

一种基于校准的模型问题域识别方法

王 路,杜玉越

(山东科技大学 信息科学与工程学院,山东 青岛 266590)

摘 要:以 Petri 网作为描述流程模型的形式化方法,利用工作流的四种基本结构描述流程中活动之间的关系;结合流程模型质量维度中的拟合度、简化度、精确度和泛化度,给出了流程模型中执行序列与事件日志中迹的校准方法。通过所提出的校准动作序列与流程模型的映射方法,可以识别流程模型的问题域。通过网上购物的实例验证了方法的有效性和实用性。

关键词:Petri 网;流程模型;质量维度;校准;动作映射;大数据

中图分类号:TP311.13

文献标志码:A

文章编号:1672-3767(2015)01-0042-05

An Alignment-Based Identifying Method of the Problem Areas Within Process Models

Wang Lu, Du Yuyue

(College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China)

Abstract: Petri nets were employed as a formalization method of process models in this research and the four basic structures of workflows were used to describe the relationships in the processing activities. A process describing on-line shopping was used as a running example to explain how to identify problem areas within a process model. The fitness, simplicity, precision, generalization of model quality dimensions were given concern to align execution sequences in process models and traces in event logs. Based on the alignment move mapping, a process model was proposed, and the negative effect parts were identified. The validity and effectiveness of the proposed method were thus illustrated through the analysis of the online shopping examples.

Key words: Petri net; process model; quality dimension; aligning; move mapping; big data

近年来,信息系统与其所支持的业务流程联系越来越密切,越来越多的数据被记录并存储在信息系统中。流程挖掘的目标就是利用这些事件数据挖掘与流程有关的信息^[1],如利用企业信息系统记录的事件自动发现一个流程模型。

事件日志作为流程挖掘的输入数据,是流程执行过程中发生的事件序列集合。日志中的每一个事件对应一个活动,并且与一个特定的实例相关。实例中的一条事件序列称为一条迹。除了活动,日志中也可能包含事件的其他信息,如执行或发起活动的资源(即人力或设备),事件发生的时间或与事件相关的其他属性(如订单大小)。

现有流程挖掘技术有流程发现、一致性检查和流程改进^[1]。流程发现是利用事件日志中观察到的行为来创建流程模型。一致性检查是用目前存在的流程模型和对应的事件日志作对比,找出执行的流程与流程

收稿日期:2014-10-28

基金项目:国家自然科学基金项目(61170078,61472228);教育部高等学校博士学科点专项科研基金博导类项目(20113718110004);青岛市科技计划基础研究项目(13-1-4-116-jch);山东科技大学科研创新团队支持计划项目(2011KYTD102)

作者简介:王 路(1989—),女,山东泰安人,硕士研究生,主要从事流程挖掘、Petri 网等方面的研究。

杜玉越(1960—),男,山东聊城人,教授,博士生导师,主要从事 Petri 网、工作流、过程挖掘等方面的研究,本文通信作者。E-mail:yydu001@163.com

模型之间存在的偏差。流程改进是利用事件日志中包含的信息对流程模型进行扩展和改善。这三种流程挖掘技术可以用来进行符合性检查,分析瓶颈,预测延迟,为流程执行时间的最小化提供建议。

1 Petri 网基本知识

目前,存在很多描述流程模型的形式化语言,可以从控制流的角度描述流程,比如,业务流程建模标记法^[2]、因果网^[1]、事件驱动流程链^[3]等。

Petri 网能够作为分布式系统建模和分析的工具,并且具有严格的数学定义以及强大的图形表达能力^[4]。Petri 网不仅能够描述流程的静态结构,还可以模拟流程运行过程中的动态行为。对于具有并行、并发、异步、分布和不确定性等性质的信息系统,可以利用 Petri 网对其进行有效描述和分析。下面给出 Petri 网的基本概念以及 Petri 网建模中常用的基本路由结构。

定义 1 四元组 $\Sigma = (S, T; F, M)$ 称为一个 Petri 网,当且仅当

① $N = (S, T; F)$ 为一个网。

② $M: S \rightarrow N. M$ 称为网 Σ 的一个标识。

③ Σ 具有下面的变迁发生规则:

(a) 对于变迁 $t \in T$, 如果 $\forall s \in \cdot t: M(s) \geq 1$, 则说变迁 t 在标识 M 下使能, 记为 $M[t >$ 。

(b) 若 $M[t >$, 则在标识 M 下, 变迁 t 可以发生, 从标识 M 引发变迁 t 得到一个新的标识 M' , 记为 $M[t > M'$, 对 $\forall s \in S$,

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

在 Petri 网的建模过程中,经常用到工作流的四种基本结构如图 1 所示:顺序结构、并行结构、选择结构、循环结构。这四种结构是流程模型的基本结构,流程所有的执行过程都可以由这四种结构组合而成。

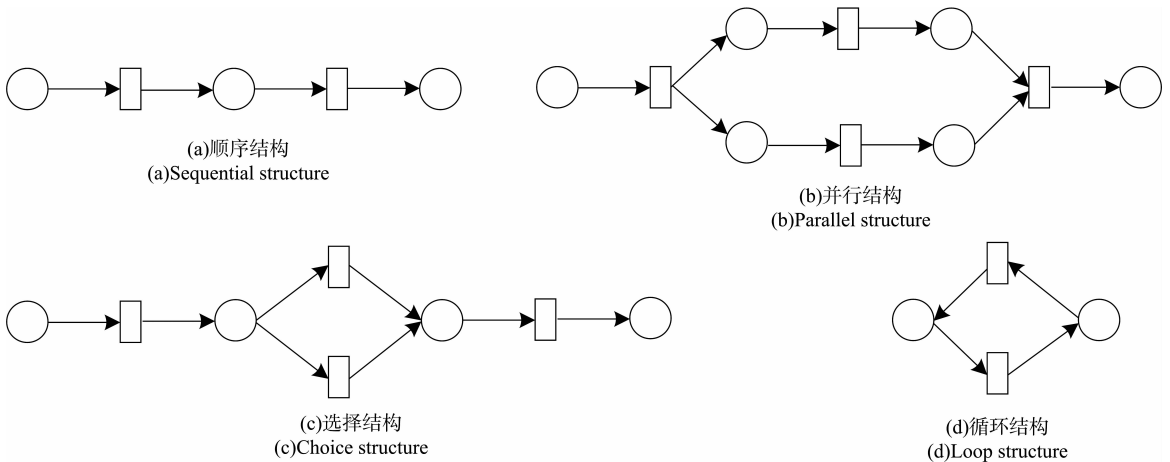


图 1 工作流的四种基本结构

Fig. 1 Four basic structures of workflows

2 流程模型的质量

流程模型描述流程的行为以及可能的执行过程,可以用于性能分析、项目规划,或仅作为了解流程的工具,以促进利益相关者之间的沟通^[1]。当评价流程模型的质量时,通常要考虑以下四个质量维度^[1,5]。

拟合度:量化流程模型正确重演事件日志中案例的程度。模型不可能有完美的拟合度,模型中包含的活动在事件日志没有记录,日志中包含的活动在模型中没有描述,或模型中描述的活动顺序和事件日志中记录

的活动顺序不同。

简化度:指奥卡姆剃刀定律,即最好的模型是能够描述事件日志中行为的最简单模型。

精确度:测量模型是否能够禁止日志中没有记录的行为的发生,如果模型不是“欠拟合”就认为该流程模型是精确的。

泛化度:指模型能够正确重演流程中未来可能发生且没有被记录的行为的程度。如果模型不是“过拟合”,就认为该流程模型是泛化的,即模型不仅说明了日志中记录的案例,而且描述了流程产生的事件日志。

现有的流程挖掘算法还不能同时满足这四个质量维度的要求,但是,一致性检测技术可以用来检查和诊断模型中行为和事件日志中行为的不同。由于一致性检测和拟合度测量的联系密切,相比之下和精确度及泛化度的测量联系较小,因此,通过一致性检测能够识别对模型质量产生消极影响的部分^[6],从而对该部分模型进行修正以提高流程模型的质量,这就是一种流程改进。

3 识别模型问题域

一致性检测技术可以识别事件日志中的行为与模型存在的偏差^[1]。一致性检测将事件日志和流程模型进行匹配,以便找到要消除日志和模型偏差需要做的最小集合的改变^[7]。本节主要介绍如何利用校准识别模型中拟合度比较低的部分。

3.1 网上购物流程

本研究以网上购物为例说明如何识别模型的问题域,该流程的 Petri 网描述如图 2 所示。顾客可以浏览网页选择商品(*select items*),选择完商品后可以直接确认订单(*confirm order*),也可以加入购物车(*add cart*)。然后,可以继续选择商品加入购物车(*select items*),选择完商品或加入购物车后如果不想购买就可以取消订单(*cancel*),网上购物流程结束。购物完成后要确认订单(*confirm order*),确认订单时要确认购买清单(*confirm list*),确认送货地址(*confirm address*)。确认订单后,可以选择货到付款(*pay on delivery*),也可以选择网上支付(*pay online*)。付款成功/失败分别用两个不可见变迁 t_6 和 t_7 来描述,成功付款后就可

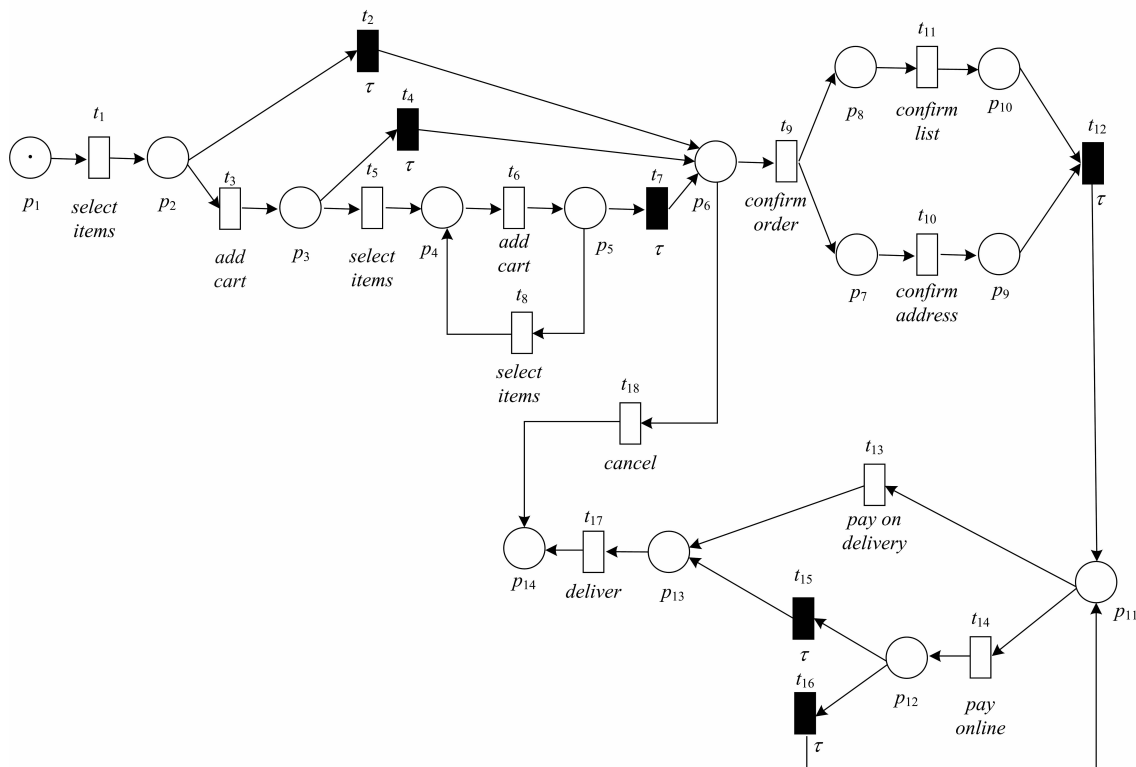


图 2 网上购物流程

Fig. 2 Online shopping process

以进行配送(*deliver*),网上购物流程结束。

3.2 校准流程模型和事件日志

为了找到模型能够重演日志所要做的改变的最小集合,需要为每一条迹 $l \in L$ 在模型 M 中找到一条执行序列 σ ,使 σ 与 l 尽可能相似。当校准 σ 与 l 时,如果在 σ 中执行的活动在 l 中没有记录,称之为模型动作(*model moves*),如果 l 中记录的事件在 σ 中没有执行,称之为日志动作(*log moves*),如果 σ 中执行的活动能够在 l 中找到与之匹配的事件,称之为同步动作(*synchronous moves*)。

假设事件日志中的迹 $l = \langle add\ cart, confirm\ order, confirm\ address, confirm\ list, pay\ online, deliver \rangle$ 与图 2 所示模型中的执行序列 $\sigma = \langle select\ items, add\ cart, \tau, confirm\ order, confirm\ list, \tau, pay\ online, \tau, deliver \rangle$ 进行校准。序列 σ 中的活动 *select items* 在迹 l 中没有出现,如果要校准 l 和 σ 时,需要跳过活动 *select items*,跳过的活动就是模型动作。符号 τ 表示在模型中执行的不可见活动,根据定义该活动不能在事件日志的迹中找到。迹 l 中的事件 *confirm address* 没有在序列 σ 中执行,所以在校准时需要插入活动 *confirm address*,插入的活动就是日志动作。活动 *add cart, confirm order, confirm list, pay online* 和 *deliver* 都包含在迹 l 和序列 σ 中,这些活动的执行就是同步动作。迹 l 和序列 σ 的校准如图 3 所示。

l	>>	<i>add cart</i>	>>	<i>confirm order</i>	<i>confirm address</i>	<i>confirm list</i>	>>	<i>pay online</i>	>>	<i>deliver</i>
σ	<i>select items</i>	<i>add cart</i>	τ	<i>confirm order</i>		<i>confirm list</i>	τ	<i>pay online</i>	τ	<i>deliver</i>
	t_1	t_3	t_4	t_9	>>	t_{11}	t_{12}	t_{13}	t_{15}	t_{17}

图 3 日志中的迹 l 与模型中执行序列 σ 的校准

Fig. 3 An alignment between trace l in an event log and execution σ in the model

3.3 校准动作映射

本节主要说明如何识别模型中对模型质量有消极影响的部分。拟合度是质量维度中最重要的部分,所以判断模型质量时主要考虑模型的拟合度^[8]。为了方便,用大写字母 A, B 等表示迹和模型中的活动。

模型和事件日志校准确定了哪些活动应该跳过,如果插入的话要确定在模型运行中活动要插入到模型的一个特定位置,因此校准提供的模型和事件日志的不一致信息是局部的。图 4 描述:①需要修正的 Petri 网模型使它能够正确描述网上购物流程;②迹 $l = \langle A, F, G, H, J, K \rangle$ 和这个不正确模型的一个最优校准;③校准中的动作和 Petri 网模型中节点的一个映射,这样可以清楚知道流程模型中存在问题的部分。

校准动作映射如图 4 所示,模型和日志动作指出需要修正的问题,校准动作映射说明需要修正的问题处在 Petri 网模型中的位置。通过把模型动作映射为可跳过的活动对模型进行修正,而日志动作可以通过在模型中的其他活动之前或之后插入额外活动对模型进行修正。校准中的动作通过以下方式映射成为 Petri 网模型中的活动:同步动作(如 A)和模型动作(如 B)映射成模型中相应的活动,日志动作(如 H)映射成该活动之前在模型中执行的最后一个活动(H 映射为 G)和该活动之后在模型中执行的第一个活动(J 映射为 I)。

根据图 4 给出的校准动作映射就可以对有问题的流程模型进行修正,但是,有时造成流程模型有问题的原因是非局部的,如图 5 所示,需要对其进行修正使模型能够符合上面提出的网上购物的例子。该模型的问题是虚框内所描述的顺序结构,需要将其改成并行结构,但是校准动作映射只能识别出模型变迁与日志活动不匹配,不能识别出模型的结构是否存在问题。

要对校准动作映射进行扩展,使其能够识别模型结构存在的问题。所有的对基本结构中结点的映射也要与结构本身进行映射,比如,对于同步、日志和模型动作与 H 和 J 映射如图 5 所示,也要与 H 和 J 之间的关系顺序结构映射。目前,没有办法确定模型中的哪些结点为修正问题提供最好的解决办法,除非尝试所有

的解决方法,然后评价所得模型质量。

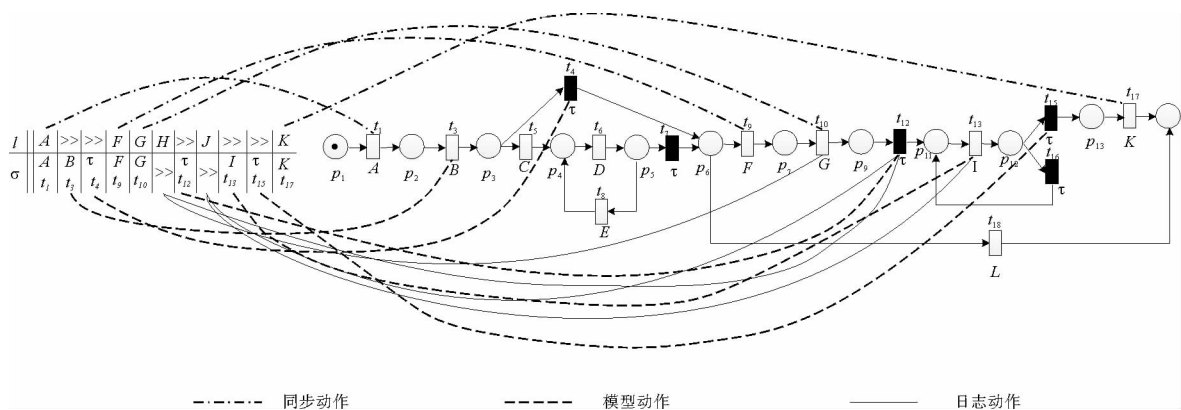


图 4 迹 $l = \langle A, F, G, H, J, K \rangle$ 与模型 1 的最优校准及校准动作映射

Fig. 4 An optimal alignment between the trace $l = \langle A, F, G, H, J, K \rangle$ and the process model 1 in need of repair, together with a mapping of the alignment onto model elements

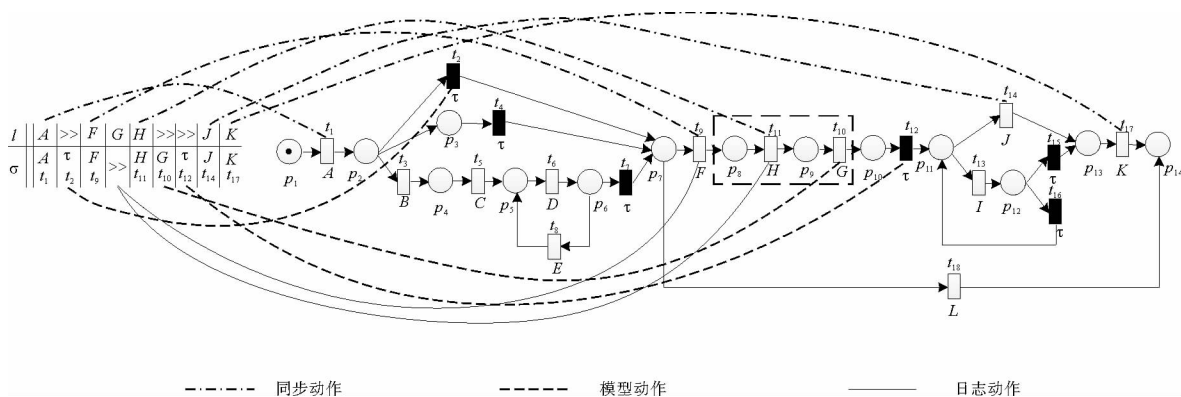


图 5 迹 $l = \langle A, F, G, H, J, K \rangle$ 与模型 2 的最优校准及校准动作映射

Fig. 5 An optimal alignment between the trace $l = \langle A, F, G, H, J, K \rangle$ and the process model 2 in need of repair, together with a mapping of the alignment onto model elements

对校准动作映射的扩展能够识别流程模型的哪一部分对模型的拟合度存在负面影响,同时也指出是否要在模型中插入额外的活动或跳过模型中执行的活动,利用这些信息可以确定如何对流程模型进行修正。

4 结束语

利用 Petri 网语言对网上购物流程的描述,给出了事件日志中记录的行为与流程模型中执行序列的校准方法,通过校准可以识别出问题模型中需要跳过或插入的活动。由于事件日志与模型校准提供的是局部信息,不能够识别出模型结构存在的问题。因此,提出校准动作与流程模型的映射,通过校准动作映射提供的信息正确识别出错误模型中的问题。

进一步的工作是利用校准动作映射提供的信息,通过移除模型中的活动、在模型中插入活动或改变模型的基本结构,对模型中不正确的问题部分进行修正。