

学校编码: 10384

分类号_____密级_____

学号: X2013231731

UDC _____

廈門大學

工 程 碩 士 學 位 論 文

醫療數據管理系統的設計與實現

Design and Implementation of Medical Data Management
System

鄭祥聰

指 導 教 師: 張志宏 助理教授

專 業 名 稱: 軟 件 工 程

論 文 提 交 日 期: 2015 年 10 月

論 文 答 辯 日 期: 2015 年 11 月

學 位 授 予 日 期: 2015 年 12 月

指 導 教 師: _____

答 辯 委 員 會 主 席: _____

2015 年 10 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

近年来,大数据的快速发展及落地,为各行各业带来了前所未有的冲击,为医疗行业的信息化带来了新的解决方案。目前,传统的医院在信息化系统及电子化健康档案系统方面已远远不能满足海量医疗数据的存储处理、分析挖掘以及多元化医疗数据服务等方面的要求。医疗数据经过生成及采集后,如何将异构、海量、实时、多样的数据进行高效存储,以实现大规模复杂健康数据的快速查询和准确响应;如何通过数据挖掘手段,对海量医疗数据进行深度分析,快速有效为用户提供有价值的医疗数据服务,达到未病先防的效果,成为大数据时代医疗领域需要解决的问题。

针对个性化的健康照护计划以及疾病并发预测的需求,提出了医疗数据的关联性挖掘算法,包括基于贝叶斯网络的健康状况推理服务实现和基于关联规则的疾病并发预测服务,应用拓展可以向个人用户提供疾病预警及防范建议,同时可用于对医疗机构作为临床决策的辅助及疾病并发的分析研究。本文在所设计的医疗数据管理系统上,将海量、多样的医疗数据进行高效存储,实现大规模复杂健康数据查询的快速和准确响应;通过对现有数据集的分析挖掘,研究病人的各种检查结果和身体状况与个体所得病症的相关性,以及疾病与疾病之间的关联性,利用所得的相关性为病人提供健康状况的推理服务。

应用软件工程的理论和技术,通过对当前存在的问题如医疗数据量大、没有充分利用、医生队伍紧缺、没有深入分析医疗数据等问题和情况进行深入分析,针对某医院在医疗数据管理工作中的实际需求,完成系统的分析和设计,构建医疗数据管理系统,利用数据挖掘技术,为病人提供健康状况的数据分析服务。论文基于海量医疗数据的特点,以及现有海量医疗数据挖掘应用服务的不足,提出了医疗数据管理系统的设计,其中包括医疗数据管理系统模型构建,医疗数据的快速统计查询方法设计,以及医疗数据的关联性挖掘服务。

关键词: 医疗数据; 医疗信息化; 数据挖掘

Abstract

In recent years, the rapid development of large data and landing, for all walks of life has brought an unprecedented impact for the medical industry, the information technology has brought a new solution. At present, the traditional hospital information system and electronic health records system has been far from meeting the requirements of mass medical data storage, analysis and data services and other medical services. After the generation and collection of medical data, how to store heterogeneous, massive, real-time and diverse data is stored in order to realize the fast query and accurate response of large scale complex health data. How to carry out the analysis of massive medical data by means of data mining, and provide valuable medical data services to users.

For personalized health care planning and the need to predict the disease, the association mining algorithm based on Bayesian network is proposed, which is based on Bayesian network. It can provide a disease warning and prevention. In this paper, data management system, the massive and diverse medical data is stored in order to realize the fast and accurate response of large scale complex health data query. Through the analysis of the existing data sets, the paper studies the relationship between the different kinds of patients.

By using the theory and technology of software engineering, the problems and situations such as medical data, lack of medical data, lack of medical data, analysis and design of medical data management system, construct data analysis service for patients with data mining technology. Based on the characteristics of massive medical data and the shortage of existing massive medical data mining applications, this paper presents the design of medical data management system, which includes the design of medical data management system model, the design of medical data and the data mining service.

Key words: Medical data; Medical Information; Data Mining

目 录

第一章 绪论	1
1.1 研究背景	1
1.2 国内外研究现状	3
1.2.1 医疗大数据研究现状.....	3
1.2.2 数据挖掘在医疗行业应用现状.....	5
1.3 论文研究目的和意义	7
1.4 论文研究内容	8
1.5 论文组织结构	9
第二章 相关技术研究	10
2.1 Hadoop 技术概述	10
2.2 Java 语言	12
2.3 数据挖掘技术	12
2.3.1 数据挖掘简介.....	12
2.3.2 数据挖掘流程.....	13
2.3.3 数据仓库.....	13
2.4 本章小结	15
第三章 需求分析	16
3.1 用户角色分析	16
3.2 功能性需求分析	16
3.2.1 健康档案子系统.....	17
3.2.2 疾病管理子系统.....	18
3.2.3 门诊和住院信息管理子系统.....	19
3.2.4 影像数据管理子系统.....	20
3.2.5 诊断分析子系统.....	21
3.2.6 系统配置管理子系统.....	22
3.3 非功能性需求分析	23
3.3.1 数据精度要求.....	23
3.3.2 响应时间要求.....	23
3.3.3 安全与保密.....	24
3.3.4 系统易用性要求.....	24

3.3.5 其他需求.....	24
3.4 运行环境分析	25
3.4.1 硬件环境.....	25
3.4.2 软件环境.....	25
3.5 本章小结	25
第四章 系统设计	26
4.1 设计目的和原则	26
4.2 系统架构设计	27
4.3 功能结构设计	30
4.3.1 健康档案管理子系统.....	31
4.3.2 疾病管理子系统.....	32
4.3.3 门诊和住院信息管理子系统.....	33
4.3.4 诊断分析管理子系统.....	33
4.3.5 影像数据管理子系统.....	34
4.3.6 系统配置管理子系统.....	35
4.4 系统流程设计	36
4.4.1 数据挖掘流程.....	36
4.4.2 数据查询流程.....	38
4.4.3 数据上传流程.....	39
4.5 系统类图设计	40
4.6 数据库设计	42
4.6.1 E-R 图.....	43
4.6.2 数据字典.....	46
4.7 数据挖掘设计	59
4.7.1 基于贝叶斯网络的疾病与参数关联性分析.....	59
4.7.2 基于关联规则的疾病与疾病关联性挖掘.....	61
4.8 本章小结	63
第五章 系统实现与测试	65
5.1 功能实现	65
5.1.1 系统登陆.....	65
5.1.2 健康档案管理子系统.....	67
5.1.3 疾病管理子系统.....	72
5.1.4 门诊和住院信息管理子系统.....	73
5.1.5 诊断分析管理子系统.....	76

5.1.6 影像数据管理子系统.....	78
5.1.7 系统配置管理子系统.....	79
5.2 系统测试	81
5.2.1 测试原则.....	81
5.2.2 测试用例及结果.....	82
5.3 本章小结	85
第六章 总结与展望	86
6.1 总结	86
6.2 展望	86
参考文献	88
致 谢.....	90

厦门大学博硕士学位论文摘要库

Contents

Chapter 1 Introduction.....	1
1.1 Research Background.....	1
1.2 Research Status	3
1.3 Research Purpose and Significance.....	7
1.4 Research Contents	8
1.5 Outline of the Dissertation	9
Chapter 2 System Related Technologies	10
2.1 Hadoop	10
2.2 Java.....	错误!未定义书签。
2.3 Data Mining.....	12
2.3.1 Introduction to data mining.....	12
2.3.2 Data mining process.....	13
2.3.3 Data warehouse	13
2.4 Summary.....	15
Chapter 3 Requirement Analysis.....	16
3.1 User Role Analysis.....	16
3.2 Function Requirement Analysis.....	16
3.2.1 Health File Subsystem	17
3.2.2 Disease management subsystem.....	18
3.2.3 Information Management Subsystem	19
3.2.4 Image data management subsystem.....	20
3.2.5 Diagnostic Analysis Subsystem	21
3.2.6 System Management Subsystem.....	22
3.3 Non-function Requirement Analysis	23
3.3.1 Data Accuracy.....	23
3.3.2 Response Time.....	23
3.3.3 Security And Confidentiality	24
3.3.4 Usabilit.....	24
3.3.5 Others	24
3.4 Runtime environment analysis	25
3.4.1 Hardware environment.....	25
3.4.2 Software environment.....	25

3.5 Summary.....	25
Chapter 4 System Design	26
4.1 Design Purpose and Principle	26
4.2 System Framework Design.....	27
4.3 System Function Design	30
4.3.1 Health File Subsystem	31
4.3.2 Disease management subsystem	32
4.3.3 Information Management Subsystem	33
4.3.4 Image data management subsystem.....	33
4.3.5 Diagnostic Analysis Subsystem	34
4.3.6 System Management Subsystem.....	35
4.4 System Process Design	36
4.4.1 Data mining process.....	36
4.4.2 Data query process	38
4.4.3 Data upload process	39
4.5 System class diagram design	40
4.6 DB Design	42
4.6.1 E-R Diagram	43
4.6.2 Data Dictionary	46
4.7 Data mining design	59
4.8 Summary.....	63
Chapter 5 System Implementation and Test	65
5.1 Function Implementation.....	65
5.1.1 System Login	65
5.1.2 Health File Subsystem	67
5.1.3 Disease management subsystem	72
5.1.4 Information Management Subsystem	73
5.1.5 Image data management subsystem.....	76
5.1.6 Diagnostic Analysis Subsystem	78
5.1.7 System Management Subsystem.....	79
5.2 System Test	81
5.2.1 Test Principle.....	81
5.2.2 Test Case and Result.....	82

5.3 Summary.....	85
Chapter 6 Conclusions and Future Work.....	86
6.1 Conclusions.....	86
6.2 Future Work.....	86
References.....	88
Acknowledgements.....	<u>90</u>

厦门大学博硕士学位论文摘要库

第一章 绪论

1.1 研究背景

随着信息技术的飞速发展，一个大规模数据生产、传播、应用的时代已经开启。无处不在的社交网络、电子商务、移动通信加速推进着数据的生产及传播。大数据，将逐步发展成为像铁路、公路、水电一样，成为社会发展必不可少的社会基础设施。大数据的意义不光体现在数据量之“大”，更为重要的价值是隐藏在内部的“金子”^[1]。云计算，是征服海量数据、挖掘数据价值最重要的计算工具。以云计算为基础的信息存储、分享和挖掘技术，可以绿色高效地解决海量、高速、多变的数据存储，数据分析及挖掘计算^[2]。教育、交通、医疗保健等与民生息息相关的领域，借助云计算、大数据挖掘技术等种创新技术，必将会给这些行业带来跨越式的发展。

近年来，随着医疗信息化的逐步发展，绝大部分医院先后建立起数字化医疗信息系统及患者电子健康档案。但目前大部分系统仅限于医疗机构内部使用，造成了较为严重的医疗信息孤岛。在国家政策的统一规划下，十二五卫生信息化建设规划明确了未来卫生信息化发展路线图，即 3521 工程，将逐步建设国家、省、地市 3 级卫生信息平台，加强公共卫生方面、医疗服务方面、新农村合作医疗方面、基本药物制度方面以及综合管理 5 方面的业务应用，建设居民健康档案和电子病历 2 个基础数据库，以及 1 个专用网络^[3]。该工程所产生、汇合的医疗数据量之大将令人难以置信，一张 CT 图像大约 150MB，一个基因组序列约 750MB，一张病理图接近 5GB，在此单位数据量的基础上乘以人口数量和平均寿命，仅单个医院就可生成和累积 TB 级以上数据。在医疗数据采集方面，随着区域医疗信息系统、医疗数据中心及数据交换平台的建设，包含结构化和非结构化的医疗及健康数据将不断扩容并呈现几何级增长。同时，医疗数据的产生不再局限于医院的单一环境，也将包括来自于家庭、私人诊所、实验检验中心、医院/乡镇卫生院、急救中心、体检中心等不同机构带来的多样的数据源。且随着物联网技术及可穿戴设备的不断发展，我们可以随时随地采集自身医疗数据^[4]。医疗数据将不再是某家医疗机构专有资源，它可向区域中以及更上层的大区域直至国家级的医疗健康信息系统提供数据共享及交换。

在医疗数据服务方面,当前我国的医疗卫生服务体系仍然存在着医疗资源配置不均衡、服务费用增长过快、服务可及性差、服务质量参差不齐以及服务效率低等问题。如何能够通过大数据处理及挖掘技术来提高数据服务的效率及多样性成为亟待解决的问题。未来我们可以直接通过手机APP查询个人在所有医院的就诊记录、医疗处方、检验结果、费用清单等;慢性病患者如糖尿病、高血压等,可随时随地将智能设备检测到的血糖、血压等指标数据上传至医疗数据中心,数据挖掘平台可自动判断其病情发展动态,并发送给医生做远程决策分析诊断;从海量医疗数据中挖掘出来大量医疗诊断信息,可用于来协助基层或者偏于地区的医生对患者做出辅助诊断和用药建议;患者的生命体征、医疗影像、诊断结果、处方医嘱、手术住院以及账单等全周期数据将进行全方位的跟踪分析,不断更迭数据挖掘的训练集,为后续治疗方案设计提供支持;同时通过大规模医疗健康数据的汇总分析,可用于慢性病及流行病的筛选、分析及疫情预警,并为公共卫生管理部门防治干预计划及行动提供分析基础和参考^[5]。

综合上述分析,医疗领域是大数据技术应用的重要领域之一,可以用大数据的4个“V”来定义和描述医疗数据:

Volume (数据量): 一个区域医疗数据中心通常保存着成百上千家规模不一医疗机构所产生的医疗健康数据,而且随着时间的发展数据量持续增长^{[6][7]}。

Velocity (实时性): 医疗数据每时每刻都在产生,医疗信息服务中可能需要实时进行处理的需求,例如:实时临床决策支持中的诊断和治疗建议、健康指标实时预警等^[8]。

Variety (数据种类): 医疗数据不仅包含各种结构化数据检测表、非结构化或半结构化的数据,还包括CT、B超等医疗影像数据,可见医疗数据的数据类型之多。**Value (数据价值):** 医疗数据与每个人生活和生命健康息息相关,其价值不言而喻^[8]。

海量医疗数据经过生成及采集,如何将海量、异构、实时、多样的医疗数据进行高效存储,以实现大规模复杂健康数据查询的快速和准确响应,将实时信息与历史数据综合分析,可以通过云服务的方式向医务人员提供诊疗参考,或直接向终端用户提供医疗护理方案;与此同时,在医疗大数据存储与处理平台上,应用数据挖掘理论和技术对历史医疗数据进行分析,如何从海量医疗大数据信息中

挖掘患者的关键生理特征，可靠、高效地发现早期疾病和预测可能存在的健康风险，为患者提供有价值的医疗服务^[10]。

1.2 国内外研究现状

蔡佳慧等表达了随着卫生领域信息化建设进程的加快，获取的医疗数据类型和规模正以前所未有的速度增加，医疗卫生领域也已进入“大数据时代”^[11]。在剖析了医疗大数据基本概念的基础上，归纳总结了在医疗大数据时代信息化建设所面临的新挑战，详细介绍上海市闸北区为应对大数据时代的挑战在医疗数据存储、管理、整合、利用等方面所采取的具体措施。周光华等^[12]首先对我国医疗卫生领域中数据资源管理的现状进行了说明，之后结合国内外大数据技术在医疗卫生领域的应用情况，探讨了大数据技术在部分医疗卫生领域包括疾病管理、公共卫生、健康管理、医药研发等方面的应用前景，并提出了在大数据时代医疗卫生领域面临的各种挑战和建议。

1.2.1 医疗大数据研究现状

医疗是大数据的富矿，过去由于各个医疗服务机构之间存在严重的信息孤岛现象，海量数据处于闲路沉睡状态。2015年3月国务院发布《全国医疗卫生服务体系规划纲要（2015—2020年）》中，明确提出要加强人口健康信息化建设，为医疗大数据的应用铺路；并首次提出“健康中国云服务计划”，积极推动健康大数据的应用。

医疗大数据是开展商保、移动医疗的基础。随着商保、移动医疗等加速入场，医疗大数据找到了用武之地。以移动医疗为例，“硬件+数据+服务”的商业模式闭环中，由于消费者付费意愿尚未激发，C端硬件推广趋于“免费”，数据运营将成为其重要的赢利点。

医保+大数据：让医保控费更有效

DRGs 受政府高度重视，有望推动大数据控费发展。DRGs 是国外发达国家行之有效的医保控费模式，其本质是将病人进行科学分组，对同组病人确定合理的医疗服务路径，从而能够有效的监控医疗服务机构的服务，达到控费的目的。全国有 8 个省市（北京市、天津市、内蒙古自治区、浙江省、安徽省、山东省、湖南省、四川省、云南省）已开展 DRGs 试点工作，政策的大力扶持有望助力

医保控费进入大数据时代。

商保入场为医疗大数据带来了全新的应用场景：1) 保险的精算定价。目前国内健康险精算定价基础薄弱，缺乏对疾病治疗费用的深度分析及对参保群体医疗费用风险的科学评估。为了便于商业保险更好的利用医疗大数据，中国保险行业协会成立了首家保险大数据公司——中国保险信息技术管理有限公司将收集和整理健康险行业数据，搭建健康险大数据平台，为险企提供服务；2) 构筑 PBM 商业模式的核心。PBM 商业模式的核心在于，通过对于处方的大数据分析，设计合理的医药报销目录，借助医药报销目录与医药上游企业进行议价。



数据来源：保监会，安信证券研究中心

图 1-1 我国商业健康险呈现加速发展趋势

医疗+大数据：开启精准医疗时代。医疗+大数据全面开启精准医疗时代。基因大数据的获取和分析是精准医疗发展的基石。

今年 3 月，科技部召开首次精准医学战略专家会议，精准医疗计划将在今年启动。此外，卫计委层面逐步放开测序试点牌照，为行业有序规范的发展奠定了基础。



数据来源：安信证券研究中心整理

图 1-2 大数据助力走向医疗 3.0 时代



数据来源：动脉网，安信证券研究中心

图 1-3 精准记录依赖于人体大数据的获取与分析

1.2.2 数据挖掘在医疗行业应用现状

面对结构复杂和种类繁多的医疗数据，传统的数据统计方法已无能为力；对于绝大部分具有无数学特征、非规范化表述形式的医疗数据，如何找到较好的数据统计分析方法成为当今医疗领域数据处理领域的研究热点。数据挖掘技术是通过相关算法对海量数据进行处理，从而实现自动的信息模式的获取。可见数据挖掘技术能够作为传统数据统计分析方法的延伸和补充，实现对复杂的医疗数据进行分析处理。而基于云计算的数据挖掘技术能够在此基础上更好的应对海量的、复杂的非结构化医疗数据计算需求^[14]。

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.