一种新的概念格结构——区间概念格

刘保相 张春英
(河北联合大学理学院 唐山 063009)

摘 要 对经典概念格、粗糙概念格的分析表明，其概念外延或者具有全部属性，或者只具备一个属性，从而造成所提取关联规则支持度和可信度严重下降。为此提出一种新的概念格结构——区间概念格 $L_\beta^\alpha(M^\alpha,M^\beta,Y)$，其概念外延是区间$[\alpha,\beta]$($0\leqslant\alpha\leqslant\beta\leqslant1$)范围内满足内涵属性的对象集。证明了当 $\alpha=\beta=1$ 时，区间概念格退化为经典概念格；当 $\beta=1,\alpha>0$ 时，区间概念格退化为粗糙概念格；其次，给出了区间概念格中概念度量的精度、覆盖度等概念，并给出了相关性质；接着，证明了区间概念格具有的一些独特性质；然后，初步给出了构造区间概念格的方法；最后，通过实例证明了区间概念格提出的必要性和实用性。

关键词 概念格，粗糙概念格，区间概念格，概念精度
中图法分类号 TP182 **文献标识码** A

New Concept Lattice Structure——Interval Concept Lattice

LIU Bao-xiang ZHANG Chun-ying
(College of Science, Hebei United University, Tangshan 063009, China)

Abstract Analysis of classic concept lattice and rough concept lattice shows that concept extension with either all the attributes or only one attribute can decrease support and confidence of the extracted association rules greatly. To solve this problem, the author put forward a new concept of lattice structure: interval concept lattice $L_\beta^\alpha(M^\alpha,M^\beta,Y)$, in which the concept extension is object sets that meet the intension property in the interval$[\alpha,\beta]0\leqslant\alpha\leqslant\beta\leqslant1$. Firstly, it was proved that interval concept lattice degenerates into classic concept lattice when $\alpha=\beta=1$, and when $\beta=1,\alpha>0$, interval concept lattice degenerates into rough concept lattice. Secondly, the measurement precision and coverage of interval concept lattice concepts were given and some related properties were discussed. Thirdly, some unique properties of interval concept lattice were proved. Fourthly, the construction method of interval concept lattice was preliminary provided. Finally, the necessity and practicability were verified through a case study.

Keywords Concept lattice, Rough concept lattice, Interval concept lattice, Concept precision

1 引言

概念格是 Wille R 教授于1982年提出的进行数据分析的一种有力工具[1]，其上的每个节点是由内涵(即概念的描述)和外延(即内涵所覆盖的对象)构成的一个形式概念。从形式背景中生成概念格的过程实质上是概念聚类的过程，并通过 Hasse 图生动地体现了概念之间的泛化和特化关系。概念格作为一种数据分析工具，具有完备性和精确性的特点，在信息检索、数字图书馆、知识发现等方面得到了广泛应用[2-4]。目前，国内外学者对概念格进行了多方面深入研究，主要包括：概念格构造算法及改进[5-7]、基于概念格的规则挖掘[8,9]，以及与其他理论如模糊理论、谓词逻辑、粗集理论等进行融合从而对概念格进行扩展，得到模糊概念格、加权概念格、约束概念格、量化概念格、扩展概念格、粗糙概念格等[10-12]。

特别地，在经典概念格及其扩展的模糊概念格、加权概念格等概念中，对概念的描述都是基于外延完全具备内涵中的所有属性来进行构建的，要查找具有部分属性的概念则要通过扫描概念格，对概念进行合并，这对庞大的概念格来说时间代价高、效率低。在粗糙概念格[12]中，利用粗糙集中上、下近似的理论，描述概念格中内涵所拥有的外延，从而对于特定的内涵，在其描述的外延中能够表示出某种属性可能覆盖的对象，及其拥有的外延具有的不确定的性质。由于其概念的上近似外延只要求属性与内涵的交集非空即可，这样可能会存在大量仅具备内涵中一个属性的对象，由此构建的关联规则支持度和可信度都将大大降低。在实际应用中，用户往往关心的是具备一定数量或比例的内涵中属性的对象集合，并由此进行关联分析，挖掘更具针对性的关联规则。例如，在超市购物系统中，推销部经理往往更关注同时购买 $k(k>1)$种及以上商品的客户以及这些客户对这些商品的潜在需求，从而有针对性地对这些客户进行产品宣传，以达到最小的宣传最大的效益。在目前已有的概念格结构中均不能直接进行这种查询，都要进行一定的合并连接或者筛选，时间代价和空间代价

到稿日期：2011-11-12 返修日期：2012-03-25 本文受河北省自然科学基金项目(A2011209046)，国家自然基金项目(61170317)资助。
刘保相(1957—)，男，教授，主要研究方向为概念格、粗糙集、数据挖掘等，E-mail：liubx5888@126.com；张春英(1969—)，女，博士，教授，主要研究方向为概念格、人工智能、多关系数据挖掘等。

较高。为此,本文提出一种新的概念格结构——区间概念格,并证明此区间概念格是经典概念格和粗糙概念格的拓展,经典概念格和粗糙概念格是区间概念格的特例。文章进一步给出区间概念格的构造算法,并通过比较证明该方法更有效。

2 基本概念

2.1 概念格理论的相关概念

定义 1[1] 称(U,A,R)为一个形式背景,其中$U=\{x_1,x_2,\cdots,x_n\}$为对象集,每个$x_i(i\leqslant n)$称为一个对象;$A=\{a_1,a_2,\cdots,a_m\}$为属性集,每个$a_j(j\leqslant m)$称为一个属性;R为U和A之间的二元关系,$R\subseteq U\times A$。若$(x,a)\in R$,则称x具有属性a,记为xRa。

定义 2[1] 对于形式背景(U,A,R),算子f,g定义为:

$\forall x\in U,f(x)=\{y|\forall y\in A,xRy\}$,即$f$是对象$x$与其具有的所有属性的映射;

$\forall y\in A,g(y)=\{x|\forall x\in U,xRy\}$,即$g$是属性$y$与其覆盖的所有对象的映射。

定义 3 对于形式背景(U,A,R),若对于$X\subseteq U,Y\subseteq A$,有$f(X)=Y,g(Y)=X$,则称序偶$\langle X,Y\rangle$是一个形式概念,简称概念。其中$X$称为概念的外延,$Y$称为概念的内涵。

定义 4 用$L(U,A,R)$表示形式背景(U,A,R)的全体概念,记:

$$(X_1,Y_1)\leqslant(X_2,Y_2)\Leftrightarrow X_1\subseteq X_2(\Leftrightarrow Y_1\supseteq Y_2) \tag{1}$$

则"$\leqslant$"是$L(U,A,R)$上的偏序关系。

定义 5 若$L(U,A,R)$中的所有概念满足"$\leqslant$"偏序关系,则称$L(U,A,R)$是形式背景(U,A,R)的概念格。

定义 6 如果$(X_1,Y_1)\leqslant(X_2,Y_2)$,且二者之间不存在概念$(X_3,Y_3)$,满足:

$$(X_1,Y_1)\leqslant(X_3,Y_3)\leqslant(X_2,Y_2) \tag{2}$$

则称(X_1,Y_1)是(X_2,Y_2)的子概念,(X_2,Y_2)是(X_1,Y_1)的父概念。

2.2 粗糙概念格

粗糙概念格[12]是将粗糙集的上、下近似引入到概念中,用上近似外延和下近似外延分别表示具有属性集$Y\subseteq U$的最大概念集合与最小概念集合,避免了对概念进行部分属性查询时概念的合并运算。

定义 7[12] 对于形式背景(U,A,R),$L(U,A,R)$是其诱导的经典概念格,$\langle X,Y\rangle$是L上的经典概念。上近似外延M定义为:

$$M=\{x|x\in U,f(x)\cap Y\neq\phi\}$$

式中,Y是概念的内涵,$f(x)$是定义 2 中的算子。M表示可能被内涵所覆盖的对象集。

定义 8[12] 对于形式背景(U,A,R),$L(U,A,R)$是其诱导的经典概念格,$\langle X,Y\rangle$是L上的经典概念。下近似外延N定义为:

$$N=\{x|x\in X,f(x)\cap Y=Y\}$$

式中,X是经典概念的外延,Y是概念的内涵,$f(x)$是定义 2 中的算子。N表示满足内涵Y中全体属性的对象集。

定义 9[12] 设形式背景(U,A,R),三元序偶(M,N,Y)为由偏序关系R产生格结构L的任一结点,其中,Y为内涵,是概念的描述;M为上近似外延,表示可能被内涵所覆盖的对象;N为下近似外延,表示内涵所覆盖的部分对象,如果满足条件:

$$M=\{x|x\in U,\exists y\in A,xRy\} \tag{3}$$

$$N=\{x|x\in U,\forall y\in A,xRy\} \tag{4}$$

$$Y=\{y|y\in A,\forall x\in M,\exists y\in A,xRy\} \tag{5}$$

则称L为由(U,A,R)所诱导的粗糙概念格,简记为RL。每个结点(M,N,Y)称为一个粗糙概念。

定义 10[12] 设$H_1=(M_1,N_1,Y_1)$,$H_2=(M_2,N_2,Y_2)$是粗糙概念格的两个不同结点,则$H_1\leqslant H_2\Leftrightarrow Y_2\subseteq Y_1$,如果不存在$H_3(M_3,N_3,Y_3)$,使得$H_1\leqslant H_3\leqslant H_2$,则$H_2$是$H_1$的父结点(直接前驱),$H_1$是$H_2$的子结点(直接后继)。

定理 1 设(M,N,Y)为粗糙概念格的任一结点,当$M=N$时,粗糙概念格退化为经典概念格,即粗糙概念格是经典概念格的扩展。

3 区间概念格结构及性质

3.1 区间概念格结构

定义 11 对于形式背景(U,A,R),$RL(U,A,R)$是其诱导的经典概念格,(M,N,Y)是RL上的粗糙概念。设有区间$[\alpha,\beta](0\leqslant\alpha\leqslant\beta\leqslant1)$,则$\alpha$上界外延$M^\alpha$为:

$$M^\alpha=\{x|x\in M,|f(x)\cap Y|/|Y|\geqslant\alpha,0\leqslant\alpha\leqslant1\} \tag{6}$$

式中,Y是概念的内涵,$f(x)$是定义 2 中的算子,$|Y|$是集合Y中包含元素的个数,即基数。M^α表示可能被Y中至少$\alpha\times|Y|$个内涵属性所覆盖的对象。

定义 12 对于形式背景(U,A,R),$RL(U,A,R)$是其诱导的经典概念格,(M,N,Y)是RL上的粗糙概念。设有区间$[\alpha,\beta](0\leqslant\alpha\leqslant\beta\leqslant1)$,则$\beta$下界外延$M^\beta$为:

$$M^\beta=\{x|x\in M,|f(x)\cap Y|/|Y|\geqslant\beta,0\leqslant\alpha\leqslant\beta\leqslant1\} \tag{7}$$

X是经典概念的外延,Y是概念的内涵,$f(x)$是定义 2 中的算子。M^β表示可能被Y中至少$\beta\times|Y|$个内涵属性所覆盖的对象。

定义 13 设形式背景(U,A,R),三元序偶称(M^α,M^β,Y)为区间概念,其中,Y为内涵,是概念的描述;M^α为α上界外延;M^β为β下界外延。

定义 14 用$L_\alpha^\beta(U,A,R)$表示形式背景(U,A,R)的全体$[\alpha,\beta]$区间概念,记:

$$(M_1^\alpha,M_1^\beta,Y_1)\leqslant(M_2^\alpha,M_2^\beta,Y_2)\Leftrightarrow Y_1\supseteq Y_2 \tag{8}$$

则"$\leqslant$"是$L_\alpha^\beta(U,A,R)$上的偏序关系。

定义 15 若$L_\alpha^\beta(U,A,R)$中的所有概念满足"$\leqslant$"偏序关系,则称$L_\alpha^\beta(U,A,R)$是形式背景(U,A,R)的区间概念格。

定理 2 设序偶(M^α,M^β,Y)为区间概念格的任一结点,则其满足以下性质:

$$M^\alpha=\{x\in U|\exists Y_1\subseteq Y,|Y_1|/|Y|\geqslant\alpha(0\leqslant\alpha\leqslant1),\forall y\in Y_1,xRy\} \tag{9}$$

$$M^\beta=\{x\in U|\exists Y_2\subseteq Y,|Y_2|/|Y|\geqslant\beta(0\leqslant\alpha\leqslant\beta\leqslant1),\forall y\in Y_2,xRy\} \tag{10}$$

$$Y=\{y\in A|\forall x\in M^\alpha,\exists Y_3\subseteq A,|Y_3|/|Y|\geqslant\alpha(0\leqslant\alpha\leqslant1),\forall y\in Y_3,xRy\} \tag{11}$$

定义 16 设区间概念格的结点$G_1=(M_1^\alpha,M_1^\beta,Y_1)$,$G_2=(M_2^\alpha,M_2^\beta,Y_2)$,则$G_1\leqslant G_2\Leftrightarrow Y_1\subseteq Y_2$,如果不存在$G_3(GM_3,GN_3,Y_3)$,使得$G_1\leqslant G_3\leqslant G_2$,则$G_2$是$G_1$的父结点(直接前驱),$G_1$是$G_2$的子结点(直接后继)。

定义 17 设$L_\alpha^\beta(U,A,R)$是区间概念格,如果结点是(除

自身外)所有结点的前驱,则称该结点为区间概念格的根结点;如果结点是(除自身外)所有结点的后继,则称该结点为区间概念格的末梢结点。

3.2 概念度量

定义 18 设序偶(M^α, M^β, Y)为区间概念格的任一结点,

$$\tau_\beta^\alpha(Y)=|M^\beta|/|M^\alpha| \tag{12}$$

表示属性集 Y 在指定的参数 α,β 下所覆盖的对象关于指定关系 R 的近似精度;

$$\rho_\beta^\alpha(Y)=1-\tau_\beta^\alpha(Y) \tag{13}$$

表示相应的粗糙度。

性质 1 $0<\tau_\beta^\alpha(Y)\leqslant 1$,当且仅当 $\alpha=\beta$ 时,$\tau_\beta^\alpha(Y)=1$。

定义 19 设 β 下界外延为区间概念格中的一个概念结点,对于 M^α 中的任意对象 $u\in M^\alpha$,u 相对概念 G 的覆盖度定义为:

$$\varphi(u|G)=\frac{|u'|}{|Y|} \tag{14}$$

式中,$|u'|$为对象 u 的属性集的基数。

性质 2 $\varphi(u|G)\leqslant 1$,当且仅当 $u'=Y$ 时等号成立。

定义 20 设 $G=(M^\alpha, M^\beta, Y)$为区间概念格中的一个概念结点,$u_i\in M^\alpha(i=1,2,\cdots,|M^\alpha|)$,$\alpha$ 上界外延对象集相对于概念 G 的最小覆盖度为:

$$\varphi_{\min}(M^\alpha|G)=\mathrm{Min}_{i=1}^{|M^\alpha|}(\varphi(u_i|G)) \tag{15}$$

α 上界外延对象集相对于概念 G 的最大覆盖度为:

$$\varphi_{\max}(M^\alpha|G)=\mathrm{Max}_{i=1}^{|M^\alpha|}(\varphi(u_i|G)) \tag{16}$$

α 上界外延对象集相对于概念 G 的平均覆盖度为:

$$\varphi_{ave}(M^\alpha|G)=\sum_{i=1}^{|M^\alpha|}(\varphi(u_i|G)) \tag{17}$$

定义 21(α 上界对象覆盖向量) 规定 $u\notin M^\alpha$ 时,$\varphi(u|G)=0$,则

$$\varphi(M^\alpha|G)=(\varphi(u_i|G)),i=1,2,\cdots,k(k=|M^\alpha|) \tag{18}$$

称为 α 上界外延对象集相对于概念 G 的覆盖向量,这是一个 k 维向量。

性质 3 $\varphi_{\min}(M^\alpha|G)\leqslant\varphi_{ave}(M^\alpha|G)\leqslant\varphi_{\max}(M^\alpha|G)$

定义 22 设(M^α, M^β, Y)为区间概念格中的一个概念结点,对于 M^β 中的任意对象 $v\in M^\beta$,β 下界外延对象 v 相对概念 G 的覆盖度定义为:

$$\omega(v|G)=\frac{|v'|}{|Y|} \tag{19}$$

式中,$|v'|$为对象 u 的属性集的基数。

性质 4 $\omega(v|G)\leqslant 1$,当且仅当 $v'=Y$ 时等号成立。

定义 23 设 $G=(M^\alpha, M^\beta, Y)$为区间概念格中的一个概念结点,$v_i\in M^\beta(i=1,2,\cdots,|M^\beta|)$,$\beta$ 下界外延对象集相对于概念 G 的最小覆盖度为:

$$\omega_{\min}(M^\beta|G)=\mathrm{Min}_{i=1}^{|M^\beta|}(\omega(v_i|G)) \tag{20}$$

β 下界外延对象集相对于概念 G 的最大覆盖度为:

$$\omega_{\max}(M^\beta|G)=\mathrm{Max}_{i=1}^{|M^\beta|}(\omega(v_i|G)) \tag{21}$$

β 下界外延对象集相对于概念 G 的平均覆盖度为:

$$\omega_{ave}(M^\beta|G)=\sum_{i=1}^{|M^\beta|}(\omega(v_i|G)) \tag{22}$$

定义 24(β 下界对象覆盖向量) 规定当 $v\notin M^\beta$ 时,$\omega(v|G)=0$,则

$$\omega(M^\beta|G)=(\omega(v_i|G)),i=1,2,\cdots,n(n=|M^\beta|) \tag{23}$$

称为 β 下界外延对象集相对于概念 G 的覆盖向量,这是一个 n 维向量。

性质 5 $\omega_{\min}(M^\beta|G)\leqslant\omega_{ave}(M^\beta|G)\leqslant\omega_{\max}(M^\beta|G)$

性质 6 $\omega_{\min}(M^\beta|G)\geqslant\varphi_{\min}(M^\beta|G)$

性质 7 $\omega_{\max}(M^\beta|G)\geqslant\varphi_{\max}(M^\beta|G)$

性质 8 $\omega_{ave}(M^\beta|G)\geqslant\varphi_{ave}(M^\beta|G)$

3.3 区间概念格性质

区间概念格是在粗糙概念格基础上进行的拓展,其适用范围更加广泛。

定理 3 区间概念格是粗糙概念格的拓展,粗糙概念格是区间概念格的特例。

证明:设形式背景(U,A,R),据其构造的区间概念格为 $L_\alpha^\beta(U,A,R)$,$G=(M^\alpha, M^\beta, Y)$是任一$[\alpha,\beta]$区间概念。而据其构造的粗糙概念格为 $RL(U,A,R)$,$H=(M,N,Y)$是任一粗糙概念。

α 上界外延:

$M^\alpha=\{x|x\in U,|f(x)\cap Y|/|Y|\geqslant\alpha,0\leqslant\alpha\leqslant 1\}$

当 $\alpha=0$,且$|f(x)\cap Y|/|Y|>\alpha$ 时,其 α 上界外延:$M^\alpha=\{x|x\in U,|f(x)\cap Y|/|Y|>0\}$,即:

$M^\alpha=\{x|x\in U,f(x)\cap Y\neq\phi\}$

此时 $M^\alpha=M$,即 α 上界外延与粗糙概念格中的上近似外延相等。

β 下界外延为:

$M^\beta=\{x|x\in X,|f(x)\cap Y|/|Y|\geqslant\beta,0\leqslant\alpha\leqslant\beta\leqslant 1\}$

当 $\beta=1$,且$|f(x)\cap Y|/|Y|=\beta$ 时,其广义下近似外延为:$M^\beta=\{x|x\in X,|f(x)\cap Y|/|Y|=1\}$,即:

$M^\beta=\{x|x\in X,f(x)\cap Y=Y\}$

此时 $M^\beta=N$,即 β 下界外延与粗糙概念格中的下近似外延相等。

定理 4 当 $\beta=1$,$|f(x)\cap Y|/|Y|=1$;且 $\alpha=0$,$|f(x)\cap Y|/|Y|=0$ 时,区间概念格即退化为经典概念格。

证明:略。

定理 5 设序偶(M^α, M^β, Y)是区间概念格的任一结点,$M^\alpha=\{u_1,u_2,\cdots,u_r\}$,$M^\beta=\{v_1,v_2,\cdots,v_s\}$,$Y=\{y_1,y_2,\cdots,y_t\}=\bigcup_{i=1}^{p}Y_i$,$P(Y)$是 Y 的幂集,$Y_i\in P(Y)$,则:

(1)$M^\alpha=g(Y_1)\cup g(Y_2)\cup\cdots\cup g(Y_p)$,$|Y_i|/|Y|\geqslant\alpha$,即针对 α 上界外延 M^α,内涵 Y 中属性子集间的逻辑关系为"或";

(2)$Y\subseteq f(u_1)\cup f(u_2)\cup\cdots\cup f(u_r)$;

(3)$M^\beta=g(Y_1)\cup g(Y_2)\cup\cdots\cup g(Y_q)$,$|Y_i|/|Y|\geqslant\beta$,即针对 β 下界外延 M^β,内涵 Y 中属性子集间的逻辑关系为"或";

(4)$Y\subseteq f(v_1)\cup f(v_2)\cup\cdots\cup f(v_s)$。

证明:(1) $M^\alpha=g(Y_1)\cup g(Y_2)\cup\cdots\cup g(Y_p)$,$|Y_i|/|Y|\geqslant\alpha$,先证后者是前者的子集:

不妨设$\exists Y_1\subseteq Y$ 且$|Y_1|/|Y|\geqslant\alpha$,使得 $g(Y_1)\not\subset M^\alpha$。假设 $x\in g(Y_1)$且 $x\notin M^\alpha$,可以得出 $Y_1\subseteq f(x)$,又因为$|Y_1|/|Y|\geqslant\alpha$,则$|f(x)\cap Y|/|Y|\geqslant\alpha$,由定义 11 得 $x\in M^\alpha$,与假设 $x\notin M^\alpha$ 矛盾。所以任意的 $Y_i\subseteq Y$ 且$|Y_i|/|Y|\geqslant\alpha$,都有 $g(Y_i)\subseteq M^\alpha$,即$\bigcup_{i=1}^{p}g(Y_i)\subseteq M^\alpha$。

再证前者是后者的子集:

由定理 2 知,对于$\forall x_i\in M^\alpha$ 都有$\exists Y_i\subseteq Y$,$|Y_i|/|Y|\geqslant\alpha$,

$\forall y_i \in Y_i$ 满足 $x_i R y_i$，即 $x_i \in g(Y_i)$，所以有 $M^{\alpha}=\bigcup_i x_i \subseteq \bigcup_i g(Y_i)$。

故此，(1)成立。

(2) $Y \subseteq f(u_1) \cup f(u_2) \cup \cdots \cup f(u_r)$。

对于 $\forall Y_i \subseteq Y$，且 $|Y_i|/|Y| \geqslant \alpha$，有 $g(Y_i) \subseteq M^{\alpha}$。不妨设 $x \in g(Y_i)$，则可以得出：$Y_i \subseteq f(x)$。由于 Y_i 与 x 取值的任意性，可知：$Y \subseteq f(u_1) \cup f(u_2) \cup \cdots \cup f(u_r)$成立。

(3) $M^{\beta}=g(Y_1) \cup g(Y_2) \cup \cdots \cup g(Y_q)$，$|Y_i|/|Y| \geqslant \beta$。

此证明方法与(1)一样，故略。

(4) $Y \subseteq f(v_1) \cup f(v_2) \cup \cdots \cup f(v_s)$。

证明方法与(2)一样，故略。

定理 6 设 Y 是一组属性集合，$Y_i \subseteq Y$ 且 $|Y_i|/|Y| \geqslant \alpha$，$Y_j \subseteq Y$ 且 $|Y_j|/|Y| \geqslant \beta$。

X 是属性 Y 所覆盖的经典概念格的外延，M，N 分别是属性 Y 所覆盖的粗糙概念格的上近似外延和下近似外延，M^{α}，M^{β} 分别是属性 Y 所覆盖的区间概念格的 α 上界外延和 β 下界外延，则存在以下关系：

$X \subseteq N \subseteq M^{\beta} \subseteq M^{\alpha} \subseteq M$

证明：(1)设 $x \in X$，则根据经典概念格的完备性特点可知，$f(x)=Y$，即 $f(x) \cap Y=Y$，根据粗糙概念格的定义 8 可知，$x \in N$，所以 $X \subseteq N$ 成立。

(2)设 $x \in N$，由定义 8 可知，$f(x) \cap Y=Y$，即 $|f(x) \cap Y|/Y=1 \geqslant \beta(1 \leqslant \beta < 0)$，由定义 12 可知，$x \in GN$，所以 $N \subseteq M^{\beta}$ 成立。

(3)设 $x \in GN$，由定义 12 可知，$|f(x) \cap Y|/Y \geqslant \beta$，$(0 \leqslant \alpha \leqslant \beta \leqslant 1)$。所以 $|f(x) \cap Y|/Y \geqslant \alpha$，由定义 11 可知，$x \in GM$，所以 $M^{\beta} \subseteq M^{\alpha}$。

(4)设 $x \in M^{\alpha}$，由定义 11 可知，$|f(x) \cap Y|/Y \geqslant \alpha(0 < \alpha \leqslant 1)$，故 $f(x) \cap Y \neq \phi$，由定义 7 可知，$x \in M$，所以 $M^{\alpha} \subseteq M$（注意：此处 $\alpha \neq 0$）。

由上，命题得证。

4 区间概念格的构造方法

算法分 5 个步骤：

(1)计算属性集合幂集

从形式背景中提取属性集合 A，生成 A 的所有子集，构成幂集 $P(A)$。

(2)确定内涵

将属性集合幂集中的每个子集对应确定为区间概念的内涵。从而生成初始的区间概念格的结点集 G。为方便算法描述及实现，设每个结点为六元组(M^{α}，M^{β}，Y，$Parent$，$Child$，No)，内涵 Y 对应初始化为属性幂集 $P(A)$中的元素，其余都初始化为空。

(3)确定 α 上界外延

设定参数 $\alpha(0 < \alpha \leqslant 1)$。针对结点集中每个结点的内涵，取内涵中的每个属性子集 $Y_i \subseteq Y$ 且 $|Y_i|/|Y| \geqslant \alpha$，遍历全部记录，找出满足 Y_i 中全部属性的所有对象，并将其加入相应结点的 α 上界外延 M^{α} 中。

(4)确定 β 下界外延

设定参数 $\beta(0 \leqslant \alpha \leqslant \beta \leqslant 1)$。针对结点集中每个结点的内涵，取内涵中的每个属性子集 $Y_i \subseteq Y$ 且 $|Y_i|/|Y| \geqslant \beta$，遍历全部记录，找出满足 Y_i 中全部属性的所有对象，并将其加入相应结点的 β 下界外延 M^{β} 中，与上一步可以合并。

(5)形成格结构

首先构造根节点和末梢结点，然后确定其他结点的前驱和后继关系，构造概念格结构。

5 应用实例

设有一个形式背景，如表 1 所列。

表 1 一个形式背景

U \ A	a	b	c	d	e
1	0	0	1	1	1
2	0	1	0	1	0
3	1	0	0	0	1
4	1	1	1	0	0

其对应的形式概念如表 2 所列。

表 2 表 1 对应的形式背景形成的形式概念

概念名称	外延	内涵	概念名称	外延	内涵
c0	φ	abcde	C5	12	d
c1	1	cde	C6	13	e
c2	2	bd	C7	14	c
c3	3	ae	C8	24	b
c4	4	abc	C9	34	a
c10	1234	φ			

构建的经典概念格如图 1 所示。

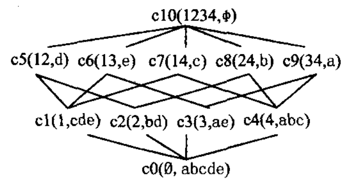

图 1 由表 1 形式背景构建的经典概念格

在此概念格中，若要查找满足 c、d、e 任一属性的结点，就要进行结点合并，即将 c5、c6、c7 合并才能找到。

形成的粗糙形式概念如表 3 所列。

表 3 表 1 对应的形式背景形成的粗糙形式概念

粗糙概念名称	上近似外延	下近似外延	内涵
c0	φ	φ	φ
c1	{12}	{12}	d
c2	{13}	{13}	e
c3	{14}	{14}	c
c4	{24}	{24}	b
c5	{34}	{34}	a
c6	{234}	{4}	ab
c7	{134}	{4}	ac
c8	{134}	{3}	ae
c9	{124}	{4}	bc
c10	{124}	{2}	bd
c11	{124}	{1}	cd
c12	{134}	{1}	ce
c13	{123}	{1}	de
c14	{1234}	{4}	abc
c15	{1234}	{1}	cde
c16	{1234}	φ	abcde

构建的粗糙概念格如图 2 所示。

在此粗糙概念格中，可同时找到满足 c、d、e 属性和满足其中一个属性的结点，即 c15。其下近似为同时满足 3 个属性的对象“1”，上近似为满足其中一个属性的对象，有“1、2、3、4”。

设 $\alpha=0.5$，$\beta=0.6$，则形成的$[\alpha,\beta]$区间概念如表 4 所列。

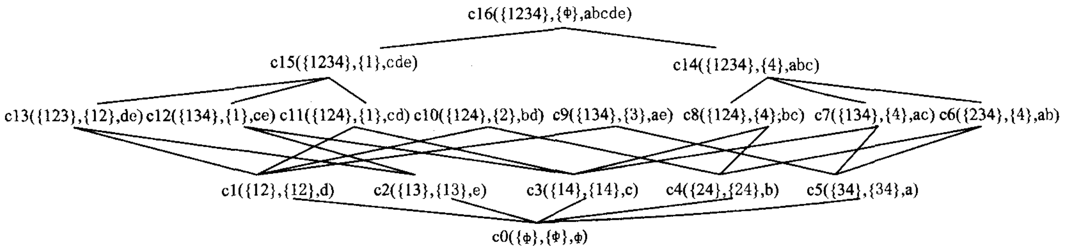

图 2　粗糙概念格

表 4　表 1 对应的形式背景形成的区间形式概念

区间概念名称	α上界外延	β下界外延	内涵
c0	Φ	Φ	Φ
c1	{12}	{12}	d
c2	{13}	{13}	e
c3	{14}	{14}	c
c4	{24}	{24}	b
c5	{34}	{34}	a
c6	{234}	{4}	ab
c7	{134}	{4}	ac
c8	{1234}	{Φ}	ad
c9	{134}	{3}	ae
c10	{124}	{4}	bc
c11	{124}	{2}	bd
c12	{1234}	{Φ}	be
c13	{124}	{1}	cd
c14	{134}	{1}	ce
c15	{123}	{1}	de
c16	{4}	{4}	abc
c17	{24}	{24}	abd
c18	{34}	{34}	abe
c19	{124}	{124}	bcd
c20	{14}	{14}	bce
c21	{1}	{1}	cde
c22	{124}	{4}	abcd
c23	{1234}	{Φ}	abde
c24	{124}	{1}	bcde
c25	{1234}	{14}	abcde

由此构造的区间概念格如图 3 所示。在此概念格中，列出了满足一定精度的区间概念，每个概念的上、下界都满足给定的精度。如结点 c24，其下近似外延“1”至少满足内涵“bcde”中的 $\beta\times4=0.6\times4=2.4\approx3$ 个属性。而上近似外延“1、2、4”至少满足内涵“bcde”中的 $\alpha\times4=0.5\times4=2$ 个属性。在实际购物系统中，可认为上近似外延为至少购买指定商品中 2 种及以上商品的客户，而下近似为至少购买 3 种及以上商品的客户。由此可对此部分客户进行重点分析，从中挖掘出潜在的客户与购买商品之间的关联规则，为商场决策提供依据。

结束语　考虑在实际应用中对对象的分析不仅仅要考虑其具备全部属性或只具备一个属性，在更多情况下，往往要求对象要具备一定精度的属性，故综合经典概念格和粗糙概念格，提出一种区间概念格，其 α 上界外延和 β 下界外延均要满足给定精度的内涵属性；定义了区间概念格，给出了其结构特征，并给出了概念度量方法——概念精度和覆盖度。证明了区间概念格是经典概念格、粗糙概念格的拓展，经典概念格、粗糙概念格是区间概念格的特例。进一步给出了区间概念格的构造方法，并通过实例说明区间概念格在实际问题中应用的有效性。下一步的研究工作主要集中在区间概念格结点特征以及基于区间概念格的关联规则提取方法，及其在实际问题中的应用方法上。

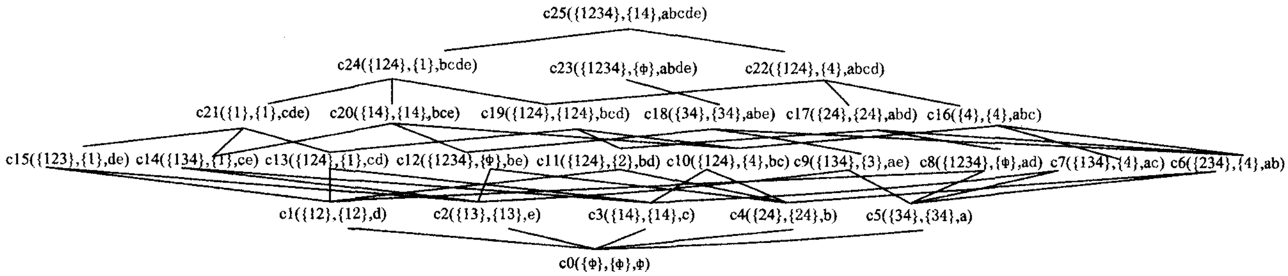

图 3　区间概念格

参考文献

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

[2] Godin R, Missaoui R, April A. Experimental compare-eison of navigation in a Galois lattice with conventional information retrieval methods[J]. International Journal of Man-machine studies, 1993, 38: 747-767

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

[4] Cole R, Stumme G. CEM-A conceptual email manager[C] // Proc 7th International Conference on Conceptual Structures (ICCS' 2000). Springer Verlag, 2000: 438-452

[5] 张春英，刘保相，郭景峰，等. 基于属性链表的关联规则格的渐进式构造算法[J]. 计算机工程与设计，2005(2): 320-322

[6] 强宇，刘宗田，林炜，等. 一种模糊概念格构造算法研究[J]. 计算机工程与应用，2004, 29: 50-53

[7] Burusco, Fuentes-Gonzlez. Construction of the L-fuzzy concept lattice [J]. Fuzzy Sets and Systems, 1998, 97(1): 109-114

[8] 谢志鹏，刘宗田. 概念格与关联规则发现[J]. 计算机研究与发展，2000, 12

[9] 仇国芳，朱朝晖. 基于经典-模糊变精度概念格的决策规则获取及其推理算法[J]. 计算机科学，2009, 36(12): 216-218

[10] 谢志鹏. 概念格及扩展模型研究[D]. 合肥：合肥工业大学计算机学院，2000: 1-20

[11] 张继福，张素兰. 加权概念格及其渐进式构造[J]. 模式识别与人工智能，2005, 18(2): 171-176

[12] 杨海峰，张继福. 粗糙概念格及构造算法[J]. 计算机工程与应用，2007, 43(24): 172-175