升袋	世新【世新(20%中/长链脂肪乳注射液C8~24)】	Y0000000016142	CZ	长期医嘱	0	1	250 ml	QD
211337	14761	41943	静配配药	三升袋	维力能【(维力能)醋酸纳林格注射液】	Y0000000016103	CZ	长期医嘱	0	1	500 ml	QD
211338	14761	41943	静配配药	三升袋	海斯维【海斯维(转化糖电解质注射液)】	Y0000000000874	CZ	长期医嘱	0	1	500 ml	QD
211339	14761	41943	静配配药	三升袋	注射用复方维生素3【(爱尔维泰)注射用复方维	Y000000016982	CZ	长期医嘱	0	1	20 mg	QD
211340	14761	41943	静配配药	三升袋	代甲【代甲(门冬氨酸钾注射液)】	Y000000016238	CZ	长期医嘱	0	1	20 ml	QD
211341	14761	41943	静配配药	三升袋	【胰岛素注射液】	Y000000002517	CZ	长期医嘱	0	1	28 IU	QD
211342	14761	41943	静配配药	三升袋	佳乐同怡【佳乐同怡(多种微量元素II注射液)】	Y000000001018	CZ	长期医嘱	0	1	10 ml	QD
211343	14761	41943	静配配药	三升袋	澳宜林【(澳宜林)复方氨基酸15双肽2液】	Y000000001562	CZ	长期医嘱	0	1	67 g	QD
211344	14761	41943	静配配药	三升袋	【50%葡萄糖注射液】	Y000000000088	CZ	长期医嘱	0	1	80 ml	QD
211345	14761	41943	静配配药	三升袋	世新【世新(20%中/长链脂肪乳注射液C8~24)】	Y0000000016142	CZ	长期医嘱	0	1	250 ml	QD
211346	14761	41943	静配配药	三升袋	维力能【(维力能)醋酸纳林格注射液】	Y0000000016103	CZ	长期医嘱	0	1	500 ml	QD
211347	14761	41943	静配配药	三升袋	海斯维【海斯维(转化糖电解质注射液)】	Y0000000000874	CZ	长期医嘱	0	1	500 ml	QD
211348	14761	41943	静配配药	三升袋	代甲【代甲(门冬氨酸钾注射液)】	Y000000016238	CZ	长期医嘱	0	1	20 ml	QD
211349	14761	41943	静配配药	三升袋	【胰岛素注射液】	Y000000002517	CZ	长期医嘱	0	1	28 IU	QD
211350	14761	41943	静配配药	三升袋	佳乐同怡【佳乐同怡(多种微量元素II注射液)】	Y000000001018	CZ	长期医嘱	0	1	10 ml	QD
211351	14761	41944	中心配药	三升袋	澳宜林【(澳宜林)复方氨基酸15双肽2液】	Y000000001562	LZ	临时医嘱	0	1	67 g	QD
211352	14761	41944	中心配药	三升袋	【软包0.9%氯化钠注射液】	Y000000001807	LZ	临时医嘱	0	1	500 ml	QD

图2 初始医疗数据截图

经过数据预处理之后,利用 weka 挖掘工具分别选取决策树、神经网络、支持向量机、贝叶斯四种分类挖掘算法对肿瘤数据进行挖掘操作,实验结束后,得到了四种算法的准确率及运算效率,分别如表1和表2所示。

表1 算法(病症)计算精度对比

算法名称	测试集准确率/%	排名
决策树算法	88.97	3
贝叶斯算法	88.21	4
支持向量机算法	92.03	2
神经网络	92.88	1

表2 算法(病症)运行时间效率对比

算法名称	建模时间/s	测试时间/总执行时间/s	排名	
决策树算法	0.08	0.01	0.09	1
贝叶斯算法	0.2	0.12	0.32	2
支持向量机	1.33	0.15	1.48	3
神经网络	3.06	0.1	3.16	4

通过以上结果可以看出,神经网络算法以及支持向量机算法在分类的准确率上要高一些,但是在表2中可以看到它们的运行时间效率要比另外两种算法低很多,尤其是神经网络算法,它的效率要比决策树算法慢大约三倍。表1中,虽然决策树算法以及贝叶斯算法在分类精确度上稍低于另外两种算法,但是在表2中,它们在时间效率上比神经网络算法和支持向量机算法要高很多。综合来看,决策树算法的精确率与另外三种算法的精确率相差较小,但是它的运算效率却

远远超过其他算法,因此决策树算法具有较高的综合性能,适用于作为创建肿瘤疾病的辅助诊断治疗的分类器。

2.3.2 挖掘潜在规则算法的选择

关联规则挖掘是用来发现大量数据中项集之间有趣的关联或相关联系的方法。关联规则算法对数据预处理的要求较低,且易于理解和操作,更重要的是可以挖掘多个变量之间的联系,无需对其变量进行区分等^[12]。基于以上关联规则的主要优势及功能以及肿瘤数据自身的特点,选取它作为实现医疗潜在规则挖掘的主要算法。

这里以肿瘤疾病的数据为例,使用 Apriori 算法对其进行关联规则挖掘,最开始先缩减肿瘤数据的属性维度,去掉一些无关紧要的属性维度,之后将肿瘤数据分为症状表以及治疗方式表,并依次进行挖掘实验。这里将 Apriori 算法的最小支持度设为 0.1,最小置信度设为 0.9,最大支持度设为 1,最终从结果中只选取排在前十位的规则。图3和图4分别是症状及治疗方式 Apriori 算法挖掘对应的实验结果。

从挖掘结果中可以得到“肋骨疼痛→髋部疼痛”、“胃部疼痛→嗝气”症状之间有强关联关系。“过敏→降血脂固醇”、“调节心率→降血脂固醇”治疗方案之间有强关联关系。经病理探究及药理实验对挖掘结果的检验,证明发现的关联关系大多具有一定的可解释性,这表明关联规则算法在发现医疗潜在规则工作上具有一定的参考价值。

Best rules found:

- 1.Leigutengtong=No ==> Kuanbutengtong=No <conf:(1)> lift:(1) lev:(0) [0] conv(0)
- 2.Kuanbutengtong=No ==> Leigutengtong=No <conf:(1)> lift:(1) lev:(0) [0] conv(0)
- 3.Yanteng=No ==> Kuanbutengtong=No <conf:(1)> lift:(1) lev:(0) [0] conv(0)
- 4.Yanteng=No ==> Leigutengtong=No <conf:(1)> lift:(1) lev:(0) [0] conv(0)
- 5.Weibutengtong=No ==> Aiqi=No <conf:(1)> lift:(1.02) lev:(0.02) [0] conv(0.98)
- 6.Aiqi=No ==> Weibutengtong=No <conf:(1)> lift:(1.02) lev:(0.02) [0] conv(0.98)
- 7.Aiqi=No ==> Kuanbutengtong=No <conf:(1)> lift:(1) lev:(0) [0] conv(0)
- 8.Aiqi=No ==> Leigutengtong=No <conf:(1)> lift:(1) lev:(0) [0] conv(0)
- 9.Shengyinsiya=No ==> Kuanbutengtong=No <conf:(1)> lift:(1) lev:(0) [0] conv(0)
- 10.Shengyinsiya=No ==> Leibutengtong=No <conf:(1)> lift:(1) lev:(0) [0] conv(0)

图 3 Apriori 算法症状挖掘结果规则

Best rules found:

- 1.kangyan=wu ==> jiangxuezhiguchun=wu <conf:(1)> lift:(1.02) lev:(0.02) [0] conv(0.9)
- 2.kangguomin=wu ==> jiangxuezhiguchun=wu <conf:(0.98)> lift:(1) lev:(-0) [0] conv(0.48)
- 3.baowei=wu ==> jiangxuezhiguchun=wu <conf:(0.98)> lift:(1) lev:(-0) [0] conv(0.48)
- 4.fangliaobuwei=wu ==> jiangxuezhiguchun=wu <conf:(0.98)> lift:(1) lev:(-0) [0] conv(0.48)
- 5.tiaojixinxlv=wu ==> jiangxuezhiguchun=wu <conf:(0.98)> lift:(1) lev:(-0) [0] conv(0.46)
- 6.kangguomin=wu baowei=wu ==> jiangxuezhiguchun=wu <conf:(0.98)> lift:(1) lev:(-0) [0] conv(0.46)
- 7.kangguomin=wu fangliaobuwei=wu ==> jiangxuezhiguchun=wu <conf:(0.98)> lift:(1) lev:(-0) [0] conv(0.46)
- 8.baowei=wu fangliaobuwei=wu ==> jiangxuezhiguchun=wu <conf:(0.98)> lift:(1) lev:(-0) [0] conv(0.46)
- 9.jiangxuezhiguchun=wu ==> kangguomin=wu <conf:(0.96)> lift:(1) lev:(-0) [0] conv(0.65)
- 10.jiangxuezhiguchun=wu ==> baowei=wu <conf:(0.96)> lift:(1) lev:(-0) [0] conv(0.65)

图 4 Apriori 算法药物挖掘结果

3 结束语

统一医疗系统的研究与开发,充分利用了医院各个系统的数据资源,做到了资源不浪费,协同起来为挖掘需要的知识提供强有力的数据支持,与此同时,结合要达到的辅助医生诊断及症状关联分析的目的进而筛选出决策树及关联规则算法嵌入到挖掘内核层中,不仅极大地提高了诊断的准确率及诊断效率,也能从海量的医疗数据库中挖掘出潜在有价值的医疗规则。该系统的设计与实现大大推动了医学知识创新的开展,推动了医学信息化及现代化的进步。

但是目前尚在起步阶段,仍有很多不足有待探索提升,基于当前不足提出以下几点展望:

研究开发更多的医药领域的数据挖掘应用算法集成至系统,以满足更多的需求分析;针对挖掘分析结果的展示,丰富图形界面,提高系统的用户体验和可用性,挖掘结果的可视化是数据挖掘领域的一个重要研究方向,对数据挖掘技术的应用起着至关重要的作用^[13]。

参考文献:

- [1] 袁琛.浅谈大数据技术在医疗信息化中的应用[J].医疗装备,2016,29(2):12-13.
- [2] ZHANG Wenjing, MA Donglai, YAO Wei. Medical diagnosis data mining based on improved apriori algorithm[J]. Journal

of Networks, 2014, 9(5):1339-1345.

- [3] 王勤,祝小芳.基于信息系统的智慧医疗体系框架及关键技术[J].华东科技:学术版,2017(12):7.
- [4] 张威,刘霞.基于云计算的医疗信息系统结构设计与研究[J].中国科技信息,2012(24):98.
- [5] 唐维维.基于云计算的区域医疗信息数据共享平台的设计与实现[D].北京:中国人民解放军医学院,2015.
- [6] 李浩,薛镭,朱岩.基于云计算的医疗信息共享体系[J].中国卫生信息管理杂志,2011(3):65-69.
- [7] 关静静,贺鹏涛.基于云计算的区域医疗数据共享平台建设思考[J].通讯世界,2017(6):55-56.
- [8] 王冬冬.基于openEHR的临床医疗数据仓库构建方法研究与实践[D].杭州:浙江工业大学,2015.
- [9] LI Ping, WU Tao, CHEN Mu, et al. A study on building data warehouse of hospital information system[J]. 中华医学杂志:英文版,2011,124(15):2372-2377.
- [10] 李波.中医临床数据挖掘系统的研究与实现[D].北京:北京交通大学,2011.
- [11] 丁中正,刘云,景慎旗,等.医疗数据挖掘综述[J].智慧健康,2016(10):54-56.
- [12] 刘木林,朱庆华.基于Hadoop的关联规则挖掘算法研究-以Apriori算法为例[J].计算机技术与发展,2016,26(7):1-5.
- [13] GUO Jianwei, LI Ying, DU Liping, et al. Research on distributed data mining system based on Hadoop platform[M]. Berlin: Springer, 2014.