

面向 Pareto 最优遗传算法的服务组合方法

胡焕耀^{1,2}, 董渭清¹, 符锐¹, 张星¹, 赵晓映¹

(1. 西安交通大学电子与信息工程学院, 710049, 西安; 2. 中兴通讯股份有限公司上海研发中心, 201203, 上海)

摘要: 为了解决 Pareto 遗传算法在每一次进化操作中都要构造当前进化群体最优解集而影响运行效率的问题, 提出了一种面向 Pareto 最优遗传算法的服务组合方法, 以实现 Web 服务组合的全局优化. 用伪二叉树法则构造目标函数的 Pareto 最优集合, 再进行 Pareto 最优解集排序, 最后采用个体相似度计算来确定遗传算法的适应度函数, 由此获得一组满足约束条件的 Pareto 最优解服务集合. 实验验证表明, 所提方法可以提高多目标遗传算法处理服务组合效率的问题, 即使在服务规模较大的情况下, 所获得的解与最优值的比率仍能接近 90% 的水平.

关键词: Web 服务组合; 全局优化; 遗传算法; Pareto 最优

中图分类号: TP305 **文献标志码:** A **文章编号:** 0253-987X(2009)12-0050-05

Pareto Optimality Based Genetic Algorithm in Web Services Composition

HU Huanyao^{1,2}, DONG Weiqing¹, FU Rui¹, ZHANG Xing¹, ZHAO Xiaoyi¹

(1. School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

2. ZTE Shanghai R&D Center, Shanghai 201203, China)

Abstract: A new method of service selection is proposed to implement the global optimal selection of Web services, and to avoid the routine in a genetic algorithm where an optimal evolving set is built for every current evolving operation to impose a negative influence on the efficiency of the algorithm. The pseudo binary tree's rule is used to produce a set of optimal Pareto solutions. Then the optimal Pareto solutions are sequenced and individuals' similarity of each Pareto solution is calculated to determine the fitness function in the genetic algorithm. Experimental results show of proposed method generates a result that has a similarity up to 90% compared with the optimal result, even in the case of a large number of services.

Keywords: Web services composition; global optimization; genetic algorithm; Pareto optimality

随着 Web 服务应用范围的不断扩大, 网络 Web 服务数量也不断增加, 由此不可避免地出现了许多具有相同功能但不同服务质量(QoS)的 Web 服务.

目前对于服务的组合研究大部分为基于 QoS 局部最优的方法, 即对服务的 QoS 参数进行加权操作, 并以此为依据在服务的组合流程中选择具体的服务, 如基于整数规划的方法^[1], 考虑了开销、响应时间、有效性、和可靠性等 QoS 属性, 并通过在目标函数的计算过程中增加对数缩变来扩展相似的属性

等. 这些方法的约束条件和目标函数具有线性的特点, 无法满足用户对于服务的复杂需求. 基于 QoS 全局最优服务动态选择方法, 利用多目标遗传算法群体进化机制^[2], 将服务的组合过程转化为一个带约束条件的多目标优化问题^[3].

在进化算法中, 基于 Pareto 的方法是通过构造当前进化群体的 Pareto 最优解, 并使之不断逼近真实的 Pareto 最优解来实现. 由于每一次进化操作都要构造当前进化群体的最优解集, 因此构造最优解集的速度直接影响了进化算法的运行效率. 本文采

用了伪二叉树法则构造满足约束条件的 Pareto 最优解集合,并通过对最优解集排序以及个体的相似度计算来获得遗传算法的适应度函数,最终产生一组满足约束条件的 Pareto 最优解服务集合.

1 组合服务的基本模型以及 QoS 计算方法

在服务的组合过程中,假设最终生成的组合服务由 n 个服务组成,且每一个服务都有 m 个备用的服务可供选择,因此在完成这个组合的过程中,会有 m^n 种不同的组合方式. 可以根据 QoS 来评价这些组合方式的优劣,从而完成服务的选择,因此需要从单个服务的 QoS 属性值来计算出组合服务的 QoS 属性值. 在服务的组合过程中,有 4 种基本的组合结构,分别为顺序结构、选择结构、并行结构和循环结构.

QoS 属性的物理含义决定了聚合规则具有结构相关性,即它们在不同结构下具有不同的叠加意义. 表 1 中给出了不同结构组合服务 QoS 属性的聚合规则. 其中, $T(S)$ 表示相应服务执行时间, $C(S)$ 表示服务费用, $A(S)$ 表示可用性, $R(S)$ 表示可靠性, S 表示组合服务, s_i 表示子服务, α 表示相应服务的执行概率, ρ 表示相应服务执行循环的次数.

表 1 QoS 属性聚合规则

QoS 属性	顺序结构	选择结构	并行结构	循环结构
$T(S)$	$\sum_{i=1}^n T(s_i)$	$\sum_{i=1}^n \alpha(s_i) T(s_i)$	$\max_{i=1}^n T(s_i)$	$\rho(s_i) T(s_i)$
$C(S)$	$\sum_{i=1}^n C(s_i)$	$\sum_{i=1}^n \alpha(s_i) C(s_i)$	$\sum_{i=1}^n C(s_i)$	$\rho(s_i) C(s_i)$
$A(S)$	$\prod_{i=1}^n A(s_i)$	$\sum_{i=1}^n \alpha(s_i) A(s_i)$	$\prod_{i=1}^n A(s_i)$	$(A(s_i))^{\rho(s_i)}$
$R(S)$	$\prod_{i=1}^n R(s_i)$	$\sum_{i=1}^n \alpha(s_i) R(s_i)$	$\prod_{i=1}^n R(s_i)$	$(R(s_i))^{\rho(s_i)}$

2 基于遗传算法的服务组合方法

2.1 遗传算法的相关操作

采用遗传算法来解决基于 QoS 的 Web 服务组合问题,根据服务组合的具体情况来确定如何对染色体进行编码,以及如何对染色体执行交叉、变异的操作.

(1)基因:通过对基因位进行分组,每一组基因位的值对应抽象服务中的某一个具体的服务.

(2)染色体:表示为一个可能来自不同抽象服务的具体服务的组合,每一个染色体由若干基因位组成,是遗传算法操作的基本对象.

(3)代:每一代都是由若干个固定数量的染色体组成.

(4)群体:由独立的并能实现一定功能的若干 Web 服务组成,其规模限制了算法的搜索空间.

(5)交叉操作:对染色体采用单点交叉,即随机选择一个交叉点,将其中一个染色体交叉点前半部分与另一个染色体的后半部分组合,后半部分与另一个染色体的交叉点的前半部分组合,从而生成 2 个新的染色体.

(6)变异操作:对染色体基因的某一位按照一定的概率随机变化,从而将其所在基因组指向的 Web 服务转变为一个新的 Web 服务.

2.2 基于 Pareto 最优排序的适应度函数计算

采用轮盘赌选择方法进行染色体的选择,每一条染色体被选中的概率和它的适应度成正比. 假设第 i 条染色体的适应度为 f_i ,而它所在的群体中个体数为 N ,那么第 i 条染色体能够被选中的概率为

$$P_i = f_i / \sum_{i=1}^N f_i, \text{ 其中 } i = 1, 2, \dots, N.$$

染色体适应度的计算过程是给每一条染色体赋予一个适应值,从而对当代群体内染色体之间的优劣关系进行排序,选择具有较高质量的染色体参与到后续的遗传操作中. 本文采用了文献[2]中对染色体适应值赋值的方法,并用伪二叉树法则^[4]对 Pareto 最优解集的构造方法进行改进. 群体染色体优劣关系排序的具体方法描述如下.

(1)基于多个目标函数在整个 Web 服务群体中生成 Pareto 最优解集,并将这些最优解集中的个体标记为顺序 1.

(2)在 Web 服务群体中移除上一步中生成的最优解集,再在余下的个体中生成最优非劣个体并将他们标记为顺序 2.

(3)上述过程持续进行,直到 Web 服务种群中所有个体都分配到次序为止.

对于多个个体具有相同排序值的情况,为了加快 Pareto 最优解的收敛速度并保持种群多样性,采用了小生境技术^[5],根据个体之间的相似程度进行个体适应值修正,通过计算个体之间的相似度,让相似度较高的个体获得较低的被选择概论,从而使所获得的近似解的分布性最好,实现种群多样性的保持. 假设个体 i 的 Pareto 最优解集顺序为 K_i ,相似度值为 D_i ,则个体 i 的适应度 $f_i = \frac{1}{(D_i K_i)}$.

从上述的讨论中可知,由于在每一次进化操作

中都要构造生成种群 Pareto 最优解,因此,构造 Pareto 最优解的速度将直接影响到遗传算法的运行效率.本文采用伪二叉树法则来构造 Pareto 最优解集,通过提高 Pareto 最优解集的构造效率来提高服务组合的效率.

假设 P 为 n 个目标的进化群体,初始时令非支配集 $N = \emptyset$, H 为非支配解集中解的个数,初始 $H = 0$. 本文构造非支配集的步骤如下.

- (1) 从进化群体 P 中选一个体 X .
- (2) 令 $P = P - X$, $N = N + X$, $H = H + 1$.
- (3) 若 P 为空,则执行步骤(6),否则从 P 中任取一个体 Y ,令 $P = P - Y$,取 $H' = H$.
- (4) 从 N 中取任一个体 Z ,令 $H' = H' - 1$.
- (5) 若 $Z > Y$,则执行步骤(3);若 $Y > Z$,则 $H = H - 1$, $N = N - Z$;若 $H' = 0$,则 $N = N + Y$, $H = H + 1$,执行步骤(3);否则执行步骤(4).
- (6) 输出非支配解集 N 中的解.

如果在一轮的比较中,新获取了个体支配集 N 中的个体,则将被支配个体从 N 中剔除,将新个体插入到 N 中,并同 N 中剩余个体进行比较,清除被其支配的个体.

由于在 N 中的个体是相互不支配的,因此在比较过程中,如果新个体 X 支配 N 中的个体 Y ,则 X 不会被 N 中其余个体支配,即可将 X 直接插入到 N 中.

最后采用最大进化代数的方法,设定遗传算法的终止条件:

- (1) 设定最大的进化代数,循环操作直到进化代数为最大进化代数时,则认为搜索算法结束;
- (2) 当进化代数还没有达到设定的最大进化代数时,如果最优化的染色体在给定的进化次数内没有发生变化,则认为搜索算法结束.

采用上述方法,对染色体反复进行选择、复制、变异等操作,使结果不断优化,同时在进化过程中,采用伪二叉树法则提高 Pareto 最优解集的构造效率,最终实现在较短时间内,有效地获得基于 QoS 的 Web 服务组合问题的最优解.

2.3 本文算法复杂度分析

本文的改进遗传算法解决服务组合的执行步骤如下:

- (1) 随机生成规模为 N 的初始种群;
- (2) 当进化代数不为 0 时,执行步骤(3),否则执行步骤(9);
- (3) 计算染色体 i 的目标函数值;

(4) 采用伪二叉树法则构造 Pareto 最优解集,并计算最优解集顺序 K_i ;

(5) 计算染色体 i 的相似度 D_i ;

(6) 计算染色体 i 的适应值 $f_i = \frac{1}{(D_i K_i)}$;

(7) 对染色体进行选择、交叉、变异操作;

(8) 执行步骤(2);

(9) 输出 Pareto 最优解集合.

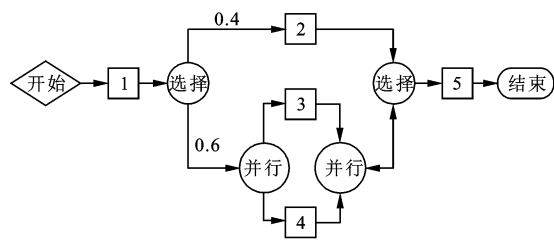
改进的遗传算法在进化迭代过程中的时间复杂度主要包括 2 个部分,即染色体适应值的分配和生成 Pareto 最优集时染色体之间的相互比较.假设目标函数的数目为 r ,约束条件为 m ,染色体的长度为 L ,群体规模为 N ,遗传算法的进化代数为 M .染色体适应值的分配时间复杂度为 $O(N^2(r+m+L))$,在最差情况下生成 Pareto 最优集时的时间复杂度为 $O\left(\frac{(r+m)N^2}{2}\right)$,因此本文算法总的时间复杂度为 $O\left(\left((r+m+L) + \frac{(r+m)}{2}\right)N^2M\right)$,文献[3]中采用遗传算法实现服务的全局最优选择的时间复杂度为 $O\left(\left((r+m+L) + \frac{(r+m)N^*}{N}\right)N^2M\right)$,其中 N^* 为辅助种群的群体规模,在多目标优化的实际测试环境下,每一代非支配个体的比率一般都会保持在 60%^[6],即 N^*/N 的值保持在 0.6,可见,本文所采用的算法在时间复杂度性能上有所提高.

3 实验结果及讨论

实验验证的环境为 Intel(R) Pentium(R) CPU 3.0 GHz 处理器,512 MB 内存,Windows XP 操作系统,在 Windows XP 平台下用 Microsoft Visual Studio 2003 编译运行.设定遗传算法的群体规模为 300,染色体交叉概率为 0.8,染色体变异概率为 0.03.每一个服务的 QoS 属性参数在规定的范围之内随机生成,参数规定的取值范围为 $0 < T < 10$ s, $0 < C < 100$ 元, $0 < R < 5$, $0 < A < 1$,将服务组合的执行时间和费用作为 2 个优化目标,可靠性和可用性作为约束条件,并根据单个服务可靠性 R 和可用性 A 的取值范围,以及图 1 所示的服务组合流程,在实验中设定组合服务的最低可靠性和可用性分别为 13 和 0.05.

实验中选择了 5 类抽象的服务,它们的组合流程如图 1 所示.

通过服务组合算法,在 5 个抽象服务类中选择具体服务生成一个组合服务,使得组合服务在满足



1~5 为抽象服务的编号;0.4、0.6 为执行概率
图 1 服务组合流程

2 个约束条件的情况下,执行时间短、费用少. 根据表 1 中的计算公式以及图 1 所示的服务组合流程,可以得到服务组合算法的目标函数和约束函数

$$\begin{aligned} T &= T_1 + 0.4T_2 + 0.6\max(T_3, T_4) + T_5 \\ C &= C_1 + 0.4C_2 + 0.6(C_3 + C_4) + C_5 \\ R &= R_1(0.4R_2 + 0.6R_3R_4)R_5 \\ A &= A_1(0.4A_2 + 0.6A_3A_4)A_5 \end{aligned}$$

3.1 有效性实验

实验通过算法执行的效率来验证本文方法解决实际问题的有效性. 在实验中设计服务群规模为 4、8 和 16,进化代数数为 100、200、300 和 400,利用本文方法来求解满足要求的组合服务的时间开销. 对于每一种情况,运行 10 次取平均值. 由图 2 可以看出,随着服务群数目的增加,在不同的进化代数下,算法运行时间只有少量的增加. 在本文的服务组合流程中有 5 个服务类、服务群规模为 16 的情况下,算法迭代 400 次得到优化解的执行时间低于 18 s. 因此,该方法的有效性还是明显的.

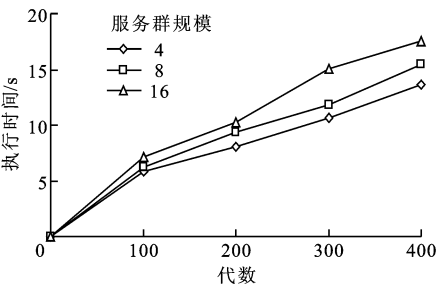


图 2 本文算法平均执行时间

3.2 可行性实验

实验目的是验证本文方法找到 QoS 全局最优组合服务的可行性. 与其相比较的方法是穷举法. 图 3 是在服务群规模为 4、8 和 16 的情况下,算法迭代 400 次所获得的 2 个目标函数的全局最优解与穷举最优值的比率 K . 算法的全局最优解是在相同的服务群体和服务群规模的前提下,运行 20 次所获得的

全局最优解的平均值,而穷举最优值解是在满足约束条件下,组合服务的 C 和 T 两个目标函数分别穷举所获得的最优值. 由图 3 可看出,在服务群规模为 4 的情况下,算法生成的全局最优解与穷举最优解相等,随着服务群规模的增大,算法生成的全局最优解与穷举最优解的比率递减. 同文献[3]的实验结果相比,在服务群规模较小,如规模不大于 10 时,获取的全局最优解与穷举最优解的比率基本保持在同一水平上,而当服务群规模较大时,全局最优解与穷举最优解的比率还能保持在接近 90% 的水平. 因此,本文算法能够获取接近穷举生成的最优解,即本文算法用来解决带约束的多目标优化问题是可行的.

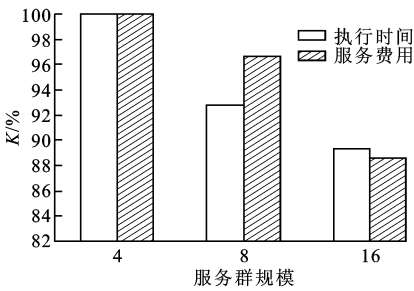


图 3 最优解与穷举最优解的比率

图 4 所示是在服务规模为 8、算法迭代 400 次所产生的一组全局最优非劣解,从时间和费用这 2 个目标函数的角度来看,这些解都是非劣的,用户可以根据自己的实际情况来选择最适合自己的解,其他的解可以作为备选方案,在服务组合过程中发生意外时采用.

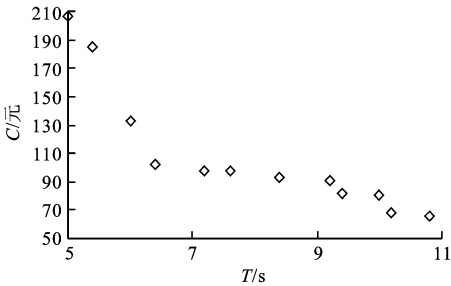


图 4 全局最优非劣解集

4 结 论

服务的全局优化是服务组合的一个重要问题,本文采用遗传算法作为一种全局最优的服务组合方法,将基于 QoS 动态组合的全局问题转化为一个带有约束条件的多目标优化问题,通过伪二叉树法则构造满足约束条件的 Pareto 最优解集合,并通过对

最优解集排序以及个体相似度的计算来获得遗传算法的适应度函数,最终产生一组满足约束条件的 Pareto 最优解服务集合. 本文通过实验验证了该方法的可行性和有效性.

参考文献:

- [1] ZHAO Zengliang, 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.
- [2] 刘书雷,刘云翔,张帆,等. 一种服务聚合中 QoS 全局最优服务动态选择算法[J]. 软件学报,2007,18(3): 656-656.
LIU Shulei, LIU Yunxiang, ZHANG Fan, et al. A dynamic Web services selection algorithm with QoS global optimal in Web services composition [J]. Journal of Software, 2007, 18(3): 656-656.
- [3] ATASHKARI K, NARIMAN-ZADEH N, PILECHI A, et al. Thermodynamic Pareto optimization of turbojet engines using multi-objective genetic algorithm [J]. International Journal of Thermal Sciences, 2005, 44(11): 1061-1071.
- [4] 胡焕耀,董渭清. 用伪二叉树法则构造多目标 Pareto 最优解集的方法[J]. 西安交通大学学报,2009,43(2):29-32.
HU Huanyao, DONG Weiqing. An approach to constructing multi-objective Pareto optimal solutions using pseudo binary tree's rule [J]. Journal of Xi'an Jiaotong University, 2009, 43(2): 29-32.
- [5] 王小平,曹立明. 遗传算法——理论应用与软件实现[M]. 西安:西安交通大学出版社,2002:74-76.
- [6] 郑金华,蒋浩,邝达,等. 用擂台赛法则构造多目标 Pareto 最优解集的方法[J]. 软件学报,2007,18(6): 1287-1297.
ZHENG Jinhua, JIANG Hao, KUANG Da, et al. An approach of constructing multi-objective Pareto optimal solutions using Arena's principle [J]. Journal of Software, 2007, 18(6): 1287-1297.

(编辑 刘杨 苗凌)