基于格结构的概念代数研究

张旭洁[1,2] 刘宗田[1] 刘 炜[1] 王 东[1] 田 垄[1]
(上海大学计算机科学与工程学院 上海 200072)[1] (云南师范大学信息学院 昆明 650031)[2]

摘 要 近年来,概念格作为一种有效的概念分析、表示与应用模型,被越来越多的领域所应用。概念代数是在概念格理论基础上提出的一种新的代数系统。目前,概念代数系统还只是雏形,很多方面的内容还需要扩展与进一步的研究。基于格结构的概念代数研究分别对Nilsson概念代数、Wille概念代数和认知信息学中的概念代数的基本概念与内容进行了介绍,对各方法进行了分析与比较,并证明Nilsson概念代数是一种广义概念代数,为进一步的应用与研究提供支持。

关键词 格,概念格,概念代数,概念表示,认知信息学
中图法分类号 TP311 **文献标识码** A

Concept Algebra Based on Lattice

ZHANG Xu-jie[1,2] LIU Zong-tian[1] LIU Wei[1] WANG Dong[1] TIAN Long[1]
(School of Computer Engineering and Science, Shanghai University, Shanghai 200072, China)[1]
(School of Information, Yunnan Normal University, Kunming 650031, China)[2]

Abstract Concept lattice, as an effective model for concept analysis, representation and application, has been widely utilized in many fields. Based upon concept lattice, concept algebra is a newly developing algebra system which needs further enhancement and extension. We introduced and analyzed the Nilsson concept algebra, Wille concept algebra as well as concept algebra in cognitive informatics. Furthermore, this paper elaborated comparative analysis among above three theories of concept algebra. In addition, we proved that Nilsson concept algebra belongs to Wille concept algebra. The findings and conclusions of this paper lay a solid foundation for further improvements and applications of concept algebra.

Keywords Lattice, Concept lattice, Concept algebra, Concept representation, Cognitive informatics

1 引言

概念在哲学中被理解为由外延和内涵两个部分组成的思想单元。形式概念分析[1]中,概念的外延被表示为属于这个概念的所有对象的集合,而内涵则被表示为所有这些对象所共有的特征(或属性)集合。概念格作为形式概念分析中核心的数据结构,实现了对概念哲学理解的形式化,从本质上描述了对象和属性之间的关系,表明了概念之间的泛化与特化关系,并采用Hasse图实现了对数据的可视化。经历了30年的发展,作为数据分析和知识处理的形式化工具,形式概念分析及概念格模型已经被广泛地应用到知识工程[2-4]、信息处理[5]等各领域[6,7]。

概念格理论奠定了关于概念表示、认知、推理等方面的理论基础,概念代数的研究正是其中的一个重要方向。Nilsson[8,9]教授1994年提出的概念代数以概念格理论为数学背景,用于概念知识表示,其能够把概念、关系、框架、对象以及继承等重要概念进行统一的代数表示[17]。2000年,Wille[10]发现在概念叙述中常常包括不具有某个属性的描述,由此在原有概念格基础上定义了弱非运算与对偶弱非运算,并由此引出概念代数与广义概念代数[11]的属性讨论与进一步的研究。在认知信息学领域,Wang[12]2008年给出了一种概念代数的形式化表示,这一方法的基础是对象、属性、关系(the object-attribute-relation theory, OAR)理论[13],但其核心结构仍然为概念格结构。该方法试图模拟人类大脑对概念知识的存储、表示与应用。

本文主要介绍基于概念格理论的不同概念代数的基本概念与内容;第2节主要介绍概念格理论、Nilsson概念代数和Wille概念代数的定义与表示方法等;第3节详细介绍认知信息学中对概念的形式化表示方法即Wang概念代数;最后给出不同概念代数的分析与总结。

2 基本概念

2.1 形式概念分析

在介绍形式概念分析(Formal Concept Analysis, FCA)之

到稿日期:2012-05-10 返修日期:2012-08-11 本文受国家自然科学基金项目(61273328,60975033),上海市科委项目(12ZR1410900),云南省教育厅科学研究基金项目(09Y0144)资助。

张旭洁(1980—),女,博士生,讲师,主要研究方向为知识表示、自然语言处理,E-mail:zhangxujie@yahoo.cn;刘宗田(1946—),男,教授,博士生导师,主要研究方向为人工智能、软件工程等;刘 炜(1978—),男,博士,副研究员,主要研究方向为语义本体、描述逻辑;王 东(1986—),男,硕士生,主要研究方向为事件网络、知识表示;田 垄(1987—),男,硕士生,主要研究方向为语言信息处理。

前，首先定义一些基本概念。

定义 1 若 M 与 N 是两个集合，则它们的笛卡尔积 $M\times N$ 的任何一个子集 $R\subseteq M\times N$，都称为是 M 与 N 之间的关系，当 $M=N$ 时称为是 M 上的关系。若 $a\in M, b\in N$ 有关系 R，即二元组 $(a,b)\in R$，则它可记作 aRb。

定义 2 集合 M 上的关系 R 称为偏序关系，当对所有 $x,y,z\in M$ 都有：(1)自反性：xRx；(2)反对称性：xRy 且 yRx，则 $x=y$；(3)传递性：xRy 且 yRz，则 xRz。偏序关系 R 常用符号“$\leqslant$”来表示，当 $x\leqslant y$ 且 $x\neq y$ 时，还记作 $x<y$。

定义 3 设 M 是一个集合，$\leqslant$ 是 M 上的偏序关系，则二元组 $(M,\leqslant)$ 称为偏序集。当 $\leqslant$ 是全序时称为全序集。$(M,\leqslant)$ 的对偶记作 $(M,\leqslant)^d$ 是偏序集 $(M,\geqslant)$。

定义 4 设 $(M,\leqslant)$ 是一个偏序集，$A\subseteq M$，若对 A 中所有元素 a 都有 $x\leqslant a(x\in M)$，则称 x 是 A 的一个下界；若对 A 中所有元素 a 都有 $a\leqslant y$ $(y\in M)$，则称 y 是 A 的一个上界。若 A 的所有下界的集合有最大元素，则称其为 A 的下确界；若 A 的所有上界的集合有最小元素，则称其为 A 的上确界，并分别记作 $\inf A$ 及 $\wedge A$ 或 $\sup A$ 及 $\vee A$。当只有两个元素 x,y 时，记作 $x\wedge y$ 及 $x\vee y$。

定义 5 设 $(M,\leqslant)$ 是一个偏序集，若 M 中的任何两个元素都有上确界及下确界，则称 $(M,\leqslant)$ 是一个格。若 M 中的任何一个子集都有上确界与下确界，则称 $(M,\leqslant)$ 是一个完全格。显然当 M 中包含有限个元素时，格一定是完全格。

在 FCA 中，数据是用形式背景来表示的，下面给出形式化定义。

定义 6 设 U 是对象的集合，M 是属性的集合，I 是两个集合 U 与 M 间的关系，则称三元组 $\mathbb{K}=(U,M,I)$ 为一个形式背景（简称背景）。$(u,m)\in I$（或写成 uIm）表示对象 u 具有属性 m。背景可以用一个二维表来表示，行表示对象，列表示属性。若 u 行 m 列的交叉处是√，则表示对象 u 有属性 m，如表 1 所列。

表 1 某形式背景

	m_1	m_2	m_3	m_4
u_1	√	√		√
u_2	√		√	
u_3		√	√	
u_4	√	√		√
u_5	√			

定义 7 设 $\mathbb{K}=(U,M,I)$ 是一个背景，若 $A\subseteq U, B\subseteq M$，令 $f(A)=\{m\in M\mid \forall u\in A,(u,m)\in I\}$ 及 $g(B)=\{u\in U\mid \forall m\in B,(u,m)\in I\}$，如果 A,B 满足 $f(A)=B, g(B)=A$，称二元组 (A,B) 是一个概念。A 是概念 (A,B) 的外延，即对象，B 是概念 (A,B) 的内涵，即属性。用 $\mathscr{B}(U,M,I)$ 或 $\mathscr{B}(\mathbb{K})$ 表示背景 $\mathbb{K}=(U,M,I)$ 上的所有概念的集合。

定理 1[14] 设 (U,M,I) 是一个背景，$A,A_1,A_2\in U; B,B_1,B_2\in M$，则有：

(1)$A_1\subseteq A_2\Rightarrow f(A_2)\subseteq f(A_1)$，(1)′$B_1\subseteq B_2\Rightarrow g(B_2)\subseteq g(B_1)$；

(2)$A\subseteq g(f(A))$，(2)′$B\subseteq f(g(B))$；

(3)$f(A)=f(g(f(A)))$，(3)′$g(B)=g(f(g(B)))$；

(4)$f(A_1)\cap f(A_2)=f(A_1\cup A_2)$，(4)′$g(B_1)\cap g(B_2)=g(B_1\cup B_2)$。

定理 2[1] 若 T 是一个指标集，而且对每个 $t\in T, A_t\subseteq U$ 都是对象的集合，则：$f(\bigcup_{t\in T}A_t)=\bigcap_{t\in T}f(A_t)$ 及 $g(\bigcup_{t\in T}B_t)=\bigcap_{t\in T}g(B_t)$。

定义 8 若 $(A_1,B_1),(A_2,B_2)$ 是某个背景 $\mathbb{K}=(U,M,I)$ 上的两个概念，而且 $A_1\subseteq A_2$，则称 (A_1,B_1) 是 (A_2,B_2) 的子概念，(A_2,B_2) 是 (A_1,B_1) 的父概念，并记作 $(A_1,B_1)\leqslant(A_2,B_2)$。如果 $A_1\subset A_2$，且不存在概念 (A_3,B_3) 使得 $A_1\subset A_3\subset A_2$，则称 (A_1,B_1) 是 (A_2,B_2) 的直接子概念，(A_2,B_2) 是 (A_1,B_1) 的直接父概念，并记作 $(A_1,B_1)\prec(A_2,B_2)$。关系 $\leqslant$ 显然是偏序关系，用 $\mathscr{B}(U,M,I)$ 来表示 (U,M,I) 的所有概念的集合 $\mathscr{B}(U,M,I)$ 按这种序形成的偏序集 $(\mathscr{B}(U,M,I),\leqslant)$。

定理 3[1] 设 (U,M,I) 是一个背景，T 是一个指标集，$\{(A_t,B_t)\mid t\in T\}\in\mathscr{B}(U,M,I)$，则 $\{(A_t,B_t)\mid t\in T\}$ 的下确界及上确界分别是 $\bigwedge_{t\in T}(A_t,B_t)=(\bigcap_{t\in T}A_t, f(g(\bigcup_{t\in T}B_t)))$ 及 $\bigvee_{t\in T}(A_t,B_t)=(g(f(\bigcup_{t\in T}A_t)),\bigcap_{t\in T}B_t)$。

$\mathscr{B}(U,M,I)$ 的任意子集都有上确界，所以偏序集 $(\mathscr{B}(U,M,I),\leqslant)$ 即 $\mathscr{B}(U,M,I)$ 是一个完全格，称之为概念格。图 1 是表 1 所列形式背景对应的概念格。

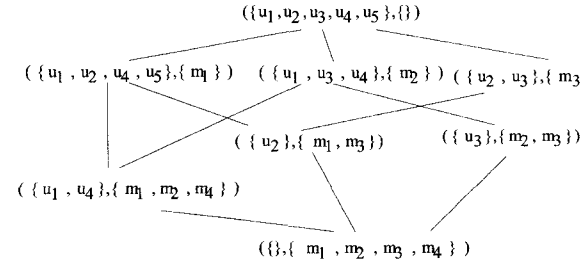

图 1 形式背景所对应的概念格

2.2 概念代数

2.2.1 Nilsson 概念代数

通常情况下，概念知识的组织均采用分类方法，利用概念之间的包含关系来构建模型，由于这些包含关系（如“is-a”、“subclass”等，“part-of”除外）构建的概念知识结构与数学的格结构是非常相似的，Nilsson 教授[9]于 1994 年提出用数学格来描述概念的思想。

由于每个代数格都是一个偏序格[14]，也就是说，如果 $(M,\leqslant)$ 是一个格，$\leqslant$ 为偏序关系，$\wedge$ 和 $\vee$ 分别表示格内的交和并，那么：$\varphi\leqslant\Psi$ 当且仅当 $\varphi=\varphi\wedge\Psi$；$\varphi\leqslant\Psi$ 当且仅当 $\Psi=\varphi\vee\Psi(\varphi\in M,\Psi\in M)$。

由于概念的包含关系也是一种偏序关系，因此可以建立代数格与概念包含之间的关系，即：$a\leqslant b$ 当且仅当 $a=a\times b$；$a\leqslant b$ 当且仅当 $b=a+b$[9,15]。

概念代数中概念的继承、交、并分别用符号 $\leqslant,\times,+$ 表示，它们分别由代数格理论中的偏序、交、并约束，引入全概念（univ）和空概念（null）来定义格的界。

公理 1 设 a,b,c 为概念格的任意概念，对于分配有界概念格存在如下概念公理：

幂等律：$a+a=a, a\times a=a$；

交换律：$a+b=b+a, a\times b=b\times a$；

结合律：$a+(b+c)=(a+b)+c, a\times(b\times c)=(a\times b)\times c$；

吸收率：$a+(a\times b)=a, a\times(a+b)=a$；

分配率：$a\times(b+c)=(a\times b)+(a\times c), a+(b\times c)=(a+b)\times(a+c)$；

有界律：$a+\text{null}=a, a\times\text{null}=\text{null}, a+\text{univ}=\text{univ}, a\times\text{univ}=a$。

概念之间除了简单的包含关系外，还存在其它复杂的关系，Nilsson[15]把概念之间的其它关系看作是概念的属性。属

性函数表示为 $r(\varphi)$，意思是属性 r 的值为 φ，用“:”运算符表示为 $r:\varphi$，通常作为概念项 $\Psi\times r(\varphi)$ 形式的一个因子出现，表示概念项 Ψ 的子概念有属性 r，值为 φ。概念属性结构的一般表示形式为：$c\times r_1(\varphi_1)\times r_2(\varphi_2)\times\cdots\times r_m(\varphi_m)$ 或 $c\begin{bmatrix} r_1:\varphi_1 \\ r_2:\varphi_2 \\ \vdots \\ r_m:\varphi_m \end{bmatrix}$。

这种表示形式也称为框架项。需要说明的是属性值 $\varphi_i(i=1,2,\cdots,m)$ 为概念项，可以递归表示，即还可以是框架项或其它概念项。

公理 2 属性满足以下公理(其中 $r_i(i=1,2,\cdots,m)$)：

$r_i(x+y)=r_i(x)+r_i(y)$

$r_i(x\times y)=r_i(x)\times r_i(y)$

$r_i(\text{null})=\text{null}$

$r_i(\text{univ})=\text{univ}$

定理 4 属性函数是单调的，即 $x\leqslant y\Rightarrow r_i(x)\leqslant r_i(y)$。文献[16]给出了证明。

定义 9 概念项由概念常量 C、概念变量 φ、属性项 $r_i(\varphi_i)$、全概念(univ)和空概念(null)，通过运算符“+”、“×”、“:”连接而成，记概念项为 $\langle C,A\rangle$，其中 C 为概念标示符，A 为属性标示符；不含概念变量的概念项称为基项。例如：概念项 $c_i\times r_1(\varphi_1)\times r_2(\varphi_2)\times\cdots\times r_m(\varphi_m)$，其中 $c_i\in C$。

概念项的定义类似于谓词公式，概念变量 φ 可以定义为任意概念项，实现了概念项的递归定义。

定义 10 概念句是表示概念相等的代数等式：$\varphi=\Psi$ 和条件等式 $\varphi=\Psi$ if $\varphi=\Psi$。例如：Hello Kitty=cat×toy(soft)×color(white)×gender(female)。

定义 11 概念代数系统是一个五元组 $(C,A,+,\times,:)$，其中 C 为概念标示符，A 为属性标示符，“+”、“×”、“:”分别表示交、并和属性函数表示符(用来描述属性值的符号)。

Nilsson 概念代数还可以通过逻辑中的 λ-表达式对概念代数的概念项和句子进行重构[8]。文献[17]对概念代数五元组进行了扩充。

2.2.2 Wille 概念代数

在 FCA 中，一个形式背景确定形式概念时，要根据概念的内涵具有哪些属性来决定，但在很多实际问题中往往有些概念是根据它不具有哪些属性来确定的，因此形式概念分析还应该研究根据不具有某些属性而形成的概念的各种性质。Wille[10]给出了概念代数的定义，研究了背景 $\mathbb{K}=(U,M,I)$ 的概念格中每个概念 (A,B) 的内涵的补集 $M\text{-}B$ 形成概念的各种性质，以及对偶的外延的补集 $U\text{-}A$ 形成概念的各种性质。

定义 12 设 $\mathbb{K}=(U,M,I)$ 是一个背景，$\mathcal{L}=\underline{\mathcal{B}}(U,M,I)$ 是它的概念格。在 $\mathcal{L}$ 上定义两个运算：$\Delta:\mathcal{L}\mapsto\mathcal{L}$ 及 $\nabla:\mathcal{L}\mapsto\mathcal{L}$。

弱非运算 $\Delta:(A,B)\mapsto(g(f(U-A)),f(U-A))$

对偶弱非运算 $\nabla:(A,B)\mapsto(g(M-B),f(g(M-B)))$

记 $(g(f(U-A)),f(U-A))$ 为 $(A,B)^{\Delta}$，称为 (A,B) 的弱非；记 $(g(M-B),f(g(M-B)))$ 为 $(A,B)^{\nabla}$，称为对偶弱非。

定义 13 设 $\mathbb{K}=(U,M,I)$ 是一个背景，$\mathcal{L}=\underline{\mathcal{B}}(\mathbb{K})$ 是其上的概念格，则称代数系统 $(\mathcal{L},\Delta,\nabla)$ 为 $\mathcal{L}$ 上的概念代数。

定理 5[10] 设 $(\mathcal{L},\Delta,\nabla)$ 是一个概念代数，则任意 $x,y\in\mathcal{L}$，有以下性质：

(1) $x^{\Delta\Delta}\leqslant x,(1)'x^{\nabla\nabla}\geqslant x$；

(2) $x\leqslant y\Rightarrow x^{\Delta}\geqslant y^{\Delta},(2)'x\leqslant y\Rightarrow x^{\nabla}\geqslant y^{\nabla}$；

(3) $(x\wedge y)\vee(x\wedge y^{\Delta})=x,(3)'(x\vee y)\wedge(x\vee y^{\nabla})=x$。

Kwuida[11]讨论了非概念格的概念代数。

定义 14[11] 设 $\mathcal{L}$ 是一个格，$\Delta:\mathcal{L}\mapsto\mathcal{L}$ 及 $\nabla:\mathcal{L}\mapsto\mathcal{L}$ 是 $\mathcal{L}$ 上的两个运算且满足 $\forall x,y\in\mathcal{L}$：

(1) $x^{\Delta\Delta}\leqslant x,(1)'x^{\nabla\nabla}\geqslant x$；

(2) $x\leqslant y\Rightarrow x^{\Delta}\geqslant y^{\Delta},(2)'x\leqslant y\Rightarrow x^{\nabla}\geqslant y^{\nabla}$；

(3) $(x\wedge y)\vee(x\wedge y^{\Delta})=x,(3)'(x\vee y)\wedge(x\vee y^{\nabla})=x$。

则称 $(\mathcal{L},\Delta,\nabla)$ 是一个“广义概念代数”，简称“概念代数”。如果一个概念代数也是一个形式背景产生的概念格上的概念代数，则称这个概念代数是可表现的。

3 认知信息学中的概念代数

在机器智能的研究过程中，人们一直在探寻一个问题：人类大脑是如何处理内部和外部信息并实现思考的，而这种思考的过程是非命令驱动方式的。在这样的需求下，认知信息学孕育而生。认知信息学[18]是一门交叉学科，它涵盖了计算机科学、信息科学、认知科学和智能科学领域研究人类大脑和自然智能的内部信息处理机制和过程，以及有关认知计算的工程应用。

Wang[12]提出的概念代数用于抽象概念的表示与计算，而这一思想也构建在概念格的基础之上。

定义 15 语义背景 $\Theta,\Theta=(\mathcal{O},\mathcal{A},\mathcal{R})$，其中 $\mathcal{O}$ 表示对象，是有限或无限的非空集合；$\mathcal{A}$ 表示属性，是有限或无限的非空集合，$\mathcal{R}$ 是 $\mathcal{O}$ 和 $\mathcal{A}$ 之间的二元关系，$\mathcal{R}::=\mathcal{O}\to\mathcal{O}|\mathcal{O}\to\mathcal{A}|\mathcal{A}\to\mathcal{O}|\mathcal{A}\to\mathcal{A}$。

定义 16 在语义背景 Θ 上定义概念项为一个五元组 $c::=(O,A,R^c,R^i,R^o)$，其中 O 是概念对象非空集合($O=\{o_1,o_2,\cdots,o_n\}\subseteq\mathbb{P}\mathcal{O}$，$\mathbb{P}\mathcal{O}$ 是 $\mathcal{O}$ 的幂集)；A 是概念属性非空集合($A=\{a_1,a_2,\cdots,a_n\}\subseteq\mathbb{P}\mathcal{A}$，$\mathbb{P}\mathcal{A}$ 是 $\mathcal{A}$ 的幂集)；R^c 是内部关系集合($R^c=O\times A$)；R^i 是输入关系集合($R^i\subseteq C'\times C$，C' 表示外部概念集合)；R^o 是输出关系集合($R^o\subseteq C\times C'$)。$c^*=A$ 表示概念的内涵，$c^+=O$ 表示概念的外延。

定义 17 概念代数是一个三元组 $CA::=(C,OP,\Theta)=(\{O,A,R^c,R^i,R^o\},\{\bullet_r,\bullet_c\},\Theta)$，其中 $OP=\{\bullet_r,\bullet_c\}$ 表示概念之间的关系与概念代数操作符。概念代数为抽象概念提供了一组关系比较与概念代数操作方法。其中 $\bullet_r=\{\leftrightarrow,\not\leftrightarrow,\prec,\succ,=,\cong,\sim,\triangleq\}$ 分别表示概念之间的关联关系、独立关系、子关系、父关系、等价关系、组成关系、相似关系、复制关系，概念之间的关系判断详见表 2 概念关系表；$\bullet_c=\{\Rightarrow,\overset{-}{\Rightarrow},\overset{+}{\Rightarrow},\overset{\sim}{\Rightarrow},\uplus,\Subset,\Leftarrow,\vdash,\mapsto\}$ 分别表示概念代数的继承、部分继承、扩展、替换、合成、分解、泛化、特化、实例化构图操作，详见表 3。

表 2 概念关系表

名称	符号	定义
关联关系	$\leftrightarrow$	$c_1\leftrightarrow c_2::=A_1\cap A_2\neq\emptyset$
独立关系	$\not\leftrightarrow$	$c_1\not\leftrightarrow c_2::=A_1\cap A_2=\emptyset$
子关系	$\prec$	$c_1\prec c_2::=A_1\supset A_2$
父关系	$\succ$	$c_1\succ c_2::=A_1\subset A_2$
等价关系	$=$	$c_1=c_2::=(A_1=A_2)\wedge(O_1=O_2)$
组成关系	$\cong$	$c_1\cong c_2::=(c_1\prec c_2)\vee(c_1\succ c_2)=(A_1\supset A_2)\vee(A_1\subset A_2)$
相似关系	$\sim$	$c_1\sim c_2::=\dfrac{\lvert A_1\cap A_2\rvert}{\lvert A_1\cup A_2\rvert}\times 100\%$
定义关系	$\triangleq$	$c_1(O_1,A_1,R_1^c,R_1^i,R_1^o)\triangleq c_2(O_2,A_2,R_2^c,R_2^i,R_2^o)::=c_1(O_1,A_1,R_1^c,R_1^i,R_1^o\mid O_1=O_2,A_1=A_2,R_1^c=O_1\times A_1,R_1^i=R_2^i,R_1^o=R_2^o)$

表 3　概念代数操作

名称	符号	定义
继承操作	$\Rightarrow$	$c_1(O_1,A_1,R_1^c,R_1^i,R_1^o)\Rightarrow c(O,A,R^c,R^i,R^o)$ $::=c(O,A,R^c,R^i,R^o \mid O=O_1, A=A_1, R^c=O\times A,$ $R^i=R_1^i\cup\{(c_1,c)\}, R^o=R_1^o\cup\{(c,c_1)\})$ $\mid\mid c_1(O_1,A_1,R_1^c,R_1^i,R_1^o \mid R_1^i=R_1^i\cup\{(c,c_1)\}, R_1^o=R_1^o\cup\{(c_1,c)\})$
部分继承	$\overset{-}{\Rightarrow}$	$c_1(O_1,A_1,R_1^c,R_1^i,R_1^o)\overset{-}{\Rightarrow}c(O,A,R^c,R^i,R^o)$ $::=c(O,A,R^c,R^i,R^o \mid O=O_1\backslash O', A=A_1\backslash A', R^c=O\times A\subset R_1^c, R^i=R_1^i\cup\{(c_1,c)\}, R^o=R_1^o\cup\{(c,c_1)\})$ $\mid\mid c_1(O_1,A_1,R_1^c,R_1^i,R_1^o \mid R_1^i=R_1^i\cup\{(c,c_1)\}, R_1^o=R_1^o\cup\{(c_1,c)\})$
扩展操作	$\overset{+}{\Rightarrow}$	$c_1(O_1,A_1,R_1^c,R_1^i,R_1^o)\overset{+}{\Rightarrow}(O,A,R^c,R^i,R^o)$ $::=c(O,A,R^c,R^i,R^o \mid O=O_1\cup O'', A=A_1\cup A'', R^c=O\times A\supset R_1^c, R^i=R_1^i\cup\{(c_1,c)\}, R^o=R_1^o\cup\{(c,c_1)\})$ $\mid\mid c_1(O_1,A_1,R_1^c,R_1^i,R_1^o \mid R_1^i=R_1^i\cup\{(c,c_1)\}, R_1^o=R_1^o\cup\{(c_1,c)\})$
替换操作	$\overset{\sim}{\Rightarrow}$	$c_1(O_1,A_1,R_1^c,R_1^i,R_1^o)\overset{\sim}{\Rightarrow}c(O,A,R^c,R^i,R^o)$ $::=c(O,A,R^c,R^i,R^o \mid O=(O_1\backslash O')\cup O'', A=(A_1\backslash A')\cup A'', R^c=O\times A, R^i=R_1^i\cup\{(c_1,c)\}, R^o=R_1^o\cup\{(c,c_1)\})$ $\mid\mid c_1(O_1,A_1,R_1^c,R_1^i,R_1^o \mid R_1^i=R_1^i\cup\{(c,c_1)\}, R_1^o=R_1^o\cup\{(c_1,c)\})$
合成操作	$\uplus$	$c(O,A,R^c,R^i,R^o)\overset{n}{\underset{i=1}{\uplus}}c_i(O_i,A_i,R_i^c,R_i^i,R_i^o)$ $::=c(O,A,R^c,R^i,R^o \mid O=\overset{n}{\underset{i=1}{\cup}}Oc_i, A=\overset{n}{\underset{i=1}{\cup}}Ac_i,$ $R^c=\overset{n}{\underset{i=1}{\cup}}(R_{\xi_i}^c\cup\{(c,c_i),(c_i,c)\}), R^i=\overset{n}{\underset{i=1}{\cup}}R_{\xi_i}^i, R^o=\overset{n}{\underset{i=1}{\cup}}R^oc_i)$ $\mid\mid \overset{n}{\underset{i=1}{R}}c_i(O_i,A_i,R_i^c,R_i^{i'},R_i^{o'} \mid R_i^{i'}=R_i^i\cup\{(c,c_i)\}, R_i^{o'}=R_i^o\cup\{(c_i,c)\})$
分解操作	$\pitchfork$	$c(O,A,R^c,R^i,R^o)\pitchfork R_{i=1}^n c_i$ $::=R_{i=1}^n\{c_i(O_i,A_i,R_i^c,R_i^{i'},R_i^{o'} \mid R_i^{i'}=R_i^i\cup\{(c,c_i)\}, R_i^{o'}=R_i^o\cup\{(c_i,c)\})$ $\mid\mid c(O,A,R^{c'},R^{i'},R^{o'} \mid R^{c'}=\overset{n}{\underset{i=1}{\cup}}(R_{\xi_i}^c\backslash\{(c,c_i)(c_i,c)\}), R^{i'}=R^i\cup\{R_{i=1}^n(c_i,c)\}, R^{o'}=R^o\cup\{R_{i=1}^n(c,c_i)\})$ $\backslash c_i(O_i,A_i,R_i^c,R_i^i,R_i^o)$ $\}$
泛化操作	$\Leftarrow$	$c(O,A,R^c,R^i,R^o)\Leftarrow c_1(O_1,A_1,R_1^c,R_1^i,R_1^o)$ $::=c(O,A,R^c,R^i,R^o \mid O\supset O_1, A\subset A_1,$ $R^c=O\times A\cup\{(c,c_1)(c_1,c)\},$ $R^i=R_1^i\cup\{(c_1,c)\}, R^o=R_1^o\cup\{(c,c_1)\})$ $\mid\mid c_1(O_1,A_1,R_1^c,R_1^i,R_1^o \mid R_1^i=R_1^i\cup\{(c,c_1)\}, R_1^o=R_1^o\cup\{(c_1,c)\})$
特化操作	$\vdash$	$c_1(O_1,A_1,R_1^c,R_1^i,R_1^o)\vdash c(O,A,R^c,R^i,R^o)$ $::=c(O,A,R^c,R^i,R^o \mid O\subset O_1, A\supset A_1,$ $R^c=O\times A\cup\{(c,c_1)(c_1,c)\},$ $R^i=R_1^i\cup\{(c_1,c)\}, R^o=R_1^o\cup\{(c,c_1)\})$ $\mid\mid c_1(O_1,A_1,R_1^c,R_1^i,R_1^o \mid R_1^i=R_1^i\cup\{(c,c_1)\}, R_1^o=R_1^o\cup\{(c_1,c)\})$
实例化操作	$\vdash\!\!\!\rightarrow$	$c(O,A,R^c,R^i,R^o)\vdash\!\!\!\rightarrow R_{i=1}^n o_i(A_{oi},R_{oi}^c,R_{oi}^i)$ $::=R_{i=1}^n o_i(A_{oi},R_{oi}^c,R_{oi}^i \mid o_i\subset O, A_{oi}\subset A, R_{oi}^i=o_i\times A_{oi}, R_{oi}^i=R_{oi}^i\cup\{(c,o_i)\})$ $\mid\mid c(O,A,R^c,R^i,R^o \mid R^i=R^i\cup\{R_{i=1}^n(o_i,c)\},$ $R^o=R^o\cup\{\overset{n}{\underset{i=1}{R}}(c,o_i)\})$

例如：在语义背景 Θ 下定义一个新概念，即：

$c::=(O,A,R^c,R^i,R^o \mid O\subset\mathcal{O}, A\subset\mathcal{A}, R^c=O\times A, R^i=\varnothing, R^o=\varnothing)$

其中 $R^i=R^o=\varnothing$，表示定义操作是一个新概念的初始化过程，输入和输出关系暂时为空，或许稍后将给出。图 2 是概念代数的层次结构，图 3 是概念代数概念网络。

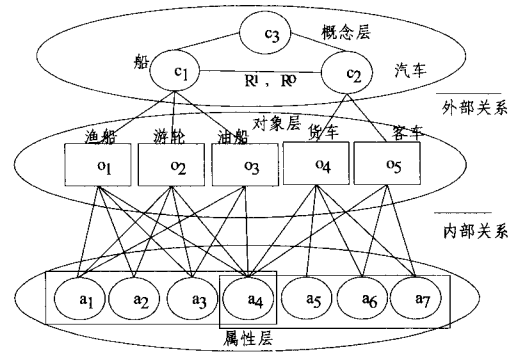

图 2　概念代数层次结构

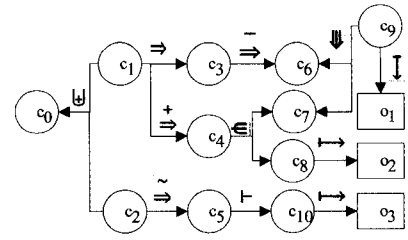

图 3　概念关系网络

4　分析与比较

根据以上对概念代数的详细描述，对各种概念代数分析比较得表 4。

可以看出，FCA 既精炼又严谨，运用哲学中概念的定义解决了对象按照属性形成概念的演算规律，揭示了概念之间继承关系的本质。但该方法除了表示概念之间的继承关系外，无法表示具有多种含义的抽象概念及其它复杂的概念关系；对象、属性、属性值的表示过于简化，不能表示属性、属性值的层次关系，过度依赖于形式背景知识，无法表示动态知识。

Nilsson 概念代数将代数格引入概念包含关系中，构成最初的概念代数的思想。该方法解决了由基本概念通过某些属性值的约束，形成子概念的演算。通过概念项与概念句的表达方式对概念进行表示与描述，由于概念项的递归定义，使其可以表示非常复杂的概念并用概念属性描述除分类关系之外的其它关系，同时这种表示方法还可以实现逻辑重构。该方法虽然采用递归定义可以对属性与属性值进行层次化的定义，但仍然没能将属性与属性值的关系描述清楚，并且缺少否定描述机制。

Wille 概念代数的提出是为了讨论不具有某些属性的概念，该方法研究了形式背景 $\mathbb{K}=(U,M,I)$ 的概念格中每个概念 (A,B) 的内涵的补集 $M\text{-}B$ 形成概念的各种性质，以及对偶的外延的补集 $U\text{-}A$ 形成概念的各种性质。Wille 概念代数仍然没有超脱 FCA 的思想，虽然增加了补运算，但实质上属性的补集仍然是一个属性集合，这与 FCA 的思想一致，只是构建概念的属性集不同。

通过上面对 Wille 概念代数的介绍，可以证明 Nilsson 概念代数是一种广义概念代数。证明过程如下：

设 $S=\{\langle C_1,A_1\rangle,\langle C_2,A_2\rangle,\cdots,\langle C_m,A_m\rangle\}$ 是概念项集，

并引入全概念(univ)和空概念(null)，则$(S\cup univ\cup null,\subseteq)$是一个格。令$\mathscr{C}=S\cup univ\cup null$，定义$\mathscr{L}=(\mathscr{C},\subseteq)$上的$\Delta$与$\nabla$运算如下，$\forall x\subseteq\mathscr{C}$，$x^{\Delta}=\mathscr{C}-x$，$x^{\nabla}=\mathscr{C}-x$，$univ^{\Delta}=null$，$univ^{\nabla}=null$，$null^{\Delta}=univ$，$null^{\nabla}=univ$，显然满足：

(1)$x^{\Delta\Delta}=\mathscr{C}-(\mathscr{C}-x)=x$，$univ^{\Delta\Delta}=null^{\Delta}=univ$，$null^{\Delta\Delta}=univ^{\Delta}=null$；

(1)$'x^{\nabla\nabla}=x$同理可证；

(2)对于$null\subseteq x$，因$null^{\Delta}=univ$，$x^{\Delta}=\mathscr{C}-x$，则$univ\supseteq\mathscr{C}-x$，即$null^{\Delta}\supseteq x^{\Delta}$；

对于$null\subseteq univ$，因$null^{\Delta}=univ$，$univ^{\Delta}=null$，则$univ\supseteq null$，即$null^{\Delta}\supseteq univ^{\Delta}$；

对于$x\subseteq y$，因$x^{\Delta}=\mathscr{C}-x$，$y^{\Delta}=\mathscr{C}-y$；则$\mathscr{C}-x\supseteq\mathscr{C}-y$，即$x^{\Delta}\supseteq y^{\Delta}$；

对于$x\subseteq univ$，因$univ^{\Delta}=null$，$x^{\Delta}=\mathscr{C}-x$，则$\mathscr{C}-x\supseteq null$，即$x^{\Delta}\supseteq univ^{\Delta}$；

(2)$'x\leqslant y\Rightarrow x^{\nabla}\geqslant y^{\nabla}$同理可证；

(3)对于$(null\wedge x)\vee(null\wedge x^{\Delta})$，即$(null\wedge x)\vee(null\wedge x^{\Delta})=null\vee null=null$；

对于$(null\wedge univ)\vee(null\wedge univ^{\Delta})$，即$(null\wedge univ)\vee(null\wedge univ^{\Delta})=null\vee null=null$；

对于$(x\wedge y)\vee(x\wedge y^{\Delta})$，即$(x\wedge y)\vee(x\wedge y^{\Delta})=(x\cap y)\cup(x\cap(\mathscr{C}-y))=x$；

对于$(univ\wedge x)\vee(univ\wedge x^{\Delta})$，即$(univ\wedge x)\vee(univ\wedge x^{\Delta})=x\vee x^{\Delta}=univ$；

(3)$'(x\vee y)\wedge(x\vee y^{\nabla})=x$同理可证。

所以由定义14可知，$((\mathscr{C},\subseteq),\Delta,\nabla)$是一个广义概念代数。

Wang[12]提出的概念代数知识表示模型可以处理复杂的知识结构，实现知识表示中概念的动态更新与扩展，且更多的内容还在研究中。与FCA相比，Wang概念代数将概念从一个二元组转变为一个三元组，其关系集合包括内部关系、输入关系和输出关系。输入输出关系使传统的静态概念与其它概念动态相关，可以表示和处理复杂的概念操作。概念项的动态可适应性说明它是一个动态结构，通过输入和输出关系使自身具有与其它概念相互关联的适应能力。

表4 概念代数的比较

	概念格	Nilsson概念代数	Wille概念代数	wang概念代数
概念项	(A,B)	(C,A)	(A,B)	c::=(O,A,R^c,R^i,R^o)
概念项的产生	算法生成	手工定义与添加	算法生成	手工定义与添加
偏序关系	$A_1\subseteq A_2$	概念之间的包含关系如："is-a"、"subclass"等，除了"part-of"	$A_1\subseteq A_2$	概念关系比较：子关系"≺"、父关系"≻"、等价关系"="、组成关系"≅"；概念代数操作：泛化操作"⇐"、特化操作"⊢"
图的结构	Hasse图，节点为概念项，边表示概念之间的偏序关系	节点为概念项，边为概念项之间的偏序关系与概念及其属性之间的关系	同概念格的图结构	节点包括概念、对象、属性，概念实例；边描述概念内部关系与外部关系。概念内部关系包括：概念与其对象、属性的关系。概念外部关系包括：概念之间的偏序与非偏序关系
运用领域	知识工程、软件工程、信息科学等	知识表示、数据库、本体等	同概念格	认知科学等
数学严谨性	非常严谨	较严谨	非常严谨	有待完善与讨论

结束语 本文在概念格的基础上介绍了Nilsson概念代数、Wille概念代数、认知信息学中的概念代数，并分析比较了各概念代数的优缺点，证明Nilsson概念代数是一种广义概念代数。目前，关于概念代数的研究还在起步阶段，在实际应用中还存在许多问题，但概念代数的提出为认知信息学、自然语言理解、知识表示、信息处理，模拟大脑对概念的存储、记忆和思考等关于概念处理的研究内容奠定了理论基础。

参考文献

[1] Wille R. Restructuring lattice theory: an approach based on hierarchies of concepts[C]// Rival I. Ordered Sets. Boston: Ridel, 1982:445-470

[2] Mineau G W, Godin R. Automatic structuring of knowledge bases by conceptual clustering. Knowledge and Data Engineering [J]. IEEE Transactions on, 1995, 7(5):824-829

[3] Cimiano P, Hotho A, Stumme G, et al. Learning concept hierarchies from text corpora using formal concept analysis[J]. Journal of Artificial Intelligence Research, 2005, 24(1):305-339

[4] Cimiano P, Hotho A, Stumme G, et al. Conceptual knowledge processing with formal concept analysis and ontologies[C]// Second International Conference on Formal Concept Analysis, ICFCA 2004. Sydney, Australia, February 2004:199-200

[5] Carpineto C, Romano G. Information retrieval through hybrid navigation of lattice representations[J]. International Journal of Human Computer Studies, 1996, 45(5):553-578

[6] Tonella P. Formal concept analysis in software engineering[C]// Proceedings of the 26th International Conference on Software Engineering. 2004

[7] 智慧来，智东杰，刘宗田. 基于概念格的概念相似度计算[J]. 计算机科学，2008，35(9)：156-157

[8] Nilsson F. An algebraic logic for concept structures[C]// Preprint of the 3rd European-Japanese Seminar on Information Modeling and Knowledge Bases, V: principles and formal techniques. 1994, 5:75

[9] Nilsson F. A concept object algebra CA[C]// Proceedings of the 2nd European-Japanese Seminar on Information Modeling and Knowledge Bases, V: principles and formal techniques. 1993

[10] Wille R. Boolean concept logic[C]// Proceedings of the 8th International Conference on Conceptual Structure, Conceptual Structures: Logical, Linguistic, and Computational Issues, ICCS 2000. 2000:317-331

[11] Kwuida L. When is a concept algebra boolean? [C]// Proceedings of the 2nd International Conference on Formal Concept Analysis, Concept Lattices, ICFCA 2004. 2004:3905-3905

[12] Wang Y. On concept algebra: A denotational mathematical structure for knowledge and software modeling[J]. International Journal of Cognitive Informatics and Natural Intelligence, 2008, 2(2):1-19

[13] Wang Y. The OAR model of neural informatics for internal knowledge representation in the brain[J]. International Journal of Cognitive Informatics and Natural Intelligence, 2007, 1(3):66-77

(下转第311页)

行结果。手写汉字自上而下依次是:思、望、永、安、秋;从左至右依次给出原始采集字、字形骨架、笔触特征三角形、美化结果。

图 7　行书书法效果美化效果

原始数据经过数据简化后,冗余数据大大减少,字形骨架与原始书写特征基本一致,在添加笔触特征三角形后,字形骨架已经初步具有书法特征,经过平滑后,得到书法效果的轮廓信息。与原字相比,书法美化后的结果中,书法效果较为明显。

4.2　隶书书法美化效果实验

本文描述方法不仅适用于一般手写体行书,对隶书字体也有较好的适应性,图 8 中给出了对一句手写隶书的古诗字体所进行的美化效果。可以看出,美化结果对于整句诗句具有良好的一致性,能够较好地体现出书写者的书写特点和意图,在细节处理方面能够较好的把握书写中的书法特点。

图 8　隶书书法效果美化效果

结束语　本文实现了一种手写汉字书法效果美化方法。该方法能够为计算机采集得到的手写字添加具有书法特征的笔触特征三角形,产生带有宽度的笔画轮廓,进而通过平滑算法,生成书法效果美化字,方法简单可靠,对行书、隶书字体均有较好地适应性;整个算法只需用到骨架汉字的坐标信息,对设备依赖性要求较低,能基本满足目前电子设备中采集手写字的需求;提出的笔触特征三角形算法能够较好地模拟毛笔运笔宽度信息,产生书法美化效果。在未来的工作中,将尝试改进数据采集模式,利用更多信息如压力和速度信息,对起笔、行笔和收笔的笔触特征三角形进行参数自适应调整,使该算法具有更好的拓展性。

参 考 文 献

[1]　王钲旋,庞云阶. 一个计算机书法系统 CCCS[J]. 计算机辅助设计与图形学学报,1991,3(1):35-40

[2]　王钲旋,庞云阶. 一个笔划填充算法及其在计算机书法中的应用[J]. 计算机辅助设计与图形学学报,1994,6(3):168-173

[3]　方逵,吴涛. 书法艺术的计算机模拟及其实现[J]. 模糊系统与数学,1996,10(1):71-74

[4]　王征,孙济洲,孙美君,等. 采用自回归模型的计算机书法的研究[J]. 工程图学学报,2006,5:38-43

[5]　Wong S T S, Leung H, Ip H H S. Model-based analysis of Chinese calligraphy images[C]// Proceedings of the 9th International Conference on Information Visualization. London, UK, 2005:221-226

[6]　Yu Jin-hui, Peng Qun-sheng. Realistic synthesis of Cao Shu of Chinese calligraphy[J]. Computer & Graphics, 2005, 29(1): 145-153

[7]　董军,徐淼,潘云鹤. 基于统计模型的书法创作模拟[J]. 计算机学报,2008,31(7):1276-1282

[8]　陈颉,朱福喜. 结合模糊支持向量机的曲线类比在书法风格仿真中的研究[J]. 中国图像图形学报,2010,15(2):340-345

[9]　夏伟平. 联机手写体汉字楷书美化方法研究[D]. 广州:华南理工大学,2010

[10]　刘丽娟. 个性化联机手写汉字隶书美化方法研究[D]. 广州:华南理工大学,2011

[11]　Douglas D, Peucker T. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature[J]. The Canadian Cartographer, 1973, 10(2): 112-122

[12]　黄襄念,程萍,杨波,等. 自然手写汉字预处理子系统[J]. 重庆大学学报:自然科学版,2000,23(4):33-37

[13]　Su Sara L, Xu Ying-qing, Shum H-Y, et al. Simulating artistic brush strokes using interval splines[C]// Proceedings of the 5th IASTED International Conference on Computer Graphics and Imaging. Kauai, Hawaii, 2002:85-90

(上接第 269 页)

[14]　Ganter B, Wille R. Formal concept analysis: mathematical foundations[M]. Springer Verlag, 1999

[15]　Nilsson J F. A logico-algebraic framework for ontologies[C]// Proceedings of the 1st International OntoQuery Workshop. 2001:32

[16]　时百胜,余泓. 概念知识表示和推理[J]. 小型微型计算机系统,2006,27(9):1618-1622

[17]　敦仁,周立柱. 概念代数—新一代数据库系统的理论[J]. 计算机研究与发展,1996,33(1):32-38

[18]　Wang Y. Cognitive informatics: towards future generation computers that think and feel[C]// Proceedings of the 5th IEEE International Conference on Cognitive Informatics, ICCI 2006. 2006:3-7