

跨组织业务流程模型挖掘与质量评估

刘 聪¹⁾ 李会玲²⁾ 曾庆田³⁾ 段 华³⁾ 闻立杰⁴⁾

¹⁾(山东理工大学计算机科学与技术学院 山东 淄博 255000)

²⁾(山东理工大学电气与电子工程学院 山东 淄博 255000)

³⁾(山东科技大学计算机科学与工程学院 山东 青岛 266590)

⁴⁾(清华大学软件学院 北京 100084)

摘 要 跨组织业务流程需要多个组织相互配合,协同工作来完成一项由单个组织无法完成的任务.由于跨组织业务流程的复杂性与分布性,其建模与分析过程是一项耗时且容易出错的任务,要求建模人员拥有丰富的经验和行业知识.流程挖掘通过分析业务信息系统执行过程中产生的日志为模型构建提供了一种自动化方法.然而,传统的流程挖掘技术仅支持单个组织的日志挖掘,无法有效地处理跨组织业务流程挖掘问题.本文针对此问题提出一种跨组织业务流程模型挖掘方法.首先扩展已有的流程挖掘方法来进行单个组织的业务流程模型挖掘;其次,定义组织间三种典型的协同模式,并提出相应算法以挖掘组织间的协同模式;再将各个组织的流程模型和协同模式集成,得到全局跨组织业务流程模型;最后采用传统的质量评估指标和提出的协同模式拟合度来量化发现的跨组织业务流程模型质量,通过四个不同的跨组织业务流程案例与已有挖掘方法进行比较,验证本文提出挖掘方法的有效性和可用性.

关键词 Petri 网;流程挖掘;跨组织协同模式;事件日志;质量评估

中图法分类号 TP301

DOI 号 10.11897/SP.J.1016.2023.00643

Discovery and Evaluation of Cross-organization Business Process Models

LIU Cong¹⁾ LI Hui-Ling²⁾ ZENG Qing-Tian³⁾ DUAN Hua³⁾ WEN Li-Jie⁴⁾

¹⁾(School of Computer Science and Technology, Shandong University of Technology, Zibo, Shandong 55000)

²⁾(School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, Shandong 255000)

³⁾(College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590)

⁴⁾(School of Software, Tsinghua University, Beijing 100084)

Abstract In a cross-organization business process, organizations need to collaborate with each other to complete missions that cannot be handled by a single organization. Due to the complexity and geographical dispersion of cross-organization business processes, the construction is a time-consuming and error-prone task that requires practitioners to have extensive experience and business background. Process mining provides an automated method for model construction by analyzing event logs collected from the execution of business information systems. However, traditional process mining technologies only support model discovery of a single organization, and cannot effectively deal with the problem of cross-organization business process mining. To solve this problem, a cross-organization business process model discovery method is proposed. More specif-

收稿日期:2021-08-13;在线发布日期:2022-02-11. 本课题得到国家自然科学基金资助项目(61902222)、山东省泰山学者工程专项基金资助项目(ts20190936,tsqn201909109)、山东省自然科学基金优秀青年基金(ZR2021YQ45)、山东省高等学校青创科技计划创新团队项目(2021KJ031)、山东科技大学领军人才与优秀科研团队计划资助项目(2015TDJH102)资助. 刘 聪,博士,教授,中国计算机学会(CCF)会员,主要研究领域为业务流程管理、流程挖掘、人工智能等. E-mail:liucongchina@163.com. 李会玲,硕士研究生,主要研究领域为流程挖掘等. 曾庆田(通信作者),博士,教授,博士生导师,主要研究领域为流程挖掘、业务流程管理、Petri 网等. E-mail:qtzeng@163.com. 段 华,博士,教授,主要研究领域为机器学习、数据挖掘等. 闻立杰,博士,副教授,博士生导师,主要研究领域为流程挖掘、流程数据管理、大数据分析.

ically, it first extends existing process mining methods to discover business process model of a single organizations. Then, three typical collaboration patterns among organization and their corresponding discovery algorithms are introduced. Next, process models and collaboration patterns are integrated to obtain a global cross-organization business process model. Finally, the traditional quality evaluation metrics and the fitness of collaboration pattern are invented to quantify the quality of discovered cross-organizational business process models. By comparing with the state-of-the-art process discovery techniques using four public cross-organization business process cases, the effectiveness and applicability of the proposed approach is illustrated.

Keywords petri net; process mining; cross-organization collaboration pattern; event log; quality evaluation

1 引 言

随着分布式计算和互联网技术的快速发展,不同的组织共同参与同一项业务也已变为可能,在此背景下,跨组织业务流程的概念应运而生.一个典型的跨组织业务流程通常包含多个组织,并且组织间协同工作,共同完成同一项业务目标.跨组织业务流程已在多个领域得到了广泛的应用,如跨组织突发事件的应急处置流程^[1],在面临突发重大灾害事件时,不同机构将会相互合作并建立应急联动系统以达到减少突发事件带来的各种可能的伤害、提高救援效率的目的.此外,跨组织业务流程同样在其他领域发挥着重要的作用.如在医疗机构^[2]中,多个部门协作为病人提供治疗服务,由此产生了跨部门医疗业务流程;同样,在物流运输行业,多种运输方式协同运输完成物流的配送工作,由此产生了跨组织多模态运输业务流程^[3].

跨组织业务流程模型的构建是一项耗时且容易出错的任务,要求建模人员拥有丰富的经验和行业知识,而流程挖掘能够通过分析业务信息系统执行过程中的日志为业务流程构建模型,但是传统的流程挖掘技术仅针对单个组织的事件日志进行挖掘,不能有效的解决跨组织协同业务流程模型挖掘的问题,并且跨组织的联动系统具有多组织协同的特性,组织间任务的协同模式也非常复杂^[1],因此,对于跨组织协同业务流程模型的挖掘,迫切需要一种自动化方法来挖掘跨组织协同业务流程模型,进而提高处置跨组织业务流程时的决策效率和效果.

根据 2011 年 IEEE 计算智能协会数据挖掘专委会制定的《流程挖掘宣言》^[4]所述跨组织业务流程挖掘主要存在如下挑战:(1)由于分布式系统的广泛

运用,跨组织业务流程中事件日志的示例标识难以统一、流程日志中存在着执行实例的多对多等复杂匹配关系,以及融合需要信息可能缺失,导致分布式系统中的日志融合具有更大的挑战^[5-7];(2)在跨组织业务流程中,组织间的协同信息同样复杂多样,如文献^[8]中提出的四种协同模式,挖掘难度大.

Petri 网是用于描述分布式系统行为的一种形式化建模语言,它相比于其他建模语言更便于描述系统中进程或部件的顺序、并发、冲突以及同步等关系,因此本文基于 Petri 网提出一种跨组织业务流程模型挖掘方法,能够挖掘出跨组织业务流程中的业务流程模型以应对《流程挖掘宣言》^[4]中的挑战.主要研究内容包括:(1)在传统 Petri 网模型的基础上扩展资源和消息元素,简记为 RMWF-net (Resource and Message Workflow Net),来描述支持信息和资源交互行为的跨组织协同业务流程;(2)以信息管理系统中记录的事件日志为输入,提出挖掘单个组织的 RMWF-net 模型的流程挖掘算法;(3)将抽象的跨组织协同工作的协同模式具体化,并提出相应的算法以挖掘组织间的协同模式,即消息传递模式、资源共享模式、任务同步模式;(4)基于组织间的协同模式与不同组织的 RMWF-net 模型,提出集成跨组织业务流程模型的流程挖掘算法,并针对挖掘得到的组织间的协同模式模型提出了新的质量评估方法;(5)通过四个背景不同的日志,针对各个日志设计了三个对比试验,分别从拟合度、准确率、 F -值和时间性能方面比较了本文所提方法的优越性.

本文做出的贡献如下:(1)提出了跨组织业务流程模型挖掘算法,能够挖掘出定义的三种协同模式,克服已有流程挖掘方法在跨组织场景下的不足;(2)创新性地定义了组织间消息传递模式拟合度和

组织间资源共享模式拟合度,对组织间的协同模式模型进行质量评估。另外,在挖掘模型之后,基于流程模型可以帮助决策者直观理解当前跨组织业务的全局,进而,基于该模型可以进行业务规范的合规性检查,以发现业务流程实际执行过程中的偏差。

本文的其他工作安排如下:第2章讨论相关工作;第3章介绍了本文中有关的 Petri 网和 WF-net 的定义;第4章以跨组织业务流程中的一个流程案例为例,介绍了跨组织业务流程模型挖掘与质量评估方法,包括此流程案例与日志、单个组织的业务流程模型挖掘、协同模式挖掘、跨组织业务流程模型集成以及跨组织业务流程模型质量评估方法;第5章中进行了实验评估,利用多个跨组织业务流程案例的事件日志,验证本文所提挖掘方法在多个跨组织业务流程案例中的适用性,并与已有挖掘方法进行对比,验证本文所提方法的优越性,并采用重新定义的协同模式拟合度的质量评估指标对挖掘得到的组织间协同模式模型进行质量评估;第6章对本文的工作做出总结并表述对未来工作的展望。

2 相关工作

本章从跨组织业务流程建模、流程挖掘与跨组织业务流程挖掘这两个方面总结了相关工作。

2.1 跨组织业务流程建模

在跨组织业务流程建模方面,国内外学者们已经做了大量的工作,纵观跨组织业务流程建模的研究,可以发现大部分跨组织业务流程建模方法是基于 Petri 网、通信技术或是将代数理论引入业务流程建模与验证中提出的。较早时期, Van Der Aalst 在[9,10]中将传统的工作流网 WF-Net 进一步扩展为 IOWF (Interorganizational Workflow),用以建模跨组织工作流,并验证了模型结构的合理性;文献[11]中为确保组织之间通信、交互和协作实施的正确性和一致性,提出了一种基于协作点的语义对业务流程单元间存在的协作关系进行划分,应用等价划分的思想对跨组织业务流程协同进行并行建模的方法;文献[12]给出了形式化建模和行为正确性验证方法;考虑到组织之间业务流程是多维、集成和动态的特性,文献[13]将 pi 演算理论引入业务流程建模和验证中,提出了一种基于跨组织协作的集成、动态的模型,并利用 pi 演算理论验证实例模型的死锁、活锁和同步性。

在跨组织业务流程建模的过程中,还对得到的

模型进行分析,文献[14]对地铁站火灾应急处置流程进行应急处置流程形式化建模并做正确性验证和分析;文献[1]针对跨组织应急联动系统的特性,在 Petri 网的基础上扩展了任务执行组织部门、资源消耗和消息传递要素,提出 OTRM-Net 模型,用于对跨组织应急联动系统中的任务协同模式和应急处置流程进行形式化建模,并对模型中时间和资源的利用情况进行分析与验证;文献[15]给出了组织间应急资源冲突的高效检测方法和动态的冲突消解策略;文献[16]在 Petri 网精细化操作的基础上给出了一种自上而下的跨组织应急处置流程形式化建模和正确性验证方法。

2.2 流程挖掘与跨组织业务流程挖掘

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有用信息,该技术为各应用领域中的流程发现、监测和改进提供了新的手段^[17]。总体来讲,流程发现技术可以分为四类:(a) 直接从活动执行轨迹中分析活动之间的顺序、冲突、并发等依赖关系,进而构造业务流程模型,该类方法的典型代表有 Alpha Miner^[18]等;(b) 基于计算智能的挖掘算法,这类算法首先设立一个种子模型,然后根据日志中实际执行轨迹与模型行为的差别不断调整模型,使其达到一个符合要求的流程模型,该类方法主要包括了 Genetic Miner^[19];(c) 两阶段流程发现算法,第一阶段是从事件日志出发构造出描述流程行为的一个“低级”模型,如自动机模型,第二阶段将“低级”模型转换为能表达并发等复杂流程控制语义的“高级”模型,该类方法的代表是文献[20]中提出的基于区域的流程挖掘算法;(d) 声明式流程规则挖掘算法,指的是针对日志中频繁出现的模式规则而非完整的流程模型为挖掘目的的方法,如以 Declare 规则挖掘为基础的 OCBC (Object-Centric Behavioral Constraints)^[21]挖掘方法。

在跨组织业务流程挖掘方面,流程执行的事件信息由属于不同组织的单个数据库进行记录和存储,这是与单一组织的流程挖掘最根本的差别。在跨组织业务流程中并未有一个集中的日志数据库来存储工作流的所有执行信息,因此,在进行跨组织业务流程挖掘时首先要解决的是事件日志集成的问题,文献[8]中提出了中间件的概念,将跨组织协作的不同组织内部的运行日志集中起来,再从集成的运行日志中挖掘跨组织的协同流程模型,但此文献中并未具体的实现和验证跨组织业务流程的挖掘;文献[3]针对跨组织业务流程挖掘过程中组织间隐私保

护问题展开研究,提出了支持隐私保护的跨组织业务流程模型挖掘方法,与本文相比,文献[3]中仅包含消息传递这一协同模式;文献[22]将流程挖掘应用于矿山瓦斯爆炸事故应急救援跨组织业务流程领域,仅从控制流的角度验证从历史事件日志中提取瓦斯爆炸事故应急救援流程模型的可行性与优越性.本文不仅从控制流方面,而且从组织的角度对跨组织业务流程进行了挖掘,提出并实现了组织间协作模式的挖掘算法,从而能够挖掘出跨组织业务流程模型.

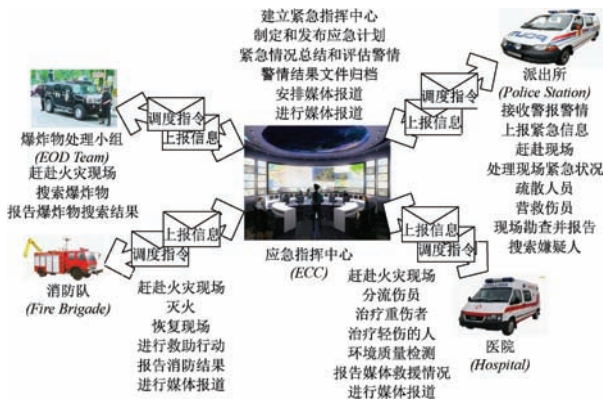


图1 跨组织火灾应急处置流程框架

3 基础知识

假设 S 是一个集合,我们用 \emptyset 表示空集, $|S|$ 表示集合 S 中的元素个数. 定义在 S 上的超集 (powerset) 记为 $P(S) = \{S' \mid S' \subseteq S\}$. 定义在 S 上的多集 (multiset) 是一个允许 S 中元素出现多次的集合. 例如, $m = [p^3, q^2]$ 是定义在集合 $S = \{p, q\}$ 上的多集, 其中 $m(p) = 3, m(q) = 2$. $B(S')$ 表示集合 S 所有多集的集合.

$f: X \rightarrow Y$ 是一个函数, 其中 $dom(f)$ 是其定义域, $cod(f) = \{f(x) \mid x \in dom(f)\}$ 是其值域. 定义在集合 S 上长度为 n 的序列 (sequence) 是一个函数 $\sigma: \{1, 2, \dots, n\} \rightarrow S$. 如果 $\sigma(1) = a_1, \sigma(2) = a_2, \sigma(3) = a_3, \dots, \sigma(n) = a_n$, 我们记为 $\sigma = \langle a_1, a_2, a_3, \dots, a_n \rangle$. $|\sigma|$ 表示序列 σ 的长度, 包括空序列 $|\langle \rangle| = 0$. S^* 表示定义在集合 S 上所有任意长度有限序列的集合.

定义 1 (Petri 网)^[23]. Petri 网是一个四元组 $\Sigma = (P, T, F, M_0)$ 需满足以下条件: $P = \{p_1, p_2, p_3, \dots, p_m\}$ 是一组有限的库所集合; $T = \{t_1, t_2, \dots, t_n\}$ 是一组有限的变迁集合; $P \cap T = \emptyset, P \cup T \cong \emptyset$; $F \subseteq (P \times T) \cup (T \times P)$ 是有限的关系集

合; $M_0: P \rightarrow \{0, 1, 2, 3, \dots\}$ 是初始标记.

定义 2 (WF-Net)^[24]. Petri 网 $\Sigma = (P, T, F, M_0)$ 是一个 WF-Net, 需满足以下条件: 有且仅有一个开始库所 $i \in P$ 使得 $\cdot i = \emptyset$; 有且仅有一个结束库所 $o \in P$ 使得 $o \cdot = \emptyset$; 任意一个节点 $x \in P \cup T$ 都在开始库所到结束库所的途径上; $\forall p \in P$, 当 $p = i, M_0(p) = 1$, 否则 $M_0(p) = 0$.

$\forall t \in T$, 如果满足 $\forall p \in \cdot t: M(p) \geq 1$, 则 t 在状态 M 能够发生, 记为 $(\Sigma, M)[t]$. 若 $(\Sigma, M)[t]$, 即 t 在状态 M 下发生, 产生新的状态 M' , 记为 $(\Sigma, M)[t](\Sigma, M')$. 如果 $p \in \cdot t \setminus t \cdot$, 则 $M'(p) = M(p) - 1$; 若 $p \in t \cdot \setminus \cdot t$, 则 $M'(p) = M(p) + 1$; 否则 $M'(p) = M(p)$. $\forall x \in P \cup T$, 记: $\cdot x = \{y \mid y \in P \cup T \wedge (y, x) \in F\}$ 是 x 的前集 (pre-set), $x \cdot = \{y \mid y \in P \cup T \wedge (x, y) \in F\}$ 是 x 的后集 (post-set); workflow 网始于一个开始库所, 最后终结于一个结束库所.

4 跨组织业务流程模型挖掘与质量评估

本章通过一个跨组织业务流程案例具体阐述在跨组织业务流程中, 组织内部业务流程模型的挖掘方法、组织间协同模式的挖掘方法、将不同组织内部流程模型与组织间协同模式集成的方法以及模型质量评估方法.

4.1 跨组织业务流程案例与日志

近年来, 城市火灾频发, 火灾的救援流程通常涉及多个组织与组织间的协作, 因此火灾应急流程是一个典型的跨组织业务流程, 通常包含以下组织: 派出所 (Police Station)、应急指挥中心 (ECC)、爆炸物处理小组 (EOD Team)、消防队 (Fire Brigade)、医院 (Hospital), 各个组织内部有各自的业务流程, 跨组织火灾应急处置流程的处置框架如图 1 所示.

在多个组织协同应对火灾应急事件时, 会产生相应的事件日志以记录整个应急处置的流程, 收集到的跨组织火灾应急处置流程事件日志称为 *FireEmergency*¹, 其中包含 126 个案例、3528 个事件、28 个应急任务、5 个组织和 1 个应急资源, 并且事件日志中的各个事件都关联相应的资源属性、消息属性和时间戳信息.

跨组织业务流程中记录的事件日志本质上是一个事件的集合. 各个事件除任务属性还包括其他属性, 如时间戳、资源、发送和接收消息等. 我们在定义 3 和 4 中形式化定义事件和属性、案例和日志.

定义 3 (事件和属性). 对于任意事件 $e \in \epsilon$, 其中 ϵ 表示事件集, N 代表事件属性集合, 对于 $n \in N$, $\#_n(e)$ 表示事件 e 属性 n 的值.

为了挖掘跨组织业务流程模型, 我们依赖于带有生命周期的事件日志. U_C 表示案例集, U_A 表示应急任务集, U_O 表示组织集, U_M 表示消息集, U_R 表示资源集, U_T 表示时间集. 对于生命周期日志, 我们假设任意一个事件 $e \in \epsilon$ 都包含如下属性: $\#_{case}(e) \in U_C$ 表示事件 e 所属的案例, 任意一个事件仅属于一个独立的案例; $\#_{act}(e) \in U_A$ 表示事件 e 的任务名; $\#_{Org}(e) \in U_O$ 表示事件 e 的组织名; $\#_{mR}(e) \in U_M$ 表示事件 e 的接收消息集合; $\#_{mS}(e) \in U_M$ 表示事件 e 的发送消息集合; $\#_{rR}(e) \in U_R$ 表示事件 e 的所需的资源集合; $\#_{time}(e) \in U_T$ 表示事件 e 的时间戳.

定义 4 (案例和日志). 定义在事件集合 ϵ 上的案例是一个有限的事件序列 $\sigma \in \epsilon^*$, 满足: 任意一个事件仅出现一次, 即 $1 \leq i < j \leq |\sigma| : \sigma(i) \neq \sigma(j)$; 同一案例中所有事件有相同的案例属性, 即 $\forall e_1, e_2 \in \sigma : \#_{case}(e_1) = \#_{case}(e_2)$. 事件日志是一个有限事件序列的集合, 即 $L \subseteq \epsilon^*$.

表 1 中给出一个跨组织火灾应急处置流程产生的事件日志片段, 其中包括 28 个事件记录. 如事件 e_2 , $\#_{case}(e_2) = 1$ 表示该事件属于案例 1, $\#_{act}(e_2) = t_2$ 表示该事件的任务名是 t_2 , $\#_{Org}(e_2) = Police Station$ 表示该事件的组织属性是 *Police Station*, $\#_{mR}(e_2) = \emptyset$ 表示该事件执行前不需要接收消息, $\#_{mS}(e_2) = \{p_{m1}\}$ 表示事件执行结束后会发送消息 p_{m1} , $\#_{rR}(e_2) = \emptyset$ 表示该事件不使用资源.

表 1 跨组织火灾应急处置流程部分运行日志示例

Org	Event	# _{case}	# _{act}	# _{mR}	# _{mS}	# _{rR}	# _{time}
Police Station	e_1	1	t_1	\emptyset	\emptyset	\emptyset	14:20 Jan 02 2018
	e_2	1	t_2	\emptyset	$\{p_{m1}\}$	\emptyset	14:21 Jan 02 2018
	e_3	1	t_3	\emptyset	\emptyset	\emptyset	14:22 Jan 02 2018
	e_4	1	t_4	\emptyset	\emptyset	\emptyset	14:27 Jan 30 2018
	e_5	1	t_5	\emptyset	\emptyset	\emptyset	14:24 Jan 02 2018
	e_6	1	t_6	\emptyset	\emptyset	\emptyset	14:28 Jan 02 2018
	e_7	1	t_7	\emptyset	$\{p_{m2}\}$	\emptyset	14:30 Jan 02 2018
	e_8	1	t_8	\emptyset	\emptyset	\emptyset	14:45 Jan 02 2018
ECC	e_9	1	t_9	$\{p_{m1}\}$	\emptyset	\emptyset	14:21 Jan 02 2018
	e_{10}	1	t_{10}	$\{p_{m2}\}$	$\{p_{m3}, p_{m4}, p_{m5}\}$	\emptyset	14:23 Jan 02 2018
	e_{11}	1	t_{11}	$\{p_{m6}, p_{m7}\}$	\emptyset	\emptyset	14:33 Jan 02 2018
	e_{12}	1	t_{12}	\emptyset	\emptyset	\emptyset	17:10 Jan 02 2018
	e_{13}	1	t_{13}	\emptyset	\emptyset	\emptyset	17:20 Jan 02 2018
EOD Team	e_{14}	1	t_{14}	\emptyset	\emptyset	\emptyset	17:25 Jan 02 2018
	e_{15}	1	t_{15}	$\{p_{m3}\}$	\emptyset	$\{p_r\}$	14:25 Jan 02 2018
	e_{16}	1	t_{16}	\emptyset	\emptyset	\emptyset	14:30 Jan 02 2018
	e_{17}	1	t_{17}	\emptyset	$\{p_{m6}\}$	\emptyset	14:40 Jan 02 2018
Fire Brigade	e_{18}	1	t_{18}	$\{p_{m4}\}$	\emptyset	$\{p_r\}$	14:25 Jan 02 2018
	e_{19}	1	t_{19}	\emptyset	\emptyset	\emptyset	14:30 Jan 02 2018
	e_{20}	1	t_{20}	\emptyset	\emptyset	\emptyset	16:40 Jan 02 2018
	e_{21}	1	t_{21}	\emptyset	\emptyset	\emptyset	16:45 Jan 02 2018
	e_{22}	1	t_{22}	\emptyset	$\{p_{m7}\}$	\emptyset	17:00 Jan 02 2018
	e_{23}	1	t_{23}	$\{p_{m5}\}$	\emptyset	$\{p_r\}$	14:25 Jan 02 2018
	e_{24}	1	t_{24}	\emptyset	\emptyset	\emptyset	14:30 Jan 02 2018
Hospital	e_{25}	1	t_{25}	\emptyset	\emptyset	\emptyset	14:45 Jan 02 2018
	e_{26}	1	t_{26}	\emptyset	\emptyset	\emptyset	14:50 Jan 02 2018
	e_{27}	1	t_{27}	\emptyset	\emptyset	\emptyset	16:40 Jan 02 2018
	e_{28}	1	t_{28}	\emptyset	\emptyset	\emptyset	17:00 Jan 02 2018
	e_{14}	1	t_{14}	\emptyset	\emptyset	\emptyset	17:25 Jan 02 2018

4.2 单个组织的业务流程模型挖掘

挖掘单个组织的业务流程模型需要获取单个组织信息管理系统中记录的运行事件日志, 并以此日

志为输入数据, 挖掘得到组织内部的 RMWF-net 模型. 由于单个组织挖掘出的流程模型是一个在 Petri 网的基础上扩充消息和资源信息的 RMWF-net 模

型,因此首先在第3章中了解 Petri 网和 WF-Net 定义的基础上,引入 RMWF-net 的概念.

定义 5 (RMWF-net). $\Sigma=(P, T, F, M_0)$ 是扩展了资源与消息相关信息的 Petri 网,需满足条件: $P = P_L \cup P_R \cup P_M, P_L \cap P_R = \emptyset, P_L \cap P_M = \emptyset, P_M \cap P_R = \emptyset, P_M \subseteq P$ 表示消息库所; $P_R \subseteq P$ 表示资源库所; $F = F_L \cup F_R \cup F_M; (P_L, T, F_L)$ 是一个 WF-Net; $\forall p \in P$, 当 $p = P_R \cup \{i\}, i \in P_L \wedge i = \emptyset, M_0(p) = 1$, 否则 $M_0(p) = 0$.

RMWF-net 的引发规则与 WF-Net 的引发规则相同,其中资源库所 (p_r) 在活动不占用时, $M(p_r) = 1$. 在定义 RMWF-net 的基础上,提出算法 1 以挖掘各组织内部的 RMWF-net 模型.

算法 1. 单个组织的 RMWF-net 模型挖掘.

输入: 单个组织的业务流程事件日志 L

输出: RMWF-net 模型 $\Sigma=(P, T, F, M_0)$

```

1.  $P \leftarrow \emptyset, P_L \leftarrow \emptyset, P_R \leftarrow \emptyset, P_M \leftarrow \emptyset, F \leftarrow \emptyset, F_L \leftarrow \emptyset, F_R \leftarrow \emptyset, F_M \leftarrow \emptyset, T \leftarrow \emptyset, M_0 \leftarrow \emptyset;$ 
   //初始化
2.  $(P_L, T, F_L) \leftarrow \text{Inductive Miner}(L);$ 
3. FOR  $t \in T$  DO
4.   FOR  $\sigma \in L$  DO
5.     FOR  $e \in \sigma$  DO
6.       IF  $t = \#_{\text{act}}(e)$  THEN
7.         IF  $\#_{mS}(e) \neq \emptyset$  THEN //消息发送库所挖掘
8.            $P_M \leftarrow P_M \cup \{\#_{mS}(e)\}; F_M \leftarrow F_M \cup \{(t, \#_{mS}(e))\};$ 
9.         END IF
10.        IF  $\#_{mR}(e) \neq \emptyset$  THEN //消息接收库所挖掘
11.           $P_M \leftarrow P_M \cup \{\#_{mR}(e)\}; F_M \leftarrow F_M \cup \{(\#_{mR}(e), t)\};$ 
12.        END IF
13.        IF  $\#_{rR}(e) \neq \emptyset$  THEN //资源库所挖掘
14.           $P_R \leftarrow P_R \cup \{\#_{rR}(e)\};$ 
15.           $F_R \leftarrow F_R \cup \{(\#_{rR}(e), t), (t, \#_{rR}(e))\};$ 
16.        END IF
17.      END IF
18.    END DO
19.  END DO
20. END DO
21.  $P \leftarrow P_R \cup P_L \cup P_M, F \leftarrow F_R \cup F_L \cup F_M, M_0 \leftarrow \{i\} \cup P_R;$ 
22. RETURN  $\Sigma=(P, T, F, M_0)$ . //输出 RMWF-
```

net 模型

本文中的原始日志是一个集中的日志,因此在应用算法 1 之前需要对跨组织业务流程事件日志进行预处理,即首先将事件日志按照组织属性进行划分,得到多个组织属性不同的子事件日志,其次以单个组织的事件日志 L 为算法 1 的输入,应用 Inductive Miner^[25] 挖掘业务流程的控制流模型;并挖掘各个任务对应的发送和接收的消息、使用的资源、以及它们对应的流关系,合并所有的库所和流关系,得到组织内部的 RMWF-net 模型.

以 *FireEmergency* 作为算法 1 的输入,使用 Inductive Miner 挖掘出业务流程的控制流模型,再分别解析出单个组织中存在的发送、接收的消息库所和使用的资源库所,最后输出组织内部的 RMWF-net 模型如表 2 中所示. 其中案例派出所中 t_2 发送消息 p_{m1}, t_7 发送消息 p_{m2} , 应急指挥中心和 t_{10} 发送消息 $p_{m3}, p_{m4}, p_{m5}, t_9$ 接收消息 p_{m1}, t_{10} 接收消息 p_{m2}, t_{11} 接收消息 p_{m6}, p_{m7} ; 爆炸物处理小组中 t_{15} 接收消息 p_{m3} , 占用资源 p_r, t_{17} 发送消息 p_{m6} ; 消防队中 t_{18} 接收消息 p_{m4} , 占用资源 p_r, t_{22} 消息发送 p_{m7} ; 医院中 t_{23} 接收消息 p_{m5} , 占用资源 p_r .

4.3 协同模式挖掘

在跨组织业务流程中,组织间存在的协同模式将各个组织的业务流程联系起来以协同完成同一项业务目标,因此协同模式的挖掘是挖掘跨组织业务流程模型的关键. 以挖掘到的不同组织的 RMWF-net 流程模型为输入,根据组织间协同模式的定义,挖掘出任意两个组织之间存在的协同模式. 首先根据组织间的协同模式^[8],定义跨组织业务流程中存在的协同模式,包括消息传递模式、资源共享模式和任务同步模式.

定义 6 (消息传递模式). $\Sigma_1=(P_1, T_1, F_1, M_{01})$ 和 $\Sigma_2=(P_2, T_2, F_2, M_{02})$ 为两个不同组织对应的 RMWF-net, 其中 $P_i = P_{Li} \cup P_{Ri} \cup P_{Mi} (i = 1, 2)$, 两个不同组织之间存在消息传递模式当且仅当: $P_{L1} \cap P_{L2} = \emptyset; P_{M1} \cap P_{M2} = \emptyset; P_{R1} \cap P_{R2} = \emptyset; T_1 \cap T_2 = \emptyset$.

即存在消息传递模式的两个组织在跨组织业务流程执行过程中,通过接收或释放同一个消息进行消息传递以完成信息交互.

定义 7 (资源共享模式). $\Sigma_1=(P_1, T_1, F_1, M_{01})$ 和 $\Sigma_2=(P_2, T_2, F_2, M_{02})$ 为两个不同组织对应的 RMWF-net, 其中 $P_i = P_{Li} \cup P_{Ri} \cup P_{Mi} (i = 1, 2)$, 两个不同组织之间存在资源共享模式当

且仅当 $P_{L_1} \cap P_{L_2} = \emptyset; P_{M_1} \cap P_{M_2} = \emptyset; P_{R_1} \cap P_{R_2} \neq \emptyset; T_1 \cap T_2 = \emptyset$.

即存在资源共享模式的两个组织在跨组织业务流程执行过程中,都占有或释放同一个资源.

定义 8 (任务同步模式). $\Sigma_1 = (P_1, T_1, F_1, M_{0_1})$ 和 $\Sigma_2 = (P_2, T_2, F_2, M_{0_2})$ 为两个不同组织对应的 RMWF-net,两个不同组织之间存在任务同步模式当且仅当: $P_1 \cap P_2 = \emptyset; T_1 \cap T_2 \neq \emptyset$.

表 2 挖掘跨组织火灾应急处置流程得到的各个组织的 RMWF-net 模型

组织	挖掘得到的组织内部 RMWF-net
派出所	
应急指挥中心	
爆炸物处理小组	
消防队	
医院	

即存在任务同步模式的两个组织在跨组织业务流程执行过程中都执行同一个任务.

根据定义 6~8 中组织间协同模式的定义,提出了算法 2.以挖掘到组织内部的 RMWF-net 流程模型为输入,通过算法 2 挖掘出组织间的协同模式,其中主要分为三个模块:(1) 消息传递模式挖掘模块:首先挖掘出两个组织间传递的消息,即两个组织中 P_M 的交集,再挖掘出此消息库所在不同组织中的发送方、接收方以及对应的流关系;(2) 资源共享模式挖掘模块:首先挖掘出两个组织间共享的资源库所,即两个组织中 P_R 的交集,再挖掘出两个组织共用此资源库所的任务以及对应的流关系;(3) 任务同步模式挖掘模块:直接挖掘出两个组织中存在的相同的任务,即两个组织中 T 的交集.

以跨组织火灾应急处置流程案例为例,将表 2 中挖掘得到各组织的 RMWF-net 作为算法 2 的输入,挖掘得到组织间协同模式模型如表 3 所示,其中派出所与应急指挥中心只存在共同的消息库所 p_{m1} 和 p_{m2} ,所以这两个组织之间只存在消息传递模式;

消防队与应急指挥中心存在两个共同的消息库所 p_{m4} 和 p_{m7} ,同时还存在相同的活动 t_{14} ,所以这两个组织之间存在消息传递模式和任务同步模式;医院与爆炸物处理小组之间只存在一个共享的资源 p_r ,所以这两个组织之间只存在资源共享模式.

4.4 跨组织业务流程模型集成

基于 Petri 网的同步合成和共享合成理论将挖掘到的各个组织的业务流程模型和组织间的协同模式模型集成,得到最终的跨组织业务流程 RMWF-net 模型,具体如算法 3 所示,将算法 1 挖掘的各个组织的业务流程模型和算法 2 挖掘到的组织间协同模式基于 Petri 网的合成理论集成,最终得到全局的跨组织业务流程模型.

以跨组织火灾应急处置流程案例为例,将挖掘到的表 2 中不同组织内部的 RMWF-net 模型与表 3 中组织间的协同模式作为算法 3 的输入,经过算法 3 的集成,最终输出如图 2 所示的火灾应急处置流程案例的跨组织业务流程模型.

4.5 跨组织业务流程模型质量评估方法

本小节中介绍跨组织业务流程模型质量评估方

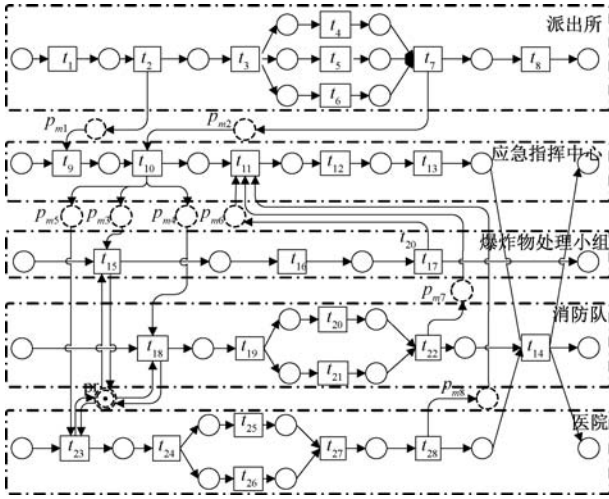


图2 火灾应急处置流程案例的跨组织业务流程模型

法,首先介绍了传统的质量评估指标,其次,针对挖掘到的组织间协同模式模型,定义了组织间消息传递模式拟合度和组织间资源共享模式拟合度,对组织间的协同模式模型进行质量评估。

算法2. 组织间协同模式挖掘。

输入: 两个组织的 RMWF-net $\Sigma_1 = (P_1, T_1, F_1, M_{01})$ 和 $\Sigma_2 = (P_2, T_2, F_2, M_{02})$

输出: 协同模式 $\Sigma = (P, T, F, M_0)$

1. $P \leftarrow \emptyset, P_L \leftarrow \emptyset, P_R \leftarrow \emptyset, P_M \leftarrow \emptyset, F \leftarrow \emptyset, F_L \leftarrow \emptyset, F_R \leftarrow \emptyset, F_M \leftarrow \emptyset, T \leftarrow \emptyset, M_0 \leftarrow \emptyset;$ // 初始化
2. IF $P_{M1} \cap P_{M2} \neq \emptyset$ THEN // 消息传递模式挖掘
3. FOR $p \in P_{M1} \cap P_{M2}$ DO
4. $P_M \leftarrow P_M \cup \{p\};$
5. IF $p' \neq \emptyset$ THEN
6. $T \leftarrow T \cup p';$
7. FOR $t \in p'$ DO
8. $F_M \leftarrow F_M \cup \{(p, t)\};$
9. END DO
10. END IF
11. IF $\cdot p \neq \emptyset$ THEN
12. $T \leftarrow T \cup \cdot p;$
13. FOR $t \in \cdot p$ DO
14. $F_M \leftarrow F_M \cup \{(t, p)\};$
15. END DO
16. END IF
17. END DO
18. END IF
19. IF $P_{R1} \cap P_{R2} \neq \emptyset$ THEN // 资源共享模式挖掘
20. FOR $p \in P_{R1} \cap P_{R2}$ DO
21. $P_R \leftarrow P_R \cup \{p\}; T \leftarrow T \cup p' \cup p;$

22. FOR $t \in p' \cup \cdot p$ DO
23. $F_R \leftarrow F_R \cup \{(p, t), (t, p)\};$
24. END DO
25. END DO
26. END IF
27. IF $T_1 \cap T_2 \neq \emptyset$ THEN // 任务同步模式挖掘
28. $T \leftarrow T \cup (T_1 \cap T_2);$
29. END IF
30. $P \leftarrow P_R \cup P_L \cup P_M, F \leftarrow F_R \cup F_L \cup F_M, M_0 \leftarrow P_R;$
31. RETURN $\Sigma = (P, T, F, M_0).$ // 输出 Σ_1 与 Σ_2 的协同模式

4.5.1 传统质量评估方法

拟合度 (*fitness*)^[26]、准确率 (*precision*)^[27] 与 F 值 (*F-measure*)^[28] 是评估流程模型质量的主流评估指标,同时由于跨组织业务流程模型挖掘方法挖掘到的模型本质也是一个流程模型,因此本文采用这三个指标来评估挖掘到的跨组织业务流程模型的质量,其中,拟合度度量了日志案例可以被流程模型从头到尾重演的比例,低的拟合度意味着日志中的很多案例在发现的流程模型中并不允许执行;准确率量化了在发现的业务流程模型中出现的案例不在日志中出现的比例,低的准确率意味着发现的业务流程模型中允许更多日志中并未出现的行为。

算法3. 基于协同模式的 RMWF-net 模型挖掘。

输入: $\Xi = \{\Sigma_x \mid \Sigma_x \text{ 是组织 } x \text{ 的 RMWF-net}\}, \Psi = \{\Sigma_{x,y} \mid \Sigma_{x,y} \text{ 是组织 } x \text{ 和 } y \text{ 的协同模型}\}$

输出: 集成的 RMWF-net $\Sigma = (P, T, F, M_0)$

1. $P \leftarrow \emptyset, P_L \leftarrow \emptyset, P_R \leftarrow \emptyset, P_s \leftarrow \emptyset, P_M \leftarrow \emptyset; F \leftarrow \emptyset, F_L \leftarrow \emptyset, F_R \leftarrow \emptyset, F_M \leftarrow \emptyset; T \leftarrow \emptyset; M_0 \leftarrow \emptyset;$ // 初始化集成模型
2. FOR $\Sigma_x \in \Xi$ DO // 模型集成
3. $P_M \leftarrow P_M \cup P_{Mx}; P_R \leftarrow P_R \cup P_{Rx}; P_L \leftarrow P_L \cup P_{Lx};$
4. FOR $p \in P_{Lx}$ DO
5. IF $\cdot p = \emptyset$ THEN
6. $P_s \leftarrow P_s \cup p;$
7. END IF
8. END DO
9. $F_L \leftarrow F_L \cup F_{Lx}; F_M \leftarrow F_M \cup F_{Mx};$
10. $F_R \leftarrow F_R \cup F_{Rx}; T \leftarrow T \cup T_x;$
11. END DO
12. FOR $\Sigma_{x,y} \in \Psi$ DO // 协同模式集成
13. $P_M \leftarrow P_M \cup P_{Mx,y}; P_R \leftarrow P_R \cup P_{Rx,y};$
14. $P_L \leftarrow P_L \cup P_{Lx,y}; F_L \leftarrow F_L \cup F_{Lx,y};$

- 15. $F_M \leftarrow F_M \cup F_{M_{x,y}}; F_R \leftarrow F_R \cup F_{R_{x,y}};$
- 16. $T \leftarrow T \cup T_{x,y};$
- 17. END DO
- 18. $P \leftarrow P_R \cup P_L \cup P_M, F \leftarrow F_R \cup F_L \cup F_M, M_0$
 $\leftarrow P_s \cup P_R;$
- 19. RETURN $\Sigma = (P, T, F, M_0).$

表 3 挖掘跨组织火灾应急处置流程得到的组织间协同模式模型

组织	挖掘的组织间协同模式	组织	挖掘的组织间协同模式
派出所与应急指挥中心		爆炸物处理小组与应急指挥中心	
消防队与应急指挥中心		医院与应急指挥中心	
消防队与爆炸物处理小组		医院与爆炸物处理小组	
医院与消防队			

拟合度与准确率之间往往存在一个平衡,有时候通过一些设置可以使得拟合度小幅下降的同时准确率可能会大幅提升,为了统一这两个指标,我们引入 F -值,其定义如下:

$$F\text{-值} = \frac{2 * fitness * precision}{fitness + precision} \quad (1)$$

4.5.2 组织间消息传递模式拟合度

消息传递模式中有发送消息与接收消息两种行为,而这两种行为本质上是一致的,仅是发送消息与接收消息的双方不同,因此这两种行为的拟合度计算方式相同。

首先计算组织间消息传递模式模型中事件 t 接收消息 p_m 这一行为的拟合度,即事件日志中事件 t 接收消息 p_m 这一行为可以被消息传递模式模型重演的比例:

$$fitness_{MR} = \frac{F_{model}(p_m, t)}{F_{log}(p_m, t)} \quad (2)$$

其中 $F_{model}(p_m, t)$ 表示在挖掘出的业务流程模型中,变迁 t 接收消息库所 p_m 的行为在模型中重演日志时出现的次数; $F_{log}(p_m, t)$ 表示在日志中事件

t 接收消息 p_m 的行为出现的次数;

其次,对于组织间消息传递模式中事件 t 发送消息 p_m 这一行为的拟合度如公式(3)所示:

$$fitness_{MS} = \frac{F_{model}(t, p_m)}{F_{log}(t, p_m)} \quad (3)$$

公式(3)中 $F_{model}(t, p_m)$ 表示在挖掘出的业务流程模型中,变迁 t 发送消息库所 p_m 的行为在模型中重演日志时出现的次数; $F_{log}(t, p_m)$ 表示在日志中事件 t 发送消息 p_m 的行为出现的次数。

4.5.3 组织间资源共享模式拟合度

在跨组织业务流程中,任何事件必须在资源未被任何其他事件占用时才可以使用该资源,且当事件使用结束并释放该资源后,其他事件才可以占用,由此可以发现,资源被事件占用与被事件释放的行为是对应的,因此,在评估组织间资源共享协同模式模型时可将事件 t 占用资源 p_r 与事件 t 释放资源 p_r ,这两种行为看作是一对行为,则组织间资源共享模式拟合度如公式(4)所示:

$$fitness_R = \frac{F_{model}(p_r, t) + F_{model}(t, p_r)}{F_{log}(p_r, t) + F_{log}(t, p_r)} \quad (4)$$

其中 $F_{model}(p_r, t)$ 表示在挖掘出的业务流程模型中, 变迁 t 占用资源 p_r 的行为在模型中重演日志时出现的次数, $F_{model}(t, p_r)$ 表示变迁 t 释放资源 p_r 的行为在模型中重演日志时出现的次数; $F_{log}(p_r, t)$ 表示在日志中事件 t 占用资源 p_r 的行为出现的次数, $F_{log}(t, p_r)$ 表示日志中事件 t 释放资源 p_r 的行为出现的次数。

5 实验评估

本章对跨组织业务流程模型挖掘方法的有效性、可用性与时间性能进行了实验评估。具体而言, 首先介绍了流程挖掘方法的工具实现, 其次, 利用四个不同领域的实际流程案例对应的跨组织业务流程事件日志, 将本文所提出的方法与现有的流程挖掘方法进行比较, 评估挖掘出模型的质量与挖掘的时间性能, 并采用本文所提出的消息传递模式拟合度和资源共享模式拟合度评估指标对挖掘的不同组织间协同模式模型进行质量评估。

5.1 工具实现

在开源流程挖掘工具平台 *ProM 6*^① 中为流程挖掘提供了一个完全可插拔的实验环境。它可以通过添加插件进行扩展, 目前包含 1600 多个插件, 该工具和所有插件都是开源的。

该跨组织业务流程模型挖掘方法^② (称为 *Cross-organization Business Process Mining*) 在开源流程挖掘工具平台 *ProM 6* 中实现。该工具的快照如图 3 所示, 它以一个跨组织业务流程事件日志作为输入, 并返回一个基于 Petri 网的跨组织业务流程模型作为输出。



图 3 跨组织业务流程模型挖掘方法工具界面

5.2 实验数据介绍

为了评估本文提出的跨组织业务流程模型挖掘方法的有效性, 本小节另外介绍了其他三个不同领域的跨组织业务流程案例中的跨组织业务流程事件

日志^③, 分别是南京市突发环境污染事件应急处置流程的事件日志 (*PollutionPro*)、跨组织多模态运输业务流程^[3]的事件日志 (*Transportation*)、商品购买流程^[29]的事件日志 (*PurchaseCase*), 这三个案例的事件日志与前述跨组织火灾应急处置流程的事件日志 (*FireEmergency*) 的详细对比信息如表 4 所示。

表 4 实验数据对比

案例 日志名	案例 数量	事件 数量	活动 数量	组织	资源	消息
<i>FireEmergency</i>	126	3528	28	5	1	8
<i>PollutionPro</i>	1952	46848	24	5	0	9
<i>Transportation</i>	19577	411117	21	4	0	12
<i>PurchaseCase</i>	33614	286300	14	2	0	6

5.3 实验设置

本小节设置了三个对比实验, 即传统流程挖掘方法 *Inductive Miner*^[25]、前期工作^[30]中提出的将各个组织内部的流程模型以并行方式合并到一个业务流程模型中的跨组织业务流程模型挖掘方法、本文所提方法, 分别对表 4 中的四个数据进行实验, 以评估挖掘模型的质量。

方法 1. 使用传统的 *Inductive Miner*^[25] 直接挖掘跨组织业务流程案例的事件日志, 得到跨组织业务流程模型;

方法 2. 首先基于跨组织业务流程案例的事件日志的组织属性, 通过传统的 *Inductive Miner*^[25] 分别挖掘不同组织的业务流程模型, 再将不同组织的业务流程模型以并行的方式合并到一个业务流程模型中, 最终得到此方法对应的跨组织业务流程模型^[30];

方法 3. 本文所提跨组织业务流程模型挖掘方法, 首先挖掘各组织内部的业务流程模型, 其次挖掘组织间存在的协同模式, 包括消息传递模式、资源共享模式和任务同步模式, 最后基于 Petri 网的同步合成和共享合成理论将挖掘到的各个组织的业务流程模型和组织间的协同模式模型集成, 得到最终的跨组织业务流程模型。

5.4 实验结果

本小节对表 4 中的数据经过 5.3 设置的三个实验方法挖掘出模型的结果进行分析。

方法 1. 使用 *Inductive Miner* 直接挖掘四个案例事件日志的跨组织业务流程模型, 其中, 以商品购

① <https://promtools.org>

② <https://svn.win.tue.nl/repos/prom/Packages/Shandong-PM/>

③ <https://pan.baidu.com/s/1ThnTSzKsiwiyMzOvdOgXdg>, 提取密码: qv0h

买流程^[29]的事件日志为例,挖掘到的跨组织业务流程模型如图 4 所示,但是此方法仅仅是从控制流的角度对各个案例的事件日志进行挖掘,挖掘出的模型与实际的流程有很大的出入,四个案例中分别采用此方法得到模型的拟合度、准确率、 F -值的具体值如表 5 所示。

方法 2. 首先根据日志的组织属性处理日志,分成各个组织的子日志,再对各个组织的子日志使用 Inductive Miner 挖掘各个组织的子业务流程模型,最后将各个组织的子业务流程模型以并行的方式合并,其中,以商品购买流程^[29]的事件日志为例,挖掘到的跨组织业务流程模型如图 5 所示,图中的两个黑色变迁分别表示合并所有不同组织的子业务流程模型时的合并点;且四个案例中分别采用此方法得到模型的拟合度、准确率、 F -值的具体值如表 5 所示。

在表 5 中可以明显看出方法 2 的拟合度,准确率, F -值都比方法 1 要高,与方法 1 相比,方法 2 增加了对日志中组织属性的约束,区分了在跨组织业务流程中不同组织的业务流程。

方法 3. 也就是本文所述方法,首先将日志依据组织属性分成不同组织的子日志,应用算法 1 挖掘各个组织的子日志对应的子业务流程模型 RMWF-net,然后再应用算法 2 挖掘跨组织业务流程中不同组织间的协同模式,最后应用算法 3,将挖掘的不同组织的业务流程模型 RMWF-net 与组织间的协同模式集成,挖掘出全局跨组织业务流程模型,在此方法中,依据组织属性分别挖掘各组织业务流程模型的同时,增加了组织间存在的消息传递模式、资源共享模式和任务同步模式的约束. 以商品购买流程^[29]事件日志为例,使用工具挖掘到的跨组织业务流程模型如图 6 所示,且四个案例中分别采用此方法得到模型的拟合度、准确率、 F -值的值在三种挖掘方法中是最高的,如表 5 所示。

与方法 2 相比,本文所述方法不仅仅只是方法 2 中简单的对不同组织内部的业务流程的控制流进行挖掘,此外还对不同组织在跨组织业务流程中组织间的协同模式进行挖掘,增加了对组织间协同模式的约束,提高了挖掘得到的跨组织业务流程模型的准确率。

表 5 四个案例的质量评估结果对比

案例 方法	FireEmergency				PollutionPro			
	拟合度	准确率	F -值	时间 /ms	拟合度	准确率	F -值	时间 /ms
方法 1	1	0.20	0.33	179.0	1	0.21	0.35	847.0
方法 2	1	0.35	0.52	2.0	1	0.42	0.59	1.0
方法 3	1	0.62	0.77	8.0	1	0.67	0.80	75.0
案例 方法	Transportation				PurchaseCase			
	拟合度	准确率	F -值	时间 /ms	拟合度	准确率	F -值	时间 /ms
方法 1	1	0.40	0.57	8544.0	1	0.85	0.92	1909.0
方法 2	1	0.46	0.63	144.0	1	0.57	0.72	479.0
方法 3	1	0.96	0.98	595.0	1	1	1	742.0

根据表 5 所示不同方法下的四个案例的拟合度、准确率、 F -值,可以得出如下结论:

(1) 三种挖掘方法挖掘出的跨组织业务流程模型的拟合度都是 1,这是由于三种挖掘方法都使用了传统的挖掘方法 Inductive Miner,而此方法挖掘出的模型可以完全重演日志中的所有行为,因此拟合度是 1;

(2) 三种挖掘方法中,方法 3,也就是本文所提方法的准确率在三种挖掘方法中是最高的;一般而言,挖掘出的模型准确率高则模型更准确,在本案例中,挖掘方法 1 仅仅是从控制流的角度挖掘模型,而挖掘方法 2 与方法 3 都增加了对跨组织业务流程中组织属性的约束,因此方法 2 与方法 3 的准确率都比方法 1 要高;而相比于方法 2 中只是简单的将

不同组织的控制流模型以并行方式合并的思想,方法 3 中不仅仅对整个跨组织业务流程的组织属性加以区分,还对组织间存在的协同模式进行挖掘,使得不同组织之间存在依赖关系,因此方法 3 的准确率要比方法 2 和方法 1 高;在所示的四个日志中,*PurchaseCase* 日志中方法 2 的准确率最低,然而对比表 4 中四个实验数据可以发现,此日志中只包含两个组织,并且包含的消息个数最少,其次观察其挖掘出的模型,如图 5 所示,可知此实验数据包含的行为较为简单,因此使用方法 1 便可达到较好的实验结果,尽管如此,方法 3 的拟合度、准确率、 F -值仍然是最高的,都达到了 1;

(3) 挖掘方法 3 的 F -值在三种挖掘方法中也是最高的, F -值是平衡了拟合度与准确率的值,因

此表明挖掘方法 3 挖掘的跨组织业务流程模型要比 方法 1 与方法 2 挖掘出的模型更好;

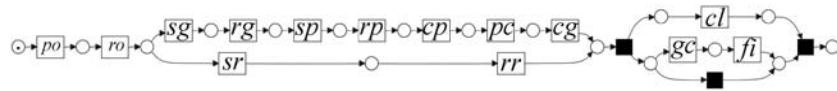


图 4 方法 1 挖掘的 PurchaseCase 的跨组织业务流程模型 (po: 下订单; rg: 收货; cg: 确认收货; cl: 关闭; rr: 拒绝收货; sp: 付款; pc: 确认付款; ro: 接收; sg: 发送货物; rp: 收款; cp: 确认收款; gc: 确认货物; fi: 完成; sr: 拒绝发货).

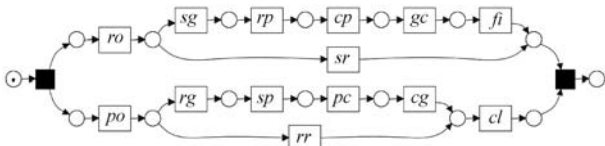


图 5 方法 2 挖掘的 PurchaseCase 的跨组织业务流程模型

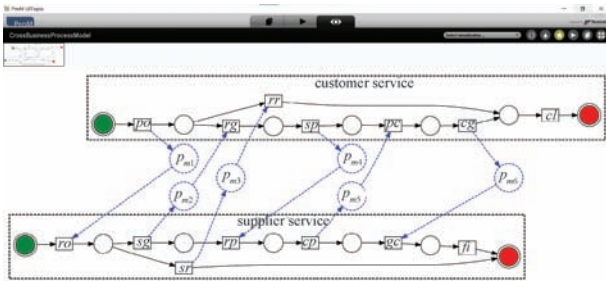


图 6 使用工具挖掘的 PurchaseCase 的跨组织业务流程模型

(4) 从时间方面来看,方法 1 的模型发现时间是最长的,方法 3 的发现时间比方法 1 缩短了很多,但是由于方法 3 是在方法 2 的基础上增加了协同模式的挖掘,因此方法 3 的模型发现时间必定是大于方法 2 的模型发现时间,以上结果表明方法 3 的模型发现时间优于方法 1.

总而言之,不论是在模型质量还是在时间性能方面,方法 3 都要优于方法 1;而方法 3 虽然在时间性能上要比方法 2 差,但挖掘出模型的质量要比方法 2 挖掘出模型的质量好得多,验证了提出跨组织业务流程模型挖掘方法的有效性和可用性.

5.5 组织间协同模式评估结果

组织间的协同模式是跨组织业务流程中重要的组成部分,在本小节中,采用重新定义的拟合度评估指标,对挖掘到的跨组织业务流程中不同组织间的协同模式进行质量评估.

利用公式(2)–(4),分别统计跨组织业务流程事件日志中存在的消息传递与资源共享对应关系的次数,并且统计在日志重演过程中挖掘出的跨组织业务流程模型中出现消息传递与资源共享对应关系的次数,可以得到 $fitness_{MR} = 1$, $fitness_{MS} = 1$, $fitness_R = 1$,这是由于采用的跨组织业务流程中事件

日志是理想的日志,不包含噪音,因此,通过本文所提的跨组织业务流程模型挖掘方法挖掘出的跨组织业务流程模型能够完全拟合日志中出现的行为,即日志中接收、发送消息,占用与释放资源的行为在模型中都可以显示,并且在重演日志时出现的频率也相同,故协同模式的拟合度为 1.

在真实的业务系统运行过程中,很难保证收集的日志是完全干净且不含噪音的,因此在含有噪音的日志中,为了提高挖掘方法的灵活性及结果的切实性,可以引入噪声处理机制,设置噪声阈值是 0.9,过滤掉原噪声日志的 10%,给出如下例子进行说明:日志中存在一个消息 p_m ,在挖掘模型时发现有两个事件 e_1, e_2 均接收此消息,但事件 e_1 接收消息 p_m 的行为出现 100 次,而事件 e_2 接收消息 p_m 的行为仅出现了 1 次,则事件 e_1 接收消息 p_m 的行为的频率为 $P_f = 100/101 \approx 0.99$,事件 e_2 接收消息 p_m 的行为的频率为 $P_f = 1/101 \approx 0.01$,在设置噪声阈值为 0.9 的情况下,将过滤事件 e_2 接收消息 p_m 这一不频繁的行为.

6 总结与展望

本文提出了一种跨组织业务流程模型挖掘方法,能够挖掘跨组织业务流程中的模型,并且通过多个不同案例验证了本文所提方法在不同案例场景下的适用性,并且通过三个模型质量评估指标和时间性能的对比,验证本文所提方法挖掘到的跨组织业务流程模型与已有方法挖掘到的模型相比具有明显的优越性.

基于本文的工作,后续还有问题需要继续研究:(1) 对组织间的协同模式进行扩展,如对组织间任务外包模式进行挖掘与分析^[8,31,32]; (2) 对于不同组织的业务流程模型与协同模式合并之后得出的跨组织业务流程模型的正确性,后续需要严谨的理论进行证明;(3) 考虑大规模跨组织业务流程事件日志,通过高效的采样方法^[33,34]来提高挖掘效率和准确率.

致 谢 感谢审稿专家及责任编辑对本文的改进提出宝贵的意见!

参 考 文 献

- [1] Zeng Qing-Tian, Lu Fa-Ming, Liu-Cong, et al. Modeling and analysis for cross-organizational emergency response systems using petri nets. *Chinese Journal of Computers*, 2013, 36(11): 2290-2302
(曾庆田, 鲁法明, 刘聪等. 基于 Petri 网的跨组织应急联动处置系统建模与分析. *计算机学报*, 2013, 36(11): 2290-2302.)
- [2] Zeng Qing-Tian, Lu Fa-Ming, Liu Cong, et al. Modeling and verification for cross-department collaborative business processes using extended petri nets. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2014, 45(02): 349-362
- [3] Liu Cong, Duan Hua, Zeng Qing-Tian, et al. Towards comprehensive support for privacy preserving cross-organization business process mining. *IEEE Transactions on Service Computing*, 2019, 12(4): 639-653
- [4] Van Der Aalst W, Adriansyah A, De Medeiros A K A, et al. *Process mining manifesto//International Conference on Business Process Management*. Berlin, Germany: Springer, 2011: 169-194
- [5] Xu Yang, Lin Qi, Li Dong. Improved log data-merging method for process mining. *Journal of South China University of Technology(Natural Science Edition)*, 2017, 45(01): 112-117
(徐杨, 林琪, 李东. 改进的面向流程挖掘的日志数据融合方法. *华南理工大学学报(自然科学版)*, 2017, 45(01): 112-117.)
- [6] Xu Yang, Yuan Feng, Lin Qi, et al. Merging event logs for process mining with a hybrid artificial immune algorithm. *Journal of Software*, 2018, 29(02): 396-416
(徐杨, 袁峰, 林琪等. 基于混合人工免疫算法的流程挖掘事件日志融合方法. *软件学报*, 2018, 29(02): 396-416.)
- [7] Lin Qi. Research and application of merging event logs for process mining. *South China University of Technology*, 2017
(林琪. 面向流程挖掘的事件日志融合研究与应用. 华南理工大学, 2017.)
- [8] Zeng Qing-Tian, Sun Sherry X., Duan Hua, et al. Cross-organizational collaborative workflow mining from a multi-source log. *Decision Support Systems*. 2013, 54(3): 1280-1301
- [9] Van Der Aalst W M P. Modeling and analyzing interorganizational workflows. *Proceedings of the International Conference on Application of Concurrency to System Design*. Washington, USA, 1998: 262-272
- [10] Van Der Aalst W M P. Interorganizational workflows: An approach based on message sequence charts and Petri nets. *Information & Management*, 2000, 37(2): 67-75
- [11] Mo Qi, Dai Fei, Zhu Rui, et al. Parallel modeling for cross-organization business process. *Computer Integrated Manufacturing Systems*, 2015, 21(04): 1051-1062
(莫启, 代飞, 朱锐等. 跨组织业务过程协同并行建模. *计算机集成制造系统*, 2015, 21(04): 1051-1062.)
- [12] Liu Cong, Zeng Qing-Tian, Cheng Long, et al. Privacy-preserving behavioral correctness verification of cross-organizational workflow with task synchronization patterns. *IEEE Transactions on Automation Science and Engineering*, 2020(99): 1-12
- [13] Hu Qing-Cheng, Xing Chun-Xiao, Yang Ji-Jiang, et al. Dynamic business process modeling and verification for inter-organizational collaboration. *Journal of Southeast University (English Edition)*, 2007, 23(03): 455-460
- [14] Li Qi-Ming, Deng Yong-Liang, Liu Cong, et al. Modeling and analysis of subway fire emergency response: An empirical study. *Safety Science*, 2016, 84: 171-180
- [15] Zeng Qing-Tian, Liu Cong, Duan Hua, et al. Resource conflict checking and resolution controller design for cross-organization emergency response. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, 2020, 50(10): 3685-3700
- [16] Duan Hua, Liu Cong, Zeng Qing-Tian, et al. Refinement-based hierarchical modeling and correctness verification of cross-organization collaborative emergency response process. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, 2020, 50(8): 2845-2859
- [17] Van D. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. New York, USA: Springer, 2011
- [18] Van Der Aalst W M P, Weijters T, Maruster L. Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 2004, 16(9): 1128-1142
- [19] Medeiros A K A D, Weijters A J M M, Van Der Aalst W M P. Genetic process mining: An experimental evaluation. *Data Mining & Knowledge Discovery*, 2007, 14(2): 245-304
- [20] Van Der Aalst W M P, Rubin V, Verbeek H M W, et al. Process mining: A two-step approach to balance between underfitting and overfitting. *Software & Systems Modeling*, 2010, 9(1): 87-111
- [21] Li G, De Carvalho R M, Van Der Aalst W M P. Configurable event correlation for process discovery from object-centric event data//*IEEE International Conference on Web Services*, 2018: 203-210
- [22] He Zhen, Wu Qiong, Wen Li-Jie, et al. A process mining approach to improve emergency rescue process of fatal gas explosion accidents in Chinese coal mines. *Safety Science*. 2019, 111: 154-166
- [23] Reisig W. *Petri Nets: An Introduction*. Springer Science & Business Media, 1985
- [24] Van Der Aalst W M P. The application of Petri nets to workflow management. *Journal of Circuits Systems & Computers*, 1998, 8(01): 21-66
- [25] Leemans S J J, Fahland D, Van Der Aalst W M P. *Discover-*

- ring block-structured process models from event logs—a constructive approach. *International Conference on Applications and Theory of Petri Nets and Concurrency*. Berlin, Germany: Springer, 2013: 311-329
- [26] Adriansyah A, van Dongen B F, van der Aalst W M P. Conformance checking using cost-based fitness analysis//*IEEE 15th International Enterprise Distributed Object Computing Conference*. 2011, 55-64
- [27] Adriansyah A, Munoz-Gama J, Carmona J, et al. Alignment based precision checking. *International Conference on Business Process Management*. Berlin, Germany: Springer, 2012, 137-149
- [28] De Weerd J, De Backer M, Vanthienen J, et al. A robust F-measure for evaluating discovered process models//*2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. 2011, 148-155
- [29] Van Der Aalst W. Service mining: Using process mining to discover, check, and improve service behavior. *IEEE Transactions on Services Computing*, 2012, 6(4): 525-535
- [30] Liu Cong, Li Hui-Ling, Zeng Qing-Tian, et al. Cross-organization emergency response process mining: An approach based on petri nets. *Mathematical Problems in Engineering*, Article ID 8836007, 2020: 1-15
- [31] Liu Cong, Cheng Long, Zeng Qing-Tian, et al. Petri net-based hierarchical business process discovery. *Computer Integrated Manufacturing Systems*, 2020, 26(6): 1525-1536 (刘聪, 程龙, 曾庆田等. 基于 Petri 网的分层过程模型挖掘方法. *计算机集成制造系统*, 2020, 26(6): 1525-1536)
- [32] Wang Ying, Liu Cong, Wen Li-Jie, et al. Hierarchical multi-instance processes model discovery approach. *Computer Integrated Manufacturing Systems*, 2022, 28(10): 3246-3255 (王颖, 刘聪, 闻立杰等. 分层多实例过程模型挖掘方法. *计算机集成制造系统*, 2022, 28(10): 3246-3255)
- [33] Liu Cong, Pei Yu-Long, Cheng Long, et al. Sampling business process event logs using graph-based ranking model. *Concurrency and Computation: Practice and Experience*, 2021, 33(5): 1-14
- [34] Su Xuan, Liu Cong, Zhang Shuai-Peng, et al. Event log sampling approach towards log completeness. *Computer Integrated Manufacturing Systems*, 2022, 28(10): 3156-3165 (苏轩, 刘聪, 张帅鹏等. 面向日志完备性的事件日志采样方法. *计算机集成制造系统*, 2022, 28(10): 3156-3165)



LIU Cong, Ph. D., professor. His research interests are in the areas of business process management, process mining, artificial intelligence.

LI Hui-Ling, M. S. candidate. Her research interests are in the areas of Petri nets, business process mining.

ZENG Qing-Tian, Ph. D., professor. His research interests are in the areas of process mining, and Petri nets.

DUAN Hua, Ph. D., professor. Her research interests are in the areas of Petri nets, formal method.

WEN Li-Jie, Ph. D., associate professor. His research interests are in the areas of process mining, process data management, and workflow for big data analysis.

Background

Cross-organization business processes have their own features, such as complex collaboration patterns and geographical dispersion of involved organizations. Therefore, their construction is a very complicated and error-prone task. Process mining provides an automated method for model construction by analyzing event logs collected from the execution of business information systems. However, traditional process mining technologies only support model discovery of a single organization, and cannot effectively deal with the problem of cross-organization business process mining.

This work proposes a novel process model discovery method for cross-organization collaborative processes, and gives a deep research for their quality evaluation. To solve this problem, a cross-organization business process model discovery method is proposed. More specifically, it first ex-

tends existing process mining methods to discover business process model of a single organization. Then, three typical collaboration patterns among organizations and their corresponding discovery algorithms are introduced. Next, process models and collaboration patterns are integrated to obtain a global cross-organization business process model. Finally, a set of evaluation metrics are invented to quantify the quality of discovered cross-organizational business process models. By comparing with the state-of-the-art process discovery techniques using four public cross-organization business process cases, the effectiveness and applicability of the proposed approach are illustrated. Based on the above-mentioned steps, a holistic discovery and evaluation approach for cross-organization business process model is proposed. This work broadens the application of process mining techniques to a multi-organization collaboration scenario.