

WSRF 环境下的模型管理方法研究

敖翔, 陈学广

(华中科技大学, 湖北 武汉 430074)

摘要: 决策支持系统(DSS)虽已取得了相当的发展,但还处于相对不完善的阶段,尤其是在模型管理方面需要改进。网格技术的出现和发展对 DSS 提供了理论和技术上的支持。在分析模型管理问题的基础上,文中提出了网格环境下基于 Web 服务资源框架的决策模型的技术体系结构和管理运作流程,同时提出了“有状态模型”及“模型属性文档”的概念,从而以一种标准的方式管理模型资源的状态,同时也为模型描述和管理标准化提供了一个发展方向。

关键词: 决策支持系统;模型管理;网格;分布式系统;Web 服务资源框架

中图分类号: TP311.5;N945.12

文献标识码: A

文章编号: 1673-629X(2006)09-0087-04

Model Management Method Based on WSRF

AO Xiang, CHEN Xue-guang

(Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: The decision support system had rapid development since 1970's. But until now, DSS is still far from perfect and needs a lot of improvement work, especially in the aspect of model management. The appearance and development of grid technology provide great opportunity to DSS. This paper first analyses some limitations and deficiencies of decision model management, and then proposes the technology architecture and management flow of decision model based on Web service resource framework. Concepts of "stateful model" and "model attributes documents" are brought forward, which provide a standard way to manage the state of model resources and a developing orientation to the standardization of model description and management.

Key words: decision support system; model management; grid; distributed system; WSRF

0 引言

自从 20 世纪 70 年代提出决策支持系统(Decision Support System, DSS)以来, DSS 已经取得了相当的发展。模型库及其管理系统是 DSS 核心,也是区别于一般管理信息系统(MIS)的重要标志,因此受到更多学者的关注。在多年的发展中,模型管理在理论研究、开发和应用等方面都取得了显著的进展。洪一帆、荣冈^[1]等人解决模型重用性较差的问题而提出了面向对象的模型管理方法。林杰、雷星晖^[2]等人提出了基于 Web 服务的分布模型管理思想,以克服采用基于组件的模型管理方式因平台语言的异构性而带来的分布式应用方面的不便。虽然上述方法理论上一定程度上推动了分布式模型管理,但是由于受模型复杂性及理论与技术发展水平的限制,模型管理还存在许多问题。其中包括决策模型的标准化描述问题、模型服务的安全访问、如何最大化各种决策资源的利用率及如何向决策者“推销”模型服务。这些问题在分布式环境下,平

台、语言和数据格式的异构性等交织作用下,非常显著地成为当前模型管理的主要困难,亟需发展新的方法和技术来适应分布环境下模型管理的需要。

网格技术的日益成熟为解决 DSS 模型管理所遇到的困难提供了思路和技术支持。文中提出了基于 WSRF 的模型服务管理框架体系,同时提出有状态模型及模型属性文档的概念,研究了基于 WSRF 的模型属性文档描述问题,探索了分布式环境下 DSS 模型管理的新方法。

1 网格服务

被称作“第三次互联网浪潮”的网格技术的出现^[3],将给工作生活的各个方面带来巨大的变革。Foster 认为“网格是一种集成的计算与资源环境,能够用于在动态变化的多个虚拟组织间共享资源和协同解决问题”^[4]。如今网格的研究重点也在从计算网格向应用网格、信息网格、服务网格转移,以求协作化、智能化和分布式地处理信息。

网格体系结构也由最早的以协议为中心的 5 层沙漏结构逐步转换为开放式网格服务结构(Open Grid Service Architecture, OGSA)。OGSA 将 Web Service 技术引入原有的网格技术,它是一种以服务为中心的体系结构。2004 年由 Globus Alliance, IBM 和 HP 等一些大企业或组织提出 Web 服务资源框架(WS-Resource Framework, WS-

收稿日期:2005-12-21

基金项目:教育部博士点基金资助项目(20040487076)

作者简介:敖翔(1981-),男,湖北武汉人,硕士研究生,研究方向为决策支持系统;陈学广,教授,博士生导师,研究方向为信息系统工程、决策支持系统、敏捷后勤、系统建模与仿真等。

RF)^[5],它强调的一点就是 Web 服务状态的重要性(WS-Resource 被定义为由 Web 服务和有状态资源所构成的实体)。Web 中广泛使用的 HTTP 协议是一种无状态的协议,即 Web 实际上是一种无状态的环境。这种环境不利于各类资源的管理。对需求迅速变化的环境,亟需一个有状态的面向服务的模型构建环境。WSRF 实际上是一系列的规范,用来定义 Web 服务资源所使用的“消息模式”或方法,以请求或者改变属性的值。Web 服务资源是一个有状态资源(比如说模型或者数据库)和与之交互的 Web 服务的组合。WSRF 的提出标志着网格与 Web 服务之间的融合。

模型是决策中的重要资源。在模型生命周期中,网格技术将提供更多的支持。语义网络的深入研究将带来海量信息的结构化管理,它的主要用途是发现处理数据的可用资源和数据、数据处理过程的集成,这将改善决策支持过程中的重要阶段——情报。

这也是模型的整个生命周期中的重要环节。而利用计算网格已成熟的技术,可以更加有效地利用模型辅助决策。例如远程沉浸,可以创建一个可视化协同分析环境;分布式超级计算可用于构建一个大型协同工作的模型仿真平台。网格技术带给决策支持系统的不仅是高性能协同计算和解决分布式环境异构性的中间件平台,更重要的是它将带给决策支持系统体系和运作流程的变革。这最终导致决策支持系统分析设计更加简单和更具业务逻辑敏捷性,从而推动决策支持系统的发展。

文中的重点是模型管理。模型的描述方式决定模型管理的方法和过程,是模型管理的基础。使用文档方式描述模型,结合网格服务新规范和技术,使得模型管理的方式将发生根本性的变革。模型作为一种决策资源,在开放性网格环境下,可以包装成服务的形式。在 Web 资源框架下,模型作为 Web 服务,与各类有状态的资源例如数据库、文件等结合起来,使模型可以与各类决策资源无缝集成,形成一种全局的控制分布式模型资源的管理模式。

2 面向服务体系的模型管理框架

模型是对客观事物的一种抽象表述。由于系统的复杂性和多样性难以建立一种通用模型来解决现实中所有的问题,然而却可以建立一个模型的快速搭建平台。利用面向服务体系正好符合这种思想。Web 服务是面向服务体系的典型结构。模型开发者(也就是服务提供者)可以将不同领域的模型包装成服务,用 Web 服务描述语言(Web Service Description Language, WSDL)描述这些模型的接口与功能,将这些服务以 WSDL 的形式发布到 UDDI(Universal Description, Discovery and Integration)服务器上。建模者通过对 UDDI 的查询,找到自己需要的模型服务,并与这些服务进行绑定,通过建立工作流定制模型服

务。

这种面向服务的模型管理框架存在一些细节需要研究。例如,模型如何用 WSDL 描述,如何搜索调用组合,如何管理服务的生命周期等问题。在网格技术推出新规范 WSRF 之后,模型管理的理论和方法得以建立在这种新的规范和技术上,可解决传统模型管理的问题。

在面向 WSRF 的模型管理框架中,这些新的规范和技术可以划分成 4 个层次:底层基础协议层,模型服务描述层,模型服务资源层和模型服务组合层。模型服务技术协议栈如图 1 所示。

模型服务组合层:

模型服务资源层:

模型服务描述层:

底层基础协议层:

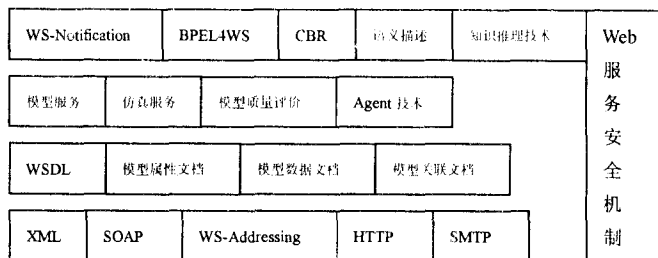


图 1 模型服务技术协议

(1)模型服务组合层中,WS-Notification 实现了 Web 服务资源基于事件的消息访问机制,BPEL4WS(Business Process Execution Language for Web Services,Web 服务的业务流程执行语言)用来实现模型服务资源的整合,CBR(Case-Based Reasoning,基于案例推理技术)^[6]结合模型服务语义及知识推理技术,得到模型服务组合方案。

(2)模型服务资源层由网格模型服务资源组成。包括模型服务、仿真服务、服务质量评价和 Agent。通过 Agent 协调各种模型服务,并评价模型仿真结果。

(3)模型服务描述层是整个框架的基础,是整个模型管理的基础。在 WSDL 基础之上,需要创建适合模型环境的资源文档。下文将详细论述 3 类模型描述文档。

(4)底层基础协议层包括各类底层协议。其中还包括 WSRF 中定位 Web 服务资源的规范 WS-Addressing 和底层消息传输协议 SOAP(Simple Object Access Protocol)。

(5)网络安全是体系中每个层次都需要涉及的问题。而在 Web 服务资源框架中提出了包括 WS-Security 在内的一系列安全规范。

在整个模型管理框架体系中,模型服务描述层是体系的核心。文中实质上提出的是一种基于 XML 的模型描述文档的分布式模型管理方式,并在 WSRF 环境中将此类文档与相关 Web 服务无缝集成,并用 WSDL 文档描述 Web 服务资源,称为模型服务资源。这种方式非常有利于控制服务状态。而借鉴到模型管理领域中,则需要建立模型资源属性文档的规范,从而可以使用标准的方式对分布式模型资源进行控制管理。

3 基于 WSRF 的模型服务资源属性文档描述

在 WSRF 环境中,模型都可以包装成有状态的模型服务。而模型管理则是对一系列资源属性文档的建立、维

护、交互和管理,其中包括模型属性文档、模型服务关联文档、模型服务数据文档。这些资源属性文档作为有状态资源和模型服务结合成 Web 服务资源。

1)模型属性文档。

有状态模型服务是通过模型属性文档来描述的。模型属性文档是一个可以用来描述模型状态的文档,实际就是一个 WSDL 文档,其中包括以下几个部分:

(1)模型动态属性(模型的生命周期、时间协同属性)。

动态模型都是在模型执行过程中动态改变的,用以控制模型的运行状态和模型服务之间的交互状态。

模型的控制就是对生命周期属性的改变,它控制着服务的调用、执行和撤销。模型调用网格服务时,需要有一定的协同工作能力,这里包括对时间和资源的协同。因此在模型属性文档中需要有一个时间协同属性,它随着模型的运行不断地改变。控制着模型中服务调用的顺序,可以对服务指定优先级。这种服务调用方式可以并行执行服务,提高模型运行的效率。

(2)模型静态属性(模型的输入输出、各类参数、功能描述、厂商信息、版本、事务及权限控制、服务安全等等)。

静态属性都是在属性文档中保持不变的。模型的输入输出参数数据有别于一般的算法,它需要指明其使用范围。因为这些数据使用范围直接影响到模型建立的成功与否。同时模型的输出应当是一个误差范围或者是一项统计参数,它的表现形式根据决策者的喜好决定。完整的安全控制体系才可以使得 DSS 的商业应用成为可能,而在 WSRF 中提出了一整套安全服务体系规范。这个服务组合参数表示为模型中事务的属性,并写到模型属性文档中。

(3)服务寻址属性。

使用 WS-Addressing 规范来对包装成服务的方法进行寻址。WS-Addressing 提供一种方式来指定关于位置的信息,而不只是一个统一资源标识符(Universal Resource Identifier,URI)。WS-Addressing 引入了 Endpoint Reference 概念。Endpoint Reference 是一种方式,用于指定消息到达适当的位置并带有适当的相关信息。

2)模型数据文档。

模型数据文档是在模型执行过程中需要使用和产生的数据集合。包括模型的总体输入输出参数、模型中方法的输入输出参数、方法之间交互的中间数据等等。模型数据及各类参数都用 XML 文档形式表示。

模型和数据库之间可以通过这类 XML 文档无缝集成。XML 文档可以存储在数据库中,也可以通过映射将 XML 文档结构映射到数据库的表结构中。当然这里需要建模者在模型生命周期中使用统一的 XML Schema 来定义 XML 文档的数据结构和类型,这样才可以使得模型之间的无缝集成成为可能。而输出数据是用 XML 文件可

以转换为很多其他的表现形式(网页,报表,数据库,图形等)。在服务组合中,存在一个上下文环境。两个紧密关联的服务使用中间数据文档,类似一个全局变量的概念。通过改变这个文档的数据属性实现服务交互。

3)模型服务关联文档。

该类文档描述了在构建一个模型时所需要的服务和资源之间及服务与资源之间的关联度。它记录模型结果(准确性和精确性)、运行参数(性能及客户综合满意度)和组合参数等,对建模者提供了参考信息。

模型关联文档的关键是模型领域知识和模型仿真结果的评价问题。首先需要建立完善的结构化领域知识,从而才能实现模型的语义描述。通过对大量模型关联文档的分析处理,并结合使用基于案例分析技术,在模型语义的基础上建立知识推理系统,可以建立一个全局的模型服务关联数据仓库。决策者便可以找到适合自身领域的模型服务组合。

模型资源文档作为资源与模型服务一同发布于 UDDI,在模型的生命周期中综合使用这些资源。模型属性文档的属性控制模型的状态的改变,包括模型的创建、运行以及撤销。它记录并监控整个模型生命周期中模型的状态,它是模型管理文档的核心。在模型运行过程中,使用模型数据文档进行服务间的数据交互,它是模型服务之间及模型与模型使用者之间的接口。而在模型仿真结束对结果进行分析后,则需要建立模型服务关联文档。它是对模型的建立使用情况的一个总体情况的分析结果,既对决策者提供更多关于模型的参考信息,使得决策者更加信任模型,同时也为将来建立相关领域的模型打下一定的基础。

以上对各类文档的描述可以视为初步的文档规范。模型属性文档需要由建模者建立,这是实现模型重用性、扩展性及模型组合的基础。

4 模型管理工作流程

在模型资源属性文档的基础之上,并综合利用模型服务技术协议栈中的各种技术,便可构建一个模型快速搭建平台,而且可以更加方便地管理模型。模型管理工作流程如下所述:

(1)建模者利用 Agent 查询模型领域关联信息;

(2)建模者的领域建模过程,并提交 Agent 模型所需的服务(利用领域知识对决策问题进行抽象,但并不涉及到底层技术和协议);

(3)Agent 的建模过程(包括模型属性、数据文档的建立,模型服务资源与底层协议的绑定);

(4)模型仿真运算,返回结果;

(5)建模者对结果评价,形成关联文档,关联信息归入 UDDI 的全局模型服务关联数据仓库。

WSRF 下的模型管理的简要流程如图 2 所示。

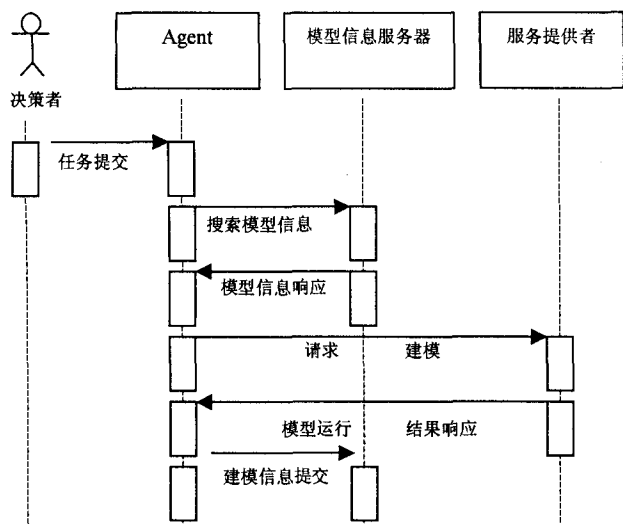


图 2 模型管理流程图

5 结束语

文中重点讲述了一种基于网格环境的模型状态管理的思路,并提出了一个模型服务技术体系框架。在网格新规范 WSRF 下,提出了用网格属性文档来管理模型的查询、建立、实施等模型生命周期的各个部分。

模型服务在分布环境下的性能和模型服务标准也是需要考

虑的问题。性能一直以来就是网格计算的优势,所以随着网格技术和硬件条件的发展它会进一步的提高。

参考文献:

- [1] 洪一帆,荣 冈.面向对象的模型管理系统[J].科技通报,2002,18(6):446-450.
- [2] 林 杰,雷星晖.基于 Web 服务的分布模型管理系统的研究[J].计算机应用,2004,24(4):80-82.
- [3] Foster I, Kesselman C. The Grid: Blueprint for a New Computing Infrastructure[M]. San Francisco, CA: Morgan Kaufmann, 1999.
- [4] Foster I, Kesselman C, Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations[J]. International Journal Supercomputer Applications, 2001, 5(3): 200-222.
- [5] Czajkowski K, Ferguson D F, Foster I. WS-Resource Framework[EB/OL]. <http://www.globus.org/wsrf/>, 2004.
- [6] Ong M, Ren X, Allan G, et al. Decision Support System on The Grid[A]. In proceedings of the Int'l Conference on Knowledge - Based Intelligent Information & Engineering Systems (KES)[C]. New Zealand: Wellington Institute of Technology, 2004.

(上接第 86 页)

义上的“数据挖掘”,实质上是挖掘建模的具体技术过程。采用 SEMMA 方法论逐步找到合适的建模技术,训练数据,最终找到规律和模式^[5]。

g. 检验模型的效果。在模型检验中,会使用历史数据中部分已有结果,以测试数据的形式与模型预测结果对比,客观地考察预测准确性。在真正的预测期间,只能等到未来的数据结果变成现实后,才能对预测结果作出对比,因此,需要有一个模型在市场环境中的试投放的时期,来检验模型真实效果。

3) 对分析结果进行理解和应用。

利用数据挖掘的最终结果和中间结果,可以深入了解企业数据的分布特征和存在的问题,进行一次性的专题分析或是周期性分析预测,还可以建立实时评分系统,如客户信用评分系统等,也可以为企业数据系统的改进提供重要的依据。

4) 评估模型的收效。

将模型的结果和投入成本与真实的业务收效相比,最终对数据挖掘过程作出综合评价。

4 小 结

数据挖掘项目在目前,特别是在国内,还处于边界条件尚未明确界分的阶段,并不是很成熟。但是数据挖掘项

目的特质之一就是动态性,这种动态性是由它与企业业务的密切结合决定的,它对于业务的辅助作用的力度和直接程度超过了传统的业务支撑系统、MIS 系统,也超过了数据仓库应用中的报表查询系统;企业对于决策信息的需求,在数据挖掘项目中,找到了前所未有的载体,因此,数据挖掘应用拥有更加广阔深远的前景。随着数据挖掘中某些应用的进一步成熟,数据挖掘将在各大行业中逐步形成有层次的产业链。

所以,不断地跟踪最新的数据挖掘知识和项目实施方法论,不断地通过数据挖掘项目实践来创造业务效益,应该作为国内信息技术领域在今后一个时期的焦点命题。

参考文献:

- [1] Berry M J A, Linoff G S. Mastering Data Mining[M]. [s. l.]: John Wiley & Sons, 2000.
- [2] Berry M J A, Linoff G S. Data Mining Techniques[M]. [s. l.]: John Wiley & Sons, 1997.
- [3] Han Jiawei, Kamber M. 数据挖掘概念与技术[M]. 北京:机械工业出版社, 2001.
- [4] 萨师焯. 数据库系统概论[M]. 北京:高等教育出版社, 2004.
- [5] 郭崇慧. 数据挖掘教程[M]. 北京:清华大学出版社, 2005.