

《智能信息处理》课程作业

## 形式概念分析理论概述

MAKHMADRASULI ERGASHI (米乐)

作业	分数
得分	

2020 年 12 月 3 日

## 形式概念分析理论概述

信息技术得迅速发展，在商业管理、政府部门、科学研究和工业数据处理等领域都收集并且存储了大量的数据，信息爆炸或者信息泛滥使当今数字化社会面临着巨大的挑战，有人说我们是数据的富翁，知识的穷人。面对如此的尴尬的局面，迫使人类不得不思考如何才能解决这个问题。形式概念分析是如何从大量的数据信息中自动地抽取出有价值的知识和信息，因而成为人工智能中非常活跃的研究领域。

### 形式概念分析产生的背景

人类在认知过程中，把所有感觉的事物的共同特点抽出来，加以概括，就称为概念。在哲学中，概念被定义为思维的基本形式之一，它反映了可观事物的一般的、本质的特性。概念被理解为由外延和内涵两个部分所组成的思想单元。对概念的这种理解源于古希腊哲学，

发展于十七世纪的近代学院派，进一步发展成德国标准，最终成为了世界标准。由于概念是思维的基本单元，因此基于对概念的研究成为人工智能的一个重要组成部分。在知识表示、知识管理、机器学习、专家系统等不同的领域，研究者们从不同的角度和观点来分析概念，形成了对概念的不同形式化描述方法。

形式概念分析是一种对数据进行分析的工具或者方法，特别是对给定的信息进行分析和处理。而数据应该是从人类有意义的 可以理解的思维单位——概念中抽取而形成的形式化的单元。形式化表明的是所有处理的数据是形式化的数学实体，不需要和人类思维中的概念完全相同，同时它也指出形式概念分析处理的基本数据形式是形式背景，形式背景是人类背景知识中的一小部分。

形式概念分析是由德国的

Darmstadt 技术大学数学系于 70 年代未发展起来的。当时 Darmstadt 的研究小组致力于系统研究并发展一种基于 Birkhoff 格理论的应用软件。自 1981 年在关于有序集合的 Banff 会议的专题演讲上对形式概念分析做了首次描述以后, 数以百计的相关论文开始出版和发表, 甚至包括一篇格理论的数学基础的书。Darmstadt 研究小组此后一直从事多达上百个基于格理论的应用软件研究, 因此研究先前的部分成员已经建立了一个很小的原型系统, 并且进一步使其逐步实用化。

Darmstadt 小组的研究显然在形式概念分析的诞生和应用过程中起到了重大的作用。他们首先把多种可能性系统的详细阐述为一种数据分析方法, 并且进行试验, 进一步开发许多应用。这个工作成功的决定性因素存在于背景的形式化和把“概念”作为外延和内涵实体的解释。在这里“概念”已经被

形式化, 而且对它的理解深深植根于哲学, 这个在文献中有详细的描述。这种思想的传统甚至在标准 DIN 2330 和 DIN 2331 (DIN 表示“Deutsche Industrienorm”, 对于由 German National Bureau of Standard 发布的标准是很典型的) 中发现了表述, Darmstadt 小组在开发初期, 多次讨论了这两个标准。

## 形式概念分析的现状

Birkhoff 的格理论是形式概念分析的重要理论, 1940 年出版的 Birkhoff 关于格理论的作者 (Lattice Theory) 第一版中明确提到了关联关系 (比如对象-属性-关系) 的解释, 并且解释和说明了如何从一个二维表中来构造完全的格结构。为了新的应用目和发展的需要, 形式概念分析在 Birkhoff 的格理论上做了进一步的扩展和更深入的研究。

基于 Birkhoff 对格理论的贡献, 德国的 Wille 教授在 1982 年作为一种数学理论首先引入了概念格

( Concept Lattice ) , 并且 在 Restructuring Lattice Theory:An Approach Based on Hierarchies of Concepts 中对形式概念分析做了全面阐述。这部出版物奠定了形式概念分析的理论基础, 将哲学的概念进行数学化的描述, 实现了概念的一种形式化描述方法。概念格理论是形式概念分析理论的核心数据结构, 被认为是知识发现和数据分析的有力数学工具。研究概念的形式化描述和定义、定理、特征等是形式概念分析的理论基础。目前仍然有研究者从事这方面的工作。

若说 Wille 教授在 1982 年发表的 Restructing Lattice Theory:An Approach Based on Hierarchies of Concepts, 标志着概念格理论的建立, 那么 Formal Concept Analysis 使概念格理论进一步成熟, 其中讨论了它的完备性等若干性质。随后又有学者研究了它的同构性; 近几年又有人提出领域、闭包、极大和随

机概念格等概念。总的来说, 关于概念格的建格算法做的研究比较多。

建格的过程实际上是概念聚类的过程, 对同一批数据, 所有生成的格也是唯一的。概念格建格算法可以分为两类: 批处理算法和增量算法。批处理算法的思想是首先生成所有概念, 然后根据它们之间的直接前驱一后续关系, 生成边, 完成概念格的构造。根据其构造格的不同方式, 批处理算法又可以分为三类: 从顶向下算法, 自底而上算法, 枚举算法。从顶向下算法首先构造格的最上层节点, 再逐渐向下, 比如 Bordat 算法; Osham 算法; 自底而上算法首先构造底部的节点, 再向上扩展, 比如 Chein 的算法. 枚举算法则是按照一定顺序枚举格的所有节点, 然后再生成格的 Hasse 图, 比如 Ganter, Nourine 算法。增量算法是把当前要插入的对象和概念格中所有的概念相交, 由

交的结果吧格中节点分为不变节点、更新节点和新增节点，根据交的结果然后采取不同的方法，比如 Godin 的算法、Capinet 算法，T.B.Ho 算法，渐进式生成算法。其中概念格的维护算法（主要涉及到对象的插入、删除、修改和属性的增加）也属于概念格的增量算法。

在概念格发展的基础上，Claudio Carpineto、胡学钢等人对概念格进行了扩展，通过引入等价内涵定义了扩展概念格（Extended Concept Lattices），并且证明扩展概念格和其对应的一般概念格是同构的；后面又定义了约简概念格（Reduced Concept Lattice）简化了扩展概念格，使其更利于大规模数据库知识发现的应用；针对约简格中的生成内涵可以从其父概念内涵中获得，对约简概念格进一步简化又提出了相对约简概念格（Relative Reduced Concept Lattice）；在概念格的内涵中引入等

价关系并且将其外延量化，得到 T 量化概念格(Quantitative Concept)，利用量化概念格可以更清晰的表示知识，从而便于挖掘包括关联规则在内的多种规则，并且用用户可以根据自己的兴趣交互的挖掘关联规则，不需要计算频繁项目集，因而提高了挖掘规则的效率。在提出这些新概念格的基础上对这类概念格的性质、构造和其维护做了一定的研究。另外结合 Zadeh 的模糊集理论（Fuzzy Sets Theory），A.Burusco 和 R.Fuentes 于 1994 年提出了一模糊概念格，讨论了它的层次结构，证明全体一模糊概念是一个完备格，并且对其算法做了一定的研究；2002 年，陈世权和程里春又提出了描述概念内涵和外延模糊关系的模糊概念格，建立了概念的内涵与外延的模糊映射，用模糊集描述模糊概念思维，并进一步探讨了模糊属性映射和模糊对象映射的隶属函数的一些基本数学

性质，证明全体模糊概念构成一个完备格，相关的研究工作对模糊概念刻画、模糊推理、模糊形式语言具有重要的理论和应用价值。

以上面的都是针对单个概念格进行的，随着网络技术尤其是互联网的飞速发展，数据的分布式存储和并行处理的需求越来越迫切，对多概念格的合并算法也有了一些研究。

概念格的研究并不仅仅限于理论方面，在应用上也有一定的发展，比如在软件工程、信息检索、知识发现、数据挖掘、规则提取、聚类分析等领域。在知识发现领域，概念格可以从关系数据中提取各种类型的知识，比如关联规则、蕴涵规则、分类规则等等；在软件工程领域，概念格可以从类库以及类库的规范说明上构造，从而对类库的可视化以及类库的重构和优化提

供支持；在知识工程领域，概念格可以用于知识库的重新结构化；在信息检索方面，概念格可以实现对信息的有机组织并过滤掉无用的信息，比如和粗糙集理论相结合对信息系统进行属性约简，来简化信息表。目前也有人在应用概念格进行搜索引擎方面的研究。而且有人指出概念格将会在生物和生命科学领域有重大应用。

## 结论

形式概念分析理论是用来进行数据分析和处理的一种数学工具，它一直是全世界学者研究的重点。经过多年的发展，已经广泛应用于许多方面，其中包括知识发现、函数依赖、信息检索、数据库管理、软件工程、数理逻辑等多方面。因此，对于概念格的理论分析和应用问题的研究都有非常重大的意义。

## 参考文献

- [1] International Organization of Standardization: ISO 704. Terminology Work-Principles and Methods, 2000.
- [2] Rudolf Wille. Begriffsdenden: Von der griechischen Philosophie bis zur Künstlichen Intelligenz heute. Diltthey-Kastanie, Ludwig-Georgs-Gymnasium Darmstadt, 1995.
- [3] Garrett Birkhoff. Lattice Theory, first edition. Amer. Math. Soc. Coll. Publ. 25, Providence, R.I 1940.
- [4] Wille R. Restructuring lattice theory: An approach based on hierarchies of concepts. In Ordered Sets, I. Rival, Ed. Reidel, 1982.
- [5] B. Ganter, R. Wille. Formal Concept Analysis. Springer, 1999.
- [6] 简宋全, 胡学钢, 蒋美华. 扩展概念格的渐进式构造. 计算机工程与应用, 2001.
- [7] 吴刚、简宋全、胡学钢等. 扩展概念格的维护. 计算机工程与应用, 2002.
- [8] Zadeh. Fuzzy Sets. Information and Control, 1965.
- [9] 王德兴、胡学钢、王洁. 基于量化概念格的关联规则挖掘, 合肥工业大学学报, 2002, 25 (5) : 678-682.
- [10] 沈夏炯, 韩道军, 刘宗田. 概念格构造算法的改进. 计算机工程与应用, 2004.
- [11] 谢志棚, 刘宗田. 概念格的快速渐进式构造算法. 计算机学报, 2002, 25 (5) 490-496.