



(12) 发明专利

(10) 授权公告号 CN 108804788 B

(45) 授权公告日 2022.06.17

(21) 申请号 201810528122.6

(22) 申请日 2018.05.25

(65) 同一申请的已公布的文献号
申请公布号 CN 108804788 A

(43) 申请公布日 2018.11.13

(73) 专利权人 浙江工业大学
地址 310014 浙江省杭州市下城区朝晖六
区潮王路18号

(72) 发明人 陆佳炜 周焕 高燕煦 吴涵
张元鸣 高飞 肖刚

(74) 专利代理机构 杭州斯可睿专利事务所有限
公司 33241
专利代理师 王利强

(51) Int. Cl.

G06F 30/20 (2020.01)

(56) 对比文件

CN 107135092 A, 2017.09.05

US 2013060775 A1, 2013.03.07

CN 106156620 A, 2016.11.23

WO 2016154460 A1, 2016.09.29

CA 2953385 A1, 2016.01.07

US 2004128669 A1, 2004.07.01

J. Cheng ET AL. Automatic composition of semantic web services based on fuzzy predicate petri nets.《IEEE Transactions on Automation Science and Engineering》.2015,第680-689页.

审查员 冯敏科

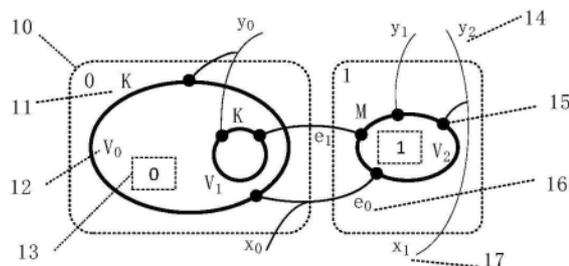
权利要求书6页 说明书17页 附图4页

(54) 发明名称

一种基于数据细胞模型的Web服务演化方法

(57) 摘要

一种基于数据细胞模型的服务演化方法,包括以下步骤:第一步、结合Bigraph理论,建立一个服务演化框架,分为信息抽取和细胞建模及服务行为演化分析两阶段;第二步、基于所述的框架提出的基于数据细胞模型的服务演化方法,过程如下:步骤2.1:Web服务的信息抽取;步骤2.2:细胞建模;步骤2.3:服务行为演化:结合服务与生物细胞特性构建DCEM,随着用户需求与资源变更的变化,数据细胞发现现有的内部服务结构已经不能提供高质量的服务能力,需要进行相应的行为演化,使之在需求频繁变更和资源动态更新的情况下也能够高效的提供服务。本发明提高系统的灵活性,使系统能够适应复杂的分析应用。



1. 一种基于数据细胞模型的Web服务演化方法,其特征在于,所述服务演化方法包括以下步骤:

第一步、结合Bigraph理论,建立一个服务演化框架,分为信息抽取和细胞建模及服务行为演化分析两阶段;

1.1、信息抽取和细胞建模:提取服务的相关描述信息,得到服务描述元数据,基于提取服务的相关描述信息,根据服务的语法信息匹配服务操作参数与资源状态,基于Bigraph理论构建数据细胞和数据细胞簇的结构及生物特性,代表服务的状态信息与服务间潜在的调用关系;

1.2、服务行为演化:将数据细胞和数据细胞簇作为服务的演化行为方式的载体,选取数据细胞和细胞簇,按照吸收消化、进化生物细胞行为定义服务的组合生长规则,模拟服务生成组合服务的行为模式;构建基于数据细胞的服务演化模型;

第二步、基于所述的框架执行服务组合方法,过程如下:

步骤2.1:Web服务的信息抽取

在Web服务的信息抽取阶段,将服务的名称映射为 s ,每个服务对应一个控制 C ,唯一,组成规则为: $C=s:control$;将服务的可用性度量映射为 U ,将服务的端口的ID、名称和类型分别映射为 pI 、 pN 和 pT ,另外将端口的控制类型映射为 pC ,代表此端口是输入、输出还是输入/输出;将此服务和其他服务的依赖状态映射为 CL ,组成规则为: $CL=\langle DL, CN \rangle$,其中, DL 是当前服务的依赖层次, CN 则是与之依赖的另一个服务的控制名称;

步骤2.2:细胞建模

对于服务的形式化建模需要根据映射规则,将服务的结构和消息交互行为加以映射到对应形式化结构和动作,从而通过形式化的方法对服务及其组合进行进程演算;

步骤2.3:服务行为演化

服务组合以服务间的按需聚合为运行模式,以动态适应、按需演化作为其运行特征,在不可预测的用户需求驱动下,各协同服务形成错综复杂的交互关系,各服务进程需要与其他服务以一系列操作序列形式进行会话,结合服务与生物细胞特性构建数据细胞演化模型DCEM,随着用户需求与资源变更的变化,数据细胞发现现有的内部服务结构已经不能提供高质量的服务能力,需要进行相应的行为演化,使之在需求频繁变更和资源动态更新的情况下也能够高效的提供服务。

2. 如权利要求1所述的一种基于数据细胞模型的Web服务演化方法,其特征在于,所述步骤2.2中,给出的形式定义如下:

定义1. 数据细胞的Bigraph定义是一个五元组 $DC=\langle S, E, Ctrl, C^P, C^L \rangle: \langle m, X \rangle \rightarrow \langle n, Y \rangle$,其中,

(1) S 是数据细胞内包含的有限的服务集, $\forall s \in S$ 称为一个数据服务;对于数据细胞 DC_i ,其内最多有 N_j 个数据服务;

(2) E 是包含服务集的有限边集合, $\forall e \in E$ 称为一条连接边;

(3) $Ctrl: S \rightarrow C$, 代表服务到控制 C 的映射图;

(4) C^P 、 C^L 为位置图和连接图,分别表示各计算服务的所在位置和服务之间的依赖关系;

(5) 内部接口 $\langle m, X \rangle$ 代表数据细胞的Bigraph形式中有 m 个根和一组内连接边集合 X ,外部接口 $\langle n, Y \rangle$ 代表有 n 个site和一组外连接边集合 Y ;

定义2. 服务的控制C是五元组 $C = \langle CN, CT, P, CL, U \rangle$, 代表该服务的上下文关系, 其中,

(1) CN和CT分别是该服务控制的名称和类型;

(2) P是当前服务控制上端口的有限集, $\forall p \in P$ 称为一个端口;

(3) $CL = \langle DL, CN \rangle$, 代表该服务当前的依赖状态, DL是当前服务的依赖层次, CN是与之依赖的控制名称;

(4) U是当前服务的可用性度量, 是一个概率值;

定义3. 服务的依赖层次DL代表该服务的组合结构, 取值来自状态集 $\{\circ, <<, >>\}$, 其中 $\circ, <<, >>$ 分别代表此服务和其他服务为无依赖、前置依赖、后置依赖关系;

定义4. 端口的定义是一个四元组 $p = \langle pI, pN, pT, pC, \rangle$, 其中,

(1) pI和pN分别是该端口的ID和名称;

(2) pT是该端口的类型, 代表接口参数的类型;

(3) pC是该端口的控制属性, 取值来自状态集 $\{\frac{in}{\leftarrow}, \frac{out}{\rightarrow}, \frac{call}{\leftrightarrow}\}$, 其中 $\frac{in}{\leftarrow}, \frac{out}{\rightarrow}, \frac{call}{\leftrightarrow}$ 分别代表该端口是输入端口、输出端口还是输入/输出端口;

数据细胞的Bigraph定义目的在于构造数据细胞的结构和生物特性, 每一个数据细胞对应一个原子服务或者复合服务, 从而体现服务的结构资源特点和相互依赖关系, 数据细胞的具体结构和元素定义详见表1, 一个数据细胞对应Bigraph中的Root, 数据细胞中的每个服务对应Bigraph中的节点, 服务之间的依赖关系为Bigraph中节点的连接边, 每个服务的基本信息和上下文关系对应Bigraph中节点的控制, 服务的每个输入、输出端口对应Bigraph中节点端口的形状;

数据细胞元素	Bigraph 对应结构	实例
DC	Root	0, 1, 2...
S	Node	$s_1, s_2, s_3 \dots$
CL	Edge	$e_0, e_1, e_2 \dots$
C	Node control	$C_1, C_2, C_3 \dots$
pC	Node Ports	$(\frac{in}{\leftarrow}), (\frac{out}{\rightarrow}), (\frac{call}{\leftrightarrow})$

表1

由于单个数据细胞提供的服务功能相对较为单一, 但在实际应用中, 为满足用户日益复杂的需求, 需要从网络上选择合适的服务并按照一定的业务规则进行组合, 构建可伸缩的松耦合的组合, 数据细胞簇基于服务组合的四种 workflow 模式, 由多个数据细胞按照功能属性和需求调用动态聚集组合而成; 数据细胞基于Bigraph的合成、并置规则从而形成数据细胞簇, 数据细胞簇同样能够消化从环境中感知吸收和其他数据细胞交流的信息, 智能地与其他数据细胞进行互动, 从而可以提供更加精确复杂的服务功能; DCEM中数据细胞簇的相关形式定义如下:

定义5. 数据细胞簇的定义是一个三元组 $DCC = \langle DCS, CS, LinkS \rangle$, 其中,

(1) DCS是一个数据细胞的有限集, $\forall DC_i \in DCS$ 称为一个数据细胞;

(2) CS是数据细胞组合结构的有限集;

(3) LinkS是数据细胞簇中各端口的连接集合, $\forall \text{Link} \in \text{LinkS}$ 称为两个端口的连接关系;

定义6. 数据细胞的组合结构cs是一个三元组 $cs = \langle DC_i, DC_j, St \rangle$, 其中,

(1) DC_i, DC_j 分别代表数据细胞i和数据细胞j, 且满足 $DC_i \cap DC_j = \Phi$;

(2) St是组合数据细胞的依赖结构, 取值来自状态集 $\{ \cdot, +, ||, * \}$, 分别代表这数据细胞 DC_i 与 DC_j 组合是顺序、选择、并行或者循环结构;

定义7. 端口的连接关系是一个二元组 $\text{Link} = \langle p_i, p_j \rangle$, 其中, p_i, p_j 分别代表两个不同数据服务的端口, 且满足 $p_i \cap p_j = \Phi$;

项语言是Bigraph动态演化的形式化规约和性质验证的基础, 不同 workflow 模式的数据细胞簇项语言定义具有通用特征, 为了更好的描述不同服务结构数据细胞簇的项语言定义, 归纳总结一系列符号和定义, 如表2所示:

符号	定义
DC_m	第m个数据细胞
S_n	第n个数据服务
e_i	第i条数据服务连接边
$i_j,$	第j个内部名
$I_j,$	用外部名替换的第j个内部名
o_k	第k个外部名

表2

通过Bigraph中项语言的表达理论和表2中基于数据细胞结构的相关定义, 对基于服务组合 workflow 模式构造的数据细胞簇分析, 在数据细胞簇项语言形式化上下文中, 定义基本代数概念, 所述基本代数概念包括基调和项集;

定义8: 数据细胞簇形式化定义上的基调 Σ 由类子: 数据细胞集DC、数据服务集S、边集E、内部连接集I、外部连接集O、地点集N和一系列操作符 $\cdot, +, ||, *$ 组成, 每个操作符f的操作数记为 $\text{ar}(f)$;

定义9: 令 Σ 是数据细胞簇形式化定义上的基调, 则基调 Σ 上的项集 $\Pi(\Sigma)$ 是满足以下条件的最小集合:

(1) 每个类子变量都是项集中的元素;

(2) $DC \subseteq \Pi(\Sigma)$

(3) 如果 $f \in \Sigma$, 且 $DC_1, DC_2, \dots, DC_{\text{iar}(f)} \in \Pi(\Sigma)$, 那么 $f(DC_1, DC_2, \dots, DC_{\text{iar}(f)}) \in \Pi(\Sigma)$;

定义10: 令 Σ 是数据细胞簇形式化定义上的基调, $\Pi(\Sigma)$ 是基调 Σ 上的项集, 若 $u_{(f)} \in \Pi(\Sigma)$, 则称 $u_{(f)}$ 为基于操作f的项; 项的定义采用结构归纳定义, 表示成以下形式:

$$u_{(f)} ::= EL_{\text{rule}} \cup \prod_{i=1}^m DC_{\text{iar}(f)}$$

其中 $::=$ 表示归纳定义, EL_{rule} 代表数据细胞簇Bigraph形式中边和连接的形式化定义, $DC_{\text{iar}(f)}$ 代表单个数据细胞的项语言定义, m为数据细胞的个数。

3. 如权利要求2所述的一种基于数据细胞模型的Web服务演化方法, 其特征在于, 所述步骤2.3中, 吸收消化和进化的流程如下:

2.3.1 吸收消化行为

在需求动态变化的环境中,当现有的数据服务不能很好的满足需求,或者发现细胞内服务的缺失时,数据细胞可以通过吸收消化行为获取其他数据细胞内的服务,从而完善自身的服务能力;吸收消化不同于进化,吸收消化行为是两个细胞之间的服务迁移组合,而进化行为则是两个数据细胞的结构重组为一个满足需求的数据细胞;

2.3.2进化行为

数据细胞的进化分为两种,一种是吞噬作用,由具有吞噬作用的细胞吞噬其他数据细胞,被吞噬的细胞成为吞噬细胞内部的一部分,从而达到内部结构更复杂,服务能力更完善的目的;另外一种为聚集作用,包含相似功能以及相近过程服务的细胞聚集在一起形成细胞簇的形式,通过聚集行为构造候选服务集,在服务发现、服务替换时可以很快定义所需服务,缩小服务的查找能耗,从而提高服务的发现效率和自适应能力。

4.如权利要求3所述的一种基于数据细胞模型的Web服务演化方法,其特征在于,所述步骤2.3.1中,数据细胞的吸收消化行为基于Bigraph的迁移操作,其流程包括:

2.3.1.1:细胞 DC_G 吸收消化细胞 DC_F 的前提是 DC_G 的服务和 DC_F 中的某服务存在连接边,其表现在服务与服务之间有关联关系,可构成服务组合,其关联关系通过端口映射实现;

2.3.1.2:查看 DC_G 和 DC_F 是否有连接边,如果这两个细胞没有连接边,则代表这两个Bigraph不相交,即 $|F| \# |G|$,则转步骤2.3.1.3,否则转2.3.1.2.1;

2.3.1.2.1:查询 DC_G 和 DC_F 的服务连接边 e ,以及该连接边对应连接的服务 s_G 和 s_F ;

2.3.1.2.2:复制 DC_F 中 s_F 服务的所有信息,包括服务名称 s 、服务控制 C 和服务到控制的映射 $Ctrl$;

2.3.1.2.3:将步骤2.3.1.2.2中复制的信息添加到 DC_G 中,在 DC_G 的服务集合 S 中添加服务 s_F ,在端口集合 P 中添加 s_F 服务的端口,在 DC_G 的位置图 G^P 和连接图 G^L 中添加 $node_{s_F}$,如果 s_F 有内/外连接或者 $site$,则对应也要在 DC_G 中添加这些元素;

2.3.1.2.4:删除 DC_F 中 s_F 服务的所有信息,包括删除 DC_F 的服务集合 S 中的服务 s_F ,端口集合 P 中 s_F 服务的端口,删除 DC_F 的位置图 F^P 和连接图 F^L 中的 $node_{s_F}$,如果 s_F 有内/外连接或者 $site$,则对应也要在 DC_F 中删除这些元素;

2.3.1.2.5:定义 DC_G 和 DC_F 中不同内接口 $I_i = \langle m_i, X_i \rangle$, $i = G, F$ 的迁移操作原则为 $\langle m_G, X_G \rightleftharpoons X_F \rangle$, DC_G 和 DC_F 中不同外接口 $J_i = \langle n_i, Y_i \rangle$, $i = G, F$ 的迁移操作原则为 $\langle n_G \cup n_F, Y_G \rightleftharpoons Y_F \rangle$;其中 $n_G \cup n_F$ 代表 DC_G 吸收消化 DC_F 后新的外部接口的 $site$ 为两个Bigraph的 $site$ 并集, \rightleftharpoons 代表 DC_F 的内/外连接边向 DC_G 迁移,则细胞 $DC_G: I_G \rightarrow J_G$ 和细胞 $DC_F: I_F \rightarrow J_F$ 发生吸收消化行为后,计算迁移操作之后的接口为 $G \oplus F: I_G \rightleftharpoons I_F \rightarrow J_G \rightleftharpoons J_F$;

2.3.1.3:该细胞 DC_G 不满足吸收消化细胞 DC_F 的条件,不能进行吸收消化行为操作。

5.如权利要求3所述的一种基于数据细胞模型的Web服务演化方法,其特征在于,所述步骤2.3.2中,吞噬和聚集的过程如下:

2.3.2.1细胞吞噬

细胞的吞噬行为基于Bigraph的合成操作,为了使服务组合能力能够具有动态适应性,对基于Bigraph合成操作的数据细胞吞噬规则进行扩展,定义最后生成的数据细胞进行接口的增加、删除;

2.3.2.2细胞聚集

细胞的聚集行为对应Bigraph的扩展乘积操作,细胞 DC_G 和细胞 DC_F 的聚集是将 DC_G 和 DC_F

作为同一类聚集在一起;细胞聚集不同于细胞吞噬,细胞聚集时细胞的总数量不变,只是细胞位置关系更加紧密。

6.如权利要求5所述的一种基于数据细胞模型的Web服务演化方法,其特征在于,所述步骤2.3.2.1中,细胞吞噬的流程如下:

2.3.2.1.1:细胞 DC_G 吞噬细胞 DC_F 的前提是 DC_F 的值域和 DC_G 的定域是相同的,其中,值域为Bigraph的外部接口,定域为Bigraph的内部接口;

2.3.2.1.2:查看细胞 DC_G 和细胞 DC_F 是否有连接边,如果这两个细胞没有连接边,则代表这两个Bigraph不相交,即 $|F|\#|G|$,则转步骤2.3.2.1.2.1,否则转2.3.2.1.3;

2.3.2.1.2.1:查看 DC_G 的接口 $I_G \rightarrow J_G$ 和 DC_F 的接口 $I_F \rightarrow J_F$,如果 $I_G = J_F$,则转步骤2.3.2.1.2.2,否则转2.3.2.1.3;

2.3.2.1.2.2:将 DC_G 中的内连接和 DC_F 的外连接相连,使之成为 DC_G 内服务的连接边;

2.3.2.1.2.3:复制 DC_F 中的服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site添加到 DC_G 中;

2.3.2.1.2.4:在 DC_G 的位置图 G^P 和连接图 G^L 中添加 DC_F 中的服务集S、边集E和步骤2.3.2.1.2.2中形成的连接边,如果S中有内连接或者site,则对应也要在 DC_G 中添加这些元素;

2.3.2.1.2.5:删除 DC_F 中的所有信息,包括删除 DC_F 中的服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site,删除 DC_F 的位置图 F^P 和连接图 F^L ;

2.3.2.1.2.6:为了使服务组合能力能够具有动态适应性,对基于Bigraph合成操作的数据细胞吞噬规则进行扩展,定义最后生成的细胞可以进行接口的增加、删除相应变化;则细胞 DC_G 吞噬细胞 DC_F 行为基于Bigraph合成规则的公式为:

$G \circ F \triangleq \langle G^P \circ F^P, G^L \circ F^L \rangle: I_{F'} \rightarrow J_{G'}$,其中 $I_{F'}$ 、 $J_{G'}$ 分别代表将合成的接口 I_F 、 J_G 进行扩展后的新接口,即 $I_{F'}$ 、 $J_{G'}$ 代表新的细胞继续保持吞噬其他细胞的能力;

2.3.2.1.3:该细胞 DC_G 不满足吞噬细胞 DC_F 的条件,不能进行吞噬进化操作。

7.如权利要求5所述的一种基于数据细胞模型的Web服务演化方法,其特征在于,所述步骤2.3.2.2中,细胞聚集的流程如下:

2.3.2.2.1:细胞 DC_G 和细胞 DC_F 的前提是 DC_G 和 DC_F 中不存在连接边,并且其内部组成的服务功能相似;

2.3.2.2.2:查看细胞 DC_G 和细胞 DC_F 是否有连接边,如果这两个细胞没有连接边,则代表这两个Bigraph不相交,即 $|F|\#|G|$,则转步骤2.3.2.2.2.1,否则转2.3.2.2.3;

2.3.2.2.2.1:复制 DC_F 中的所有信息,包括服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site添加到 DC_G 中;

2.3.2.2.2.2:基于Bigraph的扩展乘积原则,定义 DC_G 和 DC_F 中不同内接口 $I_i = \langle m_i, X_i \rangle$, $i = G, F$ 的扩展乘积操作原则为 $\langle m_G + m_F, X_G \cup X_F \rangle$, DC_G 和 DC_F 中不同外接口 $J_i = \langle n_i, Y_i \rangle$, $i = G, F$ 的扩展乘积操作原则为 $\langle n_G + n_F, Y_G \cup Y_F \rangle$;修改 DC_G 的位置图 G^P 和连接图 G^L ;

2.3.2.2.2.3:删除 DC_F 中的所有信息,包括删除 DC_F 中的服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site,删除 DC_F 的位置图 F^P 和连接图 F^L ;

2.3.2.2.2.4:修改 DC_G 和 DC_F 的接口,如果细胞 $DC_G: I_G \rightarrow J_G$ 和细胞 $DC_F: I_F \rightarrow J_F$ 不相交,它们

接口的扩展乘积为 $G \otimes F: I_G \cup I_F \rightarrow J_G \cup J_F$, 则细胞 DC_G 和细胞 DC_F 聚集行为基于Bigraph扩展乘积规则的公式为: $G \otimes F \triangleq \langle G^P \otimes F^P, G^L \otimes F^L \rangle$;

2.3.2.2.3: 该细胞 DC_G 不满足和细胞 DC_F 发生聚集行为的条件, 不能进行聚集进化操作。

8. 如权利要求1~7之一所述的一种基于数据细胞模型的Web服务演化方法, 其特征在于, 所述步骤1.2中, 数据细胞具有类似生物细胞进行智能活动的的能力, 结合服务自身实际, 数据细胞应具有以下特征: 可进行吸收消化和进化;

吸收消化: 数据细胞能够积极自发地从环境中感知信息, 有三类可以被数据细胞感知的信息, 包括查询服务的信息、查询所涉及的服务资源信息、资源状态的信息, 数据细胞根据感知的环境信息进行资源调整, 通过吸收消化其他数据细胞内的数据服务信息进行服务之间的通信交流;

进化: 基于用户需求和资源状态的变化, 数据细胞能够自发性进化, 通过进化行为, 数据细胞改变其内部服务的结构。

9. 如权利要求8所述的一种基于数据细胞模型的Web服务演化方法, 其特征在于, 数据细胞将进化分解为吞噬和聚集两类行为, 其中: 吞噬行为代表单个数据细胞可以吞噬其他细胞从而形成内部服务结构更为复杂、服务依赖关系更强的细胞, 吞噬行为具有传递性, 不仅会造成现有数据细胞结构的改变、服务功能的增强, 同时也会造成被吞噬数据细胞的消亡与其内包含服务的失效; 聚集行为代表数据细胞可以对相似的服务进行聚类, 从而聚合成功能更为庞大的数据细胞簇。

一种基于数据细胞模型的Web服务演化方法

技术领域

[0001] 本发明涉及Web服务组合领域,尤其是数据服务的组合演化方法。

背景技术

[0002] 数据即服务(Data as a Service,DaaS)是一种新兴的云计算服务模式,通过互联网上的不同协议以及时、低成本的方式向消费者按需提供数据。DaaS的主要目标是克服数据技术中先进方法的局限性,根据这些方法存储数据,并从位置已知且与共享、处理相关的存储库中访问数据。利用DaaS不仅可以随时随地支持数据访问,而且可以降低数据管理成本。不同类型的DaaS,允许消费者在大量丰富的数据源上执行数据请求和分析,提高了对数据的访问灵活性。例如Azure 的Data Market、大数据平台供应商Infochimps与开发位置数据平台Factual。以及随之产生的一系列研究系统。

[0003] 然而,由于单个数据服务功能单一,难以满足在复杂网络环境中各个消费者的需求,因此,服务组合已成为面向服务架构(Service Oriented Architecture,SOA) 中构建新服务和满足业务需求的主要手段,在业务流程的构造和重组中处于核心地位。但高动态的Web服务运行环境和多样化的用户需求增加了服务组合的动态适应难度。事实上,现有的这类服务组合系统并没有过多考虑全局设计与固定需求目标,其内的每个组成服务个体为了自身的利益最大化而提供服务,系统松散结合并根据需求变化或者自身利益而演化。因此,如何有效的对服务组合演化进行演化建模和对服务行为进行分析,成为了现有研究必须应对的问题。

[0004] 服务组合的演化不同于传统的软件演化,由于服务有着多源、异构以及自治的特点,并且具备动态特性,使得服务组合的演化面临着更严峻的挑战。目前国内外学者在服务演化建模及其行为特征方面进行了大量研究,取得了一定的成果。在服务组合演化建模方面,牛津大学的Aitor将环境的服务及其功能抽象为语义 Web服务,提出一个支持用户请求和服务行为动态推理的自适应服务组合框架,但是该组合框架的理论方法主要适用基于物联网的智能城市应用,在其他的应用场景有一定的局限性。科尔多瓦大学的Aurora基于多目标的优化分析Web服务质量(QoS),实现在复杂服务网络中选取QoS侧重点不同的服务组合方案,从而解决Web服务绑定问题的适用性。但是由于多目标算法计算复杂度较高,只能较好适用于设计阶段的服务组合,不能很好的在运行时实现优化过程。对于服务组合的行为分析主要从单个服务或者组合服务的角度,采用例如Petri网、进程代数、 π 演算等形式化的方法进行分析,并在服务组合的行为兼容性和组合有效性方面进行考察,例如,同济大学的Chen提出一种处理模糊语义服务的输入/输出兼容性和行为约束兼容性的自动Web服务组合方法,首先将用户输入/输出需求建模为Horn子句集,然后应用模糊谓词Petri网对子句集进行建模,最终构造一个FPPN模型显示所选服务的调用顺序。爱丁堡大学的Petros在经典线性逻辑中利用过程范式证明来进行复杂Web服务组合,提出一个基于线性逻辑和 π 演算严格定理证明的Web服务组合论述,但以上的研究主要集中在具体服务的抽象行为和语义分析,不能很好的支持在动态上下文环境中服务行为的有效分析,在环境发生改变时,容

易产生服务的性能下降以及组合失效等问题。南京大学的 Song 在服务组合演化中提出一套演化操作准则,只要开发人员在系统演化过程中正确修改服务组合,就可以保证 Web 服务组合的数据流的正确性,而不必重新对整个 Web 服务组合的数据流正确性进行验证。缺点是提出的演化准则较为复杂,在实际应用中可行性不高。

[0005] 生物细胞作为一种结构精密的形态整体,在生长、分化和生理活动过程中其结构能根据环境不断发生变化。在分析细胞的多种复杂的生物特性后,发现将服务组合的动态行为与生物细胞的行为方式进行比较,在某些方面两者呈现出一定的相似特效。因此可以将数据服务同生物细胞结合,通过模拟生物细胞的智能行为,来分析服务组合的演化行为。例如,北京大学的 Zhang 提出的活化细胞理论,将数据组织为一组活化细胞,每个细胞是与计算能力相结合的数据集合,随着用户需求的变化,活化细胞可以自发演变满足新的查询负载,但是其只是提出了一种理论方法,并没有对数据与细胞结合的可行性进行分析。Zhang 等人之后提出一种基于活化细胞理论的面向数据服务体系结构,将活化细胞作为系统的基本单元,细胞之间具有嵌套与层次结构,并具有一系列交互能力,在系统层次提高了数据的分析能力与系统的灵活性。但是其理论主要应用于数据服务体系的非线性任务中提高异构数据的可伸缩性和灵活性。并没有对细胞交互行为进行很好的形式化描述与演化推理。云南大学 Wei 通过对 Spark 集群和生物的 DNA/RNA 测序结合,全面研究了各生物信息学阶段对不同工作负载的性能影响,但是其主要目的在于解决 Spark 系统上如何有效管理生物信息的数据分析服务。

发明内容

[0006] 为了克服传统的形式化方法很难对数据细胞的生命活动进行准确推理的不足,本发明以数据细胞为载体,基于 Bigraph 理论,针对数据即服务的思想提出一种基于数据细胞模型的服务演化方法,将数据服务封装为具有智能行为的数据细胞,从而在数据细胞的层次进行服务组合的演化建模和动态行为分析;该方法通过对数据细胞的静态结构和动态行为模型进行设计与分析,以提高系统的灵活性,使系统能够适应复杂的分析应用。

[0007] 为了解决上述技术问题本发明所采用的技术方案是:

[0008] 一种基于数据细胞模型的服务演化方法,所述服务演化方法包括以下步骤:

[0009] 第一步、结合 Bigraph 理论,建立一个服务演化框架,分为信息抽取和细胞建模及服务行为演化分析两阶段;

[0010] 1.1、信息抽取和细胞建模:提取服务的相关描述信息,得到服务描述元数据,基于抽取的服务信息,根据服务的语法信息匹配服务操作参数与资源状态,基于 Bigraph 理论构建数据细胞和数据细胞簇的结构及生物特性,代表服务的状态信息与服务间潜在的调用关系;

[0011] 1.2、服务行为演化:将数据细胞和数据细胞簇作为服务的演化行为方式的载体,选取数据细胞和细胞簇,按照吸收消化、进化生物细胞行为定义服务的组合生长规则,模拟服务生成组合服务的行为模式;构建基于数据细胞的服务演化模型;

[0012] 第二步、基于所述的框架提出的基于数据细胞模型的服务演化方法,过程如下:

[0013] 步骤 2.1: Web 服务的信息抽取

[0014] 在 Web 服务的信息抽取阶段,将服务的名称映射为 s,每个服务对应一个控制 C,唯

一,组成规则为: $C=s:control$;将服务的可用性度量映射为 U ,将服务的端口的ID、名称和类型分别映射为 pI 、 pN 和 pT ,另外将端口的控制类型映射为 pC ,代表此端口是输入、输出还是输入/输出;将此服务和其他服务的依赖状态映射为 CL ,组成规则为: $CL=\langle DL,CN\rangle$,其中, DL 是当前服务的依赖层次, CN 则是与之依赖的另一个服务的控制名称;

[0015] 步骤2.2:细胞建模

[0016] 对于服务的形式化建模需要根据映射规则,将服务的结构和消息交互行为加以映射到对应形式化结构和动作,从而通过形式化的方法对服务及其组合进行进程演算;

[0017] 步骤2.3:服务行为演化

[0018] 服务组合以服务间的按需聚合为主要运行模式,以动态适应、按需演化作为其运行特征,在不可预测的用户需求驱动下,各协同服务形成错综复杂的交互关系,各服务进程需要与其他服务以一系列操作序列形式进行会话,结合服务与生物细胞特性构建DCEM(数据细胞模型),随着用户需求与资源变更的变化,数据细胞发现现有的内部服务结构已经不能提供高质量的服务能力,需要进行相应的行为演化,使之在需求频繁变更和资源动态更新的情况下也能够高效的提供服务。

[0019] 进一步,所述步骤2.2中,给出的形式定义如下:

[0020] 定义1.数据细胞的Bigraph定义是一个五元组 $DC=\langle S,E,Ctrl,C^P,C^L\rangle:\langle m,X\rangle \rightarrow \langle n,Y\rangle$,其中,

[0021] (1) S 是数据细胞内包含有限的服务集, $\forall s \in S$ 称为一个数据服务;对于数据细胞 DC_i ,其内最多可有 N_i 个数据服务;

[0022] (2) E 是包含服务集的有限边集合, $\forall e \in E$ 称为一条连接边;

[0023] (3) $Ctrl:S \rightarrow C$,代表服务到控制 C 的映射图;

[0024] (4) C^P 、 C^L 为位置图和连接图,分别表示各计算服务的所在位置和服务之间的依赖关系;

[0025] (5) 内部接口 $\langle m,X\rangle$ 代表数据细胞的Bigraph形式中有 m 个根和一组内连接边集合 X ,外部接口 $\langle n,Y\rangle$ 代表有 n 个site和一组外连接边集合 Y ;

[0026] 定义2.服务的控制 C 是五元组 $C=\langle CN,CT,P,CL,U\rangle$,代表该服务的上下文关系,其中,

[0027] (1) CN 和 CT 分别是该服务控制的名称和类型;

[0028] (2) P 是当前服务控制上端口的有限集, $\forall p \in P$ 称为一个端口;

[0029] (3) $CL=\langle DL,CN\rangle$,代表该服务当前的依赖状态, DL 是当前服务的依赖层次, CN 是与之依赖的控制名称;

[0030] (4) U 是当前服务的可用性度量,是一个概率值;

[0031] 定义3.服务的依赖层次 DL 代表该服务的组合结构,取值来自状态集 $\{\circ, \langle, \rangle\}$,其中 \circ, \langle, \rangle 分别代表此服务和其他服务为无依赖、前置依赖、后置依赖等关系;

[0032] 定义4.端口的定义是一个四元组 $p=\langle pI,pN,pT,pC\rangle$,其中,

[0033] (1) pI 和 pN 分别是该端口的ID和名称;

[0034] (2) pT 是该端口的类型,代表接口参数的类型;

[0035] (3) pC 是该端口的控制属性,取值来自状态集 $\{\frac{in}{\leftarrow}, \frac{out}{\rightarrow}, \frac{call}{\leftrightarrow}\}$,其中 $\frac{in}{\leftarrow}, \frac{out}{\rightarrow}, \frac{call}{\leftrightarrow}$

[0036] 分别代表该端口是输入端口、输出端口还是输入/输出端口;

[0037] 数据细胞的Bigraph定义目的在于构造数据细胞的结构和生物特性,每一个数据细胞对应一个原子服务或者复合服务,从而体现服务的结构资源特点和相互依赖关系,数据细胞的具体结构和元素定义详见表1,一个数据细胞对应Bigraph 中的root,数据细胞中的每个服务对应Bigraph中的节点,服务之间的依赖关系为Bigraph中节点的连接边,每个服务的基本信息和上下文关系对应Bigraph中节点的控制,服务的每个输入、输出端口对应Bigraph中节点端口的形状;

数据细胞元素	Bigraph 对应结构	实例
DC	Root	0, 1, 2...
S	Node	$s_1, s_2, s_3...$
CL	Edge	$e_0, e_1, e_2...$
C	Node control	$C_1, C_2, C_3...$
pC	Node Ports	$\nabla \left(\begin{matrix} in \\ \leftarrow \end{matrix} \right) \cdot \left(\begin{matrix} out \\ \rightarrow \end{matrix} \right) \bullet \left(\begin{matrix} call \\ \leftrightarrow \end{matrix} \right)$

[0038] 表1

[0040] 由于单个数据细胞提供的服务功能相对较为单一,但在实际应用中,为满足用户日益复杂的需求,需要从网络上选择合适的服务并按照一定的业务规则进行组合,构建可伸缩的松耦合的组合,数据细胞簇基于服务组合的四种工作流模式,由多个数据细胞按照功能属性和需求调用动态聚集组合而成;数据细胞基于 Bigraph的合成、并置规则从而形成数据细胞簇,数据细胞簇同样能够消化从环境中感知吸收和其他数据细胞交流的信息,智能地与其他数据细胞进行互动,从而可以提供更加精确复杂的服务功能;DCEM中数据细胞簇的相关形式定义如下:

[0041] 定义5.数据细胞簇的定义是一个三元组 $DCC = \langle DCS, CS, LinkS \rangle$,其中,

[0042] (1) DCS是一个数据细胞的有限集, $\forall DC_i \in DCS$ 称为一个数据细胞;

[0043] (2) CS是数据细胞组合结构的有限集;

[0044] (3) LinkS是数据细胞簇中各端口的连接集合, $\forall Link \in LinkS$ 称为两个端口的连接关系;

[0045] 定义6.数据细胞的组合结构cs是一个三元组 $cs = \langle DC_i, DC_j, St \rangle$,其中,

[0046] (1) DC_i, DC_j 分别代表数据细胞i和数据细胞j,且满足 $DC_i \cap DC_j = \Phi$;

[0047] (2) St是该组合数据细胞的依赖结构,取值来自状态集 $\{ \bullet, +, ||, * \}$,分别代表这数据细胞 DC_i 与 DC_j 组合是顺序、选择、并行或者循环结构;

[0048] 定义7.端口的连接关系是一个二元组 $Link = \langle p_i, p_j \rangle$,其中, p_i, p_j 分别代表两个不同数据服务的端口,且满足 $p_i \cap p_j = \Phi$;

[0049] 项语言是Bigraph动态演化的形式化规约和性质验证的基础,不同工作流模式的数据细胞簇项语言定义具有通用特征,为了更好的描述不同服务结构数据细胞簇的项语言定义,归纳总结一系列符号和定义,如表2所示:

符号	定义
DC_m	第m个数据细胞

S_n	第n个数据服务
e_i	第i条数据服务连接边
$i_j,$	第j个内部名
$I_j,$	用外部名替换的第j个内部名
o_k	第k个外部名

[0051] 表2

[0052] 通过Bigraph中项语言的表达理论和表2中基于数据细胞结构的相关定义,对上文所述的基于服务组合工作流模式构造的数据细胞簇分析,在数据细胞簇项语言形式化上下文中,定义基本代数概念,所述基本代数概念包括基调和项集;

[0053] 定义8:数据细胞簇形式化定义上的基调 Σ 由类子:数据细胞集DC、数据服务集S、边集E、内部连接集I、外部连接集O、地点集N和一系列操作符 \cdot 、 $+$ 、 $||$ 、 $*$ 组成,每个操作符f的操作数记为 $ar(f)$;

[0054] 定义9:令 Σ 是数据细胞簇形式化定义上的基调,则基调 Σ 上的项集 $\Pi(\Sigma)$ 是满足以下条件的最小集合:

[0055] (1) 每个类子变量都是项集中的元素;

[0056] (2) $DC \subseteq \Pi(\Sigma)$

[0057] (3) 如果 $f \in \Sigma$,且 $DC_1, DC_2, \dots, DC_{iar(f)} \in \Pi(\Sigma)$,那么 $f(DC_1, DC_2, \dots, DC_{iar(f)}) \in \Pi(\Sigma)$;

[0058] 定义10:令 Σ 是数据细胞簇形式化定义上的基调, $\Pi(\Sigma)$ 是基调 Σ 上的项集,若 $u_{(f)} \in \Pi(\Sigma)$,则称 $u_{(f)}$ 为基于操作f的项;项的定义也可采用结构归纳定义,表示成以下形式:

[0059] $u_{(f)} ::= EL_{rule} \cup \prod_{i=1}^m DC_{iar(f)}$

[0060] 其中 $::=$ 表示归纳定义, EL_{rule} 代表数据细胞簇Bigraph形式中边和连接的形式化定义, DC_i 代表单个数据细胞的项语言定义,m为数据细胞的个数。

[0061] 更进一步,所述步骤2.3中,吸收消化和进化的流程如下:

[0062] 2.3.1吸收消化行为

[0063] 在需求动态变化的环境中,当现有的数据服务不能很好的满足需求,或者发现细胞内服务的缺失时,数据细胞可以通过吸收消化行为获取其他数据细胞内的服务,从而完善自身的服务能力;吸收消化不同于进化,吸收消化行为是两个细胞之间的服务迁移组合,而进化行为则是两个数据细胞的结构重组为一个满足需求的数据细胞;

[0064] 2.3.2进化行为

[0065] 数据细胞的进化分为两种,一种是吞噬作用,由具有吞噬作用的细胞吞噬其他数据细胞,被吞噬的细胞成为吞噬细胞内部的一部分,从而达到内部结构更复杂,服务能力更完善的目的;另外一种为聚集作用,包含相似功能以及相近过程服务的细胞聚集在一起形成细胞簇的形式,通过聚集行为构造候选服务集,在服务发现、服务替换时可以很快定义所需服务,缩小服务的查找能耗,从而提高服务的发现效率和自适应能力。

[0066] 再进一步,所述步骤2.3.1中,数据细胞的吸收消化行为基于Bigraph的迁移操作,其流程包括:

[0067] 2.3.1.1:细胞 DC_G 吸收消化细胞 DC_F 的前提是 DC_G 的服务和 DC_F 中的某服务存在连接边,其表现在服务与服务之间有关联关系,可构成服务组合,其关联关系通过端口映射实现;

[0068] 2.3.1.2:查看 DC_G 和 DC_F 是否有连接边,如果这两个细胞没有连接边,则代表这两个Bigraph不相交,即 $|F| \# |G|$,则转步骤2.3.1.3,否则转2.3.1.2.1;

[0069] 2.3.1.2.1:查询 DC_G 和 DC_F 的服务连接边 e ,以及该连接边对应连接的服务 s_G 和 s_F ;

[0070] 2.3.1.2.2:复制 DC_F 中 s_F 服务的所有信息,包括服务名称 s 、服务控制 C 和服务到控制的映射 $Ctrl$;

[0071] 2.3.1.2.3:将步骤2.3.1.2.2中复制的信息添加到 DC_G 中,在 DC_G 的服务集合 S 中添加服务 s_F ,在端口集合 P 中添加 s_F 服务的端口,在 DC_G 的位置图 G^P 和连接图 G^L 中添加 $node_{s_F}$,如果 s_F 有内/外连接或者 $site$,则对应也要在 DC_G 中添加这些元素;

[0072] 2.3.1.2.4:删除 DC_F 中 s_F 服务的所有信息,包括删除 DC_F 的服务集合 S 中的服务 s_F ,端口集合 P 中 s_F 服务的端口,删除 DC_F 的位置图 F^P 和连接图 F^L 中的 $node_{s_F}$,如果 s_F 有内/外连接或者 $site$,则对应也要在 DC_F 中删除这些元素;

[0073] 2.3.1.2.5:定义 DC_G 和 DC_F 中不同内接口 $I_i = \langle m_i, X_i \rangle$ ($i = G, F$)的迁移操作原则为 $\langle m_G, X_G \rightleftharpoons X_F \rangle$, DC_G 和 DC_F 中不同外接口 $J_i = \langle n_i, Y_i \rangle$ ($i = G, F$)的迁移操作原则为 $\langle n_G \cup n_F, Y_G \rightleftharpoons Y_F \rangle$;其中 $n_G \cup n_F$ 代表 DC_G 吸收消化 DC_F 后新的外部接口的 $site$ 为两个Bigraph的 $site$ 并集, \rightleftharpoons 代表 DC_F 的内/连接边向 DC_G 迁移,则细胞 $DC_G: I_G \rightarrow J_G$ 和细胞 $DC_F: I_F \rightarrow J_F$ 发生吸收消化行为后,计算迁移操作之后的接口为 $G \oplus F: I_G \rightleftharpoons I_F \rightarrow J_G \rightleftharpoons J_F$;

[0074] 2.3.1.3:该细胞 DC_G 不满足吸收消化细胞 DC_F 的条件,不能进行吸收消化行为操作。

[0075] 更进一步,所述步骤2.3.2中,吞噬和聚集的过程如下:

[0076] 2.3.2.1细胞吞噬

[0077] 细胞的吞噬行为基于Bigraph的合成操作,为了使服务组合能力能够具有动态适应性,对基于Bigraph合成操作的数据细胞吞噬规则进行扩展,定义最后生成的数据细胞进行接口的增加、删除;

[0078] 2.3.2.2细胞聚集

[0079] 细胞的聚集行为对应Bigraph的扩展乘积操作,细胞 DC_G 和细胞 DC_F 的聚集是将 DC_G 和 DC_F 作为同一类聚集在一起;细胞聚集不同于细胞吞噬,细胞聚集时细胞的总数量不变,只是细胞位置关系更加紧密。

[0080] 所述步骤2.3.2.1中,细胞吞噬的流程如下;

[0081] 2.3.2.1.1:细胞 DC_G 吞噬细胞 DC_F 的前提是 DC_F 的值域和 DC_G 的定域是相同的,其中,值域为Bigraph的外部接口,定域为Bigraph的内部接口;

[0082] 2.3.2.1.2:查看细胞 DC_G 和细胞 DC_F 是否有连接边,如果这两个细胞没有连接边,则代表这两个Bigraph不相交,即 $|F| \# |G|$,则转步骤2.3.2.1.2.1,否则转2.3.2.1.3;

[0083] 2.3.2.1.2.1:查看 DC_G 的接口 $I_G \rightarrow J_G$ 和 DC_F 的接口 $I_F \rightarrow J_F$,如果 $I_G = J_F$,则转步骤2.3.2.1.2.2,否则转2.3.2.1.3;

[0084] 2.3.2.1.2.2:将 DC_G 中的内连接和 DC_F 的外连接相连,使之成为 DC_G 内服务的连接边;

[0085] 2.3.2.1.2.3:复制 DC_F 中的服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site添加到 DC_G 中;

[0086] 2.3.2.1.2.4:在 DC_G 的位置图 G^p 和连接图 G^L 中添加 DC_F 中的服务集S、边集E 和步骤2.3.2.1.2.2中形成的连接边,如果S中有内连接或者site,则对应也要在 DC_G 中添加这些元素;

[0087] 2.3.2.1.2.5:删除 DC_F 中的所有信息,包括删除 DC_F 中的服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site,删除 DC_F 的位置图 F^p 和连接图 F^L ;

[0088] 2.3.2.1.2.6:为了使服务组合能力能够具有动态适应性,对基于Bigraph合成操作的数据细胞吞噬规则进行扩展,定义最后生成的细胞可以进行接口的增加、删除等相应变化;则细胞 DC_G 吞噬细胞 DC_F 行为基于Bigraph合成规则的公式为: $G \circ F \triangleq \langle G^p \circ F^p, G^L \circ F^L \rangle: I_{F'} \rightarrow J_{G'}$,其中 $I_{F'}$ 、 $J_{G'}$ 分别代表将合成的接口 I_F 、 J_G 进行扩展后的新接口,代表新的细胞继续保持吞噬其他细胞的能力;

[0089] 2.3.2.1.3:该细胞 DC_G 不满足吞噬细胞 DC_F 的条件,不能进行吞噬进化操作。

[0090] 所述步骤2.3.2.2中,细胞聚集的流程如下:

[0091] 2.3.2.2.1:细胞 DC_G 和细胞 DC_F 的前提是 DC_G 和 DC_F 中不存在连接边,并且其内部组成的服务功能相似;

[0092] 2.3.2.2.2:查看细胞 DC_G 和细胞 DC_F 是否有连接边,如果这两个细胞没有连接边,则代表这两个Bigraph不相交,即 $|F| \# |G|$,则转步骤2.3.2.2.2.1,否则转 2.3.2.2.3;

[0093] 2.3.2.2.2.1:复制 DC_F 中的所有信息,包括服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site添加到 DC_G 中;

[0094] 2.3.2.2.2.2:基于Bigraph的扩展乘积原则,定义 DC_G 和 DC_F 中不同内接口 $I_i = \langle m_i, X_i \rangle$ ($i=G, F$)的扩展乘积操作原则为 $\langle m_G + m_F, X_G \cup X_F \rangle$, DC_G 和 DC_F 中不同外接口 $J_i = \langle n_i, Y_i \rangle$ ($i=G, F$)的扩展乘积操作原则为 $\langle n_G + n_F, Y_G \cup Y_F \rangle$;修改 DC_G 的位置图 G^p 和连接图 G^L ;

[0095] 2.3.2.2.2.3:删除 DC_F 中的所有信息,包括删除 DC_F 中的服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site,删除 DC_F 的位置图 F^p 和连接图 F^L ;

[0096] 2.3.2.2.2.4:修改 DC_G 和 DC_F 的接口,如果细胞 $DC_G: I_G \rightarrow J_G$ 和细胞 $DC_F: I_F \rightarrow J_F$ 不相交,它们接口的扩展乘积为 $G \otimes F: I_G \cup I_F \rightarrow J_G \cup J_F$,则细胞 DC_G 和细胞 DC_F 聚集行为基于Bigraph扩展乘积规则的公式为: $G \otimes F \triangleq \langle G^p \otimes F^p, G^L \otimes F^L \rangle$;

[0097] 2.3.2.2.3:该细胞 DC_G 不满足和细胞 DC_F 发生聚集行为的条件,不能进行聚集进化操作。

[0098] 所述步骤1.2中,数据细胞具有类似生物细胞进行智能活动的的能力,结合服务自身实际,数据细胞应具有以下特征:可进行吸收消化和进化;

[0099] 吸收消化:数据细胞能够积极自发地从环境中感知信息,有三类可以被数据细胞感知的信息,包括查询服务的信息、查询所涉及的服务资源信息、资源状态的信息,数据细胞根据感知的环境信息进行资源调整,通过吸收消化其他数据细胞内的数据服务信息进行服务之间的通信交流;

[0100] 进化:基于用户需求和资源状态的变化,数据细胞能够自发性进化,通过进化行

为,数据细胞改变其内部服务的结构;

[0101] 优选的,数据细胞将进化分解为吞噬和聚集两类行为,其中:吞噬行为代表单个数据细胞可以吞噬其他细胞从而形成内部服务结构更为复杂、服务依赖关系更强的细胞,吞噬行为具有传递性,不仅会造成现有数据细胞结构的改变、服务功能的增强,同时也会造成被吞噬数据细胞的消亡与其内包含服务的失效;聚集行为代表数据细胞可以对相似的服务进行聚类,从而聚合成功能更为庞大的数据细胞簇。

[0102] 本发明的有益效果主要表现在要:(1)该服务演化方法运用Bigraph理论将数据服务封装为数据细胞,对数据细胞及细胞簇进行Bigraph建模,为数据细胞的智能行为推理提供了理论依据。(2)该服务演化方法通过对该模型的动态行为分析,可以确保服务交互过程中服务交互行为与其定义属性的一致性,从而提高服务组合执行的有效性与可靠性(3)该服务演化方法参考分析生物细胞的吸收消化、进化等智能行为特点,同服务的动态演化规则相结合,并采用Bigraph理论进行演化推导,提高了服务动态演化的合理性和准确性。

附图说明

[0103] 图1示出了Bigraph元素图,其中,10为根(ROOT)也叫区域(REGION),11为控制(CONTROL),12为节点(NODE),13为地点(SITE),14为外连接名称(OUTER NAME),15为端口(PORT),16为连接边(EDGE),17为内连接名称(INNER NAME)。

[0104] 图2示出了数据细胞的Bigraph形式图。

[0105] 图3示出了数据细胞簇的Bigraph形式图。

[0106] 图4示出了数据细胞吸收消化反应规则图。

[0107] 图5示出了数据细胞吞噬进化规则图。

[0108] 图6示出了数据细胞聚集进化图。

具体实施方式

[0109] 下面结合附图对本发明作进一步描述。

[0110] 参照图1~图6,一种基于数据细胞模型的服务演化方法,所述服务演化方法包括以下步骤:

[0111] 第一步、建立一个服务演化框架,分为信息抽取和细胞建模及服务行为演化分析两阶段;

[0112] 1.1、信息抽取和细胞建模:提取服务的相关描述信息,得到服务描述元数据,基于抽取的服务信息,根据服务的语法信息匹配服务操作参数与资源状态,进行细胞相容性解析,基于Bigraph理论构建数据细胞和数据细胞簇的结构及生物特性,代表服务的状态信息与服务间潜在的调用关系;

[0113] Bigraph是由Milner等学者在2001年提出的一种基于图形的形式化理论工具,旨在强调计算(物理或虚拟)的位置和连接,Bigraph是一个二元组 $B = \langle B^P, B^L \rangle$,如图1所示,其中, B^P, B^L 分别是位置图(place graph)和连接图(link graph)。位置图用以表示Bigraph中各个节点(node)的所在位置,节点之间允许相互嵌套;连接图则忽略节点之间的嵌套关系,图中的边(edge)表示节点之间的连接关系。位置图和连接图是从两个不同的角度对同一个Bigraph观察所得到的结果,因此它们具有相对的独立性。 B^P 由图的节点集 V 、边的集合 E 和

接口组成,嵌套的节点在位置图中为父子关系,用分支关系表示节点之间的嵌套。位置图的接口分为内部接口和外部接口,分别表示根(root)和站点(site),用有穷序数 n 表示,即 $n = \{0, 1, \dots, n-1\}$ 。特别地, $0 = \Phi$ 。连接图的接口也分为内部接口(inner name)和外部接口(outer name),分别表示内部名字集和外部名字集。

[0114] 为了增强表示能力,Bigraph引进基本签名。基本签名是一个二元组 (K, ar) ,这里 K 表示控制(control)的集合, $\forall K \in K$,用巴科斯范式(BNF)表示,则有 $K ::= K | K:n$,其中, $n \in \mathbb{N}$, n 表示 K 的元数(arity), \mathbb{N} 为自然数集。每个控制都有一个映射 $ar:K \rightarrow \mathbb{N}$,表示从控制节点到 \mathbb{N} 的映射。基本签名 K 上的Bigraph,就是给每个节点指派一个控制,控制的元数就是该节点的端口(port)数,用小黑圆圈表示。

[0115] 在数学基础方面,位置图可视为态射,其对象是序数集。连接图则是对象为名字集的态射。每一个Bigraph都可以利用基本的图元(placings、wirings和ions),合成或扩展乘积操作的组合形成Bigraph范式,通过项语言可以对系统性质进行推理和演绎。

[0116] 在数学基础方面,位置图可视为态射,其对象是序数集。连接图则是对象为名字集的态射。每一个Bigraph都可以利用基本的图元(placings、wirings和ions),合成或扩展乘积操作的组合形成Bigraph范式,通过项语言可以对系统性质进行推理和演绎。这些基本的元素以及操作如表3所示,表3:Bigraph表达式中项的涵义:

项	涵义
$U V$	并列区域
$U V$	合并区域
$U \circ V$	合成
$U \otimes V$	扩展乘积
$U(V)$	嵌套, U 包含 V
$K_x^{\rightarrow}(U)$	离子, 节点的控制 k 有外部名 x
1	空的区域
$-i$	地点, 索引为 i
$/x.U$	将 U 的外部名 x 改为一条边
x/y	换名, 用外部名 x 替换内部名 y

[0117] 表3

[0119] 1.2、服务行为演化:将数据细胞和数据细胞簇作为服务的演化行为方式的载体。选取数据细胞和细胞簇,按照吸收消化、进化等生物细胞行为定义服务的组合生长规则,模拟服务生成组合服务的行为模式。构建基于数据细胞的服务演化模型。其中数据细胞具有类似生物细胞进行智能活动的的能力,结合服务自身实际,本发明认为,数据细胞应具有以下特征:可进行吸收消化、进化等行为。吸收消化:数据细胞可以积极自发地从环境中感知信息,一般来说有三类可以被数据细胞感知的信息,查询服务的信息、查询所涉及的服务资源信息、资源状态的信息,数据细胞根据感知的环境信息进行资源调整,通过吸收消化其他数据细胞内的数据服务信息进行服务之间的通信交流。

[0120] 进化:基于用户需求和资源状态的变化,数据细胞可以自发性进化,通过进化行为,数据细胞可以改变其内部服务的结构。数据细胞将进化具体分解为吞噬和聚集两类行

为,其中:吞噬行为代表单个数据细胞可以吞噬其他细胞从而形成内部服务结构更为复杂、服务依赖关系更强的细胞,吞噬行为具有传递性,不仅会造成现有数据细胞结构的改变、服务功能的增强,同时也会造成被吞噬数据细胞的消亡与其内包含服务的失效。聚集行为代表数据细胞可以对相似的服务进行聚类,从而聚合成功能更为庞大的数据细胞簇。

[0121] 第二步、基于所述的框架提出的基于数据细胞模型的服务演化方法,过程如下:

[0122] 步骤2.1:服务的信息抽取

[0123] 在服务的信息抽取阶段,将服务的名称映射为s,每个服务对应一个控制C,唯一,组成规则为: $C=s:control$;将服务的可用性度量映射为U,将服务的端口的ID、名称和类型分别映射为pI、pN和pT,另外将端口的控制类型映射为pC,代表此端口是输入端口、输入端口还是输入/输出端口;将此服务和其他服务的依赖状态映射为CL,组成规则为: $CL=\langle DL, CN \rangle$,其中,DL是当前服务的依赖层次,CN则是与之依赖的另一个服务的控制名称。至此,根据上述定义1—定义7将服务的信息映射至DC.xml结构。

[0124] 步骤2.2:细胞建模

[0125] 对于服务的形式化建模需要根据映射规则,将服务的结构和消息交互行为加以映射到对应形式化结构和动作,从而通过形式化的方法对服务及其组合进行进程演算。本发明基于Bigraph理论构建DCEM(数据细胞模型),该模型是本发明对数据服务封装并对服务组合动态行为分析的前提和基础。本质上DCEM是一个形式化的元模型,主要包含两层:数据细胞和数据细胞簇。形式定义如下:

[0126] 定义1.数据细胞的Bigraph定义是一个五元组 $DC=\langle S,E,Ctrl,C^P,C^L \rangle:\langle m,X \rangle \rightarrow \langle n,Y \rangle$ 。其中,

[0127] (1) S是数据细胞内包含有限的服务集, $\forall s \in S$ 称为一个数据服务;对于数据细胞 DC_i ,其内最多可有 N_i 个数据服务;

[0128] (2) E是包含服务集的有限边集合, $\forall e \in E$ 称为一条连接边;

[0129] (3) $Ctrl:S \rightarrow C$,代表服务到控制C的映射图;

[0130] (4) C^P, C^L 为位置图和连接图,分别表示各计算服务的所在位置和服务之间的依赖关系;

[0131] (5) 内部接口 $\langle m,X \rangle$ 代表数据细胞的Bigraph形式中有m个根和一组内连接边集合X,外部接口 $\langle n,Y \rangle$ 代表有n个site和一组外连接边集合Y。

[0132] 定义2.服务的控制C是五元组 $C=\langle CN,CT,P,CL,U \rangle$,代表该服务的上下文关系,其中,

[0133] (1) CN和CT分别是该服务控制的名称和类型;

[0134] (2) P是当前服务控制上端口的有限集, $\forall p \in P$ 称为一个端口;

[0135] (3) $CL=\langle DL,CN \rangle$,代表该服务当前的依赖状态,DL是当前服务的依赖层次,CN是与之依赖的控制名称;

[0136] (4) U是当前服务的可用性度量,是一个概率值。

[0137] 定义3.服务的依赖层次DL代表该服务的组合结构,取值来自状态集 $\{\circ, \langle, \rangle\}$ 。其中 \circ, \langle, \rangle 分别代表此服务和其他服务为无依赖、前置依赖、后置依赖等关系。

[0138] 定义4.端口的定义是一个四元组 $p=\langle pI,pN,pT,pC, \rangle$,其中,

[0139] (1) pI和pN分别是该端口的ID和名称;

[0140] (2) pT是该端口的类型,代表接口参数的类型;

[0141] (3) pC是该端口的控制属性,取值来自状态集 $\{\frac{in}{\leftarrow}, \frac{out}{\rightarrow}, \frac{call}{\leftrightarrow}\}$ 。其中 $\frac{in}{\leftarrow}$, $\frac{out}{\rightarrow}$, $\frac{call}{\leftrightarrow}$ 分别代表该端口是输入端口、输出端口还是输入/输出端口。

[0142] 数据细胞的Bigraph定义目的在于构造数据细胞的结构和生物特性,每一个数据细胞对应一个原子服务或者复合服务,从而体现服务的结构资源特点和相互依赖关系,数据细胞的具体结构和元素定义详见表1,一个数据细胞对应Bigraph中的root,数据细胞中的每个服务对应Bigraph中的节点,服务之间的依赖关系为Bigraph中节点的连接边,每个服务的基本信息和上下文关系对应Bigraph中节点的控制,服务的每个输入、输出端口对应Bigraph中节点端口的形状。

数据细胞元素	Bigraph 对应结构	实例
DC	Root	0, 1, 2...
S	Node	s ₁ , s ₂ , s ₃ ...
CL	Edge	e ₀ , e ₁ , e ₂ ...
C	Node control	C ₁ , C ₂ , C ₃ ...
pC	Node Ports	$\blacktriangledown \left(\frac{in}{\leftarrow} \right) \bullet \left(\frac{out}{\rightarrow} \right) \bullet \left(\frac{call}{\leftrightarrow} \right)$

[0143] 表1

[0145] 按照表1中数据细胞的结构定义,如图2所示,本发明构造了两个具有不同状态结构的数据细胞,其中,数据细胞DC₀中有一个服务s₁,该服务对应控制为 C₁,DC₀中包含外连接边x和site₁,说明此数据细胞可以被内连接为x的数据细胞吞噬,具体吞噬行为规则见步骤2.3.2.1;数据细胞DC₁中包含两个服务s₂和 s₃,s₂的输出是s₃的输入,两个服务各有自己对应的控制C₂和C₃,两个服务的连接边为e₁,并且DC₁中包含内连接边y和site₁,代表此数据细胞具有吞噬其他细胞进化的能力。

[0146] 由于单个数据细胞提供的服务功能相对较为单一,但在实际应用中,为满足用户日益复杂的需求,需要从网络上选择合适的服务并按照一定的业务规则进行组合,构建可伸缩的松耦合的组合,数据细胞簇基于服务组合的四种 workflow 模式,由多个数据细胞按照功能属性和需求调用动态聚集组合而成。数据细胞基于 Bigraph的合成、并置规则从而形成数据细胞簇,数据细胞簇同样能够消化从环境中感知吸收和其他数据细胞交流的信息,智能地与其他数据细胞进行互动,从而可以提供更加精确复杂的服务功能。本发明提出的DCEM中数据细胞簇的相关形式定义如下:

[0147] 定义5.数据细胞簇的定义是一个三元组DCC= $\langle DCS, CS, LinkS \rangle$,其中,

[0148] (1) DCS是一个数据细胞的有限集, $\forall DC_i \in DCS$ 称为一个数据细胞;

[0149] (2) CS是数据细胞组合结构的有限集;

[0150] (3) LinkS是数据细胞簇中各端口的连接集合, $\forall Link \in LinkS$ 称为两个端口的连接关系。

[0151] 定义6.数据细胞的组合结构cs是一个三元组cs= $\langle DC_i, DC_j, St \rangle$,其中,

[0152] (1) DC_i、DC_j分别代表数据细胞i和数据细胞j,且满足DC_i ∩ DC_j = Φ;

[0153] (2) St是该组合数据细胞的依赖结构,取值来自状态集 $\{\cdot, +, ||, *\}$,分别代表这数据细胞 DC_i 与 DC_i 组合是顺序、选择、并行或者循环结构。

[0154] 定义7.端口的连接关系是一个二元组 $Link = \langle p_i, p_j \rangle$,其中, p_i, p_j 分别代表两个不同数据服务的端口,且满足 $p_i \cap p_j = \Phi$ 。

[0155] 项语言是Bigraph动态演化的形式化规约和性质验证的基础,不同工作流模式的数据细胞簇项语言定义具有通用特征,为了更好的描述不同服务结构数据细胞簇的项语言定义,本发明归纳总结一系列符号和定义,如表2所示:

符号	定义
DC_m	第m个数据细胞
S_n	第n个数据服务
e_i	第i条数据服务连接边
$i_j,$	第j个内部名
$I_j,$	用外部名替换的第j个内部名
O_k	第k个外部名

[0157] 表2

[0158] 通过Bigraph中项语言的表达理论和表2 中基于数据细胞结构的相关定义。对上文所述的基于服务组合工作流模式构造的数据细胞簇分析,本发明在数据细胞簇项语言形式化上下文中,定义了基调、项等基本代数概念。

[0159] 定义8(基调):数据细胞簇形式化定义上的基调 Σ 由类子:数据细胞集DC、数据服务集S、边集E、内部连接集I、外部连接集O、地点集N和一系列操作符 $\cdot, +, ||, *$ 组成,每个操作符f的操作数记为 $ar(f)$ 。

[0160] 定义9(项集):令 Σ 是数据细胞簇形式化定义上的基调,则基调 Σ 上的项集 $\Pi(\Sigma)$ 是满足以下条件的最小集合:

[0161] (1) 每个类子变量都是项集中的元素;

[0162] (2) $DC \subseteq \Pi(\Sigma)$

[0163] (3) 如果 $f \in \Sigma$,且 $DC_1, DC_2, \dots, DC_{iar(f)} \in \Pi(\Sigma)$,那么 $f(DC_1, DC_2, \dots, DC_{iar(f)}) \in \Pi(\Sigma)$ 。

[0164] 定义10(项):令 Σ 是数据细胞簇形式化定义上的基调, $\Pi(\Sigma)$ 是基调 Σ 上的项集,若 $u_{(f)} \in \Pi(\Sigma)$,则称 $u_{(f)}$ 为基于操作f的项。项的定义也可采用结构归纳定义,表示成以下形式:

[0165] $u_{(f)} ::= EL_{rule} \cup \prod_{i=1}^m DC_{iar(f)}$

[0166] 其中 $::=$ 表示归纳定义, EL_{rule} 代表数据细胞簇Bigraph形式中边和连接的形式化定义, DC_i 代表单个数据细胞的项语言定义,m为数据细胞的个数。

[0167] 基于不同工作流构造数据细胞簇的Bigraph形式如图3所示,以图中的并行结构为例分析,该数据细胞簇中含有三个细胞 DC_0, DC_1, DC_2 ,从数据细胞的结构分析,其中 DC_0 和 DC_2 中各有一个服务和一个site, DC_0 和 DC_1 中的服务分别有一个内连接和外连接, DC_1 中有两个服务,其中一个服务内部有一个site;从服务调用关系来看, DC_0 中的 s_1 服务完成后并发执行 DC_1 中的 s_2, s_3 服务,然后执行 DC_2 中的 s_4 服务。 s_1 服务有一个内连接 i_1, s_4 服务有一个外连接

O_4 , 其中在形式化描述中内连接需要进行换名。该数据细胞簇对应的Bigraph项语言定义为 $u_{(1)}$, 按照定义10的描述, $u_{(1)}$ 由两部分组成, 该例中 EL_{rule} 为 $/e_1./e_2./e_3./e_4./I_1.$, 其中 e_i 代表服务之间的连接边, I_1 为 s_1 服务的内连接; $\prod_{i=1}^3 DC_{i ar(1)}$ 为:

$$[0168] \quad s_{1I_1e_1e_2}|-1|I_1/i_1||s_{2e_1e_3}|s_{3e_2e_4}(-2)||s_{4e_3e_4O_4}|-3$$

[0169] 其中 $s_{1I_1e_1e_2}$ 代表 s_1 服务有一个内连接和两个边连接, 边连接表示和其他服务有数据的输入或者输出的依赖关系, I_1/i_1 代表将内连接 i_1 进行换名, $|$ 代表 DC 内服务的同级并置关系, $||$ 代表 DC 之间的同级并置关系, $s_{3e_2e_4}(-2)$ 代表 s_3 服务有两个边连接并且其内有一个 site, 故最终该数据细胞簇的Bigraph项语言形式化定义为:

$$[0170] \quad u_{(1)} ::= /e_1./e_2./e_3./e_4./I_1.(s_{1I_1e_1e_2}|-1|I_1/i_1||s_{2e_1e_3}|s_{3e_2e_4}(-2)||s_{4e_3e_4O_4}|-3)$$

[0171] 顺序结构数据细胞簇中含有两个细胞 DC_0 、 DC_1 , 从数据细胞的结构分析, 其中 DC_0 中有两个服务, DC_1 中有一个服务和一个 site, 并且 DC_1 中的服务有一个外连接; 从服务调用关系来看, DC_0 中的 s_1 服务完成后执行 s_2 服务, 然后执行 DC_1 中的 s_3 服务, 组成顺序结构, 其中 s_3 服务有一个外连接 O_3 , 按照定义10, 该数据细胞簇对应的Bigraph项语言形式化定义为:

$$[0172] \quad u_{(1)} ::= /e_1./e_2./O_3.(s_{1e_1}|s_{2e_1e_2}||s_{3e_2O_3}|-1)$$

[0173] 选择结构数据细胞簇中含有三个细胞 DC_0 、 DC_1 和 DC_2 , 从数据细胞的结构分析, 三个数据细胞中各包含一个服务, DC_0 和 DC_2 中各有一个 site, DC_1 中的服务内部有一个 site 并且该服务有一个内连接; 从服务调用关系来看, DC_1 中的 s_2 服务执行完后选择执行 DC_0 中的 s_1 服务或者执行 DC_1 中的 s_3 服务, 其中三个数据细胞中各包含一个 site, DC_1 中的 s_2 服务具有内连接 i_2 , 该数据细胞簇对应的Bigraph项语言形式化定义为:

$$[0174] \quad u_{(1)} ::= /e_1./e_2./I_2.(s_{1e_1}|-1||s_{2e_1I_2e_2}(-2)|I_2/i_2||s_{3e_2}|-3)$$

[0175] 循环结构中包含两个数据细胞 DC_0 和 DC_1 , 从数据细胞的结构分析, DC_0 中有两个服务 s_1 和 s_2 , 这两个服务是循环关系, s_1 服务内部有一个 site 并且包含一个内连接, DC_1 中有一个服务和一个 site; 从服务调用关系来看, 其中 s_2 服务为 s_1 服务的前置服务, s_2 服务的输出是 s_1 服务的输入, 并且可以循环输入, 最终和 DC_1 中的 s_3 服务组成循环结构的数据细胞簇, s_1 服务有一个内连接 i_1 , DC_0 和 DC_1 中各包含一个 site, 该数据细胞簇对应的Bigraph项语言形式化定义为:

$$[0176] \quad u_{(1)} ::= /e_1./e_2./I_1.(s_{1e_1e_2I_1}(s_{2e_1}|-1)|I_1/i_1||s_{3e_2}|-2)$$

[0177] 步骤2.3: 服务行为演化

[0178] 目前, 服务组合以服务间的按需聚合为主要运行模式, 以动态适应、按需演化作为其运行特征, 在不可预测的用户需求驱动下, 各协同服务形成错综复杂的交互关系, 各服务进程需要与其他服务以一系列操作序列形式进行会话, 本发明结合服务与生物细胞特性构建 DCEM, 例如吸收消化、进化, 随着用户需求与资源变更的变化, 数据细胞发现现有的内部服务结构已经不能提供高质量的服务能力, 需要进行相应的行为演化, 使之在需求频繁变更和资源动态更新的情况下也能够高效的提供服务, 吸收消化和进化的具体流程如下:

[0179] 2.3.1 吸收消化行为

[0180] 在需求动态变化的环境中, 当现有的数据服务不能很好的满足需求, 或者发现细

胞内服务的缺失时,数据细胞可以通过吸收消化行为获取其他数据细胞内的服务,从而完善自身的服务能力。吸收消化不同于进化,吸收消化行为是两个细胞之间的服务迁移组合,而进化行为则是两个数据细胞的结构重组为一个满足需求的数据细胞。数据细胞的吸收消化行为基于Bigraph的迁移操作,其流程包括:

[0181] 2.3.1.1:细胞 DC_G 吸收消化细胞 DC_F 的前提是 DC_G 的服务和 DC_F 中的某服务存在连接边,其表现在服务与服务之间有关联关系,可构成服务组合,其关联关系通过端口映射实现;

[0182] 2.3.1.2:查看 DC_G 和 DC_F 是否有连接边,如果这两个细胞没有连接边,则代表这两个Bigraph不相交,即 $|F| \# |G|$,则转步骤2.3.1.3,否则转2.3.1.2.1;

[0183] 2.3.1.2.1:查询 DC_G 和 DC_F 的服务连接边 e ,以及该连接边对应连接的服务 s_G 和 s_F ;

[0184] 2.3.1.2.2:复制 DC_F 中 s_F 服务的所有信息,包括服务名称 s 、服务控制 C 和服务到控制的映射 $Ctrl$;

[0185] 2.3.1.2.3:将步骤2.3.1.2.2中复制的信息添加到 DC_G 中,在 DC_G 的服务集合 S 中添加服务 s_F ,在端口集合 P 中添加 s_F 服务的端口,在 DC_G 的位置图 G^P 和连接图 G^L 中添加 $node_{s_F}$,如果 s_F 有内/外连接或者 $site$,则对应也要在 DC_G 中添加这些元素;

[0186] 2.3.1.2.4:删除 DC_F 中 s_F 服务的所有信息,包括删除 DC_F 的服务集合 S 中的服务 s_F ,端口集合 P 中 s_F 服务的端口,删除 DC_F 的位置图 F^P 和连接图 F^L 中的 $node_{s_F}$,如果 s_F 有内/外连接或者 $site$,则对应也要在 DC_F 中删除这些元素;

[0187] 2.3.1.2.5:定义 DC_G 和 DC_F 中不同内接口 $I_i = \langle m_i, X_i \rangle$ ($i = G, F$)的迁移操作原则为 $\langle m_G, X_G \rightleftharpoons X_F \rangle$, DC_G 和 DC_F 中不同外接口 $J_i = \langle n_i, Y_i \rangle$ ($i = G, F$)的迁移操作原则为 $\langle n_G \cup n_F, Y_G \rightleftharpoons Y_F \rangle$;其中 $n_G \cup n_F$ 代表 DC_G 吸收消化 DC_F 后新的外部接口的 $site$ 为两个Bigraph的 $site$ 并集, \rightleftharpoons 代表 DC_F 的内/连接边向 DC_G 迁移,则细胞 $DC_G: I_G \rightarrow J_G$ 和细胞 $DC_F: I_F \rightarrow J_F$ 发生吸收消化行为后,计算迁移操作之后的接口为 $G \oplus F: I_G \rightleftharpoons I_F \rightarrow J_G \rightleftharpoons J_F$;

[0188] 2.3.1.3:该细胞 DC_G 不满足吸收消化细胞 DC_F 的条件,不能进行吸收消化行为操作。

[0189] 如图4所示,现有的三个数据细胞 DC_0 、 DC_1 和 DC_2 ,其中 DC_0 中包含两个服务 s_1 和 s_2 , DC_0 中有一个 $site$, s_2 服务内部有一个 $site$ 并且有一个内连接 i_2 , DC_1 中有一个服务 s_3 ,该服务有一个内连接 i_3 并且内部有一个 $site$, DC_2 中有一个服务 s_4 和一个 $site$;这三个数据细胞通过组合形成顺序结构的细胞簇, DC_0 的 s_1 服务、 DC_1 的 s_3 服务和 DC_2 的 s_4 服务依次执行,随着需求的变化, DC_1 通过对环境信息的感知分析,发现自身包含的服务已经不能很好的适配现有的需求, DC_1 从它相邻的细胞 DC_0 中吸收消化 s_1 ,之后 s_1 作为 s_3 的前置服务, s_1 的输出参数作为 s_3 的输入参数,该组合从之前的顺序结构变成循环结构,完善自己的服务功能。在项语言的形式化定义中,由于其中 DC_0 中的 s_1 服务被 DC_1 吸收消化,所以 DC_0 结构中 s_1 的形式化部分变为空,而 DC_1 中新增一个 s_1 服务,所以在形式化定义中补充 s_1 的部分,并且和 s_3 是循环结构,变化后的Bigraph形式化定义中各服务的边连接集合也要发生对应的变化。用项语言形式化表示反应过程如下:

[0190]
$$\begin{aligned} & /e_1./e_2./I_2./I_3.(s_{1e_1}|_{-1}|s_{2I_2}(-2)|I_2/i_2||s_{3e_1I_3e_2}(-3)|I_3/i_3||s_{4e_2}|_{-4}) \rightarrow \\ & /e_1./e_2./I_2./I_3.(s_{2I_2}(-2)|_{-1}|I_2/i_2||s_{3e_1I_3e_2}(s_{1e_1}|_{-3})|I_3/i_3||s_{4e_2}|_{-4}) \end{aligned}$$

[0191] 2.3.2进化行为

[0192] 数据细胞的进化分为两种,一种是吞噬作用,由具有吞噬作用的细胞吞噬其他数据细胞,被吞噬的细胞成为吞噬细胞内部的一部分,从而达到内部结构更复杂,服务能力更完善的目的;另外一种为聚集作用,包含相似功能以及相近过程服务的细胞聚集在一起形成细胞簇的形式,通过聚集行为构造候选服务集,在服务发现、服务替换时可以很快定义所需服务,缩小服务的查找能耗,从而提高服务的发现效率和自适应能力。吞噬和聚集的过程如下:

[0193] 2.3.2.1细胞吞噬

[0194] 细胞的吞噬行为基于Bigraph的合成操作,为了使服务组合能力能够具有动态适应性,本发明对基于Bigraph合成操作的数据细胞吞噬规则进行扩展,定义最后生成的数据细胞可以进行接口的增加、删除等相应变化,其流程如下:

[0195] 2.3.2.1.1:细胞 DC_G 吞噬细胞 DC_F 的前提是 DC_F 的值域和 DC_G 的定域是相同的,其中,值域为Bigraph的外部接口,定域为Bigraph的内部接口;

[0196] 2.3.2.1.2:查看细胞 DC_G 和细胞 DC_F 是否有连接边,如果这两个细胞没有连接边,则代表这两个Bigraph不相交,即 $|F| \# |G|$,则转步骤2.3.2.1.2.1,否则转 2.3.2.1.3;

[0197] 2.3.2.1.2.1:查看 DC_G 的接口 $I_G \rightarrow J_G$ 和 DC_F 的接口 $I_F \rightarrow J_F$,如果 $I_G = J_F$,则转步骤2.3.2.1.2.2,否则转2.3.2.1.3;

[0198] 2.3.2.1.2.2:将 DC_G 中的内连接和 DC_F 的外连接相连,使之成为 DC_G 内服务的连接边;

[0199] 2.3.2.1.2.3:复制 DC_F 中的服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site添加到 DC_G 中;

[0200] 2.3.2.1.2.4:在 DC_G 的位置图 G^P 和连接图 G^L 中添加 DC_F 中的服务集S、边集E 和步骤2.3.2.1.2.2中形成的连接边,如果S中有内连接或者site,则对应也要在 DC_G 中添加这些元素;

[0201] 2.3.2.1.2.5:删除 DC_F 中的所有信息,包括删除 DC_F 中的服务集S、服务的控制C、边集E、服务到控制的映射Ctrl以及内部的site,删除 DC_F 的位置图 F^P 和连接图 F^L ;

[0202] 2.3.2.1.2.6:为了使服务组合能力能够具有动态适应性,本发明对基于Bigraph合成操作的数据细胞吞噬规则进行扩展,定义最后生成的细胞可以进行接口的增加、删除等相应变化;则细胞 DC_G 吞噬细胞 DC_F 行为基于Bigraph合成规则的公式为: $G \circ F \triangleq \langle G^P \circ F^P, G^L \circ F^L \rangle: I_{F'} \rightarrow J_{G'}$,其中 $I_{F'}$ 、 $J_{G'}$ 分别代表将合成的接口 I_F 、 J_G 进行扩展后的新接口,代表新的细胞继续保持吞噬其他细胞的能力。

[0203] 2.3.2.1.3:该细胞 DC_G 不满足吞噬细胞 DC_F 的条件,不能进行吞噬进化操作。

[0204] 如图5所示,代表不同数据细胞吞噬行为的过程,对图5中Bigraph G_1 和 F_1 合成分析, G_1 中有两个数据细胞 DC_0 和 DC_1 , DC_0 中有两个服务 s_1 和 s_2 , s_1 是 s_2 的前置服务, s_2 有一个内连接x, DC_1 中有一个服务 s_3 和一个site, F_1 中包含一个数据细胞 DC_2 , DC_2 中有一个服务 s_3 ,并且该服务有一个外连接x;其中数据细胞 DC_0 和 DC_1 形成一种顺序结构, DC_0 内的 s_1 服务和 s_2 服务是一种循环结构。 s_2 服务中的内连接x和细胞 DC_1 中的site,代表 DC_0 和 DC_1 组成的Bigraph G_1 具有吞噬其他数据细胞的能力。可以吞噬一个外连接也为x的数据细胞。 G_1 的接口定义为: $\langle 1, \{x\} \rangle \rightarrow \langle 2, \emptyset \rangle$, G_1 的项语言定义为: $/e_1./e_2./m.(s_{2e_1}me_2(s_{1e_1})|m/x||s_{3e_2}|_1)$,在此例中

Bigraph G_1 吞噬一个内含单个服务 s_4 的数据细胞 DC_2 , F_1 的接口定义为: $\in \rightarrow \langle 1, \{x\} \rangle$, F_1 的项语言定义为: s_{4x} , 最终形成 Bigraph G_1' 中的 DC_0 和 DC_1 是一个并行结构, 而 DC_0 内部是循环结构的细胞簇。本发明扩展了 Bigraph 的合成规则, 在最后形成的细胞 DC_1 中新增了一个内连接 y 和 $site$, 代表 DC_1 能够具备继续吞噬的能力, 从而增强数据细胞演化行为的持续性。在项语言定义中, DC_2 被吞噬, 所以其节点元素 s_4 被添加到 DC_1 中, 并且外连接 x 和 DC_0 的内连接 x 进行连接形成内部边, 最终形成的 Bigraph G_1' 的接口定义为: $\langle 1, \{y\} \rangle \rightarrow \langle 2, \varphi \rangle$, 项语言定义为:

[0205] $/e_1./e_2./x./n.(s_{2e_1e_2x}(s_{1e_1})||s_{3e_2}|s_{4xn}|_{-1}|n/y)$

[0206] G_2 和 F_2 的吞噬进化过程与上述相似, G_2 中包含三个数据细胞 DC_0 、 DC_1 和 DC_2 , 其中 DC_0 和 DC_1 中分别包含一个服务 s_1 和 s_2 , 两个服务是顺序结构, s_2 服务有一个内连接 x , DC_2 中有一个 $site$, 代表 Bigraph G_2 具有吞噬其他数据细胞的能力, 可以吞噬一个对应外连接为 x 的数据细胞, G_2 的接口定义为: $\langle 1, \{x\} \rangle \rightarrow \langle 3, \varphi \rangle$, G_2 的项语言定义为:

$/e_1./m.(s_{1e_1}||s_{2e_1m}|m/x||_{-1})$, F_2 中有一个数据细胞 DC_3 , 该数据细胞中有一个服务 s_3 和一个

$site$, F_2 的接口定义为: $\langle 1, \varphi \rangle \rightarrow \langle 1, \{x\} \rangle$, F_2 的项语言定义为: $s_{3x}|_{-2}$, 最终形成的 Bigraph G_2' 是一个选择结构的数据细胞簇, 其接口定义为: $\langle 1, \varphi \rangle \rightarrow \langle 3, \varphi \rangle$, 项语言的形式化定义为:

[0207] $/e_1./x.(s_{1e_1}||s_{2e_1x}|s_{3x}||_{-2})$

[0208] 2.3.2.2 细胞聚集

[0209] 细胞的聚集行为对应 Bigraph 的扩展乘积操作, 细胞 DC_G 和细胞 DC_F 的聚集是将 DC_G 和 DC_F 作为同一类聚集在一起。细胞聚集不同于细胞吞噬, 细胞聚集时细胞的总数量不变, 只是细胞位置关系更加紧密, 其流程如下:

[0210] 2.3.2.2.1: 细胞 DC_G 和细胞 DC_F 的前提是 DC_G 和 DC_F 中不存在连接边, 并且其内部组成的服务功能相似;

[0211] 2.3.2.2.2: 查看细胞 DC_G 和细胞 DC_F 是否有连接边, 如果这两个细胞没有连接边, 则代表这两个 Bigraph 不相交, 即 $|F| \# |G|$, 则转步骤 2.3.2.2.2.1, 否则转 2.3.2.2.3;

[0212] 2.3.2.2.2.1: 复制 DC_F 中的所有信息, 包括服务集 S 、服务的控制 C 、边集 E 、服务到控制的映射 $Ctrl$ 以及内部的 $site$ 添加到 DC_G 中;

[0213] 2.3.2.2.2.2: 基于 Bigraph 的扩展乘积原则, 定义 DC_G 和 DC_F 中不同内接口 $I_i = \langle m_i, X_i \rangle$ ($i=G, F$) 的扩展乘积操作原则为 $\langle m_G + m_F, X_G \cup X_F \rangle$, DC_G 和 DC_F 中不同外接口 $J_i = \langle n_i, Y_i \rangle$ ($i=G, F$) 的扩展乘积操作原则为 $\langle n_G + n_F, Y_G \cup Y_F \rangle$; 修改 DC_G 的位置图 G^p 和连接图 G^l ;

[0214] 2.3.2.2.2.3: 删除 DC_F 中的所有信息, 包括删除 DC_F 中的服务集 S 、服务的控制 C 、边集 E 、服务到控制的映射 $Ctrl$ 以及内部的 $site$, 删除 DC_F 的位置图 F^p 和连接图 F^l ;

[0215] 2.3.2.2.2.4: 修改 DC_G 和 DC_F 的接口, 如果细胞 $DC_G: I_G \rightarrow J_G$ 和细胞 $DC_F: I_F \rightarrow J_F$ 不相交, 它们接口的扩展乘积为 $G \otimes F: I_G \cup I_F \rightarrow J_G \cup J_F$, 则细胞 DC_G 和细胞 DC_F 聚集行为基于 Bigraph 扩展乘积规则的公式为: $G \otimes F \triangleq \langle G^p \otimes F^p, G^l \otimes F^l \rangle$;

[0216] 2.3.2.2.3: 该细胞 DC_G 不满足和细胞 DC_F 发生聚集行为的条件, 不能进行聚集进化操作。

[0217] 如图6所示, 代表不同数据细胞聚集行为的过程, 对图6聚集进化中的 Bigraph G_1

和 F_1 分析, G_1 中包含两个数据细胞 DC_0 和 DC_1 ,数据细胞 DC_0 和 DC_1 分别包含服务 s_1 和 s_2 ,形成顺序结构,并且 DC_0 和 DC_1 中各有一个site, DC_1 中的 s_2 服务有一个内连接 x , F_1 中包含一个数据细胞 DC_2 ,数据细胞 DC_2 中含有两个服务 s_3 及 s_4 ,并且同样也是顺序结构;其中Bigraph G_1 的接口定义为: $\langle 2, \{x\} \rangle \rightarrow \langle 2, \varphi \rangle$,项语言定义为:

[0218] $/e_1./m.(s_{1e_1}|_{-1}||s_{2e_1m}|_{-2}|m/x)$

[0219] Bigraph F_1 的接口定义为: $\epsilon \rightarrow \langle 1, \varphi \rangle$,项语言定义为: $/e_2.(s_{3e_2}|s_{4e_2})$,DCEM 模型在第二阶段进行服务动态行为自演化的时,发现 s_1 、 s_2 组合而成的服务与 s_3 、 s_4 组合的服务具有一定相似性,于是将这两个组合服务进行聚类,对应Bigraph G_1 和 F_1 的扩展乘积操作,将两部分的项语言元素(包括节点、边和site)进行组合,最终形成的Bigraph G_1' 的接口定义为: $\langle 2, \{x\} \rangle \rightarrow \langle 3, \varphi \rangle$,项语言定义为:

[0220] $/e_1./e_2./m.(s_{1e_1}|_{-1}||s_{2e_1m}|_{-2}|m/x||s_{3e_2}|s_{4e_2})$

[0221] 聚集进化的 G_2 和 F_2 与之类似,Bigraph G_2 中包含两个数据细胞 DC_0 和 DC_1 , DC_0 中有一个服务 s_1 和一个site, DC_1 中有一个服务 s_2 并且该服务内部有一个site, F_2 中包含两个数据细胞 DC_2 和 DC_3 ,其中 DC_2 有一个服务 s_3 , s_3 有一个外连接边 x 和内连接边 y , DC_3 中有一个site;其中Bigraph G_2 的接口定义为: $\langle 2, \varphi \rangle \rightarrow \langle 2, \varphi \rangle$,项语言定义为: $/e_1.(s_{1e_1}|_{-0}||s_{2e_1}(-_1))$, Bigraph F_2 的接口定义为: $\langle 1, \{y\} \rangle \rightarrow \langle 2, \{x\} \rangle$,项语言定义为: $/n.(s_{3xn}|n/y|_{-2})$,同理这两个具有相似功能的组合服务进行聚类,最终形成的Bigraph G_2' 的接口定义为: $\langle 3, \{y\} \rangle \rightarrow \langle 4, \{x\} \rangle$,项语言定义为:

[0222] $/e_1./n.(s_{1e_1}|_{-0}||s_{2e_1}(-_1)||s_{3xn}|n/y|_{-2})$ 。

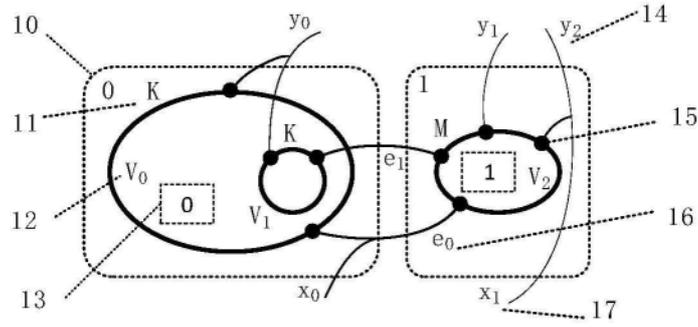


图1

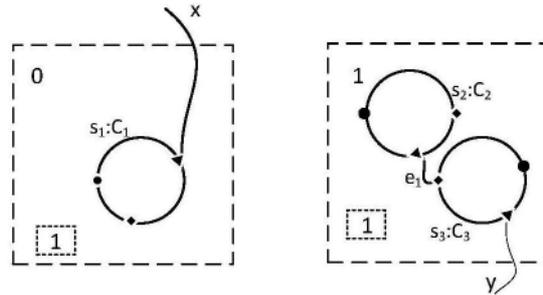


图2

服务组合结构	细胞膜	形式化描述
<p>Case 1: 并行结构</p>		<p>Graph:</p> $/e_1./e_2./e_3./e_4./l_1.(S_{11}e_{12} S_{11}e_{12} S_{21}e_{13} S_{21}e_{13} S_{31}e_{14} S_{31}e_{14} S_{41}e_{15} S_{41}e_{15})$
<p>Case 2: 顺序结构</p>		<p>Graph:</p> $/e_1./e_2./o_3.(S_{1e1} S_{2e1}e_2 S_{3e1}e_2 S_{4e1}e_3)$
<p>Case 3: 选择结构</p>		<p>Graph:</p> $/e_1./e_2./l_2.(S_{1e1} S_{1e1} S_{2e1}e_2 S_{2e1}e_2 S_{3e1}e_3)$
<p>Case 4: 循环结构</p>		<p>Graph:</p> $/e_1./e_2./l_1.(S_{1e1}e_2 S_{2e1}e_2 S_{2e1}e_2 S_{3e1}e_2)$

图3

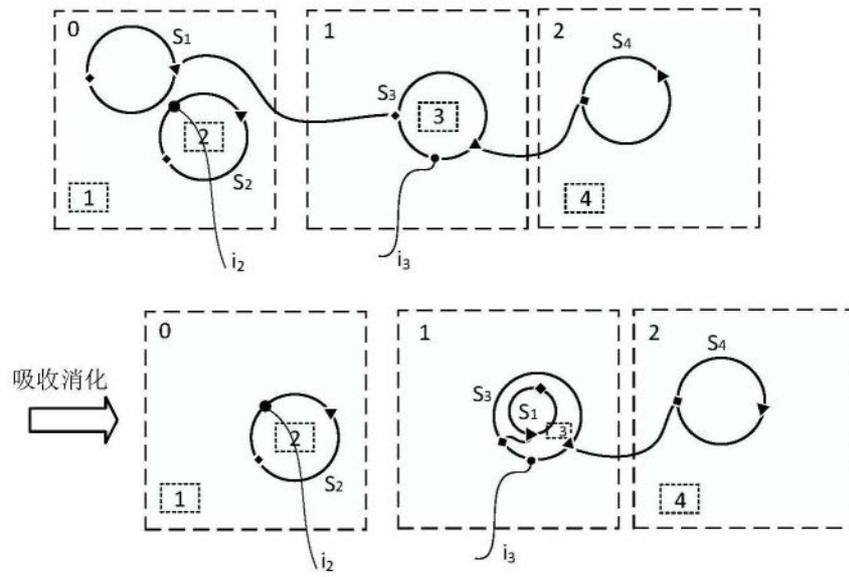


图4

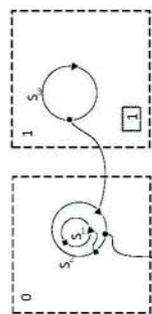
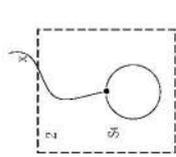
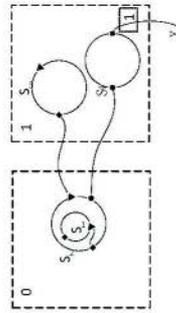
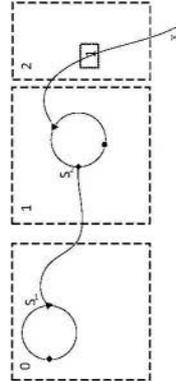
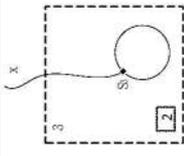
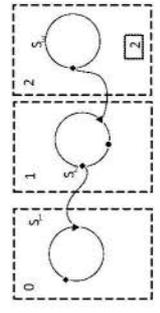
吞噬行为	进化细胞	进化细胞	进化后细胞	形式化描述
 <p>Bigraph $G_1 : \langle \langle 1, x \rangle \rangle \langle 2, \emptyset \rangle$</p>	 <p>Bigraph $F_1 : \langle \langle 1, x \rangle \rangle$</p>	 <p>Bigraph $G_2 = GeF_4$ $G_1 : \langle \langle 1, y \rangle \rangle \langle 2, \emptyset \rangle$</p>	$G_1 : /e_1/m_1(S_{in1}) m(x) \langle S_{in2} x \rangle$ $F_1 : S_{in}$ $G_1 : /e_1/e_2/x/n_1(S_{in1}, S_{in2}) S_{in2} S_{in1} x n_1(y)$	
 <p>Bigraph $G_1 : \langle \langle 1, x \rangle \rangle \langle 3, \emptyset \rangle$</p>	 <p>Bigraph $F_2 : \langle \langle 1, \emptyset \rangle \rangle \langle 1, x \rangle$</p>	 <p>Bigraph $G_2 = GeF_2$ $G_2 : \langle \langle 1, \emptyset \rangle \rangle \langle 3, \emptyset \rangle$</p>	$G_2 : /e_1/m_1(S_{in1}) S_{in1} m(x) x \rangle$ $F_2 : S_{in} x$ $G_2 : /e_1/x S_{in1} S_{in1} S_{in1} S_{in1} x \rangle$	

图5

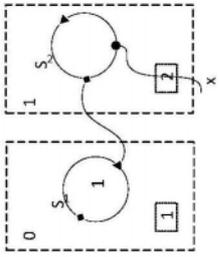
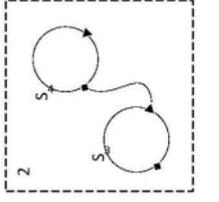
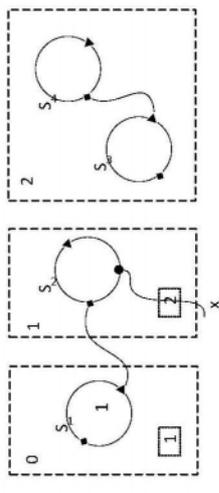
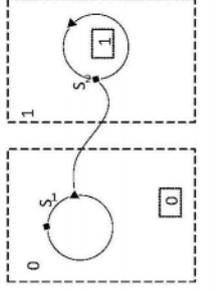
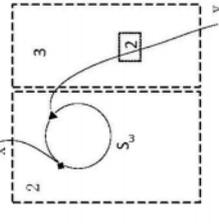
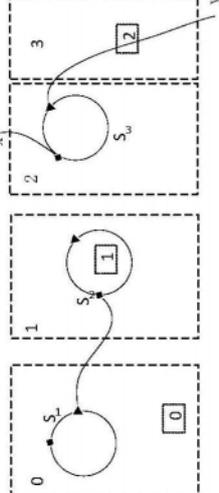
	进化细胞	进化细胞	进化后细胞	形式化描述
聚集行为	 <p style="text-align: center;">Bigraph $G_1: \langle 2, \{x \rangle \langle 2, \emptyset \rangle$</p>	 <p style="text-align: center;">Bigraph $F_1: \emptyset \langle 1, \emptyset \rangle$</p>	 <p style="text-align: center;">Bigraph $G_1 = G_0 \circ F_1$ $G_1: \langle 2, \{x \rangle \langle 3, \emptyset \rangle$</p>	$G_1: /e1./m.(S_{1e1} _{-1} S_{2e1m} _{-2} m/x)$ $F_1: /e2.(S_{3e2} S_{4e2})$ $G_1: /e1./e2./m.(S_{1e1} _{-1} S_{2e1m} _{-2} m/x S_{3e2} S_{4e2})$
	 <p style="text-align: center;">Bigraph $G_2: \langle 2, \emptyset \rangle \langle 2, \emptyset \rangle$</p>	 <p style="text-align: center;">Bigraph $F_2: \langle 1, \{y \rangle \langle 2, \{x \rangle$</p>	 <p style="text-align: center;">Bigraph $G_2 = G_0 \circ F_2$ $G_2: \langle 3, \{y \rangle \langle 4, \{x \rangle$</p>	$G_2: /e1.(S_{1e1} _{-0} S_{2e1(-1)})$ $F_2: /n.(S_{3en} n/y z)$ $G_2: /e1./n.(S_{1e1} _{-0} S_{2e1(-1)} S_{3en} n/y z)$

图6