• 人工智能 •

有限自动机在复合事件检测中的应用

袁爱平， 傅 明

(长沙理工大学 计算机与通信工程学院，湖南 长沙 410076)

摘 要：针对早期系统只提供原子事件的检测机制，不能检测由原子事件组成的复合事件的问题，提出了用有限自动机来检测复合事件的方法。说明了复合事件的组成和表达式，利用自动机原理对复合事件的检测模式进行了分析，给出了复合事件检测的具体过程：从事件表达式到不确定的有限自动机，从不确定的有限自动机到最小化确定的有限自动机，再用程序实现了确定的有限自动机。实例表明，自动机模型是检测复合事件的一种有效实现方式。

关键词：有限自动机；复合事件；检测；转化；直接代码生成

中图法分类号：TP301.1 **文献标识码**：A **文章编号**：1000-7024 (2009) 14-3393-03

Application of finite automaton in detection of composite event

YUAN Ai-ping, FU Ming

(College of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410076, China)

Abstract: Aiming at the problem that early system only provides atomic event detection and cannot detect composite event composed with atomic event, the model of using automaton to detect composite event is presented. A introduction to component and expression of composite event is provided. Detection model of composite event with automaton theory is analyzed, and the process of composite event detection is shown: From event expression to non-decided finite automaton, and from non-decided finite automaton to minimized decided finite automaton, then its realization of decided finite automaton by programming. Through instance illuminates that the model of automaton is one of effective modes to detect composite event.

Key words: finite automaton; composite event; detection; transform; code programming directly

0 引 言

很多系统需要检测两个或多个源事件，这些源事件我们称为原子事件或子事件，这些原子事件有可能互相依赖，当一个原子事件发生时，会导致另一个原子事件的发生，而在我们的系统处理中，当有多个互相依赖的原子事件发生时，只认为是某一个复合事件的发生。可以看出，复合事件是一种更高级别的事件，它由两个或多个原子事件组成。为了能正确的检测复合事件，本文提出了一种用有限自动机来检测复合事件的方法。定义复合事件的各个原子事件为有限自动机中的状态，当一系列状态发生时，如状态迁移到最终状态，则认为复合事件发生了。

1 复合事件

1.1 事件的复合操作

事件的复合操作定义如下：

(1)AND：E1 AND E2 表示 E1 和 E2 都发生；

(2)OR：E1 OR E2 表示 E1 和 E2 至少有一个发生；

(3)SEQ：E1 SEQ E2 表示 E1 和 E2 顺序发生；

(4)NOT：NOT E1 表示 E1 不发生；

(5)CLO：CLO(E1,K)表示 E1 重复发生 K 次。

1.2 复合事件表达式

每一个复合事件可以用一个“事件表达式”来表示，事件表达式的语义描述如下：

C_EVENT ::= <E_EXPR>；

E_EXPR ::= (<P_EVENT>|<E_EXPR>)<E_OP><E_EXPR>；

P_EVENT ::= <OB_EVENT>|<TR_EVENT>|<TM_EVENT>|<EX_EVENT>；

E_OP ::= AND|OR|SEQ|NOT|CLO；

上述定义中，OB_EVENT 表示对象事件，TR_EVENT 表示事务事件，TM_EVENT 表示时间事件，EX_EVENT 表示外部事件。

2 复合事件的检测

定义 1 事件表达式的自动机模型：由于事件表达式与正

收稿日期：2008-07-16；修订日期：2009-02-23。

基金项目：湖南省自然科学基金项目 (07JJ3120)。

作者简介：袁爱平 (1975－)，男，湖南邵阳人，硕士研究生，研究方向为嵌入式系统； 傅明 (1961－)，男，湖南汨罗人，博士，教授，硕士生导师，研究方向为计算机网络应用、知识工程。E-mail：yuan_ai_ping@21cn.com

规表达式的一致性,总可以构造一个FA(finite automaton)来接受某个事件表达式,这样的FA,称之为事件表达式的自动机模型。

复合事件的检测主要是构建事件表达式对应的自动机模型,可以分成4个步骤完成:

(1)确定事件的表达式;

(2)根据事件的表达式构造出一个带ε动作的NFA(non-decided finite automaton);

(3)将NFA转化成DFA(decided finite automaton),并最小化;

(4)根据最小化的DFA写出事件的检测程序。

如下以复合事件表达式E = (e1 OR e2)AND e3为例,给出其检测的过程。此事件的语义为当事件e1或e2发生后且e3发生,则复合事件E发生。

2.1 从事件表达式到NFA

由于事件表达式和正则表达式的一致性,可以直接将Thompson's construction方法应用于事件表达式,构造出相应的NFA。Thompson's construction方法的基本思想是由简单的正则表达式的NFA,通过固定的规则来构造复杂的正则表达式的NFA.。Thompson's construction方法参见文献[1]。

对于复合事件表达式E = (e1 OR e2)AND e3,首先为子事件e1、e2和e3构造NFA,如图1所示。

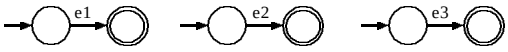

图1 e1、e2、e3各自对应的NFA

接着用选择结构为子表达式(e1 OR e2)构建NFA,如图2所示。

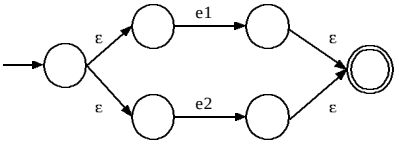

图2 子表达式(e1 OR e2)对应的NFA

然后用子表达式(e1 OR e2)和e3的NFA构造复合事件E的NFA,如图3所示。

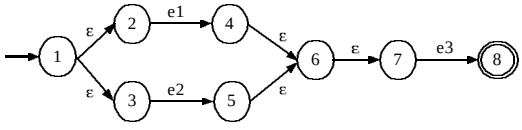

图3 表达式(e1 OR e2)AND e3对应的NFA

2.2 NFA转化成DFA

定义2 状态q的ε-闭包:对所给状态q,仅用ε转换便可到达的状态集合P称为q的ε-闭包,记为:ε-closure(q)。

定义3 状态子集I的ε-闭包:状态子集I中任何状态经过任意条ε弧到达的状态集合,记为:ε-closure(I)。显然:ε-closure(I) = $\bigcup_{q\in I}$ε-closure(q)。

定义4 状态集合I的a弧转换:状态集合I中任何状态经过一条a弧到达的状态集合,记为:Move(I,a)。

把NFA转化成DFA,主要是删除自动机中的不可到达状态和消除不确定性。下面采用子集法实现NFA到DFA的转化。

算法1 用子集法把NFA转换成DFA

(1)[初始化]

根据状态q初始化子集族C,令{ε-closure(q)}为C中的惟一未被标记的成员;

(2)[构造子集族C]

循环 当子集族C中存在未被标记的成员T,反复执行

1)标记T;

2)循环 对每个输入a,反复执行

(a)U = ε-closure(Move(T,a));

(b)若U不在C中,

则将U作为未被标记的成员加入C中

对于图3,表1用矩阵来模拟NFA到DFA的子集法转化过程。将表1状态转化矩阵中的状态重新标记得到表2矩阵表达式。

表1 状态转化矩阵

C	e1	e2	e3
1,2,3	4,6,7	5,6,7	
4,6,7			8
5,6,7			8
8			

表2 重新标记的状态转化矩阵

C	e1	e2	e3
1	2	3	
2			4
3			4
4			

进而,将表2用状态转化图得到图4。图4与图3比较可知:状态数明显减少,并且消除了不确定性。

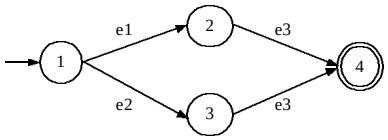

图4 表达式(e1 OR e2)AND e3对应的DFA

2.3 最小化DFA

算法2 最小化DFA[2]

(1)[初始化]

首先把状态集I分成初态集、非初非终态集和终态集3个子集。

(2)[分割状态集]

循环 反复执行下列语句

1)循环 对于每个输入a_i∈T,反复执行

循环 对于I中的子集I_i,反复执行

若子集I_i经过a_i到达的状态分别落于I中P个不同的子集,

则子集I_i被分成P个更小的子集I_{i1},I_{i2},…,I_{ip}。

2)如子集数不再增加则算法结束

根据算法2,最小化图4中表达式(e1 OR e2)AND e3对应的DFA:

(1)由于状态 1 是初态，状态 4 是终态，则先分成 3 个子集 I_1：{1}，I_2：{2，3}，I_3：{4}；

(2)对于子集 I_1 和 I_3，只含有一个状态，不能再细分。而对于子集 I_2，只能接受 e3，接受后状态转化成{4}∈I_3。很显然，子集 I_2 也不能再细分。

由此得到了下面的最小化 DFA，如图 5 所示。

重新标记图 5 中的最小化 DFA 状态，得到图 6。

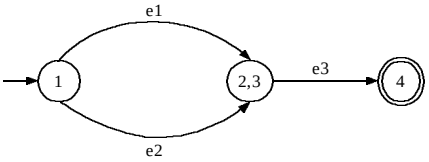

图 5 表达式(e1 OR e2) AND e3 对应的最小化 DFA

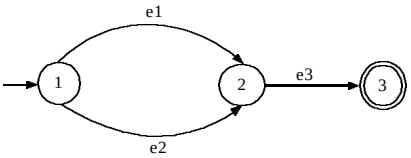

图 6 重新标记的表达式(e1 OR e2) AND e3 对应的最小化 DFA

2.4 程序实现

DFA 向程序的转化有两种代码生成算法：直接代码生成算法和表驱动代码生成算法。下面给出直接代码生成算法[3]的描述。

算法 3 图 6 中 DFA 的直接代码实现

算法描述：该算法实现了对事件表达式 E = (e1 OR e2) AND e3 的检测。变量 status 存储当前状态的值。当有某个事件发生时，算法根据所发生的事件改变状态变量 status 的值。当到达终止状态时，即 status = 3 时，表达式 E 被满足，则通知系统复合事件 E 发生。

(1)[初始化]

status = 1；

(2)[检测复合事件]

循环 当 status = 1 或 status = 2 时，反复执行

若发生的事件是 e1 或 e2 时，

则 status = 2；

否则若发生的事件是 e3 时，

则 status = 3；

(3)[判断复合事件是否发生]

若 status = 3，

则通知系统 E 发生；

否则给出出错信息；

3 结束语

本文讨论了有限自动机在复合事件检测中的应用，说明了事件的复合操作，给出了复合事件的表达式定义，描述了从事件表达式到 DFA 的完成过程，最后用程序实现了 DFA。在实际应用中，我们可以采用多种方法检测复合事件，如基于 Petri 网、基于匹配树、基于图等方法，这些方法都能达到检测复合事件的目的，具体采用哪种检测方式，需根据具体问题而定。

参考文献：

[1] Dayal U,Hsu M,Ladin R.A transaction model for long running activities[C].Barcelona,Spain:Proc of the 17th Int'l Conf Very-Large Databases,1991:113-122.

[2] 徐红.对确定有限自动机最小化算法的改进[J].桂林航天工业高等专科学校学报,2005(4):14-16.

[3] 周涛.基于有限状态自动机的复合事件检测的程序实现[J].计算机工程,2005,31(23):85-86.

[4] 郑震坤,张阔,王小鸽.支持复合事件的模型及其在中间件中的应用[J].计算机工程,2006,32(13):52-54.

[5] 张菊芳,魏峻.复合事件检测技术的综述与评价[J].计算机应用研究,2005,22(10):1-4.

[6] 周涛.基于有限状态自动机的复合事件监测模型[J].微电子学与计算机,2007,24(7):180-182.

[7] Pietzuch P R, Shand B, Bacon J. Composite event detection as generic middleware extension [J]. IEEE Trans on Network, 2004,18(1):44-55.

[8] 肖兵,卢炎生,瞿彬彬.ECA 规则中复合事件的 CCPN 建模与检测[J].计算机工程与设计,2005,26(9):2297-2299.

[9] 李钢,吴燎原,张仁斌,等.基于有限自动机的模式匹配算法及其应用研究[J].系统仿真学报,2007,19(12):2772-2775.

(上接第 3392 页)

参考文献：

[1] Zhuo Ling,Wang Cho-Li,Lau FCM.Load balancing in distributed web server systems with partial document replication [C]. Proceedings of International Conference on Parallel Processing. Washington:IEEE Computer Society,2002:305-312.

[2] Narendran B, Rangarajan S, Yajnik S. Data distribution algorithms for load balanced fault-tolerant Webaccess[C].Proceedings of the 16th Symposium on Reliable Distributed Systems.Washingtion:IEEE Computer Society,1997:97-106.??

[3] Su Zhong,Yang Qiang,Zhang Hong-Jiang,et al.Correlation-based document clustering using web logs[C].Proceedings of the 34th Annual Hawaii International Conference on System Sciences.Washington:IEEE Computer Society,2001:5022-5028.

[4] Marco Dorigo,Thomas Stutzle.蚁群优化 Ant Colony Optimization[M].张军,胡晓敏,罗旭耀,等译.北京:清华大学出版社,2007.

[5] 谈文芳,赵强,余胜阳,等.改进粒子群优化算法求解任务指派问题[J].计算机应用,2007,27(12):2892-2895.

[6] 黄茹. 一种解决指派问题的蚁群算法 [J]. 西安邮电学院学报,2006,11(3):106-109.

[7] 殷人昆,吴阳,张晶炜.蚁群算法解决指派问题的研究和应用[J].计算机工程与科学,2008,30(4):43-45.