基于不可约元的概念格属性特征识别方法*)

李鸿儒[1,2]　魏　平[2]
（西安交通大学理学院信息与系统科学研究所　西安 710049）[1]
（烟台大学数学与信息科学系　烟台 264005）[2]

摘　要　形式概念分析是数据挖掘与知识获取的一种有效的数学工具。概念格作为形式概念分析的一个核心内容，刻画了对象和属性之间的内在联系。基于概念格中外延、内涵闭系统中不可约元的性质，本文研究了概念格理论中的属性分类问题，给出了一种属性特征的识别方法。这种方法不仅揭示了交不可约元与属性特征的关系，同时为知识约简提供了一种新的途径。

关键词　形式背景，概念格，闭系统，不可约元，知识约简

A Recognition Method of Attributes in Concept Lattices Based on Irreducible Elements

LI Hong-Ru[1,2]　WEI Ping[2]
(Faculty of Science, Institute for Information and System Sciences, Xi'an Jiaotong University, Xi'an 710049)[1]
(Department of Mathematics and Information Sciences, Yantai University, Yan'tai, Shandong 264005)[2]

Abstract　Formal concept analysis is an effective mathematic tool for data mining and knowledge acquisition. Concept lattice is a key issue of formal concept analysis, and characterizes the essential relation of objects and attributes. Based on the properties of irreducible elements in the extent and intent closure systems, this paper investigates the problem of attribute classification in concept lattices and presents a recognition method of the characteristic of attributes. The method proposed not only reveals the relationships between the irreducible elements and the characteristics of attributes, but also provides a new way for knowledge reduction.

Keywords　Formal context, Concept lattice, Closure system, Irreducible element, Knowledge reduction

1　引言

R. Wille[1,2]提出的形式概念分析是以序理论和完备格理论为基础，依据数据库中提供的基本信息建立起的一种刻画对象与属性之间关系的数学结构。这种概念及概念层次的数学化使形式概念分析成为数据挖掘与知识发现的重要方法，并广泛应用于许多领域[3~8]。

形式背景与概念格是形式概念分析的两个核心内容。形式背景是由对象、属性以及对象与属性之间关系组成的数据库，是人们认识事物、获取知识的基础；概念格是根据形式背景中信息建立起的对象与属性之间满足一定条件的数据结构，这一结构从本质上揭示了对象与属性之间的关系，为描述、刻画数据提供了一种新的数学方法。在形式背景中，每个属性对概念格的刻画作用不尽相同，有的属性是绝对必要的。也就是说，去掉这种属性，将改变概念格的结构。而有些属性是多余的，即在保持概念格结构不变的前提下可以去掉这些属性。因此，对于给定的数据库，能快速、准确地识别出每个属性在概念格中的作用，对知识发现和知识约简具有重要的意义。基于格理论和形式背景中对象与属性之间的二元关系，本文研究了概念格中属性的分类问题。利用完备格中不可约元的性质，给出一种能快速识别属性特征的分类方法。实践的结果说明我们这里提出的方法是简单的、有效的。

2　形式背景与概念格

设U,A是两个非空有限集，U中的元素称为对象，A中的元素称为属性。$I\subseteq U\times A$是对象和属性之间的一个二元关系。三元组(U,A,I)被称为一个形式背景，记为$K=(U,A,I)$。

对任意的$(x,a)\in U\times A$，如果$(x,a)\in I$(或xIa)，我们称对象x具有属性a。

例如表1给出了一个形式背景$K_1=(U,A,I)$，其中，$U=\{1,2,3,4\}$，$A=\{a,b,c,d,e\}$。

表1　形式背景 K_1

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

从表1我们可以看出，对象1具有属性a,b,d,e。而拥有属性a的对象是1,2,4,等等。

对于任意的$X\subseteq U$,$B\subseteq A$记

$$X^{*}=\{a\in A\mid xIa,\ \forall x\in X\} \tag{1}$$

$$B^{*}=\{x\in U\mid xIa,\ \forall a\in B\} \tag{2}$$

*）本课题得到国家973计划资助（项目号：2002CB312200）。**李鸿儒**　副教授，博士生，研究方向：粗糙集、概念格、模糊拓扑；**魏　平**　副教授，研究方向：模糊数学、粗糙集、概念格。

其中，X^* 表示拥有对象 X 的所有属性全体构成的集合，B^* 表示 U 具有 B 中所有属性的对象全体构成的集合。对于 $x\in U$ 和 $a\in A$，我们记 $\{x\}^*=x^*$，$\{a\}^*=a^*$。

定义 1 设 $K=(U,A,I)$ 是一个形式背景，$X\subseteq U,B\subseteq A$。如果 X,B 满足条件 $X^*=B,B^*=X$，则称序对 (X,B) 为形式背景 K 的一个概念。X 称为概念 (X,B) 的外延，B 称为概念 (X,B) 的内涵。$L(U,A,I)$（或 $L(K)$）表示 K 中所有概念全体构成的集合，即

$$L(U,A,I)=\{(X,B)\in U\times A;X^*=B,B^*=X\}$$

在形式背景 K_1 中，如取 $X=\{1,2,4\},B=\{a,b\}$，容易看出 X,B 满足定义 1 中的条件，即 $X^*=\{a,b\}=B$，而且 $B^*=\{1,2,4\}=X$，因此 $(X,B)\in L(K_1)$。

设 $K=(U,A,I)$ 是一个形式背景，$\rho(U)$ 和 $\rho(A)$ 分别表示 U 和 A 的幂集，对任意的 $X,X_1,X_2\in\rho(U)$ 和 $B,B_1,B_2\in\rho(A)$，算子“$*$”具有下面的性质[2]：

1) $X_1\subseteq X_2\Rightarrow X_2^*\subseteq X_1^*$，

1)′ $B_1\subseteq B_2\Rightarrow B_2^*\subseteq B_1^*$，

2) $X\subseteq X^{**}$，

2)′ $B\subseteq B^{**}$，

3) $X^*=X^{***}$，

3)′ $B^*=B^{***}$，

4) $(X_1\cup X_2)^*=X_1^*\cap X_2^*$，

4)′ $(B_1\cup B_2)^*=B_1^*\cap B_2^*$，

5) $(X_1\cap X_2)^*\supseteq X_1^*\cap X_2^*$，

5)′ $(B_1\cap B_2)^*\supseteq B_1^*\cap B_2^*$，

6) $(X^{**},X^*)\in L(K)$，

6)′ $(B^*,B^{**})\in L(K)$。

定义 2 设 $K=(U,A,I)$ 是一个形式背景，(X_1,B_1)，$(X_2,B_2)\in L(K)$，如果 $X_1\subseteq X_2$（或 $B_2\subseteq B_1$），称 (X_1,B_1) 是 (X_2,B_2) 的子概念，记为 $(X_1,B_1)\leqslant(X_2,B_2)$。概念的运算“$\wedge$”和“$\vee$”定义为：

$$(X_1,B_1)\wedge(X_2,B_2)=(X_1\cap X_2,(B_1\cup B_2)^{**})\quad(3)$$

$$(X_1,B_1)\vee(X_2,B_2)=((X_1\cap X_2)^{**},B_1\cap B_2)\quad(4)$$

显然 $L(K)$ 关于“$\leqslant$”构成一个格，称为概念格[2]，容易证明 $L(K)$ 是一个完备格。

3 概念格中的闭系统及其性质

定义 3 设 S 是一个非空集，$U\subseteq\rho(S)$。如果 U 满足条件：

(i) $S\in\mathscr{U}$；

(ii) $\chi\subseteq\mathscr{U}\Rightarrow\bigcap\chi\in\mathscr{U}$。

我们称 $\mathscr{U}$ 为一个闭系统。

定义 3 说明闭系统关于交运算是封闭的。下面将给出概念格中的两个闭系统及它们的性质。

设 $L(K)=L(U,A,I)$ 是一个概念格，$EX(K)$ 和 $IN(K)$ 分别表示 $L(K)$ 中所有外延全体和所有内涵全体构成的集合。从上面的性质 6) 和 6)′ 容易看出，对任意的 $X\subseteq U$ 和 $B\subseteq A$，$X^*\in IN(K)$，$B^*\in EX(K)$。而且

$$X\in EX(K)\Leftrightarrow X=X^{**},B\in IN(K)\Leftrightarrow B=B^{**}$$

即

$$EX(K)=\{X\in\rho(U);X=X^{**}\}\quad(5)$$

$$IN(K)=\{B\in\rho(A);B=B^{**}\}\quad(6)$$

定理 1 设 $K=(U,A,I)$ 是一个形式背景，则

(i) $EX(K)=\{\bigcap_{a\in B}a^*;B\subseteq A\}$；

(ii) $IN(K)=\bigcap_{x\in X}x^*;X\subseteq U\}$；

(iii) $EX(K)$，$IN(K)$ 均为闭系统和完备格。

证明：(i) 由概念的定义易知，对任意的 $B\subseteq A$，$B^*=\bigcap_{a\in B}a^*$。设 $X=B^*=\bigcap_{a\in B}a^*$，则 $X^{**}=B^{***}=B^*=X$。因此 $X=\bigcap_{a\in B}a^*\in EX(K)$。

反过来，$\forall X\in EX(K)$，存在 $C\subseteq A$，使得 $(X,C)\in L(K)$。即 $X=C^*=\bigcap_{a\in C}a^*\in\{\bigcap_{a\in B}a^*;B\subseteq A\}$。由此可知，结论 (i) 成立。(ii) 的证明是类似的。(iii) 的结果可由 (i)、(iii) 以及 LK 的运算性质直接得出。证毕。

根据定理 1，容易证明 $EX(K)$ 和 $IN(K)$ 是完备格，因此它们对交、并运算是封闭的。对任意的 $X_1,X_2\in EX(K)$，B_1，$B_2\in IN(K)$，我们有

$$X_1\vee X_2=\inf\{X\in EX(K);X_1\cup X_2\subseteq X\}$$

$$X_1\wedge X_2=\sup\{X\in EX(K);X\subseteq X_1\cap X_2\}$$

$$B_1\vee B_2=\inf\{B\in IN(K);B_1\cup B_2\subseteq B\}$$

$$B_1\wedge B_2=\sup\{B\in IN(K);B\subseteq B_1\cap B_2\}$$

由 $EX(K)$ 和 $IN(K)$ 是闭系统，易知 $X_1\wedge X_2=X_1\cap X_2$，$B_1\wedge B_2=B_1\cap B_2$。但一般地，$X_1\vee X_2\neq X_1\cup X_2$，$B_1\vee B_2\neq B_1\cup B_2$。概念的交、并运算与外延集（或内涵集）的交、并运算具有下面的性质。

定理 2 设 $K=(U,A,I)$ 是一个形式背景。(X_1,B_1)，$(X_2,B_2)\in L(K)$，则

(i) $(X_1,B_1)\wedge(X_2,B_2)=(X_1\cap X_2,B_1\vee B_2)=(X_1\wedge X_2,B_1\vee B_2)$

(ii) $(X_1,B_1)\vee(X_2,B_2)=(X_1\vee X_2,B_1\cap B_2)=(X_1\vee X_2,B_1\wedge B_2)$

证明：(i) 由于 $(B_1\vee B_2)^{**}\in IN(K)$，而且它是包含 $B_1\cup B_2$ 最小的内涵[2]，因此 $(B_1\cup B_2)^{**}=\inf\{B\in IN(K);B_1\cup B_2\subseteq B\}=B_1\vee B_2$。根据 (2.3) 知结论 (i) 成立。

(ii) 的证明是类似的。

证毕。

设 T 是一个指标集，对任意的 $t\in T$，$(X_t,B_t)\in L(K)$。利用概念格的基本定理[2]和上面的定理 2，我们有

$$\bigwedge_{t\in T}(X_t,B_t)=(\bigcap_{t\in T}X_t,\bigvee_{t\in T}B_t)=(\bigwedge_{t\in T}X_t,\bigvee_{t\in T}B_t)\quad(7)$$

$$\bigvee_{t\in T}(X_t,B_t)=(\bigvee_{t\in T}X_t,\bigcap_{t\in T}B_t)=(\bigvee_{t\in T}X_t,\bigwedge_{t\in T}B_t)\quad(8)$$

等式 7 和 8 说明概念之间的运算可以通过外延（内涵）闭系统的运算来刻画。

4 不可约元与属性分类识别

定义 4 设 $K=(U,A,I)$ 是一个形式背景，$X\in EX(K)$。如果 $X\neq\vee\{Y\in EX(K);Y<X\}$，则称 X 是 $EX(K)$ 中 $\vee$-不可约元。如果 $X\neq\wedge\{Y\in EX(K);X<Y\}$，则称 X 是 $EX(K)$ 中 $\wedge$-不可约元。$EX(K)$ 中全体 $\vee$-不可约元构成的集合和 $\wedge$-不可约元构成的集合分别记为 $M_{EM}(K)$ 和 $CM_{EX}(K)$。

定理 3 设 $K=(U,A,I)$ 是一个形式背景，$Y\subseteq U$。则

$$Y\in CM_{EX}\Rightarrow\exists a\in A,\text{使得 }Y=a^*$$

证明：显然 $Y\in CM_{EX}$，蕴涵 $Y\in EX(K)$。因此存在 $C\subseteq A$，使得 $Y=\bigcap_{c\in C}c^*$。对任意的 $a\in A$，如果 $Y\neq a^*$，则 $\{c^*;c\in C\}\subseteq\{X\in EX(K)\mid Y<X\}$。

根据 $Y\leqslant\wedge\{X\in EX(K)\mid Y<X\}\leqslant\bigwedge_{c\in C}c^*=\bigcap_{c\in C}c^*=Y$，我们有 $Y=\wedge\{X\in EX(K)\mid Y<X\}$，这与 $Y\in CM_{EX}$ 是矛盾的，

因此结论成立。

证毕。

定理 3 的结果说明，闭系统 $EX(K)$ 中交不可约元全体构成的集合是每个属性外延全体构成集合的一个子集，即

$CM_{EX}\subseteq\{a^*;a\in A\}$

由此可知集合 CM_{EM} 可以表示为

$$CM_{EX}=\{a^*\mid a\in A,a^*\neq\bigcap_{b\in A}b^*\}\tag{9}$$

设 $L(U,A_1,I_1)$，$L(U,A_2,I_2)$ 是两个概念格。如果对任意的 $(X,B)\in L(U,A_2,I_2)$，总存在 $(Y,C)\in L(U,A_1,I_1)$，使得 $X=Y$。我们称 $L(U,A_1,I_1)$ 细于 $L(U,A_2,I_2)$，记作 $L(U,A_1,I_1)\leqslant L(U,A_2,I_2)$。

如果 $L(U,A_1,I_1)\leqslant L(U,A_2,I_2)$，且 $L(U,A_2,I_2)\leqslant L(U,A_1,I_1)$，则 $L(U,A_1,I_1)$ 与 $L(U,A_2,I_2)$ 是格同构，即 $L(U,A_1,I_1)\cong L(U,A_2,I_2)$。

对于一个形式背景 $K=(U,A,I)$，如果 $D\subseteq A$，则 (U,D,I_D) 也是一个形式背景，其中 $I_D=I\cap(U\times D)$。容易证明，对任意的 $D\subseteq A$，$L(U,A,I)\leqslant L(U,D,I_D)$ 总成立。如果 $L(U,D,I_D)\cong L(U,A,I)$，则称 D 为形式背景 $K=(U,A,I)$ 的一个协调集。明显地，

$$D\text{ 为协调集}\Leftrightarrow L(U,D,I_D)\leqslant L(U,A,I)\tag{10}$$

设 $K=(U,A,I)$ 是一个形式背景，$D\subseteq A$。如果 $L(U,D,I_D)\cong L(U,A,I)$，且对任意的 $d\in D$，$L(U,D-\{d\},I_{D-\{d\}})$ 不同构于 $L(U,A,I)$，我们称 D 为 (U,A,I) 的一个约简。$RED(K)$ 表示 K 中所有约简构成的集合。

知识约简是数据挖掘与知识发现的一个核心问题。概念格的知识约简就是要在属性集中找出一个最小的子集，使其确定的概念格与所有属性确定的概念格是同构的。属性特征识别就是将属性集中的元素按照在知识约简中的作用进行分类。一般来说，属性集中每个属性对概念描述的作用不尽相同，有些属性是不可缺少的，而有些属性是不必要的（或多余的）。如何识别每个属性在概念格约简中的作用，这对概念格的研究是十分重要的。下面我们将给出属性分类定义以及属性分类识别方法。

定义 5 设 $K=(U,A,I)$ 是一个形式背景，$e\in A$，$RED(K)=\{D_i\mid D_i$ 是约简，$i\in T\}$，其中 T 为指标集。记

$\mathscr{C}=\bigcap_{i\in T}D_i$，$\mathscr{K}=\bigcup_{i\in T}D_i-\mathscr{C}$，$\mathscr{L}=A-\bigcup_{i\in T}D_i$

(i) 如果 $e\in\mathscr{C}$，称 e 是绝对必要属性（或核心元）；

(ii) 如果 $e\in\mathscr{K}$，称 e 是相对必要属性；

(iii) 如果 $e\in\mathscr{L}$，称 e 是多余属性（或不必要属性）。

显然，当 $e=\varphi$ 或 $e^*=U$ 时，$e\in\mathscr{L}$。根据约简定义容易证明下面的结果：

$$e\in\mathscr{C}\Leftrightarrow L(U,A-\{e\},I_{A-\{e\}})\text{ 不同构于 }L(U,A,I)\tag{11}$$

设 $K=(U,A,I)$ 是一个形式背景，记

$$E=\{(a,b)\in A\times A;a^*=b^*\}\tag{12}$$

明显地，E 是一个 A 上的等价关系。由 E 确定的等价类记为：

$A/E=\{[a]_E;a\in A\}$，其中 $[a]_E=\{b\in A;a^*=b^*\}$

定理 4 设 $K=(U,A,I)$ 是一个形式背景，$e\in A$，则

$$e\in\mathscr{C}\Leftrightarrow e^*\in CM_{EX}\text{，且 }[e]_E=\{e\}\tag{13}$$

证明："$\Rightarrow$"任取 $(Y,B)\in L(U,A,I)$，如果 $e\notin B$，则 $(Y,B)\in L(U,A-\{e\},I_{A-\{e\}})$ 是显然的。如果 $e\in B$，且 $[e]_E\neq\{e\}$，则存在 $b\in A$，使得 $e^*=b^*$，根据(2.1)知，$e\in B\Leftrightarrow b\in B$。因为 $Y=\bigcap_{a\in B}a^*=(\bigcap_{a\in(B-\{e\})}a^*)\cap e^*=(\bigcap_{a\in(B-\{e\})}a^*)\cap b^*=(\bigcap_{a\in(B-\{e\})}a^*)=C^*$，其中 $C=B-\{e\}$，而且 $Y^*_{A-\{e\}}=Y^*\cap(A-\{e\})=B\cap(A-\{e\})=B-\{e\}=C$，从而有 $(Y,C)\in L(U,A-\{e\},I_{A-\{e\}})$。这意味 $L(U,A-\{e\},I_{A-\{e\}})\geqslant L(U,A,I)$。显然，与条件 $e\in\mathscr{C}$ 矛盾，因此 $[e]_E=\{e\}$。

假设 $e^*\notin CM_{EX}$，且 $e\in B$。我们令 $G=\{a\in A;e^*=a^*\}$，则 $e^*=\wedge\{X\in EX(K);e^*<X\}=\wedge\{a^*;a\in G\}$。根据 $(Y,B)\in L(U,A,I)$ 及 $e\in B$，我们有 $Y=\bigcap_{a\in B}a^*\subseteq e^*$，从而 $Y\subseteq a^*$ 对所有的 $a\in G$。因为 $(Y,B)\in L(U,A,I)$ 蕴涵 $B=\{a\in A,Y\subseteq a^*\}$ 和 $Y=\bigcap_{a\in B}a^*$，所以 $G\subseteq B$。而且

$Y=\bigcap_{a\in B}a^*=(\bigcap_{a\in(B-\{e\})}a^*)\cap e^*=(\bigcap_{a\in(B-\{e\})}a^*)=C^*$，$(Y,C)\in L(U,A-\{e\},I_{A-\{e\}})$

其中 $C=B-\{e\}$。因此我们得到 $L(U,A-\{e\},I_{A-\{e\}})\leqslant L(U,A,I)$，这是不可能的，因为 $e\in\mathscr{C}$。由此可知 $e^*\in CM_{EX}$。

"$\Leftarrow$"假设 $e^*\in CM_{EX}$，$[e]_E=\{e\}$，则 $e^*\neq\wedge\{a^*\mid a\in A;e^*\subset a^*\}$，而且 $e^*\neq a^*$ 对所有的 $a\in A$。由于 $(e^*,e^{**})\in L(U,A,I)$，$e^*\notin EX(U,A-\{e\},I_{A-\{e\}})=\{\bigcap_{a\in B}a^*;E\subseteq(A-\{e\})\}$，因此 $L(U,A-\{e\},I_{A-\{e\}})$ 不同构于 $L(U,A,I)$，即 $e\in\mathscr{C}$。

根据定理 4，下面的结论是显然的。

推论 5 设 $K=(U,A,I)$ 是一个形式背景，$e\in A$，若存在 $a\in A(a\neq e)$，满足条件 $e^*=a^*$，则 $e\notin\mathscr{C}$。

定理 6 设 $K=(U,A,I)$ 是一个形式背景，$e\in A$，$e^*\neq U$，$e^*\neq\varphi$，则

$e\in\mathscr{L}\Leftrightarrow e^*\notin CM_{EX}$

证明：若 $e\in\mathbb{L}$，则 $\forall D\in RED(K)$，$e\notin D$。由 $L(U,D,I_D)\leqslant L(U,A,I)$，及 $(e^*,e^{**})\in L(U,A,I)$ 可知，存在 $B\subseteq D$，使得 $e^*=\bigcap_{a\in B}a^*$。因为 $e\notin D$，$B\subseteq D$，我们有 $a\neq e$ 对所有的 $a\in B$，而且 $\{a^*;a\in B\}\subseteq\{X\in EX(K);e^*<X\}$，$e^*\leqslant\wedge\{X\in EX(K);e^*<X\}\leqslant\bigwedge_{a\in B}a^*=\bigcap_{a\in B}a^*=e^*$。

故 $e^*=\wedge\{X\in EX(K);e^*<X\}$。

反之，设 $e^*\notin CM_{EX}$，根据定理 4 的证明，我们有 $L(U,D-\{e\},I_{D-\{e\}})\leqslant L(U,A,I)$。

由此可知，e 是一个不必要属性。

证毕。

定理 7 设 $K=(U,A,I)$ 是一个形式背景，$e\in A$，则

$e\in\mathscr{K}\Leftrightarrow e^*\in CM_{EX}$，且 $[e]_E\neq\{e\}$

证明：假设 $e\in\mathscr{K}$，由定理 6 知 $e^*\in CM_{EX}$。如果 $[e]_E=\{e\}$，则 $e\in\mathscr{C}$，即必要性成立。

反之，设 $e^*\in CM_{EX}$，$[e]_E\neq\{e\}$，则存在 $b\in A$，使得 $e^*=b^*$，而且对任意的 $(X,B)\in L(U,A,I)$，$e\in B\Leftrightarrow b\in B$。如果 $e\notin B$，则 $(X,B)\in L(U,A-\{e\},I_{A-\{e\}})$。如果 $e\in B$，则 $X=\bigcap_{a\in B}a^*=(\bigcap_{a\in(B-\{e\})}a^*)\cap e^*=(\bigcap_{a\in(B-\{e\})}a^*)\cap b^*=(\bigcap_{a\in(B-\{e\})}a^*)=C^*$，其中 $C=B-\{e\}$。

容易验证 $(X,C)\in L(U,A-\{e\},I_{A-\{e\}})$。即 $L(U,A-\{e\},I_{A-\{e\}})\leqslant L(U,A,I)$，$e\notin\mathscr{C}$，因为 $e^*\in CM_{EX}$，所以 $e\notin\mathscr{L}$。由此可知，e 是一个相对必要属性。

证毕。

推论 8 设 $K=(U,A,I)$ 是一个形式背景，$e\in A$，且 $e^*\in CM_{EX}$，则 $e\in\mathscr{K}\Leftrightarrow\exists a\in A(a\neq e)$，$a^*=e^*$

例题 设 $K_2=(U,A,I)$ 是一个形式背景，其中对象集 $U=\{1,2,\cdots,6\}$，属性集 $A=\{a,b,c,d,e\}$，U 和 A 之间的关系 I

由表2给出。

表2 形式背景

	a	b	c	d	e
1		×	×	×	
2		×		×	×
3					×
4	×		×		×
5			×		
6		×		×	×

从表2可以看出，$a^*=4,b^*=126,c^*=145,d^*=126,e^*=2346$。

根据定理1及定义1，我们得到

$EX(K_2)=\{U,2346,126,145,26,4,1,\phi\}$

$IN(K_2)=\{A,bcd,bde,ace,bd,c,e,\phi\}$

$L(K_2)=\{(U,\phi),(2346,e),(126,bd),(145,c),(26,bde),(4,ace),(1,bcd),(\phi,A)\}$

因为 $a^*=c^*\cap e^*$，所以 $a^*\notin CM_{EX}$。这样

$CM_{EX}=\{b^*,c^*,d^*,e^*\}=\{2346,126,145\}$

应用定理4、定理6及定理7，可确定形式背景 K_2 中的属性可分类为：

$_e=\{c,e\},_k=\{b,d\},_t=\{a\}$

上面的结果说明，对任意的 $D\in RED(K_2),c,e\in D,a\notin D$。容易验证 $D_1=\{b,c,e\}$ 和 $D_2\{c,d,e\}$ 是概念格 $L(K_2)$ 的两个约简。

结论 在形式概念分析中，属性的分类对知识约简具有重要的意义。目前，关于概念格属性分类和约简问题的研究已取得一些成果，这些结果大多是应用序结构及辨识矩阵方法来处理。本文研究了形式背景的外延(内涵)闭系统中不可约元的性质，揭示了不可约元与属性特征的关系，从而给出了一种属性特征分类识别方法。这种方法不仅操作简单，而且为概念格约简及知识发现的进一步研究提供了新的途径。

参考文献

1 Wille R. Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival I ed. Ordered Sets. Reidel, Dordrecht, 1982. 445～470

2 Ganter B, Wille R. Formal Concept Analysis: Mathematical Foundations. New York: Springer-Verlag, 1999

3 Düntsch I, Gediga G. Algebraic aspects of attribute dependencies in information systems. Fundamenta Informaticae, 1997, 29: 119～133

4 Düntsch I, Gediga G. Approximation operators in qualitative data analysis. In: de Swart H, Orlowska E, Schmidt G, et al. eds. Theory and Application of Relational Structures as Knowledge Instruments. Springer, Heidelberg, 2003. 216～233

5 Pagliani P. From concept lattices to approximation spaces: Algebraic structures of some spaces of partial objects. Fundamenta Informaticae, 1993, 18(1): 1～25

6 Yao Y Y. Concept lattices in rough set theory. Proceedings of 23rd International Meeting of the North American Fuzzy Information Processing Society, 2004

7 Diday E, Emilion R. Maximal and stochastic Galois lattices. Discrete Applied Mathematics, 2003, 127: 271～284

8 Bělohlaávek R. Concept lattices and order in fuzzy logic. Annals of Pure and Applied Logic, 2004, 128: 277～298

9 张文修，梁怡，吴伟志. 信息系统与知识发现. 北京：科学出版社，2003

10 张文修，吴伟志，梁吉业，等. 粗糙集理论与方法. 北京：科学出版社，2001

(上接第174页)

$\vec{y}=y_1,y_2,\cdots,y_D\in R^{-1}(y)$。又设 K_i 是 $X_i\times X_i$ 上的核函数，$1\leqslant i\leqslant D$，$K_i(x_i,y_i)$ 给出 x_i 与 y_i 的相似性度量。x 与 y 的相似性度量由下面的广义卷积给出：

$$K(x,y)=\sum_{\vec{x}\in R^{-1}(x),\vec{y}\in R^{-1}(y)}\prod_{i=1}^{D}K_i(x_i,y_i)$$

上式在 $S\times S$ 上有定义，其中 $S=\{x:R^{-1}(x)非空\}\subset X$。$K$ 在 $X\times X$ 上的零扩张称为核函数 $K_1,K_2,\cdots,K_D$ 的 R-卷积，记作 $K_1*K_2*\cdots*K_D$。

定理3.3.1[12] 设 $K_1,K_2,\cdots,K_D$ 分别是 $X_1\times X_1,X_2\times X_2,\cdots,X_D\times X_D$ 上的核函数，R 是 $X_1\times\cdots\times X_D\times X$ 上的有限关系，则 $K_1*K_2*\cdots*K_D$ 是 $X\times X$ 上的核函数。

例3.7 设 Ω 是任意非空集合，映射 $f:\Omega\to X$。若 $k:X\times X\to\mathbb{R}$ 是核函数，则 $\hat{k}:\Omega\times\Omega\to\mathbb{R}$

$\hat{k}(\omega,\omega')=k(f(\omega),f(\omega'))$

是核函数。我们经常用到的 $k(x-x')=k(\|x-x'\|)$ 就属于这种类型，是由低维空间上核函数构造高维空间上核函数的常用方法。

卷积核是构造核函数，尤其是复杂结构上核函数的重要方法。很多构造核函数的方法，均可纳入卷积核这个框架。例如，设，

$x=(x_1,\cdots,x_d),x'=(x'_1,\cdots,x'_d)$

· 核函数的张量积：

$\tilde{k}(x,x')=k(x_1,x'_1)\cdots\cdots+k(x_d,x'_d)$；

· 核函数的直和：

$\tilde{k}(x,x')=k(x_1,x'_1)+\cdots\cdots+k(x_d,x'_d)$；

· 核函数的投影：

$\tilde{k}(x,x')=k((x_1,0,\cdots,0),(x'_1,0,\cdots,0))$。

灵活地定义关系 R，能得到很多重要核函数。进一步的讨论见文[12]及那里的参考文献。

结束语 本文首先对核函数的基本性质做了深入分析和概括，然后对3类重要核函数，即平移不变核函数、旋转不变核函数和卷积核，提出了简单实用的判别准则，并据此验证和构造了很多重要核函数。这些工作，为核函数的设计奠定了理论基础。下一步的工作是，把问题领域的知识与核函数的设计结合起来，即把问题领域的知识嵌入核函数之中，这对提升支持向量机的性能非常重要。

参考文献

1 Burges C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998, 2: 121～167

2 Vapnik V. Statistical learning theory. New York: John Wiley & Sons, Inc, 1998

3 Cristianini N, Shawe-Taylor J. An introduction to support Vector Machines: and other kernel-based learning methods. New York: Cambridge University Press, 1999

4 Hsu C-W, Chang C-C, Lin C-J. A practical guide to support vector classification. http://www.csie.ntu.edu.tw/~cjlin/papers.html, 2003

5 Poggio. On optimal nonlinear associative recall. Biological Cybernetics, 1975, 19: 201～209

6 Cortes C, Vapnik V. Support vector networks. Machine Learning. 1995, 20: 273～297

7 Ressel B C. Harmonic analysis on semigroups. Beijing: Springer-Verlag, World Publishing Corporation, 1984

8 Scholkopf S W. Generalization bounds for convex combinations of kernel functions: [Technical Report]. NeuroCOLT2. Technical Report Series NC2-TR-1998-022, August 1998

9 Cristianini N, Shawe-Taylor J, Kandola. On Kernel Target Alignment. In: Proceedings of the Neural Information Processing Systems, NIPS'01, MIT Press, 2002

10 Lin Hsuan-Tien, Lin Chih-Jen. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. http://www.csie.ntu.edu.tw/~cjlin/papers.html, 2003

11 Steinwart I. On the influence of the kernel on the consistency of support vector machines. The Journal of Machine Learning Research, 2002, 2: 67～93

12 Haussler D. Convolutional kernels on discrete structures: [technical report]. UCSC-CRL-99-10, Santa Cruz: Computer Science Department, University of California, 1999

13 Feller W. An introduction to probability theory and its applications. Vol 2. New York: Wiley, 1971