

# 基于混合遗传算法的工作流重构研究

田珂<sup>1</sup> 朱清新<sup>2</sup> 向培素<sup>2</sup>

(电子科技大学计算机学院 成都 610054)<sup>1</sup> (西南民族大学电气信息工程学院 成都 610041)<sup>2</sup>

**摘要** 随着知识抽取、数据挖掘等技术的兴起,基于日志的企业流程重构引起了学术界和工业界的高度重视。本文针对目前重构算法大都采用本地策略因而无法保证最优挖掘以及算法对噪声不强壮的情况,提出了一种新的工作流重构算法。该算法结合遗传算法和模拟退火算法的思想,通过对流程活动依赖关系的度量,构建因果关系矩阵映射流程实例(CASE)作为种群染色体,进而完成初始种群的构建,有效地提高了收敛速度,减少时间开销,同时避免了早熟收敛,因此该算法在有效处理噪声问题的基础上,实现了对流程模型的正确挖掘。

**关键词** 活动依赖关系度量,工作流重构,模拟退火,混合遗传算法

## Study of Workflow Reconstruction Based on Hybrid Genetic Algorithm

TIAN Ke<sup>1</sup> ZHU Qing-Xin<sup>2</sup> XIANG Pei-Su<sup>2</sup>

(CCSE, University of Electronics Science and Technology of China, Chengdu 610054)<sup>1</sup>

(CEIE, Southwest University for Nationalities, Chengdu 610041)<sup>2</sup>

**Abstract** With the rise of the technology of knowledge extraction and data mining, it is highly noted in both research and industrial field that the workflow reconstruction based on process log. However, current techniques using local strategy can not guarantee that the outcome of the locally optimal steps will result in a globally optimal process model. And they are not robust when dealing with the presence of noise in the logs. To tackle these problems, this paper proposes a completely new approach based on hybrid genetic algorithm. As can be expected, this approach is able to deal with noise and incompleteness. However, it is not easy to represent process model properly in a genetic setting. In this paper, we define activity causal matrix as a representation for individuals. With the combination of genetic algorithm and simulated annealing algorithm, the reconstruction method can effectively deal with noise and incompleteness and correctly discover the process model.

**Keywords** Activity dependency metric, Workflow reconstruction, Simulated annealing, Hybrid genetic algorithm

## 1 引言

随着企业信息化程度的深入,对现存系统中进行知识抽取的需求就越迫切,业务过程智能(Business Process Intelligence)或者业务活动监控(Business Activity Monitoring)等的提出都是这些需求的实际体现。对于在当前企业信息系统(EIS)中发挥核心作用的工作流而言,基于流程日志发现企业流程模型(即流程重构)引起了学术界和工业界的高度重视。因为通过得出的流程模型,可以分析实际流程中出现的异常行为,度量设计模型和运行模型之间的差异,这些都助于流程设计的改进,从而提高 workflow 系统的适应性。

Cook 和 Wolf 最早在软件工程方面的文献中提出了有关流程重构的思想。在文[1]中,他们提出了包括神经网络、Markovian 在内的三种流程重构方法。在文[2]中, Cook 和 Wolf 又将他们的研究工作扩展到并行过程,并在文[3]中提出了一套基于事件数据的度量方法,用以量化流程模型和实际行为之间的差异。在文[4]中,首次在工作流管理上下文中引入了流程重构的思想。随后,国外学者在工作流挖掘方面进行了积极的探索,并取得了一定的成果。其中在文[5]中,通过启发式算法( $\alpha$ 算法),实现了对 AND-SPLIT/JOIN, OR-

SPLIT/JOIN 等工作流重要模式<sup>[6]</sup>的有效挖掘。在文[7]中,通过对  $\alpha$  算法进行扩充,实现了对 short-loops 模式的有效挖掘。

流程重构问题本质上就是在由流程日志构造的备选流程模式范围中选出同流程实际行为最匹配的模式。流程在结构上是由活动构成的,随着活动数目的增加,备选流程模式将会是一个巨大的搜索空间,因此流程重构问题也是一个 NP-Hard 计算问题<sup>[8]</sup>。鉴于上面研究文献中,普遍采用对于采用概率统计和归纳推理(Inductive inference)来进行流程重构,文[6]中提出的  $\alpha$  算法就是一种典型算法。 $\alpha$  算法在计算中仅利用日志记录中活动两两之间的事件信息,所以无法保证计算结果一定全局最优。此外,对于日志的完备性要求和日志 NOISE 的敏感也限制了算法的应用。针对这些缺陷,本文通过对流程活动依赖关系的度量,构建因果关系矩阵映射流程实例(CASE);并结合遗传算法(Genetic Algorithm,简称 GA)和模拟退火算法(Simulated Annealing Algorithm,简称 SA)的思想,提出了一种新的工作流重构算法。第 2 节首先介绍流程日志、活动间依赖关系度量和因果矩阵的相关概念,第 3 节详细介绍了混合遗传算法的步骤和参数确定;第 4 节对试验结果进行了分析,最后进行了总结。

田珂 博士研究生,主要研究方向:工作流技术、最优控制与搜索、电信网络管理;朱清新 博士生导师,主要研究方向:算法设计、系统仿真、最优控制与搜索;向培素 硕士,主要研究方向:电信网络管理、工作流技术、电子政务。

## 2 相关概念

下面给出了 workflow 重构中所使用的 workflow 日志、日志噪声、活动依赖关系度量和活动因果关系矩阵的形式化定义。

### 2.1 工作流日志

$T$  为一个工作流任务的集合,  $\sigma \in T^*$  是一个工作流 TRACE,  $W \in P(T^*)$  就是一个工作流日志。一个工作流日志就是工作流事件序列集合。工作流事件  $e$  就是任务  $t$  相对于一个给定的流程实例  $c$  的执行, 定义为  $e = (c, t)$ 。

### 2.2 日志噪声

噪声是很难处理的, 本文考虑两种噪声源: (1) 不完备的事件日志, 例如日志未能反映真实情况; (2) 日志记录的事件序列反映的是一种异常行为。不完备问题就是对于很多流程而言都是存在的。假设日志包含所有可能的行为是不现实的, 对于一个有很多选择性/并发分支的流程而言, 其所有可能事件序列的总数是流程所包含活动数量的  $n$  次幂。

### 2.3 活动依赖关系的度量

$t_1, t_2$  是 workflow 日志的两个事件,  $t_1, t_2$  之间的依赖关系定义如式(1)所示。其中  $\Gamma(t_1, t_2)$  代表活动  $T1$  和  $T2$  以  $t_1, t_2$  这样的事件序列方式出现在日志的次数。如果  $\Gamma(t_1, t_2)$  远大于  $\Gamma(t_2, t_1)$ , 那么我们可以认为活动  $T1$  和  $T2$  之间存在着因果关系。对于 short-loops 模式, 我们引入两个特别符号来表示:  $L_2(t_1)$  代表活动  $T1$  以  $t_1, t_1$  这样的事件序列出现在日志中的次数,  $L_2(t_1, t_2)$  代表活动  $T1$  和  $T2$  以  $t_1, t_2, t_1$  这样的事件序列出现在日志中的次数。

$$Dep(t_1, t_2) = \begin{cases} \frac{L_2(t_1, t_2) + L_2(t_2, t_1)}{L_2(t_1, t_2) + L_2(t_2, t_1) + 1} & \text{if } t_1 \neq t_2 \text{ 且 } L_2(t_1, t_2) > 0 \\ \frac{\Gamma(t_1, t_2) - \Gamma(t_2, t_1)}{\Gamma(t_1, t_2) + \Gamma(t_2, t_1) + 1} & \text{if } t_1 \neq t_2 \text{ 且 } L_2(t_1, t_2) = 0 \\ \frac{L_2(t_1, t_2)}{L_2(t_2, t_1) + 1} & \text{if } t_1 = t_2 \end{cases} \quad (1)$$

### 2.4 活动因果关系矩阵

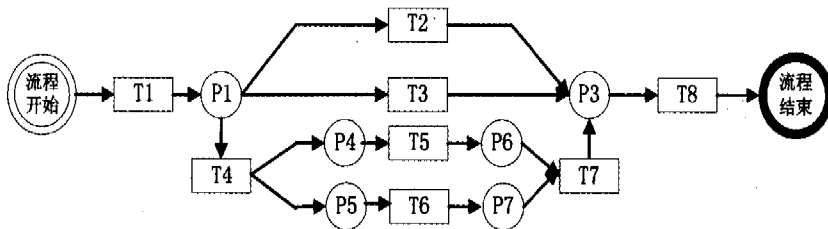


图1 一个流程模型的 Petri 网

流程模型在形式上可以表达为一个布尔表达式矩阵, 该矩阵通过布尔表达式的值来反映活动之间的因果关系, 因此也称之为因果关系矩阵。该矩阵是一个  $N \times N$  的方阵,  $N$  是流程模型中活动的个数。我们可以使用  $(\wedge)$  和  $(\vee)$  这样算子构成的布尔表达式来描述象 AND/OR-SPLIT/JOIN 等这样的工作流模式, 因此用布尔表达式来描述流程的路由结构在理论上是可行的。例如, 考虑图 1 中的  $T8$  活动。这个活动可以在活动  $T2$  或是  $T3, T7$  发生的任何时刻发生。这样, 行  $T8$  有着形如  $T2 \vee T3 \vee T7$  的布尔表达式。类似地, 输出列中的布尔表达式显示了当一个行的活动被执行后哪个活动可能被执行。例如, 行  $T4$  有着  $T5 \wedge T6$  这样的布尔表达式输出。下面我们将阐述图 1 中的 Petri 网是如何用图 2 中的因果关系矩阵来描述的。图 1 中的 Petri 网有 8 个活动(其中  $T1$  为起始活动、 $T8$  为结束活动)。因果矩阵中的元素描述了在两个活动间是否存在因果关系。如果它等于 1, 那么行活动和列活动之间存在因果关系; 如果它等于 0, 则没有因果关系。决定因果矩阵布尔值的基本思想在于, 如果对  $T1, T2$  两个活动,  $Dep(t_1, t_2)$  的值很高, 则很可能  $causal(t_1, t_2)$  为真(其值为 1)。

在活动依赖关系  $Dep(t_1, t_2)$  的基础上, 我们可以给出因果关系矩阵  $Mat_c$  的定义。  $Mat_c$  是一个四元组,  $Mat_c = (S_a, C_r, I, O)$ , 其中  $S_a$  是活动的有限集,  $C_r$  是活动之间的因果关系  $C_r \subseteq A \times A$ ,  $I$  是活动的输入约束函数,  $O$  是活动的输出约束函数。

描述一个流程必须要包括三方面的信息: (1) 流程包含哪些活动; (2) 哪些活动是前驱; 哪些活动是后继; (3) 活动之间的因果关系是以什么模式组织起来的。例如: 串行 (SEQUENCE)、并行 (PARALLEL) 还是选择 (CHOOSE) 等其它

模式。由上可知, 尽管在表达能力上因果关系矩阵没有 Petri 网那样直观, 但它包含了上面三方面的信息, 因此在描述流程能力上和 Petri 网是等价的, 而 Petri 网无法直接应用在遗传算法中。本文使用活动因果关系矩阵  $Mat_c$  完成与 Petri 网的映射, 并用  $Mat_c$  来表示遗传算法中的染色体 (Individual)。在构建初始种群的时候, 对于一个工作流日志, 遗传算法中所有的种群的染色体, 都拥有同样的活动集合  $S_a$  (该集合包含出现在日志中的所有活动), 但因果关系  $C_r$ 、输入约束函数、输出约束函数各异。初始种群可以包含定义在活动集合  $S_a$  中的任意染色体, 日志中记录的活动数目越多, 搜索空间就越大。为了对染色体进行编码, 首先需要跟踪染色体中所有活动的输入输出布尔表达式, 从布尔表达式中直接构建出因果矩阵; 同时布尔表达式还可以反映活动之间的关系是 AND 还是 OR。我们可以将图 2 中矩阵的一行作为一段基因代码, 例如:  $\{01110000\}$ , 将矩阵中所有行按顺序组合在一起, 就形成了一个染色体。

## 3 混合遗传算法

遗传算法 (Genetic Algorithm, GA)<sup>[9]</sup> 因其高度的并行处理能力、强鲁棒性和全局搜索能力而被广泛地应用于诸多领域。然而实践中, 遗传算法会表现出早熟现象、局部寻优能力较差等不足。而模拟退火 (Simulated Annealing, SA) 具有较强的局部寻优能力, 并能使搜索过程避免陷入局部最优解, 但把握整个搜索过程的能力不够, 不便于使搜索过程进入最有希望的搜索区域, 从而使得模拟退火的运算效率不高<sup>[9]</sup>。本文将模拟退火算法与遗传算法结合起来, 提出基于模拟退火的混合遗传算法, 算法主要步骤如图 3 所示。

### 3.1 创建初始种群

由第2节可知,首先通过分析 workflow 日志、度量活动间依赖关系;然后通过构造活动因果关系矩阵  $Mat_c$  作为种群的染色体,最后随机产生长度为  $N \times N$  ( $N$  为流程模型中活动数量的总和)的  $popSize$  个二进制编码,作为初始种群。为了让初始种群遍及整个解空间,尽量反映搜索空间的性质,  $popSize$  不能取太小,且随活动数目的增多而变大;不过太大会使运算时间增加,不是我们所希望的。  $popSize$  取染色体长度数的一个线性倍数是实际应用时经常采用的方法,如可取为  $N$  和  $2N$  之间的一个确定数。同时为了保持网络的放射性,以及不形成“孤岛”,在计算程序中加入了搜索判断程序,将不符合实际运行条件的染色体去掉。



图2 一个活动因果关系矩阵

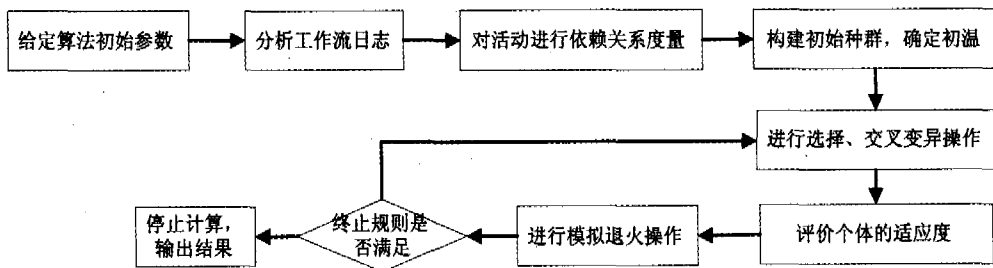


图3 算法结构图

### 3.2 适应函数和初温的确定

适应度是遗传算法指导搜索方向的依据, 首先应保证适应度不为负。其次, 因为在轮盘赌选择过程中, 适应函数值大的染色体被选中的概率大, 因此目标函数的优化方向对应适应函数值增加的方向。如果在基因种群中的染色体正确地描述了在事件日志中记录的行为, 染色体的适应度就高。在我们的方法中, 染色体适应度是日志中能够正确解析的事件序列的数量紧密相联的。但在噪声环境中, 我们不能期望流程模型能够正确地解析所有的事件序列。在下面式(2)中我们给出适应函数的定义:

$$Fit(P_m, L) = 0.4 \times \frac{NumPA}{MumAL} + 0.6 \times \frac{NumPTL}{NumTL} \quad (2)$$

其中,  $NumAL$  是日志中出现的活动数目,  $NumTL$  是日志中出现的 TRACE 的数目,  $NumPA$  是在所有  $NumTL$  中按照 Petri 网规则可以正常点火的活动 (fire) 的总数,  $NumPTL$  是指按照备选流程模型可以执行到最后一个库所 (即流程正常结束) 的 TRACE 的数目。

初温的确定, 选择初温  $t_0 = K\delta$  的形式, 其中  $K$  为充分大的数,  $\delta = Fit_{max} - Fit_{avg}$ , 其中  $Fit_{max}$  为初始种群中最大适应度,  $Fit_{min}$  为初始种群中最小适应度, 退温函数选用常用的  $t_n + 1 = \alpha t_n$  形式, 其中  $0 < \alpha < 1$ 。

### 3.3 选择复制操作

根据适应函数值的大小进行群体选择, 选择方法采用常用的轮盘赌选择法, 由于简单遗传算法不具有全局收敛性<sup>[10]</sup>, 因此对算法进行改动, 实施了最优保留策略, 即运算开始时, 把初始种群中最大适应函数值对应的染色体记录在染色体数组的第 0 位 (如  $chromo[0]$ )。每代进化完后, 计算出  $chromo[0]$  染色体在当前温度下对应的适应函数值。如果小于 1, 则把当前代最大适应函数值对应的染色体赋给  $chromo[0]$ ; 如果等于 1, 则不进行替换。染色体  $chromo[0]$  只起记录功能, 不参与遗传运算。实施了最优保留策略的遗传算法将收敛到全局最优解。

### 3.4 交叉率和变异率的确定

$$P_c = \begin{cases} K_1 + \frac{K_2 (Fit_{max} - Fit_i')}{Fit_{max} - Fit_{avg}} & \text{if } Fit_i \geq Fit_{avg} \\ K_3 & \text{if } Fit_i' < Fit_{avg} \end{cases} \quad (3)$$

$$P_m = \begin{cases} K_4 + \frac{K_5 (Fit_{max} - Fit_i)}{Fit_{max} - Fit_{avg}} & \text{if } Fit_i \geq Fit_{avg} \\ K_6 & \text{if } Fit_i < Fit_{avg} \end{cases} \quad (4)$$

上面给出了交叉率  $P_c$  和变异率  $P_m$  的计算公式, 其中  $K_1, K_2, K_3, K_4, K_5, K_6$  为常数, 且  $K_3 = K_1 + K_2, K_6 = K_4 + K_5$ ;  $Fit_{max}, Fit_{avg}$  分别为当前代进化种群的最大和平均适应度;  $Fit_i'$  为两个交叉染色体中较大的适应度。这样在操作过程中, 适应函数值小的染色体, 具有较大的交叉率和变异率, 这样有利于加快搜索速度, 且最大适应函数值对应染色体的交叉率和变异率不为零, 而为一较小值。当遗传算法陷入问题的局部极值时, 即  $Fit_{avg} \rightarrow Fit_{max}$  时, 根据式(3)和(4), 适应函数值较大的染色体对应的  $P_c, P_m$  也将增大, 这样有利于避免“早熟”。但太大的  $P_c, P_m$  有可能造成解空间过于分散。为此, 一旦  $Fit_{max} - Fit_{avg} < \epsilon$ , 例如  $\epsilon = 0.5 Fit_{avg}$  时, 就固定  $P_c, P_m$  的值, 以避免原有解空间被完全破坏。

### 3.5 交叉 (CROSSOVER) 和变异 (MUTATION) 操作

交叉方法的两个父染色体 (parent1, parent2) 通过交叉产生两个新染色体 (offspring1, offspring2)。在进行交叉操作的时候, 随机选择染色体中的基因代码片段 (对应活动 T) 作为交叉点进行交换, 当染色体 (Individual) 来自不同的因果关系矩阵的时候, CROSSOVER 操作可能会导致不一致的情况。经过 CROSSOVER 操作后, 一个活动也许没有出现在另一活动 input/output 一个布尔表达式中, 但是因果矩阵中的相应位置仍然为 1。因此在输入/输出集合被重组后, 我们需要校验对于其余活动而言, 矩阵的布尔表达式是否仍保持一致。当出现不一致的情况时, 需要对因果矩阵和布尔表达式进行更新。这样操作可以避免大量不可行解的产生, 提高计算效率。

变异操作发生在活动的  $input(t)$  和  $output(t)$  布尔表达式上。对于染色体中的每一个基因 (活动) 而言, 任意选择一个随机数  $r$ 。当  $r$  小于  $P_m$  的时候, 对活动  $t$  的输入  $input(t)$

(下转第 111 页)

表1 实验结果

|     | 实际数目 | 探测正确 | 探测错误 | 未探测到 | 准确率% | 完全率% |
|-----|------|------|------|------|------|------|
| 远景  | 789  | 743  | 32   | 46   | 95.9 | 94.2 |
| 中景  | 358  | 330  | 49   | 28   | 87.1 | 92.2 |
| 近景  | 331  | 306  | 18   | 25   | 94.4 | 92.4 |
| 场内  | 1335 | 1285 | 49   | 50   | 96.3 | 96.2 |
| 场外  | 142  | 130  | 13   | 12   | 89.7 | 91.5 |
| 慢镜头 | 105  | 97   | 10   | 8    | 90.9 | 92.3 |
| 兴奋音 | 29   | 23   | 3    | 2    | 88.5 | 92.0 |
| 事件  | 11   | 10   | 1    | 1    | 90.9 | 90.9 |

**结论与展望** 本文中,我们分析了足球视频的语义结构,每个语义事件可视为特定的镜头序列,这些镜头因而可称为“语法镜头”。对已经分割出来的物理镜头,选择相应的中间层内容,分析出语法镜头;对音频流分析出解说员的“兴奋音”。最后综合视频流和音频流的语法内容,推理出精彩事件。实验表明效果良好,这种方法也可用于其他体育视频的分析,如篮球、棒球等。但是,我们分析出的语义事件还比较粗略。在未来,我们要确定镜头的语义,如裁判员镜头;还要细化“语义事件”,如犯规、任意球等。另外,还要将这种方法应用到其他体育视频的语义分析。这将要求我们选择更多的中间层内容,并要对视频中的文字进行识别。

(上接第105页)

和  $output(t)$  进行 AND/OR 操作。操作后,染色体的因果关系矩阵并没有发生变化,改变的只是活动  $t$  的路由模式 AND/OR/JOIN-SPLIT。

### 3.4 终止规则

当满足下面任一条件时,即可认为算法收敛,停止计算,输出最优解:(a)染色体的适应度达到给定的阈值(例如1);(b)经过  $N$  代进化,当  $N$  是所允许计算代(generation)的最大值;(c)染色体数组的第0位  $chromo[0]$  连续  $q$  代没有发生变化。

### 3.6 模拟退火操作

对于经过遗传算法选择复制、交叉、变异操作的群体作为模拟退火算法的初始群体,运用基于 Metropolis 判别准则的复制策略,产生下一代群体。即在染色体的邻域中随机产生新染色体  $i$  和  $j$ ,竞争进入下一代群体的准则采用 Metropolis 判别准则:令  $\Delta F = Fit_i - Fit_j$ ,若  $\Delta F \leq 0$ ,则把染色体  $j$  复制到下一代群体,否则产生  $[0, 1]$  之间的随机数  $r$ 。如果  $r < \exp(\Delta F/t_n)$ ,则同样把染色体  $j$  复制到下一代群体,否则,把染色体  $i$  复制到下一代群体。基于 Metropolis 判别准则的复制策略,在接受优质解的同时,有限度地接受劣质解,保证了群体的多样性,进一步避免了算法陷入局部最优解的可能性。

### 3.7 试验

基于本文提出的 workflow 重构算法,对包含活动数目为 8~25 的流程日志进行测试,测试日志中包含 5%~15% 的噪声数据。初始群体规模  $popSize = 2N = 50$ ,算法中各系数确定如下:选择确定初温系数  $K = 20$ ,退温操作系数  $\alpha = 0.8$ ,交叉操作系数  $K_1 = 0.2, K_2 = 0.6, K_3 = K_1 + K_2 = 0.8$ ,变异操作系数  $K_4 = 0.01, K_5 = 0.1, K_6 = K_4 + K_5 = 0.11$ 。为防止解空间被破坏,当  $\epsilon = 0.5Fit_{avg}$  时,交叉率和变异率固定为 0.2 和 0.01,算法终止规则中  $q = 15, N = 25$ 。试验结果表明:该算法能够在包含噪声数据的工作流日志中成功进行流程重构,并且由于采用全局策略和混合遗传算法,有效地避免了结果陷入局部最优解;并通过提高收敛速度,减少了流程模型重

## 参考文献

- 1 Ekin A, Tekalp A M. Generic event detection in sports video using cinematic features. In: Second IEEE Workshop on Event Mining; Detection and Recognition of Events in Video (EVENT 2003), Madison, Wisconsin, USA, June 2003. 17~24
- 2 Yu X, Xu C, Leong H, Tian Q, Wan K. Trajectory-Based Ball Detection and Tracking with Applications to Semantic Analysis of Broadcast Soccer Video. In: Proc. of ACM Multimedia, 2003. 11~20
- 3 Babaguchi N, Kawai Y, Kitashi T. Event based indexing of broadcasted sports video by intermodal collaboration. IEEE Trans. Multimedia, 2002, 4(1): 68~75
- 4 Xie L, Xu P, Chang S F, Divakaran A, Sun H. Structure analysis of soccer video with domain knowledge and hidden Markov models. PRL(25), May 2004(7): 767~775
- 5 Liu X M, Chen T. Shot Boundary Detection Using Temporal Statistics Modeling. In: IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP2002, Orlando, FL, U. S, May 2002
- 6 Lenardi R, Migliorati P, Prandini M. Semantic Indexing of Soccer Audio-Visual Sequence, A multimodal approach based on controlled Markov chains. IEEE Trans on Circuits & System for Video Technology, 2004, (5): 634~643
- 7 Barnard M, Odobez J M, Bengio S. Multi-modal audio-visual event recognition for football analysis. In: IEEE Workshop on Neural Networks for Signal Processing, NNSP, 2003. 469~478
- 8 Baillie M, Jose J M. Audio-based Event Detection for Sports Video, CIVR2003, July 2003. 300~310

构的时间开销。

**结论** 随着知识抽取、数据挖掘等技术的兴起,基于日志的企业流程重构引起了学术界和工业界的高度重视。流程在结构上是由活动构成的,随着活动数目的增加,备选流程模式将会构成一个巨大的搜索空间。流程重构问题的实质就是在搜索空间中选出同流程实际行为最匹配的模式。本文针对目前重构算法采用本地策略以及无法有效处理噪声的情况,提出了一种新的 workflow 重构算法。该算法通过对流程活动依赖关系的度量,构建因果关系矩阵对流程实例(CASE)进行映射,然后通过因果关系矩阵来构建初始种群,结合遗传算法和模拟退火算法的思想,实现了对流程模型的有效挖掘。

## 参考文献

- 1 Cook J E, Wolf A L. Discovering Models of Software Processes from Event-Based Data. ACM Transactions on Software Engineering and Methodology, 1998, 7(3): 215~249
- 2 Cook J E, Wolf A L. Event-Based Detection of Concurrency. In: Proceedings of the Sixth International Symposium on the Foundations of Software Engineering (FSE-6), 1998. 35~45
- 3 Cook J E, Wolf A L. Software Process Validation: Quantitatively Measuring the Correspondence of a Process to a Model. ACM Transactions on Software Engineering and Methodology, 1999, 8(2): 147~176
- 4 Agrawal R, Gunopulos D, Leymann F. Mining Process Models from Workflow Logs. In: Sixth International Conference on Extending Database Technology, 1998. 469~483
- 5 Maruster L, Weijters A J M M, van der Aalst W M P, et al. Process Mining: Discovering Direct Successors in Process Logs. In: Proceedings of the 5th International Conference on Discovery Science (Discovery Science 2002), Vol 2534 of Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag, 2002. 364~373
- 6 Kiepuszewski. Expressiveness and Suitability of Languages for Control Flow Modeling in Workflows (Submitted); [Ph D Thesis]. Queensland University of Technology, 2002
- 7 de Medeiros A K A, van Dongen B F, van der Aalst W M P, et al. Process Mining: Extending the  $\alpha$ -algorithm to Mine Short Loops. BETA Working Paper Series, WP 113. Eindhoven University of Technology, Eindhoven, 2004
- 8 Hochbaum D. Approximation Algorithms for NP-hard Problems [M]. Berkeley, CA, PWS Publishing Company, 1997
- 9 周丽,孙树栋. 遗传算法原理及应用[M]. 北京:国防工业出版社, 2001
- 10 欧阳森,宋政湘,王建华,等. 一种快速收敛的遗传算法[J]. 计算机应用研究, 2003, 20(9): 50~52