

基于混合推理系统挖掘用户访问模式的 ILP 方法

孙 明, 陈 波, 周明天

(电子科技大学 计算机科学与工程学院, 四川 成都 610051)

摘 要: 提出一种结合 DL-safe 规则挖掘用户访问模式的 ILP 方法. 该方法首先定义基于事件的日志本体, 将 DL-safe 规则和日志本体构建为一个推理过程可判定的混合日志知识库, 在此基础上借助 ILP 的理论和方学习生成有效用户访问模式. 该方法充分利用日志本体中事件间特定的语义关系有效简化了挖掘过程并提高了挖掘效率, 为改进站点结构提供更有效的决策. 实验结果表明了该方法的可判定性和有效性.

关键词: 日志本体; DL-safe 规则; 混合推理; 模式发现; 归纳逻辑编程

中图分类号: TP18

文献标识码: A

文章编号: 1000-7180(2010)01-0029-05

An ILP Approach for Mining User Access Patterns Form Hybrid Reasoning System

SUN Ming, CHEN Bo, ZHOU Ming-tian

(School of Computer Science and Engineering, University of Electronic Science
and Technology of China, Chengdu 610051, China)

Abstract: We present an ILP approach for mining the user access patterns with DL-safe rules. This method defines the log ontology based on events, and combines log ontology and DL-safe rules into a decidable hybrid knowledge base. The useful user access-patterns can be generated from the knowledge base by ILP learning method. The semantics of the relations between events in log ontology can improve the results and efficiency of ontology learning effectively. The results of the method provide more decision-making for improving the structure of web sites. The experimental results show that this method is decidable and it is quite feasible to solve practical problems.

Key words: log ontology; DL-safe rule; hybrid reasoning; pattern discovering; ILP (inductive logic programming)

1 引言

模式发现是语义 Web 使用挖掘的主要过程之一. 传统的挖掘算法采用 Web 日志文件作为挖掘集, 而本体 (Ontology) 作为一种在语义和知识层面上的概念建模工具可以表示更丰富的背景知识, 用本体来表示用户与 Web 系统之间的交互能更清晰地表示用户的语义行为, 从而获得更准确的挖掘结果^[1]. 为此, 文中从分析 Web 事务序列出发, 采用基于事件的日志本体^[2]代替传统的 Web 日志文件作为用户访问模式发现的知识基础.

日志本体以描述逻辑 (Description Logic, DL) 为

基础, 拥有相对丰富的概念定义, 而缺少对概念和角色之间关系建模的能力. 克服这种缺陷的一个可行方法是与基于规则的逻辑系统相结合构成混合知识库, 但这样容易引起推理的不可判定. 已有的与规则相结合的描述逻辑系统, 在实际应用过程中存在较大的限制. DL-safe 规则^[3]是无函词 Horn 逻辑的子集, 具有概念和角色的表示推理能力, 同时对描述逻辑与规则间的接口做了限制, 要求规则满足选言规则形式同时, 规则中的变量所涉及个体被显式引入到知识库, 从而确保其推理过程的可判定性. 为此, 文中结合描述逻辑和 DL-Safe 规则将日志本体构建为一个可判定混合知识表示和推理库, 在此基础上

收稿日期: 2008-12-23; 修回日期: 2009-02-23

基金项目: 国家自然科学基金项目 (90604033); 国家“十一五”科技支撑计划重大资助项目 (2006BAH02A0407)

利用日志本体事件间特定的语义关系,借助 ILP 归纳逻辑程序设计的理论与方法合理构建用户访问模式,从而获得更确定的用户访问策略.

2 事件及日志本体

2.1 事件和事件间关系

在 Web 用户请求与被请求资源的基础之上构建的日志本体,蕴含了用户访问站点行为的潜在语义,在知识层次上反映了用户和站点交互的行为和兴趣.日志本体采用事件这一核心概念来表述用户访问站点的语义行为.

定义 1 事件(Event)事件是对一个站点商务模型所有访问任务的详细形式化描述,用集合 E 表示.

根据访问目的、访问兴趣和访问策略语义的不同,将事件在共性层次上分为原子事件和复合事件.原子事件(AtomEvent, EA)描述用户一次具体的请求,包括内容事件(ContentEvent, EC)和服务事件(ServiceEvent, ES),分别用于描述用户访问具体页面内容的语义行为和用户使用站点服务的语义行为;而复合事件(ComplexEvent, EX)是由原子事件构成的一个有序序列,在更抽象的层次上,复合事件可以被认为是事件层次体系结构中的一个语义访问路径和访问策略.

事件的生成是用户和网站系统之间交互的结果,因此我们分别从系统角度和用户角度去分析事件间抽象关系.从系统视点出发,事件之间借助用户访问顺序形成关联,具有前项关系 $\text{previousOf}(E_i, E_j)$ 和整分关系 $\text{hasPart}(E_i, E_j)$.从用户视点出发,事件关系是特定领域内的用户语义行为,主要包括:继承关系 $\text{subClassOf}(E_i, E_j)$,实例关系 $\text{instanceOf}(E_i, E_j)$.其中,整分关系表示原子事件和复合事件之间的“part-whole”关系,是挖掘用户访问模式的重要基础.

2.2 日志本体

日志本体描述站点使用信息的背景知识,蕴含了更加丰富的 Web 使用语义.在领域本体^[3]的基础上定义基于事件及事件间关系的日志本体.

定义 2 日志本体(Log Ontology, LO)若 L 为具有语义推导规则的逻辑语言,则一个领域 D 上的日志本体定义为一个六元组, $LO := \{E, R, \leq E, \leq R, \text{Rel}, A\}$.其中, E 和 R 是两个不相交的集合, E 是事件概念集,每一个事件个体都是有名个体, R 是事件关系集; $\leq E$ 是 E 上的偏序关系,表示事件层次; $\leq R$ 是 R 上的偏序关系,表示关系层次; Rel 是

函数,定义为 $R \rightarrow E \times E$,表示事件间的关系; A 表示基于 L 的推导公理集.

采用专家定义和本体学习相结合的方法,可以从 Web 日志系统中挖掘生成上述定义的日志本体,并用于后续访问模式的挖掘.

3 建立日志本体混合知识库

3.1 DL-safe 规则

日志本体采用 OWL-DL 语言表示,其描述逻辑基础是 SHOIN(D),在此基础上结合无函词 Horn 规则,如选言 Datalog 子句构成混合推理系统具有强大的推理和表示能力,十分适合于语义 WEB 挖掘推导出新的知识.为了消除推理过程的不可判定性,必须对描述逻辑和规则的结合进行限定.

定义 3 DL-safe 规则^[2] 设 KB 为一个 SHOIN(D) 知识库, N_c 为知识库中的概念集合, N_{Ra} 为抽象角色的名字集合, N_{Rc} 为类型角色的集合, N_p 为谓词符号的集合, $\{\approx\} \cup N_c \cup N_{Ra} \cup N_{Rc} \subseteq N_p$.一个 DL 项的形式如 $A(s)$,其中 $A \in N_c$,或者形如 $R(s, t)$,其中 $R \in N_{Ra} \cup N_{Rc}$.规则 r 形式如下:

$$A_1 \vee A_2 \vee \cdots \vee A_m \leftarrow B_1, B_2, \cdots, B_n \quad (1)$$

规则 r 的前件中只能包含逻辑合取运算符,谓词符号来源于 N_p 集合与非 DL 谓词,并且所有在 r 的前件中的变量必须至少出现在一个非 DL 原子中,这是为了确保所有被规则使用的个体都是被显式引入到知识库中.在实际推理过程中可以通过技术手段为规则添加特殊的非 DL 原子 $O(x)$ 并在规则库中为对应个体 a 添加基原子 $O(a)$ 的方式以满足 DL-safe 规则限制,如规则(3)是在规则(2)的基础上添加特殊非 DL 原子构成的 DL-safe 规则.

$$\text{Human}(y) \leftarrow \text{Human}(x), \text{parent}(x, y) \quad (2)$$

$$\text{Human}(y) \leftarrow \text{Human}(x), \text{parent}(x, y),$$

$$O(x), O(y) \quad (3)$$

3.2 结合 DL-Safe 规则建立日志本体混合知识库

日志本体在定义的过程中不支持标称事件,同时日志本体生成的每个事件个体都是有名个体,不支持匿名个体的存在.在推理过程中将日志本体中数量值约束归约为定性数量值约束,其描述逻辑基础就从 SHOIN(D) 限定为 SHIO(D).日志本体描述了既定的站点使用背景知识,这种表示是静态的,不能反映日志系统随站点访问量不断增加而动态变化的应用知识.为从日志本体中发现更实际用户访问模式,引入一阶规则以描述用户使用站点变化的

语义,并将其在 DL-safe 的限定下和日志本体构成混合推理知识库.接下来,定义基于 DL-safe^L 的日志本体知识库. DL-safe^L 是在描述逻辑语言 L (这里 L 限定为 SHIQ(D)) 的基础上满足 DL-safe 规则的一组选言 Datalog 语言^[4].

定义 4 DL-safe^L 知识库 $B = (KB, P)$, 其中 KB 是对应于描述逻辑 L 的日志本体知识库, P 是满足 DL-safe 限制的应用规则 r 组成的集合, r 形如:

$$h_1(X_1) \cdots, h_s(X_s) \leftarrow a_1(Y_1), \cdots, a_m(Y_m) \& b_1(Z_1), \cdots, b_n(Z_n) \quad (4)$$

式中, $h_1, \cdots, h_s$ 是任意谓词, $a_1, \cdots, a_m$ 是非 DL 谓词, $b_1, \cdots, b_n$ 是 DL 谓词, $X_1, \cdots, X_s, Y_1, \cdots, Y_m$ 是任意项的序列, $Z_1, \cdots, Z_n$ 是一元或二元项的序列.

为了在统一的规则下挖掘日志本体,需要将基于描述逻辑语言 L 的知识库 KB 等价地转换为一阶选言 Datalog 规则库 $DD(KB)$, 并在此基础上增加 DL-safe 程序 P 构成新知识库 $DD(KB) \cup P$. 参考文献[5], 定义上述转换过程如下:

首先通过消除传递公理将基于 SHIQ(D) 的日志本体知识库 KB 转换为基于 ALCHIQ(D) 的知识库 $\Omega(KB)$;

通过结构化转换, 利用映射规则 $\pi(\Omega(KB))$ 将 $\Omega(KB)$ 等价的转换为一阶逻辑形式 $\Xi(KB)$;

利用基本重叠规则 BS 优化 $\Gamma_{TR} = \Xi(KB_T) \cup KB_R$ 为 $\text{sat}(\Gamma_{TR})$;

将函词 f 用二元谓词 S_f 代替, 并为每个个体 a 在知识库中增加新的断言 $S_f(a, a_f)$, 达到消除函词的目的, 将知识库 $\text{Sat}(\Gamma_{TR}) \cup \Xi(KB_A)$ 转化为无函词一阶逻辑 $FF(KB)$;

最后改写规则将 $FF(KB)$ 等价地转换为选言 Datalog 规则库 $DD(KB)$.

日志本体知识库采用基于一阶规则的查询 - 应答的机制进行推理, 若基原子 α 是 (KB, P) 的应答当且仅当 $DD(KB) \cup P \models \alpha$.

4 基于混合知识库发现用户访问模式的 ILP 方法

语义 Web 使用挖掘过程中, 用户访问模式描述了用户在特定访问策略中发现用户使用站点行为之间的关联. 日志本体是以事件为核心概念的应用领域本体, 不同事件反映了用户使用站点的背景知识, 因此事件间的相互关系, 尤其是描述原子事件如何组成复合事件的整合关系构成了用户模式挖掘的基础. 在此基础上, 结合满足 DL-safe 限制的应用规则,

引入用户使用站点的动态语义知识, 借助 ILP 的理论可以从混合日志本体知识库中发现有效的用户访问模式.

4.1 事件间整合关系语义

日志本体的结构来自于日志文件的抽取, 即一个 Web 访问序列抽象为一个复合事件, 复合事件和原子事件之间的整合关系属于整体部分理论中组员/联合体 (component/complex) 子关系, 采用 harPart 来表示这种整合关系. 若原子事件个体 t 是构成复合事件个体 s 的一个部分, 则表示为: $\text{harPart}(t, s)$, 其中 $t \in C, s \in D, C \sqsubseteq \exists \text{hasPart}. D, C = EC \cup ES, D \sqsubset EX$. 这种角色是非传递性的简单角色, 不存在复合事件个体还可以构成其它复合事件个体的情况, 即 $C \sqsubseteq \exists \text{hasPart}. D$ 当且仅当 $C = EC \cup ES, D \sqsubset EX$.

用户访问模式考虑的是特定访问策略下用户访问行为之间的相互关系, 在描述逻辑到一阶规则转换过程中出现的函词在语义上可以看作为复合事件和原子事件间整合关系的具体体现, 因此消除整合函词时而产生的二元谓词 S_f 与整合角色等价, 在日志本体挖掘用户访问模式的过程中, 无需为每个事件个体再添加新的断言 $S_f(a, a_f)$, 从而避免了如何为事件个体 a 定位其对应 a_f 的难题. 挖掘用户访问模式就应该从表示策略语义的复合事件出发, 并以此为核心发现以原子事件表示的各种用户行为语义之间的使用规则.

4.2 挖掘用户访问模式

基于 ILP 挖掘用户访问模式的学习目标是: 在保证一致性前提下从 DL-safe^L 知识库 B 中找出访问模式 H , 使得 H 与 B 能够逻辑蕴涵观察集 $O(B \cup H'O)$, 即从复合事件 EX 出发找出与原子事件 EA 之间的语义规则. 学习过程采用查询 - 应答机制进行推理.

在挖掘过程中, 访问模式 H 被指定为一组联合查询, 其应答集包含 ALCHIQ(D) 基准事件 E_{ref} 的个体, 将这些查询称为 E- 查询. 通过对 E- 查询的应答, 发现从基准事件的 E_{ref} 个体到任务相关事件个体之间的访问关系.

定义 5 E- 查询, 对于给定基准事件 E_{ref} , 关于 DL-safe^L 知识库 B 的 E- 查询 Q 定义为具有以下形式的 Datalog 子句:

$$q(x) \leftarrow a_1(Y_1), \cdots, a_m(Y_m) \& E_{\text{ref}}(x), b_1(Z_1), \cdots, b_n(Z_n) \quad (5)$$

式中, $a_1, \cdots, a_m$ 是非 DL 谓词, $b_1, \cdots, b_n$ 是 DL 谓

词, x 是特征变量. 特别地, 约束空子句 $q(x) \leftarrow \&E_{\text{ref}}(x)$ 称为平凡 E-查询. 查询 Q 的应答是关于 Q 中变量的基置换 θ , 若 $K \models Q\theta$, 则称 θ 是查询 Q 关于知识库 B 的正确应答.

构建访问模式 H 从基准事件 E_{ref} 开始, 采用模式树^[6]的方法建立访问模式: 以基准事件 E_{ref} 为根节点, 根据容许规则为每一个节点增加 KB 中的 DL 原子或规则集 P 中的非 DL 原子作为其叶子节点生成模式树. 访问模式 H 则是模式树中从根节点出发到达叶子节点的一条路径. 根据事件关系的语义, 文中定义的增加节点的容许规则如下: 节点 N 的子节点包括其依赖节点(至少与 N 有一个共享变量), 右兄弟节点以及其复制节点, 要求增加子节点后生成的新访问模式 H' 满足 $B \cup H' \models O$. 图 1 是一棵访问模式树的例子, 满足 DL-safe 规则的一个访问模式如下:

$q(x) \leftarrow \text{DirectBuy}(x) \& \text{Buy Product}(x),$
 $\text{hasPart}(x, y), \text{Samsug}(y)$ (6)

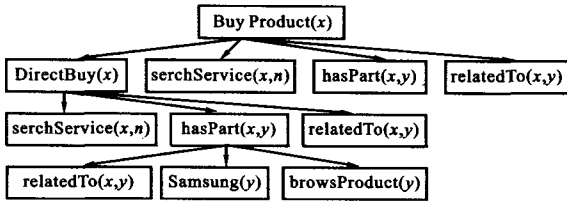


图 1 访问模式树

4.3 访问模式有效性验证

在 ILP 方法下学习生成的访问模式 H 要求 $B \cup H \models O$, 具体来说就是对于基准事件 E_{ref} 的个体 a_i 要求 $B \cup H \models O_i$, 其中 $O_i \in O$ 定义为一个二元组 $(q(a_i), A_i)$ 表示 a_i 对应的观察, $q(a_i)$ 是基原子, A_i 是一组关于 a_i 的基 Datalog 事实. 接下来给出 H 关于 B 蕴含 O_i 的定义.

定义 6 给定模式 $H \in \text{DL-safe}^L$, 知识库 B 以及观察 $O_i \in O$, H 关于 B 蕴含 O_i 当且仅当 $B \cup H \cup A_i \models q(a_i)$.

引理 1 Q 为 DL-safe^L 知识库 B 的基查询, 若 $B \vdash Q$ 当且仅当 $B \models Q$.

证明过程与参考文献[7]类似, 这里不再赘述.

定理 1 给定模式 $H \in \text{DL-safe}^L$, 知识库 B 以及观察 $O_i \in O$, 若 H 关于 B 蕴含 O_i 当且仅当 $B \cup H \cup A_i \vdash q(a_i)$.

证明: P 关于 B 蕴含 O_i

$\Leftrightarrow B \cup H \cup A_i \models q(a_i)$ (定义 5)

$\Leftrightarrow B \cup P \cup \text{body}(O_i) \vdash q(a_i)$ (引理 1)

因此验证 e_i 是 H 的正确应答就等价于是否能找到关于 $B \cup H \cup A_i$ 一个正确的应答. 为了获得更为实际和灵活的访问模式, 在模式树构建的过程中会引入应用规则中的非 DL 原子, 因此访问模式的有效性验证不仅仅是描述逻辑中的概念相容性验证或实例检查, 还涉及到异构规则的验证. 本文分两个阶段验证访问模式的有效性: 首先采用制约 SLD 否认的方法消除模式中非 DL 原子, 然后将基化后的 DL 原子提交至描述逻辑推理机完成实例检查. 参考 SPADA^[8], 我们定义 DL-safe^L 上的制约 SLD 否认.

定义 7 制约 SLD 否认 给定 DL 语言 L , $B = (KB, P)$ 是 DL-safe^L 知识库. q 是关于 B 的查询目标, 其制约 SLD 否认是一组有限的制约 SLD 演算 $d_1, \dots, d_m$ 满足下面条件, 其中 $q_0^i, \dots, q_{n_i}^i$ 是第 i 个演算 d_i 的查询目标序列, $1 \leq i \leq m$.

(1) 对于任意 i , $q_{n_i}^i$ 形如 $\& \beta_1^i, \dots, \beta_{l_i}^i$, 即: 每个演算 d_i 的最后查询目标 $q_{n_i}^i$ 是制约空子句;

(2) 对于任意 B 的模型 Γ , 至少存在某个 i , $1 \leq i \leq m$, 使得 $\Gamma \models q_{n_i}^i$, 记为 $B \models \text{disj}(q_{n_1}^1, \dots, q_{n_m}^m)$.

如果关于 B 的查询目标 q 有制约 SLD 否认, 那么我们记作 $B \vdash q$. 经过制约 SLD 否认后所有 DL 原子得以全部基化, 然后将 DL 基原子提交 DL 推理机完成实例检测验证是否为有效模式.

5 实验验证

5.1 验证过程

为验证 DL-safe^L 基础上的基于 ILP 的日志本体挖掘框架, 我们挖掘一家在线手机购物网站的日志文件, 生成日志本体构成实验的知识库 B_{OMS} . 为了形式化验证的一致性, 用户访问策略分为购买产品 BuyProduct 和浏览产品 Browse Product. 表 1 描述了部分用于推理的 B_{OMS} 描述逻辑知识库.

表 1 部分描述逻辑知识库

TBox	$\text{Shopping} \sqsubseteq \text{EX Pay} \sqsubseteq \text{ES}$ $\text{MobilePhone} \sqsubseteq \text{EC MobilePhone} \sqsubseteq \text{Product}$ $\text{Samsung} \sqsubseteq \text{MobilePhone}$ $\text{Shopping} \sqcap \exists \text{hasPart. Pay} \sqsubseteq \text{BuyProduct}$ $\text{BuyProduct} \sqcap \text{BrowseProduct} = \emptyset$
ABox	$\text{Shopping}('ex0806012')$ $\text{Pay}('espay0806012') \text{ Samsug}('E908')$ $\text{hasPart}('ex0806012', 'espay0806012')$ $\text{hasPart}('ex0806012', 'E908')$

下面应用规则满足 DL-safe 的限制, 表示在一

个具有在线支付过程的购物事件中若查询服务小于等于 3 次,则该购物事件表示直接购买策略,规则描述如下:

$\text{DicretBuy}(x) \leftarrow \text{searchService}(x, 3), O(y) \& \text{Shopping}(x), \text{hasPart}(x, y), \text{Pay}(y)$ (7)

H 是待检测的用户访问模式,

$H = q(x) \leftarrow \text{DicretBuy}(x)$

$\& \text{BuyProduct}(x) \text{hasPart}(x, y), \text{Samsug}(y)$ (8)

其基准事件 BuyProduct 的个体 'ex0806012' 的观察 $A_{\text{ex0806012}}$ 包括以下非 DL 观察事实:

$\text{searchTimes}(\text{'ex0806012'}, 3)$ (9)

$O(\text{'espay0806012'})$ (10)

分别表示购物事件 '0806012' 中用户查询次数不大于 3 以及特殊非 DL 谓词 $O(x)$ 的个体实例。

验证模式 H 关于知识库 B_{OMS} 的有效性可以归结为验证核心事件 $\text{BuyProduct}(x)$ 个体 'ex0806012' 是否是 H 的正确应答,即 $B \cup H \cup A_{\text{ex0806012}} \vdash q(\text{'ex0806012'})$. 这等同于在知识库 $B \cup H \cup A_{\text{ex0806012}} \vdash q(\text{'ex0806012'})$ 之上找一个关于查询 $Q^{(0)} = \leftarrow q(\text{'ex0806012'})$ 的正确应答。

首先将 B_{OMS} 描述逻辑知识库映射为 $\Xi(B_{\text{OMS}})$, 包括 $\neg \text{Shopping}(x) \vee \neg \text{hasPart}(x, f(x)) \vee \neg \text{Pay}(f(x)) \vee \text{BuyProduct}(x)$ (11)

消除函词后,式(11)转变为

$\text{BuyProduct}(x) \leftarrow \text{Shopping}(x), \text{hasPart}(x, x_f), \text{Pay}(x_f), S_f(x, x_f)$ (12)

通过制约 SLD 否认的方式应答 $Q^{(0)}$: 为每一个 $Q^{(i)}$ 找到一个 $\theta^{(i+1)}$, 构建从 $E^{(i+1)}$ 到 $Q^{(i)}$ 的 $Q^{(i+1)}$, 直到为制约空子句。

令 $E^{(1)}$ 为式 (8), 采用置换 $\theta^{(1)} = \{x / \text{'ex0806012'}\}$ 构建从 $E^{(1)}$ 到 $Q^{(0)}$ 的一个消解为 $Q^{(1)} = \leftarrow \text{DicretBuy}(\text{'ex0806012'}) \& \text{Samsug}(y), \text{BuyProduct}(\text{'ex0806012'}), \text{hasPart}(\text{'ex0806012'}, y)$ (13)

令 $E^{(2)}$ 为式 (7), 采用置换 $\theta^{(2)} = \{x / \text{'ex0806012'}, y / y_1\}$ 构建从 $E^{(2)}$ 到 $Q^{(1)}$ 的一个消解为

$Q^{(2)} = \leftarrow O(y_1) \& \text{BuyProduct}(\text{'ex0806012'}), \text{hasPart}(\text{'ex0806012'}, y), \text{Samsug}(y), \text{hasPart}(\text{'ex0806012'}, y_1), \text{Pay}(y_1)$ (14)

令

$E(3) = \text{hasPart}(\text{'ex0806012'}, \text{'E908'}), \text{Samsug}(\text{'E908'}), \text{hasPart}(\text{'ex0806012'}, \text{'espay0806012'}), \text{Pay}(\text{'espay0806012'})$ (15)

采用置换 $\theta^{(3)} = \{y / \text{'E908'}, y_1 / \text{'espay0806012'}\}$ 构建从 $E^{(3)}$ 到 $Q^{(2)}$ 的一个消解为

$Q^{(3)} = \leftarrow \& \text{Buy Product}(\text{'ex0806012'})$ (16)

这时 $Q^{(3)}$ 为制约空子句, 可直接提交到描述逻辑推理机进行实例检测, 或可利用 B_{OMS} 逻辑知识库对应的一阶规则进行消解: 令 $E^{(4)}$ 为式 (12), 采用置换构建从到的一个消解为

$Q^{(4)} = \leftarrow \& \text{Shopping}(\text{'ex0806012'}), \text{Pay}(\text{'espay0806012'}), \text{hasPart}(\text{'ex0806012'}, \text{'espay0806012'}), S_f(\text{'ex0806012'}, \text{'espay0806012'})$ (17)

S_f 与 hasPart 等价, 描述复合事件和原子事件之间的整分关系, 有 $\text{'ex0806012'}_f = \text{'espay0806012'}$. 因此 $Q^{(5)}$ 子句中消去 S_f 后其原子公式全部来自于知识库 B_{OMS} , 因此 $Q^{(5)}$ 为空子句, 由此证明 $B \cup H \cup A_{\text{ex0806012}} \vdash q(\text{'ex0806012'})$.

5.2 计算复杂度分析

日志本体模式发现建立在 DL-safe^L 基础上, 其描述逻辑基础和一阶逻辑分别为 $\text{ALCHIQ}(D)$ 以及 DL-safe . 的计算复杂度在考虑 TBox 后上升为 EXP-TIME - 完全, 基于选言 Datalog 的 DL-safe 程序复杂度是 $\text{EXPTIME}^{[2]}$. 作为结论, 我们认定 DL-safe^L 基础挖掘日志本体的方法限定在 EXPTIME 范畴中. 当然在以后将描述逻辑基础扩展到 $\text{SHOIN}(D)$, 解决标称事件、逆角色以及数量限制等当前工作中的难题后, 计算复杂度将上升到 NEXPTIME .

6 结束语

在日志本体挖掘过程中, 采用结合描述逻辑和 DL-safe 规则的混合逻辑推理系统能提高知识表达的适应性和演绎能力, 并且其推理过程是可判定的. 因此文中结合 DL-Safe 规则和日志本体构建混合知识库, 并在此基础上提出一种基于 ILP 的日志本体用户访问模式挖掘方法, 从基准概念出发构造有效的模式树, 通过制约 SLD 否认验证访问模式是否关于知识库蕴含既定的观察集, 达到发现其中用户访问模式的目的. 该方法充分利用了日志本体事件间特定的语义关系简化了算法的复杂度, 提高了学习效率.

参考文献:

[1] 陈沈焰, 吴军华. 基于本体的概念语义相似度计算及其应用[J]. 微电子学与计算机, 2008, 25(12): 96 - 99.

(下转第 37 页)

4 结束语

文中介绍了垃圾收集器实时化的发展过程,介绍了在 GC 实时化上做的一些研究,对单处理器 GC 性能进行改进,降低了系统内存需求,提出了 EDF 调度 GC 的思想使系统更具灵活性,同时内存需求更小,在分布式 GC 方面提出了真正适合大规模分布式系统的软实时算法,能以最短时间及时回收所有垃圾。

参考文献:

- [1] 马明全,周明全,耿国华. 分布式大规模交通网络实时系统[J]. 微电子学与计算机,2006,23(12):56-59.
- [2] 赵春云,郭煦,金戟,等. JAVA 语言及其虚拟机技术探讨[J]. 微电子学与计算机,2006,23(12):15-18.
- [3] Baker H G. List processing in real time on a serial computer[J]. Communications of the ACM, 1978, 21(4): 280-294.
- [4] Henriksson R. Scheduling garbage collection in embedded systems[D]. Sweden: Lund University, 1998.
- [5] Kim Taehyoun, Naehyuck Chang, Heonshik Shin. Bounding worst case garbage collection time for embedded real-time systems[C]// Real-Time Technology and Applications - Proceedings. Washington, USA, 2000: 46-55.
- [6] Martin Rudalics. Distributed copying garbage collection[J]. ACM Symposium on Lisp and Functional Programming, 1986(8): 364-372.
- [7] Xian Yuqiang, Xiong Guangze. Minimizing memory requirement of real-time systems with concurrent garbage collector[J]. ACM SIGPLAN Notices, 2005, 40(3): 40-48.
- [8] Zhang Ning, Xiong Guangze. Minimizing GC work by analysis of live objects[J]. ACM SIGPLAN Notices, 2006, 41(3): 20-29.
- [9] 张宁,熊光泽. 用 EDF 调度实时任务和 GC[J]. 航空学报, 2008,29(5):1226-1232.
- [10] Lang B, Queninnec C, Piquer J. Garbage collecting the world[C]// Proceedings of the Nineteenth ACM Symposium on Principles of Programming Languages. New York: ACM Press, 1992: 39-50.

作者简介:

- 张宁** 男,(1975-),博士研究生.研究方向为嵌入式实时系统、动态内存管理.
- 张春宇** 男,(1970-),讲师.研究方向为计算机控制.
- 杨霞** 女,(1978-),讲师.研究方向为实时可信计算、防危安全.
- 熊光泽** 男,(1938-),博士生导师,教授.研究方向为嵌入式实时系统、可信计算.
- 廖勇** 男,(1976-),博士,讲师.研究方向为实时计算、实时调度.

(上接第 33 页)

- [2] Sun M, Chen B, Zhou M T. An ILP approach to mine the association rules on log ontology[C]// Proceedings of ICA-CIA 2008. Chengdu, China: IEEE Press, 2008: 274-278.
- [3] Motik B, Sattler U, Studer R. Query answering for OWL-DL with rules[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2005,3(1):41-60.
- [4] 张玉花,张娜. 面向对象的本体创建方法[J]. 微电子学与计算机,2008,25(7):65-67.
- [5] Hustadt U, Motik B, Sattler U. Reducing SHIQ-description logic to disjunctive datalog programs[C]// Proceedings of the 9th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2004). Whistler, Canada, 2004:152-162.
- [6] Francesca A L. An ILP approach to ontology refinement for the semantic web[C]//Proceedings of the 4th European Semantic Web Conference, LNCS, 2007. Berlin: Springer,2007:117-129.
- [7] J'ozefowska J, Lawrynowicz A, Lukaszewski T. A study of the SEMINTEC approach to frequent pattern mining [C]// Proceedings of PriCKL 2007, ECML/PKDD' 2007. Warsaw, Poland, 2007:41-52.
- [8] Francesca A L. Principles of inductive reasoning on the semantic web: a framework for learning in AL-log[C]// Proceedings of the 3rd International Workshop on Principles and Practice of Semantic Web Reasoning. Heidelberg: Springer, Berlin, 2005:118-132.

作者简介:

- 孙明** 男,(1978-),博士研究生.研究方向为知识发现、语义 Web 挖掘以及本体学习.
- 陈波** 男,(1977-),博士.研究方向为知识发现以及粒度粗糙集理论.
- 周明天** 男,(1939-),教授,博士生导师.研究方向为中间件技术、知识发现以及网络计算.