

基于改进遗传算法的 QoS 感知 Web 服务组合

马小洁, 王晓军

(南京邮电大学 计算机学院, 江苏 南京 210003)

摘 要:传统遗传算法在种群初始化的时候,普遍采用均匀取种法或随机取种法,这些方法生成的种群的平均适应度比较低,难以保证算法的搜索效率。文中提出一种改进的遗传算法用于 QoS 敏感的 Web 服务组合,采用两种不同的算法进行服务选择,避免了随机生成初始种群给算法带来的负面影响。并且,该算法将路径模板化以减少服务组合的工作量,用染色体可变长的编码方式来解决组合服务的多路径选择问题。通过仿真实验,与传统的算法相比,所提出的算法在实现服务组合时收敛更快,最优解的适应度更高。

关键词: Web 服务组合; 服务质量; 组合计划; 模板; 遗传算法

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2012)08-0089-04

Web Service Composition Supporting QoS Based on Improved Genetic Algorithm

MA Xiao-jie, WANG Xiao-jun

(School of Computer, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)

Abstract: Even or random selecting is the common method used for generating initial population in genetic algorithm, however, the average fitness of the population generated by this method is low, and it is hard to ensure the searching efficiency of algorithm. In this study, propose a novel genetic algorithm (GA) for handling QoS-aware Web service composition, combining two initialized algorithms with GA at initialization stage to improve the algorithm effectiveness. Besides, build a path-template and variable length chromosomes service composition solution, for template paths will make the work easy and variable length chromosomes can support multi-path QoS-aware service composition. The superiority of the algorithm is analyzed theoretically and its effectiveness is demonstrated by experimental results.

Key words: Web service composition; QoS; composition plan; template; genetic algorithm

0 引言

当今, Web Service 的应用范围不断扩大, Web 服务数量也逐渐增多, 这些服务之间难免会有一些服务实现的功能是相近或相同的, 而其服务质量 QoS 却不同。如何从这些具有相同功能的服务中选择服务质量 (QoS) 最优的 Web Service 是每个客户所关心的。

目前, 国内外已经有很多学者对 Web Service 组合技术展开了一定的研究。文献[1]提出用整数规划 (Integer Programming, IP) 来实现 Web Service 组合。文献[2]在 IP 的基础上提出了一种混合整数线性规划法来进行 Web Service 组合。文献[3]给出了遗传算法在 QoS 敏感的 Web Service 组合中的应用。相对于整数

规划法, 遗传算法具有更好的性能, 特别是当候选服务数量比较多的时候。然而, 以上的这些方法都是基于单一路径的选择。文献[4]提出了采用二维编码的遗传算法来解决多路径 Web 服务组合问题, 但缺陷是它不支持某些复杂的结构, 如多路径中的选择结构。并且, 这些文献中所用的遗传算法对初始种群的选取都是随机的, 在一定情况下降低了遗传算法的性能。

针对以上的种种问题, 需要设计一种改进的遗传算法来进行服务组合。文中采用路径模板化, 基因可变长的遗传算法来实现基于 QoS 的 Web Service 组合。并且, 在种群初始化的时候, 采用两种不同的算法进行选择, 这样就避免了随机生成初始种群给算法带来的负面影响。

1 基于 QoS 属性的 Web Service 组合

1.1 QoS 定义

文中, 研究 Web 服务的 execution cost (c), execu-

收稿日期: 2011-12-08; 修回日期: 2012-03-11

基金项目: 国家科技支撑计划 (2007BAH17B04)

作者简介: 马小洁 (1986-), 女, 硕士研究生, 主要研究方向为分布计算技术与应用; 王晓军, 副教授, 硕士研究生导师, 主要研究领域为分布计算技术与应用。

tion time (t), availability (av) 和 reputation (rep) 这 4 个属性。早先的一些研究^[5]也将这几个属性作为衡量 Web 服务的度量,它们的含义定义如下:

execution cost (c): 执行一个 Web 服务的代价,由服务提供者给出。

execution time (t): 这里的 time 是执行服务的平均时间,由服务提供者不断更新。

reputation (rep): Web 服务的信用积分的平均值,计算式为 $rep = \text{服务的总积分} / \text{服务执行的次数}$ 。

availability (av): $av = \text{uptime} / (\text{uptime} + \text{downtime})$, 服务正常执行的时间与总时间的比率。

1.2 组合服务 QoS 计算

通常,一个组合计划里会有顺序、循环、并行和选择 4 个执行模式,其难点是怎么针对这些模式把各种 QoS 指标合成一个综合指标^[6]。文中 QoS 属性归一化的执行过程是:首先,对于所有的顺序模式,执行归一化;然后对并行模式和选择模式执行归一化。具体的计算公式在表 1 中列出。

表 1 组合 QoS 计算

QoS	顺序	循环	并发	选择
time	$\sum_{i=1}^n q_i^t$	$k * q_i^t$	$\max\{q_1^t, \dots, q_n^t\}$	$\sum_{i=1}^n p_i q_i^t$
cost	$\sum_{i=1}^n q_i^c$	$k * q_i^c$	$\sum_{i=1}^n q_i^c$	$\sum_{i=1}^n p_i q_i^c$
avai	$\prod_{i=1}^n q_i^a$	$(q_i^a)^k$	$\prod_{i=1}^n q_i^a$	$\sum_{i=1}^n p_i q_i^a$
rep	$\sum_{i=1}^n q_i^r / n$	q_i^r	$\sum_{i=1}^n q_i^r / n$	$\sum_{i=1}^n p_i q_i^r$

QoS 属性归一化的过程就是生成虚拟计划,即组合计划。该过程不停地迭代,到最后只有一个抽象服务存在。例如图 1,这个组合计划中有 7 个原子服务,包含了并发、选择和顺序结构。

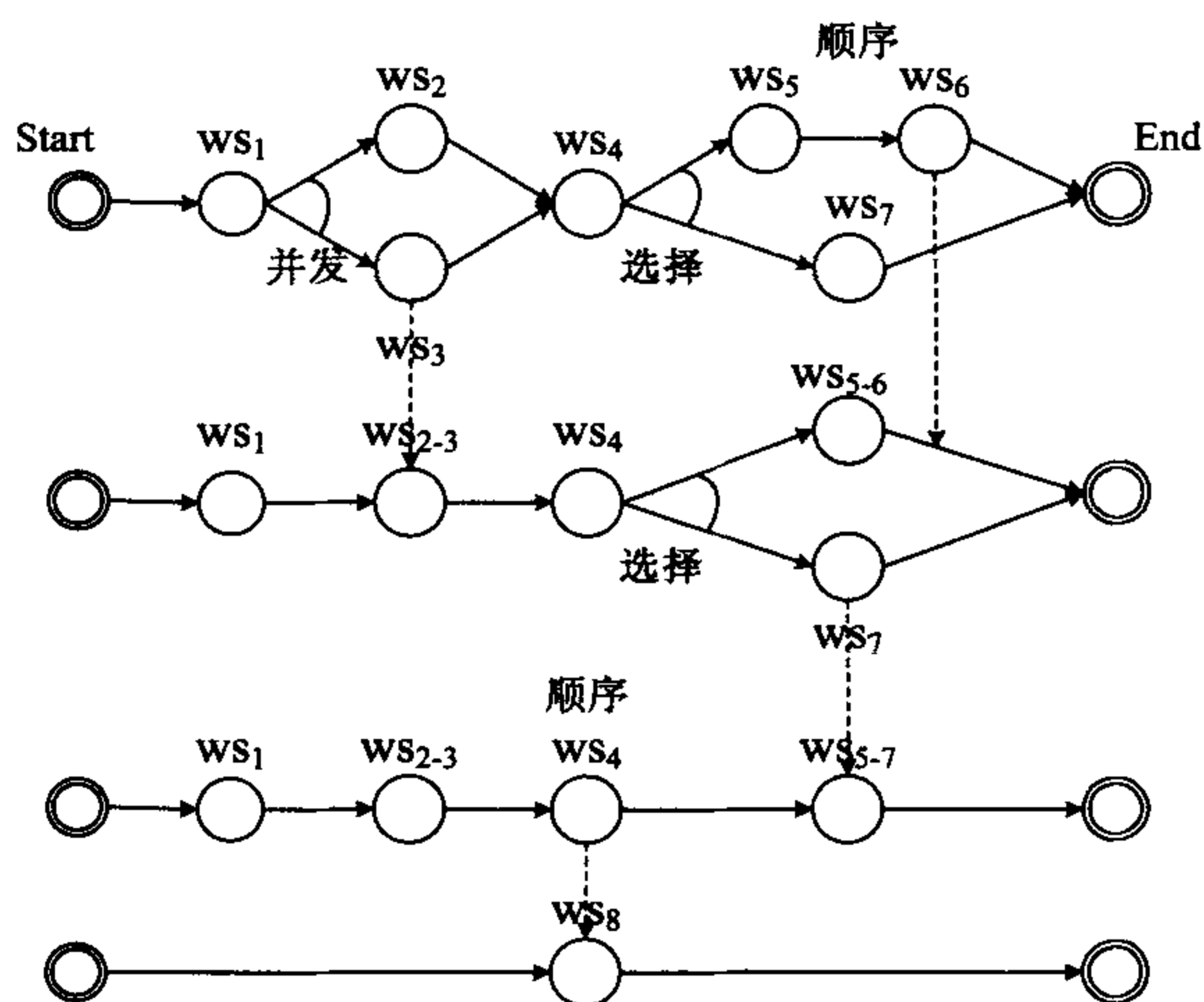


图 1 QoS 属性归一化的过程示意图

首先,将顺序执行的抽象服务 ws_5 和 ws_6 按照顺序计算公式组合成 ws_{5-6} ;然后,用选择模式计算公式将 ws_{5-6} 和 ws_7 组合成 ws_{5-7} 。类似的,用并发模式计算式

方法将服务 ws_2 和 ws_3 组合成一个服务 ws_{2-3} 。最后,根据顺序计算式,将 $ws_1, ws_{2-3}, ws_4, ws_{5-7}$ 组合成 ws_8 , ws_8 具有归一 QoS 属性值,和每一个原子服务 ws_i 有关。

2 改进的遗传算法的实现

GA (genetic algorithm) 算法将问题的求解过程表示为染色体优胜劣汰的过程^[7,8],通过种群的迭代进化,包括选择、交叉、变异等操作对种群进行组合产生下一代个体,逐步向优化的种群进化,最终求得适合环境的染色体,也就是问题的最优解或近似最优解。

2.1 染色体的表示

路由由不同数量的抽象服务组成,相应地,具有不同长度的组合计划就可以表示为长度可变的染色体。组合路径上的每个任务也称为抽象服务,每个抽象服务都包含一个与其任务对应的候选服务集,候选服务集中的每个服务再指向表示 QoS 属性的记录。采用可变长的编码方式,对于每一条染色体,因为其对应的路径不同,其染色体的长度也不同,但染色体中第一个/最后一个基因总是服务组合的起点/终点,染色体中间的每一个基因对应候选服务集中的一个具体服务,其结构如图 2 所示。

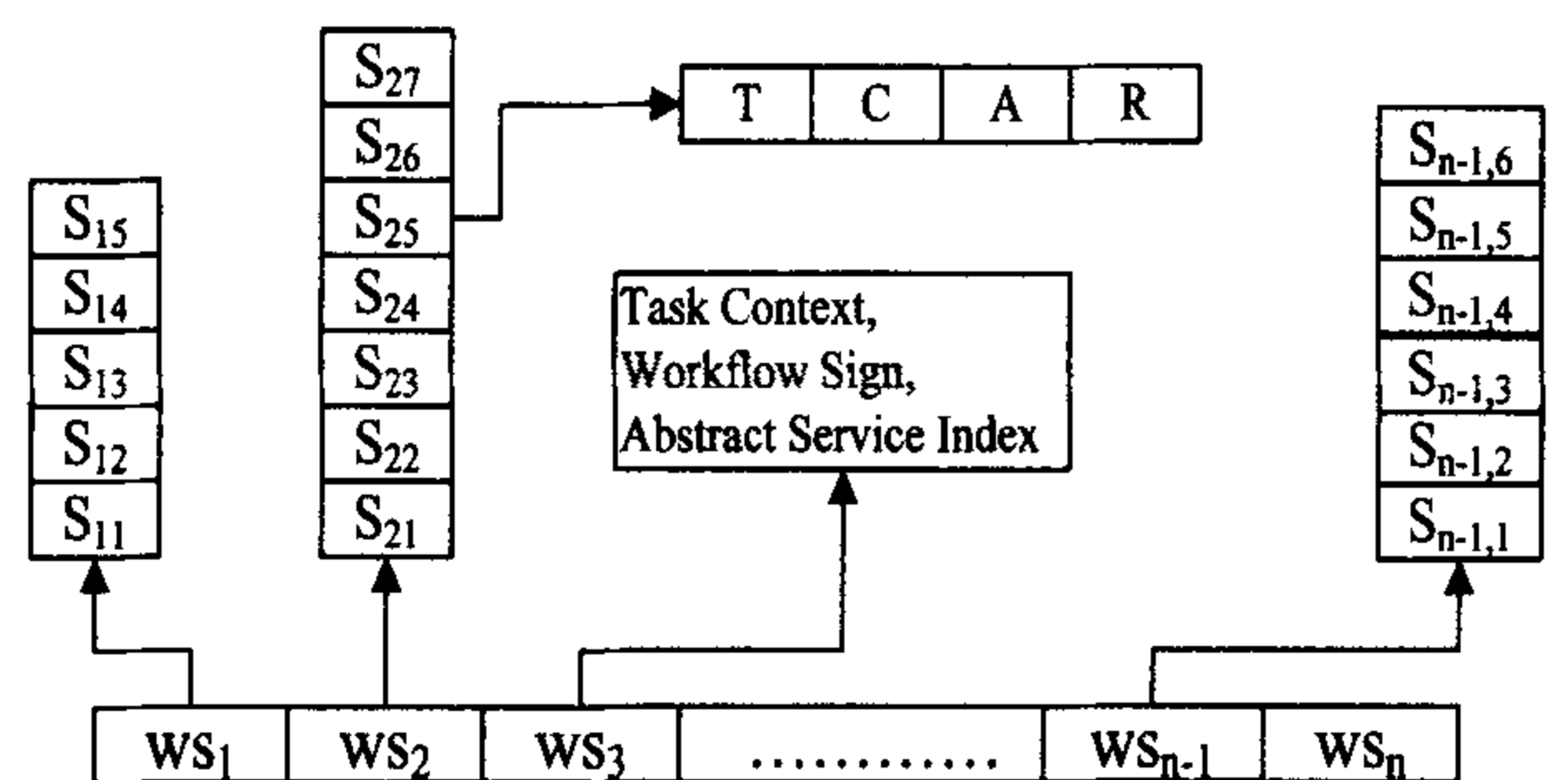


图 2 染色体结构示意图

基因的结构表示为一个具有三个成员的结构体 (task context, workflow sign, abstract service index),如图 2 所示。其中,task context 描述了抽象服务的输入/输出参数,用于进行抽象服务的参数匹配以实现交叉点的选择;workflow sign 用来检查该抽象服务的执行模式,用于选择合适的 QoS 归一化公式;而 abstract service index 代表抽象服务的编号,也指示了对应的具体服务集合。每一个服务的 QoS 由记录 (T, C, A, R) 表示, T, C, A, R 分别表示 time, cost, availability 和 reputation。

2.2 初始种群生成

众所周知,在许多启发式搜索算法中(包括遗传算法),初始种群的构造普遍采用均匀取种法或随机取种法,这些方法很简单但是具有较大的随机性,种群的平均适应度比较低,难以保证算法的搜索效率^[9,10]。所以,提供一个良好的种群初始化解方案,可以使搜

索算法具有更好的收敛性。

文中采用两种种群初始化算法。第一种生成初始种群的算法是利用 reputation 和 availability 作为选择具体服务绑定到抽象服务的标准,因为一般用户对这两个指标比较关心。在算法的选择过程中:对每一个抽象服务,到对应的具体服务集中选择 reputation score 最高的 Web Service,如果多个 Web Service 具有相同的 reputation score,则在这些 Web Service 中选择 availability 最高的 Web Service,如果有多个 Web Service 符合要求,就随机选择一个。

在第二种生成初始种群的算法中,使用了局部最优方法。首先要设计一个评分函数来评价 Web Service 的 QoS。在算法选择过程中,对于每一个抽象服务,选择服务质量评分最高的 Web Service,如果不止一个,则随机选择一个。在文中,服务的 QoS 由 4 个 QoS 属性来描述,即 $\{q^1, q^2, q^3, q^4\}$, 和 $\{time, cost, reputation, availability\}$ 所对应。算法二所用到的评分函数如下:

$$score(ws_{ij}) = \sum_{k=1}^4 w_k q_{ij}^k \quad (1)$$

在公式(1)中, w_k 为 QoS 属性 q^k 的比重系数,由客户定义。 q^k 代表服务的第 k 维属性, ws_{ij} 表示为第 i 个抽象服务所选的编号为 j 的服务。

2.3 路径模板化

组合服务的路径,是按照抽象服务的参数匹配来编排的。对于一个事务,根据输入/输出参数可能会生成若干个满足用户 QoS 需求的路径。在这样的情况下,路径的数量是一定的。由于每个抽象服务所对应的候选服务集中有多个具体服务,所以在绑定具体服务后,每条路径对应的组合计划的数量是不确定的。因此,文中把满足用户要求的路径组成一个路径集,即把路径模板化^[11,12]。这样,在交叉变异后,得到的路径还是在模板集中,不需要额外的选择判断,从而可以简化算法的工作量。

以图3所示的旅行服务为例,共有13个抽象服务,根据每个抽象服务所要求的输入/输出参数,可得到满足要求的4条路径 $\{1, 2, 4, 5, 6, 7, 8\}$, $\{1, 2,$

$4, 9, 10, 11, 12, 13\}$, $\{3, 4, 5, 6, 7, 8\}$ 和 $\{3, 4, 9, 10, 11, 12, 13\}$, 即模板集中有4条路径可选,避免了不必要的判断,方便算法的实现。

3 仿真实验和结果分析

3.1 仿真实验介绍

在仿真实验部分,用传统的遗传算法和改进的遗传算法对组合服务进行选择,并将对这两种算法的收敛性以及计算效率两方面分别比较分析。为了实验实现的简单,把并发/选择模式的服务作为一个小的服务组合,把循环模式的服务按照执行次数展开,看作顺序模式的服务^[12]。

文中所用的适应度函数定义如下:

$$fitness(cp) = \frac{w_3 q^r + w_4 q^a}{w_1 q^c + w_2 q^t} - w_5 D(cp) \quad (2)$$

$$D(cp) = \frac{gen}{mgen} \sum_{k=1}^4 \left(\frac{\Delta q^k}{q_{cons}^k} \right)^2 \quad (3)$$

在(2)式中, w_i 对应第 i 维 QoS 属性的权重, q^c , q^t , q^r , q^a 分别表示归一化后的组合服务的 QoS 属性 time, cost, reputation, availability 的值, q_{cons}^k 表示第 K 维属性的 QoS 约束,由用户定义。 Δq^k 的计算则和 QoS 属性的正负性有关。

仿真实验的软硬环境都是一样的, CPU 为 Intel Duo 2. 20GHz, 内存 2GB, 操作系统为 Windows XP 2002, 开发语言为 C, IDE 为 Microsoft Visual Studio 2008。实验所采用的数据也是一样的,以图3给出的旅行服务为例进行仿真。进行三个算法的比较, GA (随机选择), GA1 (用 reputation 和 availability) 和 GA2 (用评分函数 score)。

3.2 仿真结果分析

在第一组实验中,抽象服务数为13,服务的 QoS 属性 cost, time, reputation 和 availability 的约束值都设定为 0.5。每个抽象服务所对应的候选服务集中的服务个数是一样的,有5组数据,分别为 20, 40, 60, 80, 100。所有的结果都是程序运行10次之后的平均值。

记录了三个算法在每组候选服务下找到最优解的

时间。由图4可以看出,在候选服务的个数为20的时候,传统的遗传算法找到最优解的时间比较短,而在候选服务个数超过60的时候,文中给出的算法能够更快地找到最优组合,并且,算法 GA1 比算法 GA2 能够更快地找到最优服务组合。

在第二组实验中,假设每个任务所对应的候选服务为10,每一维

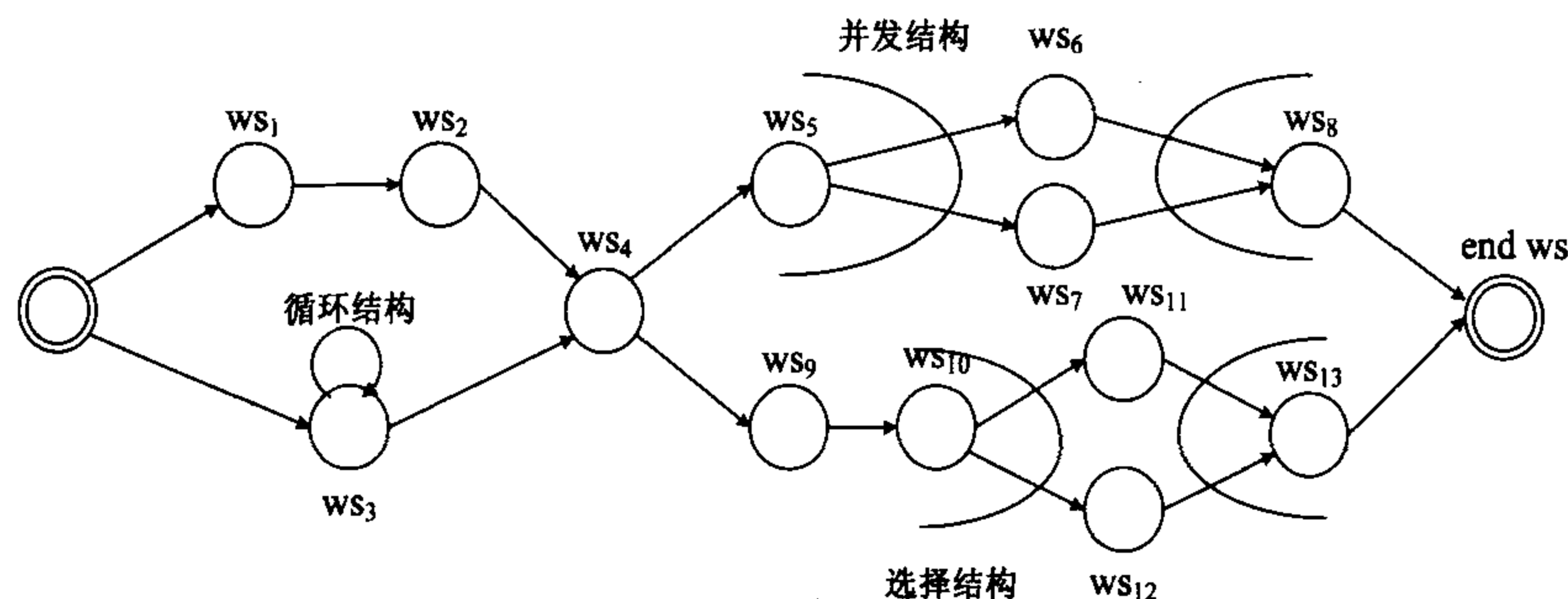


图3 旅行服务状态图

QoS 属性的值都是 0.5。遗传代数变化范围是 100 到 1000。所有的结果都是程序运行 10 次之后的平均值。

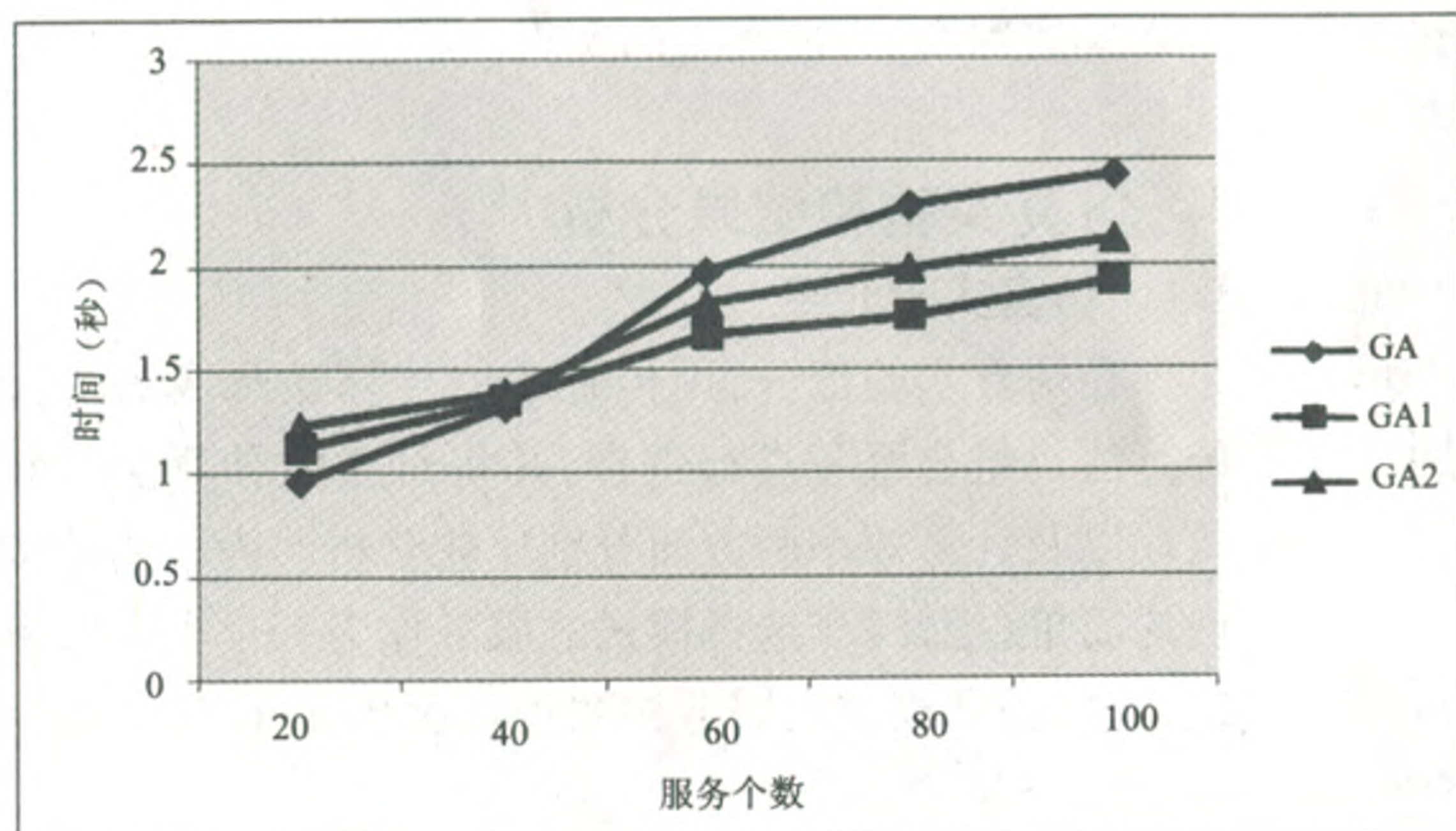


图 4 算法找到最优解的时间

图 5 显示了算法 GA, GA1, GA2 在每组测试下最优服务组合的适应度。在遗传代数小于 300 的时候, 传统的 GA 有比较好的解。但是, 在遗传代数大于 300 的时候, GA1, GA2 的解会一直优于传统 GA 的解。这说明, 相比于传统 GA, 在遗传一定代数后, GA1, GA2 能够找到更好的解, 并且, GA2 的解要优于 GA1。

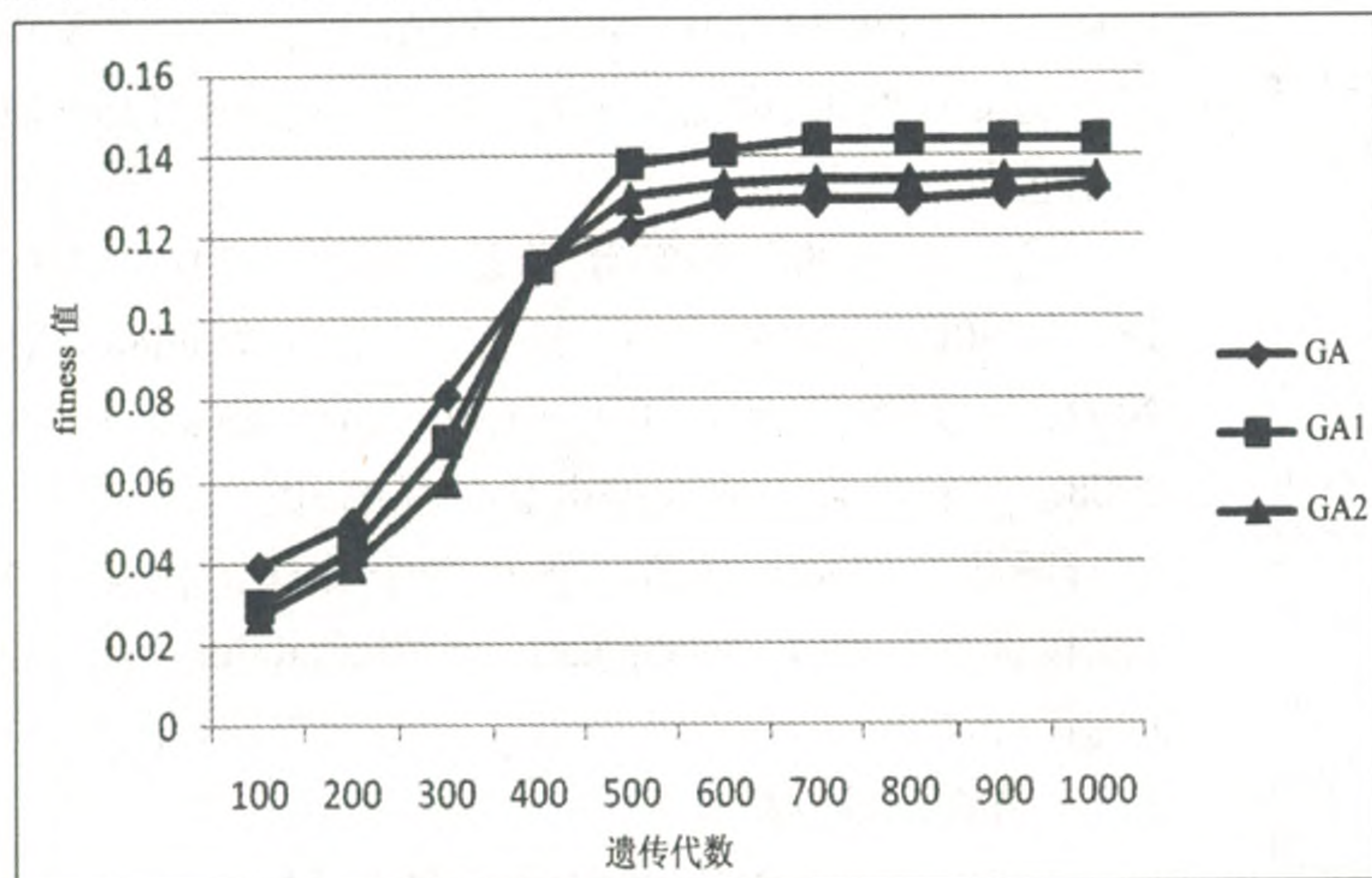


图 5 适应度的比较

4 结束语

遗传算法是一个随机算法, 初始种群的适应度值将会影响算法的收敛性和解的适应性。在初始化种群的时候, 文中给出了两种算法来生成初始群体, 提高种群的适应度。在组合路径生成方面, 根据抽象服务进行参数匹配, 找到满足用户条件的每一条路径, 形成一个路径集, 简化了服务组合的工作量。

通过仿真实验结果可知, 改进的遗传算法能够很

容易地实现多路径服务组合, 找到最优服务组合。但是抽象服务之间的参数匹配并不适合于所有的服务组合, 可以考虑将用户的偏好和 QoS 属性结合进行服务组合, 这是下一步研究的重点。

参考文献:

- [1] Zeng L Z, Boualem B, Ngu A H H, et al. QoS-aware middleware for web services composition [J]. Software Engineering, 2004, 30 (5): 311-327.
- [2] 王阳阳, 李俊, 陈志国, 等. 一种基于 QoS 全局最优的服务选择算法[J]. 计算机应用研究, 2010, 27(5): 1659-1661.
- [3] Ko J M, Kim C O, Kwon Ick-Hyun. Quality-of-service oriented web service composition algorithm and planning architecture [J]. Journal of Systems and Software, 2008, 81 (11): 2079-2090.
- [4] 张成文, 苏森, 陈俊亮. 基于遗传算法的 QoS 感知的 Web 服务选择[J]. 计算机学报, 2006, 29(7): 1029-1037.
- [5] Xu Kun, Xu Yifan. Optimization of Submarine Hydrodynamic Coefficients Based on Immune Genetic Algorithm [J]. Journal of China Ordnance, 2010(3): 200-205.
- [6] Luo Yuansheng, Qi Yong, Hou Di, et al. A novel heuristic algorithm for QoS-aware end-to-end service composition [J]. Computer Communications, 2011, 34(9): 1137-1144.
- [7] Lin Chia-Feng, Sheu Ruey-Kai, Chang Yue-Shan, et al. A relaxable service selection algorithm for QoS-based web service composition [J]. Information and Software Technology, 2011, 53(12): 1370-1381.
- [8] 陈亮, 孙敏. 基于免疫遗传算法的 Web 服务组合方法[J]. 计算机工程, 2010, 36(10): 226-230.
- [9] Zeng Liangzhao, Benatallah B, Dumas M. QoS-aware Middleware for Web Services Composition [J]. IEEE Transactions on Software Engineering, 2004, 30(5): 311-327.
- [10] 代钰, 杨雷, 张斌. 支持组合服务选取的 QoS 模型及优化求解[J]. 计算机学报, 2006(7): 1167-1178.
- [11] 唐小燕, 李斌. Web 服务集成中基于 QoS 的服务选择[J]. 计算机应用, 2006, 26(6): 242-247.
- [12] 张文博, 史维峰. 基于 BPEL 和 QoS 的动态 Web 服务组合框架研究[J]. 计算机技术与发展, 2009, 19(11): 72-75.

作者:	陈延利, 施永豪
作者单位:	陈延利(西藏大学工学院电子工程系,西藏拉萨850000), 施永豪(西南交通大学信息科学与技术学院通信与信息工程系,四川成都610031)
刊名:	计算机技术与发展
英文刊名:	Computer Technology and Development
年, 卷(期):	2012(8)

参考文献(12条)

- 1.Zeng L Z;Boualen B;Ngu A H H QoS-aware middleware for web services composition 2004(05)
- 2.王阳阳;李俊;陈志国 一种基于Qos全局最优的服务选择算法[期刊论文]-计算机应用研究 2010(05)
- 3.Ko J M;Kim C O;Kwon Ick-Hyun Quality-of-service oriented web service composition algorithm and planning architecture[外文期刊] 2008(11)
- 4.张成文;苏森;陈俊亮 基于遗传算法的QoS感知的Web服务选择[期刊论文]-计算机学报 2006(07)
- 5.Xu Kun;Xu Yifan Optimization of Submarine Hydrodynamic Coefficients Based on Immune Genetic Algorithm[期刊论文]-Journal of China Ordnance 2010(03)
- 6.Luo Yuansheng;Qi Yong;Hou Di A novel heuristic algorithm for QoS-aware end-to-end service composition 2011(09)
- 7.Lin Chia-Feng;Sheu Ruey-Kai;Chang Yue-Shan A relaxable service selection algorithm for QoS-based web service composition 2011(12)
- 8.陈亮;孙敏 基于免疫遗传算法的Web服务组合法[期刊论文]-计算机工程 2010(10)
- 9.Zeng Liangzhao;Benattallah B;Dumas M QoS-aware Middle-ware for Web Services Composition[外文期刊] 2004(05)
- 10.代钰;杨雷;张斌 支持组合服务选取的Qos模型及优化求解[期刊论文]-计算机学报 2006(07)
- 11.唐小燕;李斌 Web服务集成中基于QoS的服务选择 2006(06)
- 12.张文婧;史维峰 基于BPEL和Qos的动态Web服务组合框架研究[期刊论文]-计算机技术与发展 2009(11)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjrz201208023.aspx