一种工作流模型的模拟与分析

李 敏,曹宝香,郑 垒
(曲阜师范大学 计算机科学学院,山东 日照 276826)

摘 要:基于BPEL的工作流技术,在企业流程的管理上应用越来越广泛。为基于BPEL的工作流模型的模拟和分析提出一个框架,并给出检查用BPEL实现的工作流程正确性的方法。讨论问题主要三个:(1)如何使一个用BPEL语言实现的工作流模型可以转化为数据流网络模型;(2)如何能潜在地把不正确执行路径纳入;(3)如何用SPIN能将工作流的性能形式化地验证出来。为了实现从工作流到分析模型转变的步骤,使用了图形转变,实现分两个步骤实施工作流程-PROMELA转型,使每一个较小的一步都在抽象的水平。此验证方法方便于模型设计,而且对于验证在一个已经制定的业务流程中的小变化的执行情况也会有一定帮助。

关键词:BPEL;工作流;验证;数据流网络;故障模拟
中图分类号:TP311.5 **文献标识码**:A **文章编号**:1673-629X(2009)06-0131-04

Simulation and Analysis of a Workflow Model

LI Min, CAO Bao-xiang, ZHENG Lei
(Computer Science College, Qufu Normal University, Rizhao 276826, China)

Abstract: BPEL-based workflow technique is widely used on the aspect of the enterprise process management. Present a framework for the simulation and formal analysis of workflow models. Then, a method to check correctness properties of workflows implemented in BPEL is proposed. Discuss: (1) How a workflow model, implemented in the BPEL language, can be transformed into a dataflow network model; (2) How potentially incorrect execution paths can be incorporated; (3) How the properties of a workflow can be formally verified using the SPIN model checker. For the several model transformation steps from workflow to analysis models, use graph transformations. This allows us to implement the workflow-PROMELA transformation in two steps, each a smaller step in abstraction level. The verification method is meant to be used at design time but could also be helpful to verify the implementation of a small change in an already enacted business process.

Key words: BPEL; workflow; verification; dataflow networks; fault simulation

0 引 言

工作流是一类能够完全或部分自动执行的业务过程,根据一系列过程规则,文档、信息或任务能够在不同的执行者之间传递、执行。

目前办事处拥有重要的IT基础设施,经常使用电脑辅助业务流程的协调以提高效率和生产力。有几种语言,允许一个非常高层次的、可执行工作流的描述,比如,BPEL(业务流程执行语言)[1]或XPDL(XML过程定义语言)。工作流程的复杂性与常规编程语言的复杂性有密切联系。因此,新的问题与电子业务流程执行一起出现。

以计算机为基础的工作流的执行涉及到松散连接的信息系统之间的通信,这使得分布式工作流的测试非常困难,因为来自几个不同数据库的数据往往需要方便操纵。此外,在测试阶段产生的事务副作用必须撤销。另一种选择是要建立整个的测试环境,用多个服务器和数据库控制测试数据。这两个解决方案都是耗费时间和资金的。

因而,在工作流程制定之前,有几个语义的要求是必须满足的,比如:

1)必须没有任何的死锁。万一事件陷入死锁,工作流的执行将停顿下来。

2)所有的活动都必须是可获得的。如果活动无法获得,可能在该公司有闲置未用过的、不太可取的的资源。

3)每个变量在读之前必须要先被写入。读一个未初始化的变量可能会导致一个不可预测的结果。

收稿日期:2008-09-18;**修回日期**:2008-12-17
基金项目:山东省科技计划项目(2006GG2301001)
作者简介:李 敏(1979-),女,山东曲阜人,硕士研究生,研究方向为企业信息化;曹宝香,硕士生导师,教授,研究方向为企业信息化、图形图像、数据库技术、计算机控制、计算机辅助设计。

框架最重要的功能是，有检查上述要求中第三条的能力。还有其他的解决办法来核实前两条[2]。

文中提出一种方法用来从形式上证实一个 BPEL 模型具有上述的性能。例如，图 1 说明了工作流的高层次描述，如 BPEL，需要转化成低水平的数学符号它就可以自动验证。为此在该方法中选择了数据流网络。

数据流网络的形式是指用完全界定的语义[3]模拟复杂的、分布式计算系统。这个形式化的抽象水平，是介于 BPEL 模型和用 PROMELA 实现的验证模型之间的。它结合了基于有限状态自动机描述的状态和 Petri 网[3]的数据(令牌)流动。用数据流网络模拟的更多优势是，在附加规则和节点的状态下故障仿真也可以被执行。

工作流可以按如上述的要求核查。对于用数据流网络表示工作流的 PROMELA (Protocol Meta - Language)[4]实现而言，要核查的性能是需要确切地阐述为线性时序逻辑表达式(LTL，Linear Temporal Logic)。由 SPIN(Simple Promela Interpreter)[5]模型检查器来评价这些 LTL 表达式。

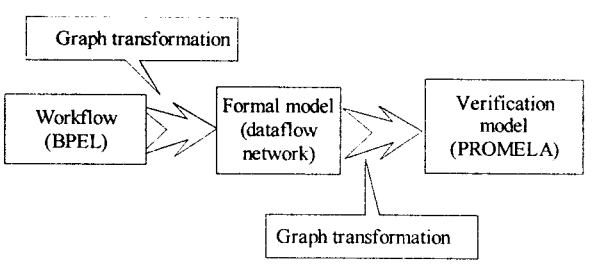

图 1 工作流分析方法

1 从工作流模式到数据流网络

1.1 工作流

工作流程描述语言非常类似于常规的结构化编程语言。工作流程的结构组件是顺序、选择、迭代和并行执行结构。

活动有输入和输出参数。这些数据元素被称为消息，它在运行于分布式环境下的不同计算机之间传递。

在 BPEL 中数据是由使用中的变量支持。变量的值可以作为消息被发送和接收，并且控制流也由它们决定。对变量的操作被称为数据处理能力。

假定要检查的工作流程在 BPEL 中只是用语言的一个子集实现的。转变处理了像那些在图 2 中所示的 BPEL 中基本的、结构化的活动[6]和数据处理，但各种事件处理是被忽略的。

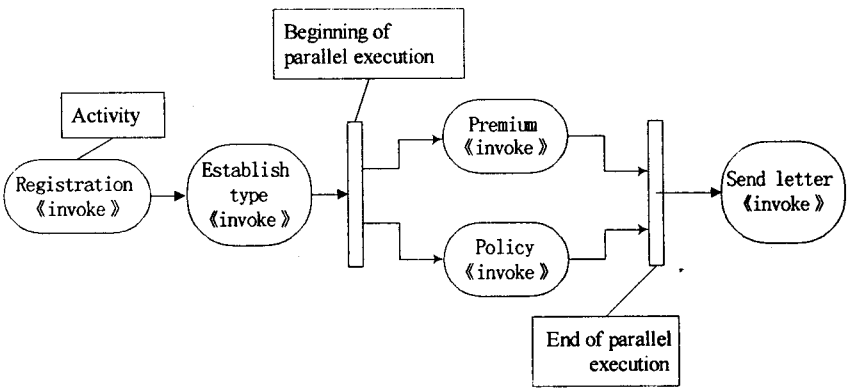

图 2 工作流程的概念和相应的 BPEL 关键词

图 2 中显示了一个保险公司的工作流程的一部分。首先，客户报告记录在案的损害；然后确立损害的类型；接下来保险业公司必须决定是否要进行赔偿损害。这需要两个独立的、可以同时执行的活动；最后向客户发送一封包含有关该决定的消息的信函。

1.2 数据流网络

数据流网络[7]被设计成模拟分布式通信系统。一个数据流网络是由信道互联的数据处理节点所组成。信道在节点之间传送令牌。一个信道不能包含令牌队列。对比原来的数据流网络的形式，规定包含这些令牌队列的端口。端口连接着节点和信道，令牌保留在输入端口，直至激发应用时它不被移走。一个端口有可能包含无数个令牌。令牌是一个用其颜色所代表的原子抽象数据单元。

每个节点是一个有限状态自动机有状态和状态转换规则。规则由两部分构成。第一部分界定了激发条件，第二部分声明了防止激发所要执行的活动。在转变期间，该节点根据条件移动令牌，改变状态，并将若干令牌放到输出端口。

1.3 映射工作流程到数据流网络

在数据流网络模型中，控制流程是用特别颜色特征来描绘：控制令牌。控制令牌的数目与在网络中并行执行的活动数目相符。

执行秩序的活动可以由结构化的活动来界定：顺序，选择，迭代和并行执行。在数据流网络模型中它们至少由两个节点表示。开头的一个节点分配控制令牌，最后一个节点收集它们以表示分支的同步。执行选择时，控制结构需要额外节点来表示条件表达式的赋值。

把转换规则应用到图 2 中的工作流程，得出图 3 表示的一个数据流网络。所用规则的归纳见表 1。

转型并不保留变量的具体值。这里只使用一个抽象的状态空间检查变量初始化、读取和写入的顺序。

2　故障仿真

工作流程执行引擎能够协调很多人的工作,并且通常连接到独立机构的多种多样的计算机上。这些电脑松散地连接也不能保证它们的可用性。因此,必须加以考虑故障问题。给工作流程增加一些冗余,然后检查是否达到了预计的容错性。

在控制流程中误差传播[8]是由故障控制令牌制作的。变量的抽象状态(即"可写","写和读",等等)由节点状态来表示。

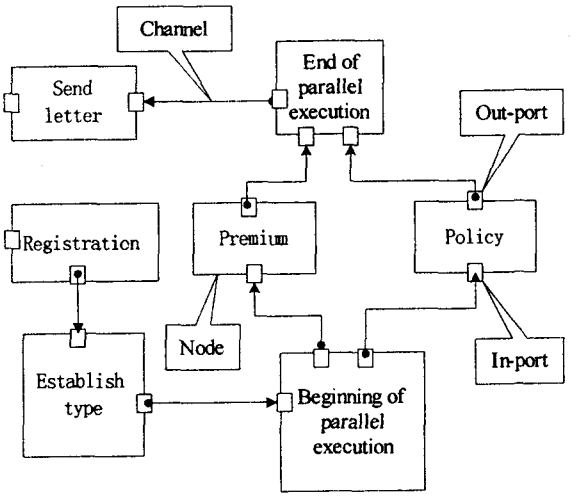

图 3　将图 2 中工作流映射成数据流网络

表 1　转换规则

工作流模式	数据流网络模式
变量	一个有适当的状态矢量空间的节点
基本活动	如果活动有输入和输出参数则有两个节点,否则只有一个
顺序	两个附加节点,一个在开始,一个在结束
并行执行	一个开始节点增加控制流,且结束节点恢复成单个控制流
选择	一个节点在开始,一个在结束。按条件赋值的顺序相互连接着各种情况
选择的一种	表示条件表达式里的每一个变量的一个节点
循环(while)	有两个分支的改进选择:一个使得活动反复执行然后再返回条件,另一个引出工作流里的下一个活动

2.1　模拟数据流错误

在这种情况下,误差仅在变量中传播,但它对控制流程并没有起作用。

需要故障插入活动,即将错误的数据写入它们的输出结果中不考虑它们的输入和收到的控制令牌的风格。假设健壮的节点总是将健壮的数据写给它们输出,除非控制或输入是错误的。这样才能确定一个错误插入变量的错误限制区域的方式。

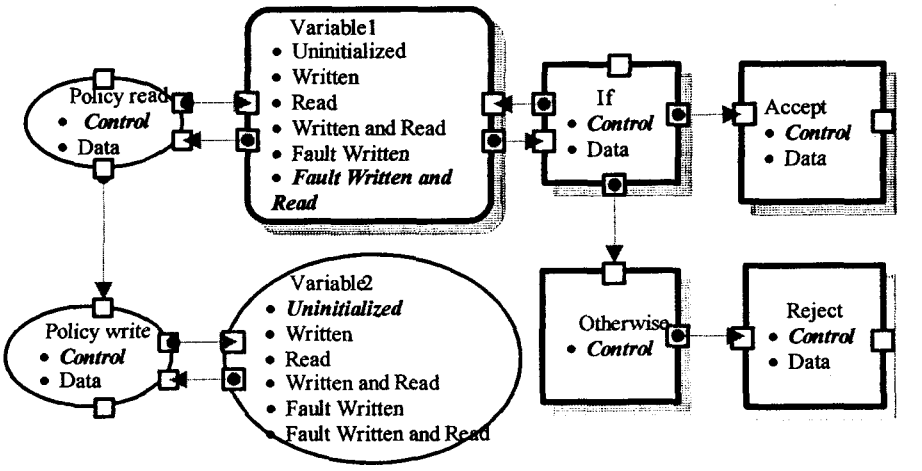

图 4　故障传播

图 4 显示了用数据流网络表示的保险公司工作流程的一部分。变量 1 包含错误的数据。椭圆节点涉及到变数之间的错误传播,正方形节点处理控制流传染。

1)第一步"Policy read"在其输入端口收到一个控制令牌。状态从"控制"改变成"数据",并发出了一个"Reading"读令牌到它的输入变量中。

2)代表输入变量(变量 1)的节点接收到"Reading"读令牌。假设变量包含着错误的数据,即输入变量的状态不是"Fault Written"就是"Fault Written and Read"。状态被切换到"Fault Written and Read"并且将错误控制令牌放置在输出端口。

3)从变量 1 收到故障控制令牌,状态从"数据"切换到"控制",并将"Fault Reading"令牌放在接近于活动的第二个节点的输出端口。

4)"Policy write"从其控制端口收到"Fault Reading"令牌,状态从"控制"切换到"数据",并发出了一个"Fault Reading"令牌给其输出变量。

5)"变量 2"切换到"Fault Written"状态并且传给活动的第二个节点一个控制令牌。

6)第二个活动节点从输出变量得到控制令牌,将状态切换到"控制"状态,并发出了一个控制令牌给其输出端口。

2.2　控制流错误的模拟

如果一个分支的条件使用了错误的变量,错误的条件变量会影响到控制流程。图 4 中的"IF"节点说明了这一点。

1)"IF"节点是分支结构中所有节点的一部分。它收到控制令牌,接着把它的控制状态改变到数据状态并且发送读令牌给"变量 1"节点。

2)一个变量节点的状态从"Fault Written"改变到"Fault Written and Read",并且将错误控制令牌送回到

选择结构节点。

3)“IF”节点接收到错误控制令牌。它不确定是将令牌传递给“Accept”活动还是传递向下一个条件。在这一点上控制令牌是变成错误控制,所有接收错误控制令牌的活动写下错误的数据,并继续传递错误控制令牌。这里假设错误控制令牌发送到“Accept”活动。

4)错误控制令牌由“Accept”接收,写入的错误数据同在 2.1 节展示的一样,并且错误控制令牌是继续传递。

3 工作流模型的认证

同在结构化编程语言中一样,在 PROMELA 中也可以使用变量和子程序。变量的类型是整数的子集以保证该计划有限的状态空间。

PROMELA 提供了一个另外的变量类型, FIFO 信道。该方式下,数据流网络的信道并没得到执行。

SPIN 是能够对 PROMELA 程序做全面的状态空间检查,在 LTL 表达式的形式下评价系统需求。这种方式下,一个数据流网络的动态特性可以得到证实。

这些转变可以保持性能,在某种意义上说,BPEL 流程的每一个执行路径在 PROMELA 程序里也可以被发现。如果性能有效,就 PROMELA 模型而言它也能为 BPEL 流程保持。

3.1 从数据流网络模型到 PROMELA 执行

将数据流网络的结构映射到 PROMELA 语言模式。

* 数据流网络中的信道映射成 PROMELA 语言的信道。由于一个信道总是最多只存在一个令牌,所以是可以实现的。

* 令牌映射成象征性的常数。

* 一个节点的状态变量都映射到全局的整数类型的变量。

* 每个节点映射到一个 Proctype 结构。初始状态由第一次的操作设定。状态的转换规则由一个无限循环实现,包括操作的有条件的原子序列这样的规则。

* 节点的一种状态转变规则映射到操作的一个原子序列。

选择 SPIN 作为验证工具主要是因为 SPIN 应用了很多高效的状态空间搜索技术,大大提高了搜索效率,SPIN 的建模语言 PROMELA 拥有基本数据类型和各种语言结构,这使得 PROMELA 语言的可读性和表达能力都非常强,它的建模方式是以进程为单位,进程异步交互运行。进程的基本要素包括赋值语句、条件语句、通讯语句、非确定性选择和循环语句。SPIN 可以随机模拟系统的一条执行路径,也可以实时自行选择执行路径。检查到性质不符合时,即得到反例时,SPIN 可以给出这个反例的执行路径,以便研究者精确发现问题所在。

3.2 证 明

在图 3 中,保险费和保险单两个活动同时执行。一个关键的问题是,一个活动的故障是否影响其他的活动。这个要求需要阐述成 LTL 公式,该公式里采取的逻辑的变量必须定义成 C 风格定义宏的形式。用 PROMELA 资源和逻辑变量定义 SPIN 来评估 LTL 公式。

例 1 图 3 描述的模型中,要求“Policy out”活动的错误不应对“Policy in”有影响。

下面说明如何确切地阐述这一要求。

```
# define Policy_out state_Policy_output! = Fault_written
# define Premium_in state_Premium_input! = Fault_written
G(! Policy_out -> Premium_in)
```

前两行界定逻辑变量的意义,决定了第三行逻辑表达式的值。变量“Policy_out ”为真,当且仅当 PROMELA 变量“state_Policy_output”的值是“Fault_written”时。“Premium_in”的定义与此类似 。该要求的 LTL 公式表示在第三行。它指出,在任何状态下沿离散的时间线当“Policy_out”出保单为假时,“Premium_in”入保险费为真。如例 1 所示公式的赋值,可能会导致正面和负面的答案。正面的结果保证该 PROMELA 模型符合要求。负面的可能是 BPEL 过程的一个好的测试案例。但是由于 BPEL 变量的具体的值,过程可能无法运行到错误执行路径。

4 结束语

提出了一个方法来检查用 BPEL 实现的工作流程的正确性。数据流网络是用来定义工作流程的形式语义。BPEL 模型映射到数据流网络,数据流网络是映射到 PROMELA 模型。这些模型的转变,创建了工作流的数据流网络模型并产生了 PROMELA 代码。每一个 BPEL 语言的重要结构都有一个图形转化规则。

以后计划支持从 BPEL 域制定的必要条件自动生成 LTL 表达式。此外,SPIN 为防止负面结果给出反例的执行路径亦是未来的目标。

参考文献:

[1] 于守健,李卫民,吴国文,等.BPEL 中基于有限状态自动机的 Web 服务自动组合[J].小型微型计算机系统,2007,28(4):742-747.

(下转第 137 页)

$\mathrm{sinc}(x)$,然后在因变量的前 10 个样本中加入噪声 $N(0,0.8)$,其余样本中加入噪声 $N(0,0.1)$,如图 1 所示。分别用标准的支持向量回归和加权支持向量回归建立预测模型。

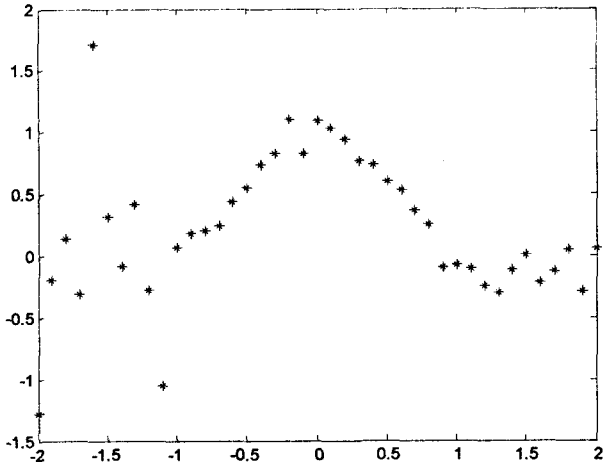

图 1　加入噪声的样本

首先对输入样本进行训练,取 $c = 1/15, \sigma^2 = 0.12$,利用线性规划得到 $\rho = 0.0931, f_{\max} = 0.1377, f_{\min} = 2.8616e-008$,则确定权值为:

$$\mu_i = \begin{cases} (1 - \dfrac{0.1377 - f(x_i)}{0.1377 - 2.8616e-008})^2 + 0.0005 \\ 2.8616e-008 \leqslant f(x_i) < 0.0931 \\ 1 - \dfrac{0.1377 - f(x_i)}{0.1377 - 2.8616e-008} \\ 0.0931 \leqslant f(x_i) \leqslant 0.1377 \end{cases} \tag{21}$$

进行 ε - SVR 训练,取 $c = 5, \sigma^2 = 1.8$,在 -2 到 2 之间取间隔为 0.13 的样本为测试样本(不同于训练样本),测试指标采用均方误差:

$$\mathrm{MSE} = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(y_i - \hat{y}_i)^2} \tag{22}$$

其中,y_i 为实际值,$\hat{y}_i$ 为预测值,k 为测试样本的数量。

标准支持向量回归的结果 MSE = 0.3530,加权支持向量回归的结果 MSE = 0.0858,说明当数据中存在噪声时,加权支持向量回归算法得到的预测值更接近真实值。

4　结束语

标准的支持向量回归在样本中无噪声点时,具有很好的学习和泛化能力。但当样本中存在噪声和野点时,回归间隔偏向噪声点移动,从而出现过拟合现象,产生误差。

文中通过引入权值,减小了噪声点的影响。在确定权值时,考虑了样本离决策超平面的距离,对不同的样本采用不同的公式计算其权值。实验表明,该方法与标准支持向量回归相比,减小了回归误差,提高了支持向量机的抗噪能力。

参考文献:

[1] Vapnik V. The nature of statistical learning theory[M]. New York: Springer - Verlag, 1995.

[2] 张　楠,范玉妹.关于支持向量机几何算法的研究[J].计算机技术与发展,2007,17(1):142 - 144.

[3] 张学工.统计学习理论的本质[M].北京:清华大学出版社, 2000.

[4] 杨　斌,陆　游.基于统计学习理论的支持向量机的分类方法[J].计算机技术与发展,2006,16(11):56 - 58.

[5] Lin C F, Wan S D. Fuzzy support vector machine[J]. IEEE Tran. on Neural Network, 2002, 13(2):464 - 471.

[6] Huang H P, Lin Y H. Fuzzy support vector machine for pattern recognition and data mining[J]. Int'l Journal of Fuzzy Systems, 2002, 4(3):826 - 835.

[7] 张　英,苏宏业,楮　建.基于数据域描述的模糊支持向量机回归[J].信息与控制,2005,34(1):1 - 6.

[8] 张讲社,郭　高.加权稳健支撑向量回归方法[J].计算机学报,2005,28(7):1171 - 1177.

[9] Bao YuKun, Zhang Rui. Fuzzy support vector machines regression for business forecasting: An application[J]. Computer science, 2006(9):1313 - 1317.

[10] 孙德山.支持向量机分类与回归方法研究[D].长沙:中南大学,2004.

(上接第 134 页)

[2] van der Aalst W, van Hee K. Workflow Management[M]. [s.l.]: The MIT Press, 2002.

[3] 周建涛,史美林,叶新铭.一种基于 Petri 网化简的工作流过程语义验证方法[J].软件学报,2005,16(7):1242 - 1251.

[4] 雷丽晖,段振华.一种基于扩展有限自动机验证组合 Web 服务的方法[J].Journal of Software, 2007,18(12):2980 - 2990.

[5] 单卓为,鱼　滨.基于 SPIN 的 CSCW 系统的验证[J].计算机技术与发展,2008,18(4):9 - 12.

[6] 范玉顺.工作流管理技术基础[M].北京:清华大学出版社,2001.

[7] Anderson A J. Data Flow Systems, In Multiple Processing: A System's Overview[M]. [s.l.]: [s.n.], 1989:441 - 488.

[8] Laprie J, Randell B, Landwehr G. Basic Concepts and Taxonomy of Dependable and Secure Computing[J]. IEEE Transactions on Dependable and Secure Computing, 2004, 1(1):11 - 33.