

文章编号: 1000-5641(2018)03-0067-10

面向食品溯源数据服务的多 QoS 约束 服务选择优化算法研究

袁培森¹, 黎薇¹, 任守纲^{1,2}, 徐焕良^{1,2}

(1. 南京农业大学 信息科学技术学院, 南京 210095;

2. 江苏省肉类生产与加工质量安全控制协同创新中心, 南京 210095)

摘要: 数据服务的理念是大数据时代一种重要的信息服务方式. 在面向服务构架 (Service-Oriented Architecture, SOA) 框架下研究了用于食品安全溯源系统中面向数据服务的多 QoS (Quality of Service) 约束下服务组合选择优化算法. 溯源服务系统是一种数据密集型服务系统, 系统的实现需要组合多源的 Web 服务. 通过提取系统的重要 Web 服务和数据服务 QoS 指标, 建立了食品安全溯源系统的以数据服务为场景的多服务属性约束下优化因子模型—MQBR (Multi-QoS based Benefit Ratio) 模型, 采用 Skyline 算法预处理并结合人工智能的启发式方法求解了满足约束条件的服务组合, 提升了系统服务选择的效率和质量. 通过实验分析证明了算法的有效性, 且实现了多 QoS 约束条件下食品安全溯源系统中服务组合选择算法, 提升了服务选择的质量和性能.

关键词: 食品安全溯源; Skyline 算法; 多 QoS 约束; 面向数据服务的服务组合; 启发式算法

中图分类号: TP391 文献标志码: A DOI: 10.3969/j.issn.1000-5641.2018.03.008

Algorithm for service optimization under multi-QoS constraints for data services in a food traceability system

YUAN Pei-sen¹, LI Wei¹, REN Shou-gang^{1,2}, XU Huan-liang^{1,2}

(1. College of Information Science and Technology, Nanjing Agricultural University, Nanjing 210095, China;

2. Jiangsu Collaborative Innovation Center of Meat Production and Processing, Quality and Safety Control, Nanjing 210095, China)

Abstract: The concept of data services plays an important role in the era of big data. In this paper, an optimization algorithm for web services in food safety traceability is

收稿日期: 2017-05-16

基金项目: 国家自然科学基金 (61502236); 中央高校基本科研业务费专项资金 (KYZ201752, KJQN 201651); 国家科技支撑计划 (2015BA1105000); 江苏省重点研发计划 (BE2016803)

第一作者: 袁培森, 男, 博士, 讲师, 研究方向为智能计算、海量数据管理.

E-mail: peiseny@njau.edu.cn.

第四作者: 徐焕良, 男, 博士, 教授, 研究方向为农业信息化与大数据技术.

E-mail: huanliangxu@njau.edu.cn.

investigated based on the SOA (Service-Oriented Architecture) framework. Traceability services are commonly data-intensive systems, which need to combine multi-source web services. In this paper, by extracting important QoS (Quality of Service) indexes from the web and data services of the system, a multi-QoS based benefit ratio (MQBR) is established, which is then used on the traceability platform for food security management. Based on the MQBR model, the skyline and heuristic method of artificial intelligence is proposed for optimizing the efficiency and quality of service selection. Experiments are conducted to prove the validity of the algorithm. The methods of our study are designed and applied to a food security management application, with multiple QoS constraints in the traceability systems, to improve the overall performance and service quality.

Keywords: food safety traceability system; skyline algorithm; multi-QoS constraints; data service composition; heuristic-based algorithm

0 引 言

随着服务计算技术 (Service Oriented Computing, SOC) 和应用的迅猛发展, Web 服务计算成为近年来的研究热点^[1-4], 服务计算已成为一种重要的跨学科的计算模式, 为企业的业务提供技术和架构支撑^[4]. 当前, 云计算、大数据、IoT 等新应用的兴起, 服务计算的理念更加深入人心^[4-5], 使得 Web 服务成为快速构建应用系统的基石^[6]. 同时, 数据作为一种核心资产, 数据服务能力成为企业的核心竞争力.

如何在大量的 Web 服务中检索到满足用户多种需求约束的服务是一个巨大的挑战^[1,7-8]. 服务质量 (Quality of Service, QoS) 使用服务等级协议 (Service-Level Agreement, SLA) 指标不仅从技术上, 而且从服务性能上来评判、衡量服务, 是重要的服务择优途径. 基于 QoS 的服务组合目的在于从服务中选择满足约束的最优服务集合, 其中多目标 QoS 约束的组合服务选择问题是 NP (Non-deterministic Polynomial) 难问题, 难点在于庞大的解空间, 无法在多项式时间内完成搜索^[9-10]. 目前多 QoS 约束的组合服务选择重点在于 QoS 模型研究和搜索算法研究上, 需要根据各领域的特点, 构建服务的 QoS 模型^[11-12]. 面向数据服务的组合服务问题, 主要的问题旨在给定的代价开销范围内选择符合要求的组合服务^[13].

食品安全溯源是农业领域食品安全管理的一个重要手段^[14], 系统确保食品质量安全及其相关信息能够被顺向追踪 (生产源头→消费终端) 或者逆向回溯 (消费终端→生产源头), 从而使食品的整个生产经营活动始终处于有效监控之中. 食品从产出到成品涉及种、养殖、生产等, 环节繁复, 食品生产加工程序多、配料多, 涵盖了生产、加工、包装、储运、销售等多环节, 每个阶段均伴随着海量数据处理需求. 因此, 该领域数据量符合“4Vs”^[15]的大数据定义, 涵盖了数据生成、数据获取、数据存储和分析完整的大数据价值链. 除此之外, 系统更重在数据服务环节, 即为公众提供数据查询服务. 以数据为中心的 Web 服务组合聚焦数据的供给和更新, 并伴随着大量的数据操作和交互^[16].

本文构建基于 SOA (面向服务构架) 框架下的食品溯源数据服务系统, 以现有 Web 服务为基础, 例如 GIS 服务^①、二维码计算服务^②等, 实现了食品安全溯源 (Food Safety Traceability, FST). FST 系统综合应用物联网、供应链信息、移动互联网、GIS 等信息技术, 定量描述食品生产全过程, 实现了生产记录可存储、产品流向可追踪、储运信息可查询、

① <http://map.baidu.com/>; <https://lbs.amap.com/api/webservice/summary/>

② <http://apistore.baidu.com/apiworks/servicedetail/2773.html>

产品质量有保证. 数据服务层使用统一的开发和数据分析工具自动地定位存储的异构数据, 独立地查询和分发应用请求的数据, 系统框架图如图1所示. FST系统的有效构建需要集成多种的服务, 服务的性能、可用性及可靠性以及数据服务的质量指标等对系统的稳定运行至关重要; 此外为了确保服务的质量和灵活性, 服务还必须满足硬约束和软约束. 因此研究在多约束条件下的服务组合对实现FST系统具有重要的现实意义. 由于Web服务海量性和多样性, 如何从众多的服务中有效地检索满足系统需求的服务成为系统实现的关键.

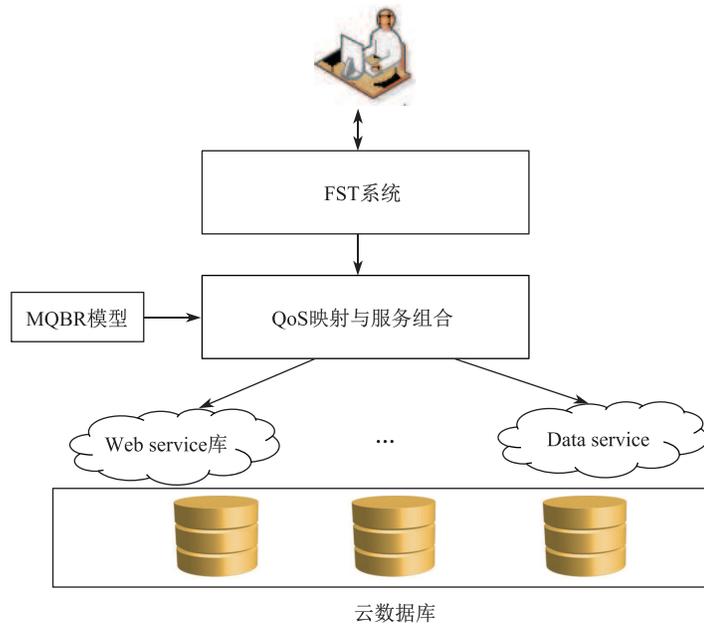


图1 基于Web service的FST系统框架图

Fig.1 Framework of FST system based on web service

针对上述问题, 本文提出了FST系统服务组合模型, 并对每种结构的QoS属性特点进行了分析. FST系统主要采用5种Web服务QoS指标和3种数据服务QoS指标, 并对其进行了定量分析, 提出了MQBR模型. 最后, 结合人工智能的启发式搜索技术, 在多QoS属性约束的Web服务组合上选择过程. 通过实验验证和分析, 本文方法能够有效实现服务组合.

1 服务组合模型

1.1 Web服务组合结构

Web服务实现了软件的部分功能, 功能复杂的软件可通过集成了众多的Web服务来构造, 这个过程就是Web服务组合^[1,17].

在Web服务组合过程中, 不可分解的功能称为原子功能, 满足原子功能的服务, 称为原子服务. 功能和服务之间存在多对多的关系, 服务又具有不同的QoS属性. 因此, 存在多个具有不同QoS属性的服务组合方案, 可以满足系统功能需要. 基于QoS的服务选择方法从众多服务中选出一个, 既符合用户QoS约束, 又是最优的方案. 基本的Web服务的组合由4种基本组合子结构, 分别为顺序关系、并行关系、分支关系和循环关系. 服务组合子结构如

图2所示,根据这些子结构的不同,所选组合服务集的QoS属性值也不同.

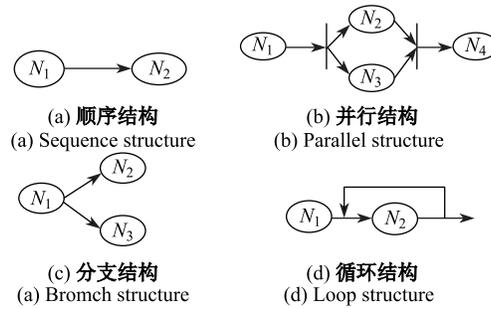


图2 服务组合子结构

Fig.2 Substructures of web service

1.2 QoS属性及度量

常用的QoS属性主要包括成本、响应时间、可靠性、可用性、安全性等^[18-19].针对数据服务的QoS除了包含一般的Web服务的QoS指标,还包括数据服务独有的指标,例如数据访问成功概率、系统数据自愈时间等.根据特性的不同,可分为可加性度量、可乘性度量、最小性度量和最大性度量这4类度量.

在服务评价的QoS属性中,主要以可加性度量和可乘性度量为主,本文主要以这两种度量为研究对象.对于以上度量,又可分为正相关和负相关.所谓正相关可乘性度量,即值越大越好;负相关可加性相反,值越小越好.对于可加性度量,在图2所示的服务组合子结构中,各QoS属性值的计算方法:给定服务 S_1 和服务 S_2 ,某个QoS属性 j ,服务 S_1 的 j 属性值为 $w_j^{S_1}$,服务 S_2 的 j 属性值为 $w_j^{S_2}$,组合服务的 j 属性值为 w_j^S .可加性与可乘性的计算如表1所示.

表1 可加性与可乘性计算

Tab.1 Evaluation of additivity and multiplicativity properties				
顺序	分支	循环	并行	
可加性	$w_j^S = w_j^{S_1} + w_j^{S_2}$	$w_j^S = w_j^{S_1}$ 或 $w_j^S = w_j^{S_2}$	$w_j^S = k \times w_j^{S_1}$, 服务 S_1 循环 k 次	$w_j^S = \max(w_j^{S_1}, w_j^{S_2})$
可乘性	$w_j^S = w_j^{S_1} \times w_j^{S_2}$	$w_j^S = w_j^{S_1}$ 或 $w_j^S = w_j^{S_2}$	$w_j^S = (w_j^{S_1})^k$, 服务 S_1 循环 k 次	$w_j^S = \max(w_j^{S_1}, w_j^{S_2})$

根据上述定义,可以计算出组合服务的QoS属性总值.顺序关系为相加或相乘;分支关系为某一分支的值;循环关系,对于可加性对量,值为循环次数与属性值的乘,对于可乘性度量,值为属性值的循环次数的指数;并行关系为最大的一个分支的属性值.

2 FST系统中的QoS属性及组合方法

2.1 FST系统

FST系统主要采用SOA.SOA是一种设计方式,也是一种定义和提供IT基础设施的方式,通过服务组合实现功能集成的系统,具有良好的模块化、跨平台、互操作和可扩展特性.FST系统包含6大服务模块,即通用业务服务(Comm Business Services)、电商交互服务(EBusiness Service)、移动平台溯源服务(Mobile Trace Service)、企业端溯源服务(Enterprise Trace Service)、溯源业务核心服务(Core Trace Service)、系统用户管理服务(User Admin Service)等,其中溯源查询、用户资料、产品信息等都需要系统支持大规模、高并发的数据访问.FST系统以工作流的方式提供物品的溯源和追踪,系统采用REST API和SOAP API两种服务调用方式.

FST系统作为面向公众的数据查询服务,数据托管在云存储之上,采用数据服务(Data as a Service, DaaS)理念,主要使用SQL Data Services. DaaS本质是Web服务在数据方面的应用,它提供了系统数据存储的抽象,提供从不同平台和系统中访问数据的编程接口.该层基于统一开发的数据分析工具,自动定位和访问存储云端数据,独立地完成查询和分发应用请求.主要的数据服务操作包括模型操作、实体操作、属性操作、事物操作、数据质量操作、数据导入导出操作及数据安全性操作等.

FST系统中对每个服务标识基本属性,如开发语言、所用算法、接口结构、开发组织、开发日期等.此外,设置每个服务的QoS属性,包括反应时间、服务可用性、服务可靠性等.FST系统生产记录的获取及存储、溯源查询等涉及多种Web服务组合.作为一种以数据服务为核心的系统,服务的吞吐量、服务的响应时间、服务的可靠性以及服务组合的代价为该系统的重要指标.

2.2 FST中的QoS属性

FST系统中的Web服务关键QoS指标和数据服务关键指标如下,其中(6)至(8)为系统数据服务的关键指标.下面对FST系统中重要的QoS属性进行分析和量化.

(1) 服务响应时间SRT(Service Response Time),记为 t_i ,该属性属于可加性度量.总的响应时间 $SRT(C) = \sum_{i=1}^N t_i$,其中, $SRT(C)$ 表示服务组合 C 的反应时间, t_i 表示服务组合里面第 i 个服务的反应时间, N 表示服务组合里面服务的个数.

(2) 服务可用性SA(Service Availability),该属性属于正相关可加性度量.该指标难以直接进行量化,因此通过间接的用户引用频率来度量,即用户采用该服务的次数越多,表面该服务的可用性也就越高.为了在不同服务之间进行属性值的相加处理,系统采用服务引用次数与满足同一功能的所有服务里面引用的服务最高次数之间的相对比率来代表引用率,为 $SA(C) = \sum_{i=1}^N u_i$,其中, SA 表示服务组合 C 的可用性, $u_i = c_i / \max(c_{i1}, \dots, c_{ij}, \dots, c_{iM})$, $i = 1, \dots, N, j = 1, \dots, M$,表示服务组合里面第 i 个服务的可用性, c_i 表示服务组合里面第 i 个服务被用户引用的次数, N 表示服务组合里面服务的个数, M 表示服务 c_i 对应的功能 F_i 、满足 F_i 的所有服务的个数.

例如,满足某一功能要求的服务有10个,其中最高的用户引用次数为10次,那么引用次数最高的服务的引用率为 $10/10=1$,若某服务引用次数为6,则引用率为 $6/10=0.6$.

(3) 服务可靠性SR(Service Reliability),该属性属于可乘性度量.系统采用用户反馈的故障率来间接度量.每个服务的故障率的计算方法为,用户反馈的故障次数与用户引用该服务的次数之比,因此故障率越低,则服务的可靠性越高.计算公式为 $SR(C) = \prod_{i=1}^N k_i$,其中, SR 表示服务组合 C 的可靠性, $k_i = 1 - f_i/c_i$, $i = 1, \dots, N$, k_i 表示服务组合里面第 i 个服务的可靠性, N 表示服务组合里面服务的个数, f_i 表示服务组合里面第 i 个服务用户反馈的故障次数, c_i 表示服务组合里面第 i 个服务被用户引用的次数.

例如,如果某个服务,用户引用了10次,而反馈了6次故障,则该服务的故障率为 $6/10=0.6$,即可靠性为0.4.

(4) 服务吞吐量ST(Service Throughput)是服务成功执行的平均次数.对于服务组合 C , $t(S)$ 为服务 S 的吞吐量,则服务组合 C 的吞吐量 $ST(C) = \min_{s \in C} t(S)$.

(5) 服务代价SC(Service Cost)代表了用户使用服务需要支付的费用,计算公式为 $SC(C) = \sum_{S \in C} c(S)$.

(6) 数据成功访问率DASR(Data Access Success Rate)是成功连接到数据库的概率.

(7) 数据自愈时间DSHT(Data Self Healing Time),即系统错误检测和自我恢复的时间.

(8) 数据吞吐量 DT (Data Throughput), 即数据服务在单位时间内的数据平均传输速率.

此外, FST 系统作为一种主要以数据服务为基础的平台, 用户对数据服务的代价比较关注, 代价可以通过计算资源的开销量化.

2.3 FST 系统的 Web 服务约束和组合选择

基于约束条件下的服务组合问题描述如下.

给定 1 组 QoS 约束 $C_{QoS} = \{c_1, c_2, \dots, c_M\}$, 在此约束条件下的一组可行服务组合 FS 定义为: 给定服务组 $P = \{S_1, S_2, \dots, S_n\}$, 从 P 中选择 N 个服务类的任一服务 $S_i, i \in [1, N]$, 满足 $C_{QoS}(S_i) \leq c_k, k \in [1, M]$. 所有的可行服务组合构成候选服务组合集 (Candidate Service Composition, CSC), $CSC = \bigcup_{i=0} FS$. 最优的服务组合是在所有的可行服务组合中选择代价最小的 1 组服务.

为了保证 FST 系统中服务的质量和灵活性, 把系统中 QoS 属性约束的 Web 服务组合选择问题抽象为: 从每个功能结点中选择一个候选服务组成候选服务集, 该候选服务集包括以下两类约束.

(1) 硬约束 (Hard Constraint) 表示候选服务集的组合属性值必须满足该约束. 硬约束可以分为组合硬约束和节点服务硬约束, 如对于反应时间, 可以给定一个组合的约束时间, 这属于组合硬约束; 而对于可靠性, 可以给定每个服务的可靠性约束, 这属于节点服务硬约束.

(2) 软约束 (Soft Constraint) 又称为偏好^[20], 表示如果候选服务集满足该约束的可能性越大, 则服务集的 QoS 总值越优, 用户满意度越高. 如服务可用性约束, 在满足硬约束的条件下, 对可用性约束做排序, 选择属性值最高的候选服务集即为最优组合服务选择.

下面举例说明系统多 QoS 约束的 Web 服务组合选择问题. 服务组合的内部结构如图 3 所示. 以 4 种 QoS 指标为例, 表 2 所示为图 3 中各功能节点可能的候选服务.

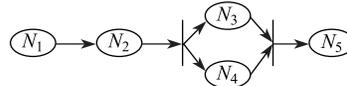


图 3 服务组合内部结构图示例

Fig. 3 Structure of the service composition

表 2 候选服务集 QoS 参数

Tab. 2 QoS parameters of the candidate web service

功能节点	候选服务	执行时间/ms	可用性(正相关)	可靠性(负相关)	代价(Cost)
N ₁	S ₁₁	70	0.6	0.7	1
	S ₁₂	80	0.7	0.7	1
	S ₁₃	60	1	0.9	1.2
N ₂	S ₂₁	50	1	0.9	0.8
	S ₂₂	60	0.4	0.8	0.9
N ₃	S ₃₁	120	0.9	0.9	0.2
	S ₃₂	100	1	0.7	0.5
	S ₄₁	201	0.8	0.8	1
N ₄	S ₄₂	243	0.9	0.7	0.7
	S ₄₃	234	1	0.6	0.4
	S ₅₁	30	1	1	1.5
N ₅	S ₅₂	40	0.7	0.9	0.6
	S ₅₃	60	0.6	0.7	1.2

表 2 包括 5 个功能点共 13 个候选服务, 每个候选服务包含 4 个 QoS 属性: 执行时间、可用性、可靠性和代价. 代价最高限额设置为 3. 该组合服务结构为分支型, 即服务功能节

点 N_3 和 N_4 属于两个分支, 则该组合服务有两条执行流程, 因此组合约束也需要指定不同的流程. 假定用户针对 $N_1 \rightarrow N_2 \rightarrow N_3 \rightarrow N_5$ 流程提出的 QoS 需求为执行时间小于 260 ms, 最高可用性和可靠性, 在此硬约束下, 候选服务集 $\{S_{11}, S_{21}, S_{31}, S_{51}\}$ 为可行性候选服务集, 且满足最高代价限额.

由上述服务选择过程可见, 多 QoS 属性约束的 Web 服务组合选择问题的计算搜索空间较大, 文献 [21] 指出服务计算中服务质量的多目标优化模型具有 NP 复杂性性质, 无法在多项式时间内得到解决, 因此需要研究有效的优化方法.

2.4 系统 MQBR 模型

根据溯源系统的特点, FST 系统的 QoS 指标需要考虑 Web 服务和数据服务的 QoS 指标. 针对多目标 QoS 优化问题, 文献 [21] 研究了服务计算中服务质量的多目标优化模型与求解. 本文提出了在多服务属性约束下优化因子 (MQBR) 模型. MQBR 的目标公式为

$$\max \left(\frac{w_1 SA(C) + w_2 SR(C) + w_3 ST(C) + w_4 DASR(C) + w_5 DT(C)}{SC(C) + DSHT(C)} \right), \quad (1)$$

且服务响应时间满足 $SRT(C) \leq \Gamma$, Γ 为服务响应时间约束阈值, 其中, w_i 为服务指标的权重, $\sum_i w_i = 1$.

公式 (1) 中的各权重取值相等, 即 $w_i = 0.2, i = 1, \dots, 5$. 公式 (1) 中 SA 为服务的可用性评价价值, 即用户对该服务的引用率; SR 是服务的可靠性评价价值, 即用户对该服务的反馈故障率; ST 是服务的吞吐量; DASR 是服务的数据成功访问率; DT 是服务的数据吞吐量; DSHT 是数据自愈时间.

3 服务选择算法

3.1 算法设计

FST 系统的多 QoS 属性约束的 Web 服务组合服务选择, 本质上是一个服务遍历过程, 可以采用深度优先, 或者宽度优先的盲目搜索算法, 但效率较低^[22-23]. FST 系统采用 QoS 约束驱动的启发式算法. 系统在满足用户多 QoS 属性约束的同时, 保证所选组合服务集是最优解或近似最优解.

为了提升计算的效率, 首先对服务进行采用 Skyline 算法预处理, 即去除掉不可能出现在候选结果中的 Web 服务. 文献 [26] 最早研究了基于 Skyline 的服务选择问题.

Skyline 算法是一个典型的多目标优化的问题^[24-25], 又称为 Pareto 最优问题, 是指从给定的 N -维空间的对象集合 O 中选择 1 个子集, 该子集中的点都不能被 O 中的任意一个其他点所控制 (Dominating).

给定 2 个服务 S_1 和 S_2 , 如果 S_2 优于 S_1 , 记为 $S_1 \prec S_2$, Skyline 服务满足条件当且仅当 $\forall k \in [1, |C_{QoS}|]: C_{QoS}^k(S_1) \leq C_{QoS}^k(S_2), \exists i \in [1, |C_{QoS}|]: C_{QoS}^i(S_1) < C_{QoS}^i(S_2)$. FST 系统中 Skyline 计算候选服务的成对 QoS 参数向量, 对服务进行初步的预处理.

在 FST 系统中, 设置了 8 个 QoS 属性, 如执行时间、可用性、代价、吞吐量等等. 服务组合的响应时间是可加性约束, 也是硬约束; 可用性是正相关的可加性约束, 可靠性是负相关的可乘性约束. 算法的目的是在硬约束组合响应服务时间 $SRT(C) \leq \Gamma$ 的约束下, 最小化服务组合代价 SC 和最小化数据自愈时间 DSHT, 即选择满足公式 (1) 的服务候选组合, 具体见算法 1.

FST 系统中可用性和可靠性评价都是正相关, 且取值范围均为 $[0-1]$ 区间, 执行时间是硬约束. 根据这个领域特点, 单个服务评价的启发性函数定义为

$$H(n) = H(n)_H + H(n)_S, \quad (2)$$

其中, $H(n)_H$ 为硬约束, 由执行时间构成, $H(n)_S$ 为软约束. $H(n)_S$ 的值反映了服务对节点 n 的适应度, $H(n)_S$ 值越大, 该服务越符合用户选择.

算法1 FST系统服务组合算法

输入: 约束条件 C_{QoS} , 服务集合 S , 节点集合 N

输出: 服务组合列表 $A(N)$

1. Compute service Pair-wise skyline
 2. Select the first node $N \in N$, sort service of N by $H(N)$;
 3. $N' = N - N$, $j = 1$, $A(N) = \emptyset$;
 4. $h_j(N) = \max H(N)$, $T = t(N_0, j)$.
 5. if $T > \Gamma$, select the service secondly by $H(N)$, go to 4;
 6. $T = t(N_0, j) + T$;
 7. if $N' \neq \emptyset$, select the next $N_i \in N$
 8. $j = 1$
 9. $h_j(N_i) = \max H(N_i)$, $T = t(N_i, j)$;
 10. if $T < \Gamma$, store the N_i^j to $A(i) // N_i^j$ 为节点 N_i 的第 j 个服务
 11. else
select the service secondly by $H(N_i)$, $T = t(N_i, j + 1) + T$, go to 6;
if service of N_i search complete, search back N_{i-1} .
 12. if $\forall a_i \in A(N)$, $a_i \neq \emptyset$, return $A(N)$.
 13. else return null.
-

算法总体分为3大步骤, 具体如下.

(1) 计算每一类服务的Skyline, 形成候选服务的成对QoS参数向量(行1).

(2) 根据用户的QoS属性, 即执行时间、可用性、可靠性和代价, 计算每个节点中每个服务的启发性函数 H 值, 并按照公式(2)计算 H 值的大小, 按照由大到小的顺序排列(行2).

(3) 根据启发式搜索算法进行搜索, 该过程采用数组 $A(N)$ 保存每个节点选择的服务, 如果该最优服务组合满足用户的执行时间约束, 则返回该服务组合, 否则将根据回溯算法, 把不满足时间要求的服务节点去掉, 选择次优服务, 直到满足执行时间的硬约束(行2~11). 如果 $A(N)$ 中任意元素不空, 则返回 $A(N)$, 否则返回 null, 即不存在满足约束的服务组合(行12~13).

由于在搜索过程中, 还需要满足执行时间的硬约束, 这样就存在着高可用性和可靠性的服务组合, 不满足执行时间的约束可能, 因此, 需要在算法中, 加入时间约束, 如果超过用户设置的时间限制, 算法需要回溯重新搜索.

3.2 实验

实验算法采用JDK1.8实现, 数据集采用QWS^①数据集, 该数据集经过预处理, 共2382个服务, 每个服务包括8个属性. 运行环境配置为Win7、CPU为i5-3210M、JVM内存512MB.

我们测试两种实验结果: ①不同的服务节点, 算法运行的效率; ②加入了Skyline预处理的启发式服务选择与直接使用启发式算法的效率.

对于不同的服务节点算法运行的实验结果如图4所示. 服务组合的节点分别设定为4~7, 横纵坐标表示响应时间阈值, 响应时间阈值 Γ 作为硬约束, 分别设定为1200、1100、1050、1000和950, 纵轴为算法运行时间, 单位为ms. 在蛮力式计算方式下, 服务组合的节点个数为7时, 服务的计算时间约为1220ms.

从实验结果可以得出, 服务选择的效率受到两个主要方面的影响: ①随着服务节点的增多, 算法在服务选择时运行的时间开销整体趋势增加. ②随着响应时间阈值 Γ 的降低, 算法运行时间整体趋势增加, 原因是算法随着硬约束阈值的减少, 满足条件的服务数量在减少, 回溯的深度和次数随硬约束阈值 Γ 的减少而增加. 与蛮力式搜索算法相比, 本文的算法可以压缩组合服务

^①<http://www.uoguelph.ca/~qmahmoud/qws/>

搜索的空间,极大地提高了服务选择的效率.

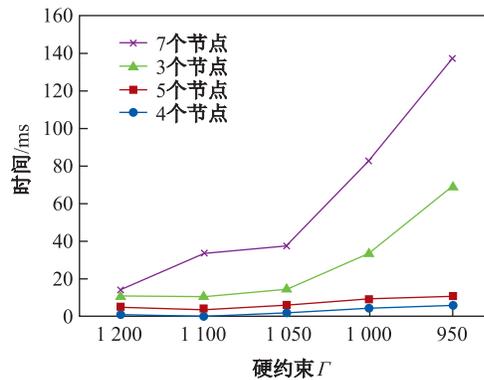


图4 服务选择算法效率

Fig. 4 Efficiency of algorithm for service selection

加入了Skyline预处理的启发式服务选择运行实验结果如图5所示.图5中的服务组合的节点个数为5,响应时间阈值 Γ 设定为1100.横轴表示服务类别,分别设置为10~50. Heuristic_NoPruning为仅使用启发式方法服务组合性能, Skyline_Heuristic为加入Skyline预处理之后的启发式方法服务组合性能.

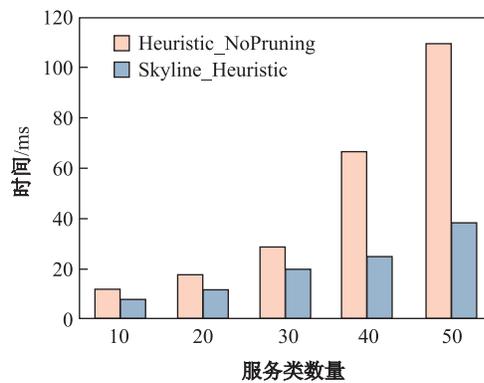


图5 服务选择算法效率

Fig. 5 Efficiency of algorithm for service selection

图5的实验结果表明,加入了基于Skyline预处理的启发式方法,服务选择效率有较大提高,尤其是随着服务类别的增加,性能提升的优势更加明显.

4 结 论

本文研究了基于Web服务的FST系统的服务组合优化问题,研究了FST系统Web服务QoS的重要指标,分析了系统中数据服务的Web服务指标,并在此基础上建立了多QoS约束的组合服务模型,采用Skyline预处理和启发式方法求解了在软硬约束下服务优化组合问题.未来的工作主要两个方面:①研究不同服务节点带有的权重值的情况,即不同的功能节点对组合服务的贡献有差别,更能体现服务组合的选择复杂性;②引入用户反馈机制,针对系统的服务有效性进行量化的研究.

[参 考 文 献]

- [1] 岳昆, 王晓玲, 周傲英. Web 服务核心支撑技术: 研究综述 [J]. 软件学报, 2004, 15(3): 428-442.
- [2] SHENG Q Z, QIAO X, VASILAKOS A V, et al. Web services composition: A decade's overview [J]. Information Sciences, 2014, 280: 218-238.
- [3] LIU R, XU X, WANG Z, et al. Probability matrix of request-solution mapping for efficient service selection [C]// 2017 IEEE International Conference on Web Services (ICWS). IEEE, 2017: 444-451.
- [4] BOUGUETTAYA A, SINGH M, HUHNS M, et al. A service computing manifesto: The next 10 years [J]. Communications of the ACM, 2017, 60(4): 64-72.
- [5] PAPAIOGLOU M P. Service-oriented computing: Concepts, characteristics and directions [C]//International Conference on Web Information Systems Engineering. IEEE Computer Society, 2003: 3-12.
- [6] HUANG K C, SHEN B J, LEE T J, et al. Resource allocation and dynamic provisioning for service-oriented applications in cloud environment [C]//International Conference on Cloud Computing Technology and Science. IEEE Computer Society, 2012: 839-844.
- [7] 杨胜文, 史美林. 一种支持 QoS 约束的 Web 服务发现模型 [J]. 计算机学报, 2005, 28(4): 589-594.
- [8] DEORA V, SHAO J, GRAY W. Modelling quality of service in service oriented computing [C]//International Symposium on Service-Oriented System Engineering. IEEE Computer Society, 2006: 95-101.
- [9] 李研, 周明辉, 李瑞超, 等. 一种考虑 QoS 数据可信性的服务选择方法 [J]. 软件学报, 2008, 19(10): 2620-2627.
- [10] 李小勇, 桂小林, 毛倩, 等. 基于行为监控的自适应动态信任度测模型 [J]. 计算机学报, 2009, 32(4): 664-674.
- [11] AL-MASRI E, MAHMOUD Q H. Discovering the best web service [C]//Proceedings of the 16th International Conference on World Wide Web. ACM, 2007: 1257-1258.
- [12] DE BARNAGHI S P, BAUER M, et al. Service modelling for the internet of things [C]//Computer Science and Information Systems. IEEE, 2011: 949-955.
- [13] YU D, LI C, YIN Y. Optimizing web service composition for data-intensive applications [J]. International Journal of Database Theory and Application, 2014, 7(2): 1-12.
- [14] 任守纲, 徐焕良, 黎安, 等. 基于 RFID/GIS 物联网的肉品跟踪及追溯系统设计与实现 [J]. 农业工程学报, 2010, 26(10): 229-235.
- [15] CHEN M, MAO S, LIU Y. Big data: A survey [J]. Mobile Networks and Applications, 2014, 19(2): 171-209.
- [16] TANAKA M, MURAKAMI Y, ZETTSU K. Data-intensive services for large-scale archive access [C]//Services Computing (SCC), 2012 IEEE 9th International Conference on. IEEE, 2012: 617-624.
- [17] SUCHITHRA M, RAMAKRISHNAN M. A Survey on web service selection techniques [J]. Wireless Communication, 2015, 7: 235-238.
- [18] WANG S, ZHU X, YANG F. Efficient QoS management for QoS-aware web service composition [J]. International Journal of Web and Grid Services, 2014, 10(1): 1-23.
- [19] WANG P W, DING Z J, JIANG C J, et al. Constraint-aware approach to web service composition [J]. Systems, Man, and Cybernetics: Systems, IEEE Transactions on, 2014, 44(6): 770-784.
- [20] LEMOS A L, DANIEL F, BENATALLAH B. Web service composition: A survey of techniques and tools [J]. ACM Computing Surveys (CSUR), 2016, 48(3): 33-74.
- [21] 林闯, 陈莹, 黄霖威, 等. 服务计算中服务质量的多目标优化模型与求解研究 [J]. 计算机学报, 2015, 38(10): 1907-1923.
- [22] PAUTASSO C. RESTful web services: Principles, patterns, emerging technologies [C]// Web Services Foundations. New York: Springer, 2014: 31-51.
- [23] FALOU M E L, BOUZID M, MOUADDIB A I, et al. A distributed planning approach for web services composition [C]//IEEE International Conference on Web Services (ICWS). IEEE, 2010: 337-344.
- [24] BORZSONY S, KOSSMANN D, STOCKER K. The skyline operator [C]//Proceedings of 17th International Conference on Data Engineering. IEEE, 2001: 421-430.
- [25] 魏小娟, 杨婧, 李翠平, 等. Skyline 查询处理 [J]. 软件学报, 2008, 19(6): 1386-1400.
- [26] ALRIFAI M, SKOUTAS D, RISSE T. Selecting skyline services for QoS-based web service composition [C]//Proceedings of the 19th International Conference on World Wide Web. ACM, 2010: 11-20.

(责任编辑: 李 艺)