

基于行为轮廓的模块化流程模型修复方法

盛梦君,方贤文,邵叱风

(安徽理工大学数学与大数据学院,安徽 淮南 232001)

摘要:模型修复是业务流程领域中新的优化方法。为了让流程模型对系统的描述更加准确,使模型修复方法系统化,且进一步降低成本,减少以往修复算法的迭代次数和时间复杂度,可运用基于行为轮廓的模块化修复新方法,由分解、过滤、选择、修复、组合五个模块构成。首先,使用极大分解算法对模型进行分解,使流程模型呈片段状,便于筛选出问题片段;然后,使用过滤算法对所有片段进行线性过滤从而缩短通过行为轮廓来选择劣子模型的时间,再使用感应挖掘算法对劣子模型进行修复;接着,再对修复后的劣子模型和优子模型进行组合,从而使得修复后的模型能够重放事件日志,且能够与原始模型尽可能保持行为相似性,最后通过实验分析验证了所提方法的可行性。

关键词:模型修复;过程挖掘;行为轮廓;分解;线性过滤

中图分类号: TP391.9 **文献标志码:** A **文章编号:** 1672-1098(2021)02-0080-07

Modular Process Model Repair Method Based on Behavioral Contours

SHENG Mengjun, FANG Xianwen, SHAO Chifeng

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

Abstract: Model repair is a new optimization method in the field of business process. To make the process model describe the system precisely, and to systemize the model repair method, cut the cost and reduce the iteration times and time consumption of the previous repair algorithm, a new modular repair method based on model decomposition and behavioral contours may be applied, which consists of five modules of decomposition, filter, selection, repair and combination. Firstly, the maximum decomposition algorithm is used to decompose the model into fragments, easy to filter out the problem fragments. Then, the filtering algorithm is used to filter all fragments linearly therefore shortening the time of selecting the inferior sub model, which will be repaired by applying the induction mining algorithm, with the behavioral contours. Finally, the repaired inferior sub model and the superior sub model are combined to make the repaired model able to replay the event log (i. e. the consistent log) and maintain the behavior similarity with the original model as much as possible. The feasibility of the proposed method is verified by experimental analysis.

Key words: process model repair; process mining; behavioral contours; decomposition; linear filtering

收稿日期:2020-11-10

基金项目:国家自然科学基金资助项目(61402011,61572035);安徽省自然科学基金资助项目(1508085MF111,1608085QF149);安徽省高校自然科学基金资助项目(KJ2016A208)

作者简介:盛梦君(1996-),女,安徽合肥人,在读硕士,研究方向:perit网理论。

模型修复是一种新的过程挖掘技术,它将事件日志和流程模型作为输入,通过对日志进行分析来发现流程模型中的偏差,进而对模型进行修复。目前,可以通过事件日志来研究系统的真实情况,运用过程挖掘技术得到流程模型。但是,由于挖掘出能够确切描述现实生活的系统模型是困难的,且流程往往会在系统的生命周期中发生变化。因此,对于模型修复问题的研究具有重要意义。

关于模型修复国内外学者已经做了很多工作。过程挖掘技术一共有三类应用:模型发现、一致性检测和过程增强,模型修复属于过程增强。评价过程模型的质量有四个标准:拟合度、精确度、泛化度和简化度。工作流网中适应性感知控制流修复的方法通过事件日志将不符合要求和符合要求的迹分开,将不合格的迹分解为不合格的子日志,每个子日志都能在相应的子流程工作流网中被发现,然后以这种方式将这些子流程添加到初始模型中,使得修复后的模型能够成功重放给定的事件日志^[1]。这种方法还允许删除不常使用的模型片段,令修复后的模型更简单,但这种删除方式略显暴力,并不是简化模型的最佳方法。基于此,通过一种简化模型结构的方法,使得模型仍然可以重放日志中描述的大多数行为^[2-3]。现有的模型修复方法大多侧重于通过一致性检验找到事件日志与模型之间的偏差,但未侧重修复偏差问题^[4]。基于对齐的概念,建议改变运营分配成本,通过调查所有可能的修复空间,提出并评估了一组最优修复算法,将 Petri 网的修复问题简化为了优化问题^[5]。通过考虑过程挖掘中应用分而治之原则的问题,提出一个过程发现的模式,该修复方法也是利用分而治之的思想进行模块化修复^[6]。通过分析事件日志来研究系统的真实情况,发现真实系统模型,工程师可以使用一致性检验技术来诊断观察事件日志和流程模型之间的差异^[7]。但一味地提高修复后模型的适应度或简化度会让修复成本大大提高,并且修复过程中事件日志遍历的次数过大,与其修复模型不如重新通过事件日志挖掘新模型。因此,进一步降低修复成本,减少以往修复算法的迭代次

数和遍历事件日志的次数以及修复算法的时间复杂度是本文针对的问题。

针对上述问题,本文提出了基于行为轮廓的流程模型修复方法,运用模型分解及线性过滤的方法对模型进行修复,尽可能地保持初始模型结构不改变、原模型可读以及结构良好的性能来进一步降低修复成本,降低修复算法的时间复杂度。此外,还给出了过滤模块以及选择模块中相应的算法,从而较好地提升了模型修复的有效性。

1 基础知识

定义 1^[8](流程模型 Petri 网)满足以下条件的四元组 $N=(P,T,F,C)$ 称作一个流程模型 Petri 网:

- (1) P 是有限非空库所集, T 是有限非空变迁集;
- (2) $P \cup T \neq \varnothing$ 且 $P \cap T = \varnothing$;
- (3) $F \subseteq (P \times T) \cup (T \times P)$ 是 N 中的流关系, $(P \cup T, F)$ 是强连通图;
- (4) $dom(F) \cup cod(F) = P \cup T$, 其中 $dom(F) = \{x \in P \cup T \mid \exists y \in P \cup T : (x, y \in F)\}$ $cod(F) = \{x \in P \cup T \mid \exists y \in P \cup T : (y, x \in F)\}$;
- (5) $C = \{\text{and}, \text{xor}, \text{or}\}$ 是流程网的结构类型。

定义 2^[10](行为轮廓) 设 $P=(A, I, C, F, T)$ 是一个流程模型 $(x, y) \in ((N \cup F) \times (N \cup F))$ 是满足下列关系中的一个:

- (1) 严格序关系 \rightarrow_p , 如果 $x >_p y$ 且 $y >_p x$;
- (2) 排他关系 $+_p$, 如果 $x >_p y$ 且 $y >_p x$;
- (3) 交叉序关系 \parallel_p , 如果 $x >_p y$ 且 $y >_p x$ 。

以上三种关系的集合是 $P=(A, I, C, F, T)$ 的行为轮廓。

定义 3(流程模型) 将流程模型设定为一个元组 $M=(P, T, F, J, \varepsilon, \sigma, M_i, M_f)$, P 是库所集 $\forall p \in P$; T 是变迁集 $\forall t \in T$; F 是连接所有直接跟随库所与变迁之间的流关系, $F \subseteq (P \times T) \cup (T \times P)$; J 是排序的有限集合; ε 是满射函数, 将每个变迁关联到一个排序, $T \rightarrow J$; σ 是指定函数, 将所有变迁映射到活动集或静默变迁中, $T \rightarrow A \cup \{\tau\}$, $\sigma(t_i) = \partial(e_i) = a$, $\forall a \in A, \tau \in T, \sigma(\tau) \notin A$; M_i 是初始标识; M_f 是终止标识。

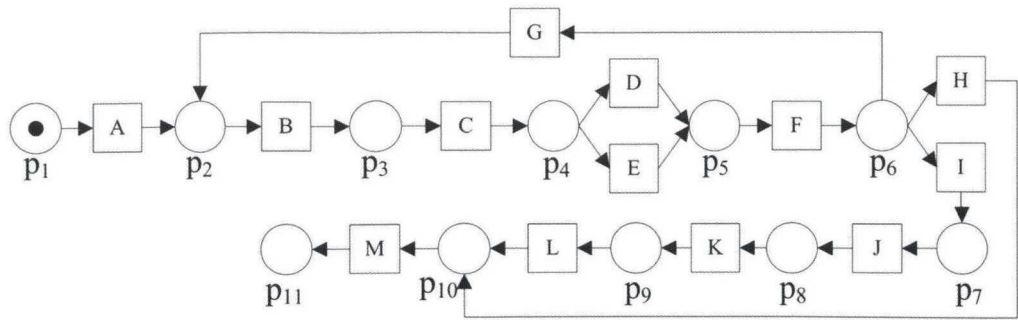


图 1 流程模型工作流网示例

表 1 图 1 示例中字母代表的活动事件

活动	活动名称	活动	活动名称
A	退换货订单生成	H	拒绝
B	提交退换货申请	I	同意
C	商家审核	J	等待物流信息
D	随意审核	K	寄件单号填写
E	彻底审核	L	商家收件
F	决定	M	交易关闭
G	重新发送请求		

定义 4(事件日志,迹)事件日志是一个数学模型,用于捕获有关动态系统执行的历史信息。从形式上讲,定义 Σ 是一组动作集,迹 $l \in \Sigma^*$ 是一个动作序列, $L \in IB\Sigma^*$ 是一个事件日志即一组迹的多元集。其中每组迹都是将系统(即流程实例)的执行描述为一系列观察到的事件。因为可以存在多个具有相同迹的情况,如果迹的频率不相关,将记录日志称为一组迹 $L = \{l_1, \dots, l_n\}$ 。

定义 5^[11](工作流网的有效分解)设 $N \in U_N$ 是一个带标签函数 l 的工作流网。 $D = \{N^1, N^2, \dots,$

$N^n\} \subseteq U_N$ 是一个有效的分解,当且仅当:
—任意 $1 \leq i \leq n, N^i = (P^i, T^i, F^i, l^i)$ 是一个 Petri 网,
—任意 $1 \leq i \leq n, l^i = l \Gamma_{\tau_i}$,
—任意 $1 \leq i < j \leq n$, 则 $P^i \cap P^j = \emptyset$,
—任意 $1 \leq i < j \leq n$, 则 $T^i \cap T^j \subseteq T_v^u(N)$,
— $N = \bigcup_{1 \leq i \leq n} N^i$
 $D(N)$ 对任意网 N 是一个有效分解集。

2 本文方案

(1) 模块化修复框架

本文考虑的修复问题:给定工作流网 N 和事件日志 L ,事件日志 L 不完全符合 N ,提出了一种基于模型分解及线性过滤的模块化修复方法,构造模型 N_r (修复后模型),让 N_r 能够重放 L 。主要思想是基于分而治之的原则筛选出需要修复的片段,修复方法由构建的块组成,每个块执行其中一个步骤,模块化模型修复方法如图 2 所示,构建的块分别为:①分解;②过滤;③选择;④修复;⑤组合;⑥案例分析。弧表示块之间的数据传输。

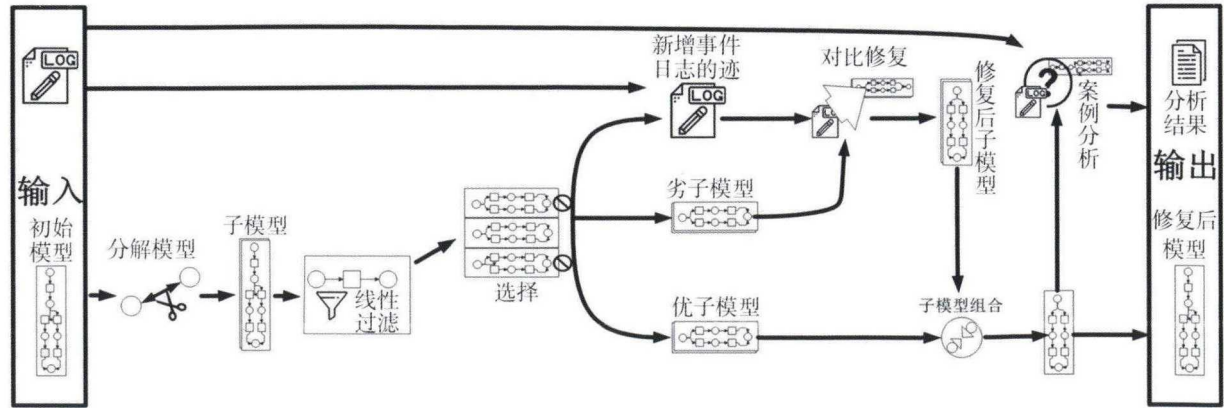


图 2 模块化修复框架

本文基于模块化模型修复的一般方法如下:该方法的输入是一组 (L, N) ,其中 L 是一组事件日志, N 是一个不完全符合事件日志的初始模型。在分解块中,使用极大分解算法^[12]将模型 N 分解成几个片段,分解结果是唯一的。在过滤块中,本文给出了一个包含在过滤块中线性过滤算法,算法将这些模型片段即子模型的直链结构进行过滤筛选,直链结构不执行下一步选择过程,从而直接保留直链的原始模型片段,最后与修复后的模型片段组合。在选择块中,选择算法将分解后的子模型的执行序列与新增的两条事件日志的迹运用行为轮廓相关知识进行比对,符合事件日志的迹即为优子模型,反之为劣子模型。在修复块中,将剩下的劣子模型与事件日志对比分析运用感应挖掘算法^[13]进行修复。在组合块中,将过滤掉的直链结构、优子模型和修复后的子模型进行组合,得到能够重放事

件日志修复后的模型。

(2)使用极大分解方法进行模块化修复

本文提出的模块化修复方法在每个构建的块中使用特定算法。其中分解块中使用的极大分解方法是一个有效分解,极大分解是基于边缘的划分,其定义如下:每个变迁严格地位于单个片段中,具有唯一标签的变迁被放置在片段的边界上,在片段之间需要有因果关系的支持。因此,对于每个变迁,一个在初始网中具有唯一标签的变迁将存在于两个或者多个片段中,最后,拆分保存了网络的初始标识。如果库所在初始网中被标记了,那么它在片段中也是被标记的。在给定的系统网中,有且仅有一个极大分解,且分解结果是唯一的。极大分解网能够很容易地通过融合有相同标签的变迁组合回完整的网。

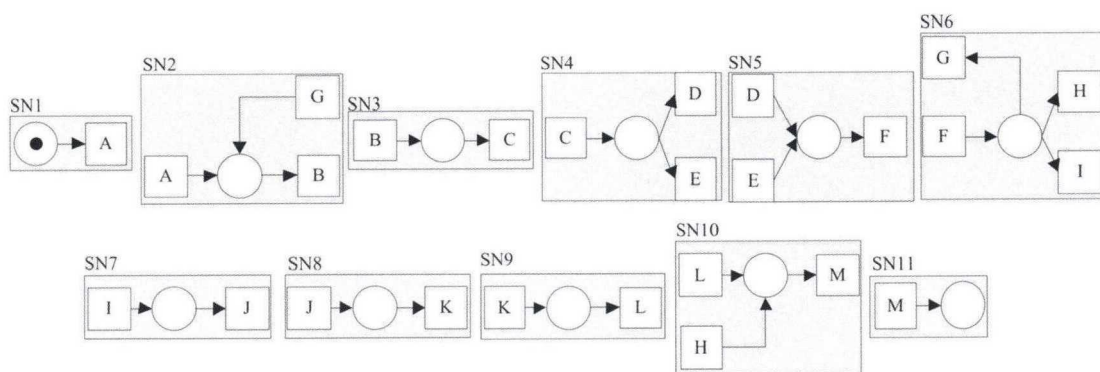


图3 图1中模型的极大分解

将文献中提出的极大分解算法应用到分解块中,并在此基础上开发了过滤块和选择块的算法,其工作原理如下:如果网片段发现可见的变迁,可依据 `decompose`、`handlePlace`、`handlePlace recursively` 三个函数的工作原理调用以后对其进行分解。即 `decompose` 函数从初始网中任意库所开始,当有库所需要遍历时调用 `handlePlace` 函数,调用 `handlePlace` 函数后,调用 `handlePlace recursively` 函数来遍历此流程中库所的前置节点和后继节点。以初始网图1中库所 p_2 为例,调用后三个函数后它的前置节点为变迁 A 和变迁 G ,后继节点为变迁 B ,因此分解后片段为图3中SN2片段。

初始网图1中完整流程模型分解后得到的全部子模型如图3所示;过滤块中运用线性过滤算法再将其进行过滤,若子模型中的库所前、后集个数分别为1或0,则可直接过滤此直链结构,直链结

构不执行下一步选择过程,过滤后的模型片段如图4所示;过滤后选择劣子模型进行修复,选择块中的选择算法将过滤后模型片段的执行序列与增加的事件日志运用Petri网中行为轮廓定义进行比对,符合的为优子模型,反之为劣子模型。过滤算法及选择算法如下:

输入: $ModelSet \mathcal{E}$

输出: $ModelSet \varphi$

$ModelSet \mathcal{E};$

$ModelSet \varphi = Null;$

1) for each part model A in $ModelSet \mathcal{E};$

2) pre set of place = prsp;

3) post set of place = posp;

4) if $|prsp| = |posp| = 1 \vee (|prsp| = 1 \wedge |posp| = 0) \vee (|posp| = 1 \wedge |prsp| = 0)$

5) part model A is null;

```

6) break;
7) else
8) Return (prsp X posp) of A
9) end if
10)  $\varphi$ . add(part model);
11) end for each
12) for each (x,y) in (prsp X posp)
13) for each Trace in New Log
14) if x in Trace and y not in pre of x;
15) return (x,y) , part Model of (x,y)
16) end for each
17) end all

```

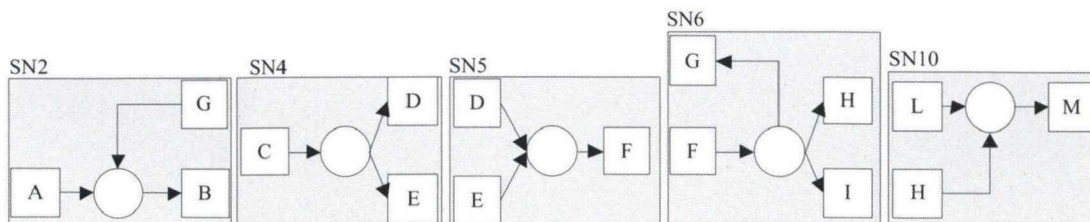


图4 图3 过滤后模型片段

文献[13]已有许多流程发现的算法,本文使用感应挖掘^[14]保证模型发现的适应性(零噪音阈值的不频繁感应挖掘)。它使用顺序和回归的感应推理来构建一个可以转化为工作流网结构的过程树。

假设图3中的模型中变迁C和E与其行为顺序不完全一致,如果二者同时出现,则在事件日志的

迹中E总是在C之前出现,根据行为轮廓,即E和C是严格序关系,记作 $E \rightarrow C$,也就是说事件日志包含迹 $\langle A, B, E, C, F, H, M \rangle$ 和 $\langle A, B, C, D, F, I, J, K, L, M \rangle$,没有C在E之前出现的迹。应用本文提出的选择算法,如图4所示,得到的片段SN4不完全适合日志,因此我们使用感应挖掘算法发现了这个正确的工作流网片段SN4'如图5所示。

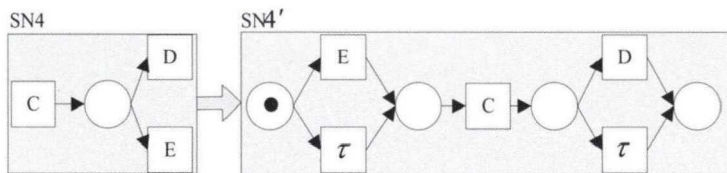


图5 通过感应挖掘发现此模型片段

在发现正确的模型片段并组合所有片段之后,得到了如图6所示的网。这个模型与事件日志的迹 $\langle A, B, E, C, F, H, M \rangle$ 和 $\langle A, B, C, D, F, I, J, K, L, M \rangle$ 完全适应。可以看到修复过的模型不是工作流网,它包含源位置的标识和初始标识中的初始库所。网的最终标识包括在接收器和库所中的两个托肯,这个模型可以通过添加静默的开始和结束变迁很

容易地转换成等效的工作流网。考虑到开始库所和结束库所限制模型的行为并可能产生死锁,从而将它们删除。但在模型运行中增加可能发生的变迁,会稍微降低模型的精度,避免这样的问题最好不要在修复过程中更改片段边界节点变迁,因此可以通过考虑更大的片段来完成修复。

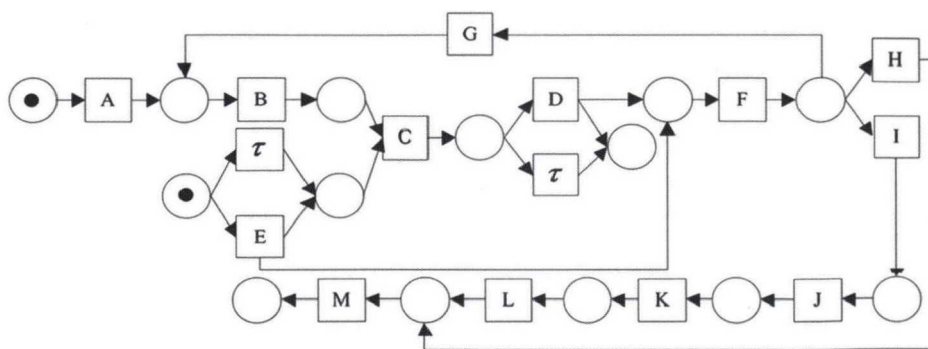


图6 变迁组合后的 Petri 网

3 案例分析

本节以上述退换货系统为例计算修复前后模型的适合度来评估所提方法的可行性,并与 Fahl-and 修复方法进行对比实验。

在给定的事件日志 L 和模型 M 上进行实验评估。如表 2 所示,事件日志 L 一共包含事件总数为:|N|=12 781,其运行次数依次为{58,87,230,270,198,205,109,26,100,12,49,210,180,33,101}。事件日志中总共有 15 条完整有效的执行序列,大小分别为{6,7,10,11,12,15,16,14,13,15,14,13,10,8,7},预定义模型上的所有发生序列大小依次为{7,7,10,10,12,15,16,14,12,15,15,12,10,7,7}。将日志与模型进行最优校准,得到 15 组最优校准。

表 2 网购退换货系统的事件日志

案例	可观测的迹	实例数
1	ABEFHM	58
2	ABECFHM	87
3	ABCDFIJKLM	230
4	ABECDFIJKLM	270
5	ABCDFGBCDFHM	198
6	ABCDFGBCDFIJKLM	205
7	ABCDFGBCDFGBCDFM	109
8	ABECDFGBECDFHM	26
9	ABECDFGBCDFHM	100
10	ABECFGBECFIJKLM	12
11	ABEFGBCDFIJKLM	49
12	ABECFGBECDFHM	210
13	ABECFIJKLM	180
14	ABECDFHM	33
15	ABCDFHM	101

分析得到的 15 组最优校准,这些校准中的不一致由偏差引起,偏差分为日志插入偏差和网 N 跳过偏差,如案例 1 即 n=1 中是日志插入偏差,n=2 中是网 N 跳过偏差。因此,日志与初始模型之间的拟合度为

$$f_{\text{初始}} = 1 - \sum_{i=1}^n \frac{L_i \sum_{j=1}^{m_i} C_{\text{cost}_j}}{(M_L(\varepsilon_i) + M_M(\delta_j))L_i} =$$
$$1 - \frac{2\,632}{21\,996} \approx 0.880\,34 < 1 \tag{1}$$

式中:L 表示每组发生序列的发生次数即实例数,n 表示校准的组数(本文有 15 组),C_{cost} 表示每组校准的偏差成本,这里将所有偏差成本设为 1,M_L 表示日志发生序列,M_M 表示模型中的发生序列,ε 表示日志发生序列的元素个数,δ 表示模型发生序列的元素个数,而日志与修复后模型之间的适合度为 1,因此本文的修复方法具有可行性。

根据本文所提算法的第 12~17 两层嵌套循环可以看出,其算法的时间复杂度为 O(MN)。运用 Fahl-and 方法对案例模型修复结果如图 7 所示,与本文所提方法的对比实验结果如表 3 所示。

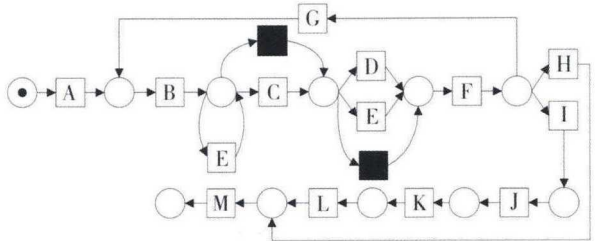


图 7 运用 Fahl-and 方法修复后模型

表 3 修复结果对比数据

方法对比	增加 F	增加 T	增加 τ	时间复杂度
Fahl-and 方法	6	1	2	O(N²)
本文方法	6	0	2	O(MN)

4 结论

本文在已有研究的基础上提出了一种基于模型分解和线性过滤的模块化流程模型修复方法,通过线性过滤算法进一步有效降低了修复的成本和修复算法的时间复杂度,并且尽可能保持了初始模型结构。首先将初始模型运用极大分解算法分解,其次使用过滤算法进行线性过滤,再使用选择算法选择劣子模型然后通过感应挖掘重新发现模型片段,最后对模型片段进行组合得到修复后模型。

本文的方法可能略微降低了模型的精度,对于此问题,计划提出更细微的分解技术以避免片段边界节点的相关问题。基于分解的模型修复也会考虑到拟合度,精确度,泛化度和简化度的组合度量,此外修复是基于模型,对模型分解的相关问题可能非常有帮助,这也是未来可以研究的方向之一。

参考文献:

- [1] FAHLAND D, VAN DER AALST, W M P. Model Repair—Aligning Process Models to Reality [J]. Information Systems, 2015, 47(2): 220–243.
- [2] FAHLAND D, VAN DER AALST, W M P. Simplifying discovered process models in a controlled manner [J]. Information Systems, 2013, 38(4): 585–605.
- [3] DE SAN PEDRO J, CARMONA J, CORTADELLA. Log-Based Simplification of Process Models [C]//Business process management: 13th International Conference. BPM 2015, Innsbruck, Austria, 2015, 9 253: 457–474.
- [4] VAN DER AALST W M P. Process mining: Discovery conformance and enhancement of business process [M]. Berlin Germany: Springer-Verlag, 2011: 64–162.
- [5] ADRIANSYAH A. Aligning observed and modeled behavior [J]. Technische Universitt Dtchen, 2014: 12(3): 29–41.
- [6] Verbeek H M W, VAN DER AALST W M P, Munoz-Gama J. Divide and conquer: a tool framework for supporting decomposed discovery in process mining [J]. Computer Journal, 2017, 60(11): 1 649–1 674.
- [7] VAN DER AALST W M P. Process Mining: Data Science in Action [J]. Springer, Berlin, Heidelberg, 2016, 10(2): 3–23.
- [8] 吴哲辉. Petri 网理论 [M]. 北京: 机械工业出版社, 2006: 6–22.
- [9] SMIRNOV S, WEIDLICH M, MENGLING J. Business Process Model Abstraction based on Behavioral Profiles [C]//International Conference on Service-Oriented Computing (ICSOC 2010). San Francisco, US, 2010: 1–16.
- [10] VAN DER AALST W M P. Decomposing Petri Nets for Process Mining: A Generic Approach [J]. Distributed and Parallel Databases, 2013, 31(4): 471–507.
- [11] MITSYUK A A, LOMAZOVA I A, et al. Process model repair by detecting unfitting fragments [C]//Supplementary Proceedings of the Sixth International Conference on Analysis of Images, Social Networks and Texts (AIST 2017). CEUR-WS. org, Moscow, Russia, 2017: 301–313.
- [12] VANDONGEN B F, CARMONA J, CHATAIN T. A unified approach for measuring precision and generalization based on anti-alignments [C]//14th International Conference on Business Process Management (BPM). Fed Univ State Rio de Janeiro, Rio de Janeiro, Brazil, Springer, 2016: 39–56.
- [13] LEEMANS S J J, FAHLAND D, VAN DER AALST W M P. Discovering block-structured process models from incomplete event logs [C]//Application and theory of petri nets and concurrency: 35th International Conference. PETRI NETS 2014, Tunis, Tunisia, Proceedings. 2014: 91–110.
- [14] DONGEN B F V, MEDEIROS A K A D, VERBEEK H M W, et al. The ProM Framework: A New Era in Process Mining Tool Support [C]//International Conference on Application & Theory of Petri Nets. Springer-Berlin Heidelberg, 2005, 3 536: 444–454.
- [15] HOMPES B F A. On Decomposed Process Discovery: How to Solve a Jigsaw Puzzle with Friends [D]. Master's thesis, Eindhoven University of Technology, De Zaal, Eindhoven, Netherlands, 2014.
- [16] Aalst W, Verbeek H. Process Discovery and Conformance Checking Using Passages [J]. Fundamenta Informaticae, 2014, 131(1): 103–138.
- [17] GNTHER C W, VAN DER AALST W M P. Fuzzy mining: Adaptive process simplification based on multi-perspective metrics [C]//5th International Conference on Business Process Management. Brisbane, Australia, BPM, 2007, 4 714: 328–343.
- [18] VAN DER AALST W M P, ADRIANSYAH A A, VANDONGEN B F. Replaying History on Process Model for Conformance Checking and Performance Analysis [J]. Wiley Interdisciplinary Reviews – Data Mining and Knowledge Discovery, 2012, 2(2): 182–192.
- [19] MACEDO N, JORGE T, CUNHA A. A Feature-Based Classification of Model Repair Approaches [J]. IEEE Transactions on Software Engineering, 2017, 43(7): 615–640.
- [20] POLYVYANYI A, VAN DER AALST W M P, Terhofstede A H M, et al. Impact-Driven Process Model Repair [J]. ACM Transactions on Software Engineering and Methodology, 2017, 25(4): 1–60.

(责任编辑:李 丽,范 君)