

基于大数据的中医养生保健平台构建研究^{*}

高翔 黄文康

商志浩

王文婷 杨珊

(广西中医药大学公共卫生与管理学院 南宁 530200)

(广西中医药大学针灸推拿学院 南宁 530200)

(广西中医药大学公共卫生与管理学院 南宁 530200)

〔摘要〕 介绍大数据概念和关键技术,从总体框架、数据层架构、服务层模式 3 方面详细阐述基于大数据的中医养生保健应用系统设计,充分发挥大数据价值,为相关研究提供参考。

〔关键词〕 大数据;中医;养生保健

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2020.03.018

Study on Building of Traditional Chinese Medicine Health Care Platform Based on Big Data GAO Xiang, HUANG Wenkang, School of Public Health and Management, Guangxi University of Chinese Medicine, Nanning 530200, China; SHANG Zhihao, School of Acupuncture - Moxibustion and Tuina, Guangxi University of Chinese Medicine, Nanning 530200, China; WANG Wenting, YANG Shan, School of Public Health and Management, Guangxi University of Chinese Medicine, Nanning 530200, China

〔Abstract〕 The paper introduces the concept and key technologies of big data, elaborates on the design of Traditional Chinese Medicine (TCM) health care application system based on big data from three aspects of overall framework, data layer architecture, and service layer mode, to give full play to the value of big data, to provide references for related study.

〔Keywords〕 big data; Traditional Chinese Medicine (TCM); health care

〔收稿日期〕 2019-07-10

〔作者简介〕 高翔,硕士,讲师,发表论文 11 篇。

〔基金项目〕 广西高校中青年教师科研基础能力提升项目“基于大数据的中医治未病慢病健康管理系統研究”(项目编号:2019KY0311);广西高校自治区级大学生创新创业训练计划项目“基于大数据的中医养生保健 app”(项目编号:XJW18098);广西高等教育本科改革工程项目“中医药院校医学信息仿真平台建设的探索与实践”(项目编号:2018JGB220)。

1 引言

随着我国经济快速发展,人民生活质量得到全面提高,但同时出现环境污染、食品安全、工作节奏加快等问题,养生保健成为备受关注的热门话题^[1]。中医养生是中国传统的养生保健手段,历史悠久,成效显著,具有西方医学无法取代的独特优势。随着互联网的迅猛发展,网络上出现大量的无结构化和半结构化信息与数据,传统的存储方式已无法满足数据分析和处理需求,从海量数据中获取有价值信息成为组织和个人在相关领域取得成果的关键。虽然利用大数据技术可处理海量数据信息,

但是由于技术尚不成熟,面临着隐私保护、相关性预知等方面的挑战,是亟需研究的问题^[2]。移动设备作为信息化技术革命的产物,其普及给人们的生活带来极大便利。本研究拟采用移动手机设备终端应用作为应用平台。在服务器端,通过 Hadoop 技术对网络海量数据进行提取、清洗和利用,根据关键词获取中医养生保健网络数据,根据用户行为习惯自动进行匹配。

2 大数据概念及其关键技术

2.1 概念

大数据即规模庞大的数据群集,在业界大数据仍然没有被广泛接受的准确定义,即定义不唯一。网络中目前已逐渐形成规模庞大的数据群集,且类型多种多样、结构复杂冗余。大数据技术拥有强大决策力、洞察力,还可以处理海量级数据以及多样化信息资源,如何处理并高效利用大数据是计算机科学技术与其他行业结合的研究重点。

2.2 关键技术

2.2.1 Hadoop 由 Apache 软件基金会开发的开源分布式系统基础架构。有两个核心组件: Hadoop 分布式文件系统 (Hadoop Distribute File System, HDFS) 为海量数据提供存储; MapReduce 为海量数据提供计算。Hadoop 运作服务器集群,依托开发者编写的程序对庞大数据进行分布式模式的管理,适用于大数据存储与分析应用,适用几万台服务器在集群模式运作,支持 PB 和 TB 级存储容量。Hadoop 最明显的优势是可储存和处理数据量庞大且类型冗余复杂的数据,而传统数据库无法完成。

2.2.2 HDFS 为 Hadoop 关键组成之一,能够实现大数据强大的存储能力。其提供强大的数据存储功能,很好地解决传统存储数据规模庞大、类型复杂的问题。HDFS 是建立于几万台机器上的数据存储系统,具有广阔的使用场景。采用主从式 (Master/Slave) 架构,一般一个 HDFS 集群由一个 NameNode 和一定数目的 DataNode 组成。NameNode 是 HDFS 集群主节点,其功能是管理数据块的映射、

HDFS 名字空间以及处理 Client 请求; DataNode 是 HDFS 集群从节点,其功能是对文件系统数据进行存储,处理数据块的读写操作。两个组成部分配合完成分布式文件存储。

2.2.3 MapReduce 一个软件框架,开发者可在框架上简单编写代码程序,代码程序能运作于由几千台业务机器构成的服务器的大规模集群上,再依托一种可行且拥有容错能力的方法并行处理 TB 级别的大量数据群集。采用 Master/Slave 架构,1 个 Master,若干 Slave。Master 运行 JobTracker 负责作业调度,定期使用“心跳”向 JobTracker 报告任务进度,同时接受新任务,Slave 运行 TaskTracker 负责具体作业处理。具体架构,见图 1。

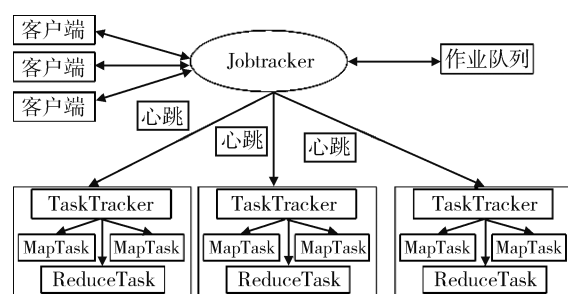


图 1 MapReduce 架构

3 基于大数据的中医养生保健应用系统设计

3.1 设计思路

3.1.1 总体框架 (图 2)

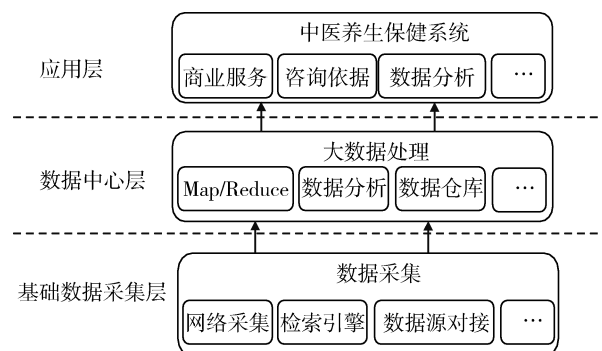


图 2 基于大数据的中医养生保健应用系统架构

3.1.2 功能结构 基于大数据的中医养生保健应用构建在安卓平台上,方便用户随时随地使用。其主要功能是提供用户个性化的中医养生保健方案、

中医膳食谱、结合线下养生馆的养生保健法等。功能结构，见图 3。

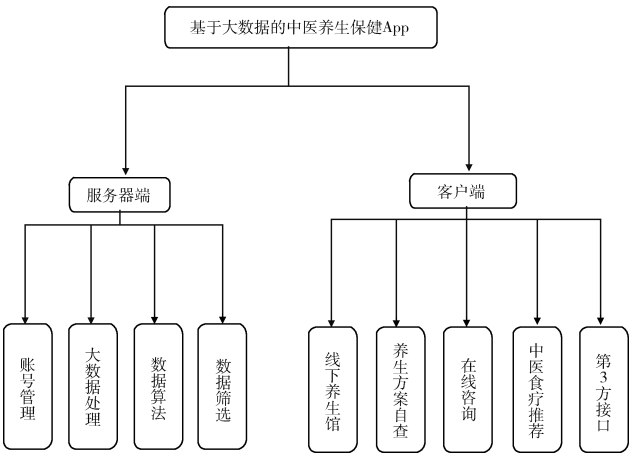


图 3 系统功能结构

3.1.3 开发环境 采用 Window + Eclipse + Android + Mysql + Hadoop^[3] 开发环境。Android 是一种基于 Linux 自由且开发源代码的操作系统，主要应用于移动设备上，具有易用性、便捷式操作等优势。结合开发工具 Eclipse 在 Window 上集成，Eclipse 开发工具的插件功能非常强大，可根据自身需求扩展 ADT 插件，满足开发者开发需求。在数据存储方面，采用 Mysql 组织和管理传统数据存储；Hadoop 为大数据存储工具，对复杂且庞大的数据集进行存储。

3.2 数据层架构

3.2.1 数据采集 一是网络检索引擎技术。网络是巨大的资源库，主要通过搜索引擎技术如 Apache Lucene 工具等检索大量的网络信息，根据中医养生保健的疗法、症状等划分需求制定规则，从而筛选出具有统计分类的中医养生保健相关信息作为数据储备。二是平台资源分享。互联网上有大量中医、养生保健等平台，可以通过平台信息资源对接方式采集相关中医养生保健信息以及较为完善的病历。三是网络爬虫。通过编写一个简单程序实现数据采集，其语言可以是 Pyhton、C + +、Java、PHP 等，捕抓网页上的信息、数据、图片、文件等，根据中医养生保健相关的匹配词进行数据采集并归纳分类，统计数据。四是数据分析。使用数据仓库对数

据进行清洗、装载、查询、展现等，决策选出业务逻辑中所需数据，实现“数据产生数据”的模式。

3.2.2 数据存储 采集的数据可分为非结构化、半结构化和结构化数据。对于结构化数据，结合中医养生保健信息高频出现、高效值的数据，优先考虑传统数据库存储方式并优先处理；对于非结构化、半结构化以及部分结构化数据，采用大数据模式，即分布式存储，将大量数据云集在服务器集群中并进行挖掘处理，做出中医养生保健决策。

3.2.3 数据算法 主要目的是对大数据进行挖掘，从而提高数据提炼的性能。海量数据较难处理，往往要达到需求量大、速度快、精确度高等要求，而传统的数据处理技术无法满足其处理需求。本研究采用 Apache Mahout^[4-5] 技术，其提供一些可扩展机器学习的经典算法，可以实现文档聚类、做出推荐和组织内容。Mahout 推荐算法分为两类：基于单机内存算法，基于 Hadoop 的分步式算法^[6]。针对大数据采用基于 Hadoop 的分步式算法。该算法是将之前单机执行的算法变成 MapReduce 模式。MapReduce 分布式并行计算的结构是将大数据均切分为小数据模块，再结合 Hadoop 集群的处理机进行并行处理，使其快捷地完成大数据挖掘、分析过程。Mahout 协同过滤推荐实现的组件，见图 4。

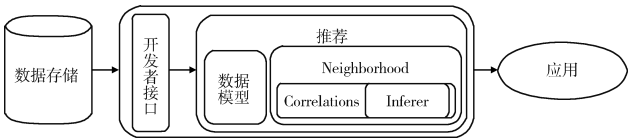


图 4 Mahout 协同过滤推荐实现的组件

3.3 服务层

3.3.1 面向用户 一是用户精准词元匹配中医养生保健服务。用户根据自身体征表现以及特点，通过选词或自定义词元将信息输入中医养生保健平台中，平台服务基于大数据筛选高效值，提供用户精准的养生保健方案，在用户实施方案期间对其追踪以反馈效果。二是中医养生保健膳食和推拿手法推荐。用户通过应用平台可查看中医养生保健膳食和手法推荐，膳食信息均由大数据通过一年四季、节

(下转第 81 页)

册、注销、定位等服务管理,提供模板解析、模型定义解析、服务自定义等基础服务,同时还包括信息自由检索、下肢动脉疾病管理文件提取、医学术语提供、下肢动脉疾病管理数据分析等各种服务。

5.4 数据存储

整个 HCDR 的核心,数据存储效率 and 安全性很大程度上决定 HCDR 产品的成功。数据存储模型支持自定义,根据模型元数据的定义,系统可扩展原来的模型定义。数据存储按下肢动脉疾病管理文档、数据两大部分存放,由模型元数据负责提供模型定义。HCDR 可以有选择地存储一些常用知识库,知识库内容可以直接提供给用户也可以作为 HCDR 内部数据分析参考。

6 结语

目前已应用平台尝试随访部分患者并在不断调试中,总体效果较好,但还需要不断改进,实现医生端-平台-患者端的无缝连接。先进的随访模式是随着个人智能设备的不断完善而发展起来的,包括智能手机、平板等,使医患之间在远距离也可以

进行点对点交流,节省大量时间。针对疾病相应特点开展随访工作更有针对性,对医护人员来说可节约随访数据的录入时间,提升准确率;对患者来说,可以在远程自助完成随访数据的输入或上传,得到相应医疗建议,极大方便患者。随着各种智能设备的不断发展,相信会建立国家层面的大数据平台,使患者随访工作得到统一管理,为国家制定相应医疗政策提供数据支持。

参考文献

- 1 肖雪雯,王谢,胡俊峰,等.基于 Android 开发的慢性疾病随访 APP 的设计与实现 [J]. 软件工程,2018,21 (2): 41-44.
- 2 叶辉.基于 IOS 移动平台的医疗预约随访软件的分析与设计 [J]. 电脑知识与技术,2014,10 (36): 8676-8677.
- 3 曹人元,任昉,曹亚茹,等.医疗随访现状及数字化随访系统的开发实践研究 [J]. 中国卫生产业,2018,15 (3): 154-155.
- 4 瓦力斯江·帕尔哈提,瓦依提·阿不力孜,买合木提·买买提,等.医院骨科随访信息处理软件的设计与实现 [J]. 电脑知识与技术,2018,14 (10): 86-89.

(上接第 77 页)

气等因素结合综合分析而得;推拿手法是根据用户已标识的自身体征匹配数据高效值来提供服务。

3.3.2 面向大数据平台 如线下养生馆,中医养生保健平台通过严格的权限管理,设置对指定场所或公司开放相关接口权限,为第 3 方提供接口服务,服务内容包括分类词元信息匹配、养生保健方案生成、计划安全性认证服务等,构建线下-线上联动模式。

4 结语

本文对基于大数据的中医养生保健平台构建进行简要论述,主要包括系统模型、数据处理方式、主要功能等。在大众养生保健的热潮中资源获取和数据有效性是用户最主要的需求。而大数据处理、

有效利用显得十分重要,是值得研究的方向,如何更好地利用数据产生更大价值需深入思考。

参考文献

- 1 于琦.中医养生知识库构建方法研究 [J]. 中国中医科学院中医药信息研究所,2015,17 (8): 1612-1614.
- 2 陈家蓉.基于大数据的城市智能公交管理系统的设计与实现 [D]. 西安:长安大学,2017.
- 3 徐凯田.基于大数据的智慧移动医疗信息系统结构研究 [D]. 青岛:青岛科技大学,2015.
- 4 黄小龙.综合性医院大数据需求分析与应用模式研究 [D]. 重庆:第三军医大学,2017.
- 5 黄倡明.大型活动安保大数据应用平台设计与分析 [D]. 南宁:广西大学,2016.
- 6 冯月春,陈惠娟.基于 Android 平台的淘医宝系统的设计与实现 [J]. 电脑知识与技术,2017,13 (25): 257-258.