

## 基于活动发生关系的流程相似性度量方法

李东月, 方欢<sup>†</sup>

(安徽理工大学 数学与大数据学院, 安徽 淮南 232001)

**摘要:** 针对现有的流程模型行为相似性度量方法大都基于变迁的发生序列, 存在时间复杂度较高的问题, 提出一种基于活动发生关系的流程相似性度量方法. 首先, 形式化左右集概念, 并引入违背度和左右集权重 $\alpha, \beta$ , 通过活动间的关系变化来确定 $\alpha$ 和 $\beta$ 的取值, 进而得到活动的相似性; 然后, 将流程中活动的相似性进行归一化, 提出基于活动发生关系的流程相似性算法(AOR), 进行业务流程的相似性度量. 最后, 使用实际事件数据案例对本文方法进行了评估, 实验结果表明: 该方法能够正确计算两个不同流程之间的相似性, 且具有较好的时空复杂度.

**关键词:** 流程相似性; 活动发生关系; 业务流程; Petri网; 算法评估

**引用格式:** 李东月, 方欢. 基于活动发生关系的流程相似性度量方法. 控制理论与应用, 2020, 37(9): 2011 – 2019

DOI: 10.7641/CTA.2020.90768

## An approach of process similarity measurement based on activity occurrence relationship

LI Dong-yue, FANG Huan<sup>†</sup>

(College of Mathematics and Big Data, Anhui University of Science and Technology, Huainan Anhui 232001, China)

**Abstract:** Aiming at the problem of high time complexity of the existing process model behavior similarity measurement methods, which are mostly based on the occurrence sequence of transitions, a kind of process similarity measurement method based on activities occurrence relationship is proposed. Firstly, the concept of left and right sets are formalized, and the degree of violation and the weights of left and right sets  $\alpha, \beta$  are introduced. According to the transformation relationship between activities, the values of  $\alpha, \beta$  are determined, and the similarities of activities are obtained. Then, the similarities of all activities in the processes are normalized, and further the processes similarity algorithm (named AOR) based on activities occurrence relationship is presented, which is used to measure the similarity of business processes. Finally, an actual event data case is used to evaluate the proposed method. The experimental results show that the proposed method can correctly calculate the similarity between two different processes, and has better space-time complexity.

**Key words:** process similarity; activity occurrence relationship; business process; Petri net; algorithm evaluation

**Citation:** LI Dongyue, FANG Huan. An approach of process similarity measurement based on activity occurrence relationship. *Control Theory & Applications*, 2020, 37(9): 2011 – 2019

### 1 引言

随着业务流程管理(business process management, BPM)在企业中的影响力越来越大, 流程模型在企业管理中得到了广泛的应用. 业务流程是对企业业务执行的一种描述, 对企业的运营具有重要的指导意义. 因此, 业务流程模型已经成为每个组织宝贵的智力资产, 企业不可缺少的一部分. 值得注意的是, 累积的流程模型已经达到了一个惊人的数字<sup>[1]</sup>, 管理动辄数以万计的流程模型已经成为人们面临的一个挑战. 流程

模型的相似性度量已经被证实是管理这些宝贵资产高效的解决方案<sup>[2]</sup>. 然而, 在处理现有业务流程的多样性时, 应根据用户需求, 选择一种适当的方法来分析业务流程的相似性. 相似性度量有3种观点: 1) 元素匹配相似度, 比较附加到流程模型上的节点和边的对应关系; 2) 比较元素标签与流程模型拓扑结构的结构相似性; 3) 比较过程模型中元素标签及其因果关系的行为相似性<sup>[3]</sup>. 其中, 流程模型的行为(执行语义)最能体现一个模型的本质.

收稿日期: 2019-09-11; 录用日期: 2020-04-30.

<sup>†</sup>通信作者. E-mail: fanghuan0307@163.com; Tel.: +86 13359018221.

本文责任编辑: 赵千川.

国家自然科学基金项目(61902002, 61402011, 61572035), 安徽省自然科学基金项目(1608085QF149), 安徽省高校优秀青年人才基金项目(gxyq ZD2018038), 安徽省博士后基金项目(2018B288)资助.

Supported by the National Natural Science Foundation of China (61902002, 61402011, 61572035), the Anhui Natural Science Foundation (1608085 QF149), the Anhui University Excellent Youth Talents Foundation (gxyqZD2018038) and the Anhui Post-doctoral Foundation (2018B288).

基于标签(元素匹配)的相似性方面往往以模式匹配和本体论为基础,例如基于标签对齐的相似性度量算法<sup>[4]</sup>和基于标签语义的相似性度量方法<sup>[1]</sup>。

现今对流程模型的相似性分析主要从行为相似性和结构相似性两个方面来展开。首先,在模型的行为相似性方面,基于行为的流程相似性度量方法主要包含以下两种:

1) 基于流程模型中活动间的时序关系,如执行顺序等:文献[5]提出了一种基于两个流程变迁紧邻关系(transition adjacency relation, TAR)集的相似性定义,对流程的相似性进行度量,该方法没有考虑活动紧邻关系的重要性,因此对顺序结构和循环结构的区分不敏感;文献[6]提出行为轮廓的概念,对比基于迹等价一致性与行为轮廓一致性的计算结果,突出行为轮廓相似性算法的优越性。文献[7]在文献[5]的基础上,将紧邻关系的重要性加以区分,提出一种基于变迁紧邻关系重要性的相似性算法TAR++。文献[8]提出一种基于任务之间的时间关系查询流程模型的方法,通过将截断事件连接到延续事件来扩展完全有限前缀,提取任务之间的所有时间关系,进而查询流程模型;文献[9]拓展变迁的紧邻关系,利用完全有限前缀,提出基于任务最短距离矩阵的流程模型行为相似性算法,但该算法需要计算两个流程模型的同维化矩阵,不能直接用于模型索引。文献[10]提供了一个基于迹语义和抽象的行为包含概念,根据网系统能否重放查询行为进行模型匹配,衡量模型的相似性,该方法只匹配了活动间的行为关系,没有考虑模型结构;文献[11]提出了一种基于业务流程内部结构的相似性度量方法,给出映射函数,通过计算单个节点(变迁和库所)的相似性,将所有变迁或库所的相似性加权求和,进而求得流程的相似性。文献[12]提出了基于Petri网的映射变迁关系相似性度量方法,给出5种变迁关系定义,利用变迁的行为关系,度量相同变迁节点在不同流程中结构上的相似性,计算流程相似性;文献[13]通过将模型的行为关系用变迁标签图来表征,结合图理论,提出一种基于可覆盖图编辑距离的行为相似性度量算法(transition-labeled graph edit distance, TAGER);文献[14]提出了一种基于包含领域知识的语义任务重要性紧邻关系(semantic task adjacency relations with importance, ISTARs)的语义工作流行为相似性度量方法。文献[15]利用Petri网的完全前缀展开理论和任务的发生关系,提出一种基于任务发生关系的流程模型相似性度量(task occurrence relation, TOR);文献[16]将行为轮廓中的交叉关系细化为6种类型,提出了细化的基于行为轮廓矩阵的相似性度量方法。

2) 基于流程模型的变迁发生序列集合<sup>[11]</sup>。文献[17]提出了一种基于主变迁序列的相似性度量方法(principal transition sequences, PTS),利用过程模型的

语义,将主变迁序列分成三类,分别定义每类主变迁序列的相似性计算方法,进而计算流程的相似性;该方法将完整的发生序列分开考虑,破坏了语义的完整性;文献[18]对文献[17]的方法进行研究,针对PTS方法存在的问题,利用完整的发生序列表征模型的行为,提出一种基于触发序列的流程相似分析方法PTS++;文献[19]提出了一种基于约束迹的过程相似性度量方法,以约束的最长公共子序列为基础,利用约束迹对齐计算流程的相似性。

其次,在流程模型的结构相似性研究方面,目前大多研究方法基于图的理论进行结构相似度计算<sup>[20-21]</sup>,而在模型结构转为图的过程中会丢失一些语义信息,导致无法有效区分图中活动之间的是选择关系还是并发关系,因此仅仅将结构转为有向图不能准确表现原有流程,必须结合一些流程行为信息,以更全面地表达原流程。为了弥补上述不足,文献[22]提出了一种基于Petri网选择性约简模型的业务流程结构相似性分析方法,以识别不同业务流程之间的相似活动,分析流程的结构相似性;文献[23]综合考虑模型结构与日志行为,提出一种基于模型结构与日志行为的流程相似性计算方法,根据模型中行为在日志中的发生概率为业务流程图加权,并根据加权业务流程图编辑距离的定义计算流程的相似性。文献[24]提出了一种基于累积分布函数的相似性计算方法。该方法通过隐马尔可夫模型(hidden Markov model, HMM)建模和状态序列引用从拓扑结构中挖掘底层拓扑语义,并根据累积分布函数计算两种拓扑结构之间的相似度。文献[25]以现有流程行为相似性度量方法的不足为出发点,提出了一种基于扩展转移关系集(etri集)的业务流程行为相似性综合度量方法。通过构造一个并发可达图获得etri集,定义一种新的流程行为相似性度量方法,进行流程相似性度量。

综上所述,现有的流程模型行为相似性算法大都基于变迁发生序列或存在计算效率较低的问题,因此,本文提出一种基于活动发生关系的流程相似性度量方法,利用活动节点的左右集及发生关系度量流程的相似性。本文的主要贡献主要有以下几点:

1) 基于标签Petri网进行业务流程建模,并提出了活动发生关系的概念和活动节点的左右集概念;

2) 在最佳活动映射的前提下,将活动在左右集上的结构特征与活动间发生关系变化与否相结合,提出了两个不同流程之间的活动相似性,反映了相同活动节点在不同结构上的相似性,并提出流程相似性度的概念来计算整个流程的相似性。

3) 提出了一种基于活动发生关系的流程相似性算法,通过实验对该算法进行了分析验证,给出了一种流程行为相似性度量的新方法。

### 2 相关知识

本文以Petri网建模语言为工具, 采用图形的方式将不同的业务流程展现出来, 对其相似性加以分析和度量.

#### 定义1 标签Petri网<sup>[13]</sup>.

满足下列条件的五元组  $LN = (P, T; F, \Phi, L)$  称为一个标签Petri网:

- 1)  $P$ 为库所的有限集合;
- 2)  $T$ 为变迁的有限集合;
- 3)  $F \subseteq (P \times T) \cup (T \times P)$ 为流关系, 即有向边的集合;
- 4)  $L : T \rightarrow \Phi \cup \{\varepsilon\}$ 是标签映射函数, 其中,  $\Phi$ 为活动标签的集合;

若满足前3个条件, 则称为一个Petri网, 记为  $N = (P, T; F)$ .

$P$ 和 $T$ 统称为网 $LN$ 的节点, 即  $P \cup T \in X$ , 若  $\exists y \in X$ , 使得  $(y, x) \in F$ , 则称 $y$ 为 $x$ 的前集节点, 所有前集节点的集合称为 $x$ 的前集, 即  $x^- = \{y \in X | (y, x) \in F\}$ ; 若  $\exists y \in X$ , 使得  $(x, y) \in F$ , 则称 $y$ 为 $x$ 的后集节点, 所有后集节点的集合称为 $x$ 的后集, 即  $x^+ = \{y \in X | (x, y) \in F\}$ . 与节点相邻的边称为节点的边, 包含输入边和输出边. 图1给出了一个最简单的标签Petri网的示例.

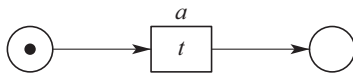


图1 最简单的标签Petri网

Fig. 1 The simplest label Petri net

当选择Petri网作为建模工具的时候, 通常会考虑Petri网的子类— workflow net (WF-net). 由于其只有一个源库所(起始库所)和一个结束库所, 流程开始于源库所(起始库所), 完成于结束库所, 并且所有节点都属于从源库所(起始库所)到结束库所的路径上. 其简洁明了的结构特性便于人们进行业务流程的分析. 此时, 变迁表示实际业务流程中的行为活动, 库所表示相关活动的执行条件或状态.

#### 定义2 workflow网<sup>[7]</sup>.

满足下列条件的标签Petri网称为一个 workflow网 (WF-net):

- 1) 只存在一个输入库所  $i \in P$ , 使得  $i^- = \emptyset$ ;
- 2) 只存在一个输出库所  $o \in P$ , 使得  $o^+ = \emptyset$ ;
- 3)  $LN = (P, T; F, \Phi, L)$ 为强连通的, 即 $N$ 的所有节点  $x \in P \cup T$ , 都属于 $i$ 到 $o$ 的一条有向路径上.

六元组  $WFN = (P, T; F, \Phi, L, M_0)$ 称为一个 workflow网系统, 其中  $M_0$ 为网系统的初始标识.

#### 定义3 迹.

令六元组  $WFN = (P, T; F, \Phi, L, M_0)$ 为一个 workflow网, 变迁的发生序列称为 workflow网的迹, 记作  $\sigma = (t_1, t_2, \dots, t_n)$ , 其中  $t_i \in T (i = 1, 2, \dots, n)$ .

在 workflow网的基础上, 任意2个变迁在整个 workflow网的结构上存在着不同的关系, 为了便于表述, 在进行活动发生关系相似性度量之前, 首先对活动之间的关系进行如下定义(见定义4).

#### 定义4 活动发生关系.

令六元组  $WFN = (P, T; F, \Phi, L, M_0)$ 为一个 workflow网,  $Tr$ 为  $WFN$ 所有可能的执行序列集合, 并且活动  $a, b \in T$ .

1) 强序关系 (strict order): 若存在一条迹  $\sigma = (\dots, t_i, t_j, \dots) \in Tr$ , 其中  $j = i + 1$ , 使得  $t_i = a, t_j = b$ , 则活动 $a$ 和活动 $b$ 称满足强序关系, 记作  $a \rightarrow b$ . 显然, 强序关系不具有对称性和传递性;

2) 并发关系 (concurrency): 至少  $\exists \sigma_1, \sigma_2 \in Tr$ ,  $\sigma_1 = (\dots, t_i, t_j, \dots), \sigma_2 = (\dots, t_j, t_i, \dots)$ , 使得  $t_i = a, t_j = b$ , 则活动 $a$ 和活动 $b$ 称满足并发关系, 记作  $a + b$ . 显然, 并发关系具有对称性, 不具有传递性;

3) 互斥关系 (mutex): 不存在  $\sigma = (\dots, t_i, \dots, t_j, \dots)$ , 使得  $t_i = a, t_j = b$ , 即活动 $a$ 与活动 $b$ 不会出现在同一条迹中, 则活动 $a$ 和活动 $b$ 称满足互斥关系, 记作  $a \times b$ . 显然, 互斥关系具有对称性.

强序关系、并发关系和互斥关系统称为活动的发生关系, 记作  $R$ .

图2对活动的强序关系、并发关系和互斥关系作了一个直观的阐释.

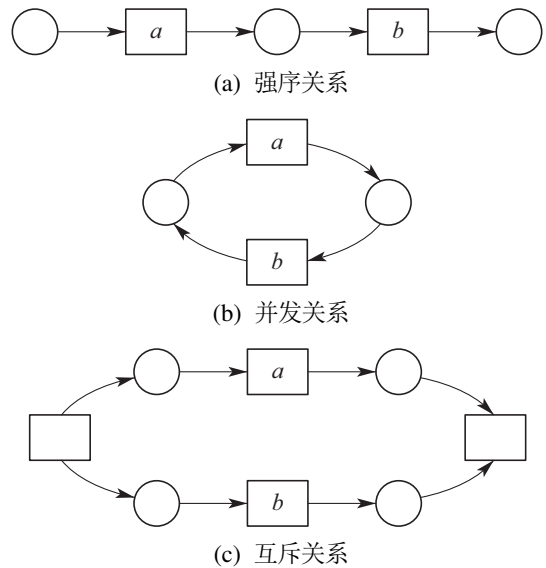


图2 活动的发生关系

Fig. 2 Occurrence relationship between activities

活动发生关系的变化在流程相似性分析中至关重要, 例如: 活动 $a$ 和活动 $b$ 在流程 $A$ 中为强序关系, 在流

程B中为互斥关系,在流程C中为并发关系,假设流程A对应的模型为原模型,那么活动a和活动b的行为关系由强序变到互斥,说明这种行为变化将本来直接跟随的两个活动直接变成了不会同时出现在一条迹中,严重破坏了流程A与流程B的相似性;如果活动a和活动b的发生关系由强序变到并发,由于并发关系保留了活动a和活动b的强序关系,所以此种行为关系变化对流程A与流程C相似性的破坏相对较小.下面将这种破坏性程度定义为违背度,并给出形式化定义:

**定义5** 违背度.

两流程中相同活动对发生关系的偏差程度称为违背度,记作 $V(R, R')$ .令流程A中活动a和活动b的发生关系为 $\rightarrow$ ,流程B中活动a和活动b的发生关系为 $\times$ ,流程C中活动a和活动b的发生关系为 $+$ ,则 $V(\rightarrow, +) < V(\rightarrow, \times)$ .

**定义6** 左集和右集.

令六元组 $WFN = (P, T; F, \Phi, L, M_0)$ 为一个工作流网,  $Tr$ 为 $WFN$ 所有可能的执行序列,并且 $a, b, c \in T$ ,若 $\exists \sigma = (\dots, t_i, t_j, t_k, \dots) \in Tr$ ,其中: $i = j - 1$ ,  $k = j + 1$ ,使得 $t_i = b, t_j = a, t_k = c$ ,则称b为a的左集活动, c为a的右集活动, a的所有左集活动组成的集合称为a的左集,记作 $L_a$ , a的所有右集活动组成的集合称为a的右集,记作 $R_a$ .

本文所讨论的Petri网,均是基于安全的工作流网.所谓安全的,是指流程在运行时,网的任意库所中最多只有一个token.

**3 基于活动发生关系的流程相似性分析方法**

在大多数相似性计算方法中,建立表示两个模型元素之间对应关系的映射函数是首要工作,而本文是在良好变迁映射的前提下,通过计算不同流程中同一活动节点(相同活动节点)在结构上的相似度,进而求得整个流程的相似度.

本文从活动节点的角度考察流程的相似性,认为一个活动节点的左集和右集共同作用,影响流程的相似性,故将活动节点相似性分为两部分:左集相似性、右集相似性,具体定义如下所示.

**定义7** 活动相似性.

令 $WFN_1 = (P_1, T_1; F_1, \Phi_1, L_1, M_0)$ ,  $WFN_2 = (P_2, T_2; F_2, \Phi_2, L_2, M'_0)$ 为两个工作流网,若 $\exists a_k \in T_1 \wedge a_k \in T_2$ ,则

$$S_{WFN_1, WFN_2}(a_k) = \alpha \cdot \frac{|L_{a_k}^{WFN_1} \cap L_{a_k}^{WFN_2}|}{|L_{a_k}^{WFN_1} \cup L_{a_k}^{WFN_2}|} + \beta \cdot \frac{|R_{a_k}^{WFN_1} \cap R_{a_k}^{WFN_2}|}{|R_{a_k}^{WFN_1} \cup R_{a_k}^{WFN_2}|}, \quad (1)$$

$L_{a_k}^{WFN_1}, R_{a_k}^{WFN_1}$ 表示工作流网 $WFN_1$ 中 $a_k$ 的左集和右集,  $L_{a_k}^{WFN_2}, R_{a_k}^{WFN_2}$ 表示工作流网 $WFN_2$ 中 $a_k$ 的左集和

右集,  $\alpha, \beta$ 分别为变迁左集和右集的权重.

特别地,若 $\exists a_l \notin T_1 \cap T_2 \wedge a_l \in T_1 \cup T_2$ ,则

$$S_{WFN_1, WFN_2}(a_k) = 0.$$

式(1)从两方面考察了不同流程中对应活动的相似性: $\frac{|L_{a_k}^{WFN_1} \cap L_{a_k}^{WFN_2}|}{|L_{a_k}^{WFN_1} \cup L_{a_k}^{WFN_2}|}$ 体现了相同元素 $a_k$ 在左集上的结构相似性.  $\frac{|R_{a_k}^{WFN_1} \cap R_{a_k}^{WFN_2}|}{|R_{a_k}^{WFN_1} \cup R_{a_k}^{WFN_2}|}$ 体现了相同元素 $a_k$ 在右集上的结构相似性.

经过反复多次实验计算,本文中取 $\alpha, \beta \in \{0.4, 0.5, 0.6\}$ ,形式化取值法则将在下文第3.1节“基于活动发生关系的流程相似性算法(activity occurrence relation, AOR)”中进行阐述.

在这里 $\alpha, \beta$ 不仅起到权重的作用,同时通过巧妙地选择左右集权重的取值,权衡了不同流程中活动的相似性程度,也体现了违背度的具体含义与文章前半部分的中心思想.同时 $\alpha, \beta$ 的具体取值要根据实际情况进行分析(结合活动发生关系的变化情况(违背度)与单个活动的结构相似性特征).

图3对权重 $\alpha$ 和 $\beta$ 选择(0.4, 0.5, 0.6)何值作出具体解释.

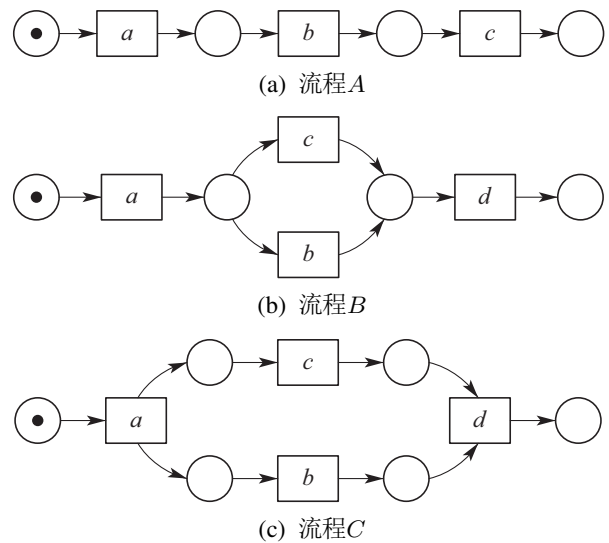


图3 流程A, B, C的Petri网模型

Fig. 3 Petri net models of process A, B and C

为了符合定义6,为图3中的流程分别人工添加一个开始库所 $p_s$ 和变迁 $a_s$ ,一个结束库所 $p_e$ 和变迁 $a_e$ ,4条边 $(p_s, a_s), (a_s, p_0), (p_4, a_e), (a_e, p_e)$ ,如图4所示.

首先,由式(1)计算流程A和流程B中,活动a的相似性: $L_a^A = \{a_s\}, R_a^A = \{b\}, L_a^B = \{a_s\}, R_a^B = \{b, c\}$ ,则 $S_{A,B}(a) = \alpha \cdot \frac{|a_s|}{|a_s|} + \beta \cdot \frac{|b|}{|b, c|} = 1\alpha + \frac{1}{2}\beta$ .

同理,  $S_{A,B}(d) = \alpha \cdot \frac{|c|}{|c, d|} + \beta \cdot \frac{|a_e|}{|a_e|} = \frac{1}{2}\alpha + 1\beta$ ,由于整个流程A, B中a和d的活动发生关系没有变化,

此时取 $\alpha = \beta = 0.5$ , 则 $S_{A,B}(a) = S_{A,B}(d) = 0.75$ ; 而活动 $b$ 的相似性:  $L_b^A = \{a\}, R_b^A = \{c\}, L_b^B = \{a\}, R_b^B = \{d\}$ , 则 $S_{A,B}(b) = \alpha \cdot \frac{|\{a\}|}{|\{a\}|} + \beta \cdot \frac{|\{\emptyset\}|}{|\{c, d\}|} = 1\alpha + 0\beta$ ; 活动 $c$ 的相似性:  $L_c^A = \{b\}, R_c^A = \{d\}, L_c^B = \{a\}, R_c^B = \{d\}$ ,  $S_{A,B}(c) = \alpha \cdot \frac{|\{\emptyset\}|}{|\{a, b\}|} + \beta \cdot \frac{|\{d\}|}{|\{d\}|} = 0\alpha + 1\beta$ .

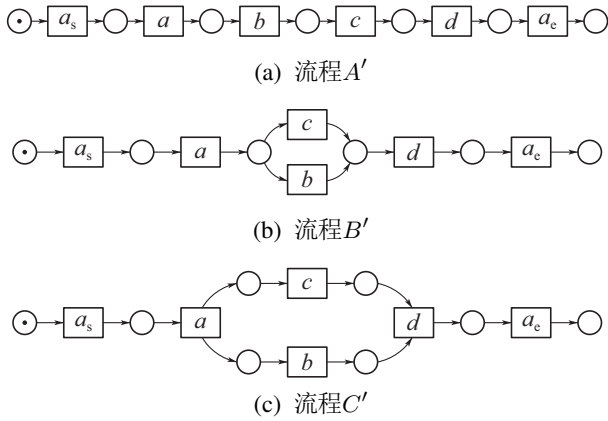


图4 人工改造后的 $A', B', C'$  Petri网模型

Fig. 4 Manual modified Petri net models of process  $A', B'$  and  $C'$

此时考察流程 $A$ 与 $C$ 中活动 $b$ 和 $c$ 的相似性: 同理可得

$$S_{A,C}(b) = \alpha \cdot \frac{|\{a\}|}{|\{a\}|} + \beta \cdot \frac{|\{\emptyset\}|}{|\{c, d\}|} = 1\alpha + 0\beta;$$

$$S_{A,C}(c) = \alpha \cdot \frac{|\{\emptyset\}|}{|\{a, b\}|} + \beta \cdot \frac{|\{d\}|}{|\{d\}|} = 0\alpha + 1\beta.$$

发现有 $S_{A,B}(b) = S_{A,C}(b), S_{A,B}(c) = S_{A,C}(c)$ , 这是因为活动 $b$ 在流程 $A$ 与 $B$ 中左右集, 与在流程 $A$ 与 $C$ 中的左右集相等, 活动 $c$ 同理, 同时也说明了引入 $\alpha, \beta$ 的必要性. 这时, 因为 $b, c$ 在流程 $A$ 中的发生关系为 $b \rightarrow c$ , 在流程 $B$ 中的发生关系为 $b \times c$ , 在流程 $C$ 的发生关系为 $b + c$ , 又由于 $V(\rightarrow, +) < V(\rightarrow, \times)$ , 即流程 $A$ 与 $B$ 中 $b, c$ 分别的相似度应该小于流程 $A$ 与 $C$ 中 $b, c$ 分别的相似度, 所以计算流程 $A$ 与 $B$ 中 $b, c$ 的相似度时, 又由于 $b, c$ 的左右结构相似性特点( $1 > 0$ 和 $0 < 1$ ), 分别取 $\alpha = 0.4, \beta = 0.6, \alpha = 0.6, \beta = 0.4$ ; 计算 $A$ 与 $C$ 中 $b, c$ 的相似度时, 分别取 $\alpha = 0.5, \beta = 0.5$ , 则

$$S_{A,B}(b) = 1\alpha + 0\beta = 0.4,$$

$$S_{A,B}(c) = 0\alpha + 1\beta = 0.4;$$

$$S_{A,C}(b) = 1\alpha + 0\beta = 0.5,$$

$$S_{A,B}(c) = 0\alpha + 1\beta = 0.5.$$

综上所述, 当活动发生关系不变或由强序变为并发时, 取 $\alpha = \beta = 0.5$ ; 发生关系由强序变为互斥时, 根据活动左右结构相似性的大小进行取值 ( $\alpha, \beta \in \{0.4, 0.6\}$ ).

需要说明的是, 上述讨论的是两个强序关系的活动变为互斥或并发的情况, 对于由弱序变为互斥或并发时, 此方法同样适用.

**定义8** 流程相似性.

令 $WFN_1 = (P_1, T_1; F_1, \Phi_1, L_1, M_0)$ 为流程1的工作流网,  $WFN_2 = (P_2, T_2; F_2, \Phi_2, L_2, M'_0)$ 为流程2的工作流网, 则流程1与流程2的相似性为

$$S_{(1,2)} = \frac{\sum_{t_i \in T_1 \cap T_2} S_{1,2}(t_i)}{\max\{|T_1| - 2, |T_2| - 2\}}, \quad (2)$$

其中分母减去2是为了消除人工初始、结束变迁的影响.

根据式(2), 图3中流程 $A$ 与 $B$ 的相似性为

$$(S_{A,B}(a) + S_{A,B}(b) + S_{A,B}(c) + S_{A,B}(d)) / (\max\{6 - 2, 6 - 2\}) = (0.75 + 0.4 + 0.4 + 0.75) / 4 = 0.48;$$

流程 $A$ 与 $C$ 的相似性为

$$(S_{A,C}(a) + S_{A,C}(b) + S_{A,C}(c) + S_{A,C}(d)) / (\max\{6 - 2, 6 - 2\}) = (0.75 + 0.5 + 0.5 + 0.75) / 4 = 0.63.$$

**3.1 基于活动发生关系的流程相似性算法(AOR)**

**算法1** 基于活动发生关系的流程相似性算法.

输入: 流程模型 $M_1$ 和 $M_2$ ;

输出:  $M_1$ 和 $M_2$ 的相似性度.

步骤如下:

**Step 1** 遍历两个流程的所有活动 $t_i \in T_{M_1} \cup T_{M_2}, 1 \leq i \leq \max\{|M_1|, |M_2|\}$ , 对 $\forall t_m \in T_{M_1} \cap T_{M_2}$ , 根据定义6, 输出 $t_m$ 在流程 $M_1$ 和 $M_2$ 中的左右集 $L_{t_m}^{M_1}$ 和 $L_{t_m}^{M_2}, R_{t_m}^{M_1}$ 和 $R_{t_m}^{M_2}$ ;

**Step 2** 根据定义4中的活动发生关系, 对 $\forall t_m \in T_{M_1} \cap T_{M_2}$ , 若 $\exists t_n \in T_{M_1} \cup T_{M_2}$ , 使得 $t_m \rightarrow t_n \in R_{M_1}, t_m \times t_n \in R_{M_2}$ , 并且

$$\left( \frac{|L_{t_m}^{M_1} \cap L_{t_m}^{M_2}|}{|L_{t_m}^{M_1} \cup L_{t_m}^{M_2}|} < \frac{|R_{t_m}^{M_1} \cap R_{t_m}^{M_2}|}{|R_{t_m}^{M_1} \cup R_{t_m}^{M_2}|} \right. \\ \left. \left( \frac{|L_{t_m}^{M_1} \cap L_{t_m}^{M_2}|}{|L_{t_m}^{M_1} \cup L_{t_m}^{M_2}|} > \frac{|R_{t_m}^{M_1} \cap R_{t_m}^{M_2}|}{|R_{t_m}^{M_1} \cup R_{t_m}^{M_2}|} \right) \right),$$

则 $\alpha = 0.6, \beta = 0.4$  ( $\alpha = 0.4, \beta = 0.6$ ).  $S_{M_1, M_2}(t_n)$ 中 $\alpha$ 和 $\beta$ 的取值同理;

**Step 3** 对 $\forall t_m \in T_{M_1} \cap T_{M_2}$ , 若 $\exists t_n \in T_{M_1} \cup T_{M_2}$ , 使得 $t_m \rightarrow t_n \in R_{M_1}, t_m + t_n \in R_{M_2}$ , 并且

$$\left( \frac{|L_{t_m}^{M_1} \cap L_{t_m}^{M_2}|}{|L_{t_m}^{M_1} \cup L_{t_m}^{M_2}|} < \frac{|R_{t_m}^{M_1} \cap R_{t_m}^{M_2}|}{|R_{t_m}^{M_1} \cup R_{t_m}^{M_2}|} \right. \\ \left. \left( \frac{|L_{t_m}^{M_1} \cap L_{t_m}^{M_2}|}{|L_{t_m}^{M_1} \cup L_{t_m}^{M_2}|} > \frac{|R_{t_m}^{M_1} \cap R_{t_m}^{M_2}|}{|R_{t_m}^{M_1} \cup R_{t_m}^{M_2}|} \right) \right),$$

则 $\alpha = \beta = 0.5$ .  $S_{M_1, M_2}(t_n)$ 中 $\alpha$ 和 $\beta$ 的取值同理;

**Step 4** 根据定义7, 计算活动 $\forall t_i \in T_{M_1} \cup T_{M_2}$ 的相似性:

$$S_{M_1, M_2}(t_i) = \alpha \cdot \frac{|L_{t_i}^{M_1} \cap L_{t_i}^{M_2}|}{|L_{t_i}^{M_1} \cup L_{t_i}^{M_2}|} + \beta \cdot \frac{|R_{t_i}^{M_1} \cap R_{t_i}^{M_2}|}{|R_{t_i}^{M_1} \cup R_{t_i}^{M_2}|};$$

**Step 5** 若 $\exists t_p \in T_{M_1} \cup T_{M_2} \wedge t_p \notin T_{M_1} \cap T_{M_2}$ , 则 $S_{M_1, M_2}(t_p) = 0$ ;

**Step 6** 对所有活动的相似度求和:  $S_{M_1, M_2}(t_j)$ ,  $t_j \in T_{M_1} \wedge t_j \in T_{M_2}$ ;

**Step 7** 根据定义8, 计算流程模型 $M_1$ 和 $M_2$ 的相似性度:

$$S_{(M_1, M_2)} = \frac{\sum_{t_i \in T_{M_1} \cap T_{M_2}} S_{M_1, M_2}(t_j)}{\max\{|T_{M_1}| - 2, |T_{M_2}| - 2\}};$$

**Step 8** END (算法结束).

### 3.2 算法时间复杂度分析

首先分析算法的时间复杂度. 给定一个工作流网 $W$ , 变迁集为 $T$ , 库所集为 $P$ , 流关系集为 $F$ , AOR算法遍历 $W$ 中除输出库所外的所有节点及流关系, 操作所需时间为 $O(|T| + |P| + |F|)$ . 条件判断、计算权重、并入操作、赋予颜色、出队、入队等操作所需时间均为 $O(1)$ , 总时间为 $O(1)$ . 因此AOR算法的时间复杂度为 $O(|T| + |P| + |F|)$ , 而目前主流算法中, 计算效率较高的基于变迁紧邻关系重要性的流程相似性(TAR++)算法的最坏时间复杂度为 $O(V + E + N!)$ , 是一个阶乘级的复杂度, 即使运用所给两种加速方式, 其时间复杂度也不可能低于 $O(|T| + |P| + |F|)$ .

### 4 案例分析

本节通过两个具体的业务流程来验证所提出相似性算法的可行性, 即利用基于活动发生关系的流程相似性算法(AOR), 度量两个业务流程的相似程度(省略人工改造后的Petri网模型), 流程模型如图5所示.

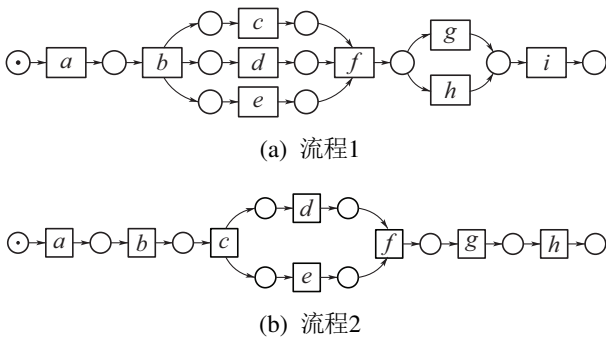


图5 流程1和2的Petri网模型

Fig. 5 Petri net models of process 1 and 2

下面以流程1和流程2的Petri网模型作为输入, 执行基于活动发生关系的流程相似性算法(AOR).

执行算法Step1, 输出相同活动的左右集:

$$\begin{aligned} L_a^1 &= L_a^2 = \{a_s\}, R_a^1 = R_a^2 = \{b\}; L_b^1 = L_b^2 = \{a\}, \\ R_b^1 &= R_b^2 = \{c\}; L_c^1 = \{b\}, R_c^1 = \{f\}; L_c^2 = \{b\}, \\ R_c^2 &= \{d\}; L_d^1 = \{b\}, R_d^1 = \{f\}; L_d^2 = \{c\}, \\ R_d^2 &= \{f\}; L_e^1 = \{b\}, R_e^1 = \{f\}; L_e^2 = \{c\}, \\ R_e^2 &= \{f\}; L_f^1 = \{c, d, e\}, R_f^1 = \{g, h\}; L_f^2 = \{d, e\}, \\ R_f^2 &= \{g\}; L_g^1 = \{f\}, R_g^1 = \{i\}; L_g^2 = \{f\}, \\ R_g^2 &= \{h\}; L_h^1 = \{f\}, R_h^1 = \{i\}; L_h^2 = \{g\}, R_h^2 = \{a_e\}; \end{aligned}$$

执行算法Step2-3, 计算每个相同活动的相似性:

$$\begin{aligned} S_{1,2}(a) &= \alpha \cdot \frac{|\{a_s\}|}{|\{a_s\}|} + \beta \cdot \frac{|\{b\}|}{|\{b\}|} = \\ &= \frac{1}{2} \times 1 + \frac{1}{2} \times 1 = 1; \\ S_{1,2}(b) &= \alpha \cdot \frac{|\{a\}|}{|\{a\}|} + \beta \cdot \frac{|\{c\}|}{|\{c, d, e\}|} = \\ &= \frac{1}{2} \times 1 + \frac{1}{2} \times \frac{1}{3} = \frac{2}{3}; \\ S_{1,2}(c) &= \alpha \cdot \frac{|\{b\}|}{|\{b\}|} + \beta \cdot \frac{|\{\emptyset\}|}{|\{d, f\}|} = \\ &= \frac{1}{2} \times 1 + \frac{1}{2} \times 0 = \frac{1}{2}; \\ S_{1,2}(d) &= \alpha \cdot \frac{|\{\emptyset\}|}{|\{b, c\}|} + \beta \cdot \frac{|\{f\}|}{|\{f\}|} = \\ &= \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2}; \\ S_{1,2}(e) &= \alpha \cdot \frac{|\{\emptyset\}|}{|\{b, c\}|} + \beta \cdot \frac{|\{f\}|}{|\{f\}|} = \\ &= \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2}; \\ S_{1,2}(f) &= \alpha \cdot \frac{|\{d, e\}|}{|\{c, d, e\}|} + \beta \cdot \frac{|\{g\}|}{|\{g, h\}|} = \\ &= \frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{1}{2} = \frac{7}{12}; \\ S_{1,2}(g) &= \alpha \cdot \frac{|\{f\}|}{|\{f\}|} + \beta \cdot \frac{|\{\emptyset\}|}{|\{i, h\}|} = \\ &= \frac{2}{5} \times 1 + \frac{3}{5} \times 0 = \frac{2}{5}; \\ S_{1,2}(h) &= \alpha \cdot \frac{|\{\emptyset\}|}{|\{f, g\}|} + \beta \cdot \frac{|\{\emptyset\}|}{|\{i, a_e\}|} = \\ &= \frac{2}{5} \times 0 + \frac{3}{5} \times 0 = 0. \end{aligned}$$

执行算法Step4, 输出无匹配活动  $t_p (t_p \in T_{M_1} \cup T_{M_2} \wedge t_p \notin T_{M_1} \cap T_{M_2})$  的相似性,  $S_{(1,2)}(t_p) = 0$ ;

执行算法Step5-7, 输出流程1与流程2的相似性:

$$\begin{aligned} S_{(1,2)} &= \frac{\sum S_{(1,2)} a_i}{\max\{|T_1| - 2, |T_2| - 2\}} = \\ &= (1 + \frac{2}{3} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{7}{12} + \frac{2}{5} + 0 + 0) / 9 = \end{aligned}$$

$$\frac{249}{540} \approx 0.46.$$

### 5 实验设计与仿真分析

实验机器环境为: Intel Core I5-7200U CPU @ 可加速至3.1 Hz, 内存为8 GB, Window10 64位操作系统.

#### 5.1 实验设计

本节实验主要分两阶段进行, 第1阶段为算法的可行性分析, 第2阶段为算法的性能分析. 实验涉及的数据集由两部分组成: 450个典型的业务流程模型来自SAP模型库, 50个人工流程模型, 为满足实验需要由人工编撰. 首先将500个业务流程模型运用本文提出的相似性算法(AOR), 计算出两两模型之间的相似性度, 利用实验结果分析以及相似性结果均属于0~1, 初步验证算法的可行性, 并对同样的模型运用主流的相似性算法, 与AOR算法结果作对比, 进一步验证算法结果的正确性. 然后, 对不同规模的数据集应用本文算法和主流的相似性算法, 通过算法运行时间对比, 验证本算法良好的性能体现.

#### 5.2 效果评估

对500个业务流程模型应用本算法, 限于篇幅, 本节仅给出10个流程模型(其中6个来自SAP模型库, 4个来自人工模型, 具体请参见链接[https://pan.baidu.com/s/1PmQDvFrEXg\\_JTzT2UoD\\_Gw](https://pan.baidu.com/s/1PmQDvFrEXg_JTzT2UoD_Gw)) 的实验结果, 如表1所示. 可以看出, 本算法得出的实验结果均在0~1范围内, 当流程完全相同时, 实验结果为1, 当流程完全不同时, 实验结果为0. 由此, 初步验证了本算法的可行性.

表1 AOR算法得到的10个流程模型的相似性度

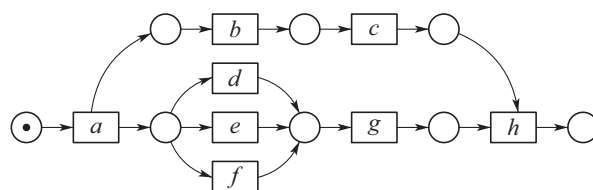
Table 1 Similarity of 10 process models derived from AOR algorithm

	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$	$M_9$	$M_{10}$
$M_1$	1	0.71	0.82	0.68	0.74	0.57	0.61	0.32	0.47	0.89
$M_2$	0.71	1	0.37	0.71	0.23	0.26	0.35	0.17	0.92	0.46
$M_3$	0.82	0.37	1	0.43	0.84	0.54	0.00	0.27	0.58	0.83
$M_4$	0.68	0.71	0.43	1	0.13	0.24	0.79	0.38	0.65	0.69
$M_5$	0.74	0.23	0.84	0.13	1	0.81	0.77	0.85	0.28	0.19
$M_6$	0.57	0.26	0.54	0.24	0.81	1	0.35	0.63	0.59	0.85
$M_7$	0.61	0.35	0.00	0.79	0.77	0.35	1	0.78	0.96	0.46
$M_8$	0.32	0.17	0.27	0.38	0.85	0.63	0.78	1	0.00	0.89
$M_9$	0.47	0.92	0.58	0.65	0.28	0.59	0.96	0.00	1	0.14
$M_{10}$	0.89	0.46	0.83	0.69	0.19	0.85	0.46	0.89	0.14	1

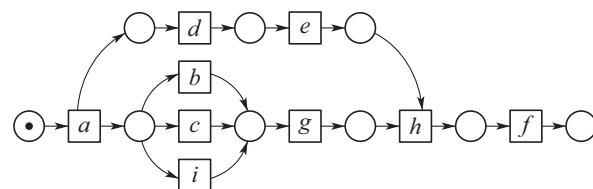
为进一步验证算法的正确性, 从500个流程模型中任意抽出300个模型运用主流的相似性算法与AOR相似性算法, 将得到的实验结果与本算法做对比. 对比算法包括基于行为的度量方法: 1) 行为轮廓相似性算

法(behavioral profile, BP)<sup>[6,8]</sup>, 细化了活动发生关系, 给出行为轮廓概念, 进而度量模型相似性, 但该算法对并发结构不敏感; 2) 任务最短跟随距离矩阵的相似性算法(shortest succession distances between tasks, SS-DT)<sup>[9]</sup>, 用两两活动间的最短跟随距离表征模型行为, 问题在于该算法需要计算两个流程模型的同维化矩阵, 降低了算法的运算效率. 基于变迁紧邻关系重要性的流程相似性算法(TAR++)<sup>[7]</sup>, 在考虑模型行为的基础上对边加权, 有效的改进了现有的算法. 基于模型内部结构的流程相似性度量方法 (total-mapping-model, TMM)<sup>[11]</sup>, 该方法采取迭代映射来映射库所和变迁, 进而识别两个模型的对应关系. 但其忽略活动文本标签的语义, 使得到的模型不够准确.

本节仅给出人工模型 $M_1$ 和 $M_2$ (图6)及其实验对比结果(表2), 根据本文提出的AOR相似性算法, 模型 $M_1$ 和 $M_2$ 的行为相似性度为0.23, 由表2可以看出, AOR算法的运算结果与主流算法比较接近, 进一步验证了算法的正确性.



(a) 流程3的Petri网模型 $M_1$



(b) 流程4的Petri网模型 $M_2$

图6 Petri网模型 $M_1$ 和 $M_2$

Fig. 6 Petri net models  $M_1$  and  $M_2$

表2 与主流算法的实验结果比较

Table 2 Comparisons with experimental results of mainstream algorithms

算法名称	流程 $M_1, M_2$ 的相似性度
SSDT	0.24
TAR++	0.18
BP	0.19
TMM	0.21
AOR	0.23

#### 5.3 性能评估

为评估AOR算法在实际应用中的性能体现, 将抽取的300个实验模型分成不同的数据集(数据集1、数

数据集2、数据集3), 分别包含70, 90和140个流程模型, 表3列出了各数据集的基本特征: 所含平均变迁数、最大变迁数、平均库所数、最大库所数、平均边数、最大边数. 目的是利用算法在不同复杂程度模型上的花费, 验证算法的性能表现.

为更加直观地比较AOR算法与其他主流算法, 根

据表3信息, 将算法在不同复杂程度数据集上的运行时间制成柱形图, 横坐标为不同的数据集, 纵坐标为运行时间, 如图7所示. 由图7可以看出, 对于任意一种复杂程度的数据集, AOR算法所花费的时间都稍低与其他算法, 运算效率最高, 而行为轮廓相似性算法(BP)的运行时间明显高于其他算法, 运算效率最低.

表3 不同规模下的算法性能体现

Table 3 Performance of algorithms at different scales

数据规模	平均变迁数/库所数/边数	最大变迁数/库所数/边数	算法运行时间/ms				
			SSDT	TMM	TAR++	BP	AOR
70	12.89/11.53/25.42	38/33/82	1.54	1.42	1.24	2.17	1.21
90	7.64/9.03/18.01	40/31/109	1.89	1.79	1.63	2.46	1.54
140	10.54/10.86/20.77	51/48/137	2.21	2.13	1.98	2.52	1.93

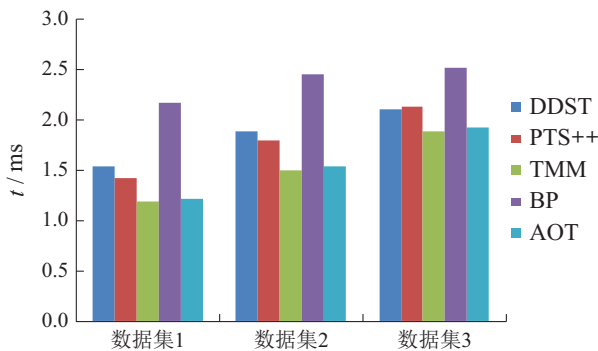


图7 不同算法的运行时间对比

Fig. 7 Comparisons of runtime for different algorithms

另外, 本文方法(AOR)与文献[11] (TMM)不同之处在于不需要定义不同流程中的活动节点映射, 本文是在最佳的变迁节点映射条件下, 进行流程的相似性度量. 并且文献[13]是从左右集的交叉元素与节点左右分支的结构考虑节点的相似性, 进而度量整个流程的相似性. 而本文是结合活动节点左右集的重复元素与活动发生关系在整个流程中是否发生变化, 来度量流程的相似性.

## 6 结语与展望

为了一定程度上解决大量流程模型的管理问题, 提高模型检索、重用以及流程合并的效率, 本文提出了一种基于模型时序关系(执行顺序)的流程行为相似性算法, 该算法在标签Petri网和良好的变迁映射基础上, 模型结构与行为相结合, 将活动在左右集上的结构相似性与活动间发生关系在不同流程的变化情况综合考虑, 进一步识别活动的相似性, 然后, 通过将活动的相似性进行归一化处理, 给出流程相似性定义, 最后, 提出基于活动发生关系的流程相似性算法(AOR)度量不同流程的相似性. 实验结果表明, 该算法能够正确的计算不同流程的相似性, 与其他相似算法相比, 具有更高的时空复杂度.

该相似性算法尚存在一些不足, 算法适用于当相同活动节点的左右集与“原模型”左右集不同时, 未考虑左右集相等的情况. 另外, 如何将AOR方法应用于工业场景, 未来将对这些工作做进一步研究, 以使算法具有更强的适应能力.

## 参考文献:

- [1] AALST V D W M P. *Process mining: Discovery, conformance and enhancement of business processes*. Germany, Heidelberg: Springer Publishing Company, Incorporated, 2011.
- [2] EHRIG M, KOSCHMIDER A, OBERWEIS A. Measuring similarity between semantic business process models. *The Conference on Conceptual Modelling*. Ballarat, Victoria, Australia: DBLP, 2007: 71 – 80.
- [3] DIJKMAN R, DUMAS M, GARCIA-BANUELOS L. Graph matching algorithms for business process model similarity search. *International Conference on Business Process Management*. Ulm, Germany: Springer-Verlag, 2009: 48 – 63.
- [4] ROSA M L, DUMAS M, UBA R, et al. *Merging business process models on the Move to Meaningful Internet Systems: OTM 2010*. Berlin: Springer Heidelberg, 2010: 96 – 113.
- [5] ZHA H P, WANG J M, WEN L J, et al. A workflow net similarity measure based on transition adjacency relations. *Computers in Industry*, 2010, 61(5): 463 – 471.
- [6] WEIDLICH M, EMENDLING J, WESKE M. Efficient consistency measurement based on behavioral profiles of process models. *IEEE Transactions on Software Engineering*, 2011, 37(3): 410 – 429.
- [7] YIN Ming, WEN Lijie, WANG Jianmin, et al. Process similarity algorithm based on importance of transition adjacent relations. *Computer Integrated Manufacturing Systems*, 2015, 21(2): 344 – 358. (殷明, 闻立杰, 王建民, 等. 基于变迁紧邻关系重要性的流程相似性算法. 计算机集成制造系统, 2015, 21(2): 344 – 358.)
- [8] SONG L, WANG J M, WEN L J, et al. Querying process models based on the temporal relations between tasks. *IEEE International Enterprise Distributed Object Computing Conference Workshops*. New York: IEEE, 2011: 213 – 222.
- [9] WANG Shuhao, WEN Lijie, WEI Daisen, et al. SSDT matrix-based behavioral similarity algorithm for process models. *Computer Integrated Manufacturing Systems*, 2013, 19(8): 1822 – 1831. (汪抒浩, 闻立杰, 魏代森, 等. 基于任务最短跟随距离矩阵的流程模



- 型行为相似性算法. 计算机集成制造系统, 2013, 19(8): 1822 – 1831.)
- [10] KUNZE M, WEIDLICH M, WESKE M. Querying process models by behavior inclusion. *Software and System Modeling*, 2015, 14(3): 1105 – 1125.
- [11] LU Y H, YU H F, MING Z, et al. A similarity measurement based on structure of Business Process. *International Conference on Computer Supported Cooperative Work in Design*. New York: IEEE, 2016: 498 – 503.
- [12] ZHANG Rongbin, SHI Jun, ZHANG Pengfei, et al. Similarity measurement of transition mapping relation using Petri net. *Journal of Zhejiang University (Engineering Science)*, 2017, 51(6): 1205 – 1213.  
(张荣斌, 石军, 张鹏飞, 等. Petri网的映射变迁关系相似性度量. 浙江大学学报(工学版), 2017, 51(6): 1205 – 1213.)
- [13] WANG Zixuan, WEN Lijie, WANG Shuhao, et al. Similarity measurement for process models based on transition-labeled graph edit distance. *Computer Integrated Manufacturing Systems*, 2016, 22(2): 343 – 352.  
(王子璇, 闻立杰, 汪抒浩, 等. 基于变迁标签图编辑距离的过程模型相似性度量. 计算机集成制造系统, 2016, 22(2): 343 – 352.)
- [14] SUN J Y, GU T L, QIAN J Y. A behavioral similarity metric for sSemantic workflows based on semantic task adjacency relations with importance. *IEEE Access*, 2017, 5: 15609 – 15618.
- [15] SONG Jinfeng, WEN Lijie, WANG Jianming. A similarity measure for process models based on task occurrence relations. *Journal of Computer Research and Development*, 2017, 54(4): 832 – 843.  
(宋金凤, 闻立杰, 王建民. 基于任务发生关系的流程模型相似性度量. 计算机研究与发展, 2017, 54(4): 832 – 843.)
- [16] WANG M M, DING Z J, LIU G J, et al. Measurement and computation of profile similarity of workflow nets based on behavioral relation matrix. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018: 1 – 18.
- [17] WANG J M, HE T F, WEN L J, et al. *A behavioral similarity measure between labeled Petri nets based on principal transition sequences*. Move to Meaningful Internet Systems: OTM 2010. Berlin: Springer Heidelberg, 2010: 394 – 401.
- [18] DONG Zihe, WEN Lijie, HUANG Haowei, et al. Behavioral similarity algorithm for process models based on firing sequence collection. *Journal of Software*, 2015, 26(3): 449 – 459.  
(董子禾, 闻立杰, 黄浩未, 等. 基于触发序列集合的过程模型行为相似性算法. 软件学报, 2015, 26(3): 449 – 459.)
- [19] TAN W A, WANG P. A method towards process similarity measurement according to constrained traces. *IEEE International Conference on Computer Supported Cooperative Work in Design*. New York: IEEE, 2017: 79 – 84.
- [20] DUMAS M, GARC A-BAN UELOS L, DIJKMAN R. Similarity search of business process models. *IEEE Data Engineering Bulletin*, 2009, 32(3): 23 – 28.
- [21] DIJKMAN R, DUMAS M, DONGEN B V, et al. Similarity of business process models: Metrics and evaluation. *Information Systems*, 2011, 36(2): 498 – 516.
- [22] SYUKRIILAH N, KUSUMO D S, WIDOWATI S. Structural similarity analysis of business process model using selective reduce based on Petri Net. *International Conference on Information & Communication Technology*. New York: IEEE, 2015: 1 – 5.
- [23] ZHOU Changhong, ZENG Qingtian, LIU Cong, et al. Business process similarity computing method based on process model structure and log behavior. *Computer Integrated Manufacturing Systems*, 2018, 243(7): 209 – 221.  
(周长红, 曾庆田, 刘聪, 等. 基于模型结构与日志行为的流程相似度计算. 计算机集成制造系统, 2018, 243(7): 209 – 221.)
- [24] LI Q, TANG W Q, LI S C. The topology similarity calculation of process plant models based on probability distribution. *Journal of Computer-Aided Design and Computer Graphics*, 2019, 31(1): 165.
- [25] LIU C, ZENG Q T, DUAN H, et al. Towards comprehensive support for business process behavior similarity measure. *IEICE Transactions on Information and Systems*, 2019, 102(3): 588 – 597.

#### 作者简介:

李东月 硕士研究生, 目前研究方向为行为分析与挖掘研究,  
E-mail: 1244321663@qq.com;

方欢 博士, 教授, 目前主要研究方向为Petri网理论与应用、智能控制, E-mail: fanghuan0307@163.com.