

基于模糊描述逻辑的模糊 XML 模型的表示与推理

张 富¹⁾ 严 丽²⁾ 马宗民¹⁾ 王 星¹⁾

¹⁾(东北大学信息科学与工程学院计算机应用技术研究所 沈阳 110819)

²⁾(东北大学软件学院软件工程研究所 沈阳 110819)

摘 要 通过分析描述逻辑与 XML 之间的关系,该文进一步研究了基于模糊描述逻辑的模糊 XML 模型的表示与推理.首先提出了模糊 XML 模型(包括模糊 XML 文档和模糊 DTD)的形式化定义.然后针对模糊 XML 模型的特点和推理需求,提出了一种新的模糊描述逻辑 f -ALCQ_{wf-reg},给出了它的语法、语义、知识库以及推理算法.在此基础上,研究了基于 f -ALCQ_{wf-reg} 的模糊 XML 模型的表示、推理以及查询问题,具体包括:实现了从模糊 XML 模型到 f -ALCQ_{wf-reg} 知识库的转化;基于转化后得到的 f -ALCQ_{wf-reg} 知识库,进一步研究了如何利用 f -ALCQ_{wf-reg} 的推理机制对模糊 XML 模型的推理任务(如一致性、包含性、等价性以及不相交性)进行推理;最后简要讨论了如何利用上述转化和推理结果来更有效地支持模糊 XML 文档库的查询问题.

关键词 模糊 XML 模型;文档和结构;模糊描述逻辑;表示;推理

中图法分类号 TP18 DOI号: 10.3724/SP.J.1016.2011.01437

Representation and Reasoning of Fuzzy XML Model with Fuzzy Description Logic

ZHANG Fu¹⁾ YAN Li²⁾ MA Zong-Min¹⁾ WANG Xing¹⁾

¹⁾(Institute of Computer Application Technology, College of Information Science and Engineering, Northeastern University, Shenyang 110819)

²⁾(Institute of Software Engineering, School of Software, Northeastern University, Shenyang 110819)

Abstract By analyzing the relationships between description logics and XML models, the paper investigates the representation and reasoning of fuzzy XML models with fuzzy description logics. The formal definition of fuzzy XML models (including fuzzy XML documents and fuzzy DTDs) is proposed. Then, aiming at the characteristics and reasoning requirement of fuzzy XML models, a kind of new fuzzy description logic called f -ALCQ_{wf-reg} is developed, and the syntax, semantics, knowledge base and reasoning algorithm are given. On this basis, representation and reasoning of fuzzy XML models with f -ALCQ_{wf-reg} is investigated, including: Fuzzy XML models are translated into f -ALCQ_{wf-reg} knowledge bases; Based on the translated f -ALCQ_{wf-reg} knowledge bases, how to reason on fuzzy XML models (e. g., conformance, inclusion, equivalence and disjointness) through the reasoning mechanism of f -ALCQ_{wf-reg} is investigated; Also it is briefly discussed that how to support query processing over a document base more efficiently based on the translation and reasoning results above.

Keywords fuzzy XML model; document and structure; fuzzy description logic; representation; reasoning

收稿日期:2010-08-23;最终修改稿收到日期:2011-05-17. 本课题得到国家自然科学基金(60873010,61073139)、教育部新世纪优秀人才支持计划(NCET-05-0288)和教育部中央高校基本科研业务费专项(N090504005, N090604012, N100604017)资助. 张 富,男,1984年生,博士研究生,主要研究方向为语义 Web、描述逻辑和 XML. E-mail: zhangfu216@gmail.com; zhangfu216@126.com. 严 丽,女,1964年生,博士,副教授,主要研究方向为数据库与 XML. 马宗民(通信作者),男,1965年生,博士,教授,博士生导师,中国计算机学会高级会员,主要研究领域为智能数据与知识工程. E-mail: mazongmin@ise.neu.edu.cn. 王 星,男,1983年生,博士,主要研究方向为语义 Web 和规则.

1 引 言

近年来,随着 Web 的广泛使用,XML 已经成为网络上各种系统和数据库进行数据描述和交换的事实上的标准,由此产生了有关 XML 数据管理的需求,如 XML 文档的表示、推理、整合以及查询等,并成为了当前学术界和工业界研究和开发的热点^[1-2].

许多研究工作指出,研究 XML 文档的表示和推理将有助于 XML 的设计、整合以及查询优化,同时也将有利于提高 Web 上信息检索的准确性和检索处理的有效性等^[3-10].

在众多知识表示和推理方法中,描述逻辑引起了人们广泛的关注.描述逻辑(Description Logic, DL)是一种适合表示关于概念和概念层次结构的知识表示语言,具有形式化语义、很强表达能力和可判定推理算法,且在服务匹配、语义 Web 以及数据库等领域得到了广泛的应用^[9,11].尤其值得一提的是,基于描述逻辑有效的推理机制,许多研究工作利用描述逻辑对各种数据模型(如 XML 模型^[3,5-9]、ER 模型^[12]和面向对象模型^[12]等)进行表示与推理(详见第 5 节).如果能够将数据模型转化为描述逻辑知识库,则可以利用描述逻辑推理机制对数据模型的相关推理问题(如 ER 模型的冗余性、XML 模型的一致性和等价性)进行推理,从而克服目前数据模型推理中存在的不足(如推理需要手工进行,且推理效率和可靠性不高等),也将有利于解决数据建模和查询优化等相关问题^[3-9,12].然而,上述已有研究工作只能处理精确知识,不能处理模糊知识.

在现实世界中,大量的不精确和不确定信息存在于各个应用领域(如数据库^[13-14]和语义 Web^[15]等),而 XML 作为 Web 环境下数据表示和交换的标准,如何扩展以实现不精确和不确定数据的管理变得十分必要,为此,文献^[16-21]研究了如何在 XML 中表示和处理不精确和不确定信息.同样,为了使描述逻辑能够表示和推理模糊知识,许多研究者对描述逻辑进行了模糊化扩展,提出了各种形式的模糊描述逻辑(fuzzy DLs),如 FALC^[22]、FALNUI^[23-24]、EFALCN^[25]、EFALC_{R+}^[26]、FSHOIQ^[27]、*f*-SHIN^[28]以及 FDLR^[29]等,并将它们应用到医疗诊断、信息检索和数据库等领域^[15](有关模糊 XML 模型和模糊描述逻辑的详细介绍见第 5 节).应当指出的是,已存在的模糊描述逻辑并不能很好地表示模糊 XML 模型,且有关利用模糊描述逻辑对模糊 XML

模型进行表示与推理的研究,目前尚未见有文章发表.如果能够将模糊 XML 模型转化为模糊描述逻辑知识库,则可以利用模糊描述逻辑推理机制对模糊 XML 模型的推理问题(如两个模糊 XML DTDs 是否等价/包含,或一个模糊 XML 文档是否与给定的模糊 XML DTD 相一致等)进行推理,从而有利于模糊 XML 建模、推理以及查询优化等相关问题的解决.

为了满足模糊 XML 数据管理的需求,本文研究基于模糊描述逻辑的模糊 XML 模型的表示、推理以及查询:

首先提出了模糊 XML 模型(包括模糊 XML 文档和文档结构模糊 DTD)的形式化定义.

然后通过分析 XML 模型与经典描述逻辑之间的对应关系^[3,5-9],并针对模糊 XML 模型的特点,对经典描述逻辑 ALCIQ_{reg}^[30-31]进行模糊化扩展,提出了一种新的模糊描述逻辑 *f*-ALCQ_{wf-reg},给出了它的语法、语义、知识库以及推理算法.

在此基础上,研究了基于 *f*-ALCQ_{wf-reg} 的模糊 XML 模型的表示、推理以及查询问题,具体包括:实现了从模糊 XML 模型到 *f*-ALCQ_{wf-reg} 知识库的转化,并给出了转化的正确性证明和转化实例;基于转化后得到的 *f*-ALCQ_{wf-reg} 知识库,进一步研究了如何利用 *f*-ALCQ_{wf-reg} 的推理机制对模糊 XML 模型的推理问题(如一致性、包含性和等价性等)进行推理;最后简要讨论了如何利用上述转化和推理结果来更有效地支持文档库的查询问题.

需要指出的是,XML 文档的结构能够通过 XML DTD 和 XML Schema 来表示^[2],本文主要集中在 XML DTD,它作为一种 XML 文档结构最为传统的表示机制,被广泛应用于许多系统(YAT、XMF 和 MIX 等^[32]),且当前许多有关模糊 XML 的研究工作都是集中在扩展 XML DTD 上(详见文献^[21]).基于此,本文也考虑使用 XML DTD 的模糊化扩展作为模糊 XML 文档结构的表示形式.

本文第 2 节给出模糊 XML 文档和模糊 DTD 的形式化定义;第 3 节提出一种新的模糊描述逻辑 *f*-ALCQ_{wf-reg};第 4 节研究基于 *f*-ALCQ_{wf-reg} 的模糊 XML 模型的表示、推理以及查询,即实现从模糊 XML 模型到 *f*-ALCQ_{wf-reg} 知识库的转化,并利用 *f*-ALCQ_{wf-reg} 的推理机制研究模糊 XML 模型的推理问题以及查询优化问题;第 5 节详细介绍相关工作;最后给出总结及下一步的研究工作.

2 模糊 XML 模型

本节首先介绍如何在 XML 模型(包括 XML 文档和文档结构 DTD)中表示模糊信息^[16-18,20],然后,提出模糊 XML 文档和模糊 DTD 的形式化定义,以便更好地在模糊 XML 模型与模糊描述逻辑之间建立对应关系。

2.1 模型 XML 文档和模糊 DTD

与 XML 文档^[2]相类似,一个模糊 XML 文档(fuzzy XML document)主要由有限对互相匹配的标签(开始标签和结束标签)以及标签内的元素(element)组成,且每个元素可有其相应的属性(attribute)。在模糊关系数据库中存在两类模糊性^[13,17]:一类是与元组的成员度相关的模糊性,表示元组属于一个关系的可能性;另一类是用可能性分布表示的属性值,表示不确定属性的精确值,只能确定属性可能取值的范围。

与模糊关系数据库相类似,在模糊 XML 文档中也存在两类模糊性^[17-18,20]:

第 1 类是元素的模糊性,可以用成员度与这样的元素关联(如图 1 中的第 4 行)。由于在一个模糊 XML 文档中,元素可以嵌套出现,因此与一个元素相关联的成员度表明了现实世界状态中,该元素(包括以该元素为根元素的子孙元素)属于其父元素的可能性。

第 2 类是元素属性值的模糊性,可以用可能性分布表示,且该可能性分布分为模糊析取(disjunctive)可能性分布和模糊合取(conjunctive)可能性分布(如图 1 中的第 8~15、18~25 行)。在现实世界中,某些属性值可能是不确定的,只能用可能性分布来表示。例如,一个人“Tom Smith”的年龄是一个单值非负的整数,如果该年龄当前是未知的,则可以用模糊析取可能性分布 $\{0.8/23, 0.6/25\}$ 来表示其年龄信息;相反,“Tom Smith”可以有多个 e-mail 地址,且 E-mail 地址为“TS@yahoo.com”的可能性是 0.6,为“TS@sina.com”的可能性是 0.85,则可以用模糊合取可能性分布来表示其地址信息。

下面通过一个例子来详细说明上述两类模糊性,图 1 给出了一个包含模糊信息的模糊 XML 文档片段 d_1 ^[17],描述的是一个国家所在地区的大学情况,所关注的是美国密歇根州的底特律(Detroit)城市地区的大学情况:

(1) Detroit Mercy University 位于 Detroit 市

中心,它属于 Detroit 地区大学的可能性是 1.0。对于可能性为 1.0 的元素,可能性标签 $\langle \text{Val Poss} = 1.0 \rangle \dots \langle / \text{Val} \rangle$ 可以省略(如图 1 中的第 30~32 行);

(2) Oakland University 位于密歇根州附近的一个郡 Oakland,对于 Oakland University 是否属于 Detroit 地区的大学则取决于如何定义 Detroit 地区,是仅包括 Detroit 城市还是也包括周边地区。假设当前对 Detroit 地区的确切定义是未知的,则 Oakland University 属于 Detroit 地区大学的可能性将不为 1.0,例如指定为 0.8(如图 1 中的第 4 行);上述 1~2 属于第 1 类模糊性,即元素的模糊性。

(3) 对于学生 Tom Smith,他的“age”属性值是模糊的。由于该属性值是一个单值非负的整数,因此可以用一个析取可能性分布 $\{0.8/23, 0.6/25\}$ 来表示(如图 1 中的第 8~15 行);

(4) 对于学生 Tom Smith,他的“E-mail”属性值也是模糊的。由于该属性值可以是多个字符串,因此可以用一个合取可能性分布来表示(如图 1 中的第 18~25 行);上述 3~4 属于第 2 类模糊性,即元素属性值的模糊性。

```

1. <universities>
2.   <university>
3.     <UName>Oakland University</UName>
4.     <Val Poss=0.8>
5.       <student>
6.         <sname>Tom Smith</sname>
7.         <age>
8.           <Dist type="disjunctive">
9.             <Val Poss=0.8>
10.              <age_value>23</age_value>
11.            </Val>
12.            <Val Poss=0.6>
13.              <age_value>25</age_value>
14.            </Val>
15.          </Dist>
16.        </age>
17.        <email>
18.          <Dist type="conjunctive">
19.            <Val Poss=0.60>
20.              <em_value>TS@yahoo.com</em_value>
21.            </Val>
22.            <Val Poss=0.85>
23.              <em_value>TS@tom.com</em_value>
24.            </Val>
25.          </Dist>
26.        </email>
27.      </student>
28.    </university>
29.  </university>
30. <university>
31.   <UName>Detroit Mercy University</UName>
32. </university>
33. </universities>

```

图 1 一个模糊 XML 文档片段 d_1

从图 1 可以看出,为了表示 XML 文档中的模糊信息,几个模糊构造子被引入:

(1) 一个取值为 $[0, 1]$ 的可能性属性“Poss”首先被引入,它与一个称作“Val”的模糊构造子共同用于说明一个给定元素存在于 XML 文档的可能性.以图 1 中的第 4 行为例, $\langle \text{Val Poss}=0.8 \rangle$ 表明给定的大学“Oakland University”(包括其子元素)属于 Detroit 地区大学的可能性为 0.8.

(2) 利用标签对 $\langle \text{Val Poss} \dots \rangle \dots \langle / \text{Val} \rangle$ 可以表示元素的可能性,而可能性分布用于表示元素的模糊属性值.为此,需要引入一个称作“Dist”的模糊构造子,以表示一个可能性分布.一个 Dist 元素典型地有多个 Val 元素作为孩子元素,每一个孩子元素带有一个相关联的可能度.此外,从图 1 可知,存在两种类型的可能性分布,因此 Dist 构造子应当指明可能性分布的类型是析取的还是合取的,则需要引入“Type”属性,如图 1 中的 8~25 行.

新的模糊构造子的引入,必将影响 XML 文档结构的定义.上述模糊 XML 文档的结构可以通过模糊 DTD (fuzzy Document Type Definition) 来描述.一个模糊 DTD 是一个元素类型定义的集合,指定了与该模糊 DTD 相一致的模糊 XML 文档必须满足的结构条件约束.下面介绍如何通过修改 XML 文档的 DTD 模式,容纳这些新引入的模糊构造子.此外,文献[2]指出一个元素的属性可以通过把属性改写为嵌套元素的形式来表达,例如 $\langle \text{Val Poss}=0.8 \rangle \dots \langle / \text{Val} \rangle$ 可以通过嵌套元素 $\langle \text{Val} \rangle \langle \text{Poss} \rangle 0.8 \langle / \text{Poss} \rangle \langle / \text{Val} \rangle$ 来表示,于是,本文后面的工作将采用嵌套元素来代替属性以便更容易地在模糊 XML 模型与模糊描述逻辑之间建立对应关系.

下面给出模糊 DTD 中元素类型的定义形式,首先给出一般元素的定义形式(与经典 DTD^[2] 相类似):

```
<!ELEMENT 元素 1 (元素 2?*+)
```

//该定义表示“元素 1”包含“元素 2”,且“元素 2”的出现次数限制如下: ? 号表示出现 0 次或 1 次, * 号表示出现 0 次或多次, + 表示出现 1 次或多次, 缺省情况下则表示出现且仅出现 1 次.

```
<!ELEMENT 元素 1 (元素 2,元素 3,...)
```

//表示“元素 1”按顺序包含“元素 2”和“元素 3”等.

```
<!ELEMENT 元素 1 (元素 2|元素 3|...)
```

//表示“元素 1”或包含“元素 2”,或包含“元素 3”等.

```
<!ELEMENT 元素 1 (#PCDATA)
```

// #PCDATA 是元素的原子类型^[1-2].

```
<!ELEMENT 元素 1 (empty)
```

//表示元素 1 是一个空元素,不包含任何内容.

然后给出上述新引入构造子的模糊 DTD 定义

形式,先给出 Val 元素的定义:

```
<!ELEMENT Val (Poss,original_definition)
```

```
<!ELEMENT Poss (#PCDATA|"1.0").
```

```
<!ELEMENT original_definition (#PCDATA|any)
```

其中 #PCDATA 是元素的原子类型, any 表示模糊 DTD 中任意序列的元素类型定义;然后给出 Dist 元素的定义:

```
<!ELEMENT Dist (Type,Val+)
```

```
<!ELEMENT Type (disjunctive|conjunctive).
```

使用上面给出的 DTD 扩展形式,可以构建出模糊 XML 文档相应的模糊 DTD.图 2 给出与图 1 中模糊 XML 文档 d_1 相应的模糊 DTD D_1 .

```
<!ELEMENT universities (university * )
<!ELEMENT university (UName,Val+)
<!ELEMENT UName (#PCDATA)
<!ELEMENT Val (Poss,student * )
<!ELEMENT student (sname?,age?,email?)
<!ELEMENT sname (#PCDATA)
<!ELEMENT age (Dist)
<!ELEMENT Dist (Type,Val+)
<!ELEMENT Type (disjunctive)
<!ELEMENT Val (Poss,age_value)
<!ELEMENT age_value (#PCDATA)
<!ELEMENT email (Dist)
<!ELEMENT Dist (Type,Val+)
<!ELEMENT Type (conjunctive)
<!ELEMENT Val (Poss,email_value)
<!ELEMENT Poss (#PCDATA|"1.0")
<!ELEMENT email_value (#PCDATA)
```

图 2 图 1 中模糊 XML 文档 d_1 相应的模糊 DTD D_1

2.2 模糊 XML 文档和模糊 DTD 的形式化定义

基于 2.1 节中的模糊 XML 模型以及文献[3-10] 中有关经典 XML 的介绍,本节提出模糊 XML 文档和模糊 DTD 的形式化定义.

从图 1 可知,一个模糊 XML 文档主要由一系列元素组成(其中属性可以由嵌套元素来表示),且一个元素 E 由一个起始标签 $\langle E \rangle$ 、元素内容和一个结束标签 $\langle /E \rangle$ 构成.用 \mathbf{E} 表示元素类型的集合, \mathbf{T} 表示终止符(如 #PCDATA).下面给出模糊 XML 文档的形式化定义.

定义 1(模糊 XML 文档). 定义在 \mathbf{T} 和 \mathbf{E} 上的模糊 XML 文档实例的集合 $d_{\mathbf{T},\mathbf{E}}$ 可以通过如下方法归纳得到

(1) 如果 $d \in \mathbf{T}$ 是一个终止符,则 $d \in d_{\mathbf{T},\mathbf{E}}$;

(2) 如果 d 是一个形如 $\langle E \rangle d_1, \dots, d_k \langle /E \rangle$ 的序列,其中 $E \in \mathbf{E}$ 是一个元素类型, $d_1, \dots, d_k \in d_{\mathbf{T},\mathbf{E}}$, 则 $d \in d_{\mathbf{T},\mathbf{E}}$.

其中:

终止符 \mathbf{T} 表示基本类型 #PCDATA,用于表示

模糊 XML 文档中元素的内容;

元素类型 E 包括基本元素类型 E_g (如图 1 中的 university 和 student 等) 和用于表示模糊信息的特殊元素类型 E_s (如图 1 中的 Val 和 Dist 等);

为了简化, 用 S 表示 $T \cup E_g \cup E_s$.

从图 2 可知, 一个模糊 DTD 是一个元素类型定义 $E \rightarrow \alpha$ 的集合, 其中 E 表示被定义的元素类型 (如图 2 中的 university 和 Dist 等), α 是定义在 S 上的与 E 相对应的内容模型 (content model, 如图 2 中的 $\langle \text{UName}, \text{Val} \rangle$ 和 $\langle \text{Type}, \text{Val} \rangle$ 等). 下面给出模糊 DTD 的形式化定义.

定义 2 (模糊 DTD). 一个模糊 DTD D 是一对 (P, r) , 其中 P 是元素类型定义 $E \rightarrow \alpha$ 的集合, r 是根元素类型用于指定某一特定的模糊 DTD, 每一个元素类型定义 $E \rightarrow \alpha$ 是通过下面的语法定义而成:

$\alpha ::= S | \text{empty} | (\alpha_1 | \alpha_2) | (\alpha_1, \alpha_2) | \alpha? | \alpha^* | \alpha+ | \text{any}$,

其中 $S = T \cup E_g \cup E_s$, empty 表示空元素, “|” 表示元素的或操作, “,” 表示元素的顺序操作, “?” 表示 0 或 1 次, “*” 表示 0 或多次, “+” 表示 1 或多次, any 表示模糊 DTD 中任意序列的元素类型定义.

3 模糊描述逻辑 $f\text{-ALCQ}_{\text{wf-reg}}$

为了表示和推理模糊 XML 模型, 需要使用一种模糊描述逻辑. 通过分析 XML 与经典描述逻辑之间的关系^[3,5-9], 并针对第 2 节中提到的模糊 XML 模型的特点, 本节对经典描述逻辑 $\text{ALCIQ}_{\text{reg}}$ ^[30-31] 进行模糊化扩展, 提出一种新的模糊描述逻辑 $f\text{-ALCQ}_{\text{wf-reg}}$, 给出它的语法、语义、知识库以及推理算法.

3.1 语法和语义

下面通过定义 3 和 4 分别给出模糊描述逻辑 $f\text{-ALCQ}_{\text{wf-reg}}$ 的语法和语义.

定义 3 (语法). 与经典描述逻辑^[3,8,30-31] 的语法相类似, $f\text{-ALCQ}_{\text{wf-reg}}$ 概念和关系按如下规则定义:

$C, D \rightarrow \top | \perp | A | \neg C | C \cap D | C \cup D | \forall R.C | \exists R.C | \omega f(R) | \geq k R.C | \leq k R.C$;

$R \rightarrow P | id(C) | R_1 \cup R_2 | R_1 \circ R_2 | R^* | R^+$.

其中:

A 表示原子概念, P 表示原子关系, C 和 D 表示任意概念, R 表示任意关系, $k \in N$ (自然数集);

$f\text{-ALCQ}_{\text{wf-reg}}$ 允许关系正规表达式 (下标 “reg”)

的出现, 即关系同一、关系的并、连接、自反传递闭包和传递闭包;

概念构造子 $\omega f(R)$ 用于表示关系链 R 中的起始个体;

关系同一构造子 $id(C)$ 允许构建一个关系使得概念 C 的个体实例通过该关系指向它本身.

模糊描述逻辑 $f\text{-ALCQ}_{\text{wf-reg}}$ 的语义是根据 Zadeh 对模糊集的语义解释方法^[33] 和经典描述逻辑的语义解释方法给出的, 下面给出 $f\text{-ALCQ}_{\text{wf-reg}}$ 的语义.

定义 4 (语义). $f\text{-ALCQ}_{\text{wf-reg}}$ 的语义由模糊解释 $FI = (\Delta^{FI}, \cdot^{FI})$ 给出, 其中 Δ^{FI} 是解释论域, \cdot^{FI} 是解释函数, 并且 \cdot^{FI} 满足:

(1) 对任意个体 d , 解释函数 \cdot^{FI} 把 d 映射到 Δ^{FI} 的一个元素, 即 $d^{FI} \in \Delta^{FI}$;

(2) 对任意个体 c 和 d , 如果 $c \neq d$, 则 $c^{FI} \neq d^{FI}$;

(3) 解释函数 \cdot^{FI} 将概念名 A 映射为一个隶属函数, 即 $A^{FI}: \Delta^{FI} \rightarrow [0, 1]$;

(4) 解释函数 \cdot^{FI} 将关系 R 映射为一个隶属函数, 即 $R^{FI}: \Delta^{FI} \times \Delta^{FI} \rightarrow [0, 1]$;

(5) $f\text{-ALCQ}_{\text{wf-reg}}$ 概念的语义解释如下 (对任意个体 $c_i, d_i, d, d' \in \Delta^{FI}$):

$\top^{FI}(d) = 1, \perp^{FI}(d) = 0$;

$(\neg C)^{FI}(d) = 1 - C^{FI}(d)$;

$(C \cap D)^{FI}(d) = \min\{C^{FI}(d), D^{FI}(d)\}$;

$(C \cup D)^{FI}(d) = \max\{C^{FI}(d), D^{FI}(d)\}$;

$(\forall R.C)^{FI}(d) =$

$\inf_{d' \in \Delta^{FI}} \{\max\{1 - R^{FI}(d, d'), C^{FI}(d')\}\}$;

$(\exists R.C)^{FI}(d) = \sup_{d' \in \Delta^{FI}} \{\min\{R^{FI}(d, d'), C^{FI}(d')\}\}$;

$\omega f(R)^{FI}(d_0) = \begin{cases} 1, & \forall d_1, d_2, \dots (ad \text{ infinitum}) \\ \exists i \geq 0: R^{FI}(d_i, d_{i+1}) < \beta \\ 0, & \text{其它;} \end{cases}$

$(\geq n R.C)^{FI}(d) =$

$\sup_{c_1, \dots, c_n \in \Delta^{FI}} \bigwedge_{i=1}^n \{\min\{R^{FI}(d, c_i), C^{FI}(c_i)\}\}$;

$(\leq n R.C)^{FI}(d) =$

$\inf_{c_1, \dots, c_{n+1} \in \Delta^{FI}} \bigvee_{i=1}^{n+1} \{\max\{1 - R^{FI}(d, c_i), C^{FI}(c_i)\}\}$.

其中:

概念构造子 $\omega f(R)$ 被解释为关系链 R 中的起始个体, 隶属度 $\beta \in (0, 1]$ 是一个给定的临界值, 用于限定关系链的有限性, 例如, 若 β 取值为 0.01, 则 $R^{FI}(d_i, d_{i+1}) < 0.01$ 表示个体 d_i 与 d_{i+1} 之间不存

在关系 R , 即 d_i 是该关系链的最后一个节点.

f -ALCQ_{wf-reg} 关系构造子的语义: $id(C)^{FI}(d, d) = C^{FI}(d)$. 由文献[3]可知 $(R^*)^{FI} = (R^{FI})^* = \bigcup_{n \geq 0} (R^{FI})^n$ 和 $(R^+)^{FI} = (R \circ R^*)^{FI}$, 于是 R^* 和 R^+ 的语义可以根据下面的关系连接和关系并的语义给出. 给定两个关系 $R_1: X \times Y \rightarrow [0, 1]$ 和 $R_2: Y \times Z \rightarrow [0, 1]$, 则 R_1 与 R_2 的连接被解释为 $(R_1 \circ R_2)^{FI}(x, z) = \sup_{y \in Y} \{\min\{R_1^{FI}(x, y), R_2^{FI}(y, z)\}\}$; 给定两个关系 $R_1: X \times Y \rightarrow [0, 1]$ 和 $R_2: X \times Y \rightarrow [0, 1]$, 两个关系的并被解释为 $(R_1 \cup R_2)^{FI}(x, y) = \max\{R_1^{FI}(x, y), R_2^{FI}(x, y)\}$.

3.2 知识库

一个 f -ALCQ_{wf-reg} 知识库(记作 KB)包含 TBox TB 和 ABox AB , 其中 TBox 定义了应用领域的内涵知识, 而 ABox 定义了应用领域的内延知识.

定义 5(TBox). 一个 f -ALCQ_{wf-reg} TBox TB 是模糊概念公理的有限集合. 模糊概念公理包括: 概念包含 $A \subseteq C$ 和概念等价 $A \equiv C$, 其中 A 表示概念名, C 表示任意概念. 模糊概念公理的语义如下:

- (1) $A \subseteq C$, 当且仅当对任意 $d \in \Delta^{FI}$, $A^{FI}(d^{FI}) \leq C^{FI}(d^{FI})$;
- (2) $A \equiv C$, 当且仅当 $A \subseteq C$ 和 $C \subseteq A$, 即对任意 $d \in \Delta^{FI}$, $A^{FI}(d^{FI}) = C^{FI}(d^{FI})$.

当一个 TBox 中的概念公理仅包括 $A \subseteq C$ 或 $A \equiv C$, 且不存在概念名 A 直接或间接引用其自身, 则称该 TBox 为简单 TBox; 否则, 当 TBox 中允许出现通用包含公理 $C \subseteq D$ 时(其中 C 和 D 为任意概念), 则称该 TBox 为一般 TBox^[22, 28]. 如果一个模糊解释 FI 满足 TBox TB 中的每一条模糊概念公理, 那么 FI 满足 TBox TB , 也称 FI 是 TB 的一个模型. 如果 TBox TB 存在一个模型, 则称 TB 是可满足的.

定义 6(ABox). 一个 f -ALCQ_{wf-reg} ABox AB 是一个模糊断言公式的有限集合. 模糊断言公式包括: 概念断言 $\langle C(d) \infty n \rangle$ 和角色断言 $\langle R(c, d) \infty n \rangle$, 其中 $\infty \in \{\geq, >, \leq, <\}$, $n \in [0, 1]$. 模糊断言的语义如下:

- (1) 一个模糊解释 FI 满足公式 $\langle C(d) \infty n \rangle$, 当且仅当 $C^{FI}(d^{FI}) \infty n$;
- (2) 一个模糊解释 FI 满足公式 $\langle R(c, d) \infty n \rangle$, 当且仅当 $R^{FI}(c^{FI}, d^{FI}) \infty n$.

如果一个模糊解释 FI 满足 ABox AB 中的每一条模糊断言公式, 那么 FI 满足 ABox AB , 也称

FI 是 AB 的一个模型. 如果 ABox AB 存在一个模型, 则称 AB 是可满足的.

定义 7(知识库 KB). 一个 f -ALCQ_{wf-reg} 知识库 KB 由 TBox TB 和 ABox AB 构成, 记作 $KB = \langle TB, AB \rangle$.

一个模糊解释 FI 满足 KB (即 FI 是 KB 的一个模型), 当且仅当 FI 是 TB 和 AB 的模型. 一个 KB 是可满足的, 当且仅当存在 KB 的模型 FI .

3.3 推理问题和算法

下面首先介绍 f -ALCQ_{wf-reg} 的推理问题, 然后给出相应的推理算法. 常见的 f -ALCQ_{wf-reg} 推理问题包括:

- (1) 概念可满足性(Satisfiability). 在 TBox TB 的约束下, 模糊概念 C 是可满足的, 当且仅当存在 TB 的模型 FI 使得某个个体 $d \in \Delta^{FI}$ 满足 $C^{FI}(d^{FI}) > 0$.
- (2) 概念包含(Subsumption). 在 TBox TB 的约束下, 概念 C 包含于概念 D (即 $C \subseteq D$), 当且仅当任意 TB 的模型 FI 都满足 $\forall d \in \Delta^{FI}, C^{FI}(d^{FI}) \leq D^{FI}(d^{FI})$.
- (3) 知识库一致性(Consistency). 知识库 $KB = \langle TB, AB \rangle$ 是一致的或是可满足的, 当且仅当存在 TB 的模型 FI 也是 AB 的模型. 此外, 若不考虑 TBox TB 约束或 TBox TB 为空时, KB 是一致的(即 ABox AB 是一致的), 当且仅当存在 AB 的模型 FI .

(4) 知识库蕴涵(Entailment). 知识库 KB 蕴含模糊概念公理或模糊断言公式 ψ (记作 $KB \models \psi$), 当且仅当 KB 的任意模型 FI 都满足 ψ .

上述推理问题之间可以相互转化, 其中概念可满足、概念包含以及知识库蕴涵问题都可以转化为知识库一致性问题. 此外, 如果 TB 是简单 TBox, 则知识库 KB 的一致性问题又可以转化为空 TBox 约束下的 ABox 一致性问题^[15, 22, 28]. 基于此, 在文献[24, 27-28, 30-31]的基础之上, 下面给出 f -ALCQ_{wf-reg} 的 ABox 一致性检测算法. 首先介绍两个基本概念, 即否定范式和共轭断言^[22, 28]:

- (1) 否定范式. 给定一个 ABox 断言集合, 假设其中所有概念都是否定范式(Negative Normal Form, NNF), 即否定只出现在概念名之前. 利用德摩根律, 任何概念都可以通过将否定内推, 从而在线性时间内转换为等价的否定范式. 例如:
 - $\neg \forall R.C \equiv \exists R.(\neg C)$; $\neg \exists R.C \equiv \forall R.(\neg C)$;
 - $\neg \geq k R.C \equiv \leq (k-1) R.C (k \geq 1)$;
 - $\neg \geq k R.C \equiv \perp (k=0)$;

$$\neg \leq kR.C \equiv \geq (k+1)R.C.$$

(2) 共轭断言. 给定 f -ALCQ_{wf-reg} 的一个模糊断言公式 ψ , 称 ψ 为 ψ 的共轭模糊断言公式, 包括如下 4 种形式:

$$\begin{aligned} & \{ \langle \alpha \geq n \rangle, \langle \alpha < m \rangle, n \geq m \}, \\ & \{ \langle \alpha > n \rangle, \langle \alpha < m \rangle, n \geq m \}, \\ & \{ \langle \alpha \geq n \rangle, \langle \alpha \leq m \rangle, n > m \}, \\ & \{ \langle \alpha > n \rangle, \langle \alpha \leq m \rangle, n \geq m \}, \end{aligned}$$

其中 $n, m \in [0, 1], \alpha$ 为 $C(d)$ 或 $R(c, d)$.

算法 1. f -ALCQ_{wf-reg} 的 ABox 一致性检测算法.

输入: f -ALCQ_{wf-reg} 的 ABox AB , 即模糊断言公式集

输出: 布尔值(真/假)

假设: 为了使算法表达更清晰, 下面用符号 ∞ 表示 $\geq, >, \leq, <$; 用 \triangleright 表示 $\geq, >$; 用 \triangleleft 表示 $\leq, <$; 用带有下标 r 的 $\infty_r, \triangleright_r$ 和 \triangleleft_r , 分别表示 ∞, \triangleright 和 \triangleleft 的反射 (reflection), 其中 \geq 和 \leq 相互反射, $>$ 和 $<$ 相互反射.

1. 对 ABox AB 中的模糊断言公式反复利用下列规则, 直到无规则可用为止:

(1) $\neg \infty$ 规则. 如果 $(\neg C)(a) \infty n \in AB$, 且 $C(a) \infty_r (1-n) \notin AB$, 则 $AB = AB \cup \{C(a) \infty_r (1-n)\}$;

(2) $\cap \triangleright$ 规则. 如果 $(C \cap D)(a) \triangleright n \in AB$, 且 $C(a) \triangleright n \notin AB, D(a) \triangleright n \notin AB$, 则 $AB = AB \cup \{C(a) \triangleright n, D(a) \triangleright n\}$;

(3) $\cap \triangleleft$ 规则. 如果 $(C \cap D)(a) \triangleleft n \in AB$, 且 $C(a) \triangleleft n \notin AB, D(a) \triangleleft n \notin AB$, 则 $AB = AB \cup \{C'\}$, 其中 $C' \in \{C(a) \triangleleft n, D(a) \triangleleft n\}$;

(4) $\cup \triangleright$ 规则. 如果 $(C \cup D)(a) \triangleright n \in AB$, 且 $C(a) \triangleright n \notin AB, D(a) \triangleright n \notin AB$, 则 $AB = AB \cup \{C'\}$, 其中 $C' \in \{C(a) \triangleright n, D(a) \triangleright n\}$;

(5) $\cup \triangleleft$ 规则. 如果 $(C \cup D)(a) \triangleleft n \in AB$, 且 $C(a) \triangleleft n \notin AB, D(a) \triangleleft n \notin AB$, 则 $AB = AB \cup \{C(a) \triangleleft n, D(a) \triangleleft n\}$;

(6) $\forall \triangleright$ 规则. 如果 $\forall R.C(a) \triangleright n \in AB$, 且存在个体 b , 使得 $\psi \in AB$, 且 $C(b) \triangleright n \notin AB$, 其中 $\psi = R(a, b) \triangleright_r (1-n)$, 则 $AB = AB \cup \{C(b) \triangleright n\}$;

(7) $\forall \triangleleft$ 规则. 如果 $\forall R.C(a) \triangleleft n \in AB$, 且不存在个体 b , 使得 $R(a, b) \triangleleft_r (1-n) \in AB$ 和 $C(b) \triangleleft n \in AB$, 则 $AB = AB \cup \{R(a, b) \triangleleft_r (1-n), C(b) \triangleleft n\}$;

(8) $\exists \triangleright$ 规则. 如果 $\exists R.C(a) \triangleright n \in AB$, 且不存在个体 b , 使得 $R(a, b) \triangleright n \in AB$ 和 $C(b) \triangleright n \in AB$, 则 $AB = AB \cup \{R(a, b) \triangleright n, C(b) \triangleright n\}$;

(9) $\exists \triangleleft$ 规则. 如果 $\exists R.C(a) \triangleleft n \in AB$, 且存在个体 b , 使得 $\psi \in AB$, 且 $C(b) \triangleleft n \notin AB$, 其中 $\psi = R(a, b) \triangleleft n$, 则 $AB = AB \cup \{C(b) \triangleleft n\}$;

(10) wf_1 规则. 如果 $wf(R)(b_0) = 1 \in AB$, 且不存在个体 $b_1, \dots, b_n, b_{n+1} (n \geq 1)$, 使得 $R(b_0, b_1) \triangleright \beta \in AB, \dots, R(b_{n-1}, b_n) \triangleright \beta \in AB, R(b_n, b_{n+1}) < \beta \in AB$, 则 $AB = AB \cup \{R(b_0, b_1) \triangleright \beta, \dots, R(b_{n-1}, b_n) \triangleright \beta\}$;

(11) $\leq \triangleright$ 规则. 如果 $\leq kR.C(a) \triangleright n \in AB$, 且存在 $k+1$

个互不相同的个体 b_1, \dots, b_{k+1} , 使得 $\psi_1 \in AB, \dots, \psi_{k+1} \in AB$ 和 $C(b_1) \triangleright n \notin AB, \dots, C(b_{k+1}) \triangleright n \notin AB$, 且 $R(a, b_1) \triangleright_r (1-n) \notin AB, \dots, R(a, b_{k+1}) \triangleright_r (1-n) \notin AB$, 其中 $\psi_1 = R(a, b_1) \triangleright n, \dots, \psi_{k+1} = R(a, b_{k+1}) \triangleright n$, 则 $AB = AB \cup \{R(a, b_1) \triangleright_r (1-n), \dots, R(a, b_{k+1}) \triangleright_r (1-n), C(b_1) \triangleright n, \dots, C(b_{k+1}) \triangleright n\}$;

(12) $\geq \triangleleft$ 规则. 如果 $\geq kR.C(a) \triangleleft n \in AB$, 则将 $\leq \triangleright$ 规则应用于 $\leq (k-1)R.C(a) \triangleleft_r (1-n)$ 上;

(13) $\geq \triangleright$ 规则. 如果 $\geq kR.C(a) \triangleright n \in AB$, 且不存在 k 个互不相同的个体 b_1, \dots, b_k , 使得 $R(a, b_i) \triangleright n \in AB$ 和 $C(b_i) \triangleright n \in AB$, 则 $AB = AB \cup \{R(a, b_i) \triangleright n, \dots, C(b_i) \triangleright n\}$, 其中 $b_i \neq b_j, 1 \leq i < j \leq k$;

(14) $\leq \triangleleft$ 规则. 如果 $\leq kR.C(a) \triangleleft n \in AB$, 则将 $\geq \triangleright$ 规则应用于 $\geq (k+1)R.C(a) \triangleleft_r (1-n)$ 上;

(15) $id(C) \infty$ 规则. 如果 $id(C)(a, a) \infty n \in AB$, 且 $C(a) \infty n \notin AB$, 则 $AB = AB \cup \{C(a) \infty n\}$;

(16) $R^* / +$ 规则. 因为 $(R^*)^{FI} = (R^{FI})^* = \bigcup_{n \geq 0} (R^{FI})^n$ 和 $(R^+)^{FI} = (R \circ R^*)^{FI}$ (见第 3.1 节), 则 $R^* / +$ 可以转化为下面的 R_\circ 和 R_\cup 规则;

(17) $R_\circ \triangleright$ 规则. 如果 $(R_1 \circ R_2)(a, c) \triangleright n \in AB$, 且不存在个体 b , 使得 $R_1(a, b) \triangleright n \in AB$ 和 $R_2(b, c) \triangleright n \in AB$, 则 $AB = AB \cup \{R_1(a, b) \triangleright n, R_2(b, c) \triangleright n\}$;

(18) $R_\circ \triangleleft$ 规则. 如果 $(R_1 \circ R_2)(a, c) \triangleleft n \in AB$, 且不存在个体 b , 使得 $R_1(a, b) \triangleleft n \in AB$ 和 $R_2(b, c) \triangleleft n \in AB$, 则 $AB = AB \cup \{R'\}$, 其中 $R' \in \{R_1(a, b) \triangleleft n, R_2(b, c) \triangleleft n\}$;

(19) $R_\cup \triangleright$ 规则. 如果 $(R_1 \cup R_2)(a, b) \triangleright n \in AB$, 且 $R_1(a, b) \triangleright n \notin AB, R_2(a, b) \triangleright n \notin AB$, 则 $AB = AB \cup \{R'\}$, 其中 $R' \in \{R_1(a, b) \triangleright n, R_2(a, b) \triangleright n\}$;

(20) $R_\cup \triangleleft$ 规则. 如果 $(R_1 \cup R_2)(a, b) \triangleleft n \in AB$, 且 $R_1(a, b) \triangleleft n \notin AB, R_2(a, b) \triangleleft n \notin AB$, 则 $AB = AB \cup \{R_1(a, b) \triangleleft n, R_2(a, b) \triangleleft n\}$.

2. 通过运用上面规则, 将会产生一个 ABox 的集合 $S = \{AB_1, \dots, AB_n\}$. 如果通过检测发现其中至少有一个 ABox AB_i 不包含冲突, 则 ABox AB 是一致的, 否则 ABox AB 是不一致的. 如果一个 ABox AB_i 中出现下列 10 种情形中的任意一种, 则称该 ABox AB_i 包含冲突:

(1) 上面提到的 4 种共轭模糊断言公式;

(2) 底概念 \perp 和顶概念 \top : $\{\perp(a) \geq n, n > 0\}, \{\perp(a) > n, n \geq 0\}, \{\top(a) \leq n, n < 1\}, \{\top(a) < n, n \leq 1\}$;

(3) $\{C(a) > 1\}, \{C(a) < 0\}$, 其中 C 为任意概念, a 是个体.

3. 算法结束.

基于文献[24, 27-28, 30-31]中的内容, 可以类似地证明上述 f -ALCQ_{wf-reg} ABox 的一致性检测算法是正确的(即可终止的、可靠的和完备的), 因此本文这里不再赘述.

4 基于 f -ALCQ_{wf-reg} 的模糊 XML 模型 表示与推理

本节研究如何利用 f -ALCQ_{wf-reg} 对模糊 XML 模型进行表示与推理,包括:(1)提出模糊 XML 模型的 f -ALCQ_{wf-reg} 表示方法,即实现从模糊 XML 模型到 f -ALCQ_{wf-reg} 知识库的转化,并给出转化的正确性证明和转化实例;(2)研究如何将模糊 XML 模型的推理问题转化为 f -ALCQ_{wf-reg} 的推理问题,并简要讨论如何利用上述转化和推理结果来更有效地支持模糊 XML 文档库的查询问题。

4.1 模糊 XML 模型到 f -ALCQ_{wf-reg} 知识库的转化

下面首先介绍模糊 XML 模型和模糊描述逻辑之间的关系,然后提出一种从模糊 XML 模型到 f -ALCQ_{wf-reg} 知识库的转化方法。

由上面讨论的内容不难看出,首先模糊 XML 模型和模糊描述逻辑表示语法有所不同,即二者采用不同的语法形式来表示应用领域的相关信息,其中前者采用有限对互相匹配的标签形式,而后者则采用定义在概念、角色以及个体上的公理和断言形式。此外,与经典 XML 相似^[1-2],模糊 XML 模型缺少语义,而模糊描述逻辑作为一阶谓词逻辑的一个可判定子集,具有基于逻辑的形式化语义和较强的推理能力^[15]。应当指出的是,尽管二者具有不同的语法和语义,但它们都具有树形化结构的特点:任何一个模糊 XML 文档都可以看作是一棵语法树^[2,21],而大多数模糊描述逻辑知识库的模型都是完全无冲突的森林(即树的集合)^[15,22,24-28]。基于上述特点以及 XML 和经典描述逻辑之间的对应关系^[3,5-9],下面给出从模糊 XML 模型到 f -ALCQ_{wf-reg} 知识库的形式化转化方法。

定义 8(转化方法). 给定一个模糊 XML 模型 (D, d) ,其中 $D=(\mathbf{P}, r)$ 是一个模糊 DTD(见定义 2), d 是一个与 D 相一致的模糊 XML 文档, \mathbf{P} 是 D 中元素类型定义的集合, r 表示 D 的根元素用于唯一标识该模糊 DTD,则与该模糊 XML 模型对应的模糊描述逻辑 f -ALCQ_{wf-reg} 知识库 $\varphi(D) = \langle FA, FT \rangle$ 按下列规则得到:

(1) FA 是一个模糊概念和模糊关系的集合,包含下面的元素:

① 对每一个元素类型定义 $E \rightarrow \alpha \in \mathbf{P}$,定义一个原子模糊概念 E_D 和两个原子模糊概念 $StartE$ 和 $EndE$,其中 $StartE$ 和 $EndE$ 用于表示元素 E 的开

始和结束标签。

② 对每一个终止符 T (如 #PCDATA),定义一个原子模糊概念 T 。

③ 定义两个原子角色 f 和 r ,其用途见下面(2)。

④ 定义两个原子模糊概念 Tag 和 $Terminal$,用于区分标签(tags)和终止符(terminals)。

(2) FT 是一个模糊公理 FT_a 和 FT_D 的集合,其中 FT_a 来自于模糊 XML 文档,而 FT_D 来自于模糊 DTD:

① FT_a , 一个模糊公理的集合,用于表示模糊 XML 文档 d 的内容。下面首先给出模糊 XML 文档 d 的树形化表示形式,在此基础之上,给出 d 相应的模糊公理集合 FT_a 。按照下面步骤(a)~(c),能够将一个模糊 XML 文档转化为一棵二叉树(即树中每一个节点至多有两个孩子节点):

(a) 模糊 XML 文档 d 中的根元素对应于树的根节点,用于唯一标识该模糊 XML 文档。

(b) 上述(1)中定义的模糊角色 f 用于表示一个节点的开始标签(即左孩子节点),而 r 用于表示该节点的其余部分(即右孩子节点及其子孙节点),见下图 3。

(c) 给定一个含有 n 个成分的模糊 XML 文档 d ,如 $\langle E \rangle \langle E_1 \rangle \cdots \langle /E_1 \rangle \cdots \langle E_n \rangle \cdots \langle /E_n \rangle \langle /E \rangle$,则根节点的 f -filler(左孩子)表示开始标签 $\langle E \rangle$, $(r \circ f)$ -filler 表示第一个成分 $\langle E_1 \rangle \cdots \langle /E_1 \rangle$, ..., $(r^n \circ f)$ -filler 表示最后一个成分 $\langle E_n \rangle \cdots \langle /E_n \rangle$, r^{n+1} -filler 表示结束标签 $\langle /E \rangle$ 。图 3 给出图 1 中模糊 XML 文档 d_1 的一个片段的树形化表示形式。

(d) 基于上面给出的树形化表示形式,模糊 XML 文档 d 相应的模糊公理集合 FT_a 如下:

$$\top \equiv (\leq 1f. \top) \cap (\leq 1r. \top) \cap \omega f(f \cup r) \quad // \top \text{ 为顶概念} \quad (1)$$

$$Tag \subseteq \forall (f \cup r). \perp \quad (2)$$

$$StartE \subseteq Tag \quad (3)$$

$$EndE \subseteq Tag \quad (4)$$

$$Terminal \subseteq \forall (f \cup r). \perp \cap \neg Tag \quad (5)$$

$$T \subseteq Terminal \quad // T \text{ 为终止符} \quad (6)$$

$$T_1 \subseteq \neg T_2 \quad (7)$$

$$\begin{cases} StartE_i \equiv StartE_j, & \text{元素 } E_i \text{ 和 } E_j \text{ 的开始和} \\ EndE_i \equiv EndE_j, & \text{结束标签等价} \end{cases} \quad (8)$$

$$\begin{cases} StartE_i \subseteq \neg StartE_j, & \text{元素 } E_i \text{ 和 } E_j \text{ 的开始和} \\ EndE_i \subseteq \neg EndE_j, & \text{结束标签不等价} \end{cases} \quad (9)$$

其中,公理(1)~(7)用于限定模糊 XML 文档的树形化结构形式,且 $\omega f(f \cup r)$ 用于限定关系链的有限性和非循环性;公理(8)~(9)用于限定两个元素类型 E_i 和 E_j 之间的等价/不等价关系,在实际应用中,尽管两个元素类型的名称不同,如“Database”和“DB”,但在某些情况下它们表示相同的元素,这一点将在本文 4.2 节中做详细介绍。

② FT_D , 一个模糊公理的集合,用于限定模糊 DTD 中的元素类型定义 $E \rightarrow \alpha$, 给定一个元素类型定义 $E \rightarrow \alpha$, 其中 $\alpha ::= S | \text{empty} | (\alpha_1 | \alpha_2) | (\alpha_1, \alpha_2) | \alpha? | \alpha^* | \alpha^+ | \text{any}$ (见定义 2), 则 FT_D 包含如下模糊公理:

$$E_D \equiv \exists f. \text{Start}E \cap \exists (r \circ \varphi(\alpha)). \text{End}E.$$

其中 $\varphi(\alpha)$ 定义如下:

$$\begin{aligned} \varphi(S) &= id(\exists f. \beta(D, S)) \circ r, \\ \varphi(\text{empty}) &= id(\top), \\ \varphi(\alpha_1 | \alpha_2) &= \varphi(\alpha_1) \cup \varphi(\alpha_2), \\ \varphi(\alpha_1, \alpha_2) &= \varphi(\alpha_1) \circ \varphi(\alpha_2), \\ \varphi(\alpha?) &= \varphi(\alpha) \cup id(\top), \\ \varphi(\alpha^*) &= \varphi(\alpha)^*, \\ \varphi(\alpha^+) &= \varphi(\alpha)^+. \end{aligned}$$

其中, $S = T \cup E = T \cup E_g \cup E_s$, $\beta(D, S)$ 是一个函数,且满足:如果 $S = E \in E$ (即为元素类型), 则 $\beta(D, S) = E_D$; 如果 $S = T \in T$ (即为终止符), 则 $\beta(D, S) = T$.

下面定理 1 给出上述转化的正确性证明,它通过建立映射(即模糊 XML 模型与转化后知识库的模型之间的相互映射)来证明上述定义 8 能够正确地将模糊 XML 模型转化为模糊描述逻辑 $f\text{-ALCQ}_{\text{wf-reg}}$ 知识库。

定理 1. 给定任意模糊 XML 模型 (D, d) , 其中 D 是一个模糊 DTD, d 是一个与 D 相一致的模糊 XML 文档, $\varphi(D)$ 是通过定义 8 得到的 $f\text{-ALCQ}_{\text{wf-reg}}$

知识库, 则

(1) 存在从模糊 XML 文档到 $\varphi(D)$ 的模型的映射 μ , 使得对任意模糊 XML 文档 d , $\mu(d)$ 是 $\varphi(D)$ 的一个模型, 且对任意 $d \in d_{T,E}$ (见定义 1), 当且仅当 $o \in \beta(D, S)^{\mu(d)}$ (见定义 8), 其中 o 是 $\mu(d)$ 的根元素;

(2) 存在从 $\varphi(D)$ 的模型到模糊 XML 文档的映射 λ , 使得对 $\varphi(D)$ 的任意模型 FI 以及个体元素 $o \in \Delta^{FI}$, $\lambda(o)$ 是与 D 相一致的模糊 XML 文档, 且对任意 $o \in \beta(D, S)^{FI}$ (见定义 8), 当且仅当 $\lambda(o) \in d_{T,E}$.

证明. 下面给出定理 1 中第二部分的证明. 假定 $FI = (\Delta^{FI}, \cdot^{FI})$ 是 $\varphi(D)$ 的一个模型, 且 $o \in \Delta^{FI}$, 则与 D 相一致的模糊 XML 文档 $\lambda(o)$ 可以通过下列规则得到:

(a) 如果 $o \in T^{FI}$, 其中 $T \in T$, 则 $\lambda(o) = T$;

(b) 如果 $o \in E^{FI}$, 其中元素 $E \in E$, 即存在 $n \geq 0$ 以及对象 o_b, o_i, o'_i 和 o_e , 使得 $o_b \in \text{Start}E^{FI}$, $o_e \in \text{End}E^{FI}$, $(o, o_b), (o_1, o'_1), \dots, (o_n, o'_n) \in f^{FI}$ 以及 $(o, o_1), (o_1, o_2), \dots, (o_{n-1}, o_n), (o_n, o_e) \in r^{FI}$, 则 $\lambda(o) = \langle E \rangle \lambda(o'_1), \dots, \lambda(o'_n) \langle /E \rangle$.

定理 1 中第一部分的证明过程与上述类似, 它们是一个相互逆过程. 事实上, 定理 1 的证明可以更直接地通过图 3 来说明. 证毕.

为了说明定义 8 中的转化过程, 下面给出一个从模糊 XML 模型到 $f\text{-ALCQ}_{\text{wf-reg}}$ 知识库的转化实例。

例 1. 给定第 2 节中的模糊 XML 模型(即图 1 中的模糊 XML 文档 d_1 和图 2 中的模糊 DTD D_1), 通过定义 8, 图 4 给出转化后的 $f\text{-ALCQ}_{\text{wf-reg}}$ 知识库(为了简化, 这里仅给出公理集合 FT_{D_1} , 而 FT_{D_1} 部分可以由定义 8 中的公理(1)~(9)类似地得到)。

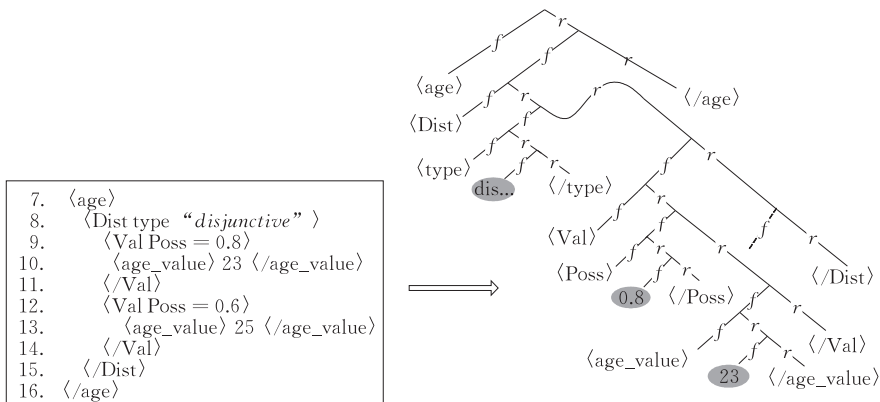


图 3 图 1 中模糊 XML 文档 d_1 的一个片段的树形化表示形式

$$\begin{aligned}
Universities_D &\equiv \exists f.StartUniversities \cap \exists (r \circ (id(\exists f.University_D) \circ r)^*). EndUniversities \\
University_{D_1} &\equiv \exists f.StartUniversity \cap \exists (r \circ id(\exists f.UName_D) \circ r \circ id(\exists f.Val_{D_1}) \circ r \circ \\
&\quad (id(\exists f.Val_{D_1}) \circ r)^*). EndUniversity \\
UName_D &\equiv \exists f.StartUName \cap \exists (r \circ id(\exists f.#PCDATA) \circ r). EndUName \\
Val_{D_1} &\equiv \exists f.StartVal \cap \exists (r \circ id(\exists f.Poss_D) \circ r \circ (id(\exists f.Student_D) \circ r)^*). EndVal \\
Student_D &\equiv \exists f.StartStudent \cap \exists (r \circ ((id(\exists f.Sname_D) \circ r) \cup id(T)) \circ ((id(\exists f.Age_D) \circ r) \cup \\
&\quad id(T)) \circ ((id(\exists f.Email_D) \circ r) \cup id(T))). EndStudent \\
Sname_D &\equiv \exists f.StartSname \cap \exists (r \circ id(\exists f.#PCDATA) \circ r). EndSname \\
Dist_D &\equiv \exists f.StartDist \cap \exists (r \circ id(\exists f.Type_D) \circ r \circ id(\exists f.Val_{D_2}) \circ r \circ \\
&\quad (id(\exists f.Val_{D_2}) \circ r)^*). EndDist \\
&\quad \vdots \\
Email_value_D &\equiv \exists f.StartEmail_value \cap \exists (r \circ id(\exists f.#PCDATA) \circ r). EndEmail_value
\end{aligned}$$

图 4 第 2 节中的模糊 XML 模型(D_1, d_1)相对应的 f -ALCQ_{wf-reg} 知识库(仅给出 FT_{D_1} 部分)

4.2 基于 f -ALCQ_{wf-reg} 的模糊 XML 模型的推理以及查询

基于上一节转化后的 f -ALCQ_{wf-reg} 知识库,本节研究如何利用 f -ALCQ_{wf-reg} 的推理机制对模糊 XML 模型进行推理以及查询优化,包括:(1)首先给出模糊 XML 模型推理问题的形式化定义;(2)将模糊 XML 模型的推理问题转化为 f -ALCQ_{wf-reg} 的推理问题,从而可以利用 f -ALCQ_{wf-reg} 的推理机制实现对模糊 XML 模型的推理,并简要讨论如何利用上述转化和推理结果来更有效地支持模糊 XML 文档库的查询问题。

4.2.1 模糊 XML 模型的常见推理问题

模糊 XML 模型中常见的推理任务包括:一致性(conformance)、包含性(inclusion)、等价性(equivalence)以及不相交性(disjointness)^[3-10,21],其中一致性用于判定是否一个模糊 XML 文档与一个模糊 DTD 相一致,而其它推理任务则是判定两个模糊 DTDs 是否包含、等价或不相交。下面分别给出其形式化定义。

定义 9(一致性). 给定一个模糊 DTD $D = (\mathbf{P}, r)$ 和一个模糊 XML 文档 d , 则当 $d \in d(D)$ 时, 称 d 相对于 D 是一致的, 其中 $d(D)$ 是一个定义在 D 上的模糊 XML 文档实例集合, 且 $d(D)$ 可以通过定义 1 中提到的方法类似给出, 即定义在 (\mathbf{P}, r) 上的集合 $d(D)$ 可以按如下方法归纳得到:

(1) 如果 $r = T \in \mathbf{T}$ 是一个终止符, 则 $d(D) = T$;

(2) 如果 r 是一个元素类型 $E \rightarrow \alpha \in \mathbf{P}$, 则 $d(D)$ 是一个形如 $\langle E \rangle d_1, \dots, d_k \langle /E \rangle$ 序列的集合, 其中 $E \in \mathbf{E}$ 是一个元素类型, d_1, \dots, d_k 是通过内容模型 α 产生的文档实例。

定义 10(包含性). 给定两个模糊 DTDs D_1 和 D_2 , 则当 $d(D_1) \subseteq d(D_2)$ 时, 称 D_1 强包含于 D_2 (记作 $D_1 \subseteq_s D_2$)。

定义 11(等价性). 给定两个模糊 DTDs D_1 和

D_2 , 则当 $d(D_1) = d(D_2)$ 时, 称 D_1 强等价于 D_2 (记作 $D_1 \equiv_s D_2$)。

定义 12(不相交性). 给定两个模糊 DTDs D_1 和 D_2 , 则当 $d(D_1) \cap d(D_2) = \emptyset$ 时, 称 D_1 强不相交于 D_2 (记作 $D_1 \otimes_s D_2$)。

从上述定义可知, 判定两个模糊 DTDs D_1 和 D_2 之间的强包含(强等价、强不相交)问题, 就是判定定义在 D_1 和 D_2 上的模糊 XML 文档集合是否完全包含(相同、不相交)。因此, DTDs 中元素类型的“名称”(即元素的标签)起到了关键性作用, 例如, 给定两个简化的模糊 DTDs(仅给出部分元素类型):

$D_{University} \langle !ELEMENT University (Uname, Position, Val+, \dots) \rangle$ 和

$D_{College} \langle !ELEMENT College (Cname, Position, Val+, \dots) \rangle$, 根据定义 9~12, 可判定 $D_{University}$ 和 $D_{College}$ 是强不相交的。然而, 在实际应用中, 尽管两个模糊 DTDs 中的元素类型名称不同, 但在某些情况下它们的结构是相关的。例如, 假如认为元素类型名称 University 和 Uname 分别与 College 和 Cname 等价, 则 $D_{University}$ 等价于 $D_{College}$ 。

基于此, 通过考虑某些元素类型名称之间的等价/不等价关系(equivalent/unequivalent relation), 可以更好地判定两个模糊 DTDs 之间的包含、等价或不相交关系, 此时, 上述模糊 XML 模型的四种推理问题被称为是 ϵ -一致(ϵ -包含、 ϵ -等价以及 ϵ -不相交), 且强包含和强等价等推理问题是 ϵ -包含和 ϵ -等价等推理问题的特例。

本文后面的工作仅考虑 ϵ -包含和 ϵ -等价等推理问题, 与定义 9~12 相类似, 下面给出 ϵ -一致性(ϵ -conformance)、 ϵ -包含性(ϵ -inclusion)、 ϵ -等价性(ϵ -equivalence)以及 ϵ -不相交性(ϵ -disjointness)推理问题的形式化定义。

定义 13(ϵ -一致性). 给定一个模糊 DTD $D = (\mathbf{P}, r)$ 和一个模糊 XML 文档 d , 则当 $d \in d_\epsilon(D)$ 时,

称 d 相对于 D 是 ϵ -一致的, 其中 $d_\epsilon(D)$ 是一个定义在 D 上的模糊 XML 文档实例集合, 且 $d_\epsilon(D)$ 可以通过定义 9 中提到的方法类似给出, 即定义在 (\mathbf{P}, r) 上的集合 $d_\epsilon(D)$ 可以按如下方法归纳得到:

(1) 如果 $r \in \mathbf{T}$ 是一个终止符, 则 $d_\epsilon(D) = \mathbf{T}$;

(2) 如果 r 是一个元素类型 $E \rightarrow \alpha \in \mathbf{P}$, 则 $d_\epsilon(D)$ 是一个形如 $\langle E' \rangle d_1, \dots, d_k \langle /E' \rangle$ 序列的集合, 其中 $E' \in [E]_\epsilon$ ($[E]_\epsilon$ 表示元素类型 E 的等价类), d_1, \dots, d_k 是通过内容模型 α 产生的文档实例.

定义 14(ϵ -包含性). 给定两个模糊 DTDs D_1 和 D_2 , 则当 $d_\epsilon(D_1) \subseteq d_\epsilon(D_2)$ 时, 称 D_1 ϵ -包含于 D_2 (记作 $D_1 \subseteq_\epsilon D_2$).

定义 15(ϵ -等价性). 给定两个模糊 DTDs D_1 和 D_2 , 则当 $d_\epsilon(D_1) = d_\epsilon(D_2)$ 时, 称 D_1 ϵ -等价于 D_2 (记作 $D_1 \equiv_\epsilon D_2$).

定义 16(ϵ -不相交性). 给定两个模糊 DTDs D_1 和 D_2 , 则当 $d_\epsilon(D_1) \cap d_\epsilon(D_2) = \emptyset$ 时, 称 D_1 ϵ -不相交于 D_2 (记作 $D_1 \otimes_\epsilon D_2$).

4.2.2 基于 f -ALCQ_{wf-reg} 的模糊 XML 模型的推理

下面的定理 2~5 能够将上一节提到的模糊 XML 模型的推理问题转化为模糊描述逻辑 f -ALCQ_{wf-reg} 的推理问题, 进而利用模糊描述逻辑的推理机制实现对模糊 XML 模型的推理.

定理 2. 给定一个模糊 DTD $D = (P, r)$ 和一个模糊 XML 文档 d , $\varphi(D)$ 是通过定义 8 得到的 f -ALCQ_{wf-reg} 知识库, 则判定 d 是否相对于 D 是 ϵ -一致的问题能多项式归约为知识库 $\varphi(D)$ 的模型检测 (model checking) 问题, 即判定定理 1 中的 $\mu(d)$ 是否是 $\varphi(D)$ 的一个模型.

上述归约过程之所以是多项式时间的, 因为从定义 8 可以看出, 对于一个给定大小的 D (D 中元素类型定义的数量) 来说, 从 D 转化为 $\varphi(D)$ 的过程是在多项式时间内完成的, 而从定理 1 可知, 从 d 构建 $\mu(d)$ 的过程与从 o 构建 $\lambda(o)$ 的过程互逆, 且都可以在多项式时间内完成. 此外, 定理 2 的证明可以通过定理 1 类似地给出, 这里不再赘述.

下面给出两个模糊 DTDs 之间的 ϵ -包含性、 ϵ -等价性以及 ϵ -不相交性推理问题的转化方法. 首先需要扩展模糊描述逻辑 f -ALCQ_{wf-reg} 知识库 $\varphi(D)$ 的定义: 给定一个模糊 DTDs 的集合 $D = \{D_1, \dots, D_n\}$, 通过定义 8 转化后的 f -ALCQ_{wf-reg} 知识库 $\varphi(D)$ 表示来自于多个模糊 DTDs, 而不是一个模糊 DTD.

定理 3. 给定模糊 DTDs 的集合 $D = \{D_1, D_2\}$, 其中 $D_1 = (P_1, r')$, $D_2 = (P_2, r'')$, r' 和 r'' 分别是

它们的根元素类型, $\varphi(D) = FT_{D_1} \cup FT_{D_2}$ 是通过定义 8 转化后的 f -ALCQ_{wf-reg} 知识库 (仅 FT_{D_i} 部分), r'_{D_1} 和 r''_{D_2} 是 $\varphi(D)$ 中相应于 r' 和 r'' 的两个模糊概念, 则 $D_1 \subseteq_\epsilon D_2$, 当且仅当 $\varphi(D) \models r'_{D_1} \subseteq r''_{D_2}$.

证明. 首先证明“ \Rightarrow ”, 假定 $\varphi(D) \not\models r'_{D_1} \subseteq r''_{D_2}$, 则存在 $o \in \Delta^{FI}$ 使得 $o \in r'_{D_1}^{FI}$ 和 $o \notin r''_{D_2}^{FI}$ 成立, 其中 $FI = (\Delta^{FI}, \cdot^{FI})$ 是 $\varphi(D)$ 的一个模型. 通过定理 1 的第二部分可知, 存在一个模糊 XML 文档 $\lambda(o)$, 且有 $\lambda(o) \in d_\epsilon(D_1)$ 和 $\lambda(o) \notin d_\epsilon(D_2)$, 即 $D_1 \subseteq_\epsilon D_2$ 不成立, 这与 $D_1 \subseteq_\epsilon D_2$ 相矛盾, 所以有 $\varphi(D) \models r'_{D_1} \subseteq r''_{D_2}$.

再证明“ \Leftarrow ”, 假定 $D_1 \subseteq_\epsilon D_2$ 不成立, 则存在 d 使得 $d \in d_\epsilon(D_1)$ 和 $d \notin d_\epsilon(D_2)$ 成立. 由定理 1 的第一部分可知, 存在 $\varphi(D)$ 的一个模型 $\mu(d)$, 且有 $r'_{D_1} \in \beta(D, S)^{\mu(d)}$ 和 $r''_{D_2} \notin \beta(D, S)^{\mu(d)}$, 即 $\varphi(D) \not\models r'_{D_1} \subseteq r''_{D_2}$, 这与 $\varphi(D) \models r'_{D_1} \subseteq r''_{D_2}$ 相矛盾, 所以有 $D_1 \subseteq_\epsilon D_2$.

证毕.

定理 4. 给定模糊 DTDs 的集合 $D = \{D_1, D_2\}$, 其中 $D_1 = (P_1, r')$, $D_2 = (P_2, r'')$, r' 和 r'' 分别是它们的根元素类型, $\varphi(D) = FT_{D_1} \cup FT_{D_2}$ 是通过定义 8 转化后的 f -ALCQ_{wf-reg} 知识库 (仅 FT_{D_i} 部分), r'_{D_1} 和 r''_{D_2} 是 $\varphi(D)$ 中相应于 r' 和 r'' 的两个模糊概念, 则 $D_1 \equiv_\epsilon D_2$, 当且仅当 $\varphi(D) \models r'_{D_1} \equiv r''_{D_2}$.

证明. 首先证明“ \Rightarrow ”, 假定 $\varphi(D) \not\models r'_{D_1} \equiv r''_{D_2}$, 即 r'_{D_1} 和 r''_{D_2} 不等价, 则存在 $o \in \Delta^{FI}$ 使得 $o \in r'_{D_1}^{FI}$ 和 $o \notin r''_{D_2}^{FI}$ 或者 $o \notin r'_{D_1}^{FI}$ 和 $o \in r''_{D_2}^{FI}$ 成立, 其中 $FI = (\Delta^{FI}, \cdot^{FI})$ 是 $\varphi(D)$ 的一个模型. 通过定理 1 的第二部分可知, 存在一个模糊 XML 文档 $\lambda(o)$, 且有 $\lambda(o) \in d_\epsilon(D_1)$ 和 $\lambda(o) \notin d_\epsilon(D_2)$ 或者 $\lambda(o) \notin d_\epsilon(D_1)$ 和 $\lambda(o) \in d_\epsilon(D_2)$, 即 D_1 和 D_2 不等价, 这与 $D_1 \equiv_\epsilon D_2$ 相矛盾, 所以有 $\varphi(D) \models r'_{D_1} \equiv r''_{D_2}$.

再证明“ \Leftarrow ”, 假定 $D_1 \equiv_\epsilon D_2$ 不成立, 即 D_1 与 D_2 不等价, 则存在 d 使得 $d \in d_\epsilon(D_1)$ 和 $d \notin d_\epsilon(D_2)$ 或者 $d \notin d_\epsilon(D_1)$ 和 $d \in d_\epsilon(D_2)$ 成立. 由定理 1 的第一部分可知, $\mu(d)$ 是 $\varphi(D)$ 的一个模型, 且有 $r'_{D_1} \in \beta(D, S)^{\mu(d)}$ 和 $r''_{D_2} \notin \beta(D, S)^{\mu(d)}$ 或者 $r'_{D_1} \notin \beta(D, S)^{\mu(d)}$ 和 $r''_{D_2} \in \beta(D, S)^{\mu(d)}$, 即 $\varphi(D) \not\models r'_{D_1} \equiv r''_{D_2}$, 这与 $\varphi(D) \models r'_{D_1} \equiv r''_{D_2}$ 矛盾, 所以有 $D_1 \equiv_\epsilon D_2$. 证毕.

定理 5. 给定模糊 DTDs 的集合 $D = \{D_1, D_2\}$, 其中 $D_1 = (P_1, r')$, $D_2 = (P_2, r'')$, r' 和 r'' 分别是它们的根元素类型, $\varphi(D) = FT_{D_1} \cup FT_{D_2}$ 是通过定义 8 转化后的 f -ALCQ_{wf-reg} 知识库 (仅 FT_{D_i} 部分), r'_{D_1} 和 r''_{D_2} 是 $\varphi(D)$ 中相应于 r' 和 r'' 的两个模糊概念, 则 $D_1 \otimes_\epsilon D_2$, 当且仅当 $\varphi(D) \models r'_{D_1} \cap r''_{D_2} \subseteq \perp$.

证明. 首先证明“ \Leftarrow ”, 假定 $D_1 \otimes_\epsilon D_2$ 不成立,

即 D_1 与 D_2 相交, 则存在 d 使得 $d \in d_\epsilon(D_1)$ 和 $d \in d_\epsilon(D_2)$. 由定理 1 的第一部分可知, $\mu(d)$ 是 $\varphi(D)$ 的一个模型, 且有 $r'_{D_1} \in \beta(D, S)^{\mu(d)}$ 和 $r''_{D_2} \in \beta(D, S)^{\mu(d)}$, 即 r'_{D_1} 和 r''_{D_2} 相交, 这与 $\varphi(D) \models r'_{D_1} \cap r''_{D_2} \subseteq \perp$ 相矛盾, 所以有 $D_1 \otimes_\epsilon D_2$.

再证明“ \Rightarrow ”, 假定 $\varphi(D) \not\models r'_{D_1} \cap r''_{D_2} \subseteq \perp$, 即 r'_{D_1} 和 r''_{D_2} 相交不为空, 则存在 $o \in \Delta^{FI}$ 使得 $o \in r'_{D_1}$ 和 $o \in r''_{D_2}$ 成立, 其中 $FI = (\Delta^{FI}, \cdot^{FI})$ 是 $\varphi(D)$ 的一个模型. 通过定理 1 的第二部分可知, $\lambda(o)$ 是一个模糊 XML 文档, 且有 $\lambda(o) \in d_\epsilon(D_1)$ 和 $\lambda(o) \in d_\epsilon(D_2)$, 即 D_1 和 D_2 相交不为空, 这与 $D_1 \otimes_\epsilon D_2$ 相矛盾, 所以有 $\varphi(D) \models r'_{D_1} \cap r''_{D_2} \subseteq \perp$. 证毕.

通过前面章节的介绍, 可以将模糊 XML 模型及其相应的推理问题转化为模糊描述逻辑 f -ALCQ_{wf-reg} 知识库和模糊描述逻辑的推理问题, 从而可以利用 f -ALCQ_{wf-reg} 的推理机制对模糊 XML 模型进行推理, 这将有助于模糊 XML 模型的设计与整合等, 并且能够在一定程度上克服目前模糊 XML 模型推理中存在的不足(如推理需要手工进行, 且存在推理效率和可靠性不高等问题). 再者, 研究基于模糊描述逻辑的模糊 XML 模型的表示与推理也将有助于支持模糊 XML 文档库的查询问题. 下面通过给出几条查询优化规则来简要说明利用上述转化和推理结果能够更有效地支持模糊 XML 文档库的查询(如有利于提高查询的准确性和有效性), 本文这里并不详细介绍有关模糊 XML 文档查询的相关技术.

一个文档库 DCB (Document Base) 是一个二元组 $DCB = \langle D_\epsilon, d_\epsilon(D_\epsilon) \rangle$, 其中 D_ϵ 是一个模糊 DTDs 的集合, $d_\epsilon(D_\epsilon)$ 是一个定义在 D_ϵ 上的模糊 XML 文档实例集合. 一个定义在 DCB 上的查询 Q 是一个简单的模糊文档类型定义(即模糊 DTD), 用于检索所有与 Q 相 ϵ -一致的文档实例, 即 $Q(DCB) = \{d \mid d \in d_\epsilon(D_\epsilon) \wedge d \in d_\epsilon(Q)\}$.

基于前面几节的转化和推理结论, 可知对于给定的每一对模糊 DTDs $D_1, D_2 \in D_\epsilon$, 关系 $D_1 \subseteq_\epsilon D_2$ 、 $D_1 \equiv_\epsilon D_2$ 或者 $D_1 \otimes_\epsilon D_2$ 是可知的, 且对于每一个文档实例 $d \in d_\epsilon(D_\epsilon)$ 和一个模糊 DTD $D_1 \in D_\epsilon$, d 是否相对于 D_1 是 ϵ -一致的同样是可知的(即 $d \in d_\epsilon(D_1)$ 或者 $d \notin d_\epsilon(D_1)$ 是可知的), 则有下面的优化规则.

规则 1. 如果在 D_ϵ 中存在一个模糊 DTD D_1 使得 $D_1 \equiv_\epsilon Q$ (且不存在 D_2 使得 $D_2 \in D_\epsilon$ 以及 $D_1 \subseteq_\epsilon D_2$ 成立), 则 $Q(DCB)$ 是与 D_1 相 ϵ -一致的文档实例的集合.

规则 2. 如果在 D_ϵ 中存在一个模糊 DTD D_1 使得 $D_1 \otimes_\epsilon Q$ (同上), 则删除模糊 DTDs 集合 D_ϵ 中所有满足条件 $D' \subseteq_\epsilon D_1$ 的模糊 DTDs D' , 并且去除所有与 D_1 相 ϵ -一致的文档实例.

规则 3. 如果在 D_ϵ 中存在一个模糊 DTD D_1 使得 $D_1 \subseteq_\epsilon Q$ (同上), 则每一个满足 $d \in d_\epsilon(D_1)$ 的文档实例 d 也属于 $Q(DCB)$, 即 $d \in Q(DCB)$, 并且删除模糊 DTDs 集合 D_ϵ 中所有满足条件 $D' \subseteq_\epsilon D_1$ 的模糊 DTDs D' .

规则 4. 如果在 D_ϵ 中存在一个模糊 DTD D_1 使得 $Q \subseteq_\epsilon D_1$ (同上), 也就是说, 每一个满足 $d \in d_\epsilon(Q)$ 的文档实例 d 也满足 $d \in d_\epsilon(D_1)$, 则能删除所有与 D' (其中 $D' \in D_\epsilon$ 且 $D' \otimes_\epsilon D_1$) 相 ϵ -一致的文档实例.

规则 5. 如果模糊 DTDs 集合 D_ϵ 中的模糊 DTD D_1 不满足上述规则 1~4 中的条件, 则删除 D_1 .

5 相关工作

基于描述逻辑较强的表达能力和有效的推理机制, 近年来, 许多研究工作者研究了如何利用描述逻辑对各种数据模型(如 ER 模型、面向对象模型以及 XML 模型等)进行表示与推理, 并指出了描述逻辑在数据建模中所具有的独特优势: 能够正确表示数据的语义, 验证数据模式的一致性以及减少数据模式表示的冗余性等^[3,23]. 文献[3,34]实现了从 ER 模型到描述逻辑知识库的转化; 文献[12]建立了描述逻辑与面向对象模型之间的对应关系. 针对 XML 模型的表示与推理; 文献[3]研究了如何利用描述逻辑对 XML 模型进行表示, 实现了从 XML 模型到描述逻辑知识库的转化; 文献[4]讨论了 XML 模型的一致性问题, 即判定一个 XML 文档是否与给定的 XML DTD 相一致; 文献[5]指出了在众多知识表示方法中描述逻辑能够很好地表示和推理 XML 模型; 文献[6]提出了一种描述逻辑 DLFD_{reg}, 实现了从 XML 到 DLFD_{reg} 知识库的转化, 并研究了 XML DTDs 的结构等价推理问题; 文献[7]研究了如何利用描述逻辑 ALCIF(D) 对 XML Schema 进行表示和推理; 文献[8]建立了描述逻辑与 XML 之间的对应关系, 利用描述逻辑的推理机制研究了 XML 的相关推理问题; 文献[9]研究了如何利用时序逻辑对 XML 进行推理; 文献[10]也研究了 XML DTDs 之间的等价以及包含等推理问题.

然而上述研究工作只能处理精确知识, 不能处

理现实世界应用中广泛存在的模糊和不精确知识. 为了使数据模型能够表示和处理模糊知识, 许多研究工作对各种数据模型进行了模糊化扩展, 提出了多种形式的模糊数据模型如模糊 ER 模型^[14,23]、模糊面向对象模型^[14]以及模糊 UML 模型^[17]等. 针对 XML 中模糊和不精确知识的处理问题, 许多研究工作对传统 XML 模型进行了模糊化扩展. Turowski 等人^[16]研究了如何模糊化扩展 XML DTDs 以满足模糊信息交换的需求; Ma 等人^[17]和 Gaurav 等人^[18]研究了如何在 XML 文档中表示模糊数据, 都给出了从模糊 XML 模型到模糊关系数据库的形式化转化方法; Oliboni 等人^[19]提出了一种基于 XML Schema 的模糊 XML 模型; Lee 等人^[20]从数据库的角度入手, 提出了一个用于模糊面向对象建模的 XML 模糊扩展模式; 最近, Ma 等人^[21]详细介绍了有关如何在 XML 中处理模糊和不精确知识的相关研究工作. 有关模糊数据模型(如模糊 ER 模型、模糊面向对象模型以及模糊 XML 模型等)的综述可见文献^[13-14,21].

同样, 针对描述逻辑中模糊和不精确知识的表示和推理问题, 许多研究者提出了多种形式的模糊描述逻辑. Straccia^[22]提出了模糊描述逻辑 FALC, 并给出了相应的推理算法; 蒋运承等人^[23,27]提出了模糊描述逻辑 FALNUI 和面向语义 Web 表示的 FSHOIQ^[27]; 李言辉等人^[25]提出了一种支持数量约束的扩展模糊描述逻辑 EFALCN; 康达周等人^[26]提出了扩展模糊描述逻辑 EFALC_{R+}, 给出了受限 TBox 约束下的 EFALC_{R+} 推理算法; Stoilos 等人^[28]提出了模糊描述逻辑 f -SHIN, 并给出了检查 ABox 一致性的完整推理算法; 王海龙等人^[35]提出了支持用户自定义数据类型表示的模糊描述逻辑 F-SHOIQ(G). 有关模糊描述逻辑的综述可见文献^[15].

基于模糊数据模型和模糊描述逻辑被广泛使用的现状, 与经典情况相类似, 利用模糊描述逻辑来表示和推理模糊数据模型也逐渐引起了人们的关注. 蒋运承等人^[23]和 Zhang 等人^[29]分别实现了从模糊 ER 模型到模糊描述逻辑 FALNUI 和 FDLR 知识库的转化, 从而可以利用模糊描述逻辑对模糊 ER 模型进行推理; Zhang 等人^[36]和 Ma 等人^[37]实现了从模糊 UML 模型到模糊描述逻辑知识库的转化, 然后借助模糊描述逻辑的推理能力研究了模糊 UML 模型的推理问题. 有关利用模糊描述逻辑对模糊 XML 模型进行表示与推理的研究, 目前尚未见有文

章发表.

6 结束语

本文研究了基于模糊描述逻辑的模糊 XML 模型的表示与推理问题. 首先给出了模糊 XML 模型(包括模糊 XML 文档和模糊 DTD)的形式化定义. 然后针对模糊 XML 模型的特点和推理需求, 提出了一种模糊描述逻辑 f -ALCQ_{wf-reg}, 给出了它的语法、语义、知识库以及推理算法. 在此基础上, 实现了从模糊 XML 模型到 f -ALCQ_{wf-reg} 知识库的转化, 并进一步利用 f -ALCQ_{wf-reg} 的推理机制研究了模糊 XML 模型的一致性、包含性以及等价性等推理问题, 从而可以利用模糊描述逻辑对模糊 XML 模型进行推理, 最后指出了利用上述转化和推理结果可以更有效地支持模糊 XML 查询. 本文的研究工作将有助于模糊 XML 建模、推理以及查询优化等相关问题.

进一步的工作包括: 研究 f -ALCQ_{wf-reg} 推理算法的实现技术(即推理机的实现)和分析算法的复杂度; 将本文提出的方法应用到 XML Schema 中, 研究基于模糊描述逻辑的模糊 XML Schema 的表示和推理问题; 研究模糊描述逻辑与其它模糊数据模型的对应关系.

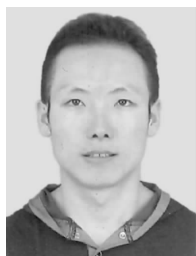
致 谢 向对本文提出宝贵意见的评审专家表示衷心的感谢!

参 考 文 献

- [1] Bray T, Paoli J, Sperberg-McQueen C M. Extensible Markup Language (XML) 1.0. W3C Recommendation. <http://www.w3.org/TR/1998/REC-xml-19980210>, 1998
- [2] Antoniou G, van Harmelen F. A Semantic Web Primer. England: MIT Press, 2004
- [3] Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P F et al. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge: Cambridge University Press, Chapters 1, 4, 2003
- [4] Arenas M, Fan W, Libkin L. On verifying consistency of XML specifications//Proceedings of the ACM Symposium on Principles of Database Systems. New York, 2002: 259-270
- [5] Calvanese D, Giacomo G D, Lenzerini M. What can knowledge representation do for semi-structured data?//Proceedings of the AAAI. Madison, Wisconsin, 1998: 205-210

- [6] Toman D, Weddell G. On reasoning about structural equality in XML: A description logic approach. *Theoretical Computer Science*, 2005, 336(11): 181-203
- [7] Wu X, Ratcliffe D, Cameron M. XML schema representation and reasoning: A description logic Method//Proceedings of the IEEE Congress on Services. Honolulu, Hawaii, USA, 2008; 487-494
- [8] Calvanese D, De Giacomo G, Lenzerini M. Representing and reasoning on XML documents: A description logic approach. *Journal of Logic Computer*, 1999, 9(3): 295-318
- [9] Libkin L, Sirangelo C. Reasoning about XML with temporal logics and automata//Proceedings of the LPAR 2008. Doha, Qatar, 2008; 97-112
- [10] Wood D. Standard Generalized Markup Language: Mathematical and philosophical issues. Jan van Leeuwen, ed. *Computer Science Today*. Germany: Springer-verlag, 1995; 344-365
- [11] Shi Zhong-Zhi, Jiang Yun-Cheng, Zhang Hai-Jun, Dong Ming-Kai. Agent service matchmaking based on description logic. *Chinese Journal of Computers*, 2004, 27(5): 625-635 (in Chinese)
(史忠植, 蒋运承, 张海俊, 董明楷. 基于描述逻辑的主体服务匹配. *计算机学报*, 2004, 27(5): 625-635)
- [12] Calvanese D, Lenzerini M, Nardi D. Unifying class-based representation formalisms. *Journal of Artificial Intelligence Research*, 1999, 11(2): 199-240
- [13] He Xin-Gui. *Fuzzy Database System*. Beijing: Tsinghua University Press, 1994(in Chinese)
(何新贵. *模糊数据库系统*. 北京: 清华大学出版社, 1994)
- [14] Ma Z M, Yan Li. A literature overview of fuzzy database models. *Journal of Information Science and Engineering*, 2008, 24(1): 189-202
- [15] Lukaszewicz T, Stracciab U. Managing uncertainty and vagueness in description logics for the semantic Web. *Web Semantics: Science. Services and Agents on the World Wide Web*, 2008, 6: 291-308
- [16] Turowski K, Weng U. Representing and processing fuzzy information-an XML-based approach. *Knowledge-Based Systems*, 2002, 15(1-2): 67-75
- [17] Ma Z M, Yan Li. Fuzzy XML data modeling with the UML and relational data models. *Data and Knowledge Engineering*, 2007, 63(3): 972-996
- [18] Gaurav A, Alhaji R. Incorporating fuzziness in XML and mapping fuzzy relational data into fuzzy XML//Proceedings of the 2006 ACM Symposium on Applied Computing. Dijon, France, 2006; 456-460
- [19] Oliboni B, Pozzani G. Representing fuzzy information by using XML schema//Proceedings of the DEXA 2008. Turin, Italy, 2008; 683-687
- [20] Lee J, Fanjiang Y Y. Modeling imprecise requirements with XML. *Information and Software Technology*, 2003, 45(7): 445-460
- [21] Ma Z M, Yan Li. *Soft Computing in XML Data Management*. Germany: Springer-verlag, 2010
- [22] Straccia U. A fuzzy description logic//Proceedings of the AAAI'98. Wisconsin, Madison, 1998; 594-599
- [23] Jiang Yun-Cheng, Tang Yong, Wang Ju. Fuzzy ER modeling with description logics. *Journal of Software*, 2006, 17(1): 20-30(in Chinese)
(蒋运承, 汤庸, 王驹. 基于描述逻辑的模糊 ER 模型. *软件学报*, 2006, 17(1): 20-30)
- [24] Jiang Yun-cheng, Tang Yong, Wang Ju, Shen Yu-Ming. A tableaux decision procedure for fuzzy description logic FALNUI. *Journal of Computer Research and Development*, 2007, 44(8): 1309-1316(in Chinese)
(蒋运承, 汤庸, 王驹, 申宇铭. 模糊描述逻辑 FALNUI 的 tableaux 推理. *计算机研究与发展*, 2007, 44(8): 1309-1316)
- [25] Li Yan-Hui, Xu Bao-Wen, Lu Jian-Jiang, Kang Da-Zhou. On computational complexity of the extended fuzzy description logic with numerical restriction. *Journal of Software*, 2006, 17(5): 968-975(in Chinese)
(李言辉, 徐宝文, 陆建江, 康达周. 支持数量约束的扩展模糊描述逻辑复杂性研究. *软件学报*, 2006, 17(5): 968-975)
- [26] Kang Da-Zhou, Xu Bao-Wen, Lu Jian-Jiang, Li Yan-Hui. Reasoning within extended fuzzy description logic supporting terminological axiom restrictions. *Journal of Software*, 2007, 18(7): 1563-1572(in Chinese)
(康达周, 徐宝文, 陆建江, 李言辉. 支持术语公理约束的扩展模糊描述逻辑推理. *软件学报*, 2007, 18(7): 1563-1572)
- [27] Jiang Yun-Cheng, Shi Zhong-Zhi, Tang Yong et al. Fuzzy description logic for semantics representation of the semantic Web. *Journal of Software*, 2007, 18(6): 1257-1269(in Chinese)
(蒋运承, 史忠植, 汤庸等. 面向语义 Web 语义表示的模糊描述逻辑. *软件学报*, 2007, 18(6): 1257-1269)
- [28] Stoilos G, Stamou G, Tzouvaras V, Pan J Z, Horrocks I. Reasoning with very expressive fuzzy description logics. *Journal of Artificial Intelligence Research*, 2007, 30(8): 273-320
- [29] Zhang Fu, Ma Z M, Yan Li. Representation and reasoning of fuzzy ER model with description logic//Proceedings of the 17th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2008). Hong Kong, 2008; 1358-1365
- [30] De Giacomo G, Lenzerini M. TBox and ABox reasoning in expressive description logics//Proceedings of the 5th International Conference on the Principles of Knowledge Representation and Reasoning (KR'96). Cambridge, Massachusetts, USA, 1996; 316-327
- [31] De Giacomo G, Lenzerini M. Boosting the correspondence between description logics and propositional dynamic logics//Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94). Seattle, WA, USA, 1994; 205-212

- [32] Lee K, Min J, Park K. A design and implementation of XML-based mediation Framework (XMF) for integration of Internet information resources//Proceedings of the 35th Hawaii International Conference on System Sciences. Big Island, Hawaii, USA, 2002: 202-210
- [33] Zadeh L. A. Fuzzy sets. *Information and Control*, 1965, 8 (3): 338-353
- [34] Xu Zhuo-Ming, Dong Yi-Sheng, Lu Yang. Semantics-preserving translation from ER schema to OWL DL ontology. *Chinese Journal of Computers*, 2006, 29(10): 1786-1796 (in Chinese)
(许卓明, 董逸生, 陆阳. 从 ER 模式到 OWL DL 本体的语义保持的翻译. *计算机学报*, 2006, 29(10): 1786-1796)
- [35] Wang Hai-Long, Ma Zong-Min, Yan Li et al. Fuzzy description logic F-SHOIQ(G) supporting representation of fuzzy data types. *Chinese Journal of Computers*, 2009, 32(8): 1511-1524 (in Chinese)
(王海龙, 马宗民, 严丽等. 支持模糊数据类型表示的模糊描述逻辑 F-SHOIQ(G). *计算机学报*, 2009, 32(8): 1511-1524)
- [36] Zhang Fu, Ma Z M et al. Fuzzy semantic web ontology learning from fuzzy UML model//Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009). Hong Kong, 2009: 1007-1016
- [37] Ma Z M, Zhang Fu, Yan Li, Cheng Jing-Wei. Representing and reasoning on fuzzy UML models: A description logic approach. *Expert Systems with Applications*, 2011, 38(3): 2536-2549



ZHANG Fu, born in 1984, Ph. D. candidate. His research interests include semantic Web, description logics and XML.

YAN Li, born in 1964, Ph. D., associate professor. Her research interests include databases and XML.

MA Zong-Min, born in 1965, Ph. D., professor, Ph. D. supervisor. His research interests include intelligent data and knowledge engineering.

WANG Xing, born in 1983, Ph. D.. His research interests include semantic Web and rules.

Background

Currently, how to efficiently reason on the data models has attracted a wide range of concerns. Description Logics (DLs), which are a family of class-based knowledge representation formalism, have the high expressive power and the effective reasoning service. On this basis, DLs have been studied in depth in classical data modeling (such as ER modeling, Object-oriented data modeling, and XML modeling). However, information imprecision and uncertainty exist in many real-world applications and hence fuzzy data modeling has been extensively investigated in various data models. In particular, XML has reached a wide acceptance as the data exchange format in these applications, and thus how to represent these non-crisp data within XML has received much

attention in recent years. Furthermore, if the relationships between fuzzy XML models and fuzzy DLs can be established, the reasoning problems of fuzzy XML models (such as conformance, equivalence, inclusion, and disjointness) may be reasoned through reasoning mechanism of fuzzy DLs. In this study, the paper investigates the representation and reasoning of fuzzy XML models with fuzzy description logic.

This work is supported by the National Science Foundation of China (60873010 and 61073139), the Program for New Century Excellent Talents in University (NCET-05-0288), and the Fundamental Research Funds for the Central Universities (N090504005, N090604012 and N100604017)