

基于群组角色协同服务聚合的 Web 服务发现

黄 黎¹, 赵 璐²

(1.江苏开放大学信息工程学院,江苏 南京 210017)

(2.南京邮电大学计算机学院,江苏 南京 210023)

[摘要] 针对群组绩效和交互成本的双目标约束优化问题,设计了一种基于群组角色协同服务聚合的多用户 Web 服务发现方法.该方法定义基于社会网络分析的协作准则,考虑群组成员间的协作性能和交互成本度量,提出一种协同候选服务间的交互成本计算方法;针对群组角色分配问题,重点考虑候选任务执行者的“角色”资源,引入群组协同的语义信息来丰富服务的语义描述;通过融合结构和语义信息的层次聚类算法寻找服务候选角色,提高服务聚合的质量,解决群组绩效和交互成本的双目标约束优化问题.实验验证该方法在服务组合发现,满足服务行为兼容性和功能可靠性等方面具有较明显的优势.

[关键词] 群智协同,服务聚合,群组角色,群组绩效,服务计算

[中图分类号] TP393 [文献标志码] A [文章编号] 1001-4616(2024)04-0135-13

Web Service Composition Discovery Based on Group Role Collaboration Service Aggregation

Huang Li¹, Zhao Lu²

(1.School of Information Engineering, Jiangsu Open University, Nanjing 210017, China)

(2.School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Towards two-objective constraint optimization problem for group performance and interaction cost, a service aggregation method based on Group Role Collaboration (GRC) is designed. Considering the collaboration performance and interaction cost among group members, an interaction cost calculation method is proposed. To resolve the problem of group role assignment, especially the “role” profile, group collaborative based semantic information is introduced to enrich the semantic description of services. A hierarchical clustering algorithm combining structure and semantic information is used to search for service candidate roles. It could improve the quality of service aggregation and resolve the dual objective constraint optimization between group performance and interaction cost. Experiments show that this method has obvious optimization effect on service aggregation in terms of satisfying service behavior compatibility and functional reliability.

Key words: crowd-based cooperation, service aggregation, group role, group performance, service computing

近年来,云计算技术的发展推动了信息技术支撑的组织领域的变革,云环境重新定义了 web 服务的部署,涉及云服务提供商 (cloud service provider, CSP) 和分布在互联网上的智能应用程序.在多云环境中,涉及同时利用由不同 CSP 提供并配备不同设置的多个云服务 (cloud service, CS),通常称为云服务组合 (cloud service composition, CSC),并形成了跨组织、松散耦合的多服务工作流聚合^[1-2].目前,拥有大量可用的 CS 可以满足特定的服务质量 (QoS) 要求,由此产生的解决方案空间变得巨大, CSC 问题具有非线性、多目标和不确定性的特点, CSC 被公认为 NP 难题^[3].业界学者提出许多创新方法,例如多准则决策^[4]、模

收稿日期:2024-10-09.

基金项目:国家自然科学基金项目 (62206114, 62403218, 62402233)、江苏省教育科学规划课题 (B/2023/01/50)、江苏省高等教育教改课题 (2023JSJG690).

通讯作者:黄黎,博士,副教授,研究方向:服务计算、数据挖掘. E-mail:huangli713@126.com

糊逻辑^[5]和元启发式算法^[6]来最大限度地提高 QoS 并解决 CSC 问题. 然而, QoS 驱动的方法在大规模系统中面临计算密集型的性质和解决问题的 NP 难度^[7]. 此外, 现有的解决方案通常涉及 CSC 资源受限的服务组合协同推荐问题, 而没有充分解决角色作为主要的底层机制来促进协作活动的实质. 在实践中, 依赖角色系统的交互成本效益, 也会导致成本效益最小化计算复杂性显著提升带来的效率下降问题^[8].

基于角色的服务推荐通常应用于统筹层面的服务资源优化问题^[9]. 本文针对群组绩效和交互成本的双目标约束优化问题, 设计了一种基于群组角色协同服务聚合的多用户 Web 服务发现方法. 基于角色的协同(role-based collaboration, RBC)将角色作为主要的底层机制来促进协作活动. 因此, 运用群组角色协同的方法解决云服务的聚合问题具有重要的研究意义和价值.

1 相关研究成果

CSC 在将单个云服务集成到更复杂、更有价值的复合服务中发挥着关键作用, 从而能够产生更加全面的解决方案来满足特定的用户需求. 在工业 4.0 背景下, 为提高边缘计算在服务组合中的健壮性与可扩展性, 学者们进行了大量的研究, 其中基于数字孪生和区块链技术的联邦学习研究取得了新的进展^[3-5]. 有学者提出一种基于上下文感知可信服务流程创建服务推荐系统, 全面结合了区块链、循环神经网络(RNN)和 Skip Gram 模型, 以提高服务流程创建与推荐的安全性、效率和质量^[10]. Zhu 等^[11]提出一种公平感知的服务推荐(FASR), 并分别为流行和不受欢迎的服务设计了不同的推荐算法. Zhang 等^[12]提出了一种即时推荐 Web 服务的方法, 以实现可持续的服务 Mashup. Yu 等^[13]提出了一种基于交叉视图和对比学习(LSRCC)的长尾服务推荐方法, 采用递归神经网络、多层感知器和注意力机制.

群组角色分配(group role assignment, GRA)是 RBC 的核心问题, 但 GRA 研究的困难在于带约束条件的角色指派. Zhu^[14]研究 E-CARGO 模型处理角色赋值指派问题时, 借助 Q 矩阵完成相对简单的约束指派, 但是 Q 矩阵难以描述现实世界中复杂多变的要求, 因此制约了 E-CARGO 模型的应用. 滕少华等^[15]将约束集(Constraint)引进 E-CARGO 模型, 提出了带约束的 EC-CARGO 模型. 对角色转换的检测, 角色识别能够反映出社会成员之间现实的或潜在的社会关系. Zhang 等^[16]通过融合了节点的高阶结构和低阶结构的影响识别社会网络中的重要节点. Mahyar 等^[17]使用压缩感知识别社交网络中信息流的中心节点. Boulmakoul 等^[18]将结构分析与 Q-analysis 和 SNA 技术相结合, 对企业组织结构进行分析, 辅助企业决策.

然而, 跨组织内部网和开源项目中的资源描述是非结构化和异构的, BPM 缺乏统一的形式化定义. 基于用户负责的安全技术角度和基于深度学习的方法解决服务组合问题均有可能存在计算复杂度较高, 计算代价高昂的问题. 科研合作和开源软件项目开发就是典型的服务聚合协同优化问题, 其特点在于自组织性和动态性, 合作者主要受到相关研究领域的驱动和自我激励, 对项目做出贡献和开展合作, 由此派生出一个社会协作网络, 其中成员或组织之间的社会关系将会影响 BPM 整体性能^[19].

由于服务数量剧增, 大规模组合服务间的交互行为日益复杂, 大规模复杂系统受到个体行为拓扑和环境的影响, 使得系统的整体行为难以预测和控制, 因此需要探索新的手段对云服务环境下松散耦合的服务过程进行有效集成.

2 问题描述

从角色和资源角度, 研究群智协同的服务聚合优化问题. 给定一个跨组织任务分配项目 PRJ 可以描述为四元组 $PRJ = [T, R, X, G]$, PRJ 由一组任务 T 组成, $T = \{t_1, \dots, t_m\}$, 其中每个任务 t_i 至少需要分配一个角色 $R = \{r_1, \dots, r_t\}$, $X_i = \{x_1, \dots, x_n\}$ 表示角色 r_j 完成任务 t_i 需要的一组技能; 若存在一组候选执行者群体 $G = \{g_1, \dots, g_u\}$, 其中群组 g_w 具备一组技能 $Y_w = \{y_1, \dots, y_v\}$. 通常在特定的环境中工作时, 每个群组中的用户在任务中都扮演一个或多个角色. 基于群组角色识别的群智协同服务模型优化问题就是找到一个可行的跨组织任务分配, 使得群组协作性能最优, 且交互成本最小.

通常在云制造服务平台中, 根据用户请求任务的复杂度将任务分解为多个子任务, 并按照一定的逻辑顺序构成任务组合执行序列^[20,21]. 因此, 在大规模社会网络中构建跨组织业务过程, 需要根据任务执行者的资源属性分析组织业务过程, 并进行服务聚合以完成复杂项目.

3 基于群组角色协同的服务聚合

本文提出了一种基于角色的协同服务过程聚合优化方法,它利用 BPaaS 理论和语义 Web 技术,从角色和资源的角度出发,研究基于群智协同的服务过程模型优化,主要包括三个方面的研究内容:(1) 基于 SNA 协同社会业务过程聚合分析;(2) 支持群组协同的多角色匹配计算,提出一种测量群组协作性能和交互成本的计算方法;(3) 提出融合结构信息和语义信息的任务角色层次聚类算法,实现服务聚合优化。

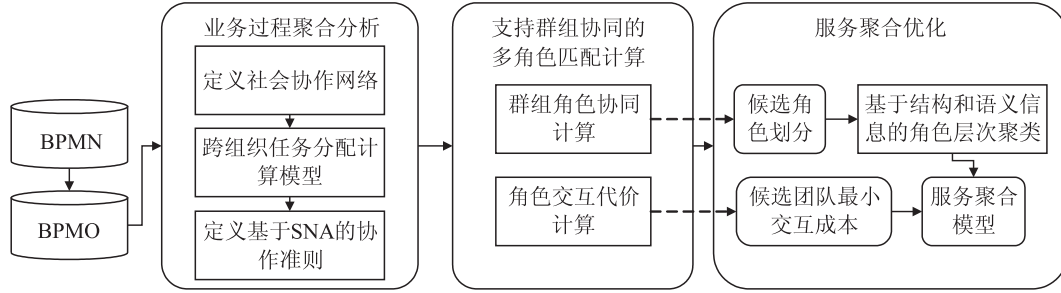


图1 基于角色的协同服务过程聚合优化方法

Fig. 1 Service aggregation based on group role collaboration

3.1 群组角色协同模型

在跨组织合作网络构建中,来自不同团队的两个候选执行者之间的业务技能相似性越高,且交互成本越低,其合作可能性越大。通常,服务聚合是通过业务过程的聚合来实现的。利用社会网络发现和再发现的方法研究面向群智协同的服务聚合,将具有相似背景和紧密关联的执行者构建成一个联盟。

定义1 业务协作群组。业务协作群组是由一组业务角色需求驱动的服务间的交互行为有向图,可以抽象为一个三元组 $G_c = (T, S, E)$, 其中, T 表示任务集合; $S \subseteq T \times T$ 是完成任务需求时连接任意两个任务之间的交互行为序列,表示任务角色分配中具有直接关联的执行者之间的关系集合,其中, w 表示交互行为关系的权重; $E: TR$ 表示一个映射函数,是指为每个任务 $t_i \in T$ 指定一个角色标签 $r_i \in R$ 。每一个群组都有一个质心节点,通过邻接矩阵识别其有向关系。

定义2 角色代理。在一个群组 $g \subset G_c$ 中,给定一组角色 $R = \{r_1, \dots, r_p\}$, 每一个角色都可以表示为 $r_i = (X_r, L_r)$, 其中, $X_r = \{x_1, \dots, x_n\}$ 表示一组属性(即技能)集合; L_r 表示角色等级的权重向量, $L_r[i] \in [0, +\infty]$, $0 \leq i \leq p$ 。在群组 g 中,每个角色代理(Role-Agent)可以分配有一个或多个角色,表示为 $(A, \{r_1, \dots, r_j\})$ 。每个用户或执行者都可以看作为一个代理,即一个代理代表存在于协作系统中的执行者。

在协同服务过程中,每个过程实例都由多个执行者参与,每个执行者可能拥有一种或多种技能,因而可能一个参与者执行过程实例中的几个事件活动。针对群组多角色分配(group multi-role assignment, GMRA)问题进行了研究,提出基于语义和结构信息的距离度量方法,来建立执行者(个人或组织团体)之间的关联。

根据定义1,将协同服务过程定义为一个业务协作网络 (G_c, C_p) , 其中, C_p 表示社会网络模型的语义信息。在语义过程模型中,语义信息 C_p 表示业务过程模型中每个执行者的特征属性,包括:执行者执行的任务集合,以及每个执行者的泳道(Lane)成员集合。因此,定义了一个角色分配的多元特征矩阵 F (nodes \times features) 来描述执行者与活动或泳道之间的关系,其中每一行表示业务过程中的每个执行者的特征,称为实例;每一列表示活动或泳道。其中 $F(i, j) \in \{0, 1\}$ 表示执行者 A_i 是否能够执行任务活动 T_j , 举例如表1。

$$F(i, j) = \begin{cases} 1 & \text{如果执行者 } A_i \text{ 隶属于活动 } T_j \text{ 或者泳道 } L_j \\ 0 & \text{否则} \end{cases} \quad (1)$$

给定一个跨组织的任务分配项目 PRJ 是一个长度为 n 的向量,它包含 n 个待分配的任务, $PRJ = \{T_1, \dots, T_n\}$, 其中 T_i 表示一个任务, $T_i = \{x_1, \dots, x_m\}$ 表示该任务需要 m 种技能 x_i 。给定一个角色代理向量 $A_j = \{p_1, \dots, p_i\}$, 其中 p_i 表示基于

表1 社会网络模型中语义信息 c_p 举例

Table 1 Example of semantic information c_p in social networks model

Nodes	Features						
	T_1	T_2	\dots	T_m	L_1	\dots	L_m
A_1	1	0	\dots	0	1	\dots	1
A_2	0	1	\dots	0	0	\dots	1
\dots	\dots	\dots	\dots	\dots	0	\dots	0
A_n	0	1	\dots	1	1	\dots	0

语义描述的执行者概要. 为了寻找任务 T_i 的候选角色代理 A_j , 通过余弦相似距离 $R(A, T)$ 计算任务向量和角色代理向量之间的语义相似性, 再将相似距离归一化为 softmax 函数计算的后验概率 $P(A^+ | T) \in [0, 1]$, 计算如下:

$$P(A^+ | T) = \frac{\exp(\gamma R(A^+, T))}{\sum_{A' \in A} \exp(\gamma R(A', T))}, \quad (2)$$

$$R(A, T) = \cos(A_i, T_j) = \frac{A_i^T T_j}{\|A_i\| \|T_j\|}, \quad (3)$$

式中, γ 表示平滑因子, A^+ 和 A' 分别表示正负样本, A 表示全部样本. 如果 $P(A^+ | T) > \lambda$, 则 A^+ 是该任务的一个候选者.

3.2 角色交互成本计算

针对上一节中获得项目的任务候选执行者集合, 需要计算交互成本来评估协同服务项目的预算, 达到交互成本最小化. 给定一组基于多角色匹配的项目候选执行者集合, 在组织内部或跨组织的合作者之间存在互操作, 因此, 这些候选者可以看作是网络节点, 互操作可以看作是边, 互操作的频率可以计算边的权重. 因此, 根据定义 1, 能够得到由候选者组成的业务协作子群图. 根据 SNA 准则^[22], 候选执行者之间的交互成本计算如下:

(1) 根据核心-外围准则, 核心节点代表权威和领导者. 通常, 核心或领导节点与其他人合作并具有良好的关联性. 通过接近中心性 C_v 可以度量节点 v 和 i 之间的关联性,

$$C_v = \frac{|U_c| - 1}{\sum_{i \neq v} d(v, i)}. \quad (4)$$

根据 C_v 的值, 选择 $Top-k$ 的领导者作为领导者集合, k 是由划分的复杂项目中子项目的数量来决定的.

(2) 根据协作准则, 复杂项目需要领导者带领团队开展协作, 因此交互成本可以表示为候选节点与领导者节点之间最短路径的加权之和, 其公式为:

$$G_c Cost = \sum_{i=1}^{|N_c|} \sum_{j=1}^k d(n_i, n_{leaderj}), \quad (5)$$

$$d(n_i, n_{leaderj}) = \frac{p(r_i, r_{leaderj} | t)}{p(r_i | t) p(r_{leaderj} | t)}, \quad (6)$$

式中, $|N_c|$ 表示候选执行者的数量, $d(n_i, n_{leaderj})$ 表示候选执行者 n_i 和领导者 $n_{leaderj}$ 之间最短路径上的边权重 ω 之和, ω 可以通过候选人 n_i 的角色 r_i 和领袖 $n_{leaderj}$ 的角色 $r_{leaderj}$ 之间的交互来度量, 它可以表示为相同任务 t 下的条件概率, $p(r_i | t)$ 表示角色 r_i 参与任务 t 的概率, $p(r_i, r_{leaderj} | t)$ 表示角色 r_i 与 $r_{leaderj}$ 共同在任务 t 中协作的概率.

4 服务聚合优化方法

4.1 聚合评价指标

在 GMRA 中, 群组绩效是衡量团队完成某一复杂任务的重要指标, 是单个候选执行者在群体中分配角色绩效的简单总和. 因此, 本文根据跨组织协同服务过程模型优化的计算要求, 将复杂任务分解为若干子服务 (Sub-service), 利用角色匹配和执行者之间的交互代价计算方法进行服务子过程的组合.

定义 3 候选匹配矩阵. 候选匹配矩阵 Q_m 是一个 $m \times n$ 矩阵, 其中, $Q_m[i, j] = \xi \in \{0, 1\}$ 代表角色代理 $i \in \mathbb{N} (0 \leq i \leq m)$ 被分配给角色 $j \in \mathbb{N} (0 \leq j \leq n)$ 之间的匹配值, $Q_m[i, j] = 0$ 表示匹配率最低, 相反, $Q_m[i, j] = 1$ 则表示匹配率最高, 非负整数 $m = |A|$ 和 $n = |R|$ 分别表示角色代理集合 A 和角色集合 R 的大小, \mathbb{N} 表示一个非负整数集合.

根据公式 (2), 角色代理 A_i 与角色 R_j 的匹配值计算如下:

$$Q_m[i, j] = P(A_i^+ | R_j). \quad (7)$$

根据文献[23]中对群组角色分配的形式化定义, 限定矩阵 Q_m 描述了角色代理集合 A 与角色集合 R

之间的匹配关系.

$$\begin{array}{ccc}
 \begin{bmatrix} 0.26 & 0.53 & 0.89 & 0.19 \\ 0.06 & 0.67 & 0.42 & 0.31 \\ 0.72 & 0.98 & 0.25 & 0.46 \\ 0.84 & 0.01 & 0.16 & 0.51 \\ 0.07 & 0.52 & 0.43 & 0.90 \end{bmatrix} & \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 2 \\ 3 \\ 1 \\ 1 \end{bmatrix}^T \\
 \text{(a) 限定矩阵 } \mathbf{Q}_m & \text{(b) 角色分配矩阵 } \mathbf{M} & \text{(c) 角色范围向量 } \mathbf{L}_r
 \end{array}$$

图 2 矩阵示例

Fig. 2 Matrix examples

举例如图 2 所示,假设存在 5 人的候选执行群组,一个项目仅需要 4 人参加,因此得到了具有 4 个角色和每个候选执行者的能力值的二维限定矩阵,每一行代表候选执行者,每一类代表角色.事实上, \mathbf{Q}_m 是描述角色和任务执行者之间的特定量化关系,它对 RBC 的角色分配问题有显著影响,即执行者的角色评价. 给定 $m \times n$ 角色分配矩阵 \mathbf{M} , $\mathbf{M}[i, j] \in \{0, 1\}$ ($0 \leq i \leq m, 0 \leq j \leq n$) 表示执行者 i 是否被分配了角色 j , 1 表示“是”, 0 表示“否”. 考虑当前和潜在的角色分配,一个角色代理可能扮演一个或多个角色,一个角色可能需要一个或多个角色代理,因此,给定 n 维角色范围向量 \mathbf{L}_r , $\mathbf{L}_r[j] = \sum_{i=0}^{m-1} \mathbf{M}[i, j]$, $\mathbf{L}_r[j] \in \mathbb{N}$, 同时给定 m 维角色限定向量 \mathbf{L}_r^a , $\mathbf{L}_r^a[i] \geq \sum_{j=0}^{n-1} \mathbf{M}[i, j]$, 表示每个代理只能被分配给有限数量的角色.

4.1.1 群组绩效优化

由若干执行者及其多角色构成的群组 g 的群组绩效 σ 可定义为候选执行者的角色匹配值的内积总和, 记为 $\sigma = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \mathbf{Q}_m[i, j] \times \mathbf{M}[i, j]$, $\sum_{i=0}^{m-1} \mathbf{M}[i, j]$ ($0 \leq j \leq n$) 表示每个角色的可执行性. 本算法使用了两个阈值 $\tau_0 \in [0, 1]$ 和 $\tau_1 \in (0, 1)$ 分别表示角色代理合理的极限和候选执行者集合的匹配范围, 其中 $\tau_1 < \tau_0$. 如果 $\mathbf{Q}_m[i, j] \geq \tau_0$, 则代理 i 符合角色 j , 反之不符合; 如果 $\mathbf{Q}_m[i, j] \geq \tau_1$, 则候选执行者 i 有资格参与角色 j , 反之不符合 ($0 \leq i \leq m, 0 \leq j \leq n$). 给定限定矩阵 \mathbf{Q}_m 、角色范围向量 \mathbf{L}_r 、角色限定向量 \mathbf{L}_r^a 和阈值 τ_0 , 基于群组角色匹配的任务分配问题就是寻找群组绩效 σ 的最优分配方案.

$$\sigma = \max \begin{cases} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \mathbf{Q}_m[i, j] \times \mathbf{M}[i, j], & \text{if } \mathbf{Q}_m[i, j] \geq \tau_1 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

满足条件(9)-(12),

$$\mathbf{T}[i, j] \in \{0, 1\} \quad (0 \leq i \leq m, 0 \leq j \leq n) \quad (9)$$

$$\sum_{i=0}^{m-1} \mathbf{M}[i, j] = \mathbf{L}_r[j] \quad (0 \leq j \leq n) \quad (10)$$

$$\sum_{j=0}^{n-1} \mathbf{M}[i, j] \leq \mathbf{L}_r^a[i] \quad (0 \leq i \leq m) \quad (11)$$

$$\mathbf{Q}_m[i, j] \times \mathbf{M}[i, j] > \tau_0 \times \mathbf{M}[i, j] \quad (0 \leq i \leq m, 0 \leq j \leq n) \quad (12)$$

4.1.2 交互成本优化

一方面, 候选执行者与角色之间的匹配运算能够反映出多角色协同分配的匹配程度; 另一方面, 候选执行者之间的交互成本计算也能够反映出群组协同的可操作性.

定义 4 交互成本矩阵. 交互成本矩阵 \mathbf{Q}_c 是一个 $m \times m$ 矩阵, 根据公式(5)协作准则量化方法, $\mathbf{Q}_c[i, j] = d(n_i, n_j)$, ($0 \leq i \leq m, 0 \leq j \leq m$), 表示代理 $i \in \mathbb{N}$ 与代理 $j \in \mathbb{N}$ 之间的交互成本, 即两者之间协作的最短路径权重.

从群体协同的交互成本考虑, 给定交互成本矩阵 \mathbf{Q}_c 、角色范围向量 \mathbf{L} 和阈值 τ_1 , 基于群组角色匹配的

任务分配问题的最优解必须满足交互成本最小的约束条件:

$$\delta = \min \left\{ \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} Q_c[i, j] \times M[i, j] \right\}. \quad (13)$$

综上,基于多角色分配的跨组织任务分配问题就是在满足双目标约束的条件下,找到一个可行的角色分配矩阵 M ,使其满足群组绩效 σ 最优和交互成本 δ 最小. 通常双目标之间会存在冲突,为了实现这两个目标之间的平衡,本文将两个目标转换成一个综合效用函数,即单位交互成本下的团队效益最大化,因此将上述双目标问题转换如下:

$$\max \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} Q_m[i, j] \times M[i, j]}{\sum_{i=0}^{m-1} \sum_{j=0}^{m-1} Q_c[i, j] \times M[i, j]} \quad (14)$$

满足条件(9)-(12).

4.2 协同服务聚合优化求解

在前述工作的基础上,基于群组角色协同的服务聚合优化求解分为两个阶段进行:

首先,融合结构和语义信息对群组角色进行层次聚类,通过角色划分获得执行者之间的关系,能够反映出具有良好合作关系和相似执行者的潜在合作团体.

其次,根据角色层次聚类的群组角色划分结果,对业务过程图进行语义标注,根据多种语义度量方法来评估多角色分配,并找到合适的候选角色. 当跨组织匹配的候选团队满足最大群组绩效和最小的交互成本的双目标约束时,即满足跨组织业务过程的分配目标.

4.2.1 融合结构和语义信息的角色层次聚类算法

本文通过融合结构和语义信息的层次聚类算法寻找服务候选角色,提高服务聚合的质量,解决群组绩效和交互成本的双目标约束优化问题. 融合结构和语义信息聚类过程如图 3 所示.

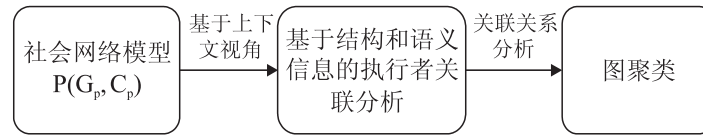


图 3 融合结构和语义信息的聚类过程

Fig. 3 Clustering based on structure and semantic information

首先,从社会实体之间的关系角度,重点关注业务过程的结构化特征和社会网络的非结构化特征,进而计算任务执行者之间的语义相似性,利用社会网络模型的语义信息 C_p ,寻找执行者的从属关系,这种从属关系包含了相似执行者的群组,它是由一个被称为“任务-泳道(Task-Lane)”所表示的语义距离 D_s 计算而获得的.

本文前期研究成果文献[24,25]在业务过程建模本体(BPMO)以及云业务过程本体(CloudPrO)^[24]的基础上构建了服务过程建模本体(SPMO).

在这种情况下,基于服务协作网络的结构化属性和基于执行者与任务之间关系的非结构化属性,都能够被 SPMO 所涵盖.

利用欧式距离计算执行者之间的语义相似性,并为 SPMO:Resource、SPMO:Action、SPMO:Link 概念之间的相似性度量赋予不同的权重,分别用 α, β, γ 表示. 具体计算如下:

$$D_s = \alpha l[(i), (j)]_{resource} + \beta l[(i), (j)]_{action} + \gamma l[(i), (j)]_{link}. \quad (15)$$

其次,在结构信息方面,根据任务执行者的图结构性特点,计算任意一对执行者之间的结构距离 D_f . 通过将结构信息与语义信息进行集成,计算语义距离 D_s 与结构距离 D_f 的乘积来表示关联执行者之间的工作流在工作流总量中的占比,形成执行者向量的距离矩阵,作为层次聚类算法的输入,从而产生了新的执行者与泳道的从属关系.

最后,利用层次聚类算法^[25]对服务过程中的任务执行者进行聚类,根据执行者特征的语义相似性及其之间的关系生成新的执行者的群组划分. 融合结构和语义信息的角色层次聚类算法具体过程如算法 1 所示.

算法 1 融合结构和语义信息的层次聚类算法

输入:用户协作群组 $G_c, G_c = \{g_1, \dots, g_n\}$;
输出: G_c 中的基于相似角色代理的群组划分 $G'_c = \{g'_1, \dots, g'_n\}$;
1:初始化:不相交聚类的层次为 $L(1) = 0$,序号为 $m = 0$;
2:根据式 (15)–(17),计算 G_c 中任意两个角色代理之间的语义欧氏距离 D_s 、结构距离 D_F 、总距离 D ;
3:寻找当前群组中距离最小的一对簇 (r) 和 (r') , $l[(r), (r')] = \min\{D\}$;
4:增加序号: $m = m + 1$. 将簇 (r) 和 (r') 合并成单个簇,形成下一个簇 m ,设置该簇的级别为 $L(m) = l[(r), (r')]$;
5:根据式 (18) 计算群组惯性因子 I ,重新计算质点;更新距离矩阵 D ,删除簇 (r) 和 (r') 对应的行和列,添加新形成的簇对应的行和列. 计算新簇 (r, r') 与旧簇 (k) 之间的距离: $l[(k), (r, r')] = \text{meanavg}[l[(k), (r)], l[(k), (r')]]$;
6:如果所有的角色代理结点都在一个群组中,那么停止,否则重复第 2 步.

为了说明上述算法思想,本文使用“软件服务交叉功能过程”为例进行描述,如图 4 所示,其中包含 7 个任务执行者、10 个任务,其业务过程中执行者和泳道的任务分配如表 2 所示, ID-Task 表示执行任务, ID Task-Lane 表示每个执行者的泳道成员.

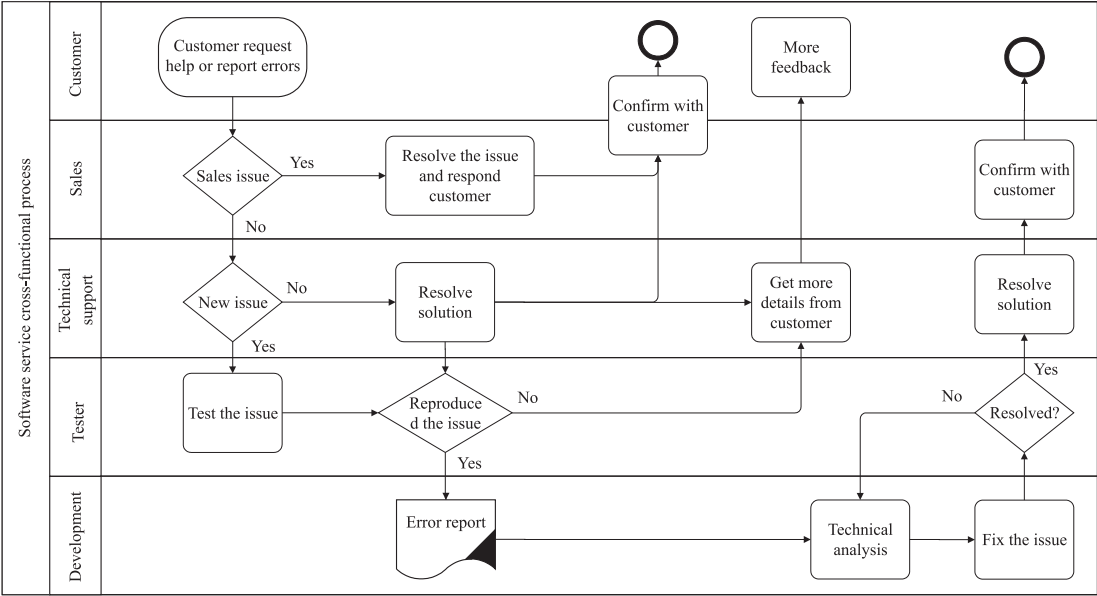


图 4 软件服务交叉功能过程图

Fig. 4 Software service cross function process diagram

表 2 角色和泳道的任务分配表

Table 2 Task assignment of role & lane

Task Label	ID Task	ID Actor	ID Task-Lane
Customer request help or report errors	T1	A1	L1
Resolve the issue and respond customer	T2	A1	L2
Confirm with customer	T3	A3	L1, L2
Resolve solution	T4	A1, A2, A3	L3
Get more details from customer	T5	A1, A3	L3
More feedback	T6	A2, A5	L1
Test the issue	T7	A4, A5, A6	L4
Error report	T8	A2, A4	L5
Technical analysis	T9	A4, A5, A6	L5
Fix the issue	T10	A7	L5

根据式 (1), 得到相应的“活动-执行者-泳道”二元从属矩阵, 从属关系网络模型在图形上表示为二部图, 如表 3 所示.

表 3 “活动-执行者-泳道”二元从属矩阵
Table 3 “Activity-Actor-Lane” binary dependency matrix

Actor	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	L1	L2	L3	L4	L5
A1	1	1	0	1	1	0	0	0	0	0	1	0	1	0	0
A2	0	0	0	1	0	1	0	1	0	0	1	0	1	0	1
A3	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0
A4	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1
A5	0	0	0	0	0	1	1	0	1	0	0	0	0	1	1
A6	0	0	0	0	0	0	1	0	1	0	0	0	0	1	1
A7	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1

根据表 3,利用欧氏距离计算出两个执行者之间的语义距离 D_s ,如表 4 所示。
根据任务执行者的图结构性特点,任意一对执行者之间的结构距离 D_F 计算如下:

$$D_F = \frac{1}{N_F + 1} - \frac{1}{N_{FT} + 1}, \quad (16)$$

式中, N_F 表示直接从一个执行者发送到另一个执行者的工作流的总数, N_{FT} 表示模型中工作流的总数. 结构距离 D_F 表示任务如何在执行者之间流转而产生的距离. 分母上加 1 是为了避免除数为 0.

根据 D_s 和 D_F ,可以得到如下计算总距离:

$$D = D_s \times D_F + \varepsilon. \quad (17)$$

式中, ε 是为了避免零距离的情况, 即当 D_F 或 D_s 其中有一个为 0 时, $\varepsilon = 0.1$.

综上,根据公式(17),可以计算得到执行者向量之间的总距离 D ,如表 5 所示.

在层次化聚类中,使用类内惯性因子(intra-class inertia)的方式实现同构不同类的分类目标:

$$I = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{c_1, c_2 \in g_i} d^2(c_1, c_2)}{\binom{|g_i|}{2}},$$
$$\binom{|g_i|}{2} = \frac{|g_i|!}{(|g_i|-2)! \times 2!}.$$

4.2.2 群智协同服务聚合优化算法

在 4.2.1 节中,融合结构和语义信息的角色层次聚类算法根据执行者特征的语义相似性及其之间的关系提出基于群组角色协同的服务聚合优化算法(GRC-based service aggregation optimization, GSAO),主要解决群组绩效和交互成本的双目标约束优化问题,从而进一步提高跨组织业务过程的服务聚合计算效果.

算法 2 基于群组角色协同的服务聚合模型优化算法

输入:任务集合 $T = \{t_1, \dots, t_m\}$, 基于相似角色代理的群组划分 $G'_c = \{g'_1, \dots, g'_n\}$;
输出:最优角色分配矩阵 M^* ;
1:遍历任务集合 T 中每一个任务 t_i 对应的角色集合 $R, r_j \in R$;
2:遍历群组集合 $G'_c, a_i \in g'_i$;
3:根据式(7)计算得到角色限定矩阵 Q_m ;
4:构造角色分配矩阵 M ;
5:根据式(8)-(12)计算群组绩效 σ ;
6:根据式(13)计算群组交互成本 δ ;
7:根据式(14)计算 σ/δ ;
8:求取 σ/δ 的最大值;
9:返回最优角色分配矩阵 M^* .

根据算法 1 得到群组 g 中的角色代理划分结果,使得具有良好协作关系的任务执行者划分在同一子群组中,为跨组织业务过程的服务聚合提供了较优的初始业务协作群组集合. 在算法 2 中,综合计算跨群组下执行者与任务角色、执行者之间的群组绩效和交互成本,当跨群组匹配的候选团队满足最大群组绩效和最小的交互成本的双目标约束时,即满足跨组织业务过程的分配目标. 在上述算法中,领导者节点的计算需要的时间复杂度为 $O(n+m)$,返回一个加权服务过程图耗费的时间复杂度为 $O(nm+n^2\log n)$,其中 n 为节点数, m 为边数. 在确定项目预算的约束条件下,本文提出的基于群组角色协同的服务聚合模型优化算法在成本最小化的同时具有最佳性能. 通过改变阈值 τ_0 和 τ_1 ,根据角色代理在任务分配中角色的匹配程度可以搜索出任务分配的最优解. 假设 n 是技能的数量, r 是项目中包含的角色的数量, m 是任务的候选个体(执行者)的数量. 那么,本节提出的算法需要在 $O(mr+nr)$ 时间内运行. 因此,本文提出的算法是可行的,并且运行效率更高.

5 实验评估

5.1 实验设置

针对基于群组角色协同的服务聚合模型优化问题,我们进行了仿真实验,验证了该方法处理实际问题的有效性. 本实验从三个方面进行算法评估:基于 SNA 的语义度量对协作过程的影响,层次聚类算法的惯性因子影响,以及基于角色识别的跨组织协作性能.

本实验使用 DBLP 数据作为构建社会图的实验数据集,包括专家集合,以及两个专家作为节点之间的边集合,两个专家之间的通信成本作为边的权重. DBLP 社交网络中的专家节点描述了若干个人信息,包括技能、工作、从属关系和职位等. 本实验选择了两个协作群组 groupA(见图 5(a))和 groupB(见图 5(b))进行仿真实验. 每组都有 30 名成员,其中 groupA 有 43 条专家之间的关系,groupB 只有 30 条.

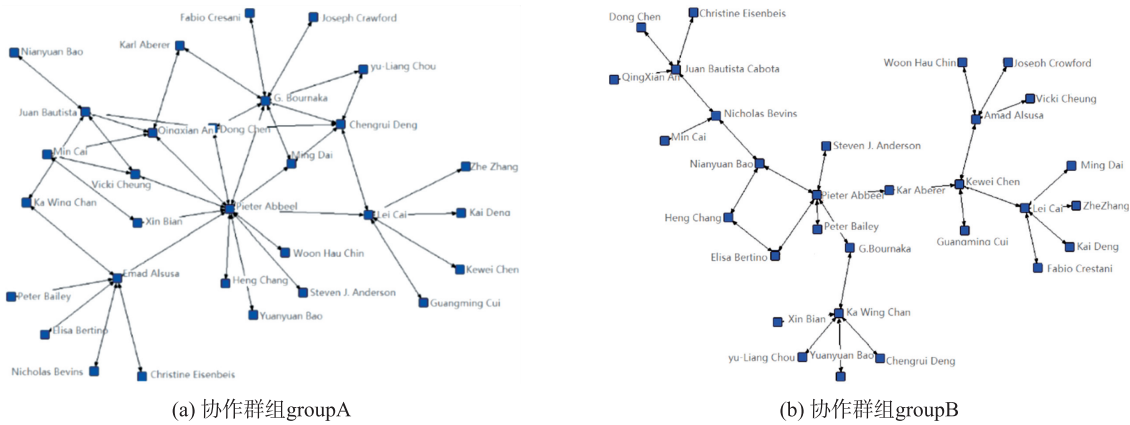


图 5 协作群组社会网络图结构

Fig. 5 Collaborative group's social networks graph

5.2 评价方法

通过比较 groupA 和 groupB 中的共识质量,评价基于 SNA 的语义度量对协同质量的影响.

$$\hat{d}(x, X) = 1 - \frac{d(x, X)}{\text{card}(X)}.$$

群组协作质量将随着群组共识质量的提高而提高,它可以基于 SNA 和本节的语义方法对协作过程进行评价. 为评估角色识别的准确性,实验将识别潜在合作群体的结果与专家确认结果进行比较. 基于概念 c 的候选专家特征向量可以表示为:

$$F = (id, v) \quad (20)$$

式中, id 表示语义概念库中实例的唯一标识符, v 表示专家概要实例的值. 此外, F_e 表示经专家验证的正确标注,因此,将第 i 组角色识别的准确性定义为:

$$\text{accuracy}(F_i, F_e) = \text{similarity}(F_i, F_e) = \frac{\sum (F_i \times F_e)}{\sqrt{\sum (F_i)^2} * \sqrt{\sum (F_e)^2}}. \quad (21)$$

5.3 实验验证与性能分析

5.3.1 基于 SNA 协作过程语义评价

图 6 显示了各群体的平均共识质量结果,基于服务过程中的资源描述,使用语义概念对执行者属性进行语义标注,因此提高了协作过程中的共识质量.如图 6 所示,groupA 比 groupB 收敛更快,且具有更好的共识质量.因为 groupA 的密度比 groupB 大,即 groupA 中节点之间具有更多的纽带和联系,使前者比后者有更多的合作机会,能够产生更多的合作关系.因此,groupA 具有更好的共识质量.

密度中心性通常被用于评价基于 SNA 的协作效果.如图 7 所示,groupA 节点之间的距离比 groupB 节点之间的距离更近.一个节点的密度中心性越大,它的权威性就越高,因此密度中心性越大的成员在完成项目时就会有更多的协作可能性,从而得到更多的项目支持与合作.

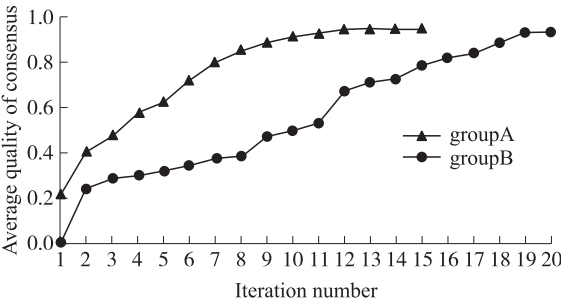


图 6 协作群体的平均共识质量结果

Fig. 6 Consensus quality for collaborative groups

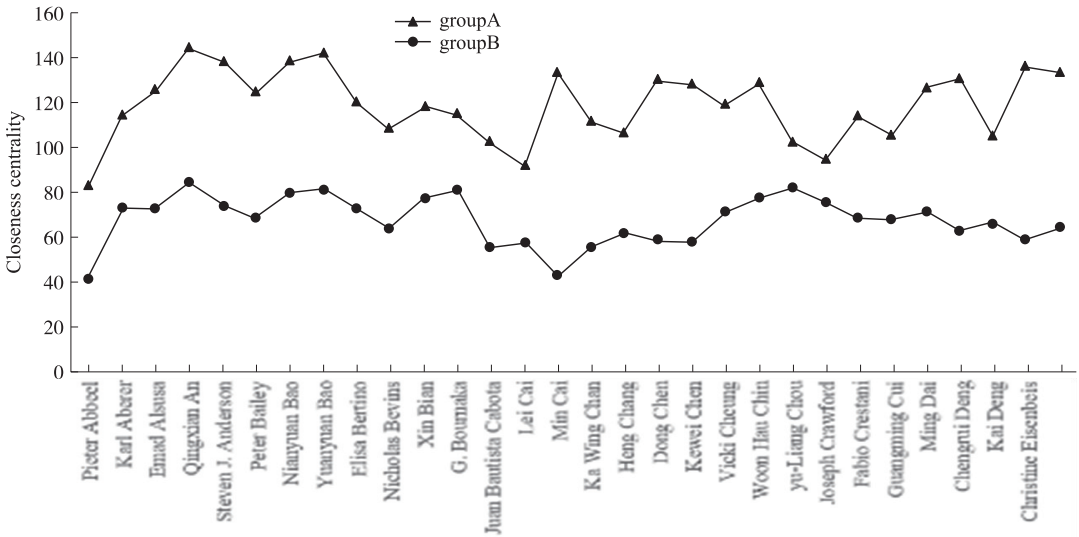


图 7 groupA 和 groupB 的密度中心性比较

Fig. 7 Comparison of closeness centrality of groupA and groupB

5.3.2 层次聚类算法的惯性因子影响

为了评估获得的惯性阈值,本文将层次聚类算法应用于 DBLP 中 50 个业务过程模型,并计算每种情况下的惯性因子.算法 1 的迭代过程如图 8 所示,可以看到连续两步迭代的惯性因子的差值从第 5 步到第 6 步逐渐增大,惯性因子在最后一步达到最大值.

实验结果表明,惯性因子在某一时刻(t)与($t-1$)的差值 $I^{(t)} - I^{(t-1)} \leq \varepsilon$,不能超过 $\varepsilon = 0.015$.在惯性因子阈值不超过 0.015 的 46 个模型中,聚类效果最好.

5.3.3 算法性能评价

为了进一步验证本文提出算法 2 的效率和有效性,针对跨组织交互成本评价算法进行了规模实验.实验具体参数如下:(1)每次规模实验均执行 20 次;(2)Agent(执行者)集合的容量 $m \in \mathbb{N}, 30 \leq m \leq 500$ 集合容量递增步长设为 20;(3)实验规模 $k \in \mathbb{N}$ 为随机数, $3 \leq k \leq 10$,且角色集合大小 $n = m/k$;(4)分配给 Agent 的角色限制 L_r^a 为 1 至 5;(5)受限制的角色等级 L_r 需进行归一化处理,计算如下:

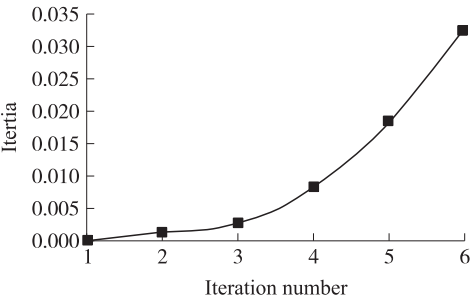


图 8 迭代过程中的惯性曲线

Fig. 8 Inertia curve during iteration

$$L'_r = \frac{L_r - \min L_r}{\max L_r - \min L_r}. \quad (22)$$

因此, L'_r 的取值范围为 $[0, 1]$; (6) 根据 ROC 曲线, 阈值 τ_0 和 τ_1 能够表示预测算法的性能, 得到最优阈值 τ_0 和 τ_1 . 本实验选取了 7 000 多个包含匹配值和预测样本的主题组成测试集, 每个 Agent 的对应的阈值 τ_0 精确到: 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, τ_1 的阈值精确到: 0.04, 0.13, 0.17, 0.21, 0.05, 0.12, 0.23.

本实验选取 5 个任务进行评价, 将所提出的模型 (GSAO) 与基于穷举搜索的群组角色分配算法 (GRA-ES)、基于线性规划的群组角色分配算法 (LPBA)、基于遗传算法的柔性编排群组角色分配算法 (GRAFF-GA) 进行对比分析, 如图 9 和图 10 所示.

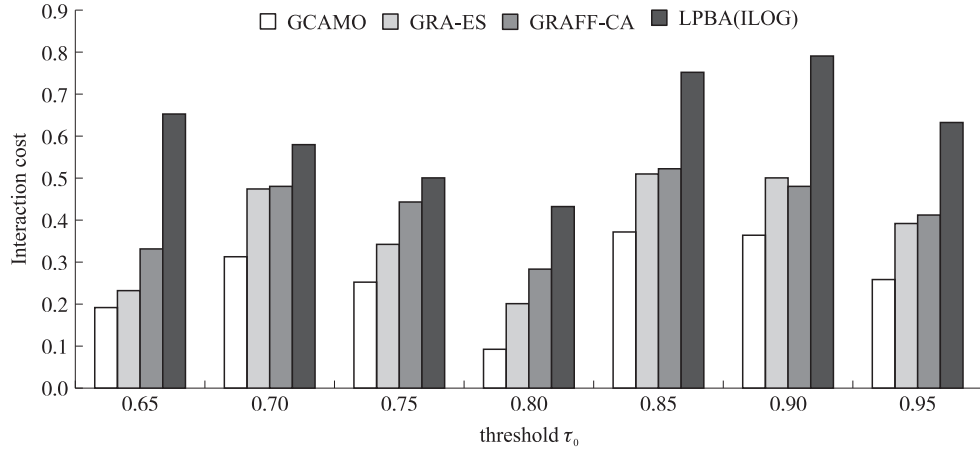


图 9 阈值 τ_0 下的候选执行者的交互成本比较

Fig. 9 Comparison of interactive cost with threshold τ_0

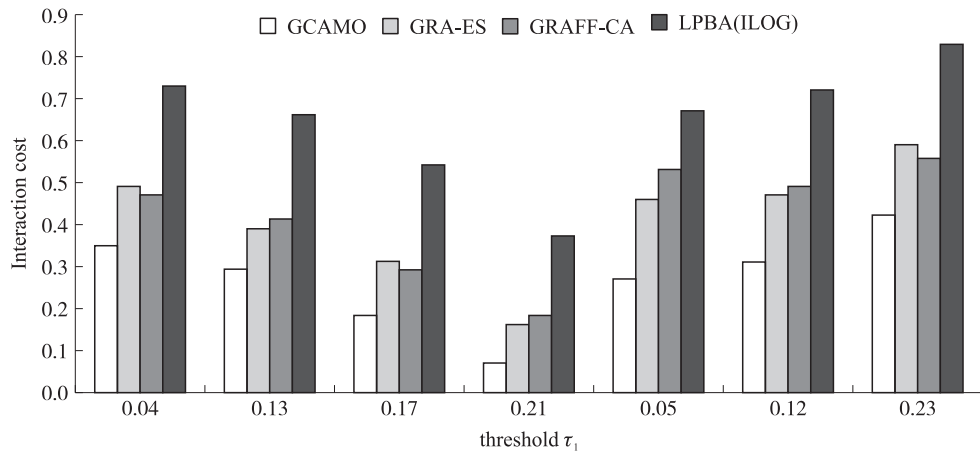


图 10 阈值 τ_1 下的候选执行者的交互成本比较

Fig. 10 Comparison of interactive cost with threshold τ_1

协同服务的交互成本受 τ_0 和 τ_1 变化的影响, 其取值范围为 $[0, 1]$. 在成本约束的条件下, 该模型会通过阈值变化来改变交互成本, 因此比其他方法在复杂社会协同服务组合应用中获得更好的效果. 从图中可以看出, 当 $\tau_0 = 0.8$ 和 $\tau_1 = 0.21$ 时, 交互成本最小.

同时, 为了验证本文提出方法的有效性, 针对 DBLP 数据集, 本实验同样选择 7 000 多个包含匹配值和预测样本的主题组成测试集, 随机选择 $K\%$ ($K = 60, 70, 80$), 10% 以及 10% 的角色评价数据分别作为训练集、验证集合测试集. 最大迭代次数为 300, 每次规模实验均执行 10 次.

采用平均绝对误差 (MAE) 和均方根误差 (RSME) 作为评价指标. 实验对比结果如表 6 所示. 表 6 展示了本文聚合优化方法在数据集上的实验效果, 并且对最佳效果进行加粗显示. 可以发现, 本文方法明显优于其他对比方法, 有效验证了基于群组角色的服务聚合方法的优越性. GRA-ES 方法在不同的训练集中的效果都明显弱于其他方法, 这主要是由于穷举搜索算法未对角色特征信息进行聚类分析, GRAFF-GA 采

用了遗传算法的柔性编排,因此其效果与本文方法的效果相当. 本文所提出的 GSAO 方法主要考虑了针对受限制的角色等级和多群组共识质量的交互评价,允许每种角色 Agent 之间的交互作用,因此有效提升了服务聚合的效果.

表 6 相关算法 MAE 和 RMSE 比较
Table 6 Comparison among correlation algorithms on MAE and RMSE

训练集	评价指标	GRA-ES	LPBA	GRAFF-FA	GSAO
60%	MAE	0.943 7±0.014 6	0.886 9±0.012 5	0.899 2±0.017 3	0.765 1±0.047 8
	RMSE	1.302 1±0.023 6	1.328 2±0.030 4	1.213 3±0.035 3	1.169 8±0.039 7
70%	MAE	0.928 5±0.008 6	0.876 4±0.008 7	0.883 8±0.008 0	0.781 1±0.036 7
	RMSE	1.287 9±0.011 3	1.287 6±0.030 1	1.198 4±0.008 9	1.167 3±0.015 2
80%	MAE	0.917 3±0.010 9	0.867 5±0.009 1	0.876 1±0.022 4	0.776 3±0.020 5
	RMSE	1.278 3±0.015 4	1.245 3±0.026 3	1.187 9±0.021 1	1.141 8±0.011 5

6 结论

本文从角色和资源的角度出发研究群智协同的服务聚合问题,提出一种基于语义的群组多角色识别和支持语义分析的群智协同服务聚合优化方法. 首先,对跨组织任务分配问题进行形式化定义,设计了跨组织的任务分配计算模型,将协同服务聚合定义为多目标约束下的寻优问题;其次,通过基于角色协同的社会协作网络分析,制定了基于 SNA 的协作准则,提出一种用于测量群组协作性能和交互成本的协作社区和实体检测计算方法;第三,提出了融合结构信息和语义信息的任务角色层次聚类算法,丰富了基于语义的服务过程模型的抽象表示,并设计基于角色协同的群智协同服务聚合优化算法. 仿真实验分别从基于 SNA 的语义度量对协作群组的共识质量、层次聚类算法的惯性因子影响,以及基于角色识别的跨组织协作交互成本等性能进行了比较,证明了本文提出方法的有效性,为有效解决面向服务的跨组织业务管理和运营中的服务过程聚合优化提供了解决方案. 在未来的工作中,将进一步挖掘服务评价中的隐含信息,采用隐私感知的时间聚合机制来保存用户的敏感时间信息,研究基于隐含语义模型与时间聚合的可信服务推荐,进一步提高服务推荐的质量和精确度.

[参考文献]

[1] TAN W,ZHAO Y,HU X,et al. A method towards web service combination for cross-organisational business process using qos and cluster[J]. Enterprise information systems,2019,13(5):631-649.

[2] ZHAO Y,TAN W,JIN T. Qos-aware web service composition considering the constraints between services[C]//In Proceedings of the 12th Chinese Conference on Computer Supported Cooperative Work and Social Computing. Chongqing,2017:229-232.

[3] IBRAHIM G J,RASHID T A,AKINSOLU M O. An energy eficient service composition mechanism using a hybrid meta-heuristic algorithm in a mobile cloud environment[J]. Journal parallel and distributed computing,2020,143:77-87.

[4] HOSSEINZADEH M,HAMA H K,GHAFOUR M Y,et al. Service selection using multi-criteria decision making;a comprehensive overview[J]. Journal of network system manage,2020,28:1639-1693.

[5] DASTJERDI A V,BUYYA R. Compatibility-aware cloud service composition under fuzzy preferences of users [J]. IEEE transactions on cloud computing,2014,2(1):1-13.

[6] ASGHARI S,NAVIMIPOUR N J. Review and comparison of meta-heuristic algorithms for service composition in cloud computing[J]. Journal of multimedia processing,2016,4(4):28-34.

[7] VAKILI A,NAVIMIPOUR N J. Comprehensive and systematic review of the service composition mechanisms in the cloud environments[J]. Journal of network and computer application,2017,81:24-36.

[8] ZHAO Y,QU Y,CHEN F,et al. Data integrity verification in mobile edge computing with multi-vendor and multi-server[J]. IEEE transactions on mobile computing,2023,23(5):5418-5432.

[9] ZHAO Y,QU Y,XIANG Y,et al. Longterm over one-of:Heterogeneity-oriented dynamic verification assignment for edge data integrity[J]. IEEE transactions on mobile computing,2024,23(5):4601-4616.

[10] GAO T L,DUAN L,FENG L F,et al. A novel blockchain-based responsible recommendation system for service process

- creation and recommendation[J]. *ACM transactions on intelligent systems and technology*, 2024, 15(4): 1–24. DOI 10.1145/3643858.
- [11] ZHU Q L, FAN Y L, WANG S G. A fairness aware service recommendation method in service ecosystem[J]. *International journal of web and grid services*, 2023, 19(4): 427–445.
- [12] ZHANG F, CHEN B M, LIU C. Web service instant recommendation for sustainable service mashup[J]. *Sustainability*, 2020, 12(20): 8563. DOI 10.3390/su12208563.
- [13] YU D J, YU T, WANG D J, et al. Long tail service recommendation based on cross-view and contrastive learning[J]. *Expert systems with applications*, 2024, 238: 121957. DOI 10.1016/j.eswa.2023.121957.
- [14] ZHU H. Role-based collaboration and E-CARGO: revisiting the developments of the last decade[J]. *IEEE systems, man, and cybernetics magazine*, 2015, 1(3): 27–36.
- [15] 滕少华, 张红, 刘冬宁, 等. E-CARGO 模型在 CSP 问题中的描述[J]. *计算机科学*, 2015, 42(2): 241–246.
- [16] ZHANG M, YAN G, WANG Y, et al. Identifying vital nodes in social networks using an evidential methodology combining with high-order analysis[J]. *International conference of pioneering computer scientists, engineers and educators*, 2020(1): 110–117.
- [17] MAHYAR H, HASHEMINEZHAD R, GHALEBI E, et al. Identifying central nodes for information flow in social networks using compressive sensing[J]. *Social network analysis & mining*, 2018, 8(33): 1–24.
- [18] BOULMAKOUL A, BESRI Z. Scoping enterprise organizational structure through topology foundation and social network analysis[C]//3rd Edition on Innovation and News Trends in Information Systems, Tanger, Morocco, 2013.
- [19] HUANG L, TAN W, SUN Y. Collaborative recommendation algorithm based on probabilistic matrix factorization in probabilistic latent semantic analysis[J]. *Multimedia tools and applications*, 2019, 78: 8711–8722.
- [20] HAO Y S, FAN Y S, ZHANG J. Service recommendation based on description reconstruction in cloud manufacturing[J]. *International journal of computer integrated manufacturing*, 2019, 32(3): 294–306.
- [21] LIU Y, GARG S, NIE J, et al. Deep anomaly detection for time-series data in industrial iot: A communication-efficient on-device federated learning approach[J]. *IEEE internet things*, 2020, 8(8): 6348–6358.
- [22] HU Y, WU L, ZHANG L, et al. Review on Theory and method of cloud manufacturing service evaluation[J]. *Computer integrated manufacturing systems*, 2017, 23(3): 640–649.
- [23] ZHU H, ZHOU M C, ALKINS R. Group role assignment via a kuhn-munkres algorithm-based solution[J]. *IEEE transactions on systems, man, and cybernetics — Part A: systems and humans*, 2012, 42(3): 739–750.
- [24] TAN W, HUANG L, ZHAO L, et al. A Role-based semantic framework for collaborative socialized process model reconstruction[C]//Computer Supported Cooperative Work and Social Computing. ChineseCSCW, 2018.
- [25] HUANG L, ZHAO L, LIU Y, et al. Hierarchical service composition via blockchain-enabled federated learning[J]. *Data sci eng*, 2024, <https://doi.org/10.1007/s41019-024-00258-7>.

[责任编辑: 陆炳新]