

大数据驱动下的智慧医疗服务系统设计

李胜旭, 王 颖

(福建中医药大学人文与管理学院, 福建 福州 350122)

✉18905011828@189.cn; 494845194@qq.com



摘 要: 针对近年来大数据、云计算、人工智能等新技术在智慧医疗领域的迅速发展, 通过分析智慧医疗设计的层次结构, 提出基于Zookeeper下的分布式Redis缓存框架实现数据平台展示及交互。针对各种不同的医疗信息实现离线批处理与在线实时计算协同处理策略。在持续高并发下实现毫秒级延迟, 并准确快速地返回用户所需的信息。实现医疗信息采集及在大数据驱动下进行多维度分析与数据挖掘, 构建智慧医疗体系云系统, 为智慧医疗服务建设提供一种可行性分析方法。

关键词: 医疗大数据; 数据挖掘; 智慧医疗

中图分类号: TP393 **文献标识码:** A

Design of Smart Medical Service System Driven by Big Data

LI Shengxu, WANG Ying

(College of Humanities and Management, Fujian University of Traditional Chinese Medicine, Fuzhou 350122, China)

✉18905011828@189.cn; 494845194@qq.com

Abstract: In view of the rapid development of big data, cloud computing, artificial intelligence and other new technologies in the field of smart medicine in recent years, this paper proposes a distributed Redis caching framework based on Zookeeper through the analysis of the hierarchical structure of smart medical design, achieving data platform display and interaction. The collaborative processing strategies of offline batch processing and online real-time computing are implemented for various medical information. The system solves the problem of realizing millisecond-level delay and returning the information required by users accurately and quickly under continuous high concurrency. Medical information collection, and multi-dimensional analysis and data mining driven by big data are realized, so that a cloud system of smart medical system is built. The proposed system provides a feasibility analysis method for smart medical service construction.

Keywords: medical big data; data mining; smart medicine

1 引言(Introduction)

近年来, 大数据、云计算、人工智能等新技术发展迅速, 大数据和云计算作为信息化建设的新理念、新方向在不同的领域得到了研究人员的广泛关注。在医疗大数据环境下, 传统的医疗模式对数据的处理滞后, 已经不能适应新时代医疗信息化发展的形势及任务发展的要求, 如何将云计算技术与大数据服务相结合, 将两者引入智慧医疗建设, 已成

为当前智慧医疗是否实现转型升级的重大机遇^[1]。

将大数据、云计算和智慧医疗三者结合起来进行, 实现医疗信息化的研究, 是摆在科研人员面前的一个新的高度^[2]。首先, 医疗大数据中蕴含着巨大的有价值的信息, 如何提炼挖掘出这些有价值的信息, 对智慧医疗的发展有很大的促进作用。比如对医疗大数据中的病历, 利用文字搜索的技术、自然语言理解的技术、数据挖掘的技术进行对比、归纳、分

需及实际数据要求状况,在海量的医疗数据中准确收集有效有用信息建立数学模型,并进行预测研究。系统面对用户运维过程中,接收到的数据量以数量级形式呈现,数据量增加迅速,因此通过建立预测模型,可以在数据变化过程中就进行分析,及时对系统进行改变。最终产生的数据结果可以在医疗过程中或医疗后发挥作用,比如对某一个时期某些疾病是否有可能爆发进行预测判断。针对新型冠状病毒肺炎疫情发展的可能性,项目提出的方案可以对病毒的后期走向建立一种预测模型,给政府部门、医疗机构提供合理的参考价值。

在医疗大数据的分析中,将以医院信息管理和医疗信息服务的应用为驱动,基于大数据系统框架模型和医疗专题数据模型,采用先进的数据分析方法,比如关键绩效指标分析、聚类缝隙分析技术、数据多维分析技术、数据报表分析技术及数据仪表分析等,提供医疗大数据分析与决策信息服务。

大数据的分析算法和方法很多,结合医疗大数据的特点,特别是分析临床医学信息系统,研究其与临床决策支持系统的关联,建立基于医疗大数据的系统框架模型,研究支持PB级海量数据的、面向专题数据仓库的高效、快速的数据存储、数据索引及数据检索方法。在大数据系统中,由于数据量庞大,高效、快速的数据存储、数据索引及数据检索在数据仓库和大数据领域非常之重要,这种方法依靠良好的优化的数据组织结构和相应的算法。一个优良的检索技术对数据库具有至关重要的作用,并且是衡量一个数据库系统性能的重要指标。本文采用协同过滤算法来分析管理存储在数据仓库里的医疗大数据,通过算法设计基于HL7消息的推荐系统接口,最后通过HL7推荐消息使系统与电子病历系统等院内其他信息系统实现应用的实时对接。

算法设计:首先,我们利用基于Sqoop的大数据收集模块分布式地从各医院信息系统中收集患者就诊信息,截取其有效数据集组成用户可信内容,并将这些可信的记录收集过程转换为一个三元组序列:

<UserID, ItemID, Preference>

然后,我们使用若干相似度度量方式计算用户之间的相似度,比如欧氏距离、皮尔逊相关系数和余弦相似度等。

欧式距离表示多维空间中两个点的真实距离,其计算如式(1)所示:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

以上 n 维欧氏空间是一个点集, $d(x, y)$ 定义了两个点 x 与 y 之间的距离。其中 $(x_1, x_2, \dots, x_n)$ 为坐标上的每个不同点。

而用欧式距离表示的相似度则为式(2)所示:

$$\text{sim}(x, y) = \frac{1}{1 + d(x, y)} = \frac{1}{1 + \sqrt{\sum_{i=1}^n (x_i - y_i)^2}} \quad (2)$$

皮尔逊相关系数表示两个三元组之间协方差和标准差之间的比率,其计算如式(3)所示:

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{N}}{\sqrt{(\sum x_i^2 - \frac{(\sum x_i)^2}{N})(\sum y_i^2 - \frac{(\sum y_i)^2}{N})}} \quad (3)$$

通过皮尔逊相关系数衡量线性相关关系。若 $\rho=0$,说明 x 与 y 之间无线性相关关系。相关系数的绝对值越大,相关度越强;相关系数越接近于1或-1,相关度越强;相关系数越接近于0,相关度越弱。

相对于欧氏距离,余弦距离更关注两个向量在方向上的差异。余弦相似度表述了向量空间中两个三元组的向量值之间夹角的余弦值,通过计算的值来衡量两者之间差异的大小,其计算如式(4)所示。

$$\cos(x, y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}} \quad (4)$$

综合上述几种相似度计算方法,通过计算得到相似度度量值,应用在智慧医疗系统中来获得毗邻的用户或者项目,也就是一种基于相似度门槛阈值的固定数量的邻居。通过以上方法,设计了一种符合智慧医疗服务的数据挖掘及分析系统。算法代码如下。

```
for data in nb:
    dist=self.gCost(uId, data)
    self.nb.append([dist, data])
    self.nb.sort(reverse=True)
    self.nb=self.nb[:self.data]
def formDict(self, uId, l):
    user={}
    for data in self.dict[uId]:
        user[data [0]]=data [1], 0]
    for Ndata in self.dict[l]:
        if(Ndata [0] not in user):
            user[Ndata [0]]=[0, j[1]]
        else:user[Ndata [0]][1]=Ndata [1]
    return user

def precision(self, uId):
    user=[data [0] for data in self.dict[uId]]
    recom=[data [1] for data in self.recomList]
    num=0.0
    if(len(user)>=len(recom)):
        for data in recom:
            if(data in user):
```

```
num+=1.0    self.cost=num / len(recom)
else:
    for data in user:
        if(data in recom):
            num+=1.0
    self.cost=num / len(user)
```

4 智慧医疗大数据云服务数据展示平台(Data display platform of smart medical big data cloud service)

在智慧医疗领域,无论是病人、医生或者管理人员,都期望把蕴藏在大数据背后对自己有价值的信息清晰地展示在面前,所以最终的数据如何展现出来是非常关键的。

通过分析智慧医疗云平台的功能,数据展现主要有几个方面:(1)医疗机构实时了解和掌握近期的医疗动态并调整目标。医疗大数据云平台帮助各个医疗服务机构实现医疗资源共享共用,信息及时沟通、信息协同共享,共同为用户实现不同的服务视图,实现类型丰富的服务方式。另外也可以根据平台提供的信息将一些医疗服务进行整合,构建一种新的服务模式,以便节约运行成本。(2)决策或政府部门对数据的需求主要通过丰富的各类报表、图表及分析结果来满足,这些数据要具有一定的说服力,在决策过程中提供数据支撑。(3)患者主要通过提供的数据综合分析个人自身信息并进行预判,根据数据或信息选择合适的医生进行咨询或者从平台中得到指导、治疗和参考,避免盲目的诊疗。(4)医生也可以根据患者及平台提供的数据分析病况,对患者规划一个有针对性的治疗目标,最终制定合适的治疗方案。

因此大数据驱动下的智慧医疗服务系统设计的最后,是在智慧医疗云平台的应用上能够建立一个友好的交互。因此项目在智慧医疗应用层架构中,通过基于Zookeeper的分布式缓存框架来实现数据展示平台,分布式缓存架构要及时准确地快速地返回不同用户的请求,根据用户需要合理地提供查询服务。

本文提出一种分布式发布订阅消息队列系统,实现针对医疗数据的多样化处理方法,解决Hadoop无法实现即时查询和交互设计的局限。即针对获得的不同的医疗信息数据源,提出离线批处理与在线实时计算相结合的医疗数据处理策略,如图2所示。

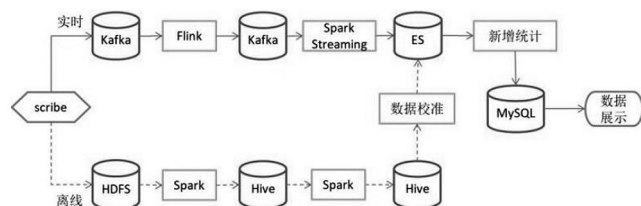


图2 离线批处理和在线实时计算的数据处理策略

Fig.2 Data processing strategies for offline batch processing and online real-time computation

为应对大量复杂的医疗大数据,尽可能降低处理延迟,实时计算部分采用了数据分层与分流相结合的技术路线,将数据计算流程拉长,采用单功能多阶段的数据处理方式将数据处理拆分为三个阶段:日志解析、产品分流和新增计算。在实时处理部分,采用了Flink和Spark Streaming相结合的方式。Flink是一种具有高吞吐、低延迟的实时离线统一的流式数据处理引擎,非常适合医疗大数据分析中第一阶段的日志解析特点。而Spark Streaming是微批处理,可以将实时数据流输入的数据划分为一个个小批次数据流,保障后续新增计算中聚合操作稳定的分钟级响应。为了将计算引擎的性能发挥到最大,将新增计算的延迟降到最低,

5 结论(Conclusion)

医疗云计算平台系统其数据的复杂性及整个系统的弹性伸缩能力与传统IT基础设施的技术有很大的差别。医疗大数据挖掘处理需要云计算技术作为平台,反之云计算技术也将计算资源作为一种服务支撑医疗大数据的有效挖掘,两者相辅相成,相得益彰,提供了各自需求的有价值信息。对于云计算理论应用,本文提出了智慧医疗云服务体系结构设计模型及其工程实现方法,将多个差异的医疗服务组织及医疗资源聚集于医疗云中,构建一个多样化、开放式、可伸缩、多用户的智慧医疗云数字化生态环境。在大数据的挖掘中,根据医疗大数据的特点,设计算法实现数据分析,给后期工作者提供以下几个方面的理论及实践研究的参考。

(1)针对不同对象提供个性化服务

通过大数据采集与关联技术,面向患者的分析结果提供具有针对性、个性化的服务方式;针对医疗机构及管理部门进行管理行为预测,提前发现可能出现的潜在问题。

(2)大数据环境下智慧医疗更加完善

在设计不同阶段对他人的不足之处提出解决方案,并通过算法分析实现,提出的方法及算法节省了运维时间和空间,节省了硬件成本。

(3)智慧医疗云平台的建立

智慧医疗云服务平台重点解决平台开放性、可持续扩展性以及服务之间松耦合的问题。在云平台架构设计时考虑了可扩展性问题,不断集成新的服务而不需要对整体架构做过多修改,最终建立智慧医疗云平台时把其中的每个服务都看作一组服务接口的集合,达到服务之间松耦合,实现整体架构的开放和易扩展性。

(下转第58页)