

基于高阶逻辑的双基融合机制

刘英华^{1,2}, 李广原¹, 刘永彬¹, 郭建威¹

(1 北京科技大学 信息工程学院, 北京 100083; 2 中国青年政治学院, 北京 100089)

摘 要: 随着知识库规模的扩大及结构的日趋复杂, 一阶逻辑表达能力有限的问题暴露得越来越严重, 基于高阶逻辑的知识表示显得尤为必要. 原有基于 KDK 和 KDK* 的研究必然导致向高阶逻辑研究的拓展. 首先概述了基于知识库的知识发现, 然后介绍了高阶逻辑知识表示语言 Escher、双基融合机制 KDK 和 KDK*, 最后在高阶逻辑背景下, 扩展双基融合机制的内涵, 给出了面向高阶逻辑的 KDK* 过程模型逻辑等价定理.

关键词: 高阶逻辑; 双基融合机制; 复杂结构数据; 知识库中的知识发现

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2011)09-0217-03

Double-Bases Cooperating Mechanism Based on Higher-Order Logic

LIU Ying-hua^{1,2}, LI Guang-yuan¹, LIU Yong-bin¹, GUO Jian-wei¹

(1 School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China;

2 China Youth University for Political Science, Beijing 100089, China)

Abstract: With expanding of the knowledge base scale and increasingly complication of its structure, the problems of the limited capability of first-order logic expose to be more and more serious, thus the application that based on knowledge representation of higher-order logic is quite necessary. The original research based on KDK and KDK* will inevitably lead to development and research of higher first-order logic. This paper first summarizes the knowledge discovery that based on the knowledge base, then introduces the representation language of the high-order logic knowledge-Escher, KDK and KDK*. At last, expands the connotation of Double-Bases Cooperating Mechanism under the background of first-order logic, and illustrate the equivalence theorem of KDK* process model logic that is facing the high-order logic.

Key words: higher-order logic; double-bases cooperating mechanism; complex structured data; knowledge discovery in knowledgebase (KDK)

1 引言

知识发现(Knowledge Discovery in Databases, KDD)出现于1989年, U. M. Fayyad给出的定义是“从数据集中识别出有效的、新颖的、潜在有用的, 以及最终可理解的模式的非平凡过程”. 数据库中的知识发现是数据库和机器学习等人工智能技术的相结合的产物.

目前, KDD技术作为一门交叉学科, 涉及到了机器学习、数据库、神经网络、统计学、高性能计算、

专家系统等众多学科领域. 应用范围也逐步延伸到计算机生物学、医学、社会网络分析、普适计算、反恐等领域.

为了大型知识库的构建及进一步从知识库中发现出更多的知识, 提出了基于知识库的知识发现(Knowledge Discovery in Knowledgebase, KDK), 但国际国内对KDK的研究还很少. 目前, 该研究领域主要分为以下几个方面:

(1) 采用二元一阶归纳逻辑方法从知识库中发现新规则. (2) 杨炳儒教授在该领域的工作, 首次在

国际上明确提出 KDK 的研究方向. (3)涉及到如何利用机器学习技术从规则库中发现新知识. (4)IBM Watson 研究中心在专家系统的研究中,首先利用专家系统构建基本知识库,然后使用机器学习方法从该基本知识库中学习新的知识,并将学习到的新知识加入到基本知识库中.

2 基于高阶逻辑的知识表示和双基融合机制

C. Giraud-Carrier, J. W. Lloyd 等人提出了类型化的基于高阶逻辑的知识表示方式. 高阶逻辑具有更加丰富的表达力,在构造中允许下层类型的建立,其中最经典的知识表示系统是 Escher 语言.

双基融合机制是基于数据库与知识库的知识发现系统,是基于知识发现系统内在机理的研究. 它将 KDD 作为一个智能体,研究其内在机理.

KDK * (基于双基融合机制的 KDK)最主要的优越性是减少了用户和专家的参与,在对假设规则的自主性评价过程中,在用户和专家参与之前,假设规则已经完成了数据库评估和 KDD * (双库协同机制)过程的评估,将意义不充分的规则删除. KDK * 结构示意图见图 1.

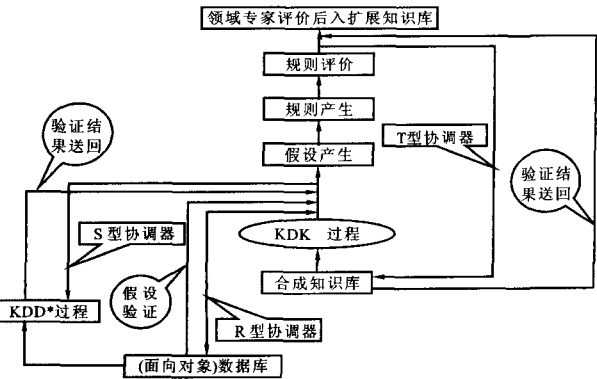


图 1 (高阶逻辑)KDK * 结构示意图

3 面向高阶逻辑的 KDK * 过程模型逻辑等价定理

3.1 知识结点与数据子类结构的泛同伦对应

在一阶逻辑中将知识结点归结为知识素结点,通过在知识库中建立知识素结点和知识合结点的关系,验证了一阶逻辑中知识结点与数据子类结构的泛同伦对应.

Escher 语言是面向高阶逻辑的语言. 这种语音的每个样例个体使用一个封闭的“项”来描述,即“个体一项”的表示方式. Escher 语言的基础是类型,包

含基本类型和非基本类型. 非基本类型是按照 Escher 语言的数据构造器依据基本类型创建的. 例如:整数、字符、字符串等均是基本类型,集合、多集、链表、树、图等均在基本类型的基础上通过数据构造器创建. Escher 语言使用“项”表示某种类型的实例,例如典型的“分子-原子”数据集表示如下:

```
Element=String;
Atom_Type=Int;
Atom=Element×Atom_Type;
No=Int;
Name=String;
Atoms={Atom};
Molecule= No×Name×Atoms;
```

Escher 语言与面向对象数据库相似,其中 Molecule 和 Atom 均是面向对象数据库中的类,Escher 语言构造器“×”表明类 Molecule 和 Atom 的组合关系. 在面向对象数据库模型中可将上述数据集表示为如下两个类:

```
class Molecule{
    int No;
    String Name;
    Set Atoms;
}
class Atom{
    String Element;
    int Atom_Type;
}
```

其中,Escher 语言与面向对象数据库存在这对应关系,Escher 中的“类型”即对应面向对象数据库中的“类”和“属性”,“函数”对应“方法”. “个人”对应“对象”,“数据构造器”对应“继承”和“组合”等.

综上所述并结合泛同伦空间概念得出以下结论:

定理 1 高阶逻辑知识表示语言 Escher 表示的知识结点与面向对象数据库是同一泛同伦空间.

3.2 扩展后的 KDK * 过程模型

如果知识库中的知识用 Escher 语言表示,那么发现新知识就是使用数据构造器构造更复杂的类型,根据定理 1 可以扩展 KDK 过程模型如下:

定义 1 定义高阶逻辑知识表示下的 KDK 过程模型为 $M=\langle Q, R, f, g \rangle$,其中:

- (1)Q 是高阶逻辑知识表示下的知识结点集,即可能世界集;
- (2) $R \subseteq Q \times Q$,可以理解为 Q 的组合、继承等关系;
- (3)f 是知识库中的正则测度函数;

(4) g 是 R 上的正则确证函数.

说明: Q 是知识结点集,则 $Q \times Q$ 是知识结点集的笛卡儿乘积. R 是通过两个或以上的知识点得到的新知识点,实际在高阶逻辑知识表示下的知识点不是通过归纳而得出新知识点,是通过 Escher 语言的组合、继承等关系生成的新知识点.

定义 2 定义面向对象数据库中 KDD 的过程模型为 $N = \langle S, F, \text{sup}, \text{rel} \rangle$, 其中:

(1) S 是面向对象数据库中的数据子类集;(2) F 是 S 上挖掘可达关系,可以理解为组合、继承等关系;(3) sup 是数据子类的支持度;(4) rel 是 F 上的挖掘可信度.

显然基于高阶逻辑知识表示下的扩展 KDK 过程中 R 与 F 分别是 Q 与 S 的组合、继承等关系. Q 与 S 是一一对应的,则 R 与 F 也是一一对应的.

定理 2 高阶逻辑知识表示下知识结点集 Q 与面向对象数据库中数据子类集 S 是一一对应的.

扩展后的 KDK * 过程模型见图 1,其中的数据库转换为面向对象数据库.

4 结束语

借鉴面向事实和面向规则的双基融合机制思想,本文对原一阶逻辑与命题逻辑(属性—值)知识表示方式下的双基融合机制的过程模型逻辑等价定理进行了扩充,解决了双基融合机制从一阶逻辑知识表示方式向高阶逻辑知识表示方式的泛化更新,从而提出了高阶逻辑下的双基融合机制,并且在高阶逻辑下对过程模型逻辑等价定理进行了证明,但是还有些问题有待解决:(1)对已有的一阶逻辑下的 R 型协调器向高阶逻辑的扩展,以验证发现的新知识在数据库中的合理性;(2)对已有的一阶逻辑下的 S 型协调器向高阶逻辑的扩展,以验证发现的新知识在现实中的合理性;(3)对已有的一阶逻辑下的 T 型协调器向高阶逻辑的扩展,以验证发现的新知识与原有知识的相容性.

参考文献:

- [1] Bowers A F, Giraud—Carrier C, Kennedy C, et al. A framework for higher—order inductive machine learning [R]. Bristol, U K:University of Bristol, 1997.
- [2] 杨炳儒. 基于内在认知机理的知识发现理论[M]. 北京:国防工业出版社,2009.
- [3] Wei Minshen. Discovering regularities from knowledge base[J]. International Journal of Intelligent Systems, 1992,24(12):55—62.
- [4] Breidenstein T, Bournaud I, Wolinski F. Knowledge discovery in rule bases[C]// Proc of the 10th European Workshop on Knowledge Acquisition, Modeling and Management. London: Lecture Notes In Computer Science,1997: 329—334.
- [5] Weiss S M, Buckley S J, Kapoor S, et al. Knowledge—based data mining[C]// Proc of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). USA: Philadelphia, 2003: 456—461.
- [6] 杨炳儒,宋威,徐章艳.基于知识发现创新技术的专家系统新构造[J]. 中国科学 E,2011,7(6):738—747.
- [7] Yang Bingru, Shen Jiangtao, Song Wei. KDK based double—basis fusion mechanism and its structural model [J]. International Journal of Artificial Intelligence Tools (IJAIT), 2011,14(3): 399—423.
- [8] Bingru Yang, Wei Song, Linna Li, et al. Research overview of regulations in the Process of Dynamic Mining [J]. Journal of Computational Information Systems, 2006,2(3):973—979.
- [9] Bingru Yang, Jing Gao, Wei Song. Cognitive physics techniques in data mining[J]. Journal of Computational Information Systems,2007,3(5):1915—1928.
- [10] Yang Bingru, Song Wei, Xu Zhangyan. New construction for expert system based on innovative knowledge discovery technology[J]. Science in China Series, 2007, 50(1):29—40.
- [11] Bingru Yang, Wei Hou. KAAPRO:an approach of protein secondary structure prediction based on KDD * in the compound pyramid prediction model [J]. Expert Systems With Applications,1992,36(1):9000—9006.

作者简介:

刘英华 女,(1975—),博士研究生,讲师.研究方向为数据挖掘、知识发现、隐私保护.

[1] Bowers A F, Giraud—Carrier C, Kennedy C, et al. A framework for higher—order inductive machine learning