

文章编号: 1004-4353(2019)01-0075-05

基于 Petri 网行为轮廓的网上购物 流程挖掘方法

王倩倩, 王丽丽

(安徽理工大学 数学与大数据学院, 安徽 淮南 232001)

摘要: 为了满足企业业务流程模型的运行效率、完善网上购物流程,提出了一种基于 Petri 网行为轮廓的网上购物流程挖掘方法. 首先根据某网上购物系统提供的事件日志,利用 Petri 网行为轮廓的弱序关系,设计出相应的流程模型;然后通过计算模型与事件日志之间的服从程度,对初始模型进行改进和优化;最后用实例分析说明了本文挖掘方法的可行性.

关键词: 流程挖掘; 行为轮廓; 服从性

中图分类号: TP391.9

文献标识码: A

The online shopping process mining based on behavioral profile of Petri net

WANG Qianqian, WANG Lili

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

Abstract: In order to meet the operational efficiency of the enterprise business process model and improve the online shopping process, an online shopping process mining method based on behavioral profile of Petri net is proposed. Firstly, according to the event log provided by an online shopping system, the corresponding process model is designed by using the weak order relationship of the behavioral profile of Petri net. Then the degree of compliance between the model and the event log is calculated, and the initial model is improved and optimized. The last example analysis shows that the algorithm has feasibility.

Keywords: process mining; behavioral profile; compliance

0 引言

流程挖掘是一种从实际业务执行日志中发现结构化流程信息的过程^[1]. 流程挖掘技术能够从现代信息系统普遍产生的事件日志中抽取信息,通过建立清晰的流程模型,保证构建的流程模型与实际流程执行过程保持一致,从而监测和改进原过程模型^[2]. 目前,过程发现技术面临的主要问题是事件日志往往不能包含一个业务流程模型的所有信息(如不完整日志),从而导致所挖掘的流程模型与实际业务流程不一致. 文献[3]给出了流程模型与事件日志一致性程度的 3 种度量维度(合适度、精度、泛化),但是这 3 种度量维度不能明确表达活动之间的因果关系和并行关系. 文献[4]提出了一种从日志的活动发生顺序来推断业务流程模型的算法,但该方法只能发现简单的业务流程模型. M. Rovani 等^[5]和高立哲等^[6]基于流程挖掘 ProM 平台和事件日志,挖掘医疗业务流程,但该方法仍不适合结构复杂的流程模

型. 针对以上问题, M. Weidlich 等^[7]以 Petri 网行为轮廓作为基线来计算新的服从指标, 该方法通过有效计算行为轮廓, 避免了基于状态空间指标的性能问题, 从而能够清晰表达活动之间复杂的业务流程关系. 本文受文献^[7]的启发, 提出一种基于 Petri 网行为轮廓的网上购物流程挖掘方法, 并结合实例验证本文方法的有效性.

1 基本概念

定义 1^[8] (流程模型) 一个网 $P = (A, a_i, a_0, C, F, T)$ 称作流程模型, 满足以下条件:

- 1) A 是活动节点, $A \neq \emptyset$; C 是控制节点, $C \neq \emptyset$.
- 2) $a_i, a_0 \in A$, a_i 是起始活动, a_0 是结束活动.
- 3) $F \subseteq ((A \setminus \{a_0\}) \cup C) \times ((A \setminus \{a_i\}) \cup C)$ 是流关系.
- 4) $T: C \mapsto \{\text{and}, \text{or}, \text{xor}\}$ 是关于控制节点类型的函数.

定义 2^[9] (流程模型的弱序关系) 若 $P = (A, a_i, a_0, C, F, T)$ 为一个流程模型, ϵ_P 是执行序列的集合, 弱序关系 $\succ_P \subseteq (A \times A)$ 包含所有 (x, y) , 存在一个执行序列 $\sigma = n_1, \dots, n_m$, $\sigma \in \epsilon_P$, $\forall i, j \in \{1, \dots, m-1\}$ 且 $j < k \leq m$, 使得 $n_i = x, n_j = y$.

定义 3^[9] (流程模型的行为轮廓) 若 $P = (A, a_i, a_0, C, F, T)$ 为一个流程模型, $\forall (x, y) \in (A \times A)$ 满足下面关系:

- 1) 若 $x \succ_P y$ 且 $y \not\prec_P x$, 则称 x 和 y 之间为严格序关系, 记作 $x \rightarrow_P y$;
- 2) 若 $x \not\prec_P y$ 且 $y \not\prec_P x$, 则称 x 和 y 之间为排他性关系, 记作 $x +_P y$;
- 3) 若 $x \succ_P y$ 且 $y \succ_P x$, 则称 x 和 y 之间为交叉序关系, 记作 $x \parallel_P y$.

称所有关系的集合为流程模型的行为轮廓, 记作 $B_P = \{\rightarrow_P, +_P, \parallel_P\}$. 若 $y \succ_P x$ 且 $x \not\prec_P y$, 则称 x 和 y 之间为逆严格序关系, 记作 $x \rightarrow_P^{-1} y$.

定义 4^[7] (执行序列, 事件日志) 若 Petri 网 $P = (A, a_i, a_0, C, F, T)$ 为一个流程模型, σ_P 为 P 的一个发生序列 $\sigma_P \in \{a_i\} \cdot A^* \cdot \{a_0\}$, 事件日志 L 是执行序列 σ_P 的集合.

定义 5^[9] (流程模型的因果行为轮廓) 若 $P = (A, a_i, a_0, C, F, T)$ 为一个流程模型, 则:

- 1) 共现关系 $\gg_P, \forall (x, y) \in (A \times A)$ 满足 $\forall \sigma = n_1, \dots, n_m, \sigma \in \epsilon_P, n_i = x, n_j = y, \forall i, j \in \{1, \dots, m\}$, 且 $i \neq j$.
- 2) 流程模型的因果行为轮廓 $B_P^+ = B_P \cup \{\gg_P\}$.

定义 6^[7] (日志的行为轮廓) 若 $L_P = n_1, \dots, n_m$ 是流程模型 $P = (A, a_i, a_0, C, F, T)$ 的一个日志, 对于 $\forall (x, y) \in (A_L \times A_L)$ 满足:

- 1) 若 $x \succ_L y$ 且 $y \not\prec_L x$, 则称 x 和 y 之间为严格序关系, 记作 $x \rightarrow_L y$;
- 2) 若 $x \succ_L y$ 且 $y \succ_L x$, 则称 x 和 y 之间为交叉序关系, 记作 $x \parallel_L y$.

称所有关系的集合为日志的行为轮廓, 记作 $B_L = \{\rightarrow_L, \parallel_L\}$.

定义 7^[7] (日志的因果行为轮廓) 若 $L_P = n_1, \dots, n_m$ 是流程模型 $P = (A, a_i, a_0, C, F, T)$ 的一个日志, 对于 $\forall (x, y) \in (A_L \times A_L)$, 共现关系 $\gg_L = (A_L \times A_L)$, 则日志的因果行为轮廓表示为 $B_L^+ = B_L \cup \{\gg_L\}$.

定义 8^[7] (包容谓词) 给定两个行为关系 $R, R' \in \{\rightarrow, \rightarrow^{-1}, +, \parallel\}$, 当 $R = R'$ 和 $R = \parallel$, ($R \in \{\rightarrow, \rightarrow^{-1}\} \wedge R' = +$) 中有一个成立时, 则称 R 和 R' 满足包容谓词关系, 记作 $S(R, R')$.

定义 9^[7] (执行顺序服从性) 若 $L_P = n_1, \dots, n_m$ 是流程模型 $P = (A, a_i, a_0, C, F, T)$ 的一个日志, $\forall (x, y) \in SR$ 且 $SR \subseteq (A_L \times A_L)$ 是日志 L_P 的行为关系, 即 $\forall R \in (B_P \cup \{\rightarrow_P^{-1}\})$, $R' \in (B_L \cup \{\rightarrow_L^{-1}\})$, $[(xRy \wedge xR'y) \Rightarrow (S(R, R') \vee (x = y \wedge R = \parallel_P \wedge R' = +_L))]$, 则流程模型与日志的执行

顺序服从性可表示为

$$\epsilon^{cLP} = \frac{|SR|}{|(A_L \times A_L)|}$$

2 基于行为轮廓的网上购物挖掘算法

本文提出的基于 Petri 网行为轮廓的网上购物流程挖掘算法,首先通过分析事件日志的活动关系确定行为轮廓关系,以此挖掘出一个初始的流程模型;然后测量流程模型与事件日志的服从程度,并将其与标准的服从程度值进行对比,以此来判断该流程模型是否需要进一步优化。

基于 Petri 网行为轮廓的网上购物流程挖掘算法的步骤如下:

输入:事件日志,服从性阈值

输出:网上购物 Petri 网模型

步骤 1 对事件日志 $L = \{n_1, n_2, \dots, n_i\}$, $i = 1, 2, \dots, k$ 进行预处理,按照发生频数 $f(x)$ 大小进行排序。

步骤 2 选择发生频数 $f(x) > M$ 的几条日志序列(M 为经验阈值),然后根据日志序列中活动间行为轮廓的弱序关系构造序列活动关系表。

步骤 3 根据活动关系表构造活动间行为轮廓表,然后利用行为轮廓表建立初始 Petri 网模型。

步骤 4 根据行为轮廓计算事件日志与初始模型的服从性。以 α 为服从性的标准值,若服从性高于 α ,则输出模型;反之,返回到步骤 2 进行模型优化,且令步骤 2 中的 $M = M - N$,其中 N 为每次减小的步长。

步骤 5 重复步骤 4,当服从性大于 α 或 $M - N < 0$,输出模型。

3 实例分析

系统中发生的事件被记录在日志中,它对理解系统的活动至关重要,特别是在用户交互较少的应用程序中。由于事件日志包含事件的属性较多,本文仅抽取事件中的活动序列来挖掘网上购物流程。

本文以某 app 网上购物系统记录的事件日志为例进行挖掘。网上购物业务流程信息系统中日志处理后所包含的活动 $P = \{A, B, C, D, E, F, G, H, I, J, K, L, Y\}$,其中 $\langle A, B, C, E, F, Y, G, I, L \rangle$ 为事件日志中的一个发生序列。以下是该 app 处理后的一个事件日志:

$$\Phi = [\langle A, B, C, E, F, Y, G, I, L \rangle^{80}, \langle A, B, C, E, F, Y, G, H, J, K, L \rangle^{56}, \langle A, B, D, E, F, Y, G, H, J, K, L \rangle^{33}, \langle A, B, C, E, F, Y, G, H, J, K, L \rangle^{20}, \langle A, B, C, E, Y, F, G, I, L \rangle^{63}, \langle A, B, D, E, Y, F, G, I, L \rangle^{43}, \langle A, B, D, E, G, I, L \rangle^{49}]$$

其中 $A, B, C, D, E, F, G, H, I, J, K, L, Y$ 分别表示登录、购买商品、在线支付、到付、下单、接单、发货、因质量拒货、收货、理赔、取消订单、交易结束、略过。这个事件日志中包含 7 个不同的发生序列,每个序列发生的次数不同,序列的上标表示发生次数,例如 $\langle A, B, C, E, F, Y, G, I, L \rangle^{80}$ 表示序列 $\langle A, B, C, E, F, Y, G, I, L \rangle$ 发生了 80 次。

下面利用本文提出的流程挖掘算法,对上述 app 给出的网上购物事件日志进行挖掘,以此验证本文提出算法的可行性。

1) 将所有的日志序列按照发生序列及发生次数进行排序: $\{ABCEFYGIL(80), ABCEYFGIL(63), ABCEFYGHJKL(56), ABDEGIL(49), ABDEYFGIL(43), ABDEFYGHJKL(33), ABCEFYGHJKL(20)\}$ 。

当经验阈值 $M = 45$ 时,选择发生次数大于等于 45 的序列建立活动关系表,并计算各活动间行为关系的个数。表 1 为根据日志中活动的行为轮廓建立的活动关系表。

表 1 活动关系表

	A	B	C	D	E	F	G	H	I	J	K	L	Y
A		4											
B			3	1									
C					3								
D					1								
E						2	1						1
F							1						2
G								1	3				
H										1			
I												3	
J											1		
K												1	
L													
Y						1	2						

2) 根据行为轮廓的定义,找出活动间的行为关系,结果如表 2 所示.

表 2 活动间的轮廓关系表

	A	B	C	D	E	F	G	H	I	J	K	L	Y
A		→											
B			→	→									
C					→								
D					→								
E						→	→						→
F							→						
G								→	→				
H										→			
I												→	
J											→		
K												→	
L													
Y							→						

3) 根据表 2 列出的行为关系,结合 Petri 网的基础结构,构造出初始模型 Model_1,如图 1 所示.

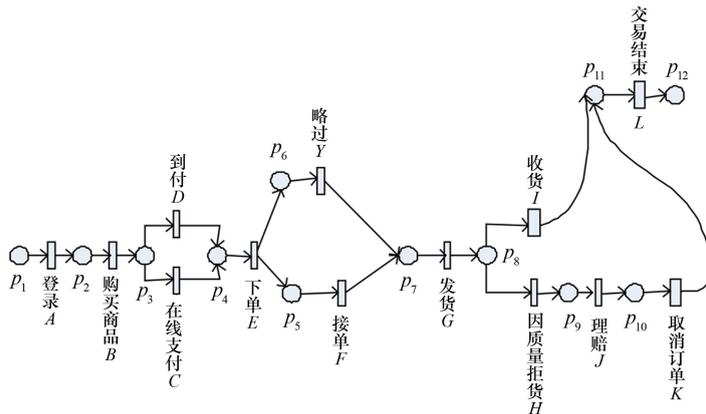


图 1 初始模型

4) 计算服从程度. 依据定义计算模型与日志之间的服从性, 得 $\epsilon_{c_{LP}} = 0.673$. 取 $\alpha = 0.95$ 作为服从性的标准值^[7], 因 $\epsilon_{c_{LP}} = 0.673 < 0.95$, 因此需要进行优化. 重新选择发生频数 M ($M \leftarrow M - N$), 其中取 $N = 10$. 优化后的模型如图 2 所示.

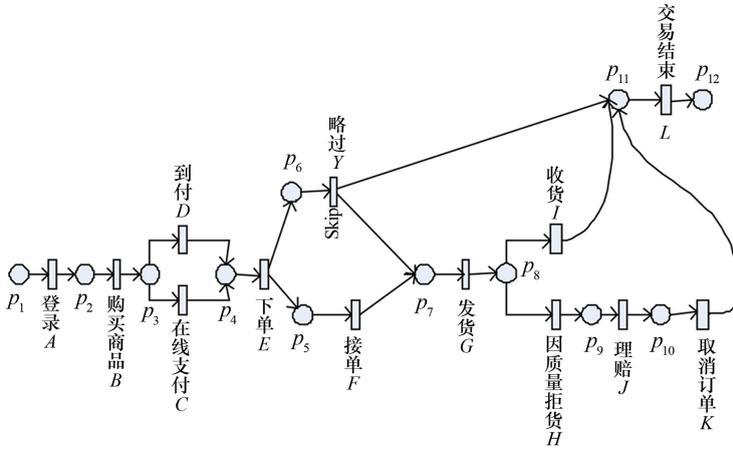


图 2 优化模型

5) 计算优化模型 Model_2 与日志的服从性, 得 $\epsilon_{c_{LP}} = 0.981$. 该值表明, 模型与事件日志具有较高的服从程度, 符合要求, 输出模型.

4 结论

本文给出的基于 Petri 网行为轮廓的网上购物流程挖掘方法, 不仅能够清晰地表达活动之间复杂的业务流程关系, 而且能够提高网上购物流程模型的运行效率, 具有一定的应用价值. 在流程挖掘过程中, 本文仅对出现在执行日志中的活动进行了研究, 但在实际执行过程中有些活动可能被隐藏或者被堵塞, 即存在隐藏变迁和阻塞变迁等问题, 因此今后将探讨隐藏变迁和阻塞变迁的问题, 以提高本文方法的有效性.

参考文献:

- [1] 宋进亮, 罗铁坚, 陈肃, 等. 一种利用聚类思想识别重复任务问题的处理方法[J]. 中国科学院研究生院学报, 2009, 26(1):107-113.
- [2] Van der AALST W M P. Process Mining: Discovery, Conformance and Enhancement of Business Processes[M]. Springer Publishing Company, Incorporated, 2011.
- [3] BUIJS J C A M, Van DONGEN B F, Van der AALST W M P. Quality dimensions in process discovery: the importance of fitness, precision, generalization and simplicity[J]. International Journal of Cooperative Information Systems, 2014, 23(1):1440001.
- [4] Van der AALST W, WEIJTERS T, MARUSTER L. Workflow mining: discovering process models from event logs[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(9):1128-1142.
- [5] ROVANI M, MAGGI F M, de LEONI M, et al. Declarative process mining in healthcare[J]. Expert Systems with Applications, 2015, 42(23):9236-9251.
- [6] 高立哲, 杜玉越, 刘伟. 医疗急诊流程挖掘与模型改进[J]. 中国科技论文, 2015, 10(2):232-236.
- [7] WEIDLICH M, POLYVYANY Y A, DESAI N, et al. Process compliance measurement based on behavioural profiles[J]. Advanced Information Systems Engineering, 2010, 6051:499-514.
- [8] GROSSMANN G, MAFAZI S, MAYER W, et al. Change propagation and conflict resolution for the co-evolution of business processes[J]. International Journal of Cooperative Information Systems, 2015, 24(1):33-40.
- [9] 吴俊枝. 基于 Petri 网行为轮廓的业务流程挖掘方法研究[D]. 淮南: 安徽理工大学, 2015.