



(12)发明专利

(10)授权公告号 CN 103793597 B

(45)授权公告日 2017. 02. 22

(21)申请号 201410019455.8

审查员 黎宾彬

(22)申请日 2014.01.16

(65)同一申请的已公布的文献号

申请公布号 CN 103793597 A

(43)申请公布日 2014.05.14

(73)专利权人 同济大学

地址 200092 上海市杨浦区四平路1239号

(72)发明人 蒋昌俊 陈阔中 闫春钢 丁志军

于汪洋

(74)专利代理机构 上海天协和诚知识产权代理

事务所 31216

代理人 叶凤

(51)Int.Cl.

G06F 19/00(2011.01)

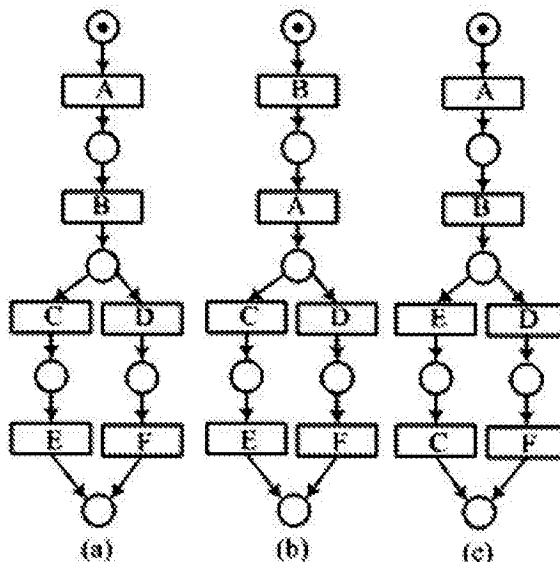
权利要求书1页 说明书3页 附图6页

(54)发明名称

基于完备主干子系统的模型相似度度量方法

(57)摘要

一种基于完备主干子系统的模型相似度度量方法,依次包括对模型分解得到完备主干子系统、为个完备主干子系统建立行为轮廓、进行完备主干子系统相似度比较、计算模型相似度。为了提高模型相似度的度量精度,本发明提供一种基于完备主干子系统的模型相似度度量方法,该方法从三个方面对因果行为轮廓进行了改造。一是将变迁对的重要度纳入度量范畴;二是更细粒度刻画因果行为轮廓;三是深化变迁对的一致度。可以从行为角度提高模型相似度度量精确,为模型检索、模型合并、模型复用等操作提供支持。本发明方法还可用于软件行为可预期性评估,有利于提高软件可信性。



1. 一种基于完备主干子系统的模型相似度量方法,其特征在于,依次包括对模型分解得到完备主干子系统、为每个完备主干子系统建立行为轮廓、进行完备主干子系统相似比较、计算模型相似度,具体步骤为:

1) 采用了基于完备主干子系统的度量方法,首先在工作流网上增加一个短循环变迁,通过求变迁不变量的外延子网,将模型分解成若干完备主干子系统;一个完备主干子系统由一个简单主干子系统和若干循环结构构成,所述完备主干子系统描述了一类实例的执行路径;

2) 将行为轮廓定义在完备主干子系统上,每个完备主干子系统拥有一个行为轮廓,这些行为轮廓的集合构成模型的行为轮廓;行为轮廓中不将并发关系并入交叉关系中,同时增加冗余变迁对,用于刻画简单主干子系统与其循环结构的公共部分;这些冗余的变迁作为计算变迁对一致度的约束;冗余变迁对由简单主干子系统生成;

3) 因果行为轮廓的相似度是建立在变迁对一致度基础上,将变迁对一致度定义为与该变迁对匹配成功的变迁对的个数和该变迁对应匹配的变迁对个数的比值,变迁对一致度被映射到 $[0, 1]$ 区间上的某个值;

4) 每个完备主干子系统依次与另一模型的完备主干子系统进行相似度计算,定义为所有变迁对一致度的和与变迁对总数的比值,最后模型相似度定义为各完备主干子系统最大相似度的平均值。

基于完备主干子系统的模型相似度度量方法

技术领域

[0001] 本发明涉及一种业务模型相似度度量方法。

背景技术

[0002] 如今,业务流程模型普遍用于业界分析业务流程或创建新模型。大规模的企业通常会有流程模型库,内含数以百计或数以千计的模型。这些模型由不同人员开发,是企业宝贵的智力财富。为了方便管理和操作这些模型,如模型搜索、模型合并等,需要研究模型相似度度量方法。

[0003] 目前已经有许多模型相似度度量方法。基于迹等价、互模拟和分支互模拟的方法只能给出是与否的二元回答,例如用1表示等价,用0表示不等价。然而在实际应用中,人们往往希望知道不同模型之间有多大程度的不同,这就使得这些方法的应用受到限制。迹相似度被定义为两个模型中相同迹的个数与迹的总数的比值。该方法与迹等价方法面临着同样的问题,就是由于模型中存在并发、选择或环结构,会造成迹的集合非常大甚至无穷,这给计算带来很大困难。而且迹相似度方法仍然过于严格,两个迹稍微不同都会造成整个迹不能匹配。为此,有学者提出了基于观测行为的度量方法,通过比较事件日志中的典型行为,计算precision和recall指标,以度量模型的相似度。该方法受制于事件日志的完备性。基于变迁邻接关系的方法可看作是对基于活动依赖图的方法的改进,但这种方法只关注直接的因果关系,忽略了间接因果关系,对模型的变化缺乏足够的敏感性。基于因果足迹的方法计算效率太低。Weidlich et al.提出了基于因果行为轮廓的度量方法,将因果轮廓定义为严序关系、交叉关系、互斥关系和共现关系的集合。该方法可以非常高效地计算自由选择网的模型相似度,但与前述方法一样,度量精度依然不足。

发明内容

[0004] 为了提高模型相似度的度量精度,本发明提供一种基于完备主干子系统的模型相似度度量方法,该方法从三个方面对因果行为轮廓进行了改造。一是将变迁对的重要度纳入度量范畴;二是更细粒度刻画因果行为轮廓;三是深化变迁对的一致度。

[0005] 本发明给出的技术方案为:

[0006] 一种基于完备主干子系统的模型相似度度量方法,其特征在于,依次包括对模型分解得到完备主干子系统、为个完备主干子系统建立行为轮廓、进行完备主干子系统相似度比较、计算模型相似度,具体步骤为:

[0007] 1)对于变迁对的重要度,本发明采用了基于完备主干子系统的度量方法。首先在工作流网上增加一个短循环变迁,通过求变迁不变量的外延子网,将模型分解成若干完备主干子系统。一个完备主干子系统由一个简单主干子系统和若干循环结构构成,它描述了一类实例的执行路径。这些路径至少执行了简单主干子系统中所有活动,还可能执行任意次循环结构中的活动。本发明并不直接对变迁对的重要度进行计算,而是通过对各完备主干子系统相似度计算中得以体现。一个变迁在各完备主干子系统中出现的次数越多,其参

与计算的次数就越多,对于模型相似度的影响也就越大,从而表明其越重要。

[0008] 2) 为了更细粒度刻画行为轮廓,本发明将行为轮廓定义在完备主干子系统上,每个完备主干子系统拥有一个行为轮廓,这些行为轮廓的集合构成模型的行为轮廓。为减少信息损失,行为轮廓中不再将并发关系并入交叉关系中,同时适当增加冗余变迁对,用于刻画简单主干子系统与其循环结构的公共部分。这样并发关系与交叉关系的交集以及严序关系与交叉关系的交集都可能不再为空。这些冗余的变迁可作为计算变迁对一致度的约束。冗余变迁对由简单主干子系统生成。

[0009] 3) 因果行为轮廓的相似度是建立在变迁对一致度基础上的。本发明将变迁一致度定义为与该变迁对匹配成功的变迁对的个数和该变迁对应匹配的变迁对个数的比值。这样变迁对一致度被映射到 $[0, 1]$ 区间上的某个值,而不再是0或1。这比原有方法在计算上又前进一步。

[0010] 4) 每个完备主干子系统依次与另一模型的完备主干子系统进行相似度计算,定义为所有变迁对一致度的和与变迁对总数的比值。最后模型相似度定义为各完备主干子系统最大相似度的平均值。

[0011] 本发明方法主要从行为的角度度量模型相似度,不考虑标签相似度和结构相似度。本发明将变迁对的重要度、变迁对的一致度都纳入相似度度量范畴,更精确定义了行为轮廓的概念。本发明具有的有益效果是:可以从行为角度提高模型相似度度量精确,为模型检索、模型合并、模型复用等操作提供支持。本发明方法还可用于软件行为可预期性评估,有利于提高软件可信性。

[0012] 本发明方法采用自由选择的工作流网来描述模型,一是因为工作流网已成为当今描述工作流程最为广泛的形式化模型;二是因为大多数工作流管理系统只允许自由选择网工作流;三是因为自由选择网已得到广泛研究且易于理解,其soundness属性验证和行为轮廓获取都可在多项式时间内完成。

附图说明

[0013] 图1为完备主干子系统求取过程。

[0014] 图2为变迁对重要度示例。

[0015] 图3为将并发关系与交叉关系分开必要性示例。

[0016] 图4为增加冗余变迁对的必要性示例。

[0017] 图5为变迁一致度细化示例。

[0018] 图6为全求解过程示例。

[0019] 图7为完备主干子系统。

具体实施方式

[0020] 下面结合实例对本发明保护的技术方案作进一步说明。

[0021] 图1说明了完备主干子系统的求取过程。图1 (a) 是一个sound自由选择工作流网,在其实库所和终止库所之间加上一个短循环变迁 t^* ,一条从终止库所指向 t^* 的弧,和一条从 t^* 指向起始库所的弧,就构成一个短循环网。(b)、(c)、(d) 是三个简单主干子系统,(e) 和(f) 是循环结构。通过合成,可得到(g)、(h)、(i) 三个完备主干子系统。

[0022] 注意到变迁对 (D, H) 在 (h) 中既属于严序关系又属于交叉关系。变迁对 (J, I) 在 (g) 中属于交叉关系, 但不属于严序关系, 而在 (h) 中既属于交叉关系又属于严序关系。下面通过几个简单的例子说明本发明方法的优越性。

[0023] 图2可以说明本发明方法能够体现变迁对的重要性。相比于模型 (a), 模型 (b) 调换了活动A和B的位置, 而模型 (c) 调换了活动C和E的位置。按照因果行为轮廓的方法, (b) 与 (a) 的相似度等于 (c) 与 (a) 的相似度。然而从图2中可以看出, (a) 中两条路径都经过A和B, 而只有一条路径经过C和E, 显然变迁对 (A, B) 比变迁对 (C, E) 更重要。运用本发明可以得到 (b) 与 (a) 的相似度为0.875, 而 (c) 与 (a) 的相似度为0.938, 很好体现了A和B调换对原模型影响更大的事实。

[0024] 图3可以说明将并发关系与交叉关系分开的好处。按照因果行为轮廓的方法, 模型 (a) 和模型 (b) 不能被区分开来, 它们的相似度是1。而运用本发明方法, 可得到它们的相似度是0.875。因为 (B, C) 在 (a) 中属于严序关系和交叉关系, 而在 (b) 中, 却属于并发关系和交叉关系。

[0025] 图4可说明增加冗余变迁对是有必要的。增加冗余关系之后, 就可以有效区分模型 (a) 和模型 (b) 了。

[0026] 图5可以说明本发明方法对变迁对一致度的深化。图5中背景和虚线表明了模型 (a) 与模型 (b) 中的活动对应关系。按照因果行为轮廓的方法, 变迁对 (B, C) 将不能成功匹配, 因为 (B, C) 在 (a) 中属于严序关系, 而在 (b) 中, (B, C1) 却属于并发关系。所以 (B, C) 的一致度是0。然而, 其他三个变迁对 (B1, C1), (B1, C2) 和 (B2, C2) 都是可以与 (B, C) 匹配的。在本发明方法技术中, (B, C) 的一致度将对应为0.75。显然, 这比因果行为轮廓的方法更合理, 计算粒度上也更精细。

[0027] 图5还说明了另一个重要问题: 共现关系并不能很好解决行为轮廓精度问题。从图4的例子中可以看出, 有时候共现关系显得太弱。而在图5中, 共现关系却又显得太严格。例如图5 (a) 中A1、A2与B是共现关系, 图5 (b) 中由于A3和A4不能同时与B1和B2是共现关系, 因此匹配时不成功, 最终模型相似度只有0.63。然而就总体而言, 这种共现关系在 (a) 和 (b) 中都是存在的。应用本发明方法技术计算所得结果是0.817, 显示了更好的合理性。

[0028] 最后, 我们通过一个简单的例子 (图6) 说明本发明方法技术操作的全过程。首先, 对模型 (a) 和 (b) 进行分解, 得到完备主干子系统, 如图7所示。

[0029] 第二步, 为个完备主干子系统建立行为轮廓。

[0030] 第三步, 进行完备主干子系统相似度比较, 表1列出了比较结果。

[0031] 表1 模型 (a) 与模型 (b) 各完备主干路径相似度比较

[0032]

相似度	a1	a2	a3	a4
b1	1	0.72	0.72	0.4
b2	0.4	0.62	0.62	1

[0033] 最后, 计算模型相似度为 $(1+0.72+0.72+1+1+1)/6=0.91$ 。

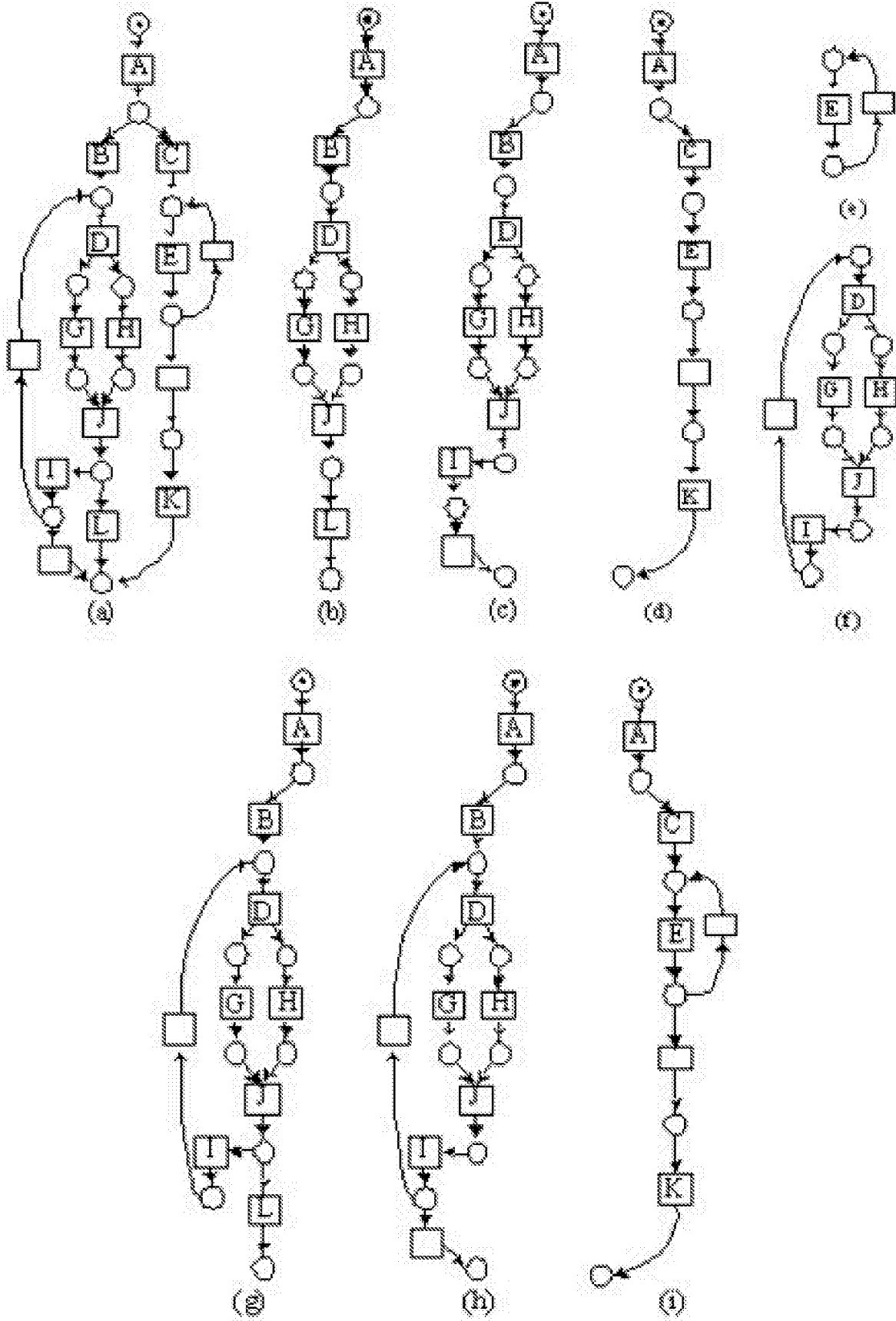


图1

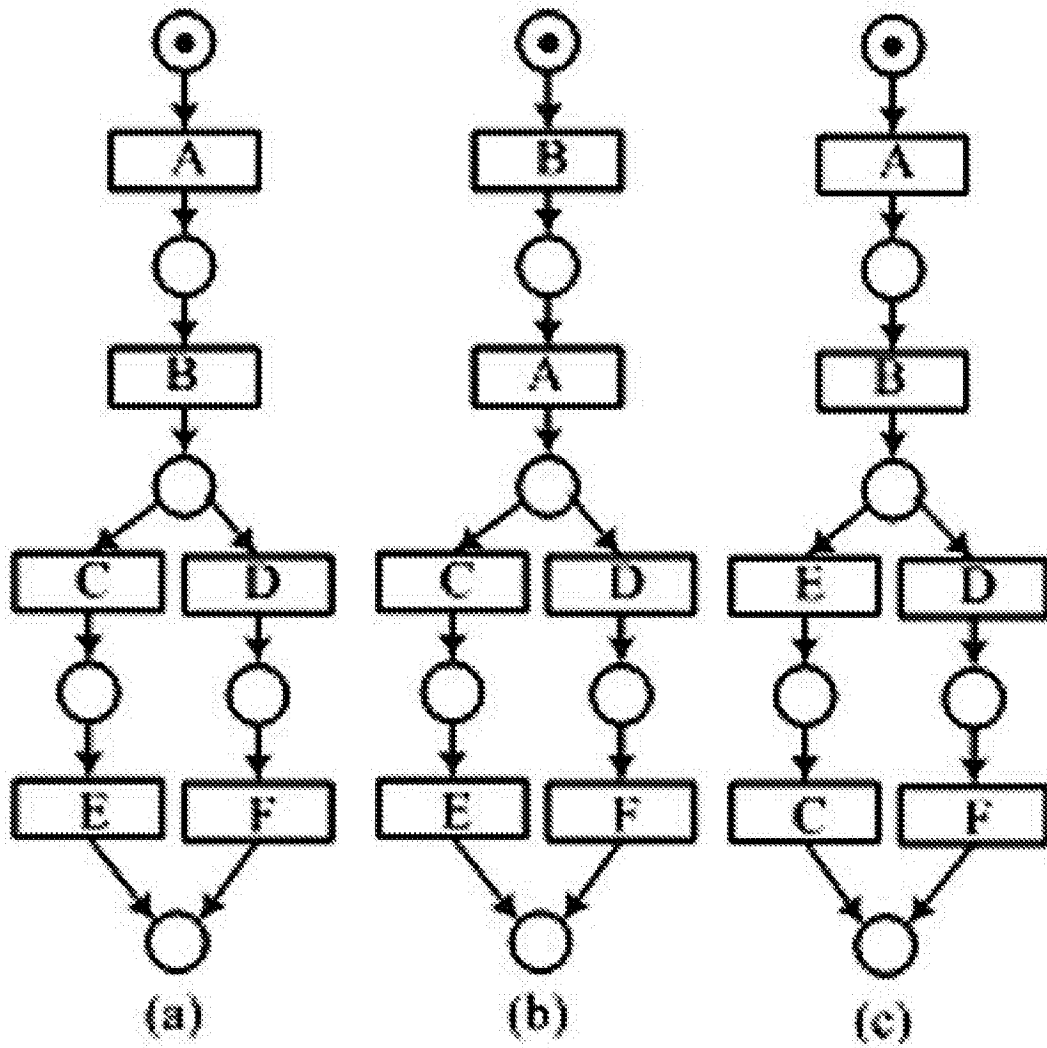


图2

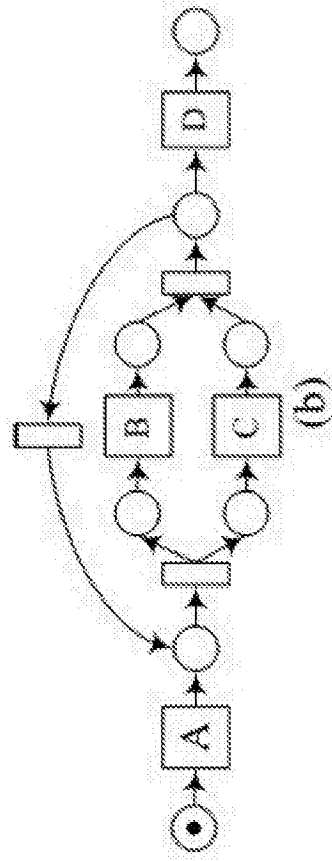


图3

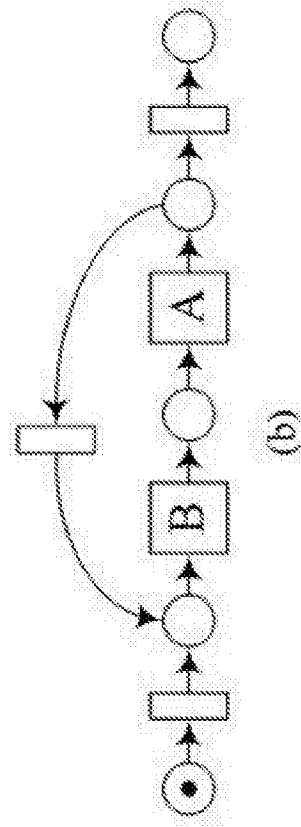


图4

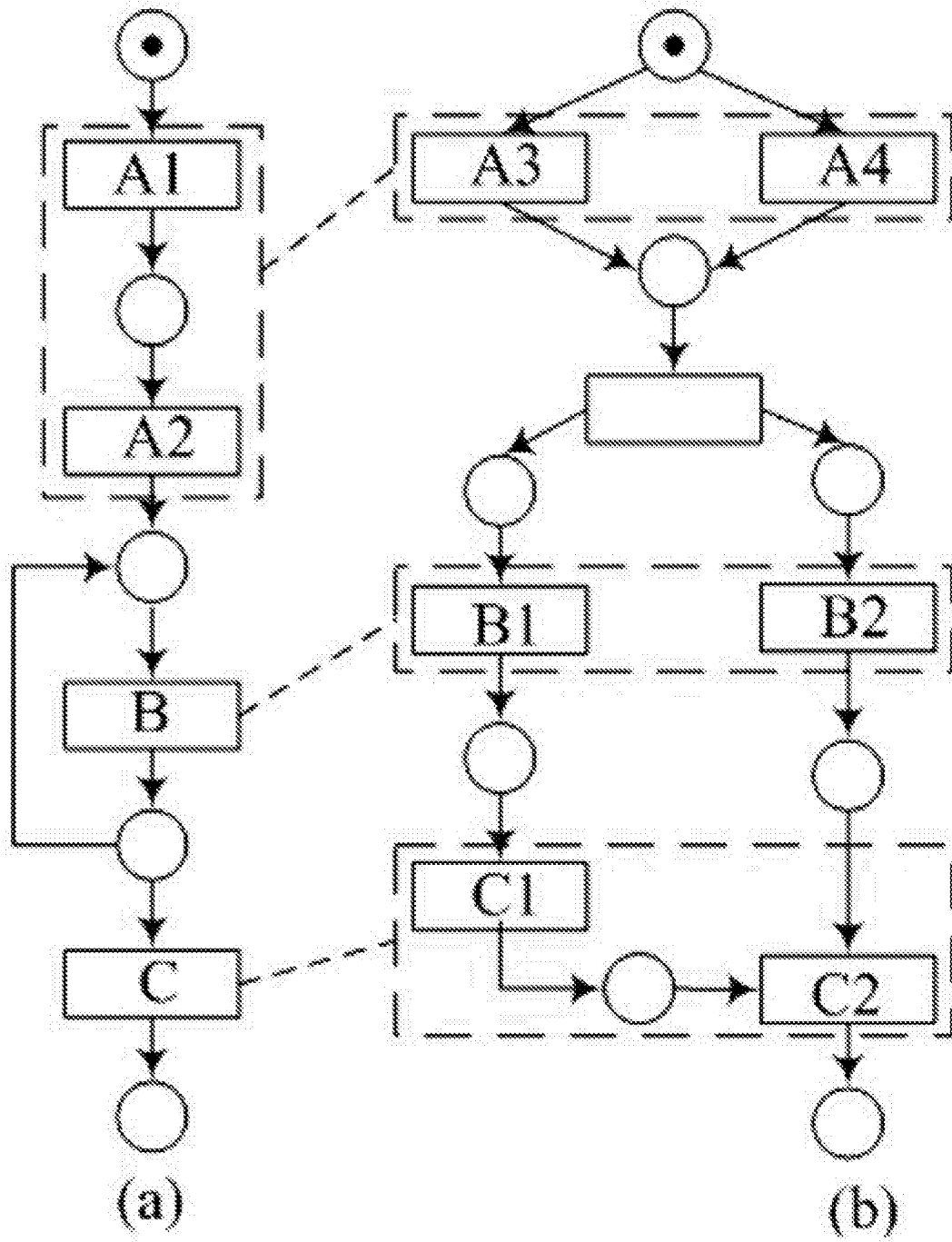


图5

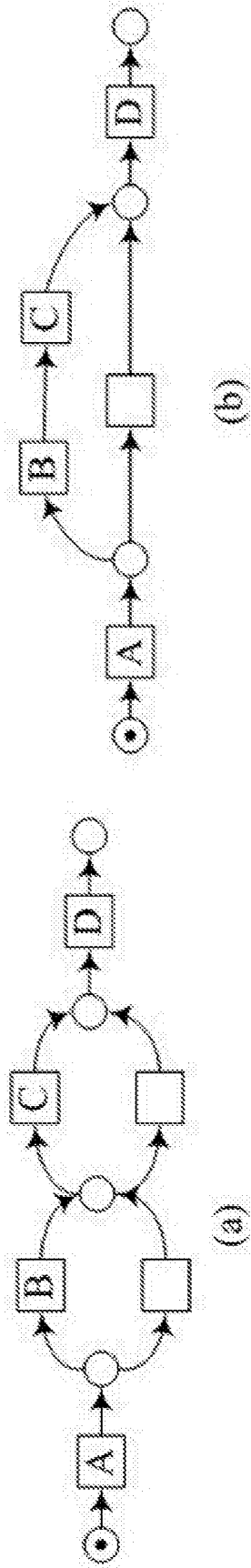


图6

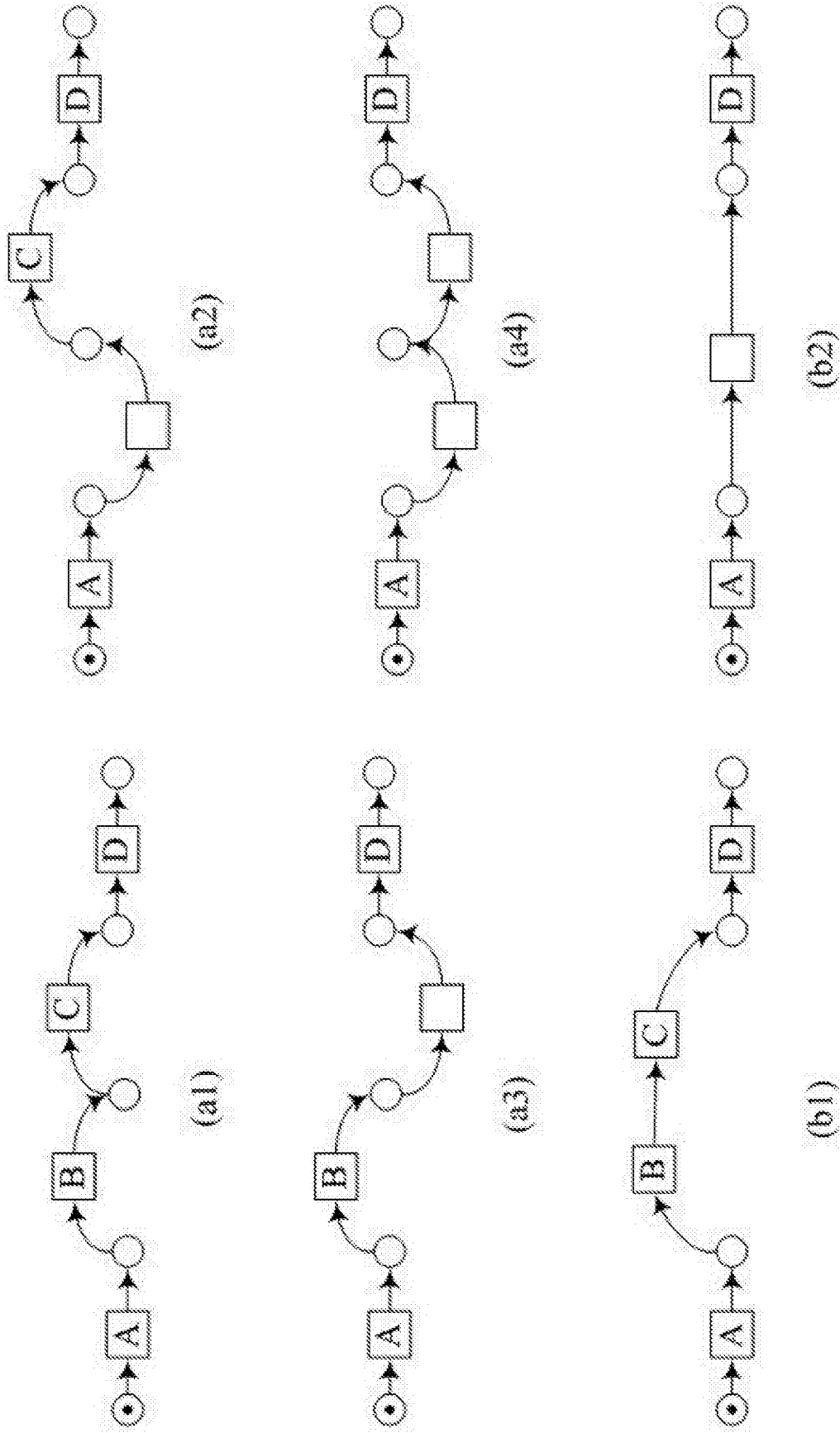


图7