概念格中多层次的属性约简理论

葛方斌[1] 杨 林[2] 王建新[2]

(解放军理工大学指挥自动化学院 南京 210007)[1] (中国电子系统工程研究所 北京 100039)[2]

摘 要 提出了形式背景的子概念格及形式背景关于概念、子概念格(包括概念格)的属性约简概念;定义了概念、子概念格(包括概念格)的特征函数,并以此为基础提出了属性约简定理,得出了具体的属性约简方法,同时研究了不同约简下属性间的关系,进一步完善了概念格属性约简理论,也为概念格用于数据处理及知识发现提供了新的工具。

关键词 形式背景,子概念格,特征函数,属性约简

Multiplayer Theory of Attributes Reduction in Concept Lattices

GE Fang-bin[1] YANG Lin[2] WANG Jian-xin[2]

(College of Command Automation, PLA University of Science and Technology, Nanjing 210007, China)[1]

(Institute of China Electronic System Engineering, Beijing 100039, China)[2]

Abstract Put forward the concepts of proper concept lattice and attributes reduction of concepts and proper concept lattices (including concept lattices) in formal contexts. Defined the characteristic functions of concepts and proper concept lattices (including concept lattices), advanced the theorems of attributes reduction based on them, and educed the idiographic methods of attributes reductions. At the same time, studied the relations between attributes under the different attributes reduction, had perfected the theory of attributes reduction of concept lattices ulteriorly, and provided the new basis for processing data and finding knowledge by concept lattices.

Keywords Formal context, Proper concept lattice, Characteristic function, Attributes reduction

德国 Wille 教授在 20 世纪 80 年代初提出了形式概念分析[1]理论,概念格是该理论中的核心数据结构形式,它可从对象与属性构成的关系数据中构造出来,能够精确而又简洁地反映出概念间的结构层次关系。作为数据分析和知识处理的形式化工具,概念格模型已广泛应用于数据挖掘、软件工程、知识工程等众多领域[2-6]。

文献[7]针对对象属性关系表中行列的可减性提出了可约简的对象与可约简的属性这两个概念,文献[8]在概念格同构与协调概念基础上提出了概念格属性约简理论。这些约简都是立足于概念格整体的。然而,某些情况下对概念格的关注可能只是其中的单个概念或部分概念,对这种概念格局部知识的简化是概念格属性约简的重要研究课题之一。本文依据概念的相似及子概念格的同构提出了形式背景关于概念、子概念格(包括概念格)的属性约简概念,定义了概念、子概念格、概念格的特征函数,得出了概念格在各种不同层次下属性约简的具体方法。同时,研究了不同约简下属性之间的关系,进一步完善了概念格属性约简理论,也为概念格用于数据处理和知识发现提供了新的工具。

1 概念格基础

在形式概念分析[7]中,形式背景是概念格的基础,其定义为:

定义 1[7] 三元组 $K=(U,D,R)$称为一个形式背景。其中,U 是非空有限的对象集合,D 是有限的属性集合,R 是 U 和 D 间的二元关系,即 $R\subseteq U\times D$。若$(x,a)\in R$,则称对象 x 具有属性 a,记为 xRa。

在形式背景(U,D,R)中,对 $\forall x\in U$, $\forall a\in D$,若用 1 表示$(x,a)\in R$,0 表示$(x,a)\notin R$,则形式背景可表示为只有 0 和 1 的表格。

对形式背景(U,D,R),定义 U,D 幂集上的映射 $f:2^U\to 2^D$, $g:2^D\to 2^U$ 如下:

$$f(X)=\begin{cases}\{a\mid a\in D,\forall x\in X,xRa\} & X\in 2^U,X\neq\phi\\ D & X=\phi\end{cases}$$

$$g(A)=\begin{cases}\{x\mid x\in U,\forall a\in A,xRa\} & A\in 2^D,A\neq\phi\\ U & A=\phi\end{cases}$$

任意 $x\in U$,若 $f(\{x\})\neq\phi$,则称(U,D,R)是 U 上正则的;任意 $a\in D$,若 $g(\{a\})\neq\phi$,则称(U,D,R)是 D 上正则的;若(U,D,R)既是 U 上正则的又是 D 上正则的,则称(U,D,R)是正规的。本文讨论的形式背景都是正规的。

定义 2[7] 设(U,D,R)为形式背景,如果一个二元组(X,A)满足 $X\subseteq U,X\neq\phi,A\subseteq D,f(X)=A,g(A)=X$,则称$(X,A)$是一个形式概念,简称概念。其中,$X$ 称为概念的外延,A 称为概念的内涵,A 中属性称为(X,A)的特征属性。

显然,任意非空的 $X\subseteq U$,$(X,f(X))$是概念,当且仅当 $g(f(X))=X$。

设(X,A)是形式背景(U,D,R)中的概念,$Y\subseteq U$,若 $Y\subseteq X$,则称(X,A)是 Y 生成的概念。

设 $L(U,D,R)$为形式背景(U,D,R)下所有概念的集合,定义 $L(U,D,R)$上的序关系:

$$(X_1,A_1)\leqslant(X_2,A_2)\Leftrightarrow X_1\subseteq X_2$$

偏序集 $L(U,D,R)$称为概念格。其中,上、下确界定义为:

葛方斌 博士研究生,主要研究方向为计算机网络、信息安全、数据挖掘。

$(X_1,A_1)\wedge(X_2,A_2)=(X_1\cap X_2,f(g(A_1\cup A_2)))$

$(X_1,A_1)\vee(X_2,A_2)=(g(f(X_1\cup X_2)),A_1\cap A_2)$

$L(U,D,R)$是一个完备格[7]。

设 $X\subseteq U$,若 $L_X(U,D,R)$是 X 生成的概念的集合,即

$L_X(U,D,R)=\{(Y,A)\mid X\subseteq Y,(Y,A)\in L(U,D,R)\}$

则对于如上定义的偏序关系及上、下确界,$L_X(U,D,R)$仍是一个完备概念格,称其为由 X 生成的子概念格。显然,$L_\varphi(U,D,R)=L(U,D,R)$。$\{x\}$生成的子概念格也称为对象 x 生成的子概念格,简记为 $L_x(U,D,R)$。易见,$L(U,D,R)=\bigcup_{x\in U}L_x(U,D,R)$。

例 1　对表 1 所示形式背景(U,D,R),给出其上的概念格及各对象生成的子概念格。

表 1　形式背景

	a	b	c	d	e
1	1	1	0	0	1
2	0	0	0	1	0
3	1	0	1	1	1
4	1	1	0	0	1

(U,D,R)上的概念格及各对象生成的子概念格为:

$L(U,D,R)=\{(3,acde),(14,abe),(23,d),(134,ae),(1234,\phi)\}$

$L_1(U,D,R)=\{(14,abe),(134,ae),(1234,\phi)\}$

$L_2(U,D,R)=\{(23,d),(1234,\phi)\}$

$L_3(U,D,R)=\{(3,acde),(23,d),(134,ae),(1234,\phi)\}$

$L_4(U,D,R)=\{(14,abe),(134,ae),(1234,\phi)\}$

定理 1　在形式背景(U,D,R)中,设 $X\subseteq U,X\neq\phi$,则$(g(f(X)),f(X))$是 X 生成的最小概念。

证明:首先证明$(g(f(X)),f(X))$是(U,D,R)中 X 生成的概念。由 $g(f(X))\supseteq X\neq\phi$ 得,$g(f(X))\neq\phi$,且 $f(g(f(X)))\subseteq f(X)$,又 $f(X)\subseteq f(g(f(X)))$,从而 $f(g(f(X)))=f(X)$,可见,$(g(f(X)),f(X))$是(U,D,R)中 X 生成的概念。其次,若(Y,A)是一概念,且 $X\subseteq Y$,则 $f(X)\supseteq f(Y)$,从而 $g(f(X))\subseteq g(f(Y))=Y$,故$(g(f(X)),f(X))$是 X 生成的最小概念。

对形式背景(U,D,R),由于$U\subseteq g(f(U))$,且 $g(f(U))\subseteq U$,可见,$(U,f(U))$一定是 $L(U,D,R)$中的概念,称其为平凡概念。任何形式背景中都存在平凡概念,只含有平凡概念的概念格称为平凡概念格。

定理 2　设(U,D,R)是形式背景,$L(U,D,R)$是非平凡概念格,当且仅当存在 $x,y\in U,a\in D$,使 $a\in f(\{x\}),a\notin f(\{y\})$。

证明:由定理 1 知 $L(U,D,R)$是非平凡概念格,当且仅当 $\exists x\in U,g(f(\{x\}))\neq U$。

而 $g(f(\{x\}))\neq U\Leftrightarrow\exists y\in Y,y\notin g(f(\{x\}))$。

$y\notin g(f(\{x\}))\Leftrightarrow\exists a\in D,a\in f(\{x\}),a\notin f(\{y\})$。

故 $L(U,D,R)$是非平凡概念格,当且仅当存在 $x,y\in U,a\in D$,使 $a\in f(\{x\}),a\notin f(\{y\})$。

定理表明,正规的形式背景表只有当其中含有“0”时,其上的概念格才是非平凡的。

定义 3　设(X_1,A_1)、(X_2,A_2)分别是 $L(U,D_1,R_1)$、$L(U,D_2,R_2)$中的概念,若 $X_1=X_2$,则称(X_1,A_1)与(X_2,A_2)相似,记为$(X_1,A_1)\triangleq(X_2,A_2)$。若对于任意$(X,A)\in L(U,D_2,R_2)$,都存在$(X',A')\in L(U,D_1,R_1)$,使$(X',A')\triangleq(X,A)$,则称 $L(U,D_1,R_1)$细于 $L(U,D_2,R_2)$,记为

$L(U,D_1,R_1)\leqslant L(U,D_2,R_2)$

若 $L(U,D_1,R_1)\leqslant L(U,D_2,R_2)$,且 $L(U,D_2,R_2)\leqslant L(U,D_1,R_1)$,则称 $L(U,D_1,R_1)$与 $L(U,D_2,R_2)$同构[8],记为

$L(U,D_1,R_1)\cong L(U,D_2,R_2)$

在形式背景(U,D,R)下,对任意 $A\subseteq D$,定义 $R_A=R\cap(U\times A)$,则(U,A,R_A)也是一个形式背景。

易见,概念格具有性质:

(1)与对象集 U 相关的概念格集合上的同构关系是等价关系;

(2)若 $A\subseteq D$,则 $L(U,D,R)\leqslant L(U,A,R_A)$[8];

(3)若 $L(U,D_1,R_1)\leqslant L(U,D_2,R_2)\leqslant L(U,D_3,R_3)$,且 $L(U,D_1,R_1)\cong L(U,D_3,R_3)$ 则

$L(U,D_1,R_1)\cong L(U,D_2,R_2)\cong L(U,D_3,R_3)$。

2　概念格中的属性约简

在形式背景(U,D,R)下,D 中的属性对概念外延的确定往往并非都是必要的,据此可建立属性约简概念,它是概念的最简特征刻画。

定义 4　设(X,A)是形式背景(U,D,R)下的概念,若 $a\in D$,若$(U,D-\{a\},R_{D-\{a\}})$中不存在与(X,A)相似的概念,则称 a 是(U,D,R)关于(X,A)的不可省属性;若$(U,D-\{a\},R_{D-\{a\}})$存在与(X,A)相似的概念,且存在 $B\subseteq D,a\in B$,(U,B,R_B)中存在与(X,A)相似的概念,但$(U,B-\{a\},R_{B-\{a\}})$中不存在与(X,A)相似的概念,则称 a 是(U,D,R)关于(X,A)的相对可省属性;否则,称 a 是(U,D,R)关于(X,A)的绝对可省属性。(U,D,R)关于(X,A)的相对、绝对可省属性统称为(U,D,R)关于(X,A)的可省属性。

定义 5　设(X,A)是形式背景(U,D,R)下的概念,$B\subseteq D$,若(U,B,R_B)中存在(X,A)的相似概念,且 $\forall a\in B$,$(U,B-\{a\},R_{B-\{a\}})$中不存在(X,A)的相似概念,则称 A 是(U,D,R)关于概念(X,A)的一个约简。若(U,D,R)关于概念(X,A)的所有约简的交集非空,则称该交集为(U,D,R)关于概念(X,A)的约简的核。

定理 3　(U,D,R)关于概念(X,A)的约简存在,当且仅当(X,A)是非平凡概念。

证明:充分性。考虑 A 的所有一元子集。若存在一元子集子$\{a\}$,使 $g(\{a\})=X$,则在$(U,\{a\},R_{\{a\}})$中,由于 $f(X)=f(g(\{a\}))\supseteq\{a\}$,且 $f(X)\subseteq\{a\}$,故$(X,\{a\})$是$(U,\{a\},R_{\{a\}})$中(X,A)的相似概念,又由于(U,ϕ,R_ϕ)中只有平凡概念,从而(U,ϕ,R_ϕ)中无(X,A)的相似概念,所以,$\{a\}$是(U,D,R)关于概念(X,A)的一个约简,于是结论成立。否则,考虑 A 的所有二元子集,若存在二元子集 B,使 $g(B)=X$,对 $\forall a\in B$,设(Y,C)是$(U,B-\{a\},R_{B-\{a\}})$上的任意概念,由 $C\subseteq B-\{a\}$ 知 $Y=g(C)\supseteq g(B-\{a\})\supseteq g(B)=X$,又 $B-\{a\}$是 A 的所有一元子集说明 $g(B-\{a\})\neq X$,故 $Y\neq X$,从而 $(U,B-\{a\},R_{B-\{a\}})$中无(X,A)的相似概念,这说明 B 是(U,D,R)关于概念(X,A)的一个约简,否则,考虑 A 的三元子集,如此进行下去,由于 A 有限,且 $g(A)=X$,因而在有限步内一定可找到(U,D,R)关于概念(X,A)的一个约简,故(U,D,R)关于概念(X,A)的约简存在。

必要性。设 B 是(U,D,R)关于概念(X,A)的一个约简,则 $\forall a\in B$,$(U,B-\{a\},R_{B-\{a\}})$中所有概念与(X,A)不同相似,由于 $L(U,B-\{a\},R_{B-\{a\}})$包含平凡概念,故(X,A)是非

平凡概念。

下面的定理反映了属性关于概念的可省性和关于概念的属性约简之间的联系。

定理 4 设(U,D,R)是一形式背景

(1)a是(U,D,R)关于(X,A)的不可省属性，当且仅当对任意(U,D,R)关于(X,A)的约简B，都有$a\in B$(即：a是(U,D,R)关于(X,A)的约简的核中的元素)。

(2)a是(U,D,R)关于(X,A)的相对可省属性，当且仅当存在(U,D,R)关于(X,A)的约简B,C，使$a\in B,a\notin C$。

(3)a是(U,D,R)关于(X,A)的绝对可省属性，当且仅当对任意(U,D,R)关于(X,A)的约简B，都有$a\notin B$。

(2)可由(1)、(3)推得，下面证明(1)、(3)。

证明：(1)充分性。假设a是(U,D,R)关于(X,A)的可省属性，则$(U,D-\{a\},R_{D-\{a\}})$存在与(X,A)相似的概念(X,A')。由(U,D,R)关于(X,A)的约简存在知，$(U,D-\{a\},R_{D-\{a\}})$关于(X,A')的约简也存在，设C是其中一个约简，C显然也是(U,D,R)关于(X,A)的约简，而由$C\subseteq D-\{a\}$知$a\notin C$，这与前提矛盾，故a是(U,D,R)关于(X,A)的不可省属性。

必要性。若a是(U,D,R)关于(X,A)的不可省属性，则显然有$X\neq U$，从而(U,D,R)关于(X,A)的约简存在。设B是其中任意一个约简，则$a\in B$，否则，若$a\notin B$，设(X,A')是(U,B,R_B)中(X,A)的相似概念，由于$B=B-\{a\}\subseteq D-\{a\}$，所以，$(U,D-\{a\},R_{D-\{a\}})$中存在$(X,A')$的相似概念$(X,A'')$，显然，$(X,A'')\triangleq(X,A)$，这与$a$是$(U,D,R)$关于$(X,A)$的不可省属性矛盾。

(3)充分性。即证任意$B\subseteq D,a\in B$，若(U,B,R_B)中存在(X,A)的相似概念(X,A')，则$(U,B-\{a\},R_{B-\{a\}})$中也存在(X,A)的相似概念。设(X,A')是(U,B,R_B)中(X,A)的相似概念，由(U,D,R)关于(X,A)的约简存在知，(U,B,R_B)关于(X,A')的约简存在。不妨设C是其中一个约简，(U,C,R_C)中(X,A')的相似概念为(X,A'')，C显然也是(U,D,R)关于(X,A)的一个约简，从而$a\notin C$，于是$C=C-\{a\}\subseteq B-\{a\}$，由此知，$(U,B-\{a\},R_{B-\{a\}})$中存在$(X,A'')$的相似概念，这些概念当然也是$(X,A)$的相似概念。故$a$是$(U,D,R)$关于$(X,A)$的绝对可省属性。

必要性。反设存在(U,D,R)关于(X,A)的约简B，使$a\in B$。由约简的定义知，(U,B,R_B)中存在(X,A)的相似概念，但$(U,B-\{a\},R_{B-\{a\}})$中不存在(X,A)的相似概念，这与a是(U,D,R)关于(X,A)的绝对可省属性矛盾，故，对(U,D,R)关于(X,A)的任意约简B，都有$a\notin B$。

定理 5 若B是(U,D,R)关于概念(X,A)的一个约简，则$B\subseteq A$。

证明：设(X',A')是(U,B,R_B)中(X,A)的相似概念，则显然有$A'\subseteq B,A'\subseteq A$。假设存在$a\in B,a\notin A$，则$A'=A'-\{a\}\subseteq B-\{a\}$，这说明$(X',A')$也是$(U,B-\{a\},R_{B-\{a\}})$中的概念，从而，$(U,B-\{a\},R_{B-\{a\}})$中存在$(X,A)$的相似概念，这与$B$是$(U,D,R)$关于概念$(X,A)$的约简矛盾，故$B\subseteq A$。

此定理说明，形式背景关于概念的约简与概念的非特征属性无关，它只由概念的特征属性决定。

推论 若$a\in D-A$，则a是(U,D,R)关于概念(X,A)的绝对可省属性。

形式背景除了有关于概念的属性约简外还有关于子概念格(包括概念格)的属性约简。

定义 6 设(U,D,R)是一形式背景，$S\subseteq U,a\in D$，若$L_S(U,D-\{a\},R_{D-\{a\}})$与$L_S(U,D,R)$不同构，则称$a$是$(U,D,R)$关于$L_S(U,D,R)$的不可省属性；若$L_S(U,D-\{a\},R_{D-\{a\}})\cong L_S(U,D,R)$，且$\exists A\subseteq D,a\in A,L_S(U,A,R_A)\cong L_S(U,D,R)$，但$L_S(U,D,R)$与$L_S(U,A-\{a\},R_{A-\{a\}})$不同构，则称$a$是$(U,D,R)$关于$L_S(U,D,R)$的相对可省属性；否则称$a$是$(U,D,R)$关于$L_S(U,D,R)$的绝对可省属性，$(U,D,R)$关于$L_S(U,D,R)$的相对、绝对可省属性统称为$(U,D,R)$关于$L_S(U,D,R)$的可省属性。

定理 6 设(U,D,R)是一形式背景，a是(U,D,R)关于$L(U,D,R)$的可省属性，当且仅当$\forall x\in U$，a是(U,D,R)关于$L_x(U,D,R)$的可省属性。

证明：充分性。反设a是(U,D,R)关于$L(U,D,R)$的不可省属性，则$L(U,D-\{a\},R_{D-\{a\}})$与$L(U,D,R)$不同构，又由于$L(U,D,R)\leqslant L(U,D-\{a\},R_{D-\{a\}})$，于是存在$(X,A)\in L(U,D,R)$，使$\forall(X',A')\in L(U,D-\{a\},R_{D-\{a\}})$，都有$X'\neq X$。设$x\in X$，则$(X,A)\in L_x(U,D,R)$，而对于$\forall(X'',A'')\in L_x(U,D-\{a\},R_{D-\{a\}})$，由于$L_x(U,D-\{a\},R_{D-\{a\}})\subseteq L(U,D-\{a\},R_{D-\{a\}})$，因而有$X''\neq X$。可见，$L_x(U,D-\{a\},R_{D-\{a\}})$与$L_x(U,D,R)$不同构。从而$a$是$(U,D,R)$关于$L_x(U,D,R)$的不可省属性，这与题设矛盾，故结论成立。

必要性。设$(X,A)\in L_x(U,D,R)$，则$(X,A)\in L(U,D,R)$，由于$L(U,D-\{a\},R_{D-\{a\}})\cong L(U,D,R)$，因此存在$(X,B)\in L(U,D-\{a\},R_{D-\{a\}})$，显然$(X,B)\in L_x(U,D-\{a\},R_{D-\{a\}})$，于是$L_x(U,D-\{a\},R_{D-\{a\}})\leqslant L_x(U,D,R)$，同理可证$L_x(U,D,R)\leqslant L_x(U,D-\{a\},R_{D-\{a\}})$，故$a$是$(U,D,R)$关于$L_x(U,D,R)$的可省属性。

推论 设(U,D,R)是一形式背景，a是(U,D,R)关于$L(U,D,R)$的不可省属性，当且仅当$\exists x\in U$，a是(U,D,R)关于$L_x(U,D,R)$的不可省属性。

定理及推论反映了属性关于概念格的可省性和关于子概念格的可省性之间的关系。实际上，属性关于概念、概念格的可省性之间以及关于概念、子概念格的可省性之间也都存在类似关系。

定义 7 设(U,D,R)是一形式背景，$S\subseteq U,A\subseteq D$，若$L_S(U,A,R_A)\cong L_S(U,D,R)$，且$\forall a\in A,L_S(U,A-\{a\},R_{A-\{a\}})$与$L_S(U,D,R)$不同构，则称$A$是$(U,D,R)$关于$L_S(U,D,R)$的一个约简。若$(U,D,R)$关于$L_S(U,D,R)$的所有约简的交集非空，则称该交集为$(U,D,R)$关于$L_S(U,D,R)$的约简的核。

取不同的S可得到关于各种层次的概念格的属性约简。$S=\phi$时就是文献[8]中的约简定义。这些约简体现了不同层次概念格的特征，这为形式背景多层次的分析奠定了理论基础。

形式背景中属性关于子概念格的可省性与关于子概念格的约简之间存在类似定理4的结论。由此结论和定理6的推论立即可知，形式背景(U,D,R)关于$L(U,D,R)$的约简的核是所有关于U中对象生成的子概念格的约简的核的并集。

定理 7 设(U,D,R)是一形式背景，$S\subseteq U$，则(U,D,R)关于$L_S(U,D,R)$的约简存在，当且仅当$L_S(U,D,R)$是非平凡概念格。

证明：充分性。由于$L_S(U,\phi,R_\phi)$是平凡概念格，因此$L_S(U,\phi,R_\phi)$与$L_S(U,D,R)$不同构。考虑D的所有一元子集，若存在$a\in D$，使$L_S(U,\{a\},R_{\{a\}})\cong L_S(U,D,R)$，则$\{a\}$显然

是(U,D,R)关于$L_S(U,D,R)$的一个约简，否则考虑D的所有二元子集，若存在二元子集$A\subseteq D$，使$L_S(U,A,R_A)\cong L_S(U,D,R)$，则$A$是$(U,D,R)$关于$L_S(U,D,R)$的一个约简，否则考虑$D$的所有三元子集，如此进行下去。由于$A$是有限集合，且$L_S(U,D,R)$与自身同构，故在有限步内必可找到$(U,D,R)$关于$L_S(U,D,R)$的一个约简，可见，$(U,D,R)$关于$L_S(U,D,R)$的约简存在。

必要性。若(U,D,R)关于$L_S(U,D,R)$的约简存在，不妨设A是(U,D,R)关于$L_S(U,D,R)$的一个约简。由约简的定义知，存在$B\subseteq A$，使，$L_S(U,B,R_B)$与$L_S(U,D,R)$不同构，又由于$L_S(U,D,R)\leqslant L_S(U,B,R_B)$，故$L_S(U,D,R)$中一定包含非平凡概念，从而，$L_S(U,D,R)$是非平凡概念格。

推论 形式背景(U,D,R)关于$L(U,D,R)$的约简存在，当且仅当存在$x\in U$，(U,D,R)关于$L_x(U,D,R)$的约简存在。

3 属性约简方法

形式背景的属性约简中，关于概念的属性约简是最基本的，它是关于子概念格（包括概念格）的属性约简的基础。概念完全由它们的特征属性所决定，针对这些特征属性我们首先引入形式背景中概念的特征函数，并由此得出确定约简的方法，然后在此基础上给出子概念格（包括概念格）的特征函数以及确定关于子概念格（包括概念格）的属性约简的方法。

设(X,A)是形式背景(U,D,R)中的概念，对每个$a\in D$，指定一个布尔变量，仍用"a"不加区分地表示。

定义 8 若 $\Delta_{(X,A)}=\begin{cases}\bigwedge\limits_{x\in M_A}(\bigvee\limits_{\substack{a\notin f(\{x\})\\a\in A}}a) & M_A\neq\phi\\ \bigvee\limits_{a\in A}a & M_A=\phi,X\neq U\end{cases}$，则称$\Delta_{(X,A)}$是$(U,D,R)$关于概念$(X,A)$的特征函数。其中，$M_A=\bigcup\limits_{a\in A}g(\{a\})-X$。

$M_A\neq\phi$时，显然有$X\neq U$，可见，上述特征函数是针对非平凡概念的。

(X,A)的特征函数是一个布尔表达式，它可化为析取范式，即由一些变量合取式的析取所组成的布尔表达式。

一个析取范式，若其各析取支互不相同，且每个析取支都包含于任何一个与它相等的析取范式的某个析取支，则称此析取范式为极小析取范式。由布尔代数理论知，任何不含变量非运算的布尔表达式都有唯一的极小析取范式。

定理 8 B是形式背景(U,D,R)关于概念(X,A)一个约简，当且仅当B是$\Delta_{(X,A)}$的极小析取范式的一个析取支。

证明：充分性。设B是$\Delta_{(X,A)}$的极小析取范式的一个析取支。若$\Delta_{(X,A)}=\bigvee\limits_{a\in A}a$，则$B$是$A$的单元子集。此时$M_A=\phi$，$X\neq U$，不妨设$B=\{b\}$，$b\in A$，由$M_A=\phi$知，$g(\{b\})=X$，从而$(X,\{b\})$是$(U,\{b\},R_{\{b\}})$中$(X,A)$的相似概念；另外，$X\neq U$又说明，$(U,\varphi,R_\varphi)$中无$(X,A)$的相似概念，可见，$B$是$(U,D,R)$关于概念$(X,A)$一个约简；若$\Delta_{(X,A)}=\bigwedge\limits_{x\in M_A}(\bigvee\limits_{\substack{a\notin f(\{x\})\\a\in A}}a)$，此时，$M_A\neq\phi$，由$\phi\neq M_A\subseteq U-X$知，$X\neq U$。为了证明$B$是$(U,D,R)$关于概念$(X,A)$一个约简，首先证明，$(U,B,R_B)$中存在$(X,A)$的相似概念。当$g(B)=X$时，显然有$(X,B)$是$(U,B,R_B)$中$(X,A)$的相似概念，所以只要证$g(B)=X$，也即证$\forall x\in M_A$，$x\notin g(B)$。对$\forall x\in M_A$，由于$B$是$\bigwedge\limits_{x\in M_A}(\bigvee\limits_{\substack{a\notin f(\{x\})\\a\in A}}a)$的极小析取范式的一个析取支，于是，$B$必包含$\bigwedge\limits_{x\in M_A}(\bigvee\limits_{\substack{a\notin f(\{x\})\\a\in A}}a)$的合取支$\bigvee\limits_{\substack{a\notin f(\{x\})\\a\in A}}a$的至少一个变量，即存在$a\in B$，$a\notin f(\{x\})$，从而$x\notin g(B)$。其次证明，$\forall a\in B$，$(U,B-\{a\},R_{B-\{a\}})$中不存在$(X,A)$的相似概念。若$\Delta_{(X,A)}=\bigvee\limits_{a\in A}a$，结论显然成立。下面证明$\Delta_{(X,A)}=\bigwedge\limits_{x\in M_A}(\bigvee\limits_{\substack{a\notin f(\{x\})\\a\in A}}a)$的情形。首先有，$\Delta_{(X,A)}$的每个合取支至少包含$B$中一个元素，否则，若存在$\Delta_{(X,A)}$的某个合取支不含$B$中元素，则当取$B$中每个变量的值为1，而$A-B$中每个变量的值为0时，$\Delta_{(X,A)}$就有一值为0的合取支，而$\Delta_{(X,A)}$的极小析取范式有一值为1的析取支，从而，$\Delta_{(X,A)}=0$；同时，$\Delta_{(X,A)}=1$，这显然是一对矛盾。反设$\exists a\in B$，$(U,B-\{a\},R_{B-\{a\}})$中存在$(X,A)$的相似概念。在此假设下先来证明，$\Delta_{(X,A)}$的每个合取支至少包含$B-\{a\}$中一个元素。反设存在$\Delta_{(X,A)}$的某个合取支不包含$B-\{a\}$中元素，由于$\Delta_{(X,A)}$的每个合取支至少包含$B$中一个元素，故存在$\Delta_{(X,A)}$的某个合取支包含$a$但不包含$B-\{a\}$中元素，这说明存在某个$x\in M_A$，使$B-\{a\}\subseteq f(\{x\})$，从而$x\in g(B-\{a\})$，而$x\in M_A$又说明$x\notin X$，可见，$g(B-\{a\})\neq X$，而在$(U,B-\{a\},R_{B-\{a\}})$中，$f(X)=B-\{a\}$，这说明$(U,B-\{a\},R_{B-\{a\}})$中不存在$(X,A)$的相似概念，这与题设矛盾。$\Delta_{(X,A)}$的每个合取支至少包含$B-\{a\}$中一个元素说明，$\Delta_{(X,A)}$的极小析取范式必有某析取支包含于$B-\{a\}$，这与$B$是$\Delta_{(X,A)}$的极小析取范式的析取支矛盾。综上知，$B$是$(U,D,R)$关于概念$(X,A)$一个约简。

必要性。若$M_A=\phi$，则$\forall a\in A$，$g(a)=X$，此时，(U,D,R)关于概念(X,A)约简B显然是A的一元子集，它当然是$\Delta_{(X,A)}=\bigvee\limits_{a\in A}a$的极小析取范式的一个分支；若$M_A\neq\phi$，由$B$是$(U,D,R)$关于概念$(X,A)$一个约简知，$g(B)=X$，于是，$\forall x\in M_A(x\notin X)$，都存在$a\in B$，使$a\notin f(\{x\})$。这说明，$\Delta_{(X,A)}$的任一合取支$\bigvee\limits_{\substack{a\notin f(\{x\})\\a\in A}}a$都包含$B$中至少一个元素，从而，$\Delta_{(X,A)}$的极小析取范式必有某析取支$C$包含于$B$。实际上，$B=C$。若非如此，则$C$是$B$的真子集。由$B$是$(U,D,R)$关于概念$(X,A)$一个约简知，$(U,C,R_C)$中不存在$(X,A)$的相似概念，但由充分性知，$C$也是$(U,D,R)$关于概念$(X,A)$一个约简，从而$(U,C,R_C)$中存在$(X,A)$的相似概念，如此形成矛盾，故$B$是$\Delta_{(X,A)}$的极小析取范式的一个析取支。

由概念的特征函数可进一步定义子概念格（包括概念格）的特征函数，它可用于生成子概念格（包括概念格）的属性约简。

定义 9 设(U,D,R)是一形式背景，$S\subseteq U$，若$\Delta_{L_S(U,D,R)}=\bigwedge\limits_{\substack{(X,A)\in L_S(U,D,R)\\X\neq U}}\Delta_{(X,A)}$，则称$\Delta_{L_S(U,D,R)}$是$(U,D,R)$关于$L_S(U,D,R)$的特征函数。其中，$\Delta_{(X,A)}$是概念$(X,A)$的特征函数。

类似于定理8，关于子概念格（包括概念格）有下面的属性约简定理。

定理 9 设(U,D,R)是一形式背景，$S\subseteq U$，B是(U,D,R)关于$L_S(U,D,R)$一个约简，当且仅当B是$\Delta_{L_S(U,D,R)}$的极小析取范式的一个析取支。

证明：充分性。首先，$\Delta_{L_S(U,D,R)}$的存在说明$L_S(U,D,R)$中存在某概念，(U,D,R)关于该概念的约简存在，从而该概念是非平凡的，$L_S(U,D,R)$是非平凡概念格。于是，(U,D,R)关于$L_S(U,D,R)$约简存在。其次，由定理8知，$\Delta_{(X,A)}$等价于(U,D,R)关于(X,A)的所有约简的析取，从而，B包含(U,D,R)关于$L_S(U,D,R)$中每个非平凡概念的至少一个约简。于是，对$L_S(U,D,R)$中每个非平凡概念(X,A)，存在

$C\subseteq B$，$L_S(U,C,R_C)$中存在(X,A)的相似概念，从而，$L_S(U,B,R_B)$中存在(X,A)的相似概念。这样就有，$L_S(U,B,R_B)\leqslant L_S(U,D,R)$，又$L_S(U,D,R)\leqslant L_S(U,B,R_B)$，故$(U,B,R_B)\cong L(U,D,R)$。另外，对任意$a\in B$，必存在$L_S(U,D,R)$中概念$(X,A)$，$B-\{a\}$不包含$(U,D,R)$关于$(X,A)$的任何约简（否则，$\Delta_{L_S(U,D,R)}$的极小析取范式必有某析取支包含于$B-\{a\}$，与$B$是$\Delta_{L(U,D,R)}$的极小析取范式的一个合取支矛盾）。于是，$L_S(U,B-\{a\},R_{B-\{a\}})$中不存在$(X,A)$的相似概念，从而$L_S(U,B-\{a\},R_{B-\{a\}})$与$L_S(U,D,R)$不同构。至此可见，$B$是$(U,D,R)$关于$L_S(U,D,R)$的一个约简。

必要性。若B是(U,D,R)关于$L_S(U,D,R)$的一个约简，则对$L_S(U,D,R)$中任意非平凡概念(X,A)，$L_S(U,B,R_B)$中存在(X,A)的相似概念(X,A')，(U,B,R_B)关于(X,A')的约简显然存在并且包含于B，这样的约简显然也是(U,D,R)关于(X,A)的约简。由此可见，对$\Delta_{L(U,D,R)}$中每个$\Delta_{(X,A)}$，B必包含它的极小析取范式的一个析取支，于是，$\Delta_{L_S(U,D,R)}$的极小析取范式必有某析取支C包含于B。而由充分性知，C也是(U,D,R)关于$L_S(U,D,R)$一个约简，从而$B=C$，否则，由B是(U,D,R)关于$L_S(U,D,R)$一个约简，C是B的真子集知，C不是(U,D,R)关于$L_S(U,D,R)$的约简，与已证充分性矛盾。可见，B是$\Delta_{L_S(U,D,R)}$的极小析取范式的一个析取支。

4 约简举例

定理8,9说明，形式背景关于概念、子概念格（包括概念格）的属性约简都能通过对相应特征函数进行布尔运算来确定。

例2 计算表1所给形式背景(U,D,R)关于概念、每个对象生成的子概念格以及概念格的属性约简。

每个非平凡概念的特征函数为：

$\Delta_{(3,acde)}=(c\vee d)(a\vee c\vee e)(c\vee d)=c\vee ad\vee de$

$\Delta_{(14,abe)}=b$，$\Delta_{(23,d)}=d$，$\Delta_{(134,ae)}=a\vee e$

所以，(U,D,R)关于$(3,acde)$的约简为：c,ad,de；关于$(14,ade)$的约简为：b；关于$(23,d)$的约简为：d；关于$(134,ae)$的约简为：a,e。

(U,D,R)关于$L_1(U,D,R)$，$L_2(U,D,R)$，$L_3(U,D,R)$，$L_4(U,D,R)$的特征函数分别为：$\Delta_{L_1(U,D,R)}=b(a\vee e)=ab\vee be$

$\Delta_{L_2(U,D,R)}=d$

$\Delta_{L_3(U,D,R)}=(c\vee ad\vee de)d(a\vee e)=ad\vee de$

$\Delta_{L_4(U,D,R)}=b(a\vee e)=ab\vee be$

所以，(U,D,R)关于$L_1(U,D,R)$，$L_2(U,D,R)$，$L_3(U,D,R)$，$L_4(U,D,R)$的约简分别为：$ab,be;d;ad,de;ab,be$。

(U,D,R)关于$L(U,D,R)$的特征函数为：

$\Delta_{L(U,D,R)}=(ab\vee be)d(ad\vee de)(ab\vee be)=abd\vee bde$

所以，(U,D,R)关于$L(U,D,R)$的约简为：abd,bde。

从以上结果可发现，(U,D,R)关于$L(U,D,R)$的约简的核$\{b,d\}$是关于所有概念约简的核的并集，同时，(U,D,R)关于$L(U,D,R)$的可省属性是(U,D,R)关于某个对象生成的子概念格的可省属性，与前面讨论的属性关系是一致的。

需要指出的是，形式背景关于子概念格的属性约简概念可以推广到任何一个含有非平凡概念的概念集合上，并且也有相应的特征函数及约简定理。文献[8]中用辨识函数求关于概念格属性约简的方法对于求子概念格属性约简也是适用的，但对于一般概念集合上的属性约简却是不适用的。可以证明的是，一个概念集合，只有当其中每个概念的所有上层概念（外延更广的概念）都在该集合中时，辨识函数求属性约简的方法才是适用的。同时辨识函数方法也无法处理关于概念的属性约简问题。这两点说明，特征函数方法比辨识函数方法更具有一般性。

结束语 形式背景关于概念、子概念格（包括概念格）的属性约简是保证相关概念外延不变的最简属性描述。各种不同层次的属性约简并不改变概念或概念集合表现出的知识，但知识的表达却更为简明，这为形式背景中发现新知识创造了条件。同时，多种层次的属性约简也进一步丰富和完善了概念格属性约简理论。

参考文献

[1] Wille R. Restructuring lattice theory: An approach based on hierarchies of concepts[C]// R Ivan Rival, ed. Ordered Sets. Dordrecht-Boston: Reidel, 1982: 445-470

[2] Godin R, Mineau G W, Missaoui R. incremental structuring of knowledge bases [C] // Proc. International Symposium on Knowledge Retrieval, Use, and Storage for Efficiency (KRUSE'95). Santa Cruz, 1995: 179-193

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

[4] Oosthuizen G D. The Application of Concept Lattice to Machine Learning[R]. Technical Report. South Africa: University of Pretoria, 1996

[5] Siff M, Reps T. Identifying modules via concept analysis[C]// Harrold M J, Visaggio G, eds. International Conference on Software Maintenance. Bari, Italy. Washington, DC: IEEE Computer Society, 1997: 170-179

[6] Ho T B. Incremental conceptual clustering in the framework of Galois lattice[C]// Lu H, Motoda H, Liu H, eds. KDD: Techniques and Applications. Singapore: World Scientific, 1997: 49-64

[7] Gander B, Wille R. Formal Concept Analysis[M]. Mathematical Foundations. Berlin: Springer, 1999

[8] 张文修，魏玲，祁建军. 概念格属性约简理论与方法[J]. 中国科学E辑：信息科学，2005，36(6)：628-639