

# 基于 Petri 网的物流派送模型修复分析

方贤文 杨慧慧 邵叱风

(安徽理工大学数学与大数据学院 安徽 淮南 232001)

**摘要:** 目前,模型修复技术应用领域非常广泛,模型修复不仅可以达到完全重放日志的目的,还可以最大程度地与原始模型保持相似,能够保留原始模型的价值。因此,越来越多的学者专注于研究模型修复。基于 Petri 网构建了一个物流派送流程模型,然后将事件日志与流程模型进行一致性检查,主要根据成本对齐测量其偏差,并针对检查到的偏差部分,根据提出的用于辅助模型修复的算法对原始模型进行修复,使得修复后的模型能够完全重放事件日志。最后通过实例分析验证了修复模型的合理性。

**关键词:** Petri 网; 一致性检查; 成本对齐; 行为轮廓; 模型修复

**中图分类号:** TP391.9 **文献标志码:** A **文章编号:** 1672-1098(2021)06-0007-06

## Repair Analysis of Logistics Delivery Model Based on Petri Net

FANG Xianwen, YANG Huihui, SHAO Chifeng

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

**Abstract:** At present, as the application field of the model repair technology is very wide, The model repair can not only achieve the purpose of completely reproducing the log, but also keep the maximum similarity with the original model and retain the value of the original model. Therefore, more and more scholars focus on the study of model repair. Based on Petri net, a logistics delivery process model was built and then checked consistently with the event log mainly according to the cost of measuring the deviation. The original model was repaired according to the proposed algorithm applied in repairing auxiliary model to be able to replay the event log. Finally, an example was given to verify the rationality of the repair model.

**Key words:** Petri net; consistency check; cost alignment; behavioral profile; model repair

流程挖掘是一门介于计算智能和数据挖掘以及过程建模和分析之间的新兴研究学科,流程挖掘的起点是事件日志,所有流程挖掘技术都假设有可能顺序地记录事件,使每个事件指向一个活动,并与特定的案例(即流程实例)相关<sup>[1]182</sup>。流程挖掘主要包括三个领域:流程发现<sup>[2-4]</sup>、一致性检查和流程增强。最近新兴的模型修复领域,也算是流程增强的一种。本文的研究侧重于一致性检查与模型修复。

一致性检查是将现有流程模型与同一流程的事件日志进行比较。一致性检查可以用来检查记录在日志中的事实是否符合模型。文献[5]中针对声明性模型提出了一种新的一致性检查方法,文中采用了基于对齐的方法,以便能够处理声明式模型固有的灵活性带来的巨大搜索空间。文献[6]中提出了一种检验序列图与状态图一致性的方法。在比较过

收稿日期:2020-10-16

基金项目:国家自然科学基金资助项目(61402011 61572035)

作者简介:方贤文(1975-)男,河南信阳人,教授,博士,研究方向:Petri网与可信计算。

程模型和事件日志有不同的质量维度,例如适应度、简单性、精度和泛化<sup>[7-8]</sup>。而在本文中,我们关注于适应度:一个具有良好适应性的模型允许在事件日志中看到的大多数行为。如果日志中的所有轨迹都能被模型从头到尾重复播放,那么模型就具有完美的适用性。

模型修复主要针对事件日志与原始模型不符合的部分进行修复,最大程度地与原始模型保持相似。文献[9]将一致性检查的信息用于模型修复,识别流程模型中已删除活动的结构、附加活动和邻接活动之间的关系以及流程模型中的不一致子流程,基于文中所提出的技术实现模型修复。文献[10]使用一个现有的一致性检查器,将日志分解为不匹配子迹的几个子日志,并对每个子日志派生出一个子流程,最后将其添加到原始模型的适当位置,从而达到修复一个流程模型的目的。文献[11]主要致力于一种模型分解修复工具,分解后的模型修复是模型增强的一种方法。模型修复可以使得事件日志与流程模型有较高的适合度甚至完美适合度,因此,越来越多的学者致力于研究模型修复。

在已有的方法中最为经典的修复方法之一是插入-跳过活动修复法,此方法对于单个日志移动来说比较容易理解和使用。但是,在大多数情况下会见到连续的日志移动,那么在同一个位置单个的插入多个活动事件,会构成模型的繁琐。因此,在已有的子流程修复算法的基础上,本文针对含有连续日志移动的流程模型修复问题,提供了一种用于辅助模型修复的有效寻找日志移动可修复的库所集的算法。从成本对齐的角度出发,根据检测到的日志移动部分借助于该算法可有效实现模型修复。

## 1 基本概念

定义1<sup>[12]</sup> (变迁发生规则) 一个网系统是一个标识网  $\Sigma = (P, T; F, M)$  并具有下面的变迁规则

(1) 对于变迁  $t \in T$ , 如果

$$\forall p \in P: p \in {}^{\cdot}t \rightarrow M(p) \geq 1$$

则说变迁  $t$  在标识  $M$  有发生权, 记为  $M[t >]$ 。

(2) 若  $M[t >]$ , 则在标识  $M$  下, 变迁  $t$  可以发生, 从标识  $M$  发生变迁  $t$  得到一个新的标识  $M'$  (记为  $M[t > M']$ ) 对  $\forall p \in P$ ,

$$M'(p) = \begin{cases} M(p) - 1 & \text{若 } p \in {}^{\cdot}t - t^{\cdot} \\ M(p) + 1 & \text{若 } p \in t^{\cdot} - {}^{\cdot}t \\ M(p) & \text{其他} \end{cases}$$

定义2<sup>[13]223</sup> (迹, 事件日志) 设  $\Phi$  是一组活动名称集合, 若  $\alpha \in \Phi^*$  是一个活动序列, 则称  $\alpha$  是一条迹。若  $L \in B(\Phi^*)$  是迹的有限非空的多重集, 则称  $L$  是一个事件日志。

定义3<sup>[13]224</sup> (对齐) 设  $N = (P, T, F, \alpha, m_0, m_f)$  是一个 Petri 网, 设  $\alpha = a_1 a_2 \cdots a_m$  是一条迹,  $\alpha$  到流程模型  $N$  之间的对齐  $\gamma = (a_1 t_1) (a_2 t_2) \cdots (a_k t_k)$  是满足下列条件的移动序列:

(1) 对齐  $\gamma$  的第一行  $(a_1 a_2 \cdots a_k)$  是事件日志中的一条迹;

(2) 对齐  $\gamma$  的第二行  $(t_1 t_2 \cdots t_k)$  是 Petri 网流程模型  $N$  中的一组变迁发生序列;

(3) 对于所有的  $i = 1, 2, \cdots, k$ , 如果  $t_i \neq \gg$ ,  $\alpha(t_i) \neq \tau$  且  $a_i \neq \gg$ , 那么  $\alpha(t_i) = a_i$ 。

这里的移动对满足以下条件:

(1) 如果  $a_i = \gg \wedge t_i \neq \gg$ , 则称为模型移动;

(2) 如果  $a_i \neq \gg \wedge t_i = \gg$ , 则称为日志移动;

(3) 如果  $a_i \neq \gg \wedge t_i \neq \gg$ , 则称为同步移动;

定义4<sup>[14]10</sup> (最优对齐) 给定一条迹  $\alpha$  和 Petri 网流程模型  $N$ , 对于  $\alpha$  和  $N$  之间所有的对齐  $\chi$ ,  $\exists \gamma \in \chi$  对  $\forall \gamma' \in \chi$  都有  $c(\gamma) \leq c(\gamma')$ 。则称  $\gamma$  是  $\alpha$  和  $N$  之间的最优对齐。

定义5<sup>[14]10</sup> (成本对齐) 给定一条迹  $\alpha$  和 Petri 网流程模型的对齐  $\gamma$ , 别赋予  $\gamma$  中同步移动, 日志移动和模型移动的成本函数值为 0、1、1。则  $\gamma$  的成本  $c(\gamma)$  为  $\gamma$  中所有移动的成本函数值的总和, 即  $c(\gamma) := \sum_{i=1}^{|\gamma|} c(\gamma_i)$ 。

定义6<sup>[15]</sup> (行为轮廓) 设  $(N, M_0)$  是一个网系统, 初始标识为  $M_0$ , 对任给的变迁对  $(x, y) \in T \times T$  满足以下关系:

(1) 若  $x > y$  且  $y \not> x$ , 则称为严格序关系, 记作  $x \rightarrow y$ ;

(2) 若  $x \not> y$  且  $y \not> x$ , 则称为排他关系, 记作  $x + y$ ;

(3) 若  $x > y$  且  $y > x$ , 则称为交叉序关系, 记作  $x \parallel y$ ;

(4) 所有关系的集合称为网系统的行为轮廓, 记作  $BP = \{ \rightarrow, +, \parallel \}$ 。

2 物流派送流程模型的一致性检查

随着国民经济的快速发展,社会物流需求也显

著增加,我国物流工程发展迅速。表 1 是从某物流公司派送包裹记录中调取的部分事件日志,表 2 是各个字母所代表的活动事件。

表 1 记录的事件日志

日志 $L_1$		日志 $L_2$	
实例数	迹	实例数	迹
294	$\langle a \ b \ c \ d \ f \ e \ i \ k \ l \rangle$	530	$\langle a \ b \ d \ c \ e \ f \ j \ k \ l \rangle$
1 310	$\langle a \ b \ d \ c \ e \ f \ j \ k \ l \rangle$	815	$\langle a \ c \ b \ e \ d \ f \ g \ h \ k \ l \rangle$
778	$\langle a \ c \ b \ e \ d \ f \ g \ h \ k \ l \rangle$	291	$\langle a \ c \ e \ b \ d \ n \ f \ j \ k \ l \rangle$
137	$\langle a \ b \ c \ d \ m \ o \ f \ e \ j \ k \ l \rangle$	108	$\langle a \ b \ c \ p \ e \ d \ g \ h \ m \ o \ f \ k \ l \rangle$

日志 $L_3$		日志 $L_4$	
实例数	迹	实例数	迹
473	$\langle a \ b \ c \ d \ f \ e \ i \ k \ l \rangle$	285	$\langle a \ b \ c \ d \ m \ o \ f \ e \ j \ k \ l \rangle$
196	$\langle a \ b \ d \ c \ n \ f \ e \ i \ k \ l \rangle$	79	$\langle a \ b \ c \ p \ e \ d \ g \ h \ m \ o \ f \ k \ l \rangle$
115	$\langle a \ b \ c \ p \ e \ d \ g \ h \ m \ o \ f \ k \ l \rangle$	132	$\langle a \ c \ e \ b \ d \ n \ f \ j \ k \ l \rangle$
187	$\langle a \ c \ e \ b \ d \ n \ f \ j \ k \ l \rangle$	314	$\langle a \ b \ c \ d \ f \ e \ i \ k \ l \rangle$

表 2 各个字母所代表的活动事件

字母	活动事件	字母	活动事件	字母	活动事件	字母	活动事件
a	物流派送	e	支付方式	i	现金支付	m	发票有误
b	开具的纸质发票	f	用户收取发票	j	网上支付	n	发票无误
c	物流送货到家	g	POS 支付	k	用户收取包裹	o	重新开具
d	物流递送发票	h	物流 POS 收费	l	用户确认收货	p	包裹完好

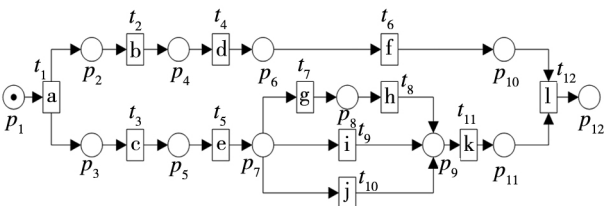


图 1 基于 Petri 网的物流派送流程模型  $N$

由表 1 提供的事件日志,得到合并后的迹及实例数为

$$\alpha_1 = \langle a \ b \ d \ c \ e \ f \ j \ k \ l \rangle^{1\ 840},$$
$$\alpha_2 = \langle a \ c \ b \ e \ d \ f \ g \ h \ k \ l \rangle^{1\ 593},$$
$$\alpha_3 = \langle a \ b \ c \ d \ f \ e \ i \ k \ l \rangle^{1\ 081},$$
$$\alpha_4 = \langle a \ c \ e \ b \ d \ n \ f \ j \ k \ l \rangle^{610},$$
$$\alpha_5 = \langle a \ b \ c \ d \ m \ o \ f \ e \ j \ k \ l \rangle^{412},$$
$$\alpha_6 = \langle a \ b \ c \ p \ e \ d \ g \ h \ m \ o \ f \ k \ l \rangle^{374},$$
$$\alpha_7 = \langle a \ b \ d \ c \ n \ f \ e \ i \ k \ l \rangle^{196}.$$

根据日志中发生频率较高的迹,即  $\alpha_1$ 、 $\alpha_2$ 、 $\alpha_3$ , 构建一个物流派送包裹的流程模型,如图 1 所示。该模型符合一般的物流派送流程,由图 1 可知,实例数较多的迹  $\alpha_1$ 、 $\alpha_2$ 、 $\alpha_3$  可以在模型中完全重放,而对于实例数较少的迹  $\alpha_4$ 、 $\alpha_5$ 、 $\alpha_6$ 、 $\alpha_7$  则与模型之间存在偏差,因此需要将实例数较少的各条迹与流程模型进行一致性检查。每条迹可能会存在多个不同的对齐,例如:迹  $\alpha_4$  与流程模型之间便存在多个不同的对齐,本文在此列出其中三个,即  $\gamma_1$ 、 $\gamma'_1$ 、 $\gamma''_1$ ,  $\gamma_1$  中包含 10 个移动,9 个同步移动和 1 个日志移动  $\langle n \gg \rangle$ ,其成本函数值的总和为 1;  $\gamma'_1$  中包含 12 个移动,9 个同步移动、1 个日志移动  $\langle n \gg \rangle$ 、2 个模型移动  $\langle \gg g \rangle$  和  $\langle \gg h \rangle$ ,其成本函数值的总和为 3;  $\gamma''_1$  中包含 11 个移动,8 个同步移动、2 个日志移动  $\langle b \gg \rangle$  和  $\langle n \gg \rangle$ 、1 个模型移动  $\langle \gg b \rangle$ ,其成本函数值的总和为 3。由于  $1 < 3 = 3$ , 所以最优对齐为  $\gamma_1$ ,因此,只考虑每条迹的最优对

齐。通过将每条迹与流程模型进行一致性检查,得到迹  $\alpha_5$ 、 $\alpha_6$ 、 $\alpha_7$  的最优对齐分别为  $\gamma_2$ 、 $\gamma_3$ 、 $\gamma_4$ ,其对齐成本分别为 2、3、1。

$$\begin{aligned}\gamma_1 &: \frac{a \ c \ e \ b \ d \ n \ f \ j \ k \ l}{a \ c \ e \ b \ d \gg f \ j \ k \ l} \\ \gamma'_1 &: \frac{a \ c \ e \ b \ d \gg \gg n \ f \ j \ k \ l}{a \ c \ e \ b \ d \ g \ h \gg f \ j \ k \ l} \\ \gamma''_1 &: \frac{a \ c \gg e \ b \ d \ n \ f \ j \ k \ l}{a \ c \ b \ e \gg d \gg f \ j \ k \ l} \\ \gamma_2 &: \frac{a \ b \ c \ d \ m \ o \ f \ e \ j \ k \ l}{a \ b \ c \ d \gg \gg f \ e \ j \ k \ l} \\ \gamma_3 &: \frac{a \ b \ c \ p \ e \ d \ g \ h \ m \ o \ f \ k \ l}{a \ b \ c \gg e \ d \ g \ h \gg \gg f \ k \ l} \\ \gamma_4 &: \frac{a \ b \ d \ c \ n \ f \ e \ i \ k \ l}{a \ b \ d \ c \gg f \ e \ i \ k \ l}\end{aligned}$$

### 3 流程模型修复分析

#### (1) 寻找日志移动的可修复库所集

通过事件日志与流程模型的一致性检查,可以看到迹  $\alpha_4$ 、 $\alpha_5$ 、 $\alpha_6$ 、 $\alpha_7$  与流程模型之间存在偏差,偏差成本分别为 1、2、3、1,因此,要想让流程模型能够重放所有的迹,必须要对流程模型进行修复。根据迹与流程模型之间的最优对齐,可以很明显地看出需要修复的部分属于日志移动。对于单个日志移动的修复,可以采用经典的“插入和跳过活动”的方法<sup>[14]</sup><sup>13</sup>对模型进行修复,然而有些日志移动是连续出现的,此时可以根据文献[13]<sup>230</sup>中提到的基于子流程的修复算法进行模型修复。

这种对单个日志移动或者连续日志移动进行模型修复的方法,都是需要插入相应的活动事件。因此在修复之前,寻找最适合插入活动事件的库所也是一个关键环节。在选择适合被插入的库所方面,文献[13]<sup>229</sup>中有简单描述,但并没有给出具体形式的算法,因此,本文针对单个日志移动或者连续日志移动在进行模型修复时如何选择对应库所做了一个算法补充,算法如下

算法 1: 寻找日志移动的可修复库所集

输入: 日志中所有的最优对齐  $\gamma_1$ 、 $\gamma_2$ 、 $\dots$ 、 $\gamma_s$ ; 一个 Petri 网流程模型  $N$

输出: 日志移动的可修复库所集  $P$

Step1 设  $p_1$  为初始库所,  $\gamma_{[i]}$  为第  $i$  个对齐中第  $j$  个移动对,  $P_{[ij]}$  为第  $i$  个对齐中第  $j$  个移动对所

对应的库所,  $m_0$  为初始标识。

Step2 如果  $\gamma_{[i]}$  是一个日志移动,则  $P_{[i]} = (p_1)$ ; 否则,  $m_0[\pi_2(\gamma_{[i]})] > m_1$ , 在  $m_1$  状态下有  $m_1(p_2) = \dots = m_1(p_r) = 1$ , 那么  $P_{[i]} = (p_2, \dots, p_r)$ 。

Step3 当  $2 \leq j \leq |\gamma_i|$  时, 如果  $\gamma_{[ij]}$  是日志移动, 则  $P_{[ij]} = P_{[i,j-1]}$ 。若  $\gamma_{[ij]}$ 、 $\gamma_{[i,j+1]}$ 、 $\dots$ 、 $\gamma_{[i,j+k]}$  是连续的日志移动, 则有  $P_{[ij]} = P_{[i,j+1]} = \dots = P_{[i,j+k]}$ 。

Step4 当  $2 \leq j \leq |\gamma_i|$  时, 如果  $\gamma_{[ij]}$  不是日志移动, 则  $m_{[j-1]}[\pi_2(\gamma_{[ij]})] > m_j$ , 在  $m_j$  状态下,  $m_j(p_{r_1}) = \dots = m_j(p_{r_n}) = 1$ , 则  $P_{[ij]} = (p_{r_1}, \dots, p_{r_n})$ , 然后执行 Step5。

Step5 每一个对齐中, 只保留日志移动对中对应的  $P_{[ij]}$ , 然后删除其余移动对所对应的库所。

Step6 观察所有保留的日志移动对, 如果不同对齐中有相同的日志移动, 即  $\gamma_{i1j_1} = \gamma_{i2j_2} = \dots = \gamma_{ijj}$ ,

其中  $1 \leq i \leq s$ ,  $1 \leq j \leq |\gamma_i|$ , 那么若  $P_{[i1j_1]} \cap P_{[i2j_2]} \cap \dots \cap P_{[ijj]} = P' \neq \emptyset$ , 则更新这些日志移动的库所表示为  $P'$ , 即  $P_{[i1j_1]} = P_{[i2j_2]} = \dots = P_{[ijj]} = P'$ ; 否则, 若交集为空集, 则无需变动。然后执行 Step7。

Step7 整理保留下来的库所集  $P_{[ij]}$ , 得到一个日志移动的可修复库所集  $P = \{P_{[ij]}\}$  算法结束。

根据上述算法, 寻找最优对齐  $\gamma_1$ 、 $\gamma_2$ 、 $\gamma_3$ 、 $\gamma_4$  中日志移动所对应的可用于修复的库所集。首先输入四个最优对齐  $\gamma_1$ 、 $\gamma_2$ 、 $\gamma_3$ 、 $\gamma_4$  和 Petri 网流程模型  $N$ , 第一步交代了字母所代表的含义和初始状态  $m_0(p_1) = 1$ ; 第二步表示在一个最优对齐中, 第一个移动对若是日志移动, 那么它所对应的可修复库所  $P_{[i1]} = (p_1)$ , 否则按照正常的变迁发生规则依次触发后面的活动事件; 第三步说明了出现连续日志的情况下, 对应可修复库所是相同的; 第四步表示在依次寻找最优对齐中所有移动对所对应的库所, 对最优对齐  $\gamma_1$ 、 $\gamma_2$ 、 $\gamma_3$ 、 $\gamma_4$  的执行结果为  $\gamma_{[1]} = \langle a, a \rangle$ ,  $P_{[11]} = (p_2, p_3)$ ;  $\gamma_{[12]} = \langle c, c \rangle$ ,  $P_{[12]} = (p_2, p_5)$ ;  $\dots$ ;  $\gamma_{[15]} = \langle d, d \rangle$ ,  $P_{[15]} = (p_6, p_7)$ ;  $\gamma_{[16]} = \langle n, \gg \rangle$ ,  $P_{[16]} = P_{[15]} = (p_6, p_7)$ ;  $\dots$ ;  $\gamma_{[25]} = \langle m, \gg \rangle$ ,  $P_{[25]} = (p_5, p_6)$ ;  $\gamma_{[26]} = \langle o, \gg \rangle$ ,  $P_{[26]} = P_{[25]} = (p_5, p_6)$ ;  $\dots$ ;  $\gamma_{[34]} = \langle p, \gg \rangle$ ,  $P_{[34]} = (p_4, p_5)$ ;  $\dots$ ;  $\gamma_{[39]} = \langle m, \gg \rangle$ ,  $P_{[39]} = (p_6, p_9)$ ;  $\gamma_{[310]} = \langle o, \gg \rangle$ ,  $P_{[310]} = P_{[39]} = (p_6, p_9)$ ;  $\dots$ ;  $\gamma_{[45]} = \langle n, \gg \rangle$ ,  $P_{[45]} = (p_5, p_6)$ ;  $\dots$ ;  $\gamma_{[410]} = \langle l, l \rangle$ ,  $P_{[410]} = (p_{12})$ ; 第五步是只保留了日志移动对所对应的库所, 此时的执行结果为:  $\gamma_{[16]}$

$= \langle n, \gg \rangle P_{[16]} = (p_6, p_7); \gamma_{[25]} = \langle m, \gg \rangle P_{[25]} = (p_5, p_6); \gamma_{[26]} = \langle o, \gg \rangle P_{[26]} = (p_5, p_6); \gamma_{[34]} = \langle p, \gg \rangle P_{[34]} = (p_4, p_5); \gamma_{[39]} = \langle m, \gg \rangle P_{[39]} = (p_6, p_9); \gamma_{[310]} = \langle o, \gg \rangle P_{[310]} = (p_6, p_9); \gamma_{[45]} = \langle n, \gg \rangle P_{[45]} = (p_5, p_6)$ ; 继而执行第六步, 观察所有保留的日志移动对, 如果不同对齐中有相同的日志移动, 则取它们对应库所的交集, 若交集为空集, 则无需变动。此步骤执行结果为:  $P_{[16]} = P_{[45]} = (p_6); P_{[25]} = P_{[39]} = (p_6); P_{[26]} = P_{[310]} = (p_6); P_{[34]} = (p_4, p_5)$ ; 最后执行第七步, 整理保留下来的库所集, 得到一个日志移动的可修复库所集

$P = \{ P_{[16]} = P_{[45]} = (p_6); P_{[25]} = P_{[39]} = (p_6); P_{[26]} = P_{[310]} = (p_6); P_{[34]} = (p_4, p_5) \}$ 。

## (2) 物流派送流程模型修复及验证分析

找到日志移动的可修复库所集, 便可根据“插入和跳过活动”的方法和基于子流程的修复算法对模型进行修复。通过观察, 日志移动  $\langle p, \gg \rangle$  是

单个日志移动, 且不同对齐中没有与之相同的日志移动, 因此, 修复模型时只需在库所  $p_4$  或  $p_5$  上插入活动事件  $p$  即可, 修复结果如图 2 所示。对于日志移动  $\langle m, \gg \rangle, \langle o, \gg \rangle$  是连续的日志移动, 而且不同对齐中有相同的日志移动, 对应库所交集为  $p_6$ , 这与日志移动  $\langle n, \gg \rangle$  所对应库所一致, 因此, 同时在同一个库所上插入三个活动事件, 需要考虑这三个事件是否具有某种行为关系, 通过分析日志可知, 活动事件  $m, o$  的每次执行都是同时出现且具有一定的顺序关系, 而且并没有与活动事件  $n$  同时出现过, 因此, 判断活动事件  $m, o$  是严格序关系, 且与活动事件  $n$  是排他序关系。因此通过它们之间的行为轮廓关系, 构建一个子流程, 如图 3 所示, 然后将子流程按照基于子流程的模型修复算法, 插入到模型中的  $p_6$  上。故最终修复模型如图 4 所示。

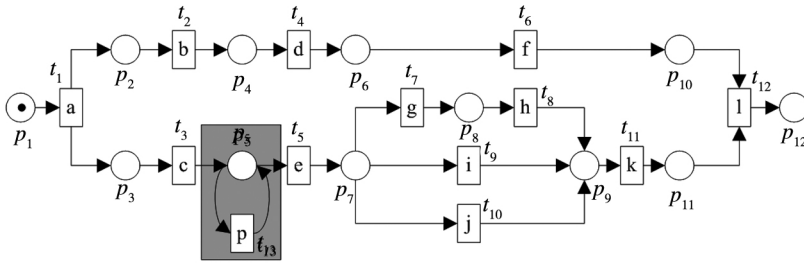


图2 插入活动事件  $p$  的修复模型  $N_1$

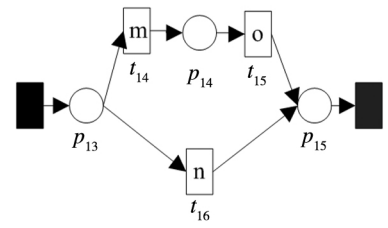


图3 根据行为轮廓关系构建的子流程

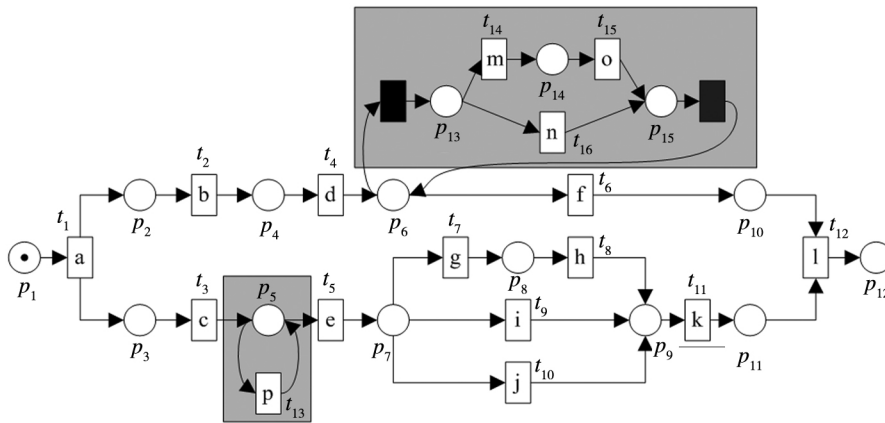


图4 最终修复模型  $N'$

为了验证修复模型的合理性, 计算了修复模型与日志的适合度。其中适合度公式<sup>[1]187</sup>如下

$$\text{fitness}(L, M) = \frac{\sum_{\alpha \in L} c(\gamma_{(\alpha)})}{\text{move}_L(L) + |L| \times \text{move}_M(M)} \quad (1)$$

其中,  $\text{move}_L(L) = \sum_{\alpha \in L} \sum_{x \in \alpha} c(x, \gg)$   $\text{move}_M(M) =$

$$\min_{\sigma_M \in \beta(M)} \sum_{\gamma \in \sigma_M} c(\gg, \gamma)。$$

现将修复模型与迹  $\alpha_4, \alpha_5, \alpha_6, \alpha_7$  进行一致性检查可得, 修复模型可以完整的重放这些迹, 所以根

据成本对齐的定义,偏差成本  $c(\gamma_{(4)}) = c(\gamma_{(5)}) = c(\gamma_{(6)}) = c(\gamma_{(7)}) = 0$ , 故  $\sum_{\alpha \in L} c(\gamma_{(\alpha)}) = c(\gamma_{(4)}) + c(\gamma_{(5)}) + c(\gamma_{(6)}) + c(\gamma_{(7)}) = 0$ , 因此  $\text{fitness}(L, M) = 1 - 0 = 1$ 。从而说明了事件日志与修复模型完全拟合。

## 4 结论

本文主要基于 Petri 网构建了一个物流派送的流程模型,并对事件日志与流程模型进行了一致性检查及模型修复分析。首先,本文根据给定事件日志构建一个流程模型。其次,主要基于成本对齐的方法,根据其成本大小判断事件日志与流程模型之间的偏差大小。在对偏差部分进行模型修复过程中,本文给出了寻找日志移动的可修复库所集的算法,该算法有助于实现模型修复。最后,根据基于子流程的模型修复方法以及行为轮廓关系对模型进行修复,使得流程模型能够完全重放事件日志,并最大可能地与原始模型相似。通过计算得到日志与模型间的适合度为 1,从而也说明修复模型具有合理性。

未来工作中,将致力于在修复模型适合度为 1 的情况下尽可能提高其精确度,并使得修复模型具有普遍适用性。

## 参考文献:

- [1] AALST W V D, ADRIANSYAH A, DONGEN B V. Replaying history on process models for conformance checking and performance analysis [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2012, 2(2): 182-192.
- [2] WIL V D A, WEIJTERS T, MARUSTER L. Workflow mining: discovering process models from event logs [J]. IEEE Educational Activities Department, 2004, 16(9): 1128-1142.
- [3] GOEDERTIER S, MARTENS D, VANTHIENEN J, et al. Robust Process Discovery with Artificial Negative Events [J]. Journal of Machine Learning Research, 2009, 10: 1305-1340.
- [4] ZELST S J V, DONGEN B F V, AALST W M P V D. Event stream-based process discovery using abstract representations [J]. Knowledge & Information Systems, 2018, 54(2): 407-435.
- [5] MASSIMILIANO D L, FABRIZIO M M, WIL M P, et al. Aligning Event Logs and Declarative Process Models for Conformance Checking [M]. Springer, Berlin: Heidelberg, 2012: 82-97.
- [6] WANG H, FENG T, ZHANG J, et al. Consistency check between behaviour models [C]//IEEE International Symposium on Communications & Information Technology. IEEE, 2006: 470-473.
- [7] MUNOZ-GAMA J, CARMONA J. Enhancing precision in process conformance: Stability, confidence and severity [C]//Computational Intelligence & Data Mining. IEEE, 2011: 184-191.
- [8] GRECO G, GUZZO A, PONTIERI L, SACCA D. Discovering expressive process models by clustering log traces [J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(8): 1010-1027.
- [9] WANG LU, DU YU-YUE, QI HONG-DA. Process Model Repair Based on Firing Sequences [J]. Journal of Computer Research and Development, 2018, 55(3): 585-601.
- [10] FAHLAND D, AALST W M P V D. Repairing Process Models to Reflect Reality [C]//International Conference on Business Process Management. Springer, Berlin: Heidelberg, 2012: 229-245.
- [11] SHUGUROV I, MITSYUK A. Iskra: a tool for process model repair [J]. Laboratory of Process-Aware Information Systems, 2015, 27(3): 237-254.
- [12] 吴哲辉. Petri 网理论 [M]. 北京: 机械工业出版社, 2006: 1-22.
- [13] FAHLAND D, VAND A W M P. Model repair-aligning process models to reality [J]. Information Systems, 2015, 47: 220-243.
- [14] POLYVYANY A, AALST W M P V D, HOFSTEDE A H M T, et al. Impact-Driven Process Model Repair [J]. Acm Transactions on Software Engineering & Methodology, 2016, 25(4): 1-60.
- [15] WEIDLICH M, MENDLING J, WESKE M. Efficient consistency measurement based on Behavioural Profiles of Process Models [J]. Software Engineering IEEE Transactions, 2011, 37(3): 410-429.

(责任编辑: 李 丽)