

# 基于概念代数的上下文本体建模

安敬民

(大连海事大学信息科学技术学院 大连 116026)

**摘 要** 本体与本体之间存在异构问题,也是目前本体在应用上的一个大问题,本体与本体之间没有较好的互操作性。之前已经有人研究过使用本体集成的方法来解决本体异构问题,主要有:本体的合并,本体映射和本体对齐三种方法。本文提出了一种基于概念代数的上下文本体建模方法,对上下文抽象化得到形式背景,利用概念代数对形式背景加深表达深度和增强其形式化程度,进而得到形式背景的概念代数表示(概念),并借鉴其运算方法将概念与概念建立起概念网的形式,使得本体有更好扩展性,进而实现不同本体间的概念成网,最终为本体互操作服务。本文只是对其的一个初步研究。

**关键字** 概念代数;上下文本体;概念网;互操作性

中图法分类号 \*\*\*\*\* DOI号 \*投稿时不提供 DOI号\* 分类号

## Context Ontology Modeling Based on Concept Algebra

An Jing min

(Department of Information Science and Technology, Dalian Maritime University, Dalian, 116026, China)

**Abstract** Inter-ontology heterogeneity is still a critical problem in ontology application. Inter-ontology Interoperability is not good. Previously ontology was used to solve the problem of ontology heterogeneity by some ones, including ontology merging, ontology mapping and ontology alignment. This paper proposes a concept algebra-based context ontology modeling method, the concepts' expression depth in ontology is deepened and formal expression is enhanced by concept algebra. And using for reference to the concept algebra operation method, the inter-concepts are established as the form of concept network to make the ontology better extensible. And the concepts of different inter-ontologies form networks, then ontology interoperability services are realized ultimately. This paper is only a preliminary study on it.

**Key words** concept algebra; context ontology; concept network; interoperability

## 0 引言

万维网是基于语法网,万维网上的信息和数据往往是人可以理解而机器不能理解的,而且随着网络不断发展,网上信息量的日益加剧,导致基于关键词的检索的查准率和查全率大幅下降;所以,在

2000年,万维网的创造者 Tim Berners-Lee<sup>[10]</sup>提出了语义网(Semantic Web)的概念。本体(Ontology)作为一种可以在语义和知识层次结构上对信息进行表示的工具,就成为了语义网的核心部分。但是本体在实际应用上存在着一个大的问题,就是本体与本体之间的异构问题,使得本体间通信困难;这是因为开发本体的团队所处的领域不同,使用的语

收稿日期:年-月-日;最终修改稿收到日期:年-月-日 \*投稿时不填写此项\*。本课题得到……基金中文完整名称(No.项目号)、……基金中文完整名称(No.项目号)、……基金中文完整名称(No.项目号)资助。作者名1(通讯作者),性别,xxxx年生,学位(或目前学历),职称,是/否计算机学会(CCF)会员(提供会员号),主要研究领域为\*\*\*\*、\*\*\*\*.E-mail: \*\*\*\*\*.作者名2(通讯作者),性别,xxxx年生,学位(或目前学历),职称,是/否计算机学会(CCF)会员(提供会员号),主要研究领域为\*\*\*\*、\*\*\*\*.E-mail: \*\*\*\*\*.作者名3(通讯作者),性别,xxxx年生,学位(或目前学历),职称,是/否计算机学会(CCF)会员(提供会员号),主要研究领域为\*\*\*\*、\*\*\*\*.E-mail: \*\*\*\*\*.(给出的电子邮件地址应不会因出国、毕业、更换工作单位等原因而变动。请给出所有作者的电子邮件)  
第1作者手机号码(投稿时必须提供,以便紧急联系,发表时会删除): …… ,E-mail: …… \*此部分6号宋体\*

言不同等等因素造成的。目前对于解决本体间的异构问题主要就是对本体集成,包括本体合并,本体映射以及本体对齐等多种方式;其中韩国栓等人提出了利用概念代数(CA)对本体进行合并,从数学角度,利用概念代数的相关理论和运算规则将两个已有的本体合并成新的本体,效果较为突出。

本文是在其基础上提出利用概念代数构建上下文本体,主要思路是:利用概念代数可以深层次的表达概念和表示的形式化程度深的特点结合其运算规则,将每一个上下文作为一个概念,构建本体,发挥概念代数的特点,使作为概念的上下文构成一个概念网(Concept Network),本体是以概念网的形式构建,具有良好的扩展性,以减少本体间异构问题。

## 1 上下文本体建模

### 1.1 上下文与形式背景

上下文表示事物的状态或属性,也已是过去的状态,也可以是现在的状态;而形式背景定义为一个三元组,形如  $B = (O, A, R)$ ,其中  $O$  为对象,  $A$  属性或状态,  $R$  为对象和属性之间的某种关系,所以根据对应关系可以把上下文视为一个形式背景,可以用形式概念分析(FCA)处理。

### 1.2 本体与基于本体建模

#### 1.2.1 本体

业界对本体做了大量的研究,对本体的定义也是不断改进,在1993年,Gruber提出:“本体是概念模型的明确规范说明”。之后Borst将“共享”和“形式化”的概念引入到本体中,提出:“本体是共享概念模型的形式化规范说明”。Studer本体是对前两者进行了归纳总结,给出:“本体是共享的概念模型的明确形式化规范说明<sup>[8]</sup>”。这是目前最被人们认可的本体定义。

Studer提出的本体概念中,包含四部分:“概念模型(conceptualization)”,是指领域的主要概念;明确的(explicit),是指概念有明确定义;形式化(formal),是指本题中的数据、信息以及概念计算机可读;“共享(share)”,是指领域内共识。

#### 1.2.2 基于本体建模

如上述1.1所述,一个上下文可以视为一个形式背景,进而可以用形式概念分析(FCA)处理,将每个FCA表示的知识、信息和概念构建某种相

互关系(联系),但是由于FCA的形式化表示程度不深,缺少语境(语义环境),会导致本体间异构问题;而概念代数恰能弥补FCA形式化不够深,表示关系不够复杂等缺点,引入概念代数(CA),将上下文深度形式化来构建本体中的每一个概念,使得机器可以理解和推理。其中,以一个概念为起点,利用CA的运算规则,逐渐扩展,将相关的概念关联起来,最终构成本体-概念网(Concept Network)。这样的本体有较好的可扩展性,能为本体与本体之间的异构解决问题上提供一个新的视角。

## 2 概念代数

根据文献<sup>[4]</sup>,概念代数是对于多概念间的关系、运算和组成复杂关联规则的进行处理的抽象数学结构。目前已经开始使用概念代数对本体进行得研究并不多,时百胜等人将概念代数用于对知识进行表示和推理<sup>[3]</sup>;宋奂寰等<sup>[1]</sup>提出基于概念代数的本体演化模型;韩国栓等<sup>[2]</sup>在本体集成来解决本体间异构的问题上提出利用概念代数合并本体;这三者虽然都是使用概念代数理论,但是后者与前两者不同的是,韩国栓等人使用的是由加拿大卡尔加里大学Yingxu Wang教授提出的概念代数,其运算规则更多,可以表达更加复杂的关联关系;不过其本质都是一样的,都是将概念抽象化,再结合代数运算对本体进行研究,而且都得到了较好的验证,所以,将本体与概念代数二者结合是一个可行的方式方法。本文在韩国栓提出的《概念代数在本体合并中的应用》基础上提出基于概念代数上下文本体建模,前者是将现有本体进行合并,而本文提出的旨在利用概念代数相关理论直接构建新的本体模型,这样本体(概念网CN)就有直接可以实现本体间互相关联的特性,效率更高。根据文献<sup>[4,5,6]</sup>,Yingxu Wang提出的概念代数的相关定义以及韩国栓等将其用于本体上对其的扩展<sup>[2]</sup>。

**定义1** 上下文(Context)是含有语义的情景(语境),如1.1所述,本文将上下文(语义环境)类比一个形式背景,形式背景(语义环境) $B = (O, A, R)$ ,其中  $O$  为对象集,  $A$  属性或状态集,  $R$  为对象和属性之间的关系集(以下的“:”均表示定义),其中  $R: O \rightarrow O \mid O \rightarrow A \mid A \rightarrow O \mid A \rightarrow A$ 。

**定义2** 根据Wille的形式概念分析<sup>[7]</sup>,形式背景可以用形式概念分析处理,定义为一个二元组

$(O, A)$ ; 对其进行扩展, 在概念代数中一个语义环境的形式概念  $C$  表示为一个五元组:  $C: (M, N, Rc, Ri, Ro)$ , 其中  $M \subseteq O$  是对象集,  $N \subseteq A$  是属性集合;  $Rc \subseteq M \times N$ , 表示对象与属性之间的关系,  $Ri \subseteq N \times N_1$  ( $N_1$  是其他概念  $C_1$  的内涵), 表示概念  $C$  与  $C_1$  的输入关系(这里的关系是从人类认知角度出发的, 是指脑内部与外部的关系, 本文是将一个概念作为脑内知识, 其他均为脑外概念), 其中  $N_1 \subseteq O$  且  $N_1 \not\subseteq N$ , 也即从其他概念到概念  $C$  的关系;  $Rc$  恰与  $Ri$  相反, 是输出关系。

**定义 3** 一个语义环境  $B$  下的概念代数  $CA$  定义为一个三元组,  $CA: (C, OP, B) = (\{O, A, Rc, Ri, Ro\}, \{OPr, OPc\}, B)$ , 其中  $\{OPr, OPc\}$  分别表示关系运算集合组合运算集, 具体如下所述:

关系运算集  $OPr: \{\leftrightarrow, \prec, \succ, =, \cong, \sim, \nleftrightarrow\}$ , 集合中的元素分别表示概念之间的相关 (relation): 若  $C_1 \leftrightarrow C_2: A_1 \cap A_2 \neq \emptyset$ ; 子概念 (sub-concept) 与超概念 (super-concept):  $C_1 \prec C_2, A_2 \subset A_1$ , 则  $C_1$  为  $C_2$  的子概念,  $C_2$  为  $C_1$  的超概念以及等价 (equivalent):  $C_1 = C_2: (A_2 = A_1) \wedge (O_1 = O_2)$ ,  $C_1$  和  $C_2$  等价; 一致 (consistent):  $C_1 \cong C_2: (C_1 \prec C_2) \vee (C_1 \succ C_2)$  和类比 (comparison):  $C_1 \sim C_2: \#(A_1 \cap A_2) / \#(A_1 \cup A_2)$ , 即内涵交集与并集的比值, 比值在 0 到 1 之间, 如果值为 1 则是等价关系; 不相关 (independent): 与相关性相反有  $A_1 \cap A_2 = \emptyset$ 。还包括因果关系 ( $\bullet \rightarrow \bullet$ ) 和行为关系 ( $do: ?$ )<sup>[2]</sup>,  $C_1 \bullet \rightarrow C_2$  表示  $C_1$  是  $C_2$  的因,  $C_1 do: C_2$  表示  $C_2$  是  $C_1$  的一种行为的被作用对象 ( $?$  表示 to do)。

利用概念的关系运算, 可判断概念之间的关系, 进而对概念进行关联组合。

组成运算集,  $OPc: \{\Rightarrow, \Rightarrow, \pm, \cong, \text{世}, \text{世}, \text{世}, \text{世}, \text{世}, \text{世}\}$ , 其中集合中的元素分别表示为继承 (inheritance):  $P \Rightarrow S$  表示概念  $S$  继承了其父概念  $P$  的所有属性和对象; 剪裁 (tailoring):  $P \Rightarrow S$ , 是继承关系的一种, 从父概念中的对象或者属性中删除一些不需要的对象或者属性, 得到概念  $S$ ; 扩展 (extension):  $P \pm S$ , 是继承的特殊关系的一种, 即从父概念中继承其对象和属性, 在这基础上添加额外的对象或者属性, 得到  $S$ ; 替换 (substitution):  $P \cong S$ , 是继承关系的一种特殊关系, 指替换了从父概念  $P$  继承来的某个属性或者对象; 组合 (composition) 和分解 (decomposition): 将两个或者多个概念组合为一个整体, 形成一个新概念, 分解恰与其相反, 二者不属于继承运算, 属于 part-of

关系; 泛化 (aggregation) 和特化 (specification),  $c_1 \Leftarrow c_2$ , 是对概念  $c_2$  进行抽象, 除去  $c_2$  内涵中的某些属性, 得到  $c_1$ ; 特化与泛化进行相反的运算; 实例化 (instantiation):  $c_1 \mapsto c_2, c_2$  是概念  $c_1$  具体的一个实例。上述提到的因果关系和行为关系分别由  $\Rightarrow$  表示和具体动作表示<sup>[2]</sup>。

上述综述的概念代数相关的运算规则在本体实际应用上已经得到了较好的验证 (《概念代数在本体合并中的应用》), 所以合理使用这些规则构建上下文本体是可行的, 而且上述各规则与已经在本体中经常使用的形式概念当中关系有着对应关系, 如继承关系对应 kind-of 关系; 组合与分解对应 part-of 关系; 泛化和特化对应 kind-of 关系; 实例化对应 instance-of 关系<sup>[2]</sup>。

**定义 4<sup>[2]</sup>** 概念网  $CN$  使用的是概念代数规则构成的,  $CN$  形式化表示为:

$$CN = R: \bigcup_{i=1}^n C_i \rightarrow \bigcup_{j=1}^n C_j, \text{ 其中 } \bigcup_{i=1}^n C_i \text{ 表示}$$

$n$  个概念的集合,  $R$  是上述综述的概念代数运算规则。

### 3 基于概念代数的上下文本体建模

上述是概念代数中的运算规则, 本文将其引入到上下文本体建模当中, 将一个上下文视为一个形式背景, 利用概念代数对形式背景进行深度形式化和利用运算规则进行复杂计算 (含有语境) 如上述 1.1 和 1.2.2 所述。将一个概念作为起点, 将与其有关联关系 (上述提到的运算关系) 的外部概念联系起来, 从一个独立的概念不断扩展成多个概念的关系集合, 最后成为一个概念网 (CN) - 本体。如下图 1 所示。

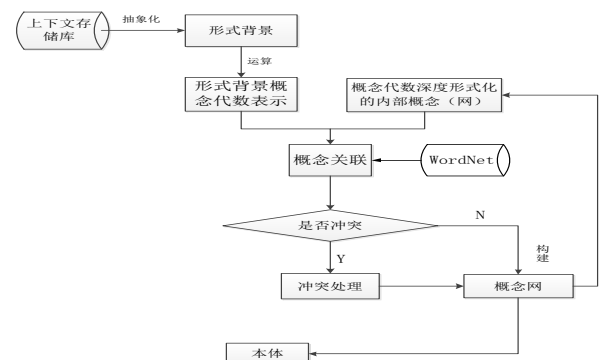


图 1 基于概念代数的上下文本体建模框架图

### 3.1 上下文抽象化

上下文抽象化：利用上下文存储库中的启发性知识判断上下文是否满足一致性，若不满足不一致性需要领域专家对不一致性处理，使用低层采集的上下文推理高层上下文的推理原则，得到高层上下文，并对高层上下文进行解析，得到对象，属性集和对象与属性的关系集；将得到的结构构建形式背景，其对应关系如下表 1 所示。

表 1 上下文与形式背景对应关系

上下文中要素	形式背景中要素
对象	形式对象
属性或状态	属性
对象与属性关系 (抽象化)	形式对象和属性关系

### 3.2 概念间的关联-概念网

对由上下文抽象化（如上述 3.1 所述）得到的形式背景，使用概念代数对其运算，得到形式背景的概念代数表示，与初始概念关联合并，借用 WordNet 知识库对概念进一步判断概念间关系，如有冲突，需要使用 WordNet 或者领域专家进行处理，否则合成构建概念网。

关于概念网中概念间相互关联的关联规则<sup>[21]</sup>如下：

**规则 1** 根据 WordNet 判断概念之间（内部概念和外部概念）关系，以当前概念为核心，若其他概念为上位词，利用泛化表示；若为下位词，用继承（扩展、裁剪或替换）表示。

**规则 2** 对整体和局部关系，以当前概念网为核心，若为上位词，利用组合表示；若为下位词利用分解表示。

**规则 3** 对于概念之间的属性关系，需要进一步判断，是属于因果关系还是行为关系。

根据以上由韩国栓等<sup>[21]</sup>提出的概念之间的关联规则，提出形式背景概念代数表示-概念与内部概念构成概念网（CN）算法描述：

(1) 初始化概念网 L，初始化内部概念 C<sub>1</sub>，以为 C<sub>1</sub> 中心。

(2) 对上下文抽象化，得到对象，属性以及其之间的关系，构建形式背景 B。

(3) 对 B 进行概念代数运算，深度形式化，记为 C<sub>i</sub>。

(i=2,3,4,……)。

(4) 将 C<sub>i</sub> 加入到概念网 L 中，与 C<sub>1</sub> 或者已有的概念

网进行关联和合并。

(5) 根据 WordNet 判断新加入的概念与 L 中已有的概念是否关联，如果不相关，删除此概念，不加入到概念网 L 中转到执行 (2)；若果相关执行 (6)。

(6) 新概念节点加入到概念网 L 中，根据概念网 L 原有的概念之间关系（局部概念网）以及 WordNet，依据人的常识以及领域专家的知识判断新节点与原有节点的关联关系（上述第 2 节所述的概念代数运算规则）。

(7) 运用概念代数的组成运算和关系运算对概念节点进行运算合并以及关联。

(8) 转到执行 (2)；上下文存储库没有新的上下文，结束操作。

利用上述构建本体算法以及第 2 节介绍的概念代数的关系运算和组成运算规则给出以 Knowledge 这个概念为内部概念（核心）的一个概念网（局部），如下图 2 所示。

E:Extension      T1:tailoring object      T:tailoring  
A:attribute-of      D:decomposion      I:instance-of  
do: 表示行为。

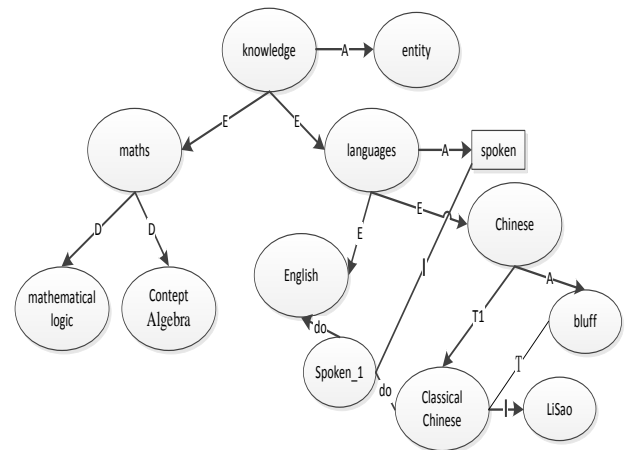


图 2 “knowledge”概念网-本体（局部）图

对图 2 给出的概念网，给出其形式化表示，如下：

$$\text{knowledge} \xrightarrow{\pm} (\text{maths} \upharpoonright (\text{mathematical -logic} \mid \text{concept algebra})) \mid (\text{Chinese} \xrightarrow{\pm} (\text{English} \mid (\text{Chinese} \xrightarrow{\pm} (\text{classical Chinese} \mapsto \text{LiSao}))));$$

classical Chinese do : spoken in English;

### 3.3 基于概念代数的上下文本体建模算法

输入：内部概念 C；上下文存储库；

输出：概念网；

N=merge (C<sub>2</sub>C<sub>1</sub>) //假定从上下文存储库中找到的 C<sub>1</sub> 与 C<sub>2</sub> 有关。

i=3;

while (ContextDB[i]>0) //上下文存储库中不空，有新的上下文

```
{  
If( hasrealitionWordNet(Ci,N)) //根据 WordNet 新概念与  
    概念网有关  
    If( hasrealitionWordNet(Ci,N) .equals(enqivalent))  
        //找到相关节点  
        N=relate(N,Ci); //利用概念代数运算规则进行关联  
    else  
        loop:i++;  
}
```

## 4 对比分析

韩国栓等人提出的利用概念代数实现本体间的合并，主要对象是已有的本体，先将现有本体转化为概念与概念之间的关系（概念网 CN），再进行实现本体间合并，已得到了较好验证；本文利用概念代数主要作用对象是上下文，将上下文抽象为形式背景，进而转化为一个概念（节点），利用一个概念节点和其他上下文构建概念网，合成新本体，合成的新本体具有可以直接可扩展的特性，而且为以后开发本体提供了一个新思路，如下表 2 所示。

表 2 概念代数应用对比对比分析表

题目	对象	方法	目标	特点
《概念代数在 本体合 并中的 应用》	已有本 体	概念代 数	合并本 体	具有可 扩展性， 减轻本 体间异 构性。
《基于 概念代 数的上 下文本 体建模》	上下文	概念代 数	新本体	具有可 可直接 扩展的 特性，减 轻本体 间异构 性。

## 5 结束语

本文是受韩国栓等人将概念代数应用到本体合并上的启发，利用概念代数将抽象化的上下文

（形式背景）构成概念网，进而形成新的本体，该新本体本身就具有可扩展性，这样的本体之间可以之间相互关联，大大减少了原来本体之间的异构性，而也不需要先将本体改成概念网在合并，提高了效率，为构建领域本体提供了一个新的视角。

但是，本文对其只是初步的研究，由于领域本体的复杂性，仍然需要进一步研究实现。

## 参 考 文 献

- [1] 宋奂寰, 张志祥. 基于概念代数的本体演化[J]. 计算机工程, 2009, 35(13): 23-25.
- [2] 韩国栓. 概念代数及其在本体合并中的应用[D]. 大连海事大学, 2013.
- [3] 时百胜, 余泓. 概念知识表示和推理[J]. 小型微型计算机系统, 2006, 27(9): 1618-1622.
- [4] Wang Y. On concept algebra: A denotational mathematical structure for knowledge and software modeling[J]. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 2008, 2(2): 1-19.
- [5] Wang Y, Tian Y, Hu K. Semantic manipulations and formal ontology for machine learning based on concept algebra[J]. 2013.
- [6] Tian Y, Wang Y, Hu K. A knowledge representation tool for autonomous machine learning based on concept algebra[M]//Transactions on Computational Science V. Springer Berlin Heidelberg, 2009: 143-160.
- [7] 甘特尔, 维勒, 垣. 形式概念分析[M]. 科学出版社, 2007
- [8] Gruber T R. A translation approach to portable ontology specifications[J]. Knowledge acquisition, 1993, 5(2): 199-220.
- [9] 李艳娜, 乔秀全, 李晓峰. 基于证据理论的上下文本体建模以及不确定性推理方法[J]. 电子与信息学报, 2010, 32(8): 1806-1811..
- [10] Berners-Lee T, Hendler J, Lassila O. The semantic web[J]. Scientific american, 2001, 284(5): 28-37.

第一作者  
照片  
(高清照片)

**Author1, ...** \*计算机学报第 1 作者提供照片电子图片，尺寸为 1 寸。英文作者介绍内容包括：出生年,学位(或目前学历),职称,主要研究领域（**与中文作者介绍中的研究方向一致**）。\*\*字体为小 5 号 Times New Roman\*

**AuthorX, ...** \*英文作者介绍内容包括：出生年,学位(或目前学历),职称,主要研究领域（**与中文作者介绍中的研究方向一致**）。\*\*字体为小 5 号 Times New Roman\*

## Background

\*论文背景介绍为**英文**，字体为小 5 号 Times New Roman 体\*

论文后面为 400 单词左右的英文背景介绍。介绍的内容包括：

本文研究的问题属于哪一个领域的什么问题。该类问题目前国际上解决到什么程度。

本文将问题解决到什么程度。

课题所属的项目。

项目的意义。

本研究群体以往在这个方向上的研究成果。

本文的成果是解决大课题中的哪一部分，如果涉及 863\973 以及其项目、基金、研究计划，注意这些项目的英文名称应书写正确。