

DOI: 10.11830/ISSN.1000-5013.202010034



激励机制下负载均衡和 QoS 感知服务组合方法

刘英, 焦竽鑫, 吴小竹

(福州大学 数字中国研究院(福建), 福建 福州 350108)

摘要: 提出一种基于激励机制的负载均衡和服务质量感知服务组合(LBQSC)方法. 首先, 构建一个全局约束分解模型, 并采用文化遗传算法求解; 其次, 考虑服务质量(QoS)和负载构造激励合同, 提出一种基于激励机制的服务选择算法, 通过不断激励 QoS 的动态调整获取最优服务; 最后, 在 QWS 2.0 综合数据集上进行对比实验. 实验结果表明: 基于激励机制的负载均衡和 QoS 感知服务组合方法能在保证负载均衡的情况下有效地获取高质量的组合服务.

关键词: 全局约束分解; 服务组合; 激励合同; 激励机制; 负载均衡

中图分类号: TP 393 **文献标志码:** A **文章编号:** 1000-5013(2021)05-0684-09

Load Balance and QoS-Aware Service Composition Method Under Incentive Mechanism

LIU Ying, JIAO Yuxin, WU Xiaozhu

(Academy of Digital China (Fujian), Fuzhou University, Fuzhou 350108, China)

Abstract: An incentive mechanism based load balance and aware service with service quality composition (LBQSC) method is proposed. Firstly, a global constraint decomposition model is constructed and solved by cultural genetic algorithm. Secondly, considering the quality of service (QoS) and load structure incentive contract, a service selection algorithm based on incentive mechanism is proposed to obtain the optimal service through constantly dynamic adjustment of incentive QoS. Finally, a set of comparative experiments on the QWS 2.0 dataset are carried out. The experimental results show that the LBQSC method can effectively obtain high-quality composite services under the condition of ensuring load balance.

Keywords: global constraint decomposition; service composition; incentive contract; incentive mechanism; load balance

近年来,随着 Web 服务技术的快速发展,越来越多的 Web 服务被发布到网络端实现共享,使具有相同功能但不同服务质量(QoS)的 Web 服务数量剧增^[1]. 单个 Web 服务由于功能简单很难满足用户的复杂需求, QoS 感知服务组合(QSC)研究应运而生^[2]. 欧阳超等^[3]结合模拟退火算法和遗传算法提出一种新的服务组合方法,加速解的收敛. 马力等^[4]针对传统缺乏语义信息支持的 Web 服务选择问题,量化 QoS 属性并将该问题转化为图搜索问题,利用 AO* 算法求解. Hwang 等^[5]将 QoS 作为具有概率质量函数的离散随机变量进行服务选择. Yuan 等^[6]提出基于自适应全局 QoS 约束分解的动态服务选择方法,采用模糊逻辑和文化遗传算法获取局部约束组合,缩小服务的搜索范围. Alayed 等^[7]提出改进蚁

收稿日期: 2020-10-23

通信作者: 吴小竹(1979-),男,讲师,博士,主要从事空间数据挖掘和分布式计算的研究. E-mail: wxz@fzu.edu.cn.

基金项目: 福建省重点科技资助项目(2015H0015)

群优化算法(EACO),该方法能有效避免局部最优,减少搜索时间.传统服务组合方法容易忽略参与者的竞争关系,部分学者引入博弈论思想.马同伟等^[8]针对分配中买卖双方的利益问题,提出双向拍卖资源分配模型,保证双方在公平公正的环境中竞争和购买.Wang 等^[9-10]认为在服务选择过程中,参与者具有自私性且不愿完全暴露自己的信息,针对信息不完整的决策问题,考虑 Web 服务存在纳什均衡竞争关系,提出一种激励机制选择最优 Web 服务;并且基于激励合同提出一种新的激励机制,在该激励机制中,请求者向提供者提供激励合同,根据提供者对激励合同的反应获取私有信息,不断迭代直到为每个任务均选择出最优服务.激励机制是将博弈论应用到服务组合中的一次大胆且成功的创新,有效地解决了服务选择中寻找最优解的问题,并且对参与者都有着正向、积极的作用.

在服务组合过程中,请求者的主要目的是在满足其功能需求下获取具有良好质量保障的组合服务,提供者则需要优化 QoS 和保证负载均衡^[11].传统服务选择策略倾向于选择 QoS 更优的 Web 服务,使该 Web 服务的负载增加.服务的负载能力往往是有限的^[12],若达到最大负载量后,其 QoS 可能无法保证,并且负载不均会导致服务出现过载或空闲的状态,这将影响系统的整体性能.李文中等^[13]提出一种自适应分布式负载均衡(LCB)算法,使用负载容率测度衡量各服务副本的负载状况.杨石等^[14]针对任务调度中的时间和通信成本问题,提出基于蜜蜂觅食行为的负载均衡(HBB-LB)算法,有效平衡虚拟机的负载和响应时间.任金霞等^[15]为缩短任务完成时间和提高负载均衡,提出一种具有 QoS 约束的模拟退火的任务调度算法,使参与者的效用均能获得满足.Pushpavati 等^[16]提出一种改进最大完成时间和迁移性能的配对树算法,提高了响应时间、成本、活力和吞吐量等综合服务质量指标.Arabinda 等^[17]提出一种基于改进粒子群算法的负载均衡技术(LBMP SO),能够缩短虚拟机完工时间,提高资源的利用率.施凌鹏等^[18]提出微服务链感知的请求负载均衡算法,减少请求响应的延迟时间,平衡主机间的负载.Muthsamy 等^[19]考虑完工时间、响应时间、执行时间和任务优先级等多种 QoS 属性,提出人工蜜蜂觅食优化任务调度(TSABF)算法,以获得虚拟机上的最优任务调度和负载均衡.Li 等^[20]考虑用户偏好、服务水平协议和成本等因素,提出边缘云资源模型,能在满足负载均衡要求的同时使成本降到最低.Mohanty 等^[21]认为负载均衡的基本目标是最小化时间和提高系统性能,采用 Jaya 算法实现了云环境中的负载均衡.Asghari 等^[22]同时考虑负载均衡和 QoS 感知问题,提出反向蚁群优化(IACO)算法,以解决云服务器的负载均衡效率低下问题.

考虑 QoS 的负载均衡策略不仅能够合理分配资源的工作负载,实现资源的有效利用,还能够获取到满足请求的优质方案.因此,QoS 感知负载均衡的研究具有必要性.基于此,本文提出一种基于激励机制的负载均衡和 QoS 感知服务组合(LBQSC)方法,通过动态调整 QoS 改变 Web 服务的选择概率,并权衡请求和提供者的双方需求,以解决服务组合的负载均衡 QoS 感知问题.

1 负载均衡和 QoS 感知服务组合问题

1.1 相关定义

定义 1 Web 服务(s)是一个独立的、功能完整的可以通过网络进行发布、定位及访问的最小资源单位,包含功能属性和非功能属性,可用 6 元组 $s = \{id, b, f, Q_{os}, l_w, l_{mw}\}$ 表示,其中, id 为 s 的唯一标识; b 为 s 的基本信息,包括提供者、名称和发布时间等; f 为 s 的功能属性,是服务被发现的首要依据; Q_{os} 为 s 的非功能属性,是服务被选择的重要评判标准; l_w 为 s 的实时负载量,是 Web 服务当前的服务对象个数; l_{mw} 为 s 的最大负载量,是被服务对象个数的最大值, $l_w \leq l_{mw}$,一旦实时工作量达到最大负载量,服务立刻停止为其他对象提供服务.

定义 2 服务质量指 Web 服务的非功能属性,包括一组 QoS 属性 $Q_{os} = \{q_1, \dots, q_v\}$. 一个 QoS 属性 q_v 可用 3 元组 $q_v = \{n_v, d_v, r_v\}$ 表示,其中, n_v 为 q_v 的名称,如响应时间、声誉、价格、可行性等, d_v 为 q_v 的初始状态默认值, r_v 为 d_v 的可变性,存在两种情况: 1) 当 $r_v = 0$ 时, d_v 不发生改变; 2) 当 $r_v \neq 0$ 且 $r_v \in (0, 1)$ 时, d_v 可发生改变,变化区间为 $[d_v \times (1 - r_v), d_v \times (1 + r_v)]$.

在服务组合中,不同 QoS 属性的数量级往往不同,因此,需要进行归一化处理.施凌鹏等^[18]将非功能属性分成积极属性和消极属性,通过简单加权(SAW)技术^[19]进行归一化.对于积极属性(声誉、可行

http://www.hdxh.hqu.edu.cn

性等),属性值越大,表示服务质量越好,按照式(1)归一化,即

$$V_{i,j}=\begin{cases} \frac{q_{i,j}-q_i^{\min}}{q_i^{\max}-q_i^{\min}}, & q_i^{\max}-q_i^{\min}\neq 0, \\ 1, & q_i^{\max}-q_i^{\min}=0. \end{cases} \tag{1}$$

对于消极属性(响应时间、价格等),属性值越小,表示服务质量越好,按照式(2)归一化,即

$$V_{i,j}=\begin{cases} \frac{q_i^{\max}-q_{i,j}}{q_i^{\max}-q_i^{\min}}, & q_i^{\max}-q_i^{\min}\neq 0, \\ 1, & q_i^{\max}-q_i^{\min}=0. \end{cases} \tag{2}$$

式(1),(2)中: q_i^{\min} 表示第*i*个抽象服务中属性*q*的最小值; q_i^{\max} 表示第*i*个抽象服务中属性*q*的最大值; $q_{i,j}$ 表示第*i*个抽象服务中第*j*个 Web 服务属性*q*的值; $V_{i,j}$ 表示第*i*个抽象服务中第*j*个 Web 服务属性*q*的归一化值.

定义 3 抽象服务(AS)代表具有相同功能、不同 QoS 的一组 Web 服务,这组服务被称为该抽象服务的候选(具体)服务集.

定义 4 工作流(*F*)可用 2 元组 $F=\{AS_set, str\}$ 表示,其中, $AS_set=\{AS_1,\cdots,AS_i,\cdots,AS_m\}$ 为一组抽象服务, AS_i 为第*i*个抽象服务且 $AS_i=\{s_{i,1},s_{i,2},\cdots,s_{i,j},\cdots,s_{i,n}\}$;str 为抽象服务的组织结构.

工作流管理联盟(WFMC)提出顺序、选择、循环和并行 4 种结构,用以支持 Web 服务组合建模^[1]. 其中,选择、循环和并行结构均可通过文献[20]中的技术转化成顺序结构.

定义 5 组合服务(CS)可用 4 元组 $CS=\{AS_set,Q_{os},w,u\}$ 表示,其中, w 表示非功能属性的权重, $w=\{w_1,w_2,\cdots,w_g\}$,满足 $\sum_{p=1}^g w_p=1$; u 为组合服务效用,计算公式为

$$u=\sum_{i=1}^m\sum_{p=1}^v(w_p\times V_i^p). \tag{3}$$

式(3)中: V_i^p 为组合服务在第*i*个抽象服务中选择服务的第*p*个 QoS 属性的归一化值; v 为 QoS 属性的个数; m 为抽象服务的个数.

组合服务的 QoS 属性值是由构成该服务的各个原子服务的 QoS 属性值聚合而成. 不同 QoS 属性在不同工作流中的聚合公式不同,不同 QoS 属性在同一工作流中的聚合公式也不同^[21]. Web 服务的声誉(*R*)、响应时间(*T*)和价格(*P*)3 个 QoS 属性在顺序工作流下的聚合公式分别为

$$R=\min\{R_i\}_{i=1}^m, \quad T=\sum_{i=1}^m T_i, \quad P=\sum_{i=1}^m p_i.$$

定义 6 服务负载(load)指服务资源利用率,用 Web 服务在某一时刻工作量与最大工作量的比值表示,即 $load=l_w/l_{mw}$.

定义 7 用户请求(*r*)用来存放请求者的需求信息,可用 2 元组 $r=\{id,C\}$ 表示,其中,id 表示 *r* 的唯一标识;*C* 表示全局 QoS 约束,是请求者对组合服务 QoS 的基本要求.

定义 8 全局 QoS 约束 $C=\{c_1,c_2,\cdots,c_g\}$,其个数不得超过 Web 服务 QoS 个数,即 $g\leq v$. 全局 QoS 约束是对组合服务 QoS 的基本约束,它要求组合服务的 QoS 应该优于用户请求中对应的全局 QoS.

定义 9 解决方案(SOL)用一个 4 元组 $SOL=\{CS_set,\delta,\overline{U},t\}$ 表示,其中, CS_set 为组合服务集, $CS_set=\{CS_k\}_{k=1}^{num}$,num 为组合服务的个数,其大小不得超过用户请求规模; \overline{U} 表示组合服务集的平均效用, $\overline{U}=(\sum_{k=1}^{num} CS_k.u)/num$, $num\leq l$; t 表示算法的时间开销; δ 为总体负载标准差, $\delta=$

$$\sqrt{\frac{\sum_{i=1}^m\sum_{j=1}^n(s_{i,j}.load-\overline{load})^2}{m\times n}}, \text{ 其中, } \overline{load} \text{ 为总体负载均值, } \overline{load}=\frac{\sum_{i=1}^m\sum_{j=1}^n s_{i,j}.load}{m\times n}, s_{i,j} \text{ 为第 } i \text{ 个抽象服务中的第 } j \text{ 个 Web 服务, } n \text{ 为每个抽象服务中的具体服务个数. 从服务资源的利用角度看,服务的负载标准差越小,表示服务资源的利用越均衡,即服务总体负载越均衡;反之,服务总体负载越失衡.}$$

1.2 问题描述

基于上述定义,提出负载均衡和 QoS 感知服务组合问题. 假设一个工作流 $F=\{AS_set, str\}$, 其中,

抽象服务集 $AS_set = \{AS_1, \dots, AS_i, \dots, AS_m\}$. 给定一组请求 $r = \{r_1, \dots, r_k, \dots, r_l\}$, 若能找到一个解决方案 $SOL = \{CS_set, \delta, \bar{U}, t\}$, 使得

$$\min \delta, \tag{4}$$

$$\max \bar{U}, \tag{5}$$

$$\min t, \tag{6}$$

$$\{\text{aggregate}(CS_k \cdot q_v)\}_{k=1}^{\text{num}} \geq \{r_k \cdot q_v\}_{k=1}^{\text{num}}, \tag{7}$$

则 SOL 为最优解决方案.

式(4)~(6)是最优解决方案的评判标准, 其中, 式(4)表示在获得 SOL 后, Web 服务的总体负载标准差最小; 式(5)表示 SOL 具有最大的平均效用; 式(6)表示获取 SOL 的执行时间最少. 式(7)是 SOL 作为最优解决方案存在的基本条件, 它要求每个组合服务 CS_k 中 QoS 属性 q_v 整合后均优于请求 r_k 中对应的全局 QoS 约束 q_v .

2 LBQSC 方法设计

2.1 LBQSC 的总体框架

基于 LBQSC 方法求解负载均衡和 QoS 感知服务组合问题主要分为全局约束分解(GCD)模型求解和局部服务选择 2 个步骤. LBQSC 的总体框架, 如图 1 所示.

GCD 模型求解被认为是组合优化问题, 文化遗传算法(CGA)在处理此类问题时具有快速收敛和防止早熟等优点^[9, 23-24], 因此, 文中采用 CGA 算法. 在局部服务过程中, Web 服务选择激励机制^[5], 通过合同激励服务提供者更改自身服务 QoS 属性, 再迭代选择 QoS 更高的 Web 服务, 从而提高服务请求者的效用. 但是这种机制并没有考虑服务的负载情况, 因此, 对合同进行改进, 使其不仅能够约束服务的 QoS 和负载, 还能提供更高报酬来激励服务提供者将本身服务的 QoS 根据当前负载整体水平进行动态调整, 在此基础上, 提出一种基于激励机制的服务选择方法. 最后, 组合局部选择的服务形成组合服务.

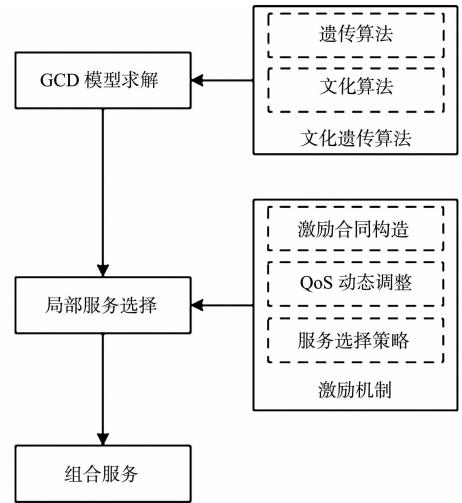


图 1 LBQSC 的总体框架

Fig. 1 Overall framework of LBQSC

2.2 GCD 模型求解

全局 QoS 约束分解过程的主要目的是将服务请求者的全局 QoS 约束合理地分解为一组局部约束, 确保满足局部约束的 Web 服务复合而成的组合服务一定满足服务请求者的全局 QoS 约束. 保留更多的组合方案考虑 2 个原则: 1) 确保满足每个局部约束的候选服务个数尽量多; 2) 确保满足每个局部约束的候选服务个数尽量均衡^[25]. 其评价公式为

$$E = \omega_{\text{sum}} \times S + \omega_{\text{mul}} \times M. \tag{8}$$

式(8)中: E 为模型评价; S 为满足局部约束的候选服务个数之和, $S = \sum_{i=1}^m n_{cs,i}$; M 为满足局部约束的候选服务个数之积, $M = \prod_{i=1}^I n_{cs,i}$, $n_{cs,i}$ 为满足第 i 个局部约束的候选服务个数; ω_{sum} , ω_{mul} 为 2 个原则各自的权重值且两者之和为 1, 取 $\omega_{\text{sum}} = 0.5$, $\omega_{\text{mul}} = 0.5$.

GCD 模型求解过程包括初始化质量等级和寻找最优局部约束组合 2 个步骤.

初始化质量等级将每个 QoS 属性划分成一组离散质量值, 给每个抽象服务 AS_i 初始化质量等级^[23-24]. 质量等级分解, 如图 2 所示. 图 2 中: $q_{i,m}^j$ 为抽象服务 AS_i 中第 j 个 QoS 属性的第 m 个值; $L_{i,d}^j$ 为抽象服务 AS_i 中第 j 个 QoS 属性的第 d 个质量等级, 并且 $\min(q_i^j) \leq L_{i,1}^j \leq L_{i,2}^j \leq \dots \leq L_{i,d}^j \leq \max(q_i^j)$.

寻找最优局部约束组合是将质量等级进行组合, 从而使 E 近似最大化的过程. 采用 CGA 求解该过程, 通过文化算法从进化种群中抽象待解决问题的知识, 反馈这些知识以指导算法的搜索过程, 从而保证解的收敛. CGA 的实现框架, 如图 3 所示. 图 3 中: 种群空间由个体组成; Update() 函数用于更新信仰空间的知识; Accept() 函数用于从种群空间选择一定数量的个体进入信仰空间以形成知识; Influence()

函数主要通过信仰空间中的知识影响社会群体空间的进化方向;Generate()函数用于构造个体;Evaluate()函数用于评价个体的适应度;Select()函数用于遗传操作中选择交叉的个体.

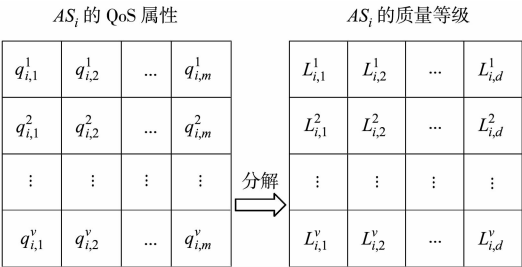


图 2 质量等级分解

Fig. 2 Quality level decomposition

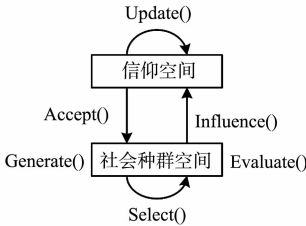


图 3 CGA 的实现框架

Fig. 3 CGA implementation framework of CGA

2.3 基于激励机制的服务选择

激励机制来源于激励理论,主要包括激励主体、激励客体和激励手段 3 部分.在激励机制中,激励主体通过各种手段引导、激发激励客体的潜能,从而实现激励主体的战略目标.在服务选择过程中,请求者、提供者和选择策略分别对应激励主体、激励客体和激励手段.于是,Wang 等^[9]提出一种 Web 服务选择激励机制,将激励机制的优势充分应用于服务选择.该机制针对不同 QoS 提供具有不同报酬的合同来调动服务提供者的积极性,能够帮助请求者快速获取质量更优的服务,并保证服务本身的基本利益.但是,该机制并没有考虑服务的负载能力.因此,对文献^[9]的方法进行改进,使之能够解决负载均衡和 QoS 感知服务组合问题.通过节 2.2 获取近似最优局部约束组合,排除不满足约束要求的候选服务,缩小服务选择的搜索范围.此时,提供一种新的合同,通过给不同的 QoS 和服务负载支付不同的报酬,激励服务提供者动态调整其自身服务的 QoS 来积极竞争合同.同时,为了选择最适合的服务,设计一个基于合同的服务选择策略.在此基础上,构建基于激励机制的服务选择算法,具体描述如算法 1.激励合同构造、QoS 动态调整和 Web 服务选择是其中的 3 个关键步骤.

算法 1 基于激励机制的服务选择算法

输入: csSet //满足局部约束的候选服务集合

输出: bos //最优可选服务

1. //M 为最大迭代次数,Size 为候选服

各个数

2. IF $j \leq M$ OR $Size! = 0$

3. //构造合同

4. $con_j = offerContract(s_set, canSet, j)$

5. //满足当前合同的候选服务集中存放

到 osSet_j

6. $osSet_j = Satisfy(con_j)$

7. //添加到 osSet 中

8. $osSet \leftarrow osSet_j$

9. //不满足的动态调整其 QoS 后存放

到 csSet 中

10. $csSet = AdjustQoSValue(Unsatisfy$

$(con_j))$

11. ENDIF

12. //选择服务

13. $bos = ServiceSelect(osSet)$

14. Return bos//最优可选服务

2.3.1 激励合同构造 合同(con)是请求者对负载和服务质量这 2 个影响因子提出的要求,并确保若服务提供者可以提供满足合同约束的 Web 服务,服务请求者会支付合同中所给定的报酬.一个合同用 4 元组 $con_{\tau} = (\tau, QoS_{\tau}, load_{\tau}, reward_{\tau})$ 表示,其中, τ 为合同的唯一标识; QoS_{τ} 为合同中除价格之外的 QoS 属性约束; $load_{\tau}$ 为合同中的服务负载约束; $reward_{\tau}$ 为针对当前的负载和服务质量,服务请求者支付的费用,也是服务价格的上限.

在合同中,2 个影响因子的数值不同,提供的报酬也随之发生变化.就服务请求者而言,在其预算范围内,他们愿意支付更多的报酬去获取 QoS 更优、负载更低的 Web 服务.对于一个抽象服务,以其局部约束为依据,提供一组合同 $Con = \{con_1, \dots, con_{\tau-1}, con_{\tau}\}$,合同 con_{τ} 的负载和服务质量均优于合同 $con_{\tau-1}$ 的负载和服务质量,且前者的报酬高于后者,则这组合同被称为激励合同.服务提供者若想获得更多的报酬,就需要调整服务自身的 QoS 来进行竞争.将满足合同约束的候选服务称作可选服务(os),

<http://www.hdxh.hqu.edu.cn>

将效用值最大的服务称作最优可选 Web 服务(bos).

2.3.2 QoS 动态调整 在不满足当前合同的情况下, 候选服务若还想获得被选择的机会且获得更高的报酬, 就必须改变自身的 QoS 属性去参与新的合同的竞争. 就候选服务来说, 若改变当前的响应时间、声誉等 QoS 属性使之较先前更优, 则其价格也会随之增长. 因此, 给定一个服务提供策略: 1) 如果候选服务的负载小于服务的总体负载均值, 则优化候选服务除价格以外的 QoS 属性, 并提高候选服务的价格; 2) 如果候选服务的负载等于服务的总体负载均值, 则保持候选服务的所有 QoS 属性不变; 3) 如果候选服务的负载大于服务的总体负载均值, 则劣化候选服务的除价格以外的 QoS 属性, 并减少候选服的价格.

2.3.3 服务选择策略 激励机制运行完毕后, 得到多组可选服务集, 提供一个服务选择策略确定最终选择的服务, 即确定可选服务 os 个数最多的可选服务集, 选择该服务集中的最优可选服务 bos.

3 实验分析

通过实验对比 LBQSC, EACO^[7], IACO^[22] 方法, 验证 LBQSC 方法在处理负载均衡和 QoS 感知服务组合问题时的优越性.

3.1 实验数据与环境

实验数据来源于综合服务数据集 QWS 2.0. 该数据集包含真实世界中 2 507 个 Web 服务数据, 每个 Web 服务有 9 个 QoS 属性, 包括响应时间、声誉、价格、吞吐量和可靠性等, 具体信息可参考文献 [21]. 考虑到对比实验中数据条件的分配情况, 仅挑选 QWS 2.0 数据集中的 2 000 个 Web 服务数据, 并重点关注 Web 服务的声誉、响应时间、价格、实时工作量和最大工作量 5 个 QoS 属性, 声誉和响应时间可从数据集中获取, 价格、实时工作量和最大工作量则采用随机算法构造. 其中, 价格取值控制在 50~500; 最大工作量取值控制在 50~100; 实时工作量取值控制在 0 到最大工作量之间; 假设响应时间和价格均是可变的 QoS 属性, 其变化率分别设置为 $r_t=0.2$, $r_p=0.2$. 在计算组合服务的效用时, 声誉、响应时间和负载的权重分别赋值为 0.3, 0.3, 0.4. 请求数据根据服务数据特征随机构造. 实验结果都是运行 10 次之后取平均值.

实验使用的算法开发语言是 Java, 实现工具为 MyEclipse 2016 CI, 运行环境的具体配置为 Windows 7 专业版操作系统, 内存为 4 GB, 处理器为 Intel(R)Core(TM) i5-2500 CPU@3.30 GHz.

3.2 实验结果与分析

采用 LBQSC, EACO 和 IACO 方法在不同的数据条件下分别进行实验, 并对各指标进行评估.

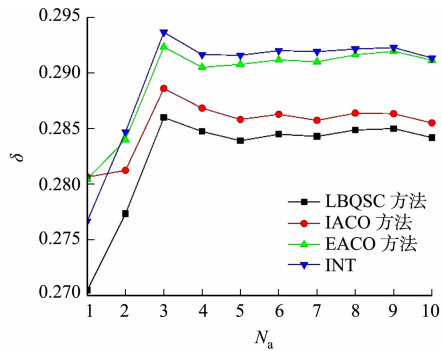
3.2.1 负载标准差 为验证 LBQSC, IACO 和 EACO 在进行大规模服务组合后, 服务的负载标准差较初始状态的优化程度, 进行 3 组对比实验, 具体如下.

1) 当抽象服务数(N_a)范围为 1~10 且步长为 1, 具体服务数(N_c)为 200, 请求数(N_r)为 100 时, 3 种方法的负载标准差(δ)对比, 如图 4(a)所示. 图 4(a)中: INT 为初始状态下服务总体的负载标准差.

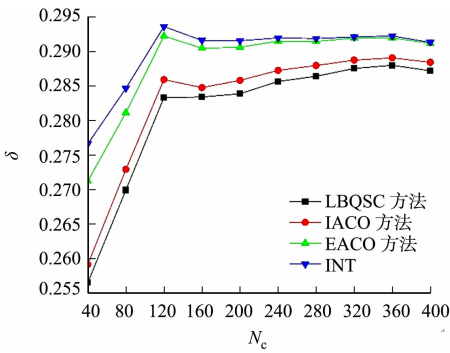
2) 当具体服务数范围为 40~400 且步长为 40, $N_a=5$, $N_r=100$ 时, 3 种方法的负载标准差对比, 如图 4(b)所示.

3) 当请求数范围为 20~200 且步长为 20, $N_a=5$, $N_c=200$ 时, 3 种方法的负载标准差对比情况, 如图 4(c)所示.

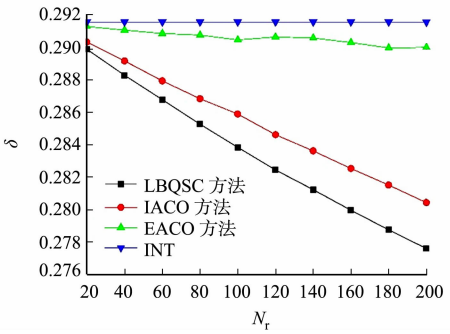
由图 4(a)可知: 在相同具体服务数和请求数情况下, 随着抽象服务数的增加, 3 种方法的负载标准差比初始状态均有所减小, 其中, LBQSC 方法减小得最明显. 由图 4(b)可知: 不论具体服务数如何变化, LBQSC 方法的负载标准差始终最小, 其次为 IACO 方法. 由图 4(c)可知: 服务的数据条件相同, 随着请求数的增加, EACO 方法的负载标准差虽低于 INT, 但基本没有太大改变; 而 LBQSC 和 EACO 方法的负载标准差都逐渐下降且很明显, LBQSC 方法下降尤为迅速.



(a) 不同抽象服务数



(b) 不同具体服务数



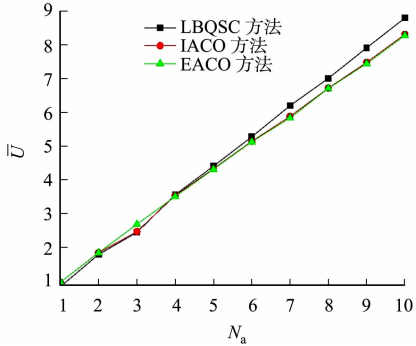
(c) 不同请求数

图 4 不同数据条件下 3 种方法的负载标准差对比

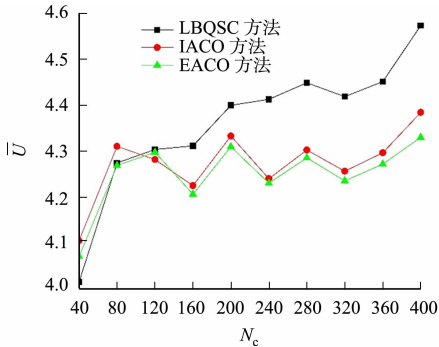
Fig. 4 Comparison of load standard deviation in three methods under different data conditions

3.2.2 平均效用 为验证 LBQSC 方法在处理 QoS 感知服务组合问题时的优越性,在节 3.2.1 的数据条件下进行 3 组对比实验. 不同数据条件下,3 种方法获取的解决方案的平均效用(\bar{U})对比,如图 5 所示.

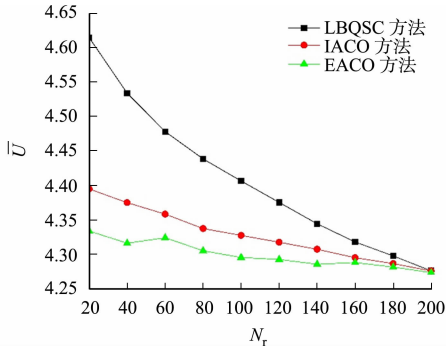
由图 5(a)可知:在具体服务数和请求数相同时,随着抽象服务数的增加,通过 3 种方法获取的解决方案的平均效用也随之增加,且平均效用与抽象服务数大小呈现近似线性关系. 由图 5(b)可知:随着具体服务数的增加,LBQSC 方法的平均效用明显高于其他两种方法,但当 $N_a < 4$ 且 $N_c < 120$ 时, LBQSC 方法的平均效用比其他两种方法差,这是因为蚁群算



(a) 不同抽象服务数



(b) 不同具体服务数



(c) 不同请求数

图 5 不同数据条件下 3 种方法获取的解决方案的平均效用对比

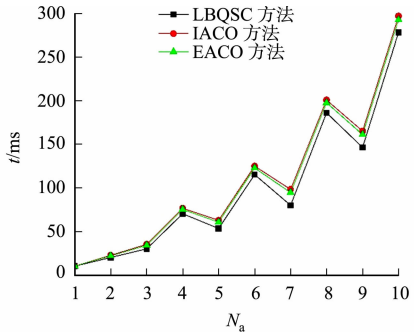
Fig. 5 Comparison of average utility in three methods under different data conditions

法在小规模服务组合下较激励机制具有更好的收敛性. 由图 5(c)可知:在服务数据相同的条件下,请求数越多,3 种方法的平均效用越差,其主要原因是在大量请求下,QoS 较高的 Web 服务被选择达到最大负载量后,后续的请求只能选择一些 QoS 较差的 Web 服务代替,使组合服务效用降低,从而使解决方案的平均效用减小.

3.2.3 执行时间 为进一步验证 LBQSC,IACO,EACO 方法的时间开销,同样采用节 3.2.1 的 3 组数据进行对比实验. 不同数据条件下,3 种方法的执行时间(t)对比,如图 6 所示.

由图 6 可知:不论数据条件如何变化,所提 LBQSC 方法的执行时间均小于其他两种方法,能快速地获取解决方案.

结合 3 个指标的对比结果可知, LBQSC 方法能够有效解决多请求下的负载均衡和 QoS 感知服务组合问题,且具有较少的的时间开销.



(a) 不同抽象服务数

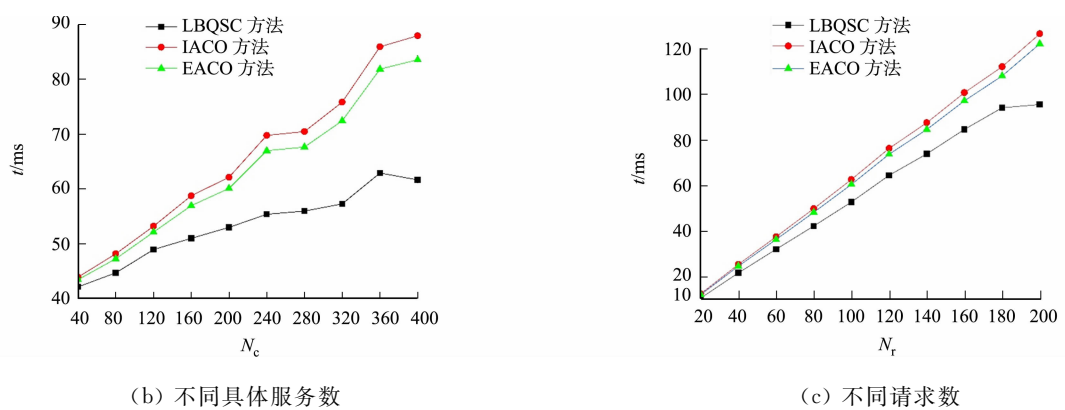


图 6 不同数据条件下 3 种方法的执行时间对比

Fig. 6 Comparison of execution time in three methods under different data conditions

4 结论

针对请求者的高 QoS 需求、提供者的负载均衡需求及 QoS 的动态性等问题,提出基于激励机制的负载均衡和 QoS 感知服务组合(LBQSC)方法. 首先,利用文化遗传算法将全局 QoS 约束分解为一组局部约束;然后,基于激励机制,在满足局部约束且具有动态 QoS 的候选服务中迭代选择出最佳 Web 服务;最后,采用 QWS 2.0 数据集进行实验. 实验结果表明:与 EACO 和 IACO 方法相比,LBQSC 方法能够有效地处理负载均衡和 QoS 感知服务组合问题.

在未来的工作中,将完善 LBQSC 方法,使其可灵活地运用到工作流的其他 3 种基本结构(并行、选择、循环);改善激励机制中的服务提供策略,使服务 QoS 的动态调整更加依赖于激励合同的变化,以更好地服务于服务选择过程;考虑用户请求的时序特征,研究面向多连续请求的负载均衡和 QoS 感知服务组合方法.

参考文献:

- [1] 谭文安,吴嘉凯. 基于改进花朵授粉算法的 Web 服务组合优化[J]. 计算机工程,2020,46(12):67-72. DOI:10.19678/j. issn. 1000-3428. 0056206.
- [2] 柳正利,李兵,强保华,等. 基于文化遗传算法的 QoS 感知的服务组合[J]. 中南大学学报(自然科学版),2018,49(11):2731-2737. DOI:10.11817/j. issn. 1672-7207. 2018. 11. 013.
- [3] 欧阳超,陈志泊,孙国栋. Web 服务组合 QoS 优化中的改进遗传算法[J]. 计算机工程,2017,43(8):231-235. DOI:10.3969/j. issn. 1000-3428. 2017. 08. 039.
- [4] 马力,邱志洋,陈彦萍,等. 基于 QoS 的语义 Web 服务选择[J]. 计算机科学,2017,44(3):226-230. DOI:10.11896/j. issn. 1002-137X. 2017. 03. 047.
- [5] HWANG S Y, HSU C C, LEE C H. Service selection for web services with probabilistic QoS[J]. IEEE Transactions on Services Computing,2017,8(3):467-480. DOI:10.1109/TSC. 2014. 2338851.
- [6] YUAN Yuan, ZHANG Weishi, ZHANG Xiuguo, et al. Dynamic service selection based on adaptive global QoS constraints decomposition[J]. Symmetry,2019,11(3):403. DOI:10.3390/sym11030403.
- [7] ALAYED H, DAHAN F, ALFAKIH T, et al. Enhancement of ant colony optimization for QoS-aware web service selection[J]. IEEE Access,2019,7:97041-97051. DOI:10.1109/ACCESS. 2019. 2927769.
- [8] 马同伟,解瑞云,廖晓飞. 云计算环境下兼顾买卖双方利益的双向拍卖资源分配算法[J]. 计算机应用研究,2016,33(3):734-740. DOI:10.3969/j. issn. 1001-3695. 2016. 03. 022.
- [9] WANG Puwei, DU Xiaoyong. QoS-aware service selection using an incentive mechanism[J]. IEEE Transactions on Services Computing,2019,12(2):262-275. DOI:10.1109/TSC. 2016. 2602203.
- [10] WANG Puwei, LIU Tao, ZHAN Ying, et al. A bayesian nash equilibrium of QoS-aware web service composition [C]//IEEE International Conference on Web Services. Honolulu:IEEE Press,2017:676-683. DOI:10.1109/ICWS. 2017. 81.
- [11] 王弦,刘建勋,曹步清,等. 面向云环境的一种负载感知的服务选择方法[J]. 小型微型计算机系统,2014,35(9):

- 1994-1998. DOI:10.3969/j.issn.1000-1220.2014.09.011.
- [12] KIL H,CHA R,NAM W. Transaction history-based web service composition for uncertain QoS[J]. International Journal of Web and Grid Services,2016,12(1):42-62. DOI:10.1504/IJWGS.2016.074180.
- [13] 李文中,郭胜,许平,等. 服务组合中一种自适应的负载均衡算法[J]. 软件学报,2006,17(5):1068-1077. DOI:10.1360/jos171068.
- [14] 杨石,王艳玲,王永利. 云计算环境下基于蜜蜂觅食行为的任务负载均衡算法[J]. 计算机应用,2015,35(4):938-943. DOI:10.11772/j.issn.1001-9081.2015.04.0938.
- [15] 任金霞,钟小康,蒋梦倩. QoS 性能约束的云任务调度算法研究[J]. 河南师范大学学报(自然科学版),2018,46(4):113-119. DOI:10.16366/j.cnki.1000-2367.2018.04.018.
- [16] PUSHPAVATI U K S,MELLO D A D. A tree based mechanism for the load balancing of virtual machines in cloud environments[J]. International Journal of Information Technology,2021,13(3):911-920. DOI:10.1007/s41870-020-00544-3.
- [17] ARABINDA P,KISHORO B S. A novel load balancing technique for cloud computing platform based on PSO[J/OL]. Journal of King Saud University:Computer and Information Sciences(2020-10-22)[2020-10-23]. <https://doi.org/10.1016/j.jksuci.2020.10.016>.
- [18] 施凌鹏,朱征,周俊松,等. 面向微服务架构的云系统负载均衡机制[J/OL]. 计算机工程(2020-10-15)[2020-10-23]. <https://doi.org/10.19678/j.issn.1000-3428.0058747>.
- [19] MUTHSAM Y G,CHANDRAN S R. Task scheduling using artificial bee foraging optimization for load balancing in cloud data centers[J]. Computer Applications in Engineering Education,2020,28(4):769-778. DOI:10.1002/cae.22236.
- [20] LI Chunlin,TANG Jianhang,LUO Youlong. Service cost-based resource optimization and load balancing for edge and cloud environment[J]. Knowledge and Information Systems,2020,62:4255-4275. DOI:10.1007/s10115-020-01489-6.
- [21] MOHANTY S,PATRA P K,RAY M,*et al.* An approach for load balancing in cloud computing using JAYA algorithm[J]. International Journal of Information Technology and Web Engineering,2019,14(1):27-41. DOI:10.4018/IJITWE.2019010102.
- [22] ASGHARI S,NAVIMPOUR N J. Cloud service composition using an inverted ant colony optimisation algorithm[J]. International Journal of Bio-Inspired Computation,2019,13(4):257-268. DOI:10.1504/IJBIC.2019.100139.
- [23] 方晨,王晋东,张恒巍,等. 基于全局 QoS 分解的多约束服务选取方法[J]. 系统仿真学报,2018,30(10):3893-3902. DOI:10.16182/j.issn1004731x.joss.201810036.
- [24] LIU Zhizhong,XUE Xiao,SHEN Jiquan,*et al.* Web service dynamic composition based on decomposition of global QoS constraints[J]. International Journal of Advanced Manufacturing Technology,2013,69:2247-2260. DOI:10.1007/s00170-013-5204-6.
- [25] 叶恒舟. 时间约束的 Web 服务组合研究[D]. 南宁:广西大学,2019.

(责任编辑:黄晓楠 英文审校:吴逢铁)