

上海大学出版社

2006年上海大学博士学位论文 57



概念格同构生成方法研究及 IsoFCA 系统实现

- 作者：沈夏炯
- 专业：控制理论与控制工程
- 导师：刘宗田



上海大学出版社

2006年上海大学博士学位论文 57



概念格同构生成方法研究及 IsoFCA 系统实现

- 作者：沈夏炯
- 专业：控制理论与控制工程
- 导师：刘宗田



图书在版编目(CIP)数据

2006 年上海大学博士学位论文. 第 2 辑/博士学位论文编辑部编. —上海: 上海大学出版社, 2010. 6

ISBN 978 - 7 - 81118 - 513 - 3

I. 2... II. 博... III. 博士—学位论文—汇编—上海市—2006 IV. G643.8

中国版本图书馆 CIP 数据核字(2009)第 162510 号

2006 年上海大学博士学位论文
——第 2 辑

上海大学出版社出版发行

(上海市上大路 99 号 邮政编码 200444)

(<http://www.shangdapress.com> 发行热线 66135110)

出版人: 姚铁军

*

南京展望文化发展有限公司排版

上海华业装潢印刷厂印刷 各地新华书店经销

开本 890 × 1240 1/32 印张 278 字数 7 760 千

2010 年 6 月第 1 版 2010 年 6 月第 1 次印刷

印数: 1—400

ISBN 978 - 7 - 81118 - 513 - 3/G · 514 定价: 880.00 元(44 册)

Shanghai University Doctoral Dissertation (2006)

Research on Isomorphism Generating Method of Concept Lattices and Implementation of IsoFCA System

Candidate: Shen Xiajiong

Major: Control Theory and Control Engineering

Supervisor: Prof. Liu Zongtian

Shanghai University Press

• Shanghai •

答辩委员会评语

本论文经答辩委员会全体委员审查,确认符合上海大学博士学位论文质量要求。

答辩委员会名单:

主任:周傲英 教授,复旦大学计算机科学与工程系 200433

委员:黄国兴 教授,华东师范大学软件学院 200062

胡运发 教授,复旦大学计算机与信息技术系 200433

苗夺谦 教授,同济大学计算机科学与技术系 200331

李青 教授,上海大学计算机工程与科学学院 200072

导师:刘宗田 教授,上海大学 200072

答辩日期:2006年4月24日

该博士论文选题新颖,逻辑严谨,方法合理,实验充分,结论正确。论文结构清晰,层次分明。论文工作表明作者掌握了相关领域宽广、扎实的基础理论和系统深入的专业知识,独立从事科学研究工作能力较强。

沈夏炯同学在答辩中论述清晰,回答问题正确。经答辩委员会表决一致同意通过其博士学位论文答辩,并建议授予工学博士学位。

学大南工

合符人前，查审员委村全会员委籍答登文尔本

。宋要量闻文尔尔学士耐学大南工

评阅人名单：

周傲英 教授，复旦大学计算机科学与工程系

黄国兴 教授，华东师范大学软件学院

汤庸 教授，中山大学信息科学与技术学院

日 13 月 4 年 2006 : 期日 稿答

答辩委员会对论文的评语

形式概念分析是知识表示和知识处理的数学基础,已有广泛深入的研究。该博士论文针对概念格构造的关键问题和知识相似性的表示问题,研究其形式概念分析方法,选题具有理论意义和实用价值。论文的主要贡献如下:

(1) 提出了一种基于形式背景核的概念格同构生成方法,与传统的批处理和渐进式算法相比,利用背景核,该方法不仅能提高概念格的构造效率,而且可以建立一种知识相似性表示模型。

(2) 提出了 n 阶形式背景核的概念及其构造算法。论文在 n 阶形式背景核的构造算法研究中,利用限定性定理,提高了构造形式背景核的效率。

(3) 设计了形式背景同构判定的等价类算法。该算法结合了不变量和测试两种方法的特点,可显著提高形式背景同构判定的效率。

(4) 开发了一个有同构生成特色的软件系统 IsoFCA,实现了论文中的全部算法和概念格三维可视化互操作模型。

该博士论文选题新颖,逻辑严谨,方法合理,实验充分,结论正确。论文结构清晰,层次分明。论文工作表明作者掌握了相关领域宽广、扎实的基础理论和系统深入的专业知识,独立从事科学研究工作能力较强。

沈夏炯同学在答辩中论述清晰,回答问题正确。经答辩委员会表决一致同意通过其博士学位论文答辩,并建议授予工学博士学位。

答辩委员会表决结果

五位评委全票同意通过论文答辩,建议授予博士学位。

答辩委员会主任:周傲英

2006年4月26日

摘 要

本文针对形式概念分析(Formal Concept Analysis, FCA)理论中概念格的构造方法进行了深入研究,并对应用形式背景核建立知识相似性表示模型做了深入探讨。FCA 是应用数学的一个领域,其主要内容是研究“概念”和“概念分层”的数学化描述,从而引发了数学思想在概念数据分析和知识处理方面的研究和应用。FCA 的主要思想是从表示为形式背景(formal context)的数据中获取形式概念(formal concept)以及形式概念之间的联系,形成一种以形式概念为元素的格结构——概念格(concept lattice)。概念格是知识系统的良好抽象,其意义主要有两方面:① 从数据中获取概念。与其他基于统计的传统数据分析方法不同,FCA 用概念表示数据分析的结果,概念格是数据的高级表示形式——知识视图。② 概念格为知识处理提供了很好的基础,比如知识表示、关联规则发现、智能搜索引擎、语义 Web 和自然语言理解等。

FCA 的一个重要问题是构造概念格。概念格的构造算法分为批处理算法和渐进式算法两大类,目前已经有许多文献致力于概念格构造算法研究和设计。然而,这些研究都集中在如何从形式背景直接构造概念格以及概念格的代数性质本身。而如何利用形式背景之间的关系生成概念格方面,除了少数文献给出简单的原理性叙述之外,尚未见有深入的研究成果发表。

本文在知识相似性的表示问题的研究中,结合形式背景同

构和概念格同构等基本原理,实现了以下创新点:

(1) 提出基于形式背景核的概念格同构生成方法。核思想是:对于任意的形式背景 K , 都可将其分解为阶数较小子背景。对每一个子背景 K_i 的简化形式,在同阶形式背景核中必存在与之同构的元素 H_j , 然后根据 H_j 的概念格导出 K_i 的全部概念,从而得到的 K_i 概念格 $\mathfrak{B}(K_i)$ 。最后合并所有的 $\mathfrak{B}(K_i)$ 得到 $\mathfrak{B}(K)$ 。

(2) 提出形式背景同构判定算法——等价类算法。形式背景同构判定是判定一个形式背景经过行、列交换能否变换成另一个形式背景,旨在寻找形式背景变换的快速算法,本文提出的等价类法能够较好的解决这个问题。

(3) 提出 n 阶形式背景核的概念及其构造算法。 n 阶形式背景核是具有 n 个属性的形式背景的集合,满足:每个元素都是简化的且任两个元素互不同构。在 n 阶形式背景核的构造算法研究中,提出限定性定理,避免对 $\mathbb{B}(n)$ 中数量接近一半而规模相对较大的背景进行穷举,从而提高构造形式背景核的效率和阶数。形式背景核也是利用有限知识表示新知识的基本模型。

(4) 提出并实现了概念格三维可视化互操作模型。根据概念格的空间伸展特性,此项研究完成了概念格在三维空间的自动布局、动态显示和互操作用户界面。

(5) 设计并实现了一个基于概念格同构生成的软件工具 IsoFCA(Isomorphism FCA)。基于创新点(1)的思想,在系统的开发中,具体实现了 n 阶形式背景核的构造、形式背景的自动分解、形式背景的同构判定、概念格的同构生成和格的并行合

并等。

概念格的同构生成具有多方面的理论意义和实用价值。首先,为概念格的获取提供了一条新的途径。其次,建立一种知识处理的分布式语义模型。第三,支持学习过程中的知识相似性的形式表示,即用一个知识系统相似地表示另一个知识系统。这一点在 IsoFCA 中得到了实践验证。

关键词 形式背景,概念格,同构,形式背景核,重构,知识相似性

Abstract

Our researches aim at construction method of concept lattices in Formal Concept Analysis (FCA), and the presentation model of knowledge similarity based on formal context kernels. FCA is a field of applied mathematics based on the mathematization of concept and conceptual hierarchy. Thereby it activates mathematical thinking for conceptual data analysis and knowledge processing. The major content in FCA is to extract formal concepts and connections between them from data in form of formal context so as to form a lattice structure of formal concepts. Concept lattices have been regarded as perfect abstraction of knowledge system. There are two significant aspects. First, the results of data analysis are concepts instead of data themselves. Differing to traditional data analysis methods based on statistics, FCA obtains knowledge view which is higher level of representation of data. Second, concept lattices provide better foundation for knowledge processing, such as association rule discovery, intelligent search engine, semantic Web and natural language processing.

One of important issues in FCA is construction of concept lattices. So far, there are two kinds of algorithms for concept lattices constructing, batch algorithms and incremental

algorithms. However, most of the researches dedicated to build concept lattices directly from contexts. Rare research papers were published in addition to [5] which give out some brief descriptions on exploiting the relations between contexts to generate concept lattices.

In this dissertation, five emphases are placed on the development of the methods including,

1. A novel method of isomorphic generating of concept lattices based on n th-order context kernels, of which the central idea is that an arbitrary context K can be split onto subcontexts in smaller size of order, then, for every subcontext K_i , there exists a isomorphic context H_j in the kernel with the same order as K_i , and concept lattice $\mathfrak{B}(K_i)$ can be yielded in terms of $\mathfrak{B}(H_j)$, and finally, the result $\mathfrak{B}(K)$ can be reconstructed from $\mathfrak{B}(K_i)$.

2. Equivalent class algorithm, a rapid algorithm for context isomorphism detecting, which is to determine whether a context can be transformed into another one by exchanging rows or columns, and focuses on finding out rapid algorithm of context transform.

3. Notion of n^{th} -order context kernel and an algorithm for kernel calculating. An n^{th} -order context kernel is a set of contexts with n attributes, and all elements in it are reducible and are not isomorphic to each other as well. The kernel is used to ensure that every subcontext of a context is isomorphic to one in the kernel so that the lattices of subcontext can be generated from corresponding lattices of kernel elements.

The theorem proposed in Chapter 4 guarantees that almost half of amount of large contexts in $\mathbb{B}(n)$ need not be scanned in the process for kernel calculating. Context kernel can be regarded as a foundation model to expressing new knowledge by using limited meta knowledge.

4. An interactive model based on concept lattices visualizing in three-dimensional space. The concept lattices visualizing includes not only the development of three-dimensional lattices display, which allows users to rotate, fold/unfold lattice, to and detail information checking, but also interactive user interfaces.

5. A software toolkit system, IsoFCA (Isomorphism FCA), based on concept lattices isomorphic generating. Functions implemented include nth-order context kernel construction, context reduction, context decomposition, context isomorphism determination, Godin algorithm improvement, concept lattice isomorphic generation and lattice parallel reconstruction.

Isomorphic generating of concept lattices is a significant method to yield lattices with great practical value. First of all, the method provides a different way to obtain concept lattices from conventional algorithms. Second, the context kernel is a distributed semantic model in knowledge processing. Last, knowledge similarity can be described with the model in knowledge learning, and a knowledge system can be much more easily expressed by the similarity of another knowledge system. The procedure is performed in

IsoFCA. The theorem proposed in Chapter 4 guarantees that half of amount of large contexts in $E(v)$ need not be scanned.

Key words formal context, concept lattices, isomorphism, context kernel, reconstruction, knowledge similarity

by using limited meta knowledge.

An interactive model based on concept lattices in this dissertation, will improve the visualization in three-dimensional space. The concept lattices visualizing includes not only the development of three-dimensional lattices display which allows users to rotate, fold/unfold lattice to and detail information checking, but also interactive user interfaces.

5. A software toolset system, IsoFCA (Isomorphism FCA), based on concept lattices isomorphic generating functions implemented include nth order context kernel construction, context reduction, context decomposition, context isomorphism determination, Godin algorithm improvement, concept lattice isomorphic generation and lattice parallel reconstruction.

Isomorphic generating of concept lattices is a significant method to yield lattices with great practical value. First of all, the method provides a different way to obtain concept lattices from conventional algorithms. Second, the context kernel is a distributed semantic model in knowledge processing. Last, knowledge similarity can be described with the model in knowledge learning and a knowledge system can be much more easily expressed by the similarity of another knowledge system. The procedure is performed in generated from most

目 录

第一章 绪论	1
1.1 知识的相似性	1
1.2 用概念格的同构表示知识的相似性	3
1.3 本文研究的内容、意义和创新点	5
1.4 本文的结构	8
第二章 形式概念分析基础	9
2.1 序论中的基本定义	9
2.2 格论中的基本定义	10
2.3 形式概念分析的理论基础	11
第三章 形式背景同构判定	16
3.1 引言	16
3.2 背景以及同构的基本概念	18
3.3 背景同构判定算法: 等价类法	19
3.4 算法分析	28
3.5 小结	30
第四章 n 阶形式背景核	31
4.1 形式背景清晰化	32
4.2 形式背景约简	32
4.3 n 阶背景基与 n 阶背景核	36
4.4 小结	42

第五章 Godin 算法的改进	44
5.1 概念格构造算法概述	44
5.2 批量算法(batch algorithm)	44
5.3 渐进式构造算法	48
5.4 Godin 算法的改进	52
5.5 小结	59
第六章 概念格同构生成	60
6.1 基于格同构的分布式知识处理	60
6.2 概念格同构生成算法的设计与实现	61
6.3 小结	65
第七章 概念格重构	66
7.1 引言	66
7.2 概念格纵向合并的基本思想及算法	67
7.3 基于同义概念的概念格纵向合并算法	68
7.4 实验及结果分析	74
7.5 小结	76
第八章 概念格的互操作及三维可视化	78
8.1 引言	78
8.2 概念格的可视化形式	79
8.3 基于三维空间的概念格自动布局	82
8.4 小结	86
第九章 IsoFCA 系统的设计与实现	87
9.1 国内外对 FCA 软件工具的研究与开发	87
9.2 系统最终方案的形式描述	89
9.3 系统设计和实现	91