

基于关键路径和响应时间约束的空间 信息服务优化组合算法

朱 庆¹ 杨晓霞¹ 李海峰¹

(1 武汉大学测绘遥感信息工程国家重点实验室,武汉市珞喻路 129 号,430079)

摘 要:空间信息服务的响应时间是网络信息服务和分布式空间信息处理等关心的核心问题之一。针对空间信息服务链的自动生成,基于关键路径法思想,定义了服务链关键路径、关键服务和响应时间余量的概念,建立了确定型和概率型响应时间计算模型,提出了顾及关键路径响应时间的服务优化组合算法,保证服务链的组合结果更好地满足响应时间约束。仿真实验结果证明了本文方法的有效性。

关键词:服务组合;服务链;响应时间约束;关键路径

中图法分类号:P208

空间信息服务是自包含、自描述的模块化空间信息应用,通过网络上的发布、定位和调用,能够实现多源空间信息的访问和集成处理^[1]。在广泛可得的各种服务中,搜索选取并优化组合恰当的服务是空间信息聚焦服务的关键问题。在空间信息服务 QoS 的各种因素中,响应时间常常是最重要而且是决定性的因素:① 响应时间是设计和管理服务链时需要考虑的关键问题和普遍问题^[2,3];② 由于环境的动态性和服务本身的不稳定性等原因,使得响应时间具有难以预测的特点,取值往往不固定,而是在一个范围内波动;③ 由于服务链的控制流中可能包含循环、并发等结构,使得其响应时间的计算较之费用、可靠性、可用性等属性更为复杂。因此,响应时间的准确计算成为空间信息服务 QoS 研究的难点。

文献[3,4]采用连续时间 Markov 链(CT-MC),文献[5]指定了任务执行时间约束,文献[6]证实了服务链的响应时间是由那些执行时间最长的分支决定的,文献[7]根据关键路径计算服务链的响应时间,Eder 等人提出了基于计划评审法(PERT)的时间管理模型^[2,8]。然而,针对空间信息服务响应时间的研究还仅仅局限于引用其他领域的服务质量概念,没有系统深入的研究。从已有资料看,现有方法既没有对空间信息服务链

的响应时间属性的不确定性进行特别考虑,在组合过程中,也没有对服务链中地位不同的服务进行区别对待。

本文将基于 QoS 的服务组合方法与运筹学中的关键路径法相结合,建立确定型响应时间计算模型和概率型响应时间计算模型,通过为关键服务选择具有合适响应时间的空间信息服务,使空间信息服务链更好地满足用户对响应时间的约束。

1 响应时间计算模型

1.1 基本概念

定义 1 响应时间指从提交服务请求到获得服务响应所花的时间。

响应时间 Time 包括处理延迟 P_{delay} 和往返通信延迟 T_{delay} 。对于空间信息服务,处理延迟 P_{delay} 是对空间信息执行指定操作的平均响应时间,包括资源等待时间 W_{time} 和处理时间 S_{time} :

$$\text{Time} = W_{\text{time}} + S_{\text{time}} + T_{\text{delay}} \quad (1)$$

在服务链中,从开始节点开始,按照服务的顺序连续不断地到达终止节点的一条通路称为路径。一个服务链可以包含一条或多条路径。路径长度是指路径上各个服务响应时间之和,不是路

径上服务节点的数目。

定义 2 服务链中响应时间之和最长的路径称为关键路径。

关键路径上的任何一个服务都是关键服务,其中任何一个服务的延迟都会导致整个服务链响应时间的延迟。

1.2 服务链响应时间计算模型

1.2.1 确定型响应时间计算模型

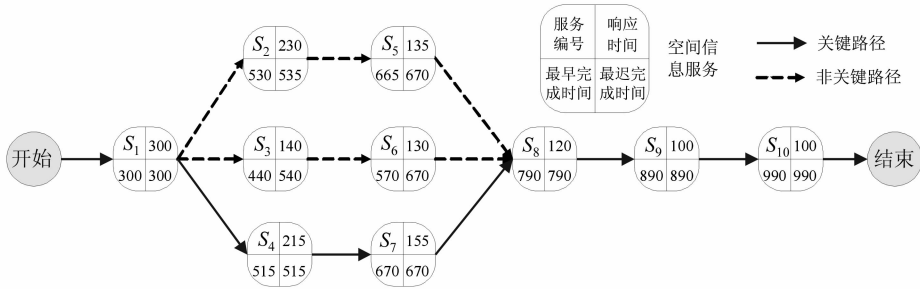


图 1 变化检测服务链
Fig. 1 Service Chain of Change Detection

表 1 变化检测组成服务

Tab. 1 Component Services of Change Detection

服务编号	服务名称	服务描述	响应时间/s
S ₁	工程定义和描述	准确定义应用问题,确定研究区域和研究对象,明确研究目标,进行成本估算	300
S ₂	数据获取	获取遥感影像 1	230
S ₃	数据获取	获取遥感影像 2	140
S ₄	数据获取	获取遥感影像 3	215
S ₅	数据预处理	影像镶嵌、影像裁剪、几何校正、辐射校正	135
S ₆	数据预处理	影像镶嵌、影像裁剪、几何校正、辐射校正	130
S ₇	数据预处理	影像镶嵌、影像裁剪、几何校正、辐射校正	155
S ₈	变化检测与分析	选用合适的变化检测方法提取和分析变化信息,生成变化分布图和其他检测结果	120
S ₉	精度评估	评价变化检测的结果精度	100
S ₁₀	结果输出	各种变化检测结果,包括图件、表格、文档的制作输出	100

1.2.2 概率型响应时间计算模型

在未知的和难以估计的因素较多的条件下,对服务响应时间的估计可以采用三点时间估计法^[9]:① 乐观时间 α ,在顺利情况下,完成服务所需要的最少时间;② 悲观时间 β ,在不顺利情况下,完成服务所需要的最多时间;③ 最可能时间 δ ,在正常情况下,完成服务所需要的时间。上述三种时间都具有一定的概率,根据经验,其概率分布可以认为近似服从于正态分布。

假设服务路径 P_i 由 m 个服务组成,则 P_i 的响应时间可以认为是一个服从以式(2)为期望值、以式(3)为方差的正态分布:

$$T_E(P_i) = \sum_{j=1}^m [(\alpha_j + 4\delta_j + \beta_j)/6] \quad (2)$$

$$\sigma_E^2(P_i) = \sum_{j=1}^m [(\beta_j - \alpha_j)/6]^2 \quad (3)$$

概率型响应时间计算模型通过加权平均算出一个期望值作为动态环境下服务的响应时间,把

确定型响应时间计算模型针对响应时间确定性高且可靠性较高的情况,每个服务只给出一个时间值。服务路径响应时间是该路径所有组成服务响应时间的和。

以变化检测为例,在从不同时间获取的多时相遥感影像中对各种地表变化进行识别和提取,将表 1 所示的空间信息服务组合起来,形成如图 1 所示的变化检测服务链。

非确定型问题转化为确定型问题来计算。从概率论的观点分析,其偏差仍不可避免,但趋向有明显的参考价值。响应时间的方差反映了响应时间的分布离散程度,用来估算服务链在指定的响应时间约束 T_K 下完成的概率:

$$\mu = (T_K - T_E)/\sigma_E \quad (4)$$

式中, μ 为概率系数。根据正态分布函数,服务链在 T_K 内完成的概率为:

$$P(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{x^2}{2}} dx \quad (5)$$

定理 服务链在给定时间 T_K 内完成的概率是由概率最小的服务路径决定的。

在 T_E 和 σ_E^2 为已知的条件下即可估算出服务链响应时间的概率,也可以估算出具有一定概率的服务链执行时间。

1.3 服务的最早开始时间、最早结束时间、最迟结束时间和最迟开始时间

1) 服务最早开始时间 b'_i 。等于从服务链的

开始服务起到服务 S_i 最长路径的长度:

$$b'_i = \begin{cases} 0, P_i^B = \emptyset \\ \max\{b'_j + t_j, \forall S_j \in P_i^B\}, P_i^B \neq \emptyset \end{cases} \quad (6)$$

式中, P_i^B 为服务 S_i 的前驱服务集合。

2) 服务最早结束时间 c'_i 为:

$$c'_i = b'_i + t_i \quad (7)$$

3) 服务最迟结束时间 c_i 。在不影响服务链最早结束时间的条件下, S_i 最迟必须结束的时间为:

$$c_i = \begin{cases} c_N, P_i^A = \emptyset \\ \min\{c_j - t_j, \forall S_j \in P_i^A\}, P_i^A \neq \emptyset \end{cases} \quad (8)$$

式中, c_N 为服务链的响应时间; P_i^A 为服务 S_i 的后继服务集合。

4) 服务最迟开始时间 b_i 为:

$$b_i = c_i - t_i \quad (9)$$

5) 服务时间余量 $TF(S_i)$ 。在不影响服务链结束时间的条件下, 服务 S_i 最早开始(或结束)时间可以推迟的时间称为服务 S_i 的服务时间余量, 即

$$TF(S_i) = b_i - b'_i \text{ 或 } TF(S_i) = c_i - c'_i \quad (10)$$

服务时间余量为零的服务, 开始和结束的时间没有机动的余地, 由这些服务所组成的路径就是服务链中的关键路径。

2 顾及关键路径响应时间的空间信息服务优化组合算法

若缩短关键服务的响应时间, 会缩短服务链的响应时间; 反之, 则会延长服务链的响应时间。而缩短非关键服务却不能使服务链的响应时间缩短, 在服务时间余量允许范围内, 延长非关键服务也不至于影响服务链的响应时间。因此, 顾及关键路径响应时间的服务优化组合算法的基本思想是在空间信息服务链中找出关键路径, 优先保证关键服务的响应时间; 而对非关键服务, 只要在不影响服务链的整体响应时间的条件下, 就可以将响应时间短的服务资源优先配置给关键服务, 以达到优化服务配置的目的。

算法的实现思路为: 当服务链的响应时间不能满足响应时间约束时, 重新选择响应时间更快的候选服务替换现有关键服务。候选服务是提供相同或相似功能但响应时间不同的服务。候选服务的选择准则为: 对于每一个关键服务 S_i , 如果存在至少一个响应时间小于当前服务实例的候选服务, 则选择其中响应时间最大的一个; 否则, 保持原来的。算法流程如图 2 所示。

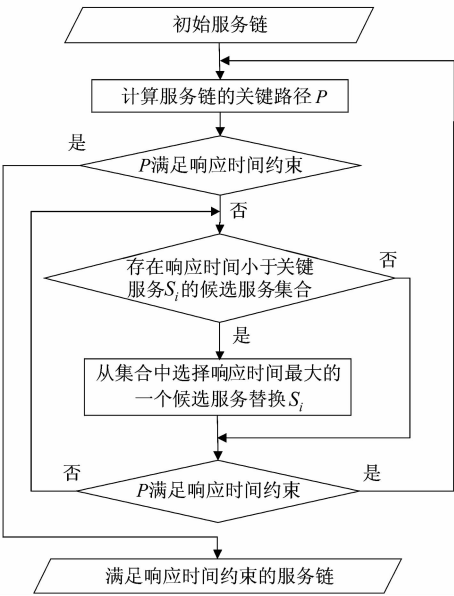


图 2 算法流程图

Fig. 2 Flowchart of Optimal Composition Algorithm

3 实验分析

定义响应时间成功率(response time success rate, RTSR)为衡量服务组合算法性能的标准。空间信息服务链能够组合成功, 除了满足功能上的需求外, 还需要满足服务的响应时间约束。假设场景中所有 n 条服务链中有 m 条服务链组合成功, 则响应时间成功率定义为 $RTSR = m/n$ 。

在仿真实验中, 设计了 10 条空间信息服务链, 每条服务链包含十余个空间信息服务, 采用三种不同的服务选择算法(穷举法、局部最优法^[7]以及本文算法), 比较算法的时间性能和优化效果。实验共模拟了 3 个场景: ① 20 种服务类型, 每个服务类型拥有 10 个服务实例; ② 20 种服务类型, 每个服务类型拥有 30 个服务实例; ③ 20 种服务类型, 每个服务类型拥有 50 个服务实例。上述 3 种方法在以上 3 个场景各进行 10 次实验, 测得的数据求取平均值, 实验结果如图 3、图 4 所示。

在服务链的生成时间上, 穷举法扩展性差; 局部最优算法和本文算法对场景内服规模不敏感,

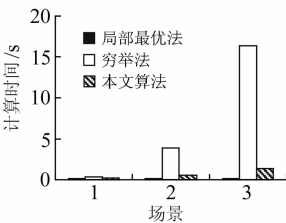


图 3 服务链生成时间

Fig. 3 Time of Service Chain Generation

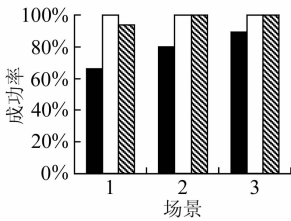


图 4 成功率

Fig. 4 RTSR of Experience Results

具有扩展性好的优点。局部最优算法虽然从各条服务链的微观角度来看是有效的,但是当服务链之间存在争夺服务实例现象时,难以在多条服务链之间实现服务的优化组合,导致 RTSR 不高。本文算法通过区分关键服务与非关键服务,从宏观角度在所有服务链之间实现服务资源的优化组合,RTSR 得到显著提高。

实验结果证明,在存在多条服务链并发组合时,相比局部最优算法,本文算法能够更好地满足所有服务链对响应时间的约束。

参 考 文 献

[1] Onchaga R. Modelling for Quality of Services in Distributed Geoprocessing[C]. The 20th Congress of the ISPRS, Istanbul, Turkey, 2004

[2] Eder J, Panagos E, Rabinovich M. Time Constraints in Workflow Systems[C]. The 11th International Conference(CAISE'99), Berlin, Germany, 1999

[3] Gillmann M, Weikum G, Wonner W. Workflow Management with Service Quality Guarantees[C]. The 2002 ACM SIGMOD International Conference on Management of Data, New York, USA, 2002

[4] Grefen P. CrossFlow: Cross-organizational Workflow Management in Dynamic Virtual Enterprises [J]. International Journal of Computer Systems Science & Engineering, 2000, 15(5): 277-290

[5] Marjanovic O, Orlowska M. On Modeling and Verification of Temporal Constraints in Production Workflows[J]. Knowledge and Information Systems, 1999, 1(2): 157-192

[6] Menasce D A. Response-time Analysis of Composite Web Services[J]. Internet Computing, IEEE, 2004, 8(1): 90-92

[7] Zeng L Z, Benatallah B, Ngu A H H, et al. QoS-aware Middleware for Web Services Composition[J]. IEEE Transactions on Software Engineering, 2004, 30(5): 311-327

[8] Eder J, Gruber W, Panagos E. Temporal Modeling of Workflows with Conditional Execution Paths [M]. Heidelberg: Springer-Verlag, 2000: 243-253

[9] 甘应爱, 田丰, 李维铮. 运筹学[M]. 北京: 清华大学出版社, 1996

第一作者简介:朱庆,教授,博士,博士生导师。现主要从事数字摄影测量、数码城市地理信息系统和虚拟地理环境研究。
E-mail:zhuq66@263.net

Optimal Composition Algorithm for Spatial Information Service Based on Critical Path and Response Time Constraint

ZHU Qing¹ YANG Xiaoxia¹ LI Haifeng¹

(1 State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: The estimation and management of the response time is an important problem of spatial information service. Aiming at the automatic generation of spatial information service chains, the determinate model and the probability model for response time estimation is established, and the concepts of critical path, critical service and service slack are defined. In order to ensure the response time of services chains satisfy the requirement of users, an optimal service composition algorithm based on critical path method is proposed to select appropriate component services. Experimental results show that this algorithm optimizes the response time of spatial information service chains.

Key words: service composition; service chain; response time constraint; critical path method

About the first author: ZHU Qing, professor, Ph.D., Ph.D supervisor, majors in digital photography, cyber city GIS and virtual geographic environment.
E-mail: zhuq66@263.net