

基于证据推理的云计算服务适应性评估*

付超^{a,b}, 桂鹏飞^{a,b}

(合肥工业大学 a. 管理学院; b. 过程优化与智能决策教育部重点实验室, 合肥 230009)

摘要: 针对云计算服务模式尽管对中小型企业吸引力巨大,但由于受到安全、技术、管理等方面的限制并不是都能适用,企业需要根据自身实际需要作出判断的问题,提出了一个云计算服务适应性评估指标体系。该体系从安全与风险、数据、服务、资源、经济五个维度来分析云计算服务的适应性问题,并利用证据推理方法对实际问题进行建模分析,验证了评估指标体系的有效性与合理性。

关键词: 云计算; 适应性评估; 证据推理; 评估模型

中图分类号: TP302 **文献标志码:** A **文章编号:** 1001-3695(2012)11-4304-05

doi:10.3969/j.issn.1001-3695.2012.11.078

Adaption assessment for cloud computing service based on evidence theory

FU Chao^{a,b}, GUI Peng-fei^{a,b}

(a. School of Management, b. Key Laboratory of the Ministry of Education on Process Optimization & Intelligent Decision Making, Hefei University of Technology, Hefei 230009, China)

Abstract: Cloud computing service model is very attractive to small and medium-sized enterprises, however whether its service capabilities can meet the actual business needs or not is difficult to measure. Based on this, this paper proposed a cloud computing service adaptability evaluation system, from security and risk, data, services, resources, economic dimensions, to analyze the adaptability of the cloud computing services, and designed a numerical example to assess and verified the validity and rationality of the evaluation index system with evidence theory.

Key words: cloud computing; adaption assessment; evidence theory; assessment model

0 引言

近年来,云计算逐渐成为计算机网络领域中一个炙手可热的新兴服务模式。其先进的服务理念表现为:将计算和相关资源以服务的形式通过网络按需提供给客户,实行按需付费,实现高可靠、可用、高扩展性服务,减少或免除前期硬件及架构资本投入,重视商业模式和网络服务^[1,2]。

正是由于其服务理念在业务模式和商业模式上的创新与突破,将推动愈来愈多的传统IT服务转向云计算形式。但是云计算并非万能钥匙可以解决所有问题。比如,目前来说云计算并不适合于科学计算,以及能源、天气预报、核爆、天文等需要高性能、高关联性计算的领域^[3,4]。需要认识到,云计算仍然处在其发展的初步阶段,在安全、信任、成本控制等方面都有待提高。而当前企业正试图探索采用这种服务模式的可能性,因此迫切地需要解决用户怎样对该模式的可用性和实用性进行准确评价这一问题。

目前,国内外关于云计算服务适应性的研究才刚刚起步,如文献[5]对云计算环境下如何评测可能存在的安全风险作了简单探讨;文献[6]制定了一些规则来评价企业是否应采用云计算模式,并设计了一个定量的一般模型计算云计算带来的投资回报等。但是这些工作或者尚停留在描述性阶段或者仅从某一个角度出发,缺乏全面、系统的评价。基于此,本文提出

了一个云计算服务适应性的评估指标体系,站在用户的角度从安全与风险、数据、服务、资源、经济五个维度对云计算的适应性进行评估,如图1所示。同时运用证据推理方法,针对某大学研究所的实际问题进行建模、分析,验证评估指标体系的有效性与合理性。

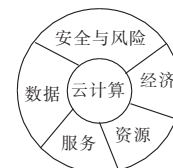


图1 评价框架

需要强调的是,安全与风险维度与数据维度中涉及的数据安全属性可能令人混淆。但是鉴于数据安全属性本身重要性显著且评价指标繁多,因此把涉及具体数据安全方面的内容独立地放在数据维度中进行探讨,而安全与风险维度只涉及对云计算环境、基础设施及法律法规方面的安全性。

1 云计算适应性评估体系概述

1.1 云计算适应性分析

尽管云计算的发展还处在初步阶段,相关技术、模式还不够成熟,并且在其发展的过程中还会不断发生变化,但是鉴于其服务模式在成本方面的优势,仍然有很多企业对云计算表现

收稿日期: 2012-03-06; **修回日期:** 2012-04-12 **基金项目:** 国家自然科学基金资助项目(71071045, 71131002); 国家教育部人文社科基金资助项目(12YJC630046); 安徽省自然科学基金资助项目(1208085QG130)

作者简介: 付超(1978-),男,湖北武汉人,副教授,硕士,主要研究方向为云计算、决策科学与技术;桂鹏飞(1988-),男,安徽宣城人,硕士研究生,主要研究方向为云计算、信息服务(hfuuaeif@126.com)。

出浓厚兴趣。那么,企业作为用户方应该如何去判断云计算到底能否满足自身需求、适应自身业务结构呢?这就需要有一套较为系统的评估体系来帮助企业构建分析模型,并选择合适的决策方法得到结论为企业决策提供支持和建议。

对于云计算适应性问题,首先有两种类型的企业需要加以区分,即拥有数据中心的大型企业和中小型企业。根据规模经济原理,企业拥有的 IT 基础设施资源量越大,其运行这些资源的平均成本越低^[7]。规模经济效应使得企业拥有和维护既有资源的平均成本降低,从长远看来管理好自身资源比采用云计算方式更加节省成本和可靠,因此云计算对拥有数据中心的大型企业来说吸引力有限。而中小型企业和技术、资金等方面能力有限,云计算提供的服务模式使其无须花费大量资本在 IT 设施的购置、存放、电力、冷却、扩展、维护等方面,只需支付所使用服务的费用就能解决问题,在大大节省成本的同时又使得企业有更多精力专注于核心业务发展。因而云计算对中小型企业来说没有进入壁垒,更适合中小型规模的企业^[2]。

需要指出的是,在云计算环境中包括云计算技术提供商、云服务提供商、云环境租赁商、云方案提供商等,这里不做区分统称为云计算服务提供商。

1.2 云计算适应性评估指标体系的建立

罗军舟等人^[8]给出了一个云计算体系架构,本文在此基础上加以改进,将其划分为物理层、服务层、传输层、终端应用层,如图 2 所示。

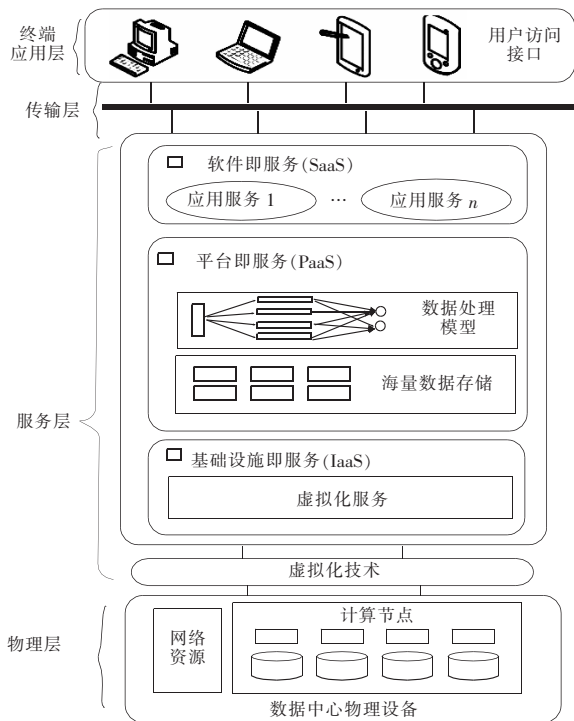


图 2 云计算体系架构

本文提出的评估指标体系的划分主要立足于云计算服务全局,综合考虑服务性能、质量和成本因素,结合服务过程中涉及的云计算体系架构中各个层次的地位和作用,从安全与风险、数据、服务、资源、经济五个维度四个层次对云计算服务适应性进行系统化的分析、评价,评估体系如表 1 所示,具体如下。

1) 安全与风险

国际合作发展机构 IDC 的一项调查表明,74% 的 IT 执行

官和 CIO 认为安全问题是制约他们采用云计算模式的首要问题^[9]。而 Amazon、Google 等爆出的各种安全事故更加加剧了人们的担忧。安全性贯穿于云计算服务架构中的各个层次,云计算提供的各种服务仍需要服务器、计算机等硬件设施来实现。一般情况下这些硬件设备损坏的可能性通常很小,而在硬件层之上虚拟出来的资源(逻辑上的主机)其安全风险往往更加让人担忧,因此这里的安全性更多的是考虑云计算体系架构中物理层以上的安全性:a) 服务层是云计算服务的实现部分,其中的虚拟化技术是实现云计算服务的关键,对于物理资源的利用往往都是通过虚拟主机以及自动化托管程序实现的^[10]; b) 传输层是服务双方传递信息和实现交互的唯一通道;c) 终端应用层是用户得到服务的平台,是用户身份与权限认证^[11]进行识别的起点。因此每个层次的安全性都会影响到云计算整体服务的安全可靠性,企业需要根据自身的实际要求进行评估。

风险性主要是来自于服务提供商方面,把自身的重要信息交给提供服务的他方管理是一项充满风险的行为,用户必须谨慎选择,尽量减小风险及损失。比如用户应当充分考虑其财务状况,以及当用户账户存在欠债时服务提供商制定有哪些关于用户数据及应用程序使用的策略。除此之外,服务提供商行为的透明度也至关重要,服务提供商会不会私自修改用户数据,用户能否追溯到服务提供商对用户信息操作的每一个行为,明确安全事故责任时服务提供商的行为是否可审计^[12]等。

2) 数据

数据维度评价的对象主要是处在物理层设备中存储的静态数据,以及传输过程中的动态数据、一些必需的临时数据。

针对数据的管理和控制活动是企业过程中最为频繁的活动,将数据交给他方管理虽然能为企业减轻负担并缩减成本,但却不让人放心。在传统的模式中,这些重要而敏感的数据通常存放在企业自身域内严格管理,因为这些数据往往关系到企业的核心利益。云计算服务模式中数据与服务高度耦合,数据的安全、管理、操纵对云计算的所有层面都至关重要,而用户却不能直接掌控和操纵数据。因此对用户而言,应该像评估传统的基础设施一样,对云中的数据安全性和数据管理与操纵特性进行充分评估。实际上,由于加密技术的限制,为了降低数据安全的风险,唯一可行的选择是确保不要将敏感数据或受管制数据放到公共云计算上^[13]。

3) 服务

服务维度主要针对的是两个方面:a) 架构中服务层各种服务的实现质量;b) 处在终端用户层的用户接受服务时对服务费用、效率等因素的评价和感受,以及服务提供商对所提供服务的管理和协作能力。

云计算的最终目的是依靠良好的网络,以基础设施、平台、应用软件等形式向用户提供各种服务。其中一些服务是免费的,但响应速度、可靠性、连续性等无法保证,另外一些服务则可能存在安全隐患或者服务的访问成功率以及负载弹性较低。服务策略是否灵活对用户也很重要,比如用户由于资金方面的原因出现欠费情况时,提供商能提供怎样的服务,能提供多少宽松的政策等。不仅如此,现代企业的经营已经不仅仅以自身为利益主体,往往需要实现多个相关企业甚至竞争对手之间的协作才能实现目标,这些都对服务提供商的服务共享和交互能力提出了挑战,云计算服务用户需要根据实际需求作出判断。

4) 资源

资源维度评估的对象主要包括服务提供商对物理层硬件资源的驾驭能力和企业用户自身硬件、技术、人力等资源的既有能力。

处在信息化时代的企业或多或少都会拥有一些 IT 基础设施、技术人员等资源,那么企业何时可以放弃自身资源转而采用云计算服务模式呢?首先,企业必须明确自身拥有资源的规模、资源负荷承载量,以及资源需求波动频率等,从而判断当前资源能否满足需求或者是否需要提高硬件设施能力;其次,在

自身资源不能满足需求的情况下,如果采用云计算服务,服务提供商对资源的驾驭能力如何,能否为用户提供合适且完善的服务。

5) 经济

在云计算的大潮下,有很多企业希望借助云计算在降低成本方面的优势使自身在企业信息化方面得到一个完善和提升,然而也有一些负面的调查显示,云计算并不像它看上去的那样省钱,采用云计算模式也是一种投资,因此,只要是投资,就要对其花费的成本和获得的收益进行充分衡量。

表 1 云计算服务适应性评估指标体系(模型)

维度	一级指标	二级指标	三级指标	评估等级	评估数据
安全与风险 (0.3)	网络层面 (0.3)	访问控制(0.4)	身份与权限管理机制(0.55) T_1	{优秀,良好,一般,较差}	良好
			网络访问控制机制(0.45) T_2	{很强,良好,一般,较弱,很弱}	良好
		网络环境安全(0.6)	互联设备可靠性(0.6) T_3	{很好,较好,一般,较差,很差}	很好
			网络入侵预防与监测(0.4) T_4	{非常优秀,优秀,良好,一般,较差,很差}	良好
	硬件层面 (0.2)	主机安全(0.6)	虚拟化服务器可靠性(1) T_5	{很强,较强,一般,较弱,很弱}	一般
			托管程序可用性和完整性(1) T_6	{非常优秀,优秀,良好,一般,较差,很差}	优秀
		应用程序安全(0.55)	Web 应用可靠性(1) T_7	{很强,较强,一般,较弱,很弱}	较强
			终端用户安全(0.45)	{非常优秀,优秀,良好,一般,较差,很差}	优秀
	应用层面 (0.25)	行为透明性(0.45)	接入终端可控性(0.4) T_{19}	{很好,较好,一般,较差,很差}	较好
			事件审计和报告(1) T_{10}	{0% ~ 100%}, { $H_1 \sim H_5$ }	32%
		提供商可用性(0.55)	部署与交付模式(0.5) T_{11}	{合适,基本合适,不合适}	合适
			服务商破产及应变计划(0.5) T_{12}	{风险很小,较小,较大,风险很大}	较小
数据 (0.25)	数据/存储安全 (0.45)	数据保护(0.55)	数据传输安全(0.45) D_1	{很好,较好,一般,较差,很差}	一般
			数据可追溯性(0.55) D_2	{20% ~ 100%}, { $H_1 \sim H_4$ }	55%
		数据加密机制(0.45)	密钥与证书管理(1) D_3	{优秀,良好,一般,较差,很差}	良好
			数据完整性(0.4) D_4	{80% ~ 100%}, { $H_1 \sim H_4$ }	86%
	数据管理 与控制 (0.55)	数据存储(0.45)	数据可用性(0.4) D_5	{50% ~ 100%}, { $H_1 \sim H_5$ }	85%
			数据分类与隔离(0.2) D_6	{20% ~ 100%}, { $H_1 \sim H_4$ }	32%
		数据操纵(0.55)	数据查询效率(0.4) D_7	{很高,较高,一般,较低,很低}	很高
			数据校验(0.25) D_8	{80% ~ 100%}, { $H_1 \sim H_4$ }	55%
			数据可审计性(0.35) D_9	{0% ~ 100%}, { $H_1 \sim H_5$ }	52%
服务 (0.2)	服务质量 (0.5)	性能度量(0.3)	吞吐量(0.2) S_1	{很高,较高,一般,较低,很低}	较高
			响应时间(0.3) S_2	{(2 s, 4 s], (4 s, 10 s], (10 s, 30 s], (30 s, 60 s], > 60 s}	2.2 s
		可用性(0.3)	负载弹性(0.2) S_3	{很大,较大,一般,较小,很小}	很大
			服务完整性(0.3) S_4	{很完整,较完整,一般,不完整,很不完整}	较完整
	可靠性(0.4)		访问成功率(0.5) S_5	{70% ~ 100%}, { $H_1 \sim H_6$ }	99%
			服务定制化能力(0.2) S_6	{高,中,低}	中
			服务连续性与稳定性(0.3) S_7	{很高,较高,一般,较低,很低}	较高
			灾难与故障恢复能力(0.5) S_8	{很快,较快,一般,较慢,很慢}	一般
	服务管理 (0.5)	服务动态管理(0.4)	服务透明度(0.5) S_9	{高,中,低}	低
			性价比(0.6) S_{10}	{非常高,很高,较高,一般,较低,非常低}	较高
		服务合作管理(0.3)	服务策略(0.4) S_{11}	{苛刻,严格,合理,灵活}	严格
			资源动态分配(0.4) S_{12}	{优秀,良好,一般,较差,很差}	良好
资源 (0.15)	用户 IT 资源承载能力 (0.6)	资源规模(0.3)	数据迁移(转入转出)(0.6) S_{13}	{0% ~ 100%}, { $H_1 \sim H_5$ }	35%
			互操作性(0.45) S_{14}	{优秀,良好,一般,较差,很差}	较差
		资源负载(0.3)	数据共享与转换(0.55) S_{15}	{0% ~ 100%}, { $H_1 \sim H_5$ }	50%
	资源需求(0.4)		硬件设备数量(0.6) R_1	{(10, 100], (100, 1000], (1000, 10000], > 10000}	44
			技术人才储备(0.4) R_2	{(1, 50], (50, 100], (100, 1000], > 1000}	22
			资源实际运行能力(0.55) R_3	{很高,较高,一般,较低,很低}	较低
			资源利用率(0.45) R_4	{10% ~ 100%}, { $H_1 \sim H_5$ }	90%
	提供商对 资源驾驭能力 (0.4)	不同组件系统整合能力(0.4)	需求波动频率(1) R_5	{每 4 个月, 每 4 个星期, 每 4 天, 每 4 个小时}	4 个月
			客户群大小		
		用户内部资源整合能力(0.3)	不同组件系统整合能力(1) R_6	{优秀,良好,一般,较差,很差,非常差}	较差
			用户内部资源整合能力(1) R_7	{优秀,良好,一般,较差,很差,非常差}	良好
经济 (0.1)	经济成本(0.4)	应用多租赁(0.2)	应用多租赁(1) R_8	{很好支持,基本支持,不支持}	支持
		部署、升级、撤销费用(0.4)	部署、升级、撤销费用(1) E_1	{很高,较高,一般,较低,很低}	很低
			管理、协商费用(1) E_2	{很高,较高,一般,较低,很低}	很低
	投资收益(0.6)	投资比重(0.25)	投资比重(1) E_3	{10% ~ 100%}, { $H_1 \sim H_5$ }	40%
		业务效率提升(0.4)	业务效率提升(1) E_4	{-10% ~ 100%}, { $H_1 \sim H_6$ }	25%
			投资回报率(0.6)	{-10% ~ 100%}, { $H_1 \sim H_6$ }	40%

2 基于证据推理的云计算服务适应性评估

2.1 证据理论基础

设 $\Omega = \{H_1, \dots, H_N\}$ 是为所有属性定义的一组共同的评价框架,称为信念函数的识别框架。一个基本信念分配 (BBA) 表示为映射 $m: 2^\Omega \rightarrow [0, 1]$, 满足 $\sum_{A \subseteq \Omega} m(A) = 1$ 。在 Shafer 的原始定义中,上述映射称为基本概率分配,要求 $m(\emptyset) = 0$ 。而在 Smets 提出的传递信念模型 (TBM 体系) 中,上述映射被称为基本信念分配,不要求 $m(\emptyset) = 0$ 。满足 $m(\emptyset) = 0$ 的 BBA 称为正规 (归一) BBA 或 mass 函数。设 m_1 和 m_2 为定义在 Ω 上来自两相异源的两 BBA, $m_{12} = m_1 \oplus m_2$ 为使用 Dempster 规则的合成信念,则

$$m_{12}(A) = \frac{\sum_{B, C \subseteq \Omega, B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B, C \subseteq \Omega, B \cap C = \emptyset} m_1(B) m_2(C)}$$

$\forall A \subseteq \Omega, A \neq \emptyset$, 其中 $\sum_{B, C \subseteq \Omega, B \cap C = \emptyset} m_1(B) m_2(C) \neq 1$ 。Dempster 合成规则满足交换性和结合性,允许空 BBA 合成。设 $m_{12}(\varphi) = \sum_{B, C \subseteq \Omega, B \cap C = \varphi} m_1(B) m_2(C)$ 称为 Dempster 规则中正规化前分配给空集的合成信念 mass。

2.2 基于证据推理的云计算服务适应性评估建模

设多级分布式适应性评估的总评价记为 V , 安全与风险、数据、服务、资源、经济五个维度记为向量 (T, D, S, R, E) , 如图 3 所示, 将 V 细分为五个维度 (T, D, S, R, E) 共 M 层来评价, 各层的节点数为 $N_1, N_2, \dots, N_M$ 。第 h 层上的节点记为 $D_{i,j}^h (1 \leq i \leq N_h, 1 \leq j \leq N_{h-1}, 1 \leq h \leq M)$, 表示 h 层上的第 i 个节点, 其父节点为 $(h-1)$ 层上的第 j 个节点。另外, 各层次上每个节点的权重记为 $W_{i,j}^h (1 \leq i \leq N_h, 1 \leq j \leq N_{h-1}, 1 \leq h \leq M)$, 满足 $0 \leq W_{i,j}^h \leq 1$, 且 $\sum_{i,j=k} W_{i,j}^h = 1, 1 \leq k \leq N_{h-1}$ 。为每一个叶节点设置合适的识别框架, 记为 $\Omega_1, \Omega_2, \dots, \Omega_N$, 允许识别框架的长度相同但实际评价意义不同。依此, 各节点的评估数据可以记为 $B(D_{i,j}^h) = \{(H_{ni}, \theta_{ni,i}), ni = 1, \dots, NI, NI = |\Omega_i|\}, \theta_{ni,i} \geq 0$ 。如果 $\sum_{ni=1, \dots, NI} \theta_{ni,i} = 1$, 则 $B(D_{i,j}^h)$ 为完全评价; 否则称为不完全评价。

对叶节点评估数据进行由下至上逐级合成时, 需要将评估数据统一到公共识别框架 Ω 上。对于定性评价结果, 可以直接由 BBA 表示, 而对于定量评估数据 (成本型和效益型), 需要先正规化再表示成一个 BBA。效益型定量评估值的效用正规化方法为

$$u(D_{i,j}^h(y)) = \frac{y - y_i^{\min}}{y_i^{\max} - y_i^{\min}} \quad i \in \Omega_i$$

类似地, 成本型定量评估值的效用正规化方法为

$$u(D_{i,j}^h(y)) = \frac{y_i^{\max} - y}{y_i^{\max} - y_i^{\min}} \quad i \in \Omega_i$$

同时, 一个定量评估值 y_i 为一定可以用两个相邻的精确评估等级所示, 即 $y_{ni,i} \leq y_i \leq y_{ni+1,i}$, 令

$$\theta_{ni,i} = \frac{y_{ni+1,i} - y_i}{y_{ni+1,i} - y_{ni,i}}, \theta_{ni+1,i} = \frac{y_i - y_{ni,i}}{y_{ni+1,i} - y_{ni,i}}$$

以上 $u(D_{i,j}^h(y))$ 、 $y_i^{\min}$ 、 $y_i^{\max}$ 、 $\theta_{ni,i}$ 、 $\theta_{ni+1,i}$ 分别表示叶节点 $D_{i,j}^h(y)$ 上定量评估值 y 的效用、最大值、最小值、信念度下限和上限, 则定量评估值 y_i 可以表示为

$$y_i \Leftrightarrow \{(H_{ni}, \theta_{ni,i}), (H_{ni+1}, \theta_{ni+1,i})\}$$

各叶节点评估结果转换为 BBA 后, 由专家给出各识别框

架相对于公共识别框架的转换矩阵, 将不同识别框架上的评估数据映射到公共识别框架 Ω 上^[14]。接着对统一到公共识别框架上的数据进行由下至上逐级合成。这里的合成算法采用 Wang 等人在文献[15]中提出的解析算法, 该算法在考虑了各个叶节点的相对权重之后可以有效合成冲突数据, 得到的根节点评估数据记为 $B(V) = \{(H_n, \theta_n), n = 1, \dots, N\}$, 满足 $\theta_n \geq 0$ 且 $\sum_{n=1, \dots, N} \theta_n = 1$ 。最后选取效用准则对根节点评估数据进行评估, 将 $[0, 1]$ 区间划分为 N 个子区间, 合成 $B(V)$ 效用并比较对应的子区间得出结论, 合成规则为

$$u(B(V)) = \sum_{n=1}^N u(H_n) \times \theta_n + \frac{u(H_1) + u(H_N)}{2} \times (1 - \sum_{n=1}^N \theta_n)$$

$$\text{且 } 0 \leq u(B(V)) \leq 1$$

本文证据是专家对云计算服务适应性的评价。专家是具备特定问题相关的专业知识、信息、数据、方法、模型的智能体, 因此认为证据间是相互独立的, 可以直接合成。

2.3 评估步骤

基于以上分析, 下面给出云计算服务适应性评估的具体流程:

- 根据实际问题的需要, 选取合适的指标构建评估模型, 确定各级指标的权重, 并确定所有叶节点的识别框架、公共识别框架, 以及框架集上各等级的专家效用;
- 邀请相关专家对各个叶节点进行评价, 收集专家评估数据, 这些数据包括定性评估数据和定量评估数据;
- 把收集到的专家评估数据中的定量精确评估数据按照各自叶节点的框架等级效用转换为 BBA;
- 将所有叶节点评估数据 BBA 按照效用转换方法统一为公共识别框架上的评估数据;
- 运用合成算法由下至上对所有评估数据逐层合成并形成最终评估数据 $B(V)$;
- 利用效用准则来判断最终评估数据 $B(V)$ 的等级, 从而确定服务适应性程度。

3 实例分析

某大学研究所一直从事商务智能方面的研究, 经过近些年的积累在领域内取得了较好的成就, 认可度也不断提升。但是目前遭遇的瓶颈是商务智能方面的实验经常需要处理海量数据, 运算负荷大, 对硬件性能要求高, 而研究所当前状况是仅拥有 3 台老旧式服务器, 实验用计算机 44 台, 及打印设备、扫描仪、网络设备若干, 越来越不能满足实验的需要。现考虑到云计算模式的逐渐成熟, 研究所想要评估云计算服务模式是否适合解决研究所目前所面临的对硬件资源的需求问题, 以及项目研究过程中与合作单位频繁交流及数据共享与转换的问题。

针对上面某研究所的实际情况及业务需要, 根据前文提出的云计算服务适应性评估指标体系构建评估模型, 给出各级指标的权重、识别框架及专家评估数据等, 限于篇幅, 直接在评估指标体系表上构建模型, 给出相应数据, 如表 1 所示。

由表 1 可知存在四种识别框架, 即 $\Omega_1 = \{H_{11}, H_{12}, \dots, H_{16}\}$, $\Omega_2 = \{H_{21}, H_{22}, \dots, H_{25}\}$, $\Omega_3 = \{H_{31}, H_{32}, \dots, H_{34}\}$, $\Omega_4 = \{H_{41}, H_{42}, H_{43}\}$, 各识别框架上评价等级集的专家效用依次为 $(0, 0.35, 0.55, 0.7, 0.9, 1)$ 、 $(0, 0.4, 0.7, 0.8, 1)$ 、 $(0, 0.4, 0.8, 1)$ 、 $(0, 0.6, 1)$ 。设定公共识别框架为 $\Omega = \{H_1, H_2, \dots, H_5\} = \{\text{极差, 较差, 一般, 良好, 优秀}\}$ 。 Ω_2 上的评估数据无须

转换直接等价于 Ω 下的评估数据,其他识别框架上的评估数据需根据规则转换矩阵转化到 Ω 上,其中规则转换矩阵分别为

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}_{5 \times 5}, \begin{pmatrix} 0.8 & 0.2 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 \end{pmatrix}_{5 \times 5}, \begin{pmatrix} 0.8 & 0.2 & 0 & 0 & 0 \\ 0 & 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 \end{pmatrix}_{3 \times 5}$$

运用效用准则划分区间 $[0,1] = [0,0.2] \cup [0.2,0.45] \cup [0.45,0.6] \cup [0.6,0.8] \cup [0.8,1]$,分别对应 Ω 的五个等级。按照前文叙述的步骤和方法逐步进行转换、合成,由下至上逐级合成评估模型中的评估数据得 $B(V) = \{(H_1, 0.0785), (H_2, 0.2024), (H_3, 0.2818), (H_4, 0.4454), (H_5, 0.1264)\}$, $u(B(V)) = 0.71$ 。

根据效用准则可得评估等级对应 Ω 中的“良好”,因此建议采用云计算模式解决研究所的当前问题。评估结果没有呈现出显著优势的一个重要原因在于用户对云计算服务模式在安全和服务透明度方面的担忧,这也与现实情况相符合。

4 结束语

尽管云计算目前在商业界取得了一定的应用,但其目前的技术和管理水平离开放、互通、融合、安全和标准都尚有一段距离,这些问题都会对云计算的广泛应用产生重大影响,因此用户需要对云计算的综合能力进行有效评估才能作出正确的决策。本文首先构建了一套云计算服务适应性评估指标体系,而后运用证据推理的方法选择一个算例进行验证,并取得了比较合理的结果,进一步证明了本文构建的评估指标体系的有效性和合理性。

下一步将进一步研究针对不同应用场景下的服务适应性以及服务提供商选择的问题。

参考文献:

- [1] MARSTON S, LI Zhi, BANDYOPADHYAY S, *et al.* Cloud computing: the business perspective [J]. *Decision Support Systems*, 2011, 51(1): 176-189.

(上接第4286页)缺失数据的各节点经过第1次线性组合融合后,再利用第2次自适应加权融合的误差较小。

4 结束语

针对含有缺失数据的无线传感器网络数据融合问题,提出了一种简单易行的二次数据融合算法(TFA)。本文充分考虑无线传感器网络的时空特性,对缺失数据利用线性组合融合方法进行第1次数据融合,融合效果要比只考虑时间相关性的线性插值方法和只考虑空间相关性的回归分析方法效果好。为了能够形成比单一传感器更准确、更全面的观测值,对各传感器同次采集的数据进行第2次自适应加权融合,融合效果较好。

参考文献:

- [1] 李建中, 李金宝, 石胜飞. 传感器网络及其数据管理的概念、问题与进展[J]. *软件学报*, 2003, 14(10): 1717-1727.
- [2] 崔莉, 鞠海玲, 苗勇, 等. 无线传感器网络研究进展[J]. *计算机研究与发展*, 2005, 42(1): 163-174.
- [3] 张强, 杨涛. 用于环境监测的自供电传感器网络[J]. *仪表技术与传感器*, 2008(2): 34-36.

- [2] ARMBRUST M, FOX A, GRIFFITH R, *et al.* Above the clouds: a berkeley view of cloud computing [EB/OL]. [2007-02-10]. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html>.
- [3] 迟学斌. 云计算不适合科学计算 [EB/OL]. [2008-12-11]. <http://server.chinabyte.com/139/8620139.shtml>.
- [4] “云计算”行业热点问题分析——五大问题 [EB/OL]. [2011-01-12]. <http://vccbook.blog.51cto.com/1589501/478151>.
- [5] 蔡盈芳. 基于云计算的信息系统安全风险评估模型 [J]. *中国管理信息化*, 2010, 13(12): 75-77.
- [6] MISRA S C, MONDAL A. Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment [J]. *Mathematical and Computer Modelling*, 2011, 53(3-4): 504-521.
- [7] KRUGMAN P R, OBSTFELD M. *International economics: theory and policy, economies of scale* [M]. 6th ed. [S. l.]: Imperfect Competition & International Trade, 2003.
- [8] 罗军舟, 金嘉晖, 宋爱波, 等. 云计算: 体系架构与关键技术 [J]. *通信学报*, 2011, 32(7): 3-20.
- [9] DARGHA R. Cloud computing, infosys white paper [S]. 2009.
- [10] 里特豪斯. 云计算实现、管理与安全 [M]. 田思源, 赵学锋, 译. 北京: 机械工业出版社, 2010: 105-123.
- [11] 冯登国, 张敏, 张妍, 等. 云计算安全研究 [J]. *软件学报*, 2011, 22(1): 71-83.
- [12] Distributed Management Task Force. DSP-ISO101, Interoperable clouds, a white paper from the open cloud standards incubator [S]. 2009.
- [13] 马泽尔. 云计算安全与隐私 [M]. 刘戈舟, 杨译明, 刘宝旭, 译. 北京: 机械工业出版社, 2011: 63-66.
- [14] YANG Jian-bo. Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties [J]. *European Journal of Operational Research*, 2001, 131(1): 31-61.
- [15] WANG Ying-ming, YANG Jian-bo, XU Dong-ling. Environmental impact assessment using the evidence reasoning approach [J]. *European Journal of Operational Research*, 2006, 174(3): 1885-1913.

- [4] 王飞. 无线传感器网络中基于模糊理论的决策级数据融合技术研究 [D]. 北京: 北京交通大学, 2010.
- [5] 孙凌逸, 黄先祥, 蔡伟, 等. 基于神经网络的无线传感器网络数据融合算法 [J]. *传感技术学报*, 2011, 24(1): 122-127.
- [6] 马守明, 王汝传, 叶宁. 基于主元分析的无线传感器网络数据融合研究 [J]. *计算机工程与科学*, 2010, 32(11): 44-46, 65.
- [7] 张书奎, 崔志明, 龚声蓉, 等. 基于 Bayes 序贯估计的无线传感器网络数据融合算法 [J]. *电子与信息学报*, 2009, 31(3): 716-721.
- [8] 任秀丽, 田洋. 传感器网络中基于预处理证据理论的数据融合 [J]. *计算机应用*, 2011, 31(7): 1992-1994.
- [9] HALATCHEV M, GRUENWALD L. Estimating missing values in related sensor data streams [C]//Proc of the 11th International Conference on Management of Data. 2005: 83-94.
- [10] JIANG Nan, GRUENWALD L. Estimating missing data in data streams [C]//Proc of the 12th International Conference on Database Systems for Advanced Applications. 2007: 981-987.
- [11] 何丽娟, 周鸣争, 陶皖, 等. 无线传感器网络中不确定数据的估计算法 [J]. *计算机工程与应用*, 2011, 47(28): 100-102, 161.
- [12] 蔡科, 左宪章, 康健, 等. 无线传感器网络测试系统数据融合研究 [J]. *计算机仿真*, 2010, 27(7): 1-4, 9.

其中: β_L 和 β_R 为线性插值插补和回归分析插补的组合融合系数。可利用求解出的 β_L 和 β_R , 根据式(9)对缺失数据 $x_i(k)$ 进行线性组合融合。

$$\hat{x}_i(k) = \beta_L \hat{x}_{i,L}(k) + \beta_R \hat{x}_{i,R}(k) \quad (9)$$

2 自适应加权融合算法

由于无线传感器网络规模大且分布密度高, 这样会引起监测区域的重叠, 导致临近节点报告的信息存在一定程度的冗余性。为了能够将来自不同节点的测量数据进行综合分析和处理, 在减少能量消耗的同时形成比单一传感器更可靠、更全面的融合值, 需要对各节点的数据进行第二次融合。

将缺失数据线性组合融合的结果记录成矩阵的形式:

$$Y = \begin{pmatrix} y_1(1) & y_2(1) & \cdots & y_n(1) \\ y_1(2) & y_2(2) & \cdots & y_n(2) \\ \vdots & \vdots & & \vdots \\ y_1(\tau) & y_2(\tau) & \cdots & y_n(\tau) \end{pmatrix} \quad (10)$$

如果 $x_i(k)$ 为缺失数据, 则 $y_i(k) = \hat{x}_i(k)$, 否则 $y_i(k) = x_i(k)$ 。为了得到第 k 次各节点采集数据的综合结果, 本文对第 k 行进行多传感器数据融合。如果 $y_i(k)$ 与第 k 行其余数据的支持度越高, 则 $y_i(k)$ 的真实性就越高; 反之, 真实性就越低。为了量化不同传感器在同一时刻观测值的支持程度, 利用模糊数学中的隶属度函数来定义第 k 次测量数据中传感器 i 与 j 之间的支持度为

$$s_{i,j}(k) = 2 \arccot(|y_i(k) - y_j(k)|) / \pi \quad i, j = 1, 2, \dots, n \quad (11)$$

根据实际问题应用背景, 可以设置一个支持度为 0 的距离阈值 $\delta \geq 0$; 当 $|y_i(k) - y_j(k)| \geq \delta$ 时, $s_{i,j}(k) = 0$ 。则 n 个传感器之间第 k 次采集数据的支持度矩阵 $S(k)$ 为

$$S(k) = \begin{pmatrix} s_{1,1}(k) & s_{1,2}(k) & \cdots & s_{1,n}(k) \\ s_{2,1}(k) & s_{2,2}(k) & \cdots & s_{2,n}(k) \\ \vdots & \vdots & & \vdots \\ s_{n,1}(k) & s_{n,2}(k) & \cdots & s_{n,n}(k) \end{pmatrix} \quad (12)$$

$s_{i,j}(k)$ 仅表示测得数据 $y_j(k)$ 对 $y_i(k)$ 的支持程度, 并不能反映系统中所有传感器所测得数据对 $y_i(k)$ 的支持程度, 而需要综合 $S(k)$ 的第 i 行全部元素, 因而各传感器节点对应的加权系数 $w_i(k)$ 应综合各子系统 $s_{i,1}(k), s_{i,2}(k), \dots, s_{i,n}(k)$ 的全部信息, 所以需要求出一组非负数 $a_1(k), a_2(k), \dots, a_n(k)$ 使

$$w_i(k) = a_1(k)s_{i,1} + a_2(k)s_{i,2} + \cdots + a_n(k)s_{i,n} \quad (13)$$

假设 $W = [w_1(k), w_2(k), \dots, w_n(k)]^T$, $A = [a_1(k), a_2(k), \dots, a_n(k)]^T$, 将式(13)改写成矩阵形式为

$$W = S(k)A \quad (14)$$

因为 $s_{i,j}(k) \geq 0$, 所以矩阵 $S(k)$ 是一个非负矩阵。由非负对称阵的性质可知, $S(k)$ 存在最大模特征值 $\lambda (\lambda > 0)$ 及其对应的正特征向量 A , 使得 $\lambda A = S(k)A$, 则 λA 就可以作为多个节点测量数据的综合关联程度的度量, 从而权重 $w_i(k)$ 即为 A 归一化后的结果, 因此

$$w_i(k) = \lambda a_i / \sum_{i=1}^n \lambda a_i = a_i / \sum_{i=1}^n a_i \quad (15)$$

则第 k 次各节点采集的数据加权融合的结果为

$$y(k) = \sum_{i=1}^n w_i(k) y_i(k) \quad (16)$$

3 仿真实验

本文在 MATLAB 6.5 的环境下进行实验, 以文献[12]中四个液压传感器节点对不同的压力状况进行实验, 其测试值如表 1 所示。

表 1 四个传感器节点的八次测量数据

采集次数	传感器 1	传感器 2	传感器 3	传感器 4	真实值
第 1 次	4.93	4.98	4.94	4.97	5
第 2 次	4.94	5.01	4.94	5.01	5
第 3 次	7.96	7.88	7.98	7.86	8
第 4 次	9.96	9.84	9.96	9.75	10
第 5 次	11.96	11.94	11.96	11.61	12
第 6 次	17.98	17.92	17.99	17.91	18
第 7 次	22.93	22.76	21.02	22.75	22
第 8 次	25.03	24.87	25.03	24.88	25

为了测试实验效果, 随机地将表 1 中节点 1 第 3 次采集的缺失数据和节点 4 第 5 次采集的缺失数据标记成缺失值 NaN1 和 NaN2, 并对这些标记为缺失值的数据进行线性组合融合估计, 通过用估计值与其对应的真实值进行比较来评价本文算法性能。

对 NaN1 利用其前后的真实数据得到线性插值插补系数和回归分析插补系数分别为 -0.042 3 和 1.053 7, 对 NaN2 利用其前后的真实数据得到线性插值插补系数和回归分析插补系数分别为 -0.068 3 和 1.069 7, 从而采用本文方法进行第 1 次线性组合融合的结果为 NaN1 = 7.9887, NaN2 = 11.8053。

为了评价算法的性能, 本文利用线性组合融合后的结果与原始值的均方根误差 RMSE 作为度量标准, 并与线性插值插补算法、回归分析插补算法、平均值插补算法进行了比较, 如表 2 所示。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{x}_i(k) - x_i(k))^2}{n}}$$

其中: n 为缺失数据的个数。

表 2 第 1 次融合结果及各种方法估计误差

估计结果	线性插值	回归分析	平均值	线性组合融合	真实值
NaN1	7.450 0	7.880 5	7.906 7	7.988 7	7.96
NaN2	13.830 0	11.918 9	11.953 3	11.805 3	11.61
RMSE(Pa)	1.610 7	0.225 5	0.245 7	0.139 6	

从表 2 中可以看出, 综合考察感知数据的时空相关性, 使用本文线性组合融合后的效果较单独基于时间相关性的线性插值算法和单独基于空间相关性的回归分析算法要好, 且比使用平均值算法要好。

对利用线性组合融合后的数据进行自适应加权融合, 以第 1 次各传感器的测量值为例, 得到最大模特征值为 3.942 7, 对应的特征向量为 0.499 2、0.499 2、0.500 8、0.500 8, 从而得到第 1 次测量各传感器的权 $w = [0.2496, 0.2496, 0.2504, 0.2504]$, 融合结果为 4.955 0, 误差为 0.045。同理, 可得其他次融合结果及误差, 如表 3 所示。

表 3 第 2 次融合结果及误差

采集次数	融合结果	误差	采集次数	融合结果	误差
第 1 次	4.955 0	0.045 0	第 5 次	11.917 6	0.082 4
第 2 次	4.975 0	0.025 0	第 6 次	17.950 0	0.050 0
第 3 次	7.927 2	0.072 8	第 7 次	22.562 4	0.562 4
第 4 次	9.878 5	0.121 5	第 8 次	24.952 6	0.047 4

从表 3 中可以看出, 对第 3 次和第 5 次含有 (下转第 4308 页)