

结合语义与模糊推理的两阶段服务组合

王 雷, 齐美石

(中国科学技术大学 自动化系, 安徽 合肥 230027)

摘 要: 如何选择 Web 服务使组合服务的功能性属性与非功能性属性都能满足实际需求, 是当前服务组合领域研究的一个重要问题. 通过引入语义和服务质量, 给出服务组合的形式化分析方法. 该方法采用模糊 Petri 网建模, 利用模糊 Petri 网的模糊推理能力, 在语义层次上对服务组合进行功能性匹配. 提出服务的服务质量 QoS 匹配度模型, 对组合服务进行服务质量匹配. 仿真结果表明, 该方法可以解决综合考虑语义和服务质量的组合问题.

关键词: 语义 Web 服务; 模糊 Petri 网; 服务组合; 形式化方法; 服务质量

中图分类号: TP31

文献标识码: A

文章编号: 1000-7180(2012)01-0001-05

Combination of Semantic and Fuzzy Reasoning Two-stage Service Composition

WANG Lei, QI Mei-shi

(Department of Automation, University of Science and Technology of China, Hefei 230027, China)

Abstract: Choose services to satisfy the demand of function and Nonfunctional property of a service composition is a very important question in the domain of web service composition. Give a formalization method by using semantic and service quality to solve this problem. Match the composition's function property in semantic using Fuzzy Petri Net. Match the composition's QoS property using a Service nonfunctional match model. The simulation experiment show that the method can resolve the problem of service composition considering its semantic property and QoS.

Key words: semantic web service; fuzzy petri net; web service composition; formalization method; QoS

1 引言

Web 服务被定义为能够通过网络描述、发布、定位、调用的自描述的模块化程序^[1], 是 SOA 架构的实施基础. 随着 Web 服务技术的日益成熟, 越来越多的稳定易用的 Web 服务共享在网络上. 然而单个的 Web 服务往往比较简单且功能有限, 必须把多个 Web 服务组合起来形成组合服务, 才能满足用户需求. 当前的服务组合领域存在如下问题:

(1) 当前关于语义 Web 服务的研究主要把服务之间的语义匹配关系划分为四个等级^[2], 这种定性的研究方法只能用于基于语义的服务发现, 不能在组合的层次上对组合服务的整体进行语义分析.

(2) 目前将语义和服务质量相结合进行服务组合的研究较少.

Petri 网具有直观的图形表示、形式化的语义、状态和事件显示表示、丰富的分析技术等优势, 成为目前服务组合验证中最主要的工具之一. 模糊 Petri 网则是在普通 Petri 网的基础上, 引入模糊产生式规则, 可以更好地对具有模糊性的系统行为进行推理仿真.

本文提出 Web 服务的语义匹配度模型和服务质量匹配度模型量化研究 Web 服务的语义匹配和服务质量匹配. 使用模糊 Petri 网作为建模与仿真工具, 以模糊推理算法为基础提出两阶段服务组合算法, 得到了服务质量匹配度和语义匹配度都较高的服务组合路径.

收稿日期: 2011-05-12; **修回日期:** 2011-06-10

基金项目: 国家“八六三”计划项目(2008AA01A317)

2 模糊 Petri 网与匹配度模型

2.1 匹配度模型

为定量研究 Web 服务的语义匹配和服务质量匹配,提出以下模型:

2.1.1 Web 服务语义匹配度模型

Web 服务的语义信息可以划分为四类:服务的基本属性、服务的联系属性、服务的功能属性和服务的情境属性^[3]. 根据 Web 服务模型中的语义信息,构建本体树,根据词汇语义相似度,计算服务之间的语义匹配度. 语义相似度计算公式为^[4]:

$$Sim(o_1, o_2) = \frac{\alpha \times (l_1, l_2)}{(Dis(o_1, o_2) + \alpha) \times \max(|l_1 - l_2|, 1)}$$

其中, o_1, o_2 是两个服务的本体, l_1, l_2 是 o_1, o_2 分别所处的层次, α 是相似度为 0.5 时 o_1, o_2 之间的距离, 是一个可调节的参数.

2.1.2 Web 服务质量匹配度模型

Web 服务的服务质量包括服务的执行时间、代价等多种因素, 且与服务所在节点的性能(CPU、带宽等)、负载情况有关. 为简化服务质量匹配度模型, 使用资源数的概念来抽象服务质量, 资源数的计算可视侧重点不同而定, 如可以定义为 Web 服务所需要的带宽大小和 CPU 消耗百分比的加权平均.

在实际服务组合过程中, 对于目标组合服务中的每个目标服务 WS_i , 都有一系列备选的服务 $WS_{i,j}, j=1, 2, \dots, n$, 分别位于 n 台服务器上, 为便于比较这些服务的服务质量差别, 文中采用隶属度函数, 将各个服务器的资源数、Web 服务所消耗的资源数映射到 0~1 之间, 于是得到归一化后的资源数.

定义 Web 服务所需要消耗的资源数为 λ' , 服务器节点所具有的资源数为 λ'' . Web 服务的服务质量与 λ', λ'' 都相关, 则可定义服务质量匹配度公式为:

$$\lambda = 1 - \lambda' / \lambda''$$

其中, λ' / λ'' 表征该服务所消耗的资源占节点所有资源的百分比, 所以 λ 值越大, 表明该服务相对于该节点的服务质量越好.

服务器节点的资源数不是固定不变的, 随着组合过程的进行, 被使用的节点会随着组合服务的访问而发生变化, 需要进行相应的更新. 使用以下节点性能更新公式对组合服务的资源数变化进行估计:

$$\lambda'' = \lambda'' - \lambda' / \phi$$

其中, ϕ 为访问频繁因子, 用于表征该组合服务在一

定时间内被访问的频率, λ' 为该服务器节点上被该组合服务使用的服务需要消耗的资源数.

2.2 模糊 Petri 网

根据 Web 服务的语义匹配度模型和服务质量匹配度模型, 可以对单个 Web 服务的语义匹配、服务质量匹配进行定量的分析. 但对于组合服务, 需要根据组合服务中单个 Web 服务的匹配度推理出组合服务整体的匹配度, 才可以对不同的组合路径进行比较, 为此考虑使用模糊 Petri 作为服务组合的建模工具.

模糊 Petri 网是对经典 Petri 网的扩展, 引入了命题集合、置信度及模糊推理规则. 命题与库所一一对应, 命题对应一个 $[0, 1]$ 之间的数, 表示该命题的真实度, 1 表示该命题成立, 0 表示该命题不成立, 命题的真实度可以理解为该命题成立的概率; 用置信度来表示变迁对命题真实度的影响程度; 用模糊推理规则来判定随着变迁的发生, 相关命题真实度的变化情况.

定义 1 (模糊 Petri 网) 是对普通 Petri 网的扩展. 定义为一个八元组^[5]:

$$FOB = (P, T, D, F, I, O, f, \alpha, \beta)$$

(1) P : 模糊库所的有限集合;

(2) T : 模糊变迁的有限集合;

(3) D : 命题的有限集合;

(4) F : 弧的集合;

(5) I/O : 分别为表示变迁的输入矩阵和输出矩阵. $I = [\alpha_{ij}]$, $\alpha_{ij} \in \{0, 1\}$, 当 $(P_i, T_j) \in F$ 时, $\alpha_{ij} = 1$, 否则 $\alpha_{ij} = 0$; $O = [\beta_{ij}]$, $\beta_{ij} \in \{0, 1\}$, 当 $(T_j, P_i) \in F$, $\beta_{ij} = 1$, 否则, $\beta_{ij} = 0$;

(6) f : 变迁的阈值或置信度 (Confidence Factors);

(7) α : 命题到一个 $[0, 1]$ 之间的数值;

(8) β : 库所与命题之间的映射.

模糊 Petri 中一个变迁及其前后的库所集合模拟了一条模糊推理规则, 变迁前的库所集合 (与命题一一对应) 代表前提条件, 变迁后的库所集合代表推理的结果, 变迁的置信度表示前提条件对该推理的支持度. 在模糊 Petri 网中, 共有三类基本的模糊推理规则:

类型 1: IF d_{i1} And d_{i2} And $\dots$ d_{in} THEN d_k ($CF = \mu_j$), d_i ($1 \leq i \leq n$) $\in D$, 它们对应的真实度为 y_{i1} ($1 \leq i \leq n$), 推理后, 命题 d_k 的真实度为 $\min(y_{i1}, y_{i2}, \dots, y_{in}) \times \mu_j$.

类型 2: IF d_i THEN d_{k1} And d_{k2} And $\dots$ d_{kn}

$(CF=\mu_j), d_{kl} (1 \leq l \leq n) \in D$, 命题 d_i 的真实度为 y_i , 推理后命题 d_{kl} 的真实度 $y_{kl} = y_i \times \mu_i$.

类型 3: IF d_{i1} OR d_{i2} OR $\dots$ d_{in} THEN d_k ($CF = \mu_j$), $d_{kl} (1 \leq l \leq n) \in D$, 命题 d_i 的真实度为 y_i , 推理后命题 d_k 的真实度 $y_k = \max(y_{i1} \times \mu_i, y_{i2} \times \mu_i, \dots, y_{in} \times \mu_i)$.

使用模糊 Petri 网建模 Web 服务组合, 用变迁表示一个 Web 服务, 库所表示变迁发生(即服务选择或调用)前后的状态, 变迁的置信度表示该服务组合路径中的 Web 服务与目标 Web 服务的语义匹配度或服务质量匹配度, 命题即库所的真实度, 表示该组合路径在语义层次上或服务质量层次上与目标组合服务的匹配度. 那么就可以使用模糊 Petri 网及其推理规则对组合服务的匹配度进行推理, 从而选择匹配度较高的服务组合路径.

3 基于模糊推理的两阶段服务组合

本文提出的两阶段服务组合以模糊推理算法为基础, 根据 Web 服务的语义匹配度和服务质量匹配度, 推理出组合服务的语义和服务质量综合匹配度, 从所有可能的组合路径中得到语义匹配度和服务质量匹配度都较高的服务组合路径.

3.1 模糊推理算法

模糊推理算法作为模糊 Petri 网最基本的推理算法, 大体可划分为两类: 正向模糊推理算法和逆向模糊推理算法, 两者都是根据模糊推理规则对系统进行推理, 区别是逆向模糊推理从最终目标库所向起始库所进行逆向推理, 可以减小问题规模, 提高算法效率, 本文采用逆向模糊推理算法. 为介绍逆向模糊推理算法, 首先定义模糊推理运算符.

定义 2 模糊推理运算符^[6]

(1) $\oplus: X \oplus Y = Z$, X, Y 和 Z 都是 $m \times n$ 维矩阵, $z_{ij} = \max(x_{ij}, y_{ij})$;

(2) $\otimes: X \otimes Y = Z$, X, Y 和 Z 都是 $m \times s, s \times n, m \times n$ 维矩阵, $z_{ij} = \max(x_{ij} \times y_{ij}), 1 \leq k \leq s$;

(3) $\odot: X \odot Y = Z$, X, Y 和 Z 都是 $m \times n$ 维矩阵, $z_{ij} = \min(x_{ij}, y_{ij})$;

(4) $\ominus: X \ominus n = Z$, X, Z 都是 $m \times n$ 维矩阵, n 是一个数字, $z_{ij} = 1$ IF $x_{ij} \times n = 1$; $z_{ij} = 0$ IF $x_{ij} \times n \neq 1$.

$\Theta: X \Theta n = Z$, X, Z 分别是 $m \times n$ 维矩阵, n 是一个数字, $z_{ij} = 1$ IF $x_{ij} \times n = 1$; $z_{ij} = 0$ IF $x_{ij} \times n \neq 1$.

逆向模糊推理算法由三个阶段组成: 第一阶段将 FPN 结构用代数形式表示, 第二阶段完成逆向推

向推理, 简化 FPN 网结构. 第三阶段计算目标状态匹配度. 具体算法如下:

Input: FPN 结构, 目标库所的名字.

Output: 目标库所的匹配度.

(1) 将 FPN 结构用矩阵形式表示.

Δ 为输入矩阵, T 为输出矩阵; 向量 $S_0 = [\alpha(p_1), \alpha(p_2), \dots, \alpha(p_n)]^T$ 为命题的真实度; 向量 G 表示需要计算的目标库所, $G = \{x_i\}, x_i = 1$, 如果 p_i 是结果命题, 否则, $x_i = 0$. 向量 U 表示变迁的置信度向量: $U = [f(t_1), f(t_2), \dots, f(t_n)]^T$.

(2) 逆向推理, 简化输入输出矩阵.

$T' = T^T \otimes G; z = 0; T_x = T'; P_x = G;$

$T = (O_1, O_2, \dots, O_n)^T$

While($T' \neq T'$) { $z = z + 1; T = T'; P' = \Delta \otimes T; P_x = P' \times z + P_x; T' = T^T \otimes P'; T_x = T_x + T' \}$

$z = z - 1;$

去掉 Δ 和 T 的 T_x, P_x 对应的行列中为零的项, 得到两个新的矩阵 Δ' 和 T' , 去掉 S_0 对应 P_x 中为零的项, 得到向量 Q_0 , 去掉 P_x 中为零的项, 得到向量 P'_x , 去掉 U 中对应 T_x 为零的项, 得到 U' ;

$\Delta x = \Delta'^T \times (M \times P'_x);$

(3) 计算目标库所的匹配度: M 为一个大小为 $|U'|$ 的单位矩阵.

for($k = 0; k < z; k++$) { $Q_{k+1} = Q_k \oplus [(T' \times M \times U') \otimes (\Delta x \ominus \frac{1}{(z-k)} \odot Q_k)] \}$

3.2 两阶段服务组合

为对服务组合的语义及服务质量进行匹配, 提出两阶段匹配算法. 第一阶段离线规划, 产生候选组合路径, 即根据 Web 服务语义匹配度模型, 使用模糊推理算法计算组合路径的语义匹配度, 筛选出语义匹配度大于某一阈值的组合路径; 第二阶段在线组合, 即实施动态组合, 根据 Web 服务服务质量匹配度模型, 使用模糊推理算法计算组合路径的服务质量匹配度, 从第一阶段得到的候选路径中确定服务质量较优的最终组合路径. 两阶段都使用逆向模糊推理算法作为基本的推理算法. 两阶段服务组合策略的具体步骤如下:

(1) 初始化: 归一化各个服务器资源数 λ'' 和各 Web 服务所需消耗的资源数 λ' 、初始化语义匹配度阈值、目标服务组合路径数 n .

(2) 离线规划

① 构建本体树, 计算语义本体相似度, 得到单个服务的语义匹配度;

② 由模糊推理算法和模糊产生式规则,得到所有可能的组合服务路径的语义匹配度;

③ 选取语义匹配度大于语义匹配度阈值的组合服务路径,得到简化的问题解空间。

(3) 在线组合:

① 由 Web 服务质量匹配度模型,确定各个服务的服务质量匹配度;

② 将各服务的语义匹配度和服务质量匹配度加权平均,等到 Web 服务的综合匹配度,再使用模糊推理算法得到包含服务质量指标的综合匹配度,从 2 中的服务组合路径中选择综合匹配度最高的一条服务组合路径;

③ 根据 2.2 中的服务器资源数更新公式更新被选中的服务器资源数,重复步骤②、③,直到选出 n 条服务组合路径;

算法描述如图 1 所示。

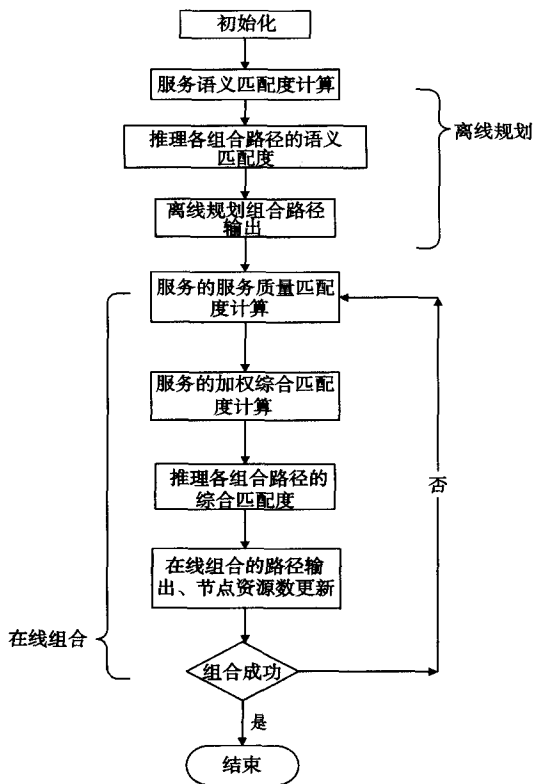


图 1 两阶段服务组合算法

4 实验结果与分析

4.1 实验

使用 PIPE^[7] 作为仿真建模工具,进行仿真实验。

实验规模:模拟 40 个节点,分为 4 组,每组 10 个。第一组节点部署服务 A、M、X,第二组节点部署

服务 B、N、Y,第三组节点部署服务 C、O、Z,第四组节点部署服务 D、P、W。组合服务由四个简单服务组成,这四个服务由这四组节点提供,仿真实验如图 2 所示。

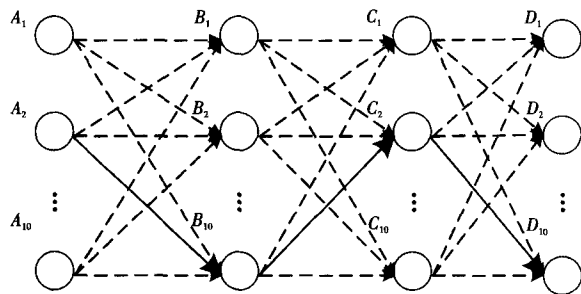


图 2 寻找较优的路径

初始值:为简化问题,不考虑节点异构性,假定每个节点在组合前的资源数相同,量化为 1,语义匹配度阈值为 0.6;每个服务消耗的资源数随机产生。每个服务与实际组合所需服务的语义匹配度随机产生。需要产生的组合路径数为 5。

实验过程:由上述服务中任意选取四个组成顺序流程。根据两阶段匹配算法,第一步进行语义匹配,化简问题解空间;第二步,根据服务质量和语义平均加权后的综合匹配度,选取最优的一条路径,更新节点资源数信息。重复组合多次,直到得到 5 条路径,为最终的服务组合路径。

4.2 结果与分析

仿真实验中,分别得到了组合服务 ABCD, MNOP, XYZW, AN CZ, BMNO 的 5 条综合匹配组合路径和整个系统中节点的使用情况。为了进行比较,从离线规划阶段的输出结果中为每个组合服务选取了仅考虑语义匹配度而得到的 5 条较优的组合路径,并记录了此时整个系统中节点的使用情况。与综合匹配的组合结果比较,只进行语义匹配时得到的组合路径,其使用的节点重合度比语义与服务质量综合匹配时得到的组合路径使用的节点重合度高,即语义匹配时每种组合服务得到的 5 条组合路径集中在几乎相同的四至五个节点之上。而综合匹配时,5 条组合路径使用的节点在网络中分布均匀。通过对整个网络中各节点的使用情况进行分析,得到实验数据如图 3 所示。

图中横坐标为 40 个节点的 ID,纵坐标为各节点用于组合的次数。蓝线为仅做语义匹配时节点使用次数,红线为语义与服务质量综合匹配时节点使用次数。由图可以看出,仅做语义匹配节点的选择集

中,个别节点的使用频率较高,这些节点会因为负载过重,成为该组合服务的性能瓶颈,使整个组合的性能下降.而做综合匹配时网络中节点的使用较为均衡,整个网络具有一定的负载均衡能力.

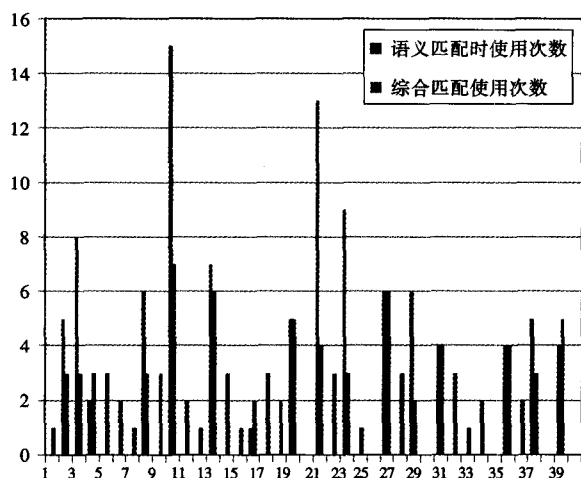


图3 系统中节点使用次数

5 结束语

本文提出一种综合考虑语义和服务质量的两阶段匹配算法,对组合服务的语义匹配和服务质量匹配进行了定量分析.基于模糊 Petri 网及其模糊推理算法,将服务组合分为两个阶段:离线规划阶段,结合语义对组合服务的功能属性进行匹配;在线组合阶段,对组合服务的功能属性和非功能属性进行综合匹配.

仿真结果表明此方法可以得到语义和服务质量都较高的服务组合路径.综合匹配与仅使用语义匹配相比较,系统具有一定的负载均衡能力.将来的工作是根据组合服务的访问频率、Web 服务各服务质量指标、服务器的性能细化服务质量匹配度模型,用于实际应用环境.

参考文献:

- [1] Gottschalk K. Web services architecture overview: the next stage of evolution for E-business [EB/OL]. [2011-05-15]. <http://www-128.ibm.com/developer-workks/>
- [2] Matthias Klusch, Benedikt Fries, Katia Sycara. OWLS-MX: a hybrid semantic web service matchmaker for OWL-S services [J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2009, 7 (2): 121-133.
- [3] 伍伟,李兵,曾诚.基于本体的语义 Web 服务注册平台研究[J].微电子学与计算机,2008,25(10):4-6.
- [4] 吴健,吴朝晖,李莹,等.基于本体论和词汇语义相似度的 Web 服务发现 [J]. 计算机学报,2005,28(4): 1593-1899.
- [5] 李智生,李俊山,耿志.模糊 Petri 网的推理式算法在雷达电子战装备效能评估中的应用 [J]. 微电子学与计算机,2007,24(5):83-86.
- [6] MenPeng, Duan Zhang-hua. Backward reasoning algorithm for petri nets based on algebra[J]. Journal of System Simulation, 2007(19):161-163.
- [7] Bonet P, Catalina M Llado, Ramon P, et al. PIPE v2.5: a petri nNet tool for performance modeling [C] // Proc. 23 rd Latin American conference on Informatics(CLEI 2007). Costa Rica, 2007.

作者简介:



王 雷 男,(1972—),博士,副主任,副教授.研究方向为网络多媒体、复杂系统建模、仿真、优化控制.

齐美石 男,(1985—),硕士研究生.研究方向为网络多媒体.