

# 基于社会网络的 Web 服务选择算法的研究

苏佰川, 张国义, 许胤龙

(中国科学技术大学 计算机科学学院, 安徽 合肥 230026)

**摘要:** 针对现有服务选择技术的不足, 提出一种应用服务社会网络(SSN(Service Social Network))机制的方法。算法主要通过社会网络原理, 将 Web 候选服务根据联系紧密程度和服务本身的 QoS 属性选择满足用户需求的组合 Web 服务, 理论分析和实验表明算法的有效性和可行性。

**关键词:** 服务选择; 社会网络; 服务质量

中图分类号: TP393

文献标识码: A

文章编号: 1674-7720(2012)06-0046-04

## Web service selection algorithm based on social network

Su Baichuan, Zhang Guoyi, Xu Yinlong

(Computer Science and Technology, University of Science & Technology of China, Hefei 230026, China)

**Abstract:** This paper proposes a new service social network algorithm (SSN Algorithm) for the optimal Web service selection problem. The algorithm applies the theory of social network based on contact intensity of the service nodes and the QoS attribute to select the Web services. Theoretical analysis and experimental results show the effectiveness and feasibility of the algorithm.

**Key words:** service selection; social network; QoS

随着面向服务计算技术的深入发展, 通过组合 Web 服务建立分布式应用系统逐渐成为基于 Web 应用开发的主流技术。在众多可获得 Web 服务资源中, 如何选择合适的服务实例进行组合, 使之既能满足应用的功能需求, 又能满足用户提出的 QoS 需求, 是一个备受关注的问题, 并已得到了广泛的研究。

QoS 约束的 Web 服务选择与组合是一类组合优化问题。到目前为止, 求解该问题比较典型的方法有线性规划算法<sup>[1-2]</sup>和遗传算法<sup>[3-6]</sup>。线性规划算法在候选服务集规模较小时是非常有效的, 但是随着规模的加大, 计算耗时增长速度明显加快, 可扩展性变差。遗传算法被用于 QoS 的全局组合优化已有许多成功的案例, 但是, 由于在诸多的研究中一般采用随机交叉和变异操作, 没有考虑候选服务实例之间实际存在的连接偏好, 组合服务的可靠性难以得到有效的保证。连接频率反映了服务之间的连接偏好, 频率越高, 下一次连接的可能性就越大, 在一定程度上可增强组合服务的可靠性。如何利用候选服务实例之间的连接偏好优化服务的随机选择, 提高服务组合的效率与成功率是本文主要关注的问题。

根据软件工程的一般常识可知, 一个复杂的组合服

务包含平行、分支、选择、循环等结构, 可以利用数据流聚合方法形成颗粒较大的结构块。而对于一个大规模的候选服务集合, 应用社会网络社区形成机制<sup>[7]</sup>, 也可以形成与服务流程聚合块功能相一致的服务子集, 即所谓的社区。这样, 服务选择就演化成社区选择问题, 运用遗传算法能较好地保证结构性选择, 且依据连接权重形成的服务社区能充分体现服务之间的连接偏好。根据这一思想, 本文提出了一个基于社区形成机制和遗传算法的服务选择方法, 与传统遗传算法不同的是, 设计了一个依据连接偏好的效用函数估算随机选择概率。初步的实验结果证明, 该方法行之有效, 且具有较高的计算效率。

### 1 问题描述

对于用户提出的服务需求首先建立组合服务流程, 图 1 所示为一个有 8 个原子服务  $S_1, S_2, \dots, S_8$  的服务流程。

如图 1 所示, 一般一个组合 Web 服务流程由多个 Web 原子服务组成, 每个 Web 原子服务又对应有多个候选服务。这些候选服务可能由不同的服务提供者提供, 具有相同的功能和不同的 QoS 度量值。

## 网络与通信 Network and Communication

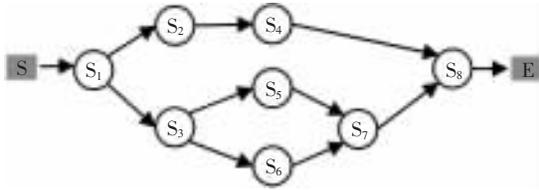


图1 组合 Web 服务的流程实例

定义(服务流程, SP)1: 对于任意一个组合 Web 服务流程 SP 可表示为  $SP=(S, \leq_s)$ , 其中

$$S=\{S_i|i=1, 2, \dots, n\} \quad (1)$$

其中  $S_i$  表示第  $i$  个原子服务,  $\leq_s$  表示各个服务  $S_i, S_j$  之间的内部关系, 可表示为:  $\leq_s=(Pre, Suc, R)$ , 其中 Pre 和 Suc 分别表示当前服务在服务流程中的前驱和后继。同时如果某个  $S_i.Pre=\wedge$ , 则此结点为开始结点, 如果某个  $S_i.Suc=\wedge$ , 则此结点为结束结点。  $R(0 < R \leq 1)$  表示当前原子服务在服务流程中的结构关系,  $R=1$  表示此服务在服务流程中是串行、并行或者循环结构,  $0 < R < 1$  则表示此服务属于服务流程选择结构的一个分支。

假定每个原子服务  $S_i$  包含  $m$  个候选服务, 即:

$$S_i=\{S_i^j \quad j=1, 2, \dots, m\} \quad (2)$$

对于每个候选服务  $S_i^j$  都有不同的服务质量(QoS)<sup>[3,8]</sup>, 如响应时间(Response time)、代价(Cost)、可靠性(Reliability)、信誉度(Reputation)等, 假定每个候选服务  $S_i^j$  的服务质量(QoS)维度为  $k$ , 即:

$$QoS=\{q_1, q_2, \dots, q_k\} \quad (3)$$

在此, 以  $T_i^j, C_i^j$  和  $R_i^j$  分别表示候选服务  $S_i^j$  的响应时间、代价和可靠性。

定义(服务选择实例, WS)2:  $WS=(S_1^i, S_2^j, \dots, S_n^k), i, j, k \in (1, 2, \dots, m)$ , 每个 WS 是满足用户 QoS 约束条件的前提下, 目标实现最优化的一组候选服务的集合, 具体目标函数及约束条件定义如下:

$$\text{目标函数: } F=\text{Min}(T) \quad (4)$$

$$\text{满足约束: } T \leq T_{QoS}, C \leq C_{QoS}, R \leq R_{QoS} \quad (5)$$

其中  $T, C, R$  为某一服务选择实例(WS)完成组合服务要经历的总的响应时间、总代价及可靠性。而  $T_{QoS}, C_{QoS}$  及  $R_{QoS}$  为用户提出的约束条件。组合 Web 服务流程<sup>[3]</sup>的 4 种基本结构——串行、并行、选择和循环中 QoS 度量值  $(T, C, R)$  计算公式在文中已给出, 此处不再叙述<sup>[8]</sup>。

## 2 算法描述

### 2.1 服务社会网络

要应用社会网络<sup>[9-10]</sup>, 首先要将候选服务结点集合转换为以网络为主的服务社会网络, 因此要在结点之间加入边及其边的权值的确定。

定义(服务结点, Node)3: 对于组合 Web 服务选择问题中的每个候选服务  $S_i^j$ , 都称之为一个候选服务结点(Node), 每个 Node 除了候选服务本身的 QoS 之外, 另外

加入边的权值 (Weight) 属性, 具体数据结构表示为 Node (Name, Type, Weight, QoS)。为描述方便, 将 Node.Name 由之前的  $S_i^j$  转化为  $N_k(k=1, 2, \dots, n \times m)$ , 每个服务结点  $N_i$  的服务类型  $N_i.Type=S_j(j=1, 2, \dots, n)$ , 同时  $N_i.Weight$  是一个三元组, 表示为  $(Name, Type, w)$ 。

设  $Edge(S_i, S_j)$  为如果在服务流程 S 中  $S_i$  到  $S_j$  存在边相连, 则  $Edge(S_i, S_j)$  为 1, 否则为 0。

定义(服务社会网络, SSN)4: 对于特定的服务流程 S, 用户每提出一个 QoS 需求, 都会对应一个 WS 满足用户的功能需求及 QoS 需求, 对每个满足需求的 WS 作如下判断:

$$(\forall (S_i^j, S_k^l) \in WS) \wedge (Edge(S_i, S_k)=1) \rightarrow \begin{cases} N_x.Weight.w+1 & \text{if } (\exists (N_x.Weight.Name=N_y)) \\ N_x.Weight(N_y, S_k, 1) & \text{else} \end{cases} \quad (6)$$

其中候选服务  $S_i^j, S_k^l$  分别对应于候选服务结点  $N_x, N_y$ 。

通过以上的定义, 将之前的服务结点之间加入边及权值, 形成服务社会网络, 如图 2 所示。

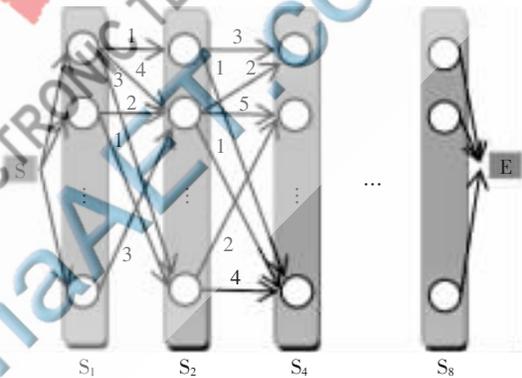


图2 服务社会网络

### 2.2 服务社会网络选择算法

Web 服务的部分 QoS 是负相关的, 即取值越小越好 (如响应时间、代价), 而同时部分 QoS 则是正相关的, 即取值越大越好 (如可靠性等)。因此对于负相关的 QoS 属性和服务结点之间的权值归一化:

定义(归一化, Normalized)5:

$$w_{ij} = \frac{w_{ij}}{\sum_{(N_i \in N, Weight.Name) \wedge (N_i.Type=N_j.Type)} w_{ik}} \quad (7)$$

$$T_i^j = \begin{cases} \frac{T_{\max} - T_i}{T_{\max} - T_{\min}} & \text{if } (T_{\max} - T_{\min} \neq 0) \\ 1 & \text{if } (T_{\max} - T_{\min} = 0) \end{cases} \quad (8)$$

$$C_i^j = \begin{cases} \frac{C_{\max} - C_i}{C_{\max} - C_{\min}} & \text{if } (C_{\max} - C_{\min} \neq 0) \\ 1 & \text{if } (C_{\max} - C_{\min} = 0) \end{cases} \quad (9)$$

结点的效用函数  $F_i$  的定义由 QoS 参数组成, 其中的 QoS 参数是与用户提出的目标函数中的 QoS 参数相对应的, 由式(4)可知目标函数使整个组合 Web 服务的执行时间最小, 故效用函数的定义如定义 6 所示。

# 网络与通信

Network and Communication

定义(效用函数,  $F_i$ ):  $F_i = T_i$

服务结点之间的权值归一化后, 结点  $N_i$  与  $N_j$  之间的归一化权值  $w_{ij}$  与  $w_{ji}$  不相同, 因此结点间的选择概率定义如下:

定义(选择概率,  $P_{ij}$ ):

$$P_{ij} = v_1 w_{ij} + v_2 w_{ji} + v_3 F_j \quad v_1 + v_2 + v_3 = 1$$

$P_{ij}$  越大, 表明结点  $N_i$  与  $N_j$  被选择在一起用于解决问题的概率越大。

给出  $P_{ij}$  之后, 同时结合 Web 服务选择问题和遗传算法, 提出了如下的服务社会网络算法(SSN Algorithm):

Begin

初始化 TBest= $\infty$ , CEnd=0, REnd=0;

While(循环条件为真)

    随机形成一个服务选择实例 WS

    计算整个流程的 T, C, R;

    If(TBest>T && C<C<sub>QoS</sub> && R>R<sub>QoS</sub>)

        TBest=T, CEnd=C, REnd=R;

    End if

    While(结束条件为假)

        计算 WS 各结点与其前驱后继的平均选择概率  $P$ ;

        在 WS 中选取选择概率最小的变异结点  $N_i$ ;

        在服务社会网络中随机选择与变异结点  $N_i$  有相同功能的 Web 服务  $N_j$ ;

        将  $N_i$  替换为  $N_j$  形成新的服务实例 WS';

        计算当前实例的 T, C, R;

        if(TBest>T && C<C<sub>QoS</sub> && R>R<sub>QoS</sub>)

            TBest=T, CEnd=C, REnd=R;

        WS=WS';

        End if

    End while

End while

End

某一结点  $N_i$  在某一服务实例 WS 中关于前驱后继的平均选择概率  $P$  的计算公式如下(前驱后继可能不止一个, Num 函数用于计算某一原子服务  $S_i$  在某一服务流程中前驱与后继的个数):

$$P = \frac{\sum_{i, Pre \in WS} P_{i, Pre} + \sum_{i, Suc \in WS} P_{i, Suc}}{2 \cdot \text{Num}(N_i, \text{Type}.Pre) + \text{Num}(N_i, \text{Type}.Suc)} \quad (10)$$

## 3 实验分析

实验环境为 100 Mb/s 的局域网, 微机的配置为 Intel Pentium Dual 2.0 GHz CPU, 2 GB 内存, 操作系统为 Windows XP, 算法编译环境为 VC++ 6.0。实验分为 4 组, 每组的服务类型个数及候选服务个数如表 1(其中  $n$  表示原子服务个数,  $m$  表示每种原子服务对应的候选服务个数)所示, 分别从以下几个方面进行实验。

(1) 服务结点规模不同时对选择算法执行时间的

影响。

选择算法中的外层循环设置为 1 000 次, 内层循环条件为当 TBest 有连续 50 次不发生变化或者循环满 3 000 次时结束。算法在 4 个 group 下的执行时间如图 3 所示, 由图可知, 当整个社会网络中总结点个数为 3 000 时, 所用的时间刚刚 1 s, 而且随着网络规模的增加, 执行时间的增长速率要缓慢得多。因此说网络规模越大, SSN 算法的执行效率越高。

表 1 各组原子服务类型规模及候选服务规模

Group	$n$	$m$
group1	5	10
group2	10	20
group3	20	50
group4	30	100

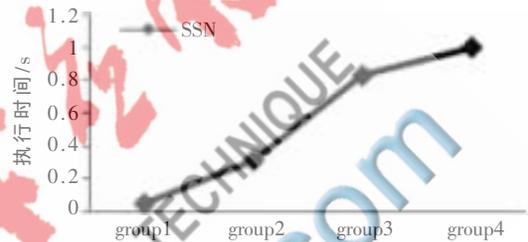


图 3 不同 group 下算法执行时间

(2) 随着社会网络权值的不同找到最优解的概率

分析在不同权值下各个 group 运行 SSN 算法 100 次, 能找到的目标函数的较优值与最优解之间的比值, 通过图 4 可以看到随着网络中权值的增大, 4 个 group 找到的较优解与最优解之间的差距越来越小。当网络中结点的规模不是很大(group1, group2)时, 即使权值不是很大, 找到最优解概率也非常高。随着权值的增加, 能找到最优解的次数也不断增多, 当用户提出次数到达 40 次左右时, group1 与 group2 基本每次都可以找到最优解。当网络规模较大(group3, group4)的情况下, 虽然权值很小时找到最优解的概率不是很高, 大概 85%, 但随着权值的增加, 能找到最优解的概率迅速加大, 当权值为 30 时, 概率达到了 96% 左右, 权值为 50 时, 概率达到了 98%, 因此随着用户 QoS 需求的增加, 能找到最优解的概率能够迅速加大, 同时对于结点规模很大的社会网络

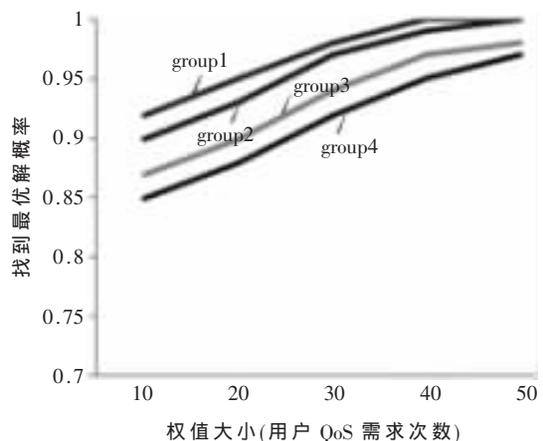


图 4 各个 group 下随着权值增大找到最优解的概率

来说,效果更佳。

随着服务计算的迅速发展,可利用的 Web 服务资源越来越多,候选服务结点的规模也越来越大。本文总结了近期相关的工作,提出了一种基于社会网络社区形成机制的 Web 服务选择方法。将 Web 服务选择问题转换到社会网络中,利用社会网络特有的性质结点联系强度等来解决服务选择问题。提出了一种服务社会网络概率选择方法,此方法与遗传算法的不同之处在于变异结点的选择不是随机的,而是根据结点之间的连接频率,它反映了服务之间的连接偏好,频率越高,下一次连接的可能性就越大。

虽然最终算法给出了很好的性能,但还有待进一步改进,如本文算法是基于单目标优化的,希望今后在多目标优化方面有所进展。同时本文算法是基于历史信息(服务实例)的,因此当历史信息较少或刚开始没有历史信息时性能方面可能会很差,希望今后在这方面也有所研究。

#### 参考文献

- [1] Zeng Liangzhao, BENATALLAH B, DUMAS M, et al. Quality driven Web services composition[C]. Proceedings of the 12th International Conference on World Wide, 2003.
- [2] ARDAGNA D, PERNICI B. Adaptive service composition in flexible processes[J]. IEEE Trans. Software Eng., 2007, 33(6): 369-384.
- [3] HWANG S Y, WANG H, TANG J, et al. A probabilistic approach to modeling and estimating the QoS of Web-Services-based Workflows[J]. Information Sciences, 2007, 23(177): 5484-5503.
- [4] KLEIN A, ISHIKAWA F, BAUER B. A probabilistic approach

to service selection with conditional contracts and usage patterns[C]. ICSOC-ServiceWave2009, 2009: 253-268.

- [5] Tang Maolin, Ai Lifeng. A hybrid genetic algorithm for the optimal constrained Web service selection problem in Web Service composition[C]. WCCI 2010 IEEE World Congress on Computational Intelligence, 2010.
- [6] Liu Shulei, Liu Yunxiang, Jing Ning, et al. A dynamic Web service selection strategy with QoS global optimization based on multi-objective genetic algorithm[C]. GCC2005, 2005: 84-89.
- [7] KUMPULA J M, ONNELA J P, SARAMAKI J, et al. Model of community emergence in weighted social networks[J]. Computer Physics Communication, 2009(180): 517.
- [8] YU T, ZHANG Y, LIN K J. Efficient algorithms for Web services selection with end-to-end qos constraints[J]. ACM Transactions on the Web, 2007, 1(1): 1-26.
- [9] NEWMAN M E J. Analysis of weighted networks[J]. Physical review, 2004, 70(05): 1-9.
- [10] WASSERMAN S, FAUST K. Introduction to social network analysis[M]. Cambridge Cambridge Univ. Press, 1994.

(收稿日期: 2011-11-22)

#### 作者简介:

苏佰川,男,1987年生,硕士研究生,主要研究方向:服务计算。

张国义,男,1965年生,讲师,主要研究方向:服务计算,云计算。

许胤龙,男,1963年生,教授,博士生导师,主要研究方向:网络路由算法与协议,网络编码,并行算法等。