一种基于FCA的面向关系数据库的本体学习方法

欧阳纯萍[1,2] 胡长军[1] 李 扬[1] 刘振宇[1]
(北京科技大学信息工程学院 北京 100083)[1] (南华大学计算机科学与技术学院 衡阳 421001)[2]

摘 要 从已有的数据模型中进行语义提取,经过一定的规则映射生成本体的过程称为本体学习。关系数据库模型是当前数据的存取与组织的主要模型,从中学习得到本体,一直是本体工程领域研究的热点之一。利用手工定义的E-R模型到本体的映射规则来完成本体的构建,是国内外大部分学者采用的方法。但这样获得的本体概念层次关系主观依赖性强,不利于本体的实际应用。为了能更加客观地获取数据之间的概念层次关系与语义信息,提出了一种基于FCA(形式概念分析)从关系数据库进行本体学习的方法。该方法既保持了关系数据表中原有的数据语义关系,又发挥了FCA自动提取语义信息的特点,提高了最终本体生成的质量,有利于在具体的领域应用中使用本体。最后结合材料服役安全数据库的数据信息,演示了运用所提出的方法学习得到领域本体的过程。

关键词 FCA,概念格,关系数据库,本体
中图法分类号 TP391 **文献标识码** A

Approach of Ontology Learning from Relational Database Based on FCA

OUYANG Chun-ping[1,2] HU Chang-jun[1] LI Yang[1] LIU Zhen-yu[1]
(School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China)[1]
(School of Computer Science and Technology, University of South China, Hengyang 421001, China)[2]

Abstract Ontology learning is a process, which extracts semantic information from the existing data model and generates ontology using a set of predefined mapping rules. Relational database is the main model of data access and management, and extracting ontology from relational database is one of the research hotspots in ontology engineering field. A common method adopted by the domestic and foreign scholars is that ontology is constructed by using mapping rules between E-R model and ontology elements. But this method is subjective and it goes against the application of ontology. To address this issue, an approach of ontology learning from relational database based on formal concept analysis was proposed, which could obtain the hierarchical relation of concept and semantic relation of data objectively. The proposed method not only keeps the semantic information of relational data tables, but also shows the advantage of FCA in automatic extraction of semantic information. Thus the quality of final ontology is improved and the application field of ontology is extended. A case study of the proposed method combined with materials service safety database was also presented.

Keywords FCA, Concept lattice, Relational database, Ontology

1 引言

本体通过定义领域概念及其之间的层次关系,能够清晰表达领域数据的语义信息,提高数据源之间的语义交互性。因此,本体作为一种有效的领域数据语义模型和知识表示形式,被广泛地应用于领域数据集成中。基于本体的数据集成可以很好地解决数据源在语法、模型和语义上的异构等问题。

本体学习的过程实际上就是从已有的各类数据模型中进行语义提取,自动组织并生成本体的一个过程。本体学习的对象有词典、知识库、文本、关系模式、半结构化数据(XML Schema)[1,2]。而关系数据模型是现有信息系统中数据的主要存取和组织模型,因此从关系数据库中学习得到本体,再辅助完成异构系统间数据集成的方法,已经成为领域数据集成中的研究热点。

从数据库中抽取概念和关系是基于关系数据库的本体学习方法的核心任务[3]。这些方法大部分是通过读取数据字典来获取关系数据库模式信息(源 Schema),然后再根据源 Schema 与目标 Schema(本体元素)间的对应关系,定义一组映射规则,从而实现关系模式向本体的转换。如 Kashyap[4] 采用逆向工程方法从数据库 Schema 中学习本体。Rubin[5] 等提出了一种从外部关系数据源中学习本体的实例和属性值的方法。Ljiljana Stojanovi[6] 等采用数据库反向工程与语义 Web 自动标注两种技术的结合来得到语义网本体模型。国内也有一些学者致力于从关系数据库到本体的映射研究。如

到稿日期:2011-01-14 返修日期:2011-03-29 本文受国家863高技术研究发展计划基金项目(2008AA01Z109),国家"十一五"科技支撑计划(2006BAK11B03),国家科技基础条件平台(2005DKA32800)资助。
欧阳纯萍(1979—),女,博士生,讲师,主要研究方向为语义 Web 与领域数据集成等,E-mail:ouyangcp@gmail.com;胡长军(1963—),男,教授,博士生导师,主要研究方向为数据工程、高性能计算等;李 扬(1983—),女,博士后,主要研究方向为数据集成与进化计算等。

许卓明等提出了从 E-R 模型到 DL 本体语义保持的翻译[7]。王洪伟等提出了一种面向关系模式的基于逆向工程的领域本体获取方法[8]。这些方法主要考虑的是本体概念与概念之间的层次关系(如父类-子类关系)的提取与表示。但是,靠手工从数据表或 E-R 模型中获取的概念层次关系会过于扁平,概念之间的语义信息不能充分体现,因此由此类方法所得到的本体大都属于轻量级,给本体的进一步应用和重用带来了困难。

为了能更加客观地获取数据之间的概念层次关系与语义信息,本文提出了一种基于 FCA 从关系数据库进行本体学习的方法。首先对数据表中的数据进行预处理,然后针对数据表中的元组数据建立"对象-属性"的单值形式背景,最后运用形式概念分析的方法对形式背景中的"对象-属性"进行概念格构造,从中发现隐含的概念及其关系,进而生成本体原型。本文第 2 节将介绍形式背景与概念格的基本定义,分析 FCA 与本体的关系;第 3 节以材料服役数据为实例详细描述如何从关系数据模式中进行本体学习的方法;最后给出结论。

2 FCA 与本体

FCA(Formal Concept Analysis)理论最初由 Rudolf Wille[9,10]提出,也称之为形式概念分析,是一种从数据集合(data sets)中发现概念结构的数据分析方法,并且将这些概念结构以图形化的方式表现出来,以探索数据之间的关联关系。

2.1 形式背景与概念格

定义 1 一个形式背景 C(Formal Context)是由 (G,M,I) 三元组构成的,其中 G 是对象(Objects)的集合,M 是属性(Attributes)的集合,I 是 G 和 M 之间的二元关系的组合。可以把形式背景表示为 $C:=(G,M,I)$,且 $I\in G\times M$。

定义 2 G 和 M 的二元关系可以定义为 $(g,m)\in I$,表示"对象 g 的属性为 m",则对所有的 $A(A\subseteq G)$ 来说,可以定义为 $A'=\{m\in M\mid \forall g\in A:(g,m)\in I\}$,表示 A 中全体对象所共有的属性集合。同理,对于所有的 $B(B\subseteq M)$,也可以定义为 $B'=\{g\in G\mid \forall m\in B:(g,m)\in I\}$,表示包含所有 B 中属性的全体对象的集合。

表 1 所列是一个用矩阵表示的关于人与食物的形式背景。行用来表示对象,列用来表示属性,行 g 和列 m 的交叉处则表示对象 g 具有属性 m。

表 1 人与食物的形式背景

对象	属性			
	鱼肉	牛肉	猪肉	鸡肉
Jack	×			×
John	×	×	×	
Katy	×		×	×
Mary	×		×	×

定义 3 形式背景 (G,M,I) 的一个形式概念可以用对 (A,B) 来表示,其中 $A\in G,B\in M$,且满足 $A'=B,B'=A$,A 称为概念 (A,B) 的外延,B 称为概念 (A,B) 的内涵。对于 (G,M,I) 的所有概念,如果存在 $A_1\subseteq A_2$(等同于 $B_2\subseteq B_1$),则记为 $(A_1,B_1)\leqslant(A_2,B_2)$,这些具有层次关系的概念集合被记作 $\beta(G,M,I)$,也叫做形式背景 (G,M,I) 的概念格(concept lattice)。

概念格从本质上描述了概念之间的泛化和特化的关系,非常适合于发现数据间潜在的关系。从形式背景中生成概念格的过程实质上是一种概念聚类的过程[11]。由表 1 的形式背景生成的概念格如图 1 所示,图中每个节点表示一个概念节点。一个概念节点包含 2 个特征,分别是对象特征 A 和属性特征 B。

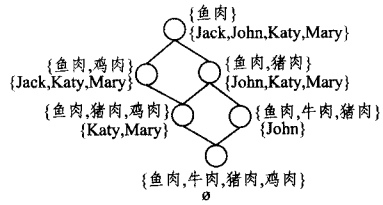

图 1 人与食物的概念格

综合以上关于形式背景和概念格的探讨,可以发现概念格是利用对象与属性间的关系由上而下建立的,并以图形化的形式呈现。在图形化的概念格表示中,可以发现每个可能性的抽象概念及分类,本体的构建过程也可以利用此重复抽象概念的过程,由上而下选择领域知识中感兴趣的分类,用于作为构建领域本体的指导原则。

2.2 本体与 FCA 的联系

目前,FCA 作为一种有效的数据分析方法已经在信息检索、软件工程、知识发现等领域得到了广泛的应用。近年来,在本体工程领域也开始有人利用 FCA 来辅助本体概念的分析和建模,并逐渐成为了研究的热点。在实际应用当中,本体的构建与 FCA 的分析过程之间是既有联系又有区别的。

FCA 与本体的相同之处在于它们都来源于哲学,同样采用形式化方法描述概念及概念之间的关系,都强调模型的形式说明的必要性。但是它们之间又存在着许多不同的地方。构建本体的目的是对现实世界建立共享的概念模型,从而支持知识密集型的领域应用;而 FCA 则不是为现实建模,是为人工世界建模,目的是支持用户在给定数据的基础上进行领域分析和建模。领域本体的构建可以在没有任何数据的前提下完成,但是 FCA 则必须是在给定数据集(形式背景)的基础上进行分析[12]。本体主要关注概念间关系的形式化表示,而 FCA 则集中在概念的分类表示上。所以两者可以相互结合,互为补充。

Philipp 在文献[13]中指出 FCA 和 Ontology 存在着双向的互动关系,FCA 可以是本体论工程的一项技术,藉由概念格来辅助获取结构化的数据信息,或者可以从已知的对象数据集中提取出有用的概念层次,作为构建本体的基础。把 FCA 和本体论进行结合,可以帮助领域知识本体的建模,在以下两个方面进行改善。

(1)解决了类别架构不够弹性的问题。无论是现有的信息分类系统,还是 semantic web 上的本体分类系统,当类别架构固定了就通常很难进行变更。不管是资料库形式还是目录形式的信息系统,都必须根据事先定义好的类别架构对信息进行分类,然而随着信息的快速增加和变动,这种不够弹性的预定义分类系统的缺点就逐渐显现出来。

(2)改进了概念相关性与信息资源重要性的问题。现有的类别架构所表现的概念相关性通常是绝对性的,无法表示出相同类别层次的概念的相对重要性,从而导致概念搜索的

路径固定且容易遗漏重要的信息。因此，如果能够区分不同类别的概念相关性和资源重要性，则能提高查询的准确率。

3 基于 FCA 和关系数据库的本体学习方法

目前，关系数据库仍然是众多应用软件系统进行数据存储和管理的主要模式，它的特点可以概括为：

(1)关系数据库主要是由二维数据表和存储在数据表中的元组构成的。数据库中的一个表就是一个关系，表和表之间则通过主键和外键来约束它们的关系。

(2)关系数据库的设计面向特定的应用，因此表与表之间的关系与领域应用的目标有很大的相关度，它能提供面向某个特定领域应用的领域概念及部分关系。

(3)虽然关系数据库表和表之间的约束描述了一部分数据的语义信息，但是由于其受到已定义的表结构的约束，对于复杂对象的语义信息无法表述，因此不适合用于数据类型众多而语义关联复杂的领域信息系统建模。

对比关系模型，本体则是一种能表达复杂语义的数据模型。因此，本文所提出的基于关系数据库进行本体学习的主要任务就是利用 FCA 分析关系模型中蕴涵的语义信息，并映射成本体中相应的概念和属性。如图 2 所示，基于 FCA 和关系数据库的本体学习方法可以分为 3 个阶段：关系数据表预处理；形式背景构造；概念格到本体模型的映射。

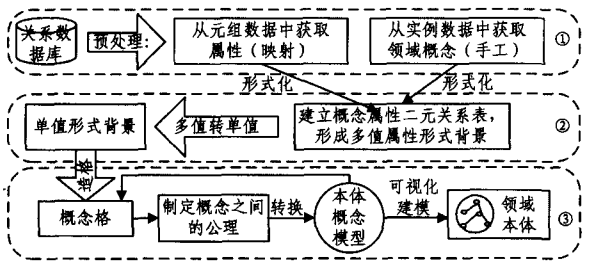

图 2 基于 FCA 的本体学习过程

3.1 关系数据表预处理

Oracle 数据库一直是国内外各大企业使用的主流数据库管理系统，因此，本研究主要针对 Oracle 数据库中的数据类型预处理关系数据表。

首先，给出关系数据表的相关定义。

定义 4 用 R 表示一张二维关系表，则 $T(R)=\{R_1,R_2,\cdots,R_n\}$ 表示一个数据库中所有关系表的集合；$P(R)$ 表示 R 的主码属性集合；$F(R)$ 表示 R 的外码属性集合。

基于上述定义，按照数据表与数据表之间的依赖关系，把数据库的关系数据表分为两类：$T1$，$T2$，其形式化定义如下：

(1)$T1=\{R\mid(\exists R_i(P(R)=F(R_i))\vee\exists R_j(P(R_j)=F(R)))\wedge R,R_i,R_j\in T(R)\}$。$T1$ 中包含的表 R 均满足：任意的 R_i 均能找到与之对应的 R_j，且关系表 R_i 的主码是关系表 R_j 的外码，或关系表 R_j 的主码是关系表 R_i 的外码。

(2)第二类表定义为 $T2=T(R)-T1$，即表示关系数据库中除 $T1$ 以外的关系数据表。

根据以上关系表的分类，$T1$ 当中的表是需要进行形式概念分析的对象，而 $T2$ 当中的表则因为没有依赖关系的存在，暂时不需要进行 FCA 分析，可以在后期的本体建模中手工添加一些 $T2$ 中的元组信息。

针对 $T1$ 当中的表，利用它们的 $P(R)$ 和 $F(R)$ 之间的依赖关系，合并成一张新的二维关系表，步骤如下：

i)把 $T1$ 中所有待处理的数据表依次装入一个有序集合 $list$ 中，$list=\langle R_1,R_2,\cdots,R_n\rangle$，$n$ 为 $T1$ 中表的数目总和，同时定义一个队列 Q；

ii)定义一张空的新表 R_{new}，取 $list$ 集合中的第一个元素 R_1，令 $R_{new}=R_1$，再把 R_1 存入队列 Q 中，同时从有序集合 $list$ 中删除 R_1；

iii)从队列 Q 取出队头元素 R，然后从有序集合 $list$ 的第一个元素开始扫描，寻找 R_i，它满足条件 $P(R)=F(R_i)$ 或者 $F(R)=P(R_i)$，则令 $R_{new}=R_{new}\infty R_i$，同时把 R_i 存入 Q 的队尾，且从 $list$ 中删除 R_i，重复这一步操作直至 $list$ 扫描完毕；

iv)当 Q 不为空的时候，一直重复步骤 iii)；如果 Q 为空，且 $list$ 不为空，则重复然步骤 ii)至 iv)；

v)如果最后 $list$ 集合为空，则停止操作。

经过上述步骤的预处理之后，则可能出现两种情况，第一种情况，整个数据库的表经过预处理后得到的就是一张 R_{new}，可以直接构成对象关系数据库的形式背景 C。第二种情况，得到多个没有定义关联关系的 R_{new}，最后可以根据应用目的，通过手工对 R_{new} 表进行合并操作，形成原始的形式背景 C。

3.2 多值向单值背景的转换

由于数据表中的数据大部分都是多值的，因此在上一节中获取的形式背景 C 也是一个关于对象数据库的多值形式背景。对于多值形式背景定义如下。

定义 5 一个多值形式背景可以用四元组 (G,M,W,I) 表示，它是由对象集合 G、多值属性集合 M、属性值域 W 以及这几个集合之间的一个三元关系 $I(I\in G\times M\times W)$ 构成的，且满足当 $(g,m,w)\in I$ 及 $(g,m,v)\in I$ 时，$w=v$ 成立。

由于形式概念分析方法处理的对象是单值形式背景，因此需把 Oracle 数据库经过预处理后得到的一个多值形式背景 C 进行单值背景转换，用 K_C 表示与 C 对应的单值形式背景。Oracle 数据库中，主要的数据类型可以分为 4 类：数值型(double，float，number)、字符型(char，varchar，nvarchar，raw，long)、日期型(date，timestamp)、“大对象”型(blob，clob，nclob)。因此，根据 Oracle 数据库对象中不同的数据类型，将多值背景 C 转换成单值背景 K_C 的过程中所要遵循的规则定义如下。

规则 1 对属性集合 M 中数据类型为日期型的多值属性列，按照年、月进行分类，再根据不同的应用目的，选取年或者月，或者全部作为形式背景中的一列属性。

规则 2 对属性集合 M 中数据类型为“大对象”型的多值属性列，由于存储的“大对象”型数据本质上是非结构化的，因此不能直接用 FCA 处理，直接去除这一列属性。

规则 3 对属性集合 M 中数据类型为字符型的多值属性列，直接拆分多值属性。设符合条件的集合为 $A(A\subseteq M)$，则 A 中所有元素 a 均有 w_a 与之相对应，其中 $w_a=(w\mid wRa,a\in A\wedge w\in W)$，$R$ 表示集合 M 与集合 W 之间的一个二元关系。那么多值属性拆分后，w_a 的值将取代 a 成为集合 A 中的新的属性元素。

规则 4 对属性集合 M 中数据类型为数值型的多值属性列，利用 K-means 算法对每列中的数据值按给定的 K 值进行聚类，根据具体的应用目的参考 K 值的定义，然后将聚类结

果作为新的单值属性列。

按照以上规则生成的单值形式背景 K_C，实际上也是一张稀疏大表。最后，可以根据属性值的空缺情况，对 K_C 进行属性约简，只留下能够完全确定形式背景上的概念及其层次结构的最小属性集，以便能更好地发现形式背景中隐含的信息。

3.3 概念格和本体的生成

概念格的最大优点就是，对于同一批数据，无论对象或者属性的排列顺序如何变化，都只能生成唯一的一个概念格。因此，它非常适合用来发现数据中的潜在关系。从形式背景生成概念格的算法大致可分为两类：批处理构造算法和渐进式构造算法[14,15]。

其中批处理算法根据其构造格的不同方式，还可以分为自顶向下、自底向上、枚举算法。比较著名的批处理算法有 NextConcept 算法、FastConcept 算法、Bordat 算法、Chein 算法等。渐进式造格算法的基本思想是将当前要插入的对象和格中所有的概念做交集运算，根据运算的结果采取不同的操作。比较著名的渐进式造格算法有 Godin 算法、Capineto 算法、T. B. Ho 算法等。建造完成的概念格还可以用图形化的 Hasse 图来表示，使概念格模型具有更直观的表达概念层次关系的可视化效果。

由于自底向下的概念格构造方法具有构造简单明了、易于生成 Hasse 图的特点，因此本文采用自底向上的构造方法构造概念格，算法如表 2 所列。

表 2 概念格构造算法

输入：C,P,N,分别表示形式背景表中对象集合、属性集合和单个对象具有的最大属性数目(概念格的有效层数)；C^P，P^P 分别表示 C 和 P 的幂集；Nodes：存储概念格节点集合。输出：概念格 CL

```
1. for( i=N;i≥1;i--){
2.   for( j=1;j < sizeOf(P^P);j ++ ){
3.     Att=P^P(j);//依次获取属性集合 Att
4.     Obj=f(Att);//获取共同拥有 Att 属性集合的类集合
5.     Att'=g(Obj);//获取 Obj 集合中共同拥有属性集合
6.     Obj'=f(Att');//获取共同 Att's 属性集合的类集合
7.     if((Att=Att') &(Obj=Obj')){
8.         Node(i,j). Attribute=Att;
9.         Node(i,j). Obj=Obj; //构造概念格 CL 中节点
10.          Nodes. add(Node);
11.          NodeNext=Nodes. find(i-1);//获取第 i-1 层的节点集合
12.          for( k=1;k < sizeOf(NodeNext);k ++ ){
13.              if NodeNext(k). Attbute. include(Att) NodeNext(k). parent=Node(i,j);
                 //根据属性集合之间的包含关系，确定上下层节点之间的父子关系
14.          }}}}
15. return Nodes;
```

根据生成的概念格 CL，得到其相应的图形化 Hasse 图，再对 Hasse 图进行循环修剪，去除不合理的概念。从可视化的 Hasse 图中，可以发现概念格中超概念是子概念的泛化，而子概念则是超概念的例化。概念格节点之间的层次关系对应于本体中父类和子类的关系，一个节点就可以定义为本体中的一个类。其次，一个概念格节点所包含的对象，则相当于本体中类的实例；而概念格节点中对象所共有的属性，则相当于本体中类的属性。因此，根据形式背景生成的概念格就可以看作是一个本体的雏形，但仍需根据领域知识做进一步扩充。具体的转换步骤如下：

i)首先移除 Hasse 图中概念节点的底下元素；

ii)为每个移走底下元素的概念节点添加子概念节点；

iii)把得到的概念层次模型直接映射成本体的层次模型，每一个概念节点映射成一个 Class，层次关系映射成 subClassof 关系，概念节点中的属性映射成 Class 的 DataProperty；

iv)把在数据预处理阶段集合 $T2$ 中的每张表都映射成一个本体中的 Class；

v)定义相应的谓词和规则来添加概念之间的非层次关系，如添加 ObjectProperty、属性的 Restriction 等的表示。这一步骤可以反复进行，并由领域专家协助，直至满意为止。

4 实例验证

为了验证所提出的面向关系数据库的基于 FCA 的本体学习方法的有效性，将此方法应用于从材料服役安全数据库中学习得到一个面向材料服役安全评价应用的领域本体的过程。

① 对材料服役数据库的数据表进行分析和预处理。

首先，根据数据表归类标准，选取出 $T1$ 集合中的表，由于篇幅原因，只选取了材料服役安全数据库中的部分表，则有 $T1$={Base_PipeInfor, Com_SoilCorrosion, Com_ProtCoating, Com_PipeCorrosion}，然后根据这些表之间的依赖关系，可以合并成一张新的关系表，如表 3 所列。

表 3 从集合 T(1)得到的一张新表(部分)

属性对象	类型	材料	土壤	电阻率	PH 值	防腐层	破损点	腐蚀类型	埋设时间
L001	中压	309	S01	112.050	8.2	P01	562	疲劳	2009
L002	低压	310	S02	30.578	7.1	P02	229	晶间	2008
L003	低压	314	S01	112.050	8.2	P03	582	应力	2009
L004	高压	314	S03	157.137	6.8	P04	230	疲劳	2008
L005	中压	310	S02	30.578	7.1	P03	650	应力	2009
L006	低压	309	S01	112.050	8.2	P06	680	晶间	2009

由于建表过程中无法自动排除无关属性，因此还需对所建立的新表手工进行一些修剪，比如去除土壤 ID 号和防腐层 ID 号之类的无意义属性列，则得到了一个基于材料服役安全数据库的多值形式背景。

② 把多值形式背景转换为单值形式背景。

先扫描多值形式背景中每一个属性列，然后按照 3.2 节中定义的规则处理每一个属性列，如拆分文本型属性列、聚类数值型属性列等，得到相应的单值形式背景如表 4 所列。

表 4 基于材料服役安全数据库的形式背景(部分)

对象	类型			材料			电阻率			PH 值			防腐层			腐蚀类型			年份	
	L	M	H	M1	M2	M3	S1	S2	S3	P1	P2	P3	C1	C2	C3	T1	T2	T3	Y1	Y2
1		×		×				×				×		×		×				×
2	×				×		×				×		×				×		×	
3	×					×		×				×		×				×		×
4			×			×			×	×			×				×			×	
5		×			×		×				×				×			×		×
6	×			×				×				×			×	×				×

L：=低压管线，M：=中压管线，H：=高压管线，M1：=309 钢，M2：=310 钢，M3：=314 钢，S1：=电阻率 0-50 的土壤，S2：= 电阻率 100-150 的土壤，S3：= 电阻率 150-200 的土壤，P1：=PH 值 6.1-7.0 的土壤，P2：= PH 值 7.1-8.0 的土壤，P3：= PH 值 8.1 以上的土壤，C1：=破损点为 500 个以下的防腐层，C2：= 破损点为 500-600 个的防腐层，C3：= 破损点为 600 个以上的防腐层，T1：=疲劳腐蚀，T2：=晶间腐蚀，T3：=应力腐蚀，Y1：=2008 年，Y2：=2009 年。

③ 生成形式背景对应的 Hasse 图和本体。

利用概念格构造算法从表 4 的形式背景中得到基于材料

服役安全数据库的概念格模型，并以 Hasse 图的形式可视化地展示，如图 3 所示。

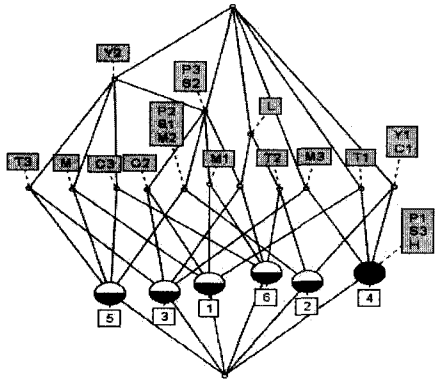

图 3　基于材料服役安全数据库的简化标号 Hasse 图

图 3 中的 Hasse 图是一个简化了标号的概念格视图，实际上每一个节点都包含了一个对象集合和一个属性集合。简化后的 Hasse 图的每个节点的对象集合则由这个节点下所有子节点中出现的对象标号构成，而每个节点的属性集合则由这个节点的所有父节点中出现的属性标号构成。例如，图 3 中的标识 $T3$，即表示了相应的概念（{5,3}，{$Y2$，$T3$}）；而图 3 中的标识 6，即表示相应的概念（{6}，{$Y2$，$C3$，$P3$，$S2$，$M1$，L，$T2$}）。最后，遵照 Hasse 图向本体转换的规则，再根据具体情况进行一些适当的修改，形成了图 4 所示的关于材料服役安全数据库的一个本体模型。此模型是独立于本体描述语言的，是对概念及关系的建模。当然，最终利用工具或者编程语言实现此本体模型时，还会反复地修正本体模型，以达到能真正应用的目的。

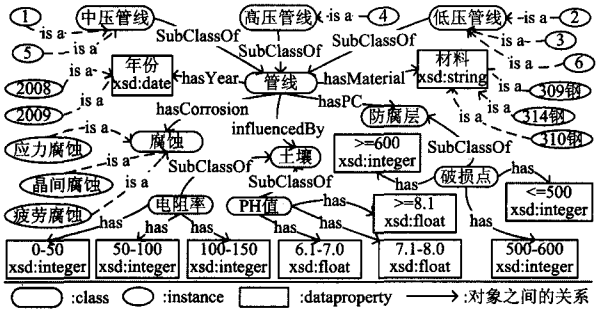

图 4　基于材料服役安全数据库的本体模型

结束语　通过定义 E-R 模型与本体之间的映射规则，从关系数据库中学习得到本体的过程过于依赖规则制定者的主观判断，使得本体不能在具体应用中很好地被利用。针对上述问题，本文提出了一种基于 FCA 的面向关系数据库的本体学习方法。经过关系数据表预处理、多值形式背景向单值形式背景转换、形式背景到 Hasse 图和生成本体模型 3 大步骤，实现了利用 FCA 方法自动地从关系数据库中学习得到语义信息，从而指导本体建立的功能。本方法不仅发挥了 FCA 的自动客观提取语义的特点，还在保持原有关系数据库表间语义的基础上增加了数据间的语义学习，使最终建立的本体具有更丰富的语义信息。通过阐述一个面向材料服役安全数据库进行本体学习的完整实例，说明了本方法的可行性，并为基于 FCA 构建其它领域本体提供了指导方案。下一步的工作是要进一步细化 Hasse 图向本体转换的规则，并完善初始本体建立后的公理约束及推理机制，以期能更好地服务于具体的领域应用。

参 考 文 献

[1] 杜小勇，李曼，王珊. 本体学习研究综述[J]. 软件学报，2006，17(9)：1837-1847

[2] Hu Chang-jun, Ouyang Chun-ping. NON-structured materials science data sharing based on semantic annotation [J]. Data Science Journal, 2009, 8: 52-61

[3] Maedche A, Staab S. Ontology Learning for the Semantic Web [J]. IEEE Intelligent Systems and Their Applications, 2005, 16(2): 72-79

[4] Kashyap V. Design and Creation of Ontologies for Environment Information Retrieval [C]// Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management (KAW'99). Banff, Alberta, Canada, 1999

[5] Rubin D L, Hewett M, Oliver D E, et al. Automatic Data Acquisiton into Ontologies from Pharmacogenetics Relational Data Source Using Declarative Object Definitions and XML [C]// Proceedings of the Pacific Symposium on Biology. Lihue, HI, 2002

[6] Stojanovie L, Stojanovie N, Volz R. Migrating data-intensive Web Sites into the Semantic Web [C]// Proceedings of the 17th ACM Symposium on Applied Computing (SAC2002). ACM Press, 2002: 1100-1107

[7] 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[J]. 计算机学报，2006，29(10)：1786-1796

[8] 王洪伟，伊磊，王洪滨. 面向关系模式的领域本体获取[J]. 计算机工程，2007(3)：1-3，23

[9] Wille R. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concept [C]// Proceedings of the 7th International Conference on Formal Concept Analysis (ICFCA 2009). Berlin: Springer-Verlag, 2009: 314-339

[10] Gater B, Wille R. Formal concept analysis: Mathematical foundations [M]. Berlin Heidelberg, 1999

[11] 周文，刘宗田，陈慧琼. FCA 与本体结合研究综述[J]. 计算机科学，2006，33(2)：8-12

[12] 周文，刘宗田. 基于形式概念分析的领域本体构建方法研究[J]. 计算机科学，2006，33(1)：210-212，239

[13] Cimiano P, Hotho A, Stumme G, et al. Conceptual Knowledge Processing with Formal Concept Analysis and Ontologies [C]// Proceedings of the Second International Conference on Formal Concept Analysis (ICFCA 2004). Berlin: Springer-Verlag, 2004: 189-207

[14] Kuznetsov S O, Obiedkov S A. Comparing Performance of Algorithms for Generating Concept Lattices [J]. Journal of Experimental & Theoretical Artificial Intelligence, 2002, 14(23): 189-216

[15] Baixeries J, Szathmary L, Valtchev P, et al. Yet a faster algorithm for building the Hasse diagram of a concept lattice [C]// Proceedings of the 7th International Conference on Formal Concept Analysis (ICFCA 2009). Berlin: Springer-Verlag, 2009: 162-177