

基于 Web Service 技术用于医院信息分析算法库的设计与实现^①

赵晴峰

(浙江省肿瘤医院信息中心, 杭州 310022)

摘要: 分析了当前 HIS 在信息分析处理领域面临的数据来源复杂多变以及信息分析算法在定制、共享和动态维护等方面存在的主要问题, 采用元数据实现对信息分析算法的描述, 基于 Web Service 架构设计了信息分析算法管理模块和信息分析算法库系统。相对于传统模式的信息分析算法, 实现了数据和算法的松散耦合和对算法库的动态管理, 并进行了实现和验证, 研究表明, 系统能够有效改善 HIS 中信息分析和数据处理的可维护性和通用性。

关键词: Web Service; 信息分析; 算法库; 医院信息系统

Design and Implementation of Hospital Information Analysis Algorithms Library Based on Web Service

ZHAO Qing-Feng

(Information Center of Zhejiang Province Tumor Hospital, Hangzhou 310022, China)

Abstract: The paper discussed the main problems about the complex data sources and customization, sharing and dynamic maintenance of information analysis algorithms in the field of HIS information analysis. The meta-data description of information analysis algorithms was given. The information analysis algorithm management module and information analysis algorithms library system based on Web Service were built. The paper realized the loose coupling of the data and the algorithm and the dynamic management of information analysis algorithms library compared with the traditional information analysis algorithms library. Experiment results show that the system can improve the maintainability and generality of information analysis in HIS.

Key words: Web service; information analysis; algorithms library; hospital information system; HIS

早在 1995 年, 国务院卫生部开始部署实施国家卫生信息化建设工程, 即“金卫工程”, 其目标是要构建基于计算机和网络技术的能够对医学信息实施科学管理的现代化国家卫生信息系统。进入 21 世纪, 在国家金卫工程的推动下, 我国医院信息系统(Hospital Information System: HIS)建设取得了很大的发展, 大量新的技术标准、应用算法、软件架构等都在医院信息系统中得到广泛应用, 对于提高医院整体医疗质量、改善服务效率以及提高医疗服务水平发挥了重要作用。

随着 HIS 规模日趋扩大以及需求日渐增加, 其数据来源也更加复杂, 涉及的各种数据种类繁多, 包括各种临床诊疗数据(如电子病历、病史信息、医学影像、病理参数、化验结果等)、各种管理数据(如医院日常综合管理、人事相关数据、医疗政策研究、医保服务、医疗资源利用评价等)、财务数据(如患者诊疗产生的各种不同类型费用、医院财务与核算数据等)、药品管理数据(如临床用药的各种服务和审核、药房、药库的发货管理等)等。这样的背景下, HIS 中对信息分析的应用需求越来越多, 与此相关的研究也不断展开。张胜

^① 收稿时间:2015-06-17;收到修改稿时间:2015-08-24

昌(2012)研究了一种改进的 Apriori 算法在 HIS 中的应用,并在导诊台和医生诊断方面进行了实例应用,结果表明改进的 Apriori 算法具有降低系统开销和改善效率的优势^[1]. 贝乾(2013)在医院信息化领域中有关数据仓库和数据挖掘方面的研究中,针对不同系统的异构数据库源,提出采用 DOP(基于领域操作平台)工具的解决思路,并指出数据挖掘技术能够对医院管理决策和临床医疗提供有效支持^[2]. 唐佑萍等(2013)以中山市人民医院为例研究了数据挖掘的应用情况,并具体开发了 DIVER 数据挖掘和分析工具,实现了对医疗质量评价、合理用药监控以及耗材管理等数据的整合、查询和分析等操作^[3]. 相关研究表明,随着 HIS 应用的普及和展开,医院信息系统中所涉及的信息类型正在呈现多样化的发展趋势,表现在整型、字符型的结构化数据和声音、图像等非结构化的数据共存的局面,对相关的分析算法提出了更高的要求.

对各种不同特点和结构的数据进行科学处理是 HIS 中信息分析和数据处理的基本要求,需要与之相对应的算法实施处理. 传统模式下, HIS 在实施信息分析时往往将数据、算法和系统紧密捆绑在一起,例如康骞(2012)提出的数据挖掘分析方法在 HIS 中应用的方案就属于这一应用思路^[4]. 这一模式虽然具有数据处理针对性强,数据与应用之间的联系紧密等优势,但在应用中灵活性差和升级维护困难的劣势明显. 例如一旦出现对新类型的数据处理需求或者有新的算法加入,就需要对 HIS 的数据库结构以及信息分析算法进行代码层次的修改,难以满足快速灵活和可维护性好的需求,对 HIS 的发展形成了很大的制约. 因此,在 HIS 中构建一个能够实现定制、共享以及可动态维护的信息分析算法库,已经成为 HIS 实现高效数据管理所亟待解决的问题. Web Service 是面向服务的体系结构(SOA)的具体实现方法,较好的解决了软件工程领域中的异构和不断变化的数据两个难题^[5]. 考虑到 HIS 对信息分析算法库的客观需求,应用 Web Service 不但有利于实现对医院中不断产生的大量异构数据源进行处理,还可以采用 Web Service 对不同的算法进行封装、保存及调用,实现数据、算法以及算法库系统之间的松散耦合,发挥易用性、可重用性以及动态性等方面的优势^[6].

1 基于Web Service的信息分析算法库在HIS中的设计

1.1 信息分析算法的元数据描述

目前对于 HIS 中复杂多样的数据,涉及的分析算法主要包括统计方法、人工神经网络方法、贝叶斯网络方法、决策树技术、关联聚类算法,模糊逻辑分析、遗传算法、规则发现算法等,而随着技术的进步,未来还会出现更多的算法,因此需要按照一定的规范来存储和管理这些算法. 为了方便算法库的管理需要以及帮助用户识别、评价和使用算法,需要制定算法管理的元数据描述规范. 算法的元数据描述是对算法库中算法的科学描述,主要描述算法的属性,包括算法的内容、表示方式、管理方式以及数据集的特征等.

HIS 中的信息分析算法元数据的描述涉及很多要素,如算法的应用信息(方便用户选择算法时参考)、算法的输入输出信息(方便用户使用算法时对相关参数的输入)、算法的名称和 URL 位置信息(方便用户动态调用)等. 以下给出了对于“遗传算法”描述的元数据实例:

```
<? xml version="1.0" encoding="utf-8" ?>
    <service location="http://localhost/算法服务
/Service.asmx" name="遗传算法"></service>
    <usage>遗传算法分析病例服务</usage>
    <input>
        <param name=" num " type="double" text="种群中的
个体数目 " valid="^(?!d+)(\d+)?$"></param>
        <param name=" bounds " type="double" text="变量
的上下界的矩阵 "
valid="^(?!d+)(\d+)?$"></param>
        <param name=" eevalFN " type="double" text="适应
度函数" valid="^(?!d+)(\d+)?$"></param>
        <param name=" eevalOps " type="double" text="传递
给适应度函数的参数"
valid="^(?!d+)(\d+)?$"></double>
        <param name=" startPop " type="double" text="初
始种群 " valid="^(?!d+)(\d+)?$"></param>
        .....
    </input>
    </output>
        <output type=" X " text="最优解 ">
        <output type="endPop " text="最终得到的种群 ">
```

.....
</output>

在 HIS 的信息分析算法库中, 利用元数据对算法所涉及的数据类型以及算法信息等进行描述, 提供一种规范和普遍的描述基准和模式, 有利于有效实现数据与算法的松散耦合, 对于算法库的柔性化组织和管理具有重要作用.

1.2 信息分析算法管理模块的设计

在采用元数据描述信息分析算法的基础上, 将不同的算法加以封装, 即可构建基于 Web Service 的 HIS 信息分析算法管理模块^[7]. 根据 Web Service 的技术架构, HIS 的信息分析算法管理模块包括算法注册中心、算法发布模块和算法发现模块三个主要部分, 图 1 给出了 HIS 中信息分析算法管理模块的组成结构.

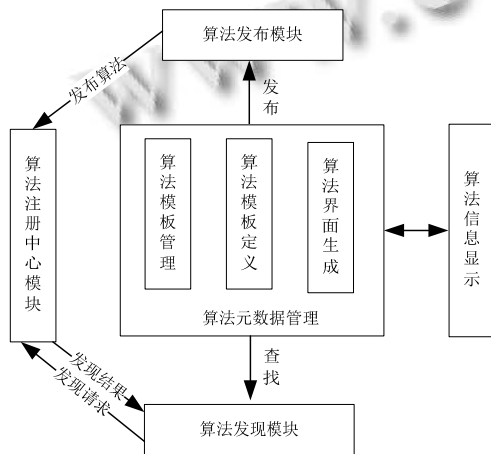


图 1 算法管理模块及其与相关模块的关系

其中, 算法注册中心模块的主要功能是用于实现算法注册以及算法发现. 算法提供者发布的算法中采用元数据的形式描述了信息分析算法的各种基本信息, 这些信息均由算法注册中心保存, 因而算法注册中心实质上提供了算法提供者与算法调用者相互联系的平台. 算法注册中心模块运行的基本流程如下: 来自算法发布子模块的请求到达算法注册中心模块后被接收进而生成信息分析算法的 WSDL 文档描述, 连同算法的服务名、URL 等信息共同保存在算法注册中心服务器上; 算法发现子模块请求算法服务的信息被算法注册中心接收后, 在信息分析算法库中进行查询, 得到算法库中现存且满足需求条件的算法后返回相关信息. 算法发布子模块提供了算法发布界面, 算法提供者可经由该界面输入算法的相关信息, 将信息分析

算法进行 Web Service 封装并发布给算法注册中心, 算法发布子模块可向算法注册中心发送请求, 包括向信息分析算法库添加一个新的算法或者对已提交的算法实施删除和修改等操作. 算法请求者可通过算法发现子模块到算法库中查找所需的算法, 基本流程为: 用户提供需要调用的算法基本信息, 算法发现子模块根据这些信息向算法注册中心发送查询请求, 算法注册中心查找满足条件的信息分析算法, 如果找到, 返回算法的相关信息(包括算法服务名、算法版权信息、算法 URL、Web Service 的 WSDL 文档等). 如没找到, 返回提示信息. HIS 中算法管理的核心是算法元数据的管理模块, 其中主要功能是实施算法模板管理、算法模板定义和算法界面生成. 对于算法元数据的管理主要是基于算法描述模板来进行, 算法描述模板通过建立的算法模板描述文件库来建立与相应算法之间的相互联系. 另外, 不同的信息分析算法对应不同的算法描述模板. 算法模板定义模块提供了算法的编辑界面, 可供用户对算法描述模板加以自行定义. 算法界面生成模块能够针对不同的算法生成相关的显示、编辑和查询界面.

从以上基本功能和流程分析, 算法元数据管理模块是算法注册、算法发布和算法发现等模块的实现基础. 以下给出一个典型的基于元数据模式的信息分析算法入库流程, 如图 2 所示.

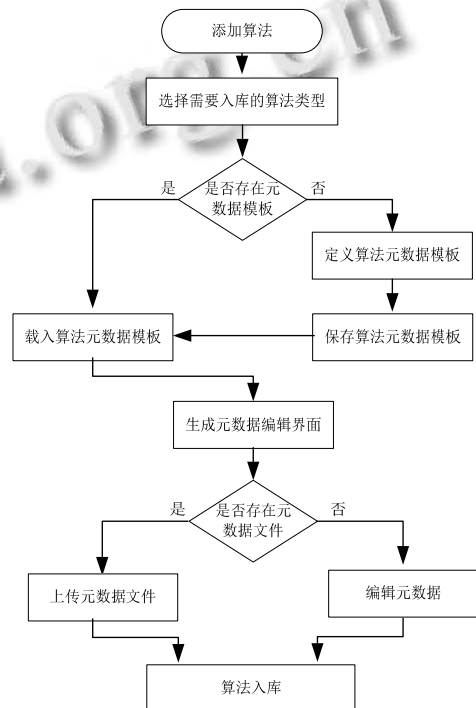


图 2 基于元数据的信息分析算法入库过程

其中,用户在添加信息分析算法时需要判断算法类型,当存在相关的算法元数据模板,即可直接载入加以调用,生成与之对应的元数据编辑界面来录入算法元数据.如果不存在相关的算法元数据模板,用户需要自行定义算法元数据模板,完成后即可在线填写算法元数据.系统基于 XSD 可生成 XML 文件,还可基于 XSD 来自行编写 XML 文件,并以文件形式上传,对 XSD 实施有效性检查后可将其存储于算法库中.

1.3 HIS 中信息分析算法库的设计

根据现代软件工程中开放性和互操作性的原则,合理的算法库要具有在不同环境和平台下方便共享的能力,同时还需要在数据与算法间实现分离并具有较高的可维护性.因此,对于 HIS 的信息处理应用,需要

以算法元数据管理的逻辑框架为基础,基于 Web Service 技术,开发一个基于 Web 的、层次化和模块化的算法库管理系统.

采用 Web Service 架构构建的信息分析算法库系统就是通过在信息分析层插入算法注册中心来改造信息分析层的基本结构.算法注册中心的作用是对不同信息分析算法的服务信息进行存储,同时还提供对算法的发布和查询等功能^[8].该架构中,基于 Web Service 的方式实现对算法的封装,采用 XML 进行算法元数据描述,并依据算法元数据自动创建 Web 界面组件.通过这一模式来实现系统中数据、算法以及交互界面的分离.基于 Web Service 的信息分析算法库模块在系统中的层次关系和结构如图 3 所示.

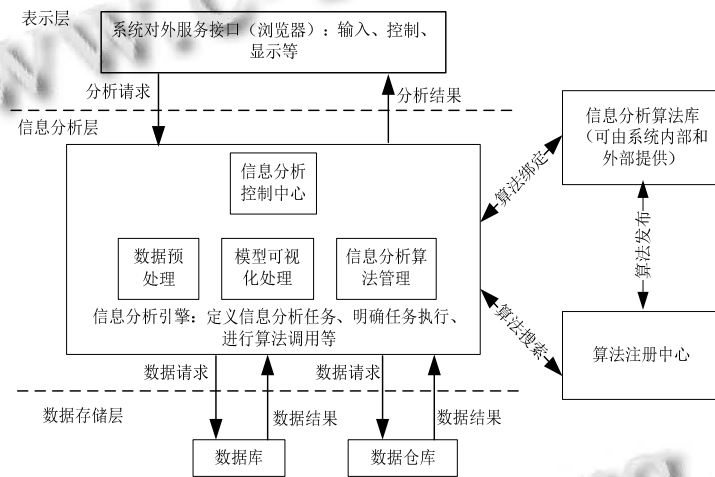


图 3 基于 Web Service 的信息分析系统算法库模块的基本架构

逻辑结构上,该信息分析算法库管理系统分为表示层、信息分析层和数据存储层三个层次.其中表示层是实现操作的交互界面,数据存储层是清洗和加工后具有统一格式的规范的数据库、数据仓库,信息分析层作为系统的核心,由不同性能和功能的分析算法组成.将 Web Service 引入信息分析系统算法库的开发,所有的算法均以 Web Service 的形式封装,构成算法库,在算法注册中心注册,成为算法库的组成部分,算法库可以分别运行在不同的 Web 服务器上.这种采用 Web Service 的方式对算法进行封装,基于 XML 进行算法元数据的描述,并根据算法元数据自动创建 Web 界面组件,实现了数据、算法以及界面的分离,大大减少了算法库模块开发的复杂性,有效提高算法库的可维护性,使系统具有灵活的动态结构,能够实现对算

法库的动态管理^[9].

2 实现及验证

目前来看,尽管许多商业化的数据挖掘和信息分析软件及平台最大程度的集成了许多经典的算法,能够为 HIS 提供多种选择的数据处理和分析服务.但鉴于 HIS 中数据的复杂性以及多样化应用需求的特殊性,对科学有效的信息分析算法的需求始终没有停止.这一背景下,本文开发了基于 Web Service 的信息分析算法库,将其应用于 HIS 中,力求能够满足 HIS 中信息处理伸缩性的需求.

2.1 系统实现的总体框架

本文设计实现的系统是在医院原有 HIS 的基础上加入基于 Web Service 的信息分析算法库模块,系统设

设计的总体框架如图 4 所示。

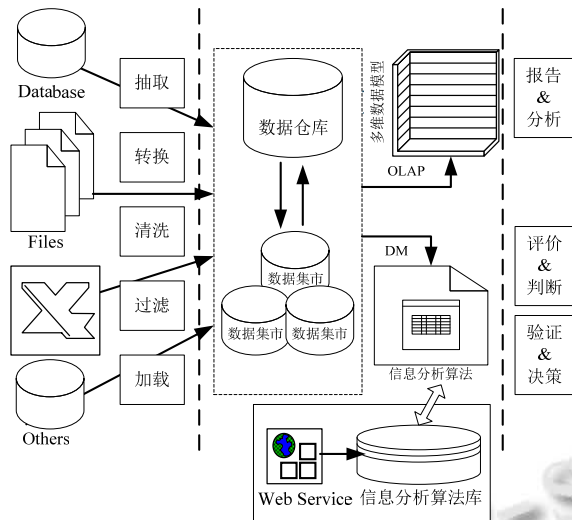


图 4 系统设计的总体框架

图 4 中, 在常规 HIS 的基础上, 扩展了基于 Web Service 的信息分析算法库模块. 系统在总体框架设计上将数据、信息分析算法以及算法库模块之间实现松散耦合, 另外, 该算法库模块具有根据不同的目标数据集和信息分析处理需求来动态和实时封装不同信息分析算法并提供给 HIS 实施信息分析处理的功能. 系统用户不但可以根据不同的分析目标以及数据源特征选择不同的信息分析方法进行信息分析, 而且还可以通过有针对性的研究分析动态添加新的算法研究成果, 充分改善系统的适应性和易用性.

2.2 系统主要功能模块

由于在传统的 HIS 中引入基于 Web Service 的信息分析算法库模块, 因此, 在信息分析方面的功能模块发生一定的变化, 根据以上给出的系统总体框架, 可分为以下几个核心功能模块, 如图 5 所示.

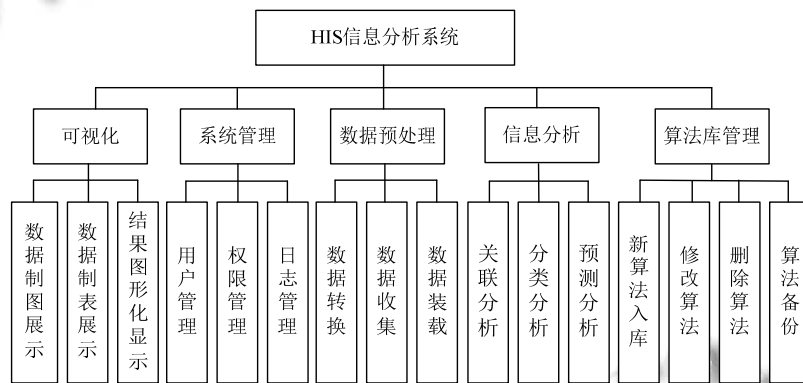


图 5 HIS 信息分析系统功能框架

其中可视化模块主要是面向用户实施交互应用, 包括给出数据展现的制图、制表以及对信息分析结果进行图形化显示; 系统管理模块主要涉及对系统的操作和日常维护管理; 数据预处理模块主要是对 HIS 中大量的数据进行预处理的相关工作, 包括数据转换、数据收集、数据装载等; 信息分析模块主要提供了主流的信息分析功能, 包括关联分析、分类分析以及预测分析等功能; 信息分析算法库管理模块提供了对信息分析算法管理的功能, 主要包括添加新的算法、修改、删除现有的算法以及算法描述等.

2.3 开发及验证

根据以上分析的系统框架和功能结构, 采用 MVC 架构来实现 HIS 信息分析算法库的开发, 其中开发平

台采用 MyEclipse6.0 和 Tomcat5.0, 开发实现上基于 SSH(Struts+Spring+Hibernate)三层框架集成的模式进行, SSH 的配置情况为: Struts2、Spring2 和 Hibernate3.1. 在 SSH 集成的框架体系中, 利用 Struts 框架完成对 MVC 的分离, Spring 用于对业务逻辑实施处理, Hibernate 实现对业务逻辑层和数据持久层的分离, SSH 集成的实现架构在具体应用中具有开发成本低、方便实现层间松耦合、便于维护和扩展性好等优势.

HIS 信息分析算法库运行的主要目标是实现对信息分析算法的管理, 而信息分析算法的生成是其中的核心功能. 入库的信息分析算法涉及到对 HIS 中相关数据输入/输出的控制、输入/输出参数或变量的确定、算法计算分析模型及方法、输出结果的展示等, 这些

变量和模型构成了完整的对于信息分析算法的描述。基于普适性的原则,在算法的设计中,需要对这些变量和模型进行抽象和组合,并提取出模型、参数和接口等的公用性描述方法,形成信息分析算法的通用描述模型。基于通用性的思路采用 UML 对信息分析算法的描述如图 6 所示。

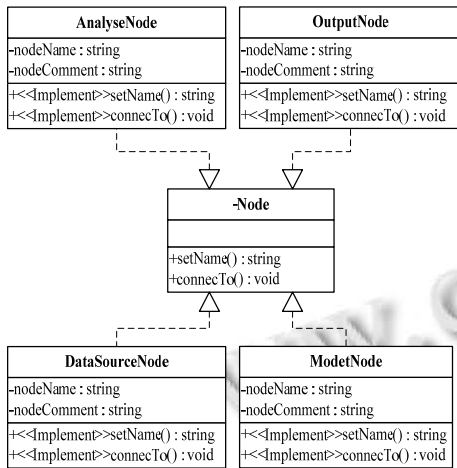


图 6 信息分析算法的 UML 描述

采用上述模式,实现信息分析算法库管理的主界面如图 7 所示。

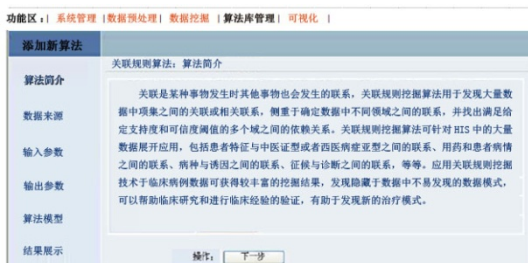


图 7 信息分析算法库管理的部分界面

在图 7 的操作界面上,用户按照系统提示输入相关参数和信息,根据算法的实际需求组合不同的变量和模型,进行个性化设置。系统根据输入的参数和模型等,可创建相应的信息分析算法执行脚本,最后形成算法的执行模型,保证在应用中能够调用该算法实施信息分析处理。算法入库后,在信息分析的操作界面上,用户通过对关联规则算法相关步骤和数据等进

行设定,即可调用该算法完成信息分析工作。

3 结语

本文在 HIS 中构建信息分析算法库用以解决对复杂动态医院信息的数据处理问题,并引入 Web Service 技术,实现一种基于 Web Service 的信息分析算法库系统,与传统方式采用的信息分析算法库相比,实现了数据、算法的分离,这一松散耦合的模式减少了信息分析系统中算法库开发的复杂性,便于对算法库的动态管理,对于改善医院信息处理的科学性以及系统的可维护性和通用性都具有重要意义。

参考文献

- 1 张胜昌.数据挖掘在医院信息系统中的应用[学位论文].青岛:山东科技大学,2012.
- 2 贝乾.数据仓库技术在医院系统集成中的应用[硕士学位论文].苏州:苏州大学,2013.
- 3 唐佑萍,卓绮雯,叶霭莹,王亚娜.基于医院信息系统实施数据挖掘.医学信息学杂志,2013,34(9):51-54,69.
- 4 康蹇.基于医院信息系统的数据挖掘探索[硕士学位论文].南昌:南昌大学,2012.
- 5 邢少敏,周伯生.SOA 研究进展.计算机科学,2008,35(9): 13-20.
- 6 汪维华,汪维富.基于 Web 服务的分布式计算模型研究.计算机工程与设计,2008,29(17):4623-4625.
- 7 Bao SH, Li R, Yu Y, Cao YB. Competitor mining with the Web. IEEE Trans. on Knowledge and Data Engineering, 2008, 20(10): 1297-1310.
- 8 Davis N, Demetriou G, Gaizauskas R, Guo Y, Roberts I. Web service architectures for text mining: An exploration of the issues via an e-science demonstrator. International Journal of Web Services Research, 2006, 3(4): 95-112.
- 9 Cong Y, Ji CX. Application of web-based data mining in personalized online learning system. International Conference on E-Business. Wuhan, China. 2007. 150-156.