



(12) 发明专利

(10) 授权公告号 CN 113255295 B

(45) 授权公告日 2024. 04. 09

(21) 申请号 202110457578.X

G06F 40/30 (2020.01)

(22) 申请日 2021.04.27

(56) 对比文件

(65) 同一申请的已公布的文献号

申请公布号 CN 113255295 A

CN 102663190 A, 2012.09.12

CN 110705316 A, 2020.01.17

CN 111581953 A, 2020.08.25

(43) 申请公布日 2021.08.13

CN 111767739 A, 2020.10.13

(73) 专利权人 西安电子科技大学

JP 2017111411 A, 2017.06.22

地址 710071 陕西省西安市太白南路2号西安电子科技大学

JP H0744560 A, 1995.02.14

KR 20120079930 A, 2012.07.16

(72) 发明人 王小兵 李戈 李春奕 段振华
赵亮 田聪 张南

US 5966686 A, 1999.10.12

吴保民; 郭永辉; 王炳锡. 英汉机器翻译中基于规则的句子结构分析与转换. 信息工程大学学报. 2007, (01), 全文.

郭艳华, 周昌乐. 一种汉语语句依存关系网协同生成方法研究. 杭州电子工业学院学报. 2000, (04), 全文.

(74) 专利代理机构 西安长和专利代理有限公司
61227

专利代理师 何畏

审查员 吴海旋

(51) Int. Cl.

G06F 40/151 (2020.01)

G06F 40/211 (2020.01)

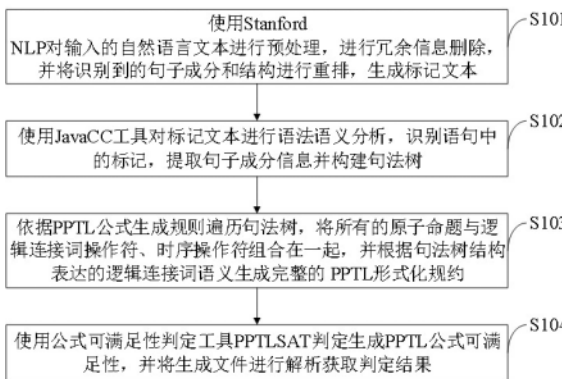
权利要求书2页 说明书10页 附图5页

(54) 发明名称

一种自然语言到PPTL形式化规约自动生成方法及系统

(57) 摘要

本发明属于计算机辅助设计技术领域,公开了一种自然语言到PPTL形式化规约自动生成方法及系统,所述自然语言到PPTL形式化规约自动生成方法包括:利用自然语言处理技术解析自然语言性质文本并生成语法树,遍历该树进行句子成分的提取、重排和标记等预处理操作,并生成标记文本;使用JavaCC工具对标记文本进行语法语义分析,生成含有子句、连接词和时序信息的句法树,遍历句法树生成原子命题及组合PPTL公式;使用PPTLSAT工具判定生成公式的可满足性。本发明能够帮助用户从自然语言描述的性质中提取形式化规约用于模型检测,将用户描述的自然语言文本性质转化为PPTL公式,为普通用户使用模型检测技术提供方便。



1. 一种自然语言到PPTL形式化规约自动生成方法,其特征在于,所述自然语言到PPTL形式化规约自动生成方法包括:

使用StanfordNLP对输入的自然语言文本进行预处理,进行冗余信息删除,并将识别到的句子成分和结构进行重排,生成标记文本;

使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树;

依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的PPTL形式化规约;

使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果;

所述语法语义分析,包括:使用JavaCC工具对标记文本进行语法语义分析,识别标记文本中的标记,提取句子成分信息,包括构成原子命题的简单句成分提取、整体句法结构的提取以及逻辑连接词语义和时序语义的提取等,将以上提取的信息记录并构建句法树;

所述PPTL公式生成过程,包括:遍历句法树,结合PPTL公式生成规范为每个子句生成原子命题;将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的PPTL形式化规约;其中,所述遍历包括:

(1) 若遇到命题节点,根据标记为命题添加时序操作符,并将命题描述信息取出,按照命题变元和命题描述存到一个哈希表中;

(2) 若遇到连接符节点,判断父节点的描述信息,根据描述信息确定逻辑操作符,并将对应命题连接起来;

所述PPTL公式生成的可满足性判定过程,包括:外部调用可满足性判定工具PPTLSAT,将生成公式和公式的非作为输入参数进行公式可满足性判定,然后将工具生成的文件进行解析并提取该工具的判定结果;其中,所述判定过程包括:

(1) 需要判定生成公式P,如果P没有构造出与之对应的LNFG,则表示P是矛盾的和不可满足的;否则P是可满足的,接着执行第(2)步;

(2) 需要判定生成公式的非Q,如果Q没有LNFG,表示Q是不可满足的,而P是一个重言式,判定失败;否则判定P既不是矛盾的,也不是重言式,判定成功。

2. 如权利要求1所述的自然语言到PPTL形式化规约自动生成方法,其特征在于,所述输入的自然语言文本进行预处理,包括:

(1) 使用Stanford NLP处理输入自然语言文本生成语法树;

(2) 遍历语法树提取句子成分,并对句子成分进行重排,去除结构歧义;

(3) 将语法树中的词性标记进行提取并整合到文本中,生成标记文本。

3. 如权利要求2所述的自然语言到PPTL形式化规约自动生成方法,其特征在于,所述句子成分的提取和重排,包括:所述句子成分提取的一般形式:英语句子的主干是主语、谓语和宾语,然后加上修饰句子的副词或介词短语;对于主从复合句的句子结构提取的形式为:一般形式是主从连词、从句、主句的形式,对于并列复合句一般形式是子句、并列连词、子句的形式;

所述句子成分重排的流程为:首先,将每个树结构下的叶子节点,分支最深位置节点作为相应的句子成分,句子成分标记由叶子节点的父节点的节点含义进行定义;然后,对抽取

出的句子成分进一步优化处理,并删除其中的冗余信息;最后,根据句子成分提取定义的句法形式对提取到的句子成分、主从复合句和并列复合句结构进行重排。

4.一种实施权利要求1~3任意一项所述的自然语言到PPTL形式化规约自动生成方法的自然语言到PPTL形式化规约自动生成系统,其特征在于,所述自然语言到PPTL形式化规约自动生成系统包括:

预处理模块,用于将输入的自然语言描述的性质进行预处理操作,将冗余无意义信息删除,去掉对后续文本分析和处理造成干扰的部分,以及句子成分提取定义的句法形式对句子成分和整体主从句结构进行重排;

语法语义分析模块,用于使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树;

PPTL公式生成模块,用于依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的PPTL形式化规约;

PPTL可满足性判定模块,用于使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果。

5.一种计算机设备,其特征在于,所述计算机设备包括存储器和处理器,所述存储器存储有计算机程序,所述计算机程序被所述处理器执行时,使得所述处理器执行权利要求1~3任意一项所述的自然语言到PPTL形式化规约自动生成方法,包括如下步骤:

使用Stanford NLP对输入的自然语言文本进行预处理,删除冗余信息,并将识别到的句子成分和结构进行重排,生成标记文本;使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树;

依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的PPTL形式化规约;使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果。

6.一种计算机可读存储介质,存储有计算机程序,所述计算机程序被处理器执行时,使得所述处理器执行权利要求1~3任意一项所述的自然语言到PPTL形式化规约自动生成方法,包括如下步骤:

使用Stanford NLP对输入的自然语言文本进行预处理,删除冗余信息,并将识别到的句子成分和结构进行重排,生成标记文本;使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树;

依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的PPTL形式化规约;使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果。

7.一种信息数据处理终端,其特征在于,所述信息数据处理终端用于实现如权利要求4所述的自然语言到PPTL形式化规约自动生成系统。

一种自然语言到PPTL形式化规约自动生成方法及系统

技术领域

[0001] 本发明属于计算机辅助设计技术领域,尤其涉及一种自然语言到PPTL形式化规约自动生成方法及系统。

背景技术

[0002] 目前,伴随着计算机软件技术的发展和人工智能时代的来临,以及5G和物联网(IOT)技术的提出,社会发展正朝向万物互联、互通的方向迈进。随之而来的是万物互联体系无声的融入人们的日常生活,未来被链接到万物互联的系统的个体将会越来越多。从国防、航空、航天等安全关键系统领域到工业生产和个人生活领域,计算机软件已经悄然渗透到社会的方方面面。随着计算机软件在各行业中越来越不可或缺,软件因设计存在的漏洞引发的系统崩溃或者遭受的损失与国家安全和个人隐私、财产和生命安全的联系也日渐紧密,如何解决软件因漏洞和错误导致的可靠性与安全性缺失被视为学术界和工业界亟待解决的热点问题。

[0003] 形式化方法是一种提高软硬件系统安全性与可靠性的方法,它使用形式化的语言来描述软硬件系统的需求和特征,并且通过严格的数学推理验证来保证最终的产品满足这些需求和具备这些特征。模型检测是形式化方法中的一种重要方法,其作为一类面向有穷状态系统进行自动验证的形式化技术,原理是借助对待验证模型的状态空间进行搜索以此验证该模型对于期望性质能否满足。

[0004] 在模型检测中,定义待验证性质的形式化规约对用户来说难度较大,普通用户不具备形式化方法的知识背景,更倾向于用自然语言来描述系统的待验证性质。为了定义出符合形式语法与期望语义的性质规约公式,用户需要经过长期的学习和严格训练,这限制了模型检测技术在各行业软硬件系统安全性验证上的应用。

[0005] 通过上述分析,现有技术存在的问题及缺陷为:

[0006] (1)在模型检测中,定义待验证性质的形式化规约对用户来说难度较大,普通用户不具备形式化方法的知识背景,更倾向于用自然语言来描述系统的待验证性质。

[0007] (2)为了定义出符合形式语法与期望语义的性质规约公式,用户需要经过长期的学习和严格训练,这限制了模型检测技术在各行业软硬件系统安全性验证上的应用。

[0008] 解决以上问题及缺陷的难度为:以上两个问题的难度主要在于用户描述文本的语法结构多变且可能存在歧义,若不对用户输入文本的语法结构进行约束,自然语言灵活的语法结构会导致语法树出现结构歧义错误,最终导致生成错误的PPTL形式化规约。另一方面,PPTL形式化规约是将原子命题和各种逻辑连接词操作符、时序操作符组合而构成,但是PPTL中逻辑连接词语义和时序操作符语义与自然语言文本所表达语义的关系不是简单的一一对应。

[0009] 解决以上问题及缺陷的意义为:解除了普通用户使用模型检测对软硬件系统验证时定义形式化规约时的专业限制,扩大模型检测技术在各行业中的应用,提高软件开发过程中的可靠性与安全性。

发明内容

[0010] 针对现有模型检测在实际应用中存在的问题,本发明提供了一种自然语言到PPTL形式化规约自动生成方法及系统。

[0011] 本发明是这样实现的,一种自然语言到PPTL形式化规约自动生成方法,所述自然语言到PPTL形式化规约自动生成方法包括以下步骤:

[0012] 步骤一,使用Stanford NLP对输入的自然语言文本进行预处理,进行冗余信息删除,并将识别到的句子成分和结构进行重排,生成标记文本,便于后续转换器对句法成分的识别;

[0013] 步骤二,使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树,句法树为自然语言文本与PPTL形式化规约的中间表示形式,为自然语言文本提供了数据结构表示形式,便于后续生成PPTL形式化规约;

[0014] 步骤三,依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的PPTL形式化规约,将中间表示形式句法树转换为PPTL形式化规约,为下一步公式可满足判定提供输入;

[0015] 步骤四,使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果,只有公式可满足性判定成功后的形式化规约进行模型检测才有意义,避免对不满足和重言式进行无意义检测。

[0016] 进一步,步骤一中,所述输入的自然语言文本进行预处理,包括:

[0017] (1) 使用Stanford NLP处理输入自然语言文本生成语法树;

[0018] (2) 遍历语法树提取句子成分,并对句子成分进行重排,去除结构歧义;

[0019] (3) 将语法树中的词性标记进行提取并整合到文本中,生成标记文本。

[0020] 进一步,所述句子成分提取和重排,包括:

[0021] 句子成分提取的一般形式:英语句子的主干是主语、谓语和宾语,然后加上修饰句子的副词或介词短语;对于主从复合句的句子结构提取的形式为:一般形式是主从连词、从句、主句的形式,对于并列复合句一般形式是子句、并列连词、子句的形式。

[0022] 句子成分重排的流程为:首先,将每个树结构下的叶子节点,分支最深位置节点作为相应的句子成分,句子成分标记由该叶节点的父节点的节点含义进行定义;然后,对抽取出的句子成分进一步优化处理,并删除其中的冗余信息;最后,根据句子成分提取定义的句法形式对提取到的句子成分、主从复合句和并列复合句结构进行重排。

[0023] 进一步,步骤二中,所述语法语义分析,包括:

[0024] 使用JavaCC工具对标记文本进行语法语义分析,识别标记文本中的标记,提取句子成分信息,包括构成原子命题的简单句成分提取、整体句法结构的提取以及逻辑连接词语义和时序语义的提取等,将以上提取的信息记录并构建句法树。

[0025] 进一步,步骤三中,所述PPTL公式生成过程,包括:

[0026] 遍历句法树,结合PPTL公式生成规范为每个子句生成原子命题;将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的 PPTL形式化规约;其中,所述遍历包括:

[0027] (1) 若遇到命题节点,根据标记为命题添加时序操作符,并将命题描述信息取出,

按照命题变元和命题描述存到一个哈希表中；

[0028] (2) 若遇到连接符节点,判断父节点的描述信息,根据描述信息确定逻辑操作符,并将对应命题连接起来。

[0029] 进一步,步骤三中,所述生成公式的可满足性判定过程,包括:

[0030] 外部调用可满足性判定工具PPTLSAT,将生成公式和公式的非作为输入参数进行公式可满足性判定,然后将工具生成的文件进行解析并提取该工具的判定结果;其中,所述判定过程包括:

[0031] (1) 需要判定生成公式P,如果P没有构造出与之对应的LNFG,则表示P是矛盾的和不可满足的;否则P是可满足的,接着执行第(2)步;

[0032] (2) 需要判定生成公式的非Q,如果Q没有LNFG,表示Q是不可满足的,而P是一个重言式,判定失败;否则判定P既不是矛盾的,也不是重言式,判定成功。

[0033] 本发明的另一目的在于提供一种应用所述的自然语言到PPTL形式化规约自动生成方法的自然语言到PPTL形式化规约自动生成系统,所述自然语言到PPTL形式化规约自动生成系统包括:

[0034] 预处理模块,用于将输入的自然语言描述的性质进行预处理操作,将冗余无意义信息删除,去掉对后续文本分析和处理造成干扰的部分,以及句子成分提取定义的句法形式对句子成分和整体主从句结构进行重排;

[0035] 语法语义分析模块,用于使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树;

[0036] PPTL公式生成模块,用于依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的 PPTL形式化规约;

[0037] PPTL可满足性判定模块,用于使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果。

[0038] 本发明的另一目的在于提供一种计算机设备,所述计算机设备包括存储器和处理器,所述存储器存储有计算机程序,所述计算机程序被所述处理器执行时,使得所述处理器执行如下步骤:

[0039] 使用Stanford NLP对输入的自然语言文本进行预处理,删除冗余信息,并将识别到的句子成分和结构进行重排,生成标记文本;使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树;

[0040] 依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的 PPTL形式化规约;使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果。

[0041] 本发明的另一目的在于提供一种计算机可读存储介质,存储有计算机程序,所述计算机程序被处理器执行时,使得所述处理器执行如下步骤:

[0042] 使用Stanford NLP对输入的自然语言文本进行预处理,删除冗余信息,并将识别到的句子成分和结构进行重排,生成标记文本;使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树;

[0043] 依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的 PPTL形式化规约;使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果。

[0044] 本发明的另一目的在于提供一种信息数据处理终端,所述信息数据处理终端用于实现所述的自然语言到PPTL形式化规约自动生成系统。

[0045] 结合上述的所有技术方案,本发明所具备的优点及积极效果为:本发明提供的自然语言到PPTL形式化规约自动生成方法,能够帮助用户从自然语言描述的性质中提取形式化规约用于模型检测,有效地将用户描述的自然语言文本性质转化为PPTL公式,为普通用户使用模型检测技术提供了方便。

[0046] 本发明提供了将用户自然语言描述的性质转化为PPTL形式化规约的方法,用户只需要提供自然语言描述的性质文本,无需进行长期的学习和严格训练,也可得到符合形式语法与期望语义的性质规约公式。

[0047] 本发明所述方法对用户提供的性质文本语法结构无严格约束,并支持用户使用修饰短语、主从复合句进行性质描述,并针对PPTL的核心操作符pr j投影的监控语义描述定义了专用语法结构。

[0048] 本发明所述方法集成了公式可满足性判定工具PPTLSAT,在公式生成的基础上判定该公式是否可满足,根据判定结果提示用户是否进行下一步模型检测工作或进行公式二次修改。

附图说明

[0049] 为了更清楚地说明本发明实施例的技术方案,下面将对本发明实施例中所需要使用的附图做简单的介绍,显而易见地,下面所描述的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下还可以根据这些附图获得其他的附图。

[0050] 图1是本发明实施例提供的自然语言到PPTL形式化规约自动生成方法流程图。

[0051] 图2是本发明实施例提供的自然语言到PPTL形式化规约自动生成系统结构框图;

[0052] 图中:1、预处理模块;2、语法语义分析模块;3、PPTL公式生成模块;4、PPTL公式可满足性判定模块。

[0053] 图3是本发明实施例提供的语法树经过预处理的句子成分提取和重排操作后的示例图。

[0054] 图3(a)是本发明实施例提供的原始语法树示意图。

[0055] 图3(b)是本发明实施例提供的重排后语法树示意图。

[0056] 图4是本发明实施例提供的标记文本经过语法语义分析后生成的句法树示例图。

[0057] 图5是本发明实施例提供的生成的PPTL公式进行可满足性判定生成的LNFG示例图。

[0058] 图6是本发明实施例提供的公式的非进行可满足性判定生成的LNFG示例图。

具体实施方式

[0059] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0060] 针对现有技术存在的问题,本发明提供了一种自然语言到PPTL形式化规约自动生成方法及系统,下面结合附图对本发明作详细的描述。

[0061] 如图1所示,本发明实施例提供的自然语言到PPTL形式化规约自动生成方法包括以下步骤:

[0062] S101,使用Stanford NLP对输入的自然语言文本进行预处理,进行冗余信息删除,并将识别到的句子成分和结构进行重排,生成标记文本;

[0063] S102,使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树;

[0064] S103,依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的 PPTL形式化规约;

[0065] S104,使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果。

[0066] 如图2所示,本发明实施例提供的自然语言到PPTL形式化规约自动生成系统包括:

[0067] 预处理模块1,用于将输入的自然语言描述的性质进行预处理操作,将冗余无意义信息删除,去掉对后续文本分析和处理造成干扰的部分,以及句子成分提取定义的句法形式对句子成分和整体主从句结构进行重排;

[0068] 语法语义分析模块2,用于使用JavaCC工具对标记文本进行语法语义分析,识别语句中的标记,提取句子成分信息并构建句法树;

[0069] PPTL生成模块3,用于依据PPTL公式生成规则遍历句法树,将所有的原子命题与逻辑连接词操作符、时序操作符组合在一起,并根据句法树结构表达的逻辑连接词语义生成完整的 PPTL形式化规约;

[0070] PPTL可满足性判定模块4,用于使用公式可满足性判定工具PPTLSAT判定生成PPTL公式可满足性,并将生成文件进行解析获取判定结果。

[0071] 下面结合实施例对本发明的技术方案作进一步描述。

[0072] PPTL是把PTL中的部分命题成分去除,比如:量词、谓词和变量等,因此它是PTL的一个具有可判定性的子集,而且其具有一套完备的公理系统,并且它在表达能力上是等价于正则表达式的,所以PPTL是非常适合用来描述模型检测中的待验证性质,现已经广泛在软件系统安全性验证中使用。

[0073] 下述英语文本为用户描述的自然语言性质文本的示例:

[0074] When an anchor vehicle appears in front, the system sets the time in the next state, then the system will send the message, then when the driver presses the switch button, the vehicle switches modes.

[0075] 该文本定义了一个L3自动驾驶车辆接管行为的安全性质,接管请求时间被认为是影响驾驶员接管行为的首要数据指标,其一般被描述为当汽车的自动驾驶系统发出接管请

求警告提示信号时,由系统计算得出汽车与系统识别的障碍物或行人相撞的预期时间。学界将接管时间视作评估驾驶员接管灵敏度效果的重要参数,一般规定为一个特定的时间间隔,这个时间间隔是从系统发出提示警告信号开始到驾驶员关闭自动驾驶模式所耗费的时间。对于驾驶员来说,驾驶员解除自动驾驶模式,主要是借助驾驶员手动制动和转向或者点击方向盘上的按钮实现(由系统预设驾驶员旋转方向盘一定角度或踩制动踏板超过一定角度的阈值来触发自动驾驶关闭信号)。本发明对该文本进行处理如下:

[0076] 第一步,预处理:自然语言因其灵活的语法结构以及一词多义等特性,增加了软件处理的难度,而本发明提出方法面向的使用群体是不具备形式化验证背景知识的普通用户,这些用户提供的自然语言性质描述不免存在或多或少的冗余信息以及模糊描述,因此也提高了生成形式化规约的错误几率,所以要对处理的自然语言语句进行文本预处理,在其中提取对生成形式化规约有帮助的信息。

[0077] 自然语言预处理流程分为如下步骤:首先,对自然语言语句进行Stanford NLP语法树转换,预处理操作是在生成语法树的数据结构基础上进行;然后,分析句子的语法结构,提取句子成分,并按照句子成分提取定义的句法形式,对句子成分进行重排,去除结构歧义对后续语句分析的干扰;最后,将语法树中的词性标记进行提取整合,生成标记文本。

[0078] 图3表示经过句子成分提取并重排操作后的效果图,本发明可以对比(a)、(b)得出,原始句法树经过预处理操作后,原始句子主干成分已经被提取并重排,主从句均按照主语、谓语、宾语、修饰成分的顺序进行句子成分重排,主从复合句结构也被解析并识别,从句结构已经从介词短语子结构中提取出来,并按照从句引导词、从句、主句的顺序进行了结构重排。这样,语法树的消歧操作就完成了,由于解决了嵌套子结构的识别问题,因此生成的标记文本中去除了结构歧义。

[0079] 生成的标记文本如下:

[0080] WRB When DT an NN anchor NN vehicle VBZ appears IN in NN front,DT the NN system VBZ sets DT the NN time IN in DT the JJ next NN state ,RB then DT the NN system MD will VB send DT the NN message ,RB then WRB when DT the NN driver VBZ presses DT the NN switch NN button ,DT the NN vehicle VBZ switches NNS modes .

[0081] 可以看到,输出的字符串中,将每个单词已经分词,并且将每个单词的词性标记在前。

[0082] 第二步,语法语义分析:经过预处理后的标记文本在语法上是规范且结构化的,易被转换器接受并处理。因此,本发明使用JavaCC编写了一个可以识别符合结构化英语语法文本的转换器,将输入的标记文本转化为结构清晰的句法树,根据转换器识别到的句法成分连接并生成对应的特定子句树节点,并将识别到的复合句成分构建对应的特定复合句树结构。

[0083] 结构化英语语法是根据预处理操作中成分重排句法结构进行定义,对于名词、动词、分词、形容词和副词,本发明不再分解定义它们。该结构化英语语法规则允许根据时态、副词和介词短语提取时序语义信息。根据主从复合句和并列复合句提取逻辑连接词语义信息,并构造了为pr j时序操作符识别服务的特殊语法结构。根据该语法规则的定义,用从句的标记clause定义了简单句的语法结构,并且一个句子至少由一个从句组成,表示标记

文本中至少存在一个简单句。一个从句至少由主谓两个成分组成,用于表达句子的核心意思,表示简单句的组成成分可以是:主语+谓语或者主语+谓语+宾语的形式,还可以选择使用主从复合句或并列复合句等复杂句法结构来扩展句子的意思,这个结构化语法规则对应了预处理的句子成分和结构重排操作后的标记文本中的语法结构。

[0084] 图4为原始语法树经过预处理和语法语义分析后生成的结构清晰的句法树,句法树相比原始语法树已经结构优化,并针对相应的子句之间的结构语义进行了子结构重定义。从句法树的结构中可以发现,叶子结点存储的是原子命题以及时序描述信息,内部节点和根节点存储的是从属连词和并列连词。并且,本发明使用不同类型的节点以及节点标记,来区分节点的作用。根据when连词的语义:其有前置状态触发后置状态的语义来对应逻辑符号->(蕴含),并且存在前置状态和后置状态的触发具有先后顺序,对应时序操作符() (next),将主句和子句进行连接,作为when节点的左右子树。根据副词then的语义:前置状态和后置状态的触发具有先后顺序的语义,使用时序逻辑符号(:) (chop)将语法树当前节点的左右子节点进行连接。根据短语“in the next state”的语义:其语义获取是在标记识别阶段,当转换器识别出这个短语后,会进行同义词匹配操作,在其同义词集合中获得含“next”的语义,表示当前命题在下一个状态成立,将时序操作符() (next)修饰当前原子命题。根据时态标志助动词“will”的语义:当前命题在将来的某一个状态成立,使用时序操作符<>(sometimes)修饰当前原子命题。

[0085] 第三步,PPTL生成:经过预处理和语法语义分析后,从自然语言文本中已经得到了关键的语法语义信息,并储存在结构清晰的句法树中,接下来将根据句法树生成PPTL公式。PPTL的基本组成是原子命题,因此PPTL公式生成过程分为两个步骤:遍历树结构,首先,将句法树中每个子句节点的子句信息转换成原子命题;然后,将树节点中的时序信息、树结构表示的逻辑关系与原子命题结合,生成完整的PPTL公式。

[0086] 基于后序遍历算法遍历生成的句法树。在遍历句法树时,将“when”连词节点映射成逻辑连接符和时序操作符的组合->()。“then”节点映射成逻辑符号;(chop)。短语“in the next state”映射成() (next)。时态助动词will映射成<>(sometimes)。将一个句子拆分成若干个原子命题,并将原子命题根据句子结构映射的逻辑连接词语义组成PPTL公式,文本中“when”和“in the next state”存在时序语义重复,本发明在生成PPTL公式时进行了语义简化操作。生成的原子命题以及PPTL公式如下所示:

[0087] P: an anchor vehicle appears

[0088] Q: the system sets the time

[0089] R: the system send the message

[0090] S: the driver presses the switch button

[0091] T: the vehicle switches modes

[0092] PPTLformula: (((P) -> () (Q)); (<>R)); ((S) -> () (T))

[0093] 生成的原子命题中已经去除了连词和时序描述信息,并且标记了命题变元,生成的公式也符合句法树中节点信息。

[0094] 第四步,PPTL可满足性判定:公式生成成功后,会外部调用公式可满足性判定工具PPTLSAT将生成的PPTL公式进行可满足性判定,输出判定结果以及生成标记范式图LNFG,也就是该公式对应的状态转换图。工具构造的LNFG就是用来判定公式可满足性的依据,该工

具构造的LNFG如图5所示,对该公式取非,并生成LNFG如图6。通过两图发现公式以及公式的非均成功构造出了LNFG,说明生成的公式既不是矛盾的也不是重言式,表明公式可满足性判定成功,可以继续下一步模型检测工作。

[0095] 证明部分(具体实施例/实验/仿真/能够证明本发明创造性的正面实验数据等)

[0096] 实验部分在四个领域共678条由自然语言描述的性质数据集上进行处理。将这些自然语言文本逐条处理,进行PPTL形式化规约生成,并对生成的PPTL公式进行人工评定公式正确性。从数据集中总结的五条结构上具有代表性的自然语言性质描述以及对应的PPTL公式生成结果如下表所示:

自然语言性质描述	PPTL 公式生成结果
<p>Whenever the button to some floor is pushed, the elevator will reach there.</p> <p>If Air_Ok_signal remains low, auto_control_mode is terminated in 3 seconds.</p>	<p>P:the button to some floor is pushed</p> <p>Q:the elevator reach</p> <p>PPTL.formula:([] (P->()Q))</p> <p>P:Air_Ok_signal remains low</p> <p>Q:auto_control_mode is terminated in 3 seconds</p> <p>PPTL.formula:(P->()Q)</p>
<p>When auto_control_mode is entered, eventually the cuff will be inflated.</p>	<p>P:auto_control_mode is entered</p> <p>Q:the cuff is inflated</p> <p>PPTL.formula:(P->fm(Q))</p>
<p>When registration_terms are displayed, if registration_form is not requested, next introduction_screen is displayed.</p>	<p>P:egistration_terms are displayed</p> <p>Q:registration_form is not requested</p> <p>R:introduction_screen is displayed</p> <p>PPTL.formula:(P->(Q->()R))</p>
<p>After local_printing is executed, registration_form is filled.</p>	<p>P:local_printing is executed</p> <p>Q:registration_form is filled</p> <p>PPTL.formula:(P->Q)</p>

[0097]

[0098] 将实验生成的公式让5位形式化验证领域专业人员进行交叉评定,保证每条公式都经过两位以上评定人的评定,并统计评定结果。使用生成公式的正确性作为评估可用性指标,评估实验结果如下表所示:

[0099]

总数	正确数	错误数	正确率
----	-----	-----	-----

678	635	43	93.66%
-----	-----	----	--------

[0100] 678条自然语言文本中,有635条文本被正确转化为PPTL形式化规约,其余文本为转换公式错误或转换失败,该工具达到了可以接受的公式生成正确率。

[0101] 在上述实施例中,可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用全部或部分地以计算机程序产品的形式实现,所述计算机程序产品包括一个或多个计算机指令。在计算机上加载或执行所述计算机程序指令时,全部或部分地产生按照本发明实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一个计算机可读存储介质传输,例如,所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如同轴电缆、光纤、数字用户线(DSL)或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输)。所述计算机可读取存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质,(例如,软盘、硬盘、磁带)、光介质(例如,DVD)、或者半导体介质(例如固态硬盘Solid State Disk(SSD))等。

[0102] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,都应涵盖在本发明的保护范围之内。

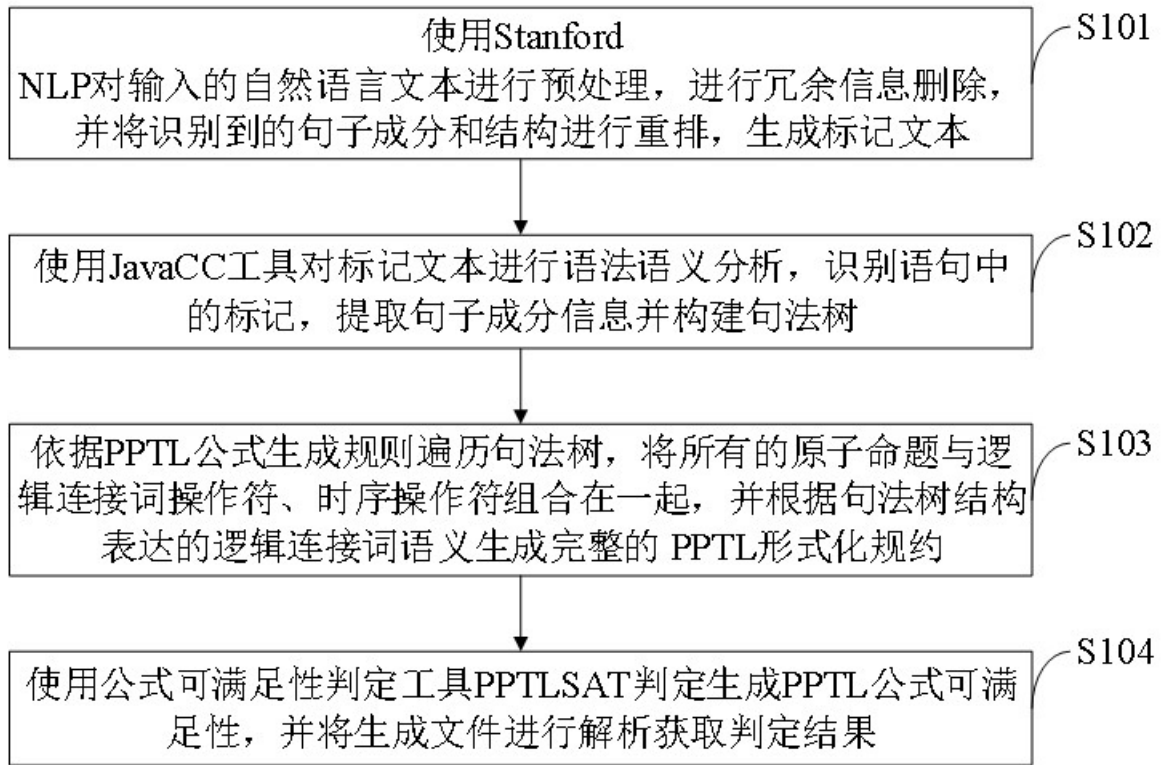


图1

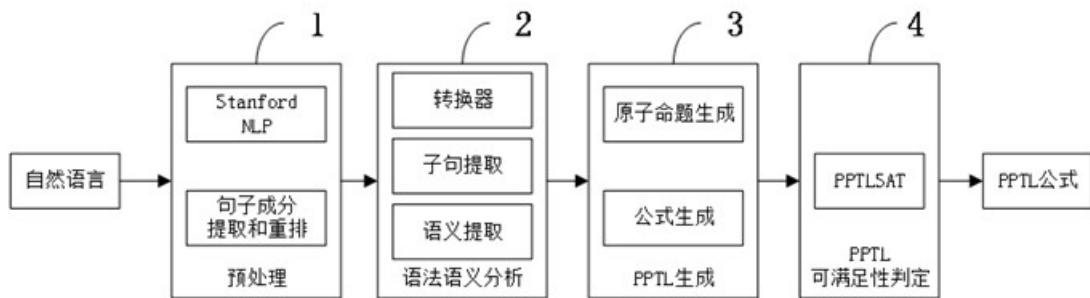


图2

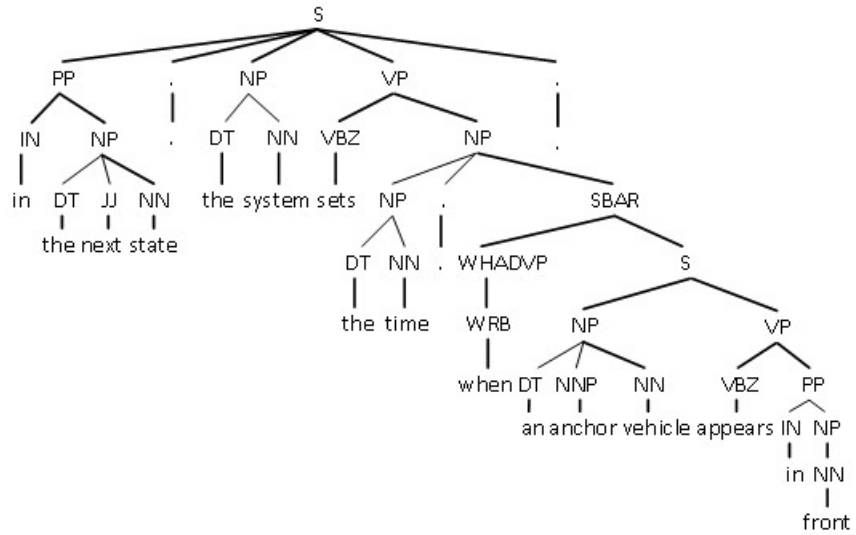


图3a

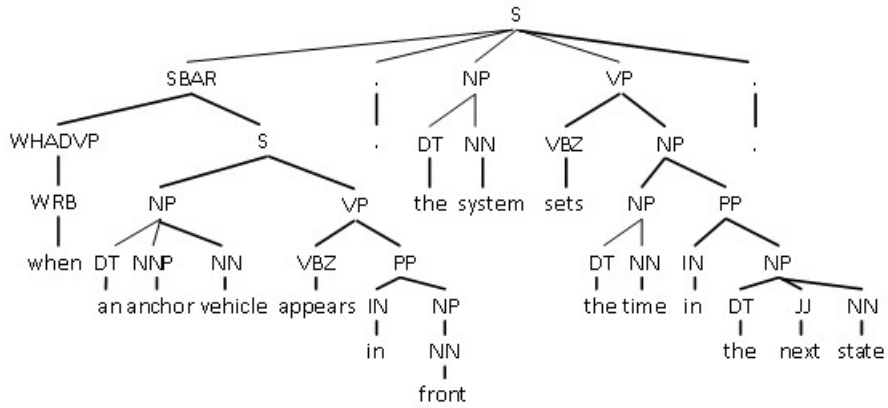


图3b

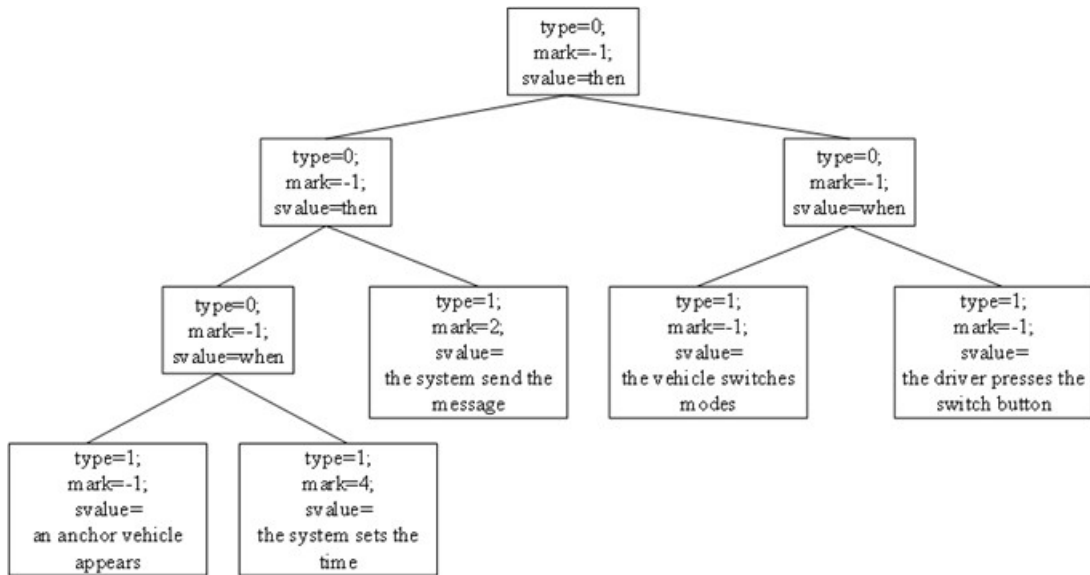


图4

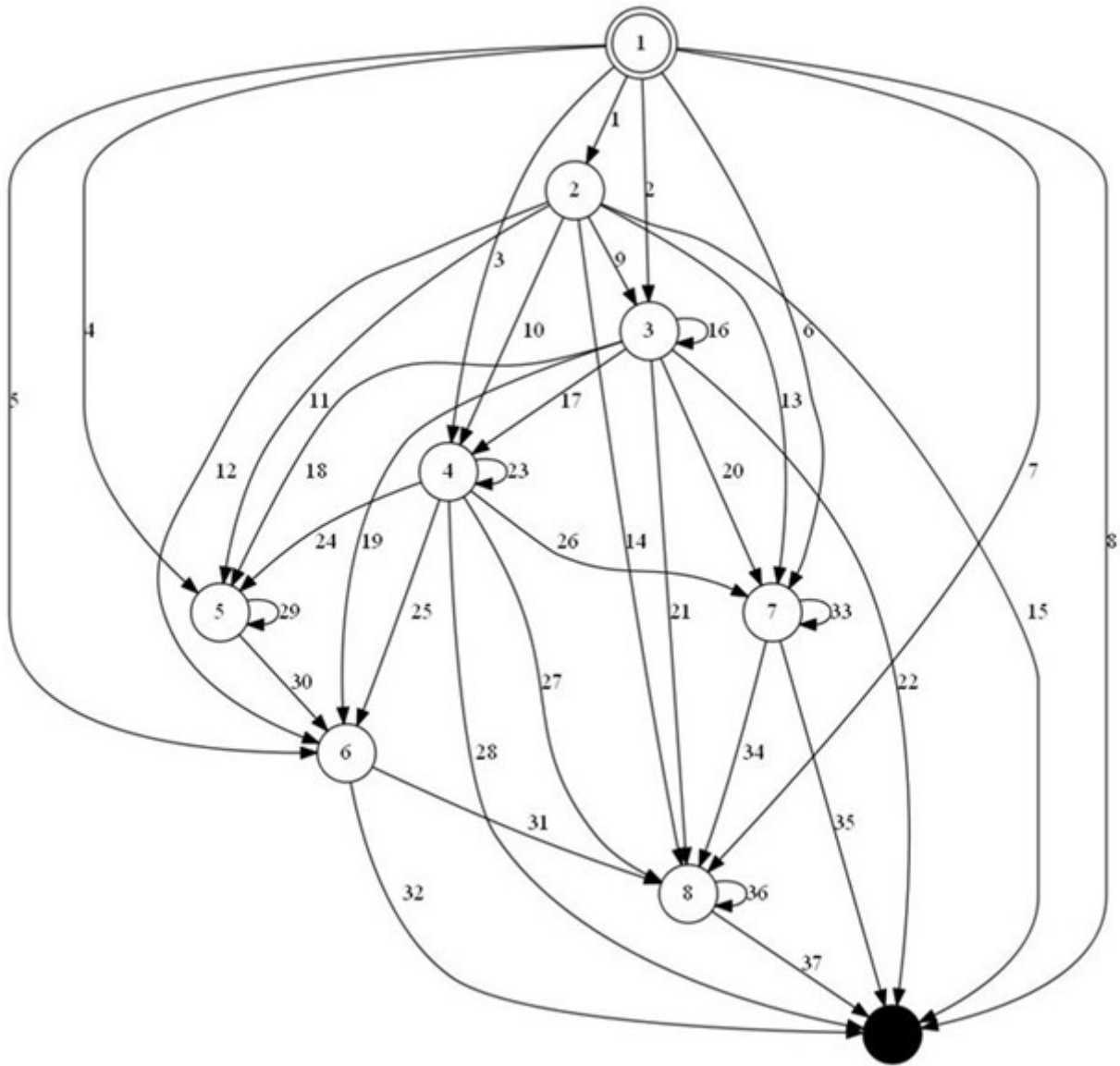


图5

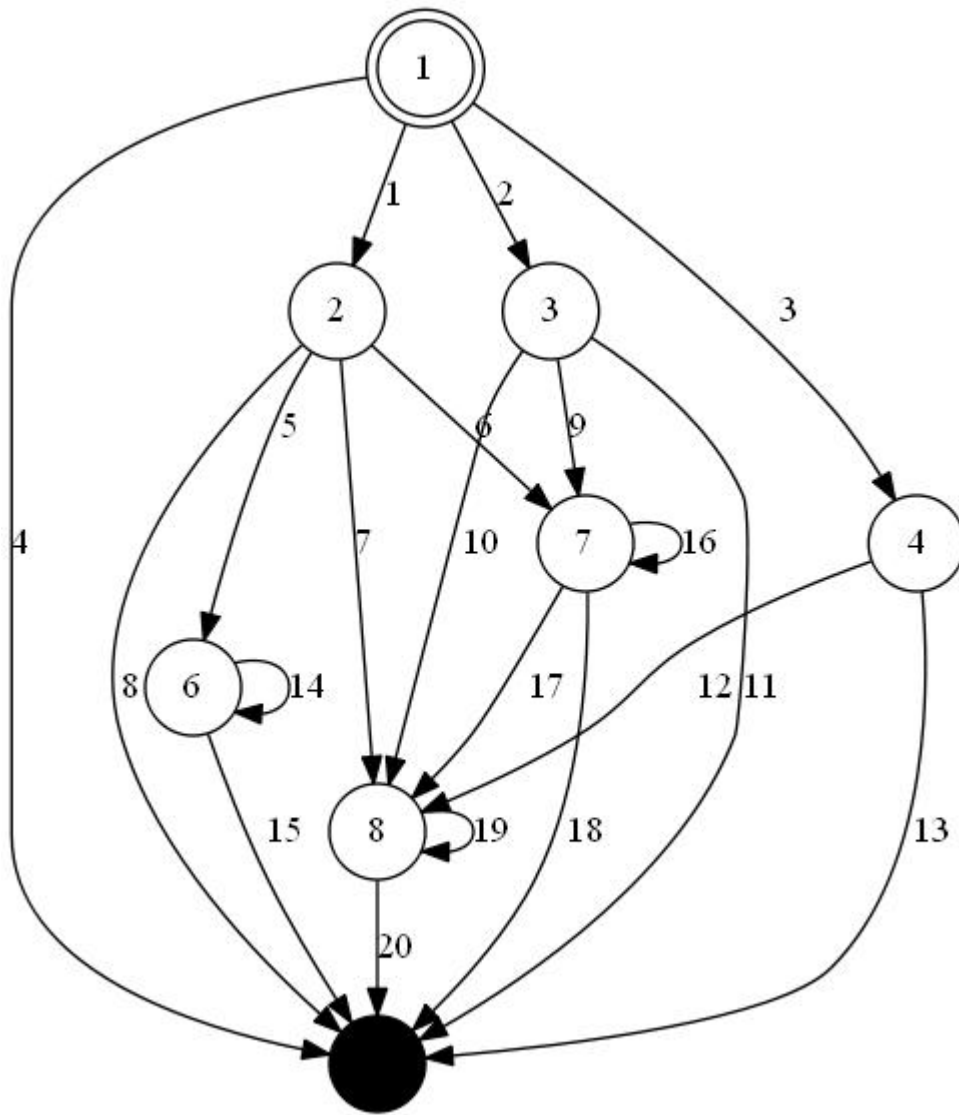


图6