《智能信息处理》课程作业

形式概念分析的发展与应用

曾 敏

作业	分数[20]
得分	

2016 年 11 月 10 日

形式概念分析的发展与应用

曾 敏

（大连海事大学信息科学与技术学院 大连 116026）

摘 要 作为一种数据挖掘工具，形式概念分析得到国内外的关注的同时，也得到了越来越广泛的应用，尤其是在软件工程中，形式概念分析的应用愈加成熟。本论文主要介绍了形式概念分析的概念、属性，并且对形式概念在国内外的发展形势以及各个领域的应用进行了一定的分析与说明。

关键词 形式概念分析，概念格，语境，对象集，属性集，软件工程

Developments and Applications of Formal Concept Analysis

Zeng Min

（Dalian Maritime University，Computer Science and Technology，Liaoning，Dalian，116026，China）

Abstract As a kind of data mining tools,Formal Concept Analysis has gained so much attentions.Also,it has been used more and more widely,especially in software engineering.The developments of Formal Concept Analysis is becoming more mature.The paper is mainly about the introductions of the concepts and attritubes of the Formal Concept Analysis, as well as the analysis of the developments and applications of the Formal Concept Analysis.

Key words Formal Concept Analysis, concept lattice, syntactics, object set, attribute set, software engineering

1 形式概念分析

1.1 概念

德国学者 Wille 于 1982 年首次在全球提出了形式概念分析这一概念，它是应用数学知识与格论的一种综合，是建立在概念以及概念层次的数学化的基础之上的。Wille 的著作中有这样一段话：“The aim and meaning of FCA is to support the rational communication of humans by mathematically developing appropriate conceptual structures which can be logically activated.” [1] 简单的翻译一下，就是：“形式概念分析的目标与意义就是，要用数学的理论创建一种逻辑上可行的概念结构，以支持人类的行为情绪交流的表达”。研究的所有对象都具有相同的特征与属性是便概念的深刻内涵，从而实现概念的形式化。在形式概念分析中其核心数据结构便是概念格。

1.2 背景

随着网络技术的发展，使得各个领域的信息量在飞速增加。这无疑给信息检索与操作带来了十分大的困难。而为了能够在这些杂乱无章的数据中查找到有用的有效信息，数据挖掘(DM) 和数据库知识发现(KDD)逐渐发展成型并得到了越来越广泛的应用。

其实，最早，是由德国 Darmstadt 研究小组便开始了对于一种以格论为基础的应用软件的研究。并且研究出了一个简单的原型系统。其中，在 Birkhoff 的格论中，对于从一个二维表中构造格结构有了一定的解释说明，而形式概念分析正是在这种情况下诞生的，并且，在 1981 年关于有序集合的 Banff 会议中被首次描述。[2-6]

1982 年，Wille 教授将形式概念分析作为一种数学理论提出来，并且获得了人们的关注，逐渐在之后的许多领域中得到了实践与应用。

1.3 基本概念

形式概念分析由几个基本的概念组成，而概念格则是形式概念分析中最基础也是最重要的概念之一。

1.3.1 语境和概念

定义一 语境（context）：一个形式化的语境 k=(G ， M ， I),包含两个集合(G 和 M)和一个二元关系(G 和 M 之间的关系 I)。在语境中，G 中的元素称为对象，M 中的元素称为属性。用 gIm ，或者 (g ,m)∈ I 来表达对象 g 和属性 m 的关系，读作“对象 g 具有属性 m” 。

根据定义一 ，可以用矩阵来表示语境。每行的开头是对象名，每列的开头是属性名。行 g 和列 m 的交叉表示对象 g 具有属性 m ,一个简单的语境如图 1 所示。

对象/属性	…	m	…	…
…	…	…	…	…
g	…	X	…	…
…	…	…	…	X

图 1 语境的矩阵表示

定义二 对一个对象集 A ，定义 A'={m ∈ M |gIm ,对所有的 g ∈ A}(即 A 中所有的对象共有的属性集合)。相应地,对一个属性集 B ,定义 B' ={g∈ G|gIm,对所有的 m ∈ B}(即包含所有 B 中属性的对象集合)。

定义三 语境(G ， M ， I)的形式概念(formal concept)是个集合对(A ， B),其中 A⊆G ， B⊆M ， 并且 A' =B ,B' = A 。A ， B 分别称作概念(A ， B)的范围(extent)和含义(intent)。 β(G ， M ， I)表示语境(G ， M ， I)中的所有概念集。

命题一 如果(G ， M ， I)是一个语境，对象集 A ， A1 ,A2⊆G ,属性集 B ， B1 ， B2⊆M ,那么有如下结论:

1) A⊆A2 ⇒ A2′ ⊆A1′

1')B1⊆B2 ⇒ B2′ ⊆B1′

2)A⊆A″

2')B⊆B″

3) A′ = A″ ′

3')B′ = B″ ′

4)A⊆B′ ⇔ B⊆A′ ⇔ AxB⊆I

1.3.2 概念格

概念格是 FCA 的核心数据结构。

概念格的每个节点是一个概念，由外延和内涵组成。外延是概念所覆盖的实例；而内涵是概念的描述，是该概念所覆盖实例的共同特征。另外，概念格可以通过其 Hasse 图生动简洁地体现概念之间的泛化和例化关系。[8]

形式中概念分析的数据主要由形式背景表示，一个简单的形式背景如下：

	红色	黄色	甜	酸
苹果	X		X	X
香蕉		X	X	
樱桃	X		X	
柠檬		X		X

图 2 简单的形式背景

定义 4 如果(A1 ， B1)，(A2 ， B2) 都是语境中的概念，并且 A1⊆A2 ，那么(A1 ， B1)被称作(A2 ， B2)的子概念(subconcept)，(A2 ， B2) 则是(A1 ， B1) 的超概念(superconcept)，记为(A1 ， B1)≤ (A2 ， B2)。“ ≤” 反映了概念间的层次关系。由层次关系搭构的所有(G ， M ， I) 的概念记作 β(G ， M ， I),被叫作概念格(concept lat tice)。如图 2 所示为形式背景的概念格。

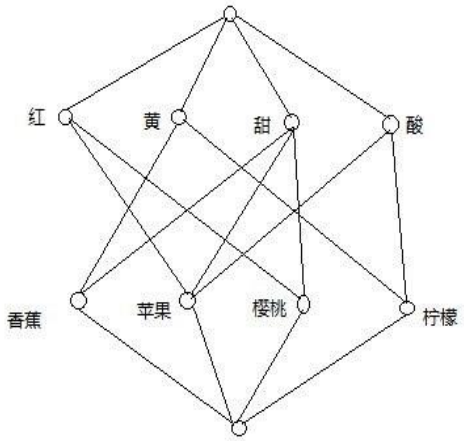

图 3 概念

比较著名的概念格构造的算法有很多，比如：Ganter's Next Concept Algorithm ， Christian Lindig 以及 Fast Concept Analysis 等。

2 形式概念分析的发展

2.1 国外研究成果

国外的主要研究成果如下：

斯坦福大学的 Sahalni 提出 Rulearner 系统，该算法采用一种“标记法”从格中提取规则。根据属性构造出概念格，而后从格中提取出分类规则用于支持对象的分类。

意大利的 Carpineto 建立了 Galois 系统，它只是一个概念聚类系统，主要用于浏览检索之中。

TOSCAN 是 Wille 所在研究所发展的一个基于概念格的数据分析系统。

Njiwoua 等设计的 LEGAL—E 系统使用学习参数来生成半格，然后采用投票的方式对新对象的分类进行群体决策。LEGAL 通过引进两个参数 T 和 U,改进 Bordat 的算法, 将其应用于分类任务。

Godin 等提出了一种增量式概念格建造方法(如前所述)。并在其基础上提出在概念

格上提取蕴含规则的算法。

Pasqllier 等研究了关联规则的提取问题，提出了用于提取确定性关联规则的 Dequenne--Guiglle 基，以及用于近似关联规则的适当基和结构基。

HOTB 研究了基于概念格的概念聚类方法，并实现了一些学习系统，包括 OSIB 和 INCOSHAM。

2.2 国内研究成果

国内研究成果如下：

张文修等引入形式背景的可辨识矩阵，提出了概念格的属性约简理论，给出了概念格约简的判定定理例。

胡可云等提出了利用概念格进行分类和无冗余规则的提取。

谢志鹏等提出利用概念格的层次关系提取关联规则。

刘宗田等提出利用容差关系建立广义概念格提取近似规则。

赵奕等针对概念格与 Rough 集之间的联系提出了 Rough 概念格，并在此基础上提取蕴含规则。

侯锦等提出了利用概念格进行 Rough 集理论中重要的组成部分一属性约简。

2.3 国内外发展状况比较

从目前的研究状况来看：在形式概念分析领域中，国内的发展还是在发展的初级阶段，国内的相关论文并不是很多，且大多数的论文都是在一些计算机领域的综合会议提出来的。国内目前来说还没有特别权威的形式概念分析专门会议。

而国际上，形式概念分析经过许多年的探索与研究，目前已经进入了十分成熟的应用阶段。概念格领域最权威的会议 ICFCA(International Conference of Formal ConceptAnalysis)从 2003 年起每年举行一次，对于形式概念分析方面所提出的新理论以及所存在的问题进行相应的研究。会议每年都发表一部论文集，反映关于形势概念分析的最新研究进展和应用水平。[7]

通过对于往年论文集的分析整理，也可以发现，国外关于形式概念分析的理论研究已经相当成熟，因而已经投入大量精力于形式概念分析的应用；而国内目前来说还是处于形式概念分析的理论学习阶段，真正的相关应用设计相对较少。

3 形式概念分析的应用

通过许多年来的发展，形式概念分析凭借其对于概念形式简单直接的分类处理方式，而在各种领域当中得到了越来越广泛的发展。

目前来讲，形式概念分析主要还是应用于四个主要的领域：本体的构建、软件工程、Web 语义检索以及知识发现。

从本体的角度来说，形式概念分析主要应用于本体的构建、映射以及合并等方面，形式概念分析对于概念以及概念之间关系的呈现方面的所具有的优势，对于本体的知识共享具有非常大的促进作用。

软件工程是近几年来形式概念分析的热点之一，但由于发展时间较短，目前来说，形式概念分析的应用还集中于软件维护以及细节设计之中，但是形式概念分析对于软件开发过程中的组件重构等方面有很大的优势，对于软件的开发维护等方面都拥有很大的促进作用，相信会在之后的研究中，逐渐成为发展趋势。

在 Web 语义检索方面，形式概念分析对于检索结果的优化以及对于不限定领域的检索都有着其得天独厚的优势。

而在只是发现领域，形式概念分析能够以直观地形式展现视图，同时对于信息检索的速度提升效果显著，因而在大型数据库的开发方面也得到很多的应用。

本论文主要针对形式概念分析在软件工程方面的应用进行一些简要的介绍。

3.1 软件工程中的应用

3.1.1 软件工程相关概念

对于软件工程，不同的学者与机构有着不同的定义，但目前来说比较认可的一种说法是：软件工程是研究和应用如何以系统性的、规范化的、可定量的过程化方法去开发和维护软件，以及如何把经过时间考验而证明正确的管理技术和当前能够得到的最好的技术方法结合起来。

根据 IEEE 的定义，一般来说，软件开发过程被分为以下八个步骤：需求分析概要设计详细设计编码单元测试集成测试系统测试维护，而形式概念分析主要应用于需求分析、设计以及维护过程，目前来说主要以软件维护为主。

3.1.2 形式概念在需求分析中的应用

需求分析，就是对要开发的软件的一些需求进行详细的应用分析，也就是通过对于软件应用环境的分析，对于软件应用信息进行收集的过程，以方便于之后软件在各种应用环境下的具体使用。

在需求分析过程中，形式概念分析的主要操作便是为软件构造项目特征集合。通过软件工程中各个部分的工作集和它的属性的对应关系的分析，运用前文提到过的形式概念分析的语境与属性集构造方法即可。

3.1.3 形式概念在结构设计过程中的应用

软件的结构设计，主要是在需求分析的基础上，通过合理的结构化设计，进行详尽的分析设计，得出一个合理的设计方法。

利用假设的方法说明形式概念分析的概念构造原理，通过相关的分析器分析每一个项目在项目过程中的使用情况。[9]

通过对于向量之间相关性的分析，分析其项目特征，形成相关的概念，从而形成系统的概念格。

3.1.4 形式概念在系统设计过程中的应用

在系统设计阶段，最重要的目的便是要设计出概念格，可以利用一些比较著名的概念格构造算法，计算出概念格的所有相关概念以及与此相关的层次结构，继而得出一个完备的概念格即可。[10]

3.1.5 软件维护中的应用

基本的软件维护过程应包括如下几个活动：与修改请求相关的软件理解，修改影响分析，修改实施（包括重构以及修改传播分析），修改后系统的调试与测试。

在软件工程的过程中，形式概念分析在软件维护方面的应用目前来说最为成熟，其应用过程比较简单，包括三步:首先构造形式背景，确定形式对象与形式属性以及两者之间的二元关系；然后根据概念格构造算法生成概念格；最后根据概念格上的特征，与具体应用相结合。

结束语 网络技术的快速发展，带来了大量的数据，纷繁复杂的数据为人工智能带来挑战的同时，带来的也是巨大的发展机遇。

形式概念分析在计算机技术方面的应用愈加重要，随着科技的快速发展，形式概念分析也在迅速发展，随着技术的不断更新，形式概念分析在各个领域的应用也越来越广泛，虽然现在形式概念分析在国内的研究力量尚且薄弱，但相信假以时日，形式概念分宜一定会得到越来越多的关注，在计算机领域大放光彩。

参 考 文 献

[1] Wille R．Restructuring Lattice Theory: an Approach Based on Hierarchies of Concept M.Dordrecht-Boston:Techn．Hochsch．Fach bereich Math，1982: 445－470．

[2] Oosthuizen GD．The Application of Concept Lattice to Machine Learnin［R]．South Africa: University of Pretoria，South Africa，1996．

[3] Ho TB.Incremental Conceptual Clustering in the Framework of Galois Lattice［M]/ /Lu H，Motode H，Liu H，et al．KDD: Thechniques and applications．Singapore: World Scientific，1997，49－64．

[4] Kent R E，Bowman C M．Digital Libraries，Conceptual Knowledge Systems and the Nebula Interface［R］. Arkansas : University of Arkansas，1995.

[5] Carpineto C，Romano G. A Lattice Conceptual Clustering System and Its Application to Browsing Retrieva［J］．Machine Learning，1996,24(2) : 95－122．

[6] Cole R，Eklund P．Scalability in Formal Concept Analysis［J］.Computational Intelligence，1999，15(1) : 11－27．

[7] David Hand,Heikki Mannila.Principles of Data Mining..entRE,BowmanCM.Digital Libraries,Conceptual nowledge systems and the Nebula interface..1995[3]Godin R,Mineau G,Missaoui R,etal.Applying oncept formation methods to software reuse. nt J Software Engand Knowledge Eng. 1996

[8] Godin R,Mineau G,Missaoui R, etal.Applying concept formation methods to software reuse. Int J Software Engand Knowledge Eng . 1996 [5] Njiwoua P,Nguifo E M.Forwarding the choice of bias LEGALF: using feature selection to reduce the complexity of LEGAL. Proceedings of BEELEARN-97,ILK and INFOLAB . 1997

[9] Meunier J.G,Bouchaffre,D.A Markov Mesh Modeling Uncertain Galois Lattice. Analisis Statistica dei Datin Testuali . [7] Mohammed J Zaki,Nagender Parimi,Nilanjana De,Feng Gao,Benjarath Phoophakdee,Joe Urban,Vineet Chaoji,Mohammad Al Hasan,Saeed Salem.Towards Generic Pattern Mining. ICFCA . 2005 .