智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于信息熵的对象加权概念格

张晓鹤， 陈德刚， 米据生

引用本文:
张晓鹤, 陈德刚, 米据生. 基于信息熵的对象加权概念格[J]. 智能系统学报, 2020, 15(6): 1097–1103.
ZHANG Xiaohe, CHEN Degang, MI Jusheng. Object-weighted concept lattice based on information entropy[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(6): 1097–1103.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202006043

DOI: 10.11992/tis.202006043

基于信息熵的对象加权概念格

张晓鹤[1], 陈德刚[1], 米据生[2]

(1. 华北电力大学 控制与计算机工程学院, 北京 102200; 2. 河北师范大学 数学科学学院, 河北 石家庄 050024)

摘 要: 在大数据时代, 由于数据规模越来越大, 导致构造概念格的难度越来越高。在能够客观反映数据隐藏信息的前提下需删除冗余对象及属性, 降低数据规模, 构造更为简单的概念格, 从而便于用户更高效地获取知识。为避免主观因素, 本文由形式背景中属性的信息熵来获取单属性权重, 采用均值方法计算对象权重, 并用标准差计算对象重要性偏差值。通过设定的属性权重、对象权重和对象重要度偏差阈值, 构造对象加权概念格。通过实例验证了, 该方法可有效删除冗余概念, 简化概念格构造过程。

关键词: 形式背景; 概念; 信息熵; 粒计算; 概念格; 决策规则; 权值; 数据挖掘

中图分类号: TP18; O236 **文献标志码:** A **文章编号:** 1673−4785(2020)06−1097−07

中文引用格式: 张晓鹤, 陈德刚, 米据生. 基于信息熵的对象加权概念格 [J]. 智能系统学报, 2020, 15(6): 1097–1103.

英文引用格式: ZHANG Xiaohe, CHEN Degang, MI Jusheng. Object-weighted concept lattice based on information entropy[J]. CAAI transactions on intelligent systems, 2020, 15(6): 1097–1103.

Object-weighted concept lattice based on information entropy

ZHANG Xiaohe[1], CHEN Degang[1], MI Jusheng[2]

(1. School of Control and Computer Engineering, North China Electric Power University, Beijing 102200, China; 2. College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, China)

Abstract: In the era of big data, it is becoming increasingly difficult to construct concept lattices due to the increasingly large scale of data. To objectively reflect hidden information, redundant objects and attributes should be deleted and data size should be reduced to construct simple concept lattices, thus, facilitating users to acquire knowledge efficiently. In this study, to prevent subjective factors, the information entropy of an attribute in the formal context is used to obtain a single attribute weight and the attribute weight of the object is, then, calculated using the mean value method and the importance deviation of the object is calculated by standard deviation. By setting the attribute weight, object weight, and object importance deviation threshold, an object-weighted concept lattice is constructed. An example is provided to verify the effectiveness of this method in removing redundant concepts and simplifying the construction of concept lattices.

Keywords: formal context; context; information entropy; granular computing; concept lattice; decision rules; weight value; data mining

概念格理论[1]是在形式背景中进行数据分析的一个重要工具, 由 Wille 教授在 1982 年提出, 从本质上描述了对象集和属性集的内在联系。

目前, 概念格理论已在信息检索[2-4]、数据挖掘[5-7]、机器学习[8-10]等领域取得了广泛应用。吴伟志等[11]将粒计算与概念格理论进行结合, 研究了保持概念格的粒结构不变的属性约简。米据生等分别从变精度[12]和公理化[13]角度考虑概念格约简问题。针对大型数据集, 陈锦坤等[14-15]将图论理论与概念格理论进行结合, 提出了一种快速属性约简方法。邵明文等[16]从形式概念分析理论角度提出一种从信息表中提取决策规则的方法。李金海等[17-18]针对决策形式背景提出了一种新的知识认知和约简框架, 并给出约简算法, 进一步提出了保持由全体对象集构造的决策规则不

收稿日期: 2020−06−24.
基金项目: 国家自然科学基金项目 (12071131, 62076088).
通信作者: 陈德刚. E-mail: zxhzxh93@126.com.

变的情况下的属性约简方法[19]。三支形式概念分析[20]的提出丰富了概念格理论，魏玲和任睿思进一步研究了三支概念格的属性约简[21]与决策规则提取[22]。李俊余等[23]研究了基于同余关系的不协调决策形式背景的属性约简。

如何简化概念格结构，从而便于提取用户所需信息是该领域的一个重要问题。上述研究中，往往默认形式背景中所有属性重要度是相同的，然而在实际问题中，可能仅需针对特定属性进行数据挖掘，即不同属性的重要性是不同的。张继福等[24]通过对概念格的内涵引入权值，提出一种加权概念格，拓广了概念格的结构。张素兰等[25]在此基础上提出了一种基于信息熵和偏差分析的加权概念格的内涵权重获取方法。但上述两种加权概念格并没有考虑到对象权重的问题。

概念格需储存由形式背景获取的全部概念及属性间的偏序关系，使概念格构造过程变得异常困难。同时随着大数据时代的到来，需要分析的数据越来越多，如果想要将其中所有概念提取出来，并根据概念间的偏序关系构成概念格，难度进一步加大。因此，如何删除无用的属性和对象，降低构造概念格的难度，使提取的规则更加简洁就成为我们应该考虑的重要问题。本文通过信息熵给出单个属性权重后，进一步给出对象权重和重要度偏差的定义，利用 3 个阈值对冗余数据进行删除，缩小了概念格的规模，提高了构造概念格的时间效率，并且能够让决策规则的提取过程更为简洁高效。

1 预备知识

首先简单介绍形式背景中的相关知识。

定义 1 设 $F=(U,A,I)$ 为形式背景，其中 $U=\{x_1,x_2,\cdots,x_n\}$，$A=\{a_1,a_2,\cdots,a_m\}$，$I\subseteq U\times A$。如果 $(x,a)\in I$，则称 x 具有属性 a。用 $P(U)$ 表示 U 的幂集，$P(A)$ 表示 A 的幂集。$\forall X\in P(U)$，$B\in P(A)$，定义：

$$f(X)=\{a\in A:\ \forall x\in X,\ (x,a)\in I\}$$
$$g(B)=\{x\in U:\ \forall a\in B,\ (x,a)\in I\}$$

定义 2 设 $F=(U,A,I)$ 为形式背景，对于 $U'\subseteq U$，$A'\subseteq A$，可以得到形式背景 $F'=(U',A',I')$，称为 F 的子背景，其中 $I'=I\cap(U'\times C')$。

定义 3 设 $F=(U,A,I)$ 为形式背景，若二元组 $(X,B)\in P(U)\times P(A)$ 能够满足 $f(X)=B$ 且 $g(B)=X$，则 (X,B) 称为形式概念或概念。其中，X 称为外延，B 称为内涵。由形式背景 (U,A,I) 构造的概念格记为 $L(U,A,I)$。

定义 4 设 (X_1,B_1)，$(X_2,B_2)\in L(U,A,I)$，如果 $X_1\subseteq X_2$ 或者 $U_2\subseteq U_1$ 则称 (X_1,B_1) 是 (X_2,B_2) 的子概念，(X_2,B_2) 是 (X_1,B_1) 的父概念，记为

$$(X_1,B_1)\leqslant(X_2,B_2)$$

另外，两个概念的上下确界定义分别为

$$(X_1,B_1)\vee(X_2,B_2)=(gf(X_1\cup X_2),B_1\cap B_2)$$
$$(X_1,B_1)\wedge(X_2,B_2)=(X_1\cap X_2,fg(B_1\cup B_2))$$

定义 5 设 $F=(U,A,I)$ 为形式背景，其中 $U=\{x_1,x_2,\cdots,x_n\}$，$x_i\in U$，$a\in A$，则 $P(a/x_i)$ 表示对象 x 提供给属性 a 的平均信息量，即属性 a 的信息量为

$$H(a)=-\sum_{i=1}^{n}P(a/x_i)\log_2P(a/x_i)$$

信息熵值越大，表示对象集 U 具有该属性的不确定性越大，也即该属性的信息量越大，信息熵从平均意义上表示了属性的总体特性。

2 对象加权概念格

2.1 对象权重及对象重要度偏差

定义 6 设 $F=(U,A,I)$ 为形式背景，且 $U=\{x_1,x_2,\cdots,x_n\}$，$A=\{a_1,a_2,\cdots,a_m\}$，每个属性 a_i 的权重均已知，记为 $w(a_i)$。

$\forall x\in U$，$f(x)=\{a_{s_1},a_{s_2},\cdots,a_{s_t}\}$，对象 x 的权重定义为

$$d(x)=\frac{\sum_{j=1}^{t}w(a_{s_j})}{t}$$

通过定义 6 获取的对象权重能够反映该对象的总体信息量，但是却忽略了该结果可能存在偏差，不利于获得具有一定偏差的信息。

定义 7 对象 x 的重要度偏差定义为

$$D(x)=\sqrt{\frac{1}{t-1}\sum_{j=1}^{t}(w(a_{s_j})-d(x))^2}$$

由重要度偏差定义，特规定当 $t=1$ 时，有 $D(x)=0$。

通过定义对象权重和对象重要度偏差能够更全面地考虑形式背景中隐含的知识。

2.2 对象加权概念格及其构造

在无专家给定属性权重时，需通过信息熵公式求出属性对应的信息量，再进行归一化处理获得每个属性的权重。按照对象对于属性是否感兴趣，给出属性权重阈值为 $\alpha(0\leqslant\alpha\leqslant1)$，对于形式背景 $F=(U,A,I)$ 上的任意属性 $a\in A$，如果 $w(a)<\alpha$，则称该属性冗余。给定对象权重阈值为 $\beta(0\leqslant\beta\leqslant1)$，对于形式背景 (U,A,I) 上的任意对象 $x\in U$，如果 $d(x)<\beta$，则称该对象冗余。删除冗余概念及对象获取的子形式背景记为 $F_d=(U_d,A_d,I_d)$，其构造的概念格称为对象加权概念格。

设对象重要度偏差阈值为 $\delta(0\leqslant\delta\leqslant1)$，对于

形式背景 (U,A,I) 上的任意对象 $x\in U$, 如果 $D(x)>\delta$, 则称 x 为偏差对象。不包含冗余属性、冗余对象、偏差对象的子形式背景, 记为 $F_D=(U_D,A_D,I_D)$。由此构造的概念格称为对象强加权概念格。

通过删除不满足给定的属性权重阈值 α 的属性和不满足对象权重阈值 β 或对象重要度偏差阈值 δ 的对象能够节省形式背景存储空间, 从而获得简化的概念格, 提升概念格构造效率, 并且获取的概念形式更为简单, 有利于规则提取。

算法 1　对象强加权概念格的构造算法

输入　$F=(U,A,I)$, α, β, δ (其中 $0\leqslant\alpha\leqslant 1$, $0\leqslant\beta\leqslant 1$, $0\leqslant\delta\leqslant 1$);

输出　对象强加权概念格。

1) for $1\leqslant i\leqslant|A|$ do

2) $w(a_i)=\dfrac{H(a_i)}{\sum\limits_{i=1}^{|A|}H(a_i)}$

3) end for

4) initialize $d'(x_j)=0$

5) for $1\leqslant i\leqslant|A|$, $1\leqslant j\leqslant|U|$ do

6) if $(x_j,a_i)\in I$ then

7) $d'(x_j)\leftarrow w(a_i)$

8) end if

9) compute $d(x_j)=\dfrac{d'(x_j)}{|f(x_j)|}$

10) $D(x_j)=\sqrt{\dfrac{1}{|f(x_j)|-1}\sum(w(a_i)-d(x_j))^2}$

11) end for

12) if $w(a_i)<\alpha$, $d(x_j)<\beta$ or $D(x_j)>\delta$ then

13) delete x_j, a_i

14) end if

15) $F_D=(U/x_j,A/a_i,I')$

16) return $L(F_D)$

2.3　基于对象强加权概念格的决策规则提取

通过 2.2 节中给出的对象强加权概念格的构造过程, 能够帮助我们获取更为简单的概念格。如在决策形式背景中, 利用该方法对于条件概念格进行简化, 则能够大大降低概念提取的难度。

定义 8　设 (U,A,I,D,J) 为决策形式背景, 由 $R_D=\{(x,y)\in U\times U: d_l(x)=d_l(y)\ \ (\forall d_l\in D)\}$, 可产生 U 上的一个划分:

$$U/R_D=\{[x]_D: x\in U\}=\{D_1,D_2,\cdots,D_r\}$$

其中: $[x]_D=\{y\in U:(x,y)\in R_D\}$; $\forall d_l\in D$。如 $x,y\in D_k$, 必有 $d_l(x)=d_l(y)$, 显然属于同一决策类的对象具有相同的决策值, 记 $T_k=\{d_1(D_k),d_2(D_k),\cdots,d_{|D|}(D_k)\}$ 为 D_k 的决策值。(D_k,T_k) 称为决策概念。

设 (U,A,I,D,J) 为决策形式背景, 则由形式背景 $F=(U,A,I)$ 能够获得子形式背景 $F_D=(U_D,A_D,I_D)$, 进一步可构造对象强加权概念格, 记为 $L(U_D,A_D,I_D)$。称 $(X,B)\in L(U_D,A_D,I_D)$ 为条件概念。

X、B均为非空集合, 如果有 $\dfrac{|X\cap D_k|}{|X|}=1$, 则称 $(X,B)\Rightarrow(D_k,T_k)$ 为决策形式背景的粒决策规则。简记为 $B\Rightarrow T_k$。

下面给出具体算法。

算法 2　粒决策规则获取算法

输入　(U,A,I,D,J)

输出　粒决策规则

1) compute $L(F_D)$

2) $\forall x,y\in U$

3) for $1\leqslant k\leqslant|D|$ do

4) if $d_k(x)=d_k(y)$ then

5) $(x,y)\in R_D$

6) end if

7) $U/R_D=\{[x]_D: x\in U\}=\{D_1,D_2,\cdots,D_r\}$

8) end for

9) compute (D_t,T_t)

10) for all $(X,B)\in L(F_D)$, $1\leqslant t\leqslant r$ do

11) if $\dfrac{|X\cap D_t|}{|X|}=1$ then

12) $(X,B)\Rightarrow(D_t,T_t)$

13) end if

3　实验和分析

3.1　应用举例

由表 1 给出人才市场大学生就业信息, 处理后构成形式背景 $F=(U,A,I)$, 其中对象集为 6 位求职者。属性集为 $A=\{a,b,c,d,e\}$, 分别代表 5 个属性, 即应届生、数学专业、教师资格证、英语四级、计算机二级。

表 1　形式背景 $F=(U,A,I)$

Table 1　**Formal context**: $F=(U,A,I)$

序号	a	b	c	d	e
1	1	1	—	—	—
2	1	—	—	1	1
3	1	1	1	—	1
4	1	—	—	—	—
5	—	1	—	—	1
6	1	1	—	1	—

1) 直接构造概念格

由表 1 给出的形式背景直接构造的概念格如图 1 所示, 共 12 个概念结点。

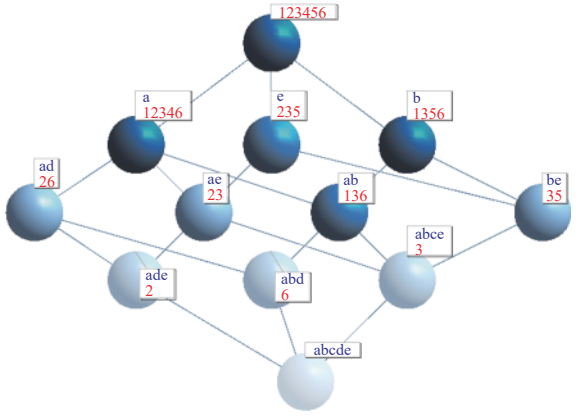

图 1　由 $F=(U, A, I)$ 构造的概念格

Fig. 1　Concept lattice of $F=(U,A,I)$

2) 对象加权概念格

给定 $\alpha=0.14, \beta=0.14$。在表 1 给定的形式背景中通过信息熵获取属性权重。由表 1 可获得属性 a 出现的概率为$P_1(a)=\dfrac{5}{6}$, 利用定义 5 可得属性 a 的信息熵:

$$H_1(a)=-P_1(a)\times\log_2 P_1(a)=\ =-\frac{5}{6}\times\log_2\frac{5}{6}=0.219\,2$$

由计算可知:

$$P_1(b)=\frac{4}{6},\ H_1(b)\ =-\frac{4}{6}\times\log_2\frac{4}{6}=0.389\,9$$

$$P_1(c)=\frac{1}{6},\ H_1(c)\ =-\frac{1}{6}\times\log_2\frac{1}{6}=0.430\,8$$

$$P_1(d)=\frac{2}{6},\ H_1(d)\ =-\frac{2}{6}\times\log_2\frac{2}{6}=0.528\,3$$

$$P_1(e)=\frac{3}{6},\ H_1(e)\ =-\frac{3}{6}\times\log_2\frac{3}{6}=0.5$$

$$H(A)=H_1(a)+H_1(b)+H_1(c)+H_1(d)+H_1(e)=2.068\,2$$

可获得各属性权重, 分别为: $w_1(a)$= 0.11(小于 α, 故该属性冗余需删去), $w_1(b)$= 0.19, $w_1(c)$= 0.21, $w_1(d)$= 0.25, $w_1(e)$ = 0.24 。

对于对象 1 来说, $f(1)=\{a,b\}$。则

$$d_1(1)=\frac{w_1(a)+w_1(b)}{2}=0.15$$

$$d_1(2)=\frac{w_1(a)+w_1(d)+w_1(e)}{3}=0.2$$

$$d_1(3)=\frac{w_1(a)+w_1(b)+w_1(c)+w_1(e)}{4}=0.187\,5$$

$d_1(4)=w_1(a)=0.11$(小于 β, 故该对象冗余需删去)

$$d_1(5)=\frac{w_1(b)+w_1(e)}{2}=0.215$$

$$d_1(6)=\frac{w_1(a)+w_1(b)+w_1(d)}{3}=0.183\,3$$

则可获得子形式背景 $F_d=(U_d,A_d,I_d)$, 其中 $U_d=\{1,2,3,5,6\}$, $A_d=\{b,c,d,e\}$, 具体数据如表 2 所示。

表 2　形式背景 $F_d=(U_d,A_d,I_d)$

Table 2　Formal context: $F_d=(U_d,A_d,I_d)$

序号	b	c	d	e
1	1	—	—	—
2	—	—	1	1
3	1	1	—	1
5	1	—	—	1
6	1	—	1	—

由表 2 给出的形式背景构造的概念格如图 2 所示, 共 9 个概念结点, 即为对象加权概念格。

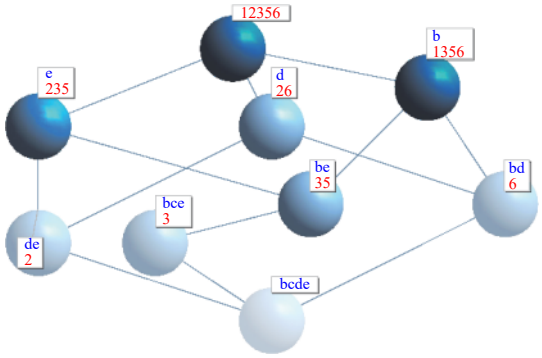

图 2　由 $F_d=(U_d,A_d,I_d)$ 构造的概念格

Fig. 2　Concept lattice of $F_d=(U_d,A_d,I_d)$

3) 对象强加权概念格

给定 $\delta=0.075$, 则通过计算可知

$$D_1(1)=\sqrt{(w_1(a)-d_1(1))^2+(w_1(b)-d_1(1))^2}=0.057$$

$$D_1(2)=\sqrt{\frac{(w_1(a)-d_1(2))^2+(w_1(d)-d_1(2))^2+(w_1(e)-d_1(2))^2}{3-1}}=0.078$$

由于 $D_1(2)>\delta$= 0.075, 则 2 为偏差对象, 需删除。

$$D_1(3)=\sqrt{\frac{(w_1(a)-d_1(3))^2+(w_1(b)-d_1(3))^2+(w_1(c)-d_1(3))^2+(w_1(e)-d_1(3))^2}{4-1}}=0.056$$

$$D_1(5)=\sqrt{(w_1(b)-d_1(5))^2+(w_1(e)-d_1(5))^2}=0.035$$

$$D_1(6)=\sqrt{\frac{(w_1(a)-d_1(6))^2+(w_1(b)-d(6))^2+(w_1(d)-d_1(6))^2}{3-1}}=0.07$$

则可获得子形式背景 $F_D=(U_D,A_D,I_D)$, 其中: U_D= {1,3,5,6}, $A_D=\{b,c,d,e\}$, 详见表 3。

表 3　形式背景 $F_D=(U_D,A_D,I_D)$
Table 3　Formal context: $F_D=(U_D,A_D,I_D)$

序号	b	c	d	e
1	1	—	—	—
3	1	1	—	1
5	1	—	—	1
6	1	—	1	—

表 3 给出的形式背景构造的概念格如图 3 所示, 共 5 个概念结点, 即为对象强加权概念格。

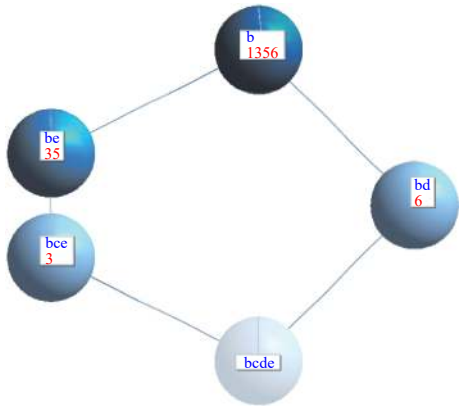

图 3　由 $F_D=(U_D,A_D,I_D)$ 构造的概念格
Fig. 3　Concept lattice of $F_D=(U_D,A_D,I_D)$

通过上述计算能够看到本文提出的对象强加权概念格能够通过删除不必要的对象和属性, 缩小形式背景的规模, 从而实现简化概念格构造过程的目的。且通过这种方法能够简化概念形式, 便于我们在之后进行规则提取。

4) 粒决策规则提取

在表 1 的基础上给出决策形式背景 (U,A,I,D,J), 其中对象集为 6 位求职者。条件属性集仍为 $A=\{a,b,c,d,e\}$, 决策属性集为 $D=\{f\}$, 取值为 1 代表教育机构工作录取该求职者, 否则为拒绝录取。具体数据如表 4所示。

表 4　决策形式背景 (U,A,I,D,J)
Table 4　Decision formal context: (U,A,I,D,J)

序号	a	b	c	d	e	f
1	1	1	—	—	—	—
2	1	—	—	1	1	—
3	1	1	1	—	1	1
4	1	—	—	—	—	—
5	—	1	—	—	1	1
6	1	1	—	1	—	—

由表 4 可知 $U/R_D=\{D_1,D_2\}$, $D_1=\{1,2,4,6\}$, $D_2=\{3,5\}$。有两个决策概念: $(\{1,2,4,6\},\{0\})$ 和 $(\{3,5\},\{1\})$。

由上述计算过程可直接获取 $L(U_D,A_D,I_D)$, 共有 5 个概念, 分别为: $(\{1,3,5,6\},\{b\})$、$(\{6\},\{b,d\})$、$(\{3,5\},\{b,e\})$、$(\{3\},\{b,c,e\})$、$(\varnothing,\{b,c,d,e\})$。

由算法 2 可获得粒决策规则:

$$\{b,e\}\Rightarrow\{1\}$$
$$\{b,c,e\}\Rightarrow\{1\}$$
$$\{b,d\}\Rightarrow\{0\}$$

可进一步简化规则得:

$$\{b,e\}\Rightarrow\{1\}$$
$$\{b,d\}\Rightarrow\{0\}$$

通过本文算法对于概念格进行简化后提取规则, 可知本岗位更倾向于招聘数学专业且拥有教师资格证的应聘者, 与现实背景相符。

对于没有删除冗余属性、冗余对象、偏差对象的情况, 也即直接用表 2 所示的决策形式背景获取条件概念格, 然后利用算法 2 获取粒决策规则, 结果如下:

$$\{a,d\}\Rightarrow\{0\}$$
$$\{b,e\}\Rightarrow\{1\}$$
$$\{a,d,e\}\Rightarrow\{0\}$$
$$\{a,b,d\}\Rightarrow\{0\}$$
$$\{a,b,c,e\}\Rightarrow\{1\}$$

进一步简化规则可得

$$\{a,d\}\Rightarrow\{0\}$$
$$\{b,e\}\Rightarrow\{1\}$$

可知本岗位更倾向于招聘拥有教师资格证的数学专业的应聘者, 这也是符合实际情况的。可见, 通过本文构造的对象强加权概念格仍能够保持原决策规则的关键信息, 即教育机构更愿意录取数学专业有教师资格证的应聘者, 而且能够简化规则获取的过程。

3.2　实验分析

在内存为 8 GB, 操作系统为 Windows 10 的计算机上, 用 Matlab 软件实现了本文算法、经典概念格构造算法及文献 [25] 的算法。选用 UCI 数据集中的 ZOO 数据集进行实验, 该数据集共有 101 条记录, 18 项属性。对数据集的预处理, 包括去掉决策属性, 将数值属性布尔化等, 最后获取了共 100 条记录, 20 项属性的形式背景, 以 20 条动物记录为单位将其划分为 5 个子形式背景。将子形式背景依次进行合并, 分别采用本文算法、经典算法及文献 [25] 的算法构造合并后的整体概念格, 对比其分别获取的概念结点个数, 以此对比其执行效率, 具体如表 5 所示。

表 5　3 种算法执行效率对比

Table 5　Comparison of the execution efficiencies of the three algorithms

动物记录/条	经典算法格结点数	文献[25]算法格结点数	本文算法格结点数
20	65	62	34
40	136	122	63
60	187	139	69
80	312	227	47
100	353	250	68

由实验结果可知，记录数逐渐增多，概念结点也会随之递增。由于经典算法在构造概念格的过程不会删除任何信息，因此该算法会获得最多的概念结点，同样会导致构造时间最长，执行效率最低。而文献 [25] 中的算法虽然减少了 19.4% 的冗余结点，但其构造的概念格规模仍然较为复杂。本文算法相较于经典算法减少了 66% 的冗余结点，相较于文献 [25] 中的算法减少了 59.2% 的冗余结点，极大缩短了构造概念格的时间，提升了概念格的构造效率，有利于提取用户关心的知识信息。

4　结束语

删除冗余概念，提升概念格构造效率是概念格理论的重要问题。本文通过信息熵获取了形式背景中数据隐含的属性权重，利用属性权重给出了对象权重及对象重要度偏差，并进一步给出满足阈值要求的对象强加权概念格的构造算法。本文的方法能够有效避免产生冗余概念，获得结构更简单的概念格，同时也能够进一步简化规则获取过程。

但本文仍有许多不足，比如在给出对象权重时采取的是均值法，把对象具有的属性作为个体去考虑，并没有将考虑对象具有的属性作为一个整体具有的特征。而且本文对于 2.3 节中给出的决策规则时仅从条件属性集角度出发构造了对象强加权概念格，并没有充分考虑决策属性隐含的信息。未来将进一步研究这些问题。

参考文献：

[1] WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts[M]//RIVAL I. Ordered Sets. Dordrecht: Springer, 1982: 445−470.

[2] SUTTON A, MALETIC J I. Recovering UML class models from C++: a detailed explanation[J]. Information and software technology, 2007, 49(3): 212–229.

[3] 黄微, 高俊峰. 基于概念格的 Web 学术信息搜索结果的二次组织 [J]. 现代图书情报技术, 2010(5): 8–12.
HUANG Wei, GAO Junfeng. A second organization of academic retrieved results based on concept lattice[J]. New technology of library and information service, 2010(5): 8–12.

[4] 沈夏炯, 叶曼曼, 甘甜, 等. 基于概念格的信息检索及其树形可视化 [J]. 计算机工程与应用, 2017, 53(3): 95–99.
SHEN Xiajiong, YE Manman, GAN Tian, et al. Information retrieval based on concept lattice and its tree visualization[J]. Computer engineering and applications, 2017, 53(3): 95–99.

[5] MISSAOUI R, GODIN R, BOUJENOUI A. Extracting exact and approximate rules from databases[C]//Proceedings of SOFTEKS Workshop on Incompleteness and Uncertainty in Information Systems. Berlin, Heidelberg: Springer, 1993: 209−222.

[6] 康向平, 苗夺谦. 一种基于概念格的集值信息系统中的知识获取方法 [J]. 智能系统学报, 2016, 11(3): 287–293.
KANG Xiangping, MIAO Duoqian. A knowledge acquisition method based on concept lattice in set-valued information systems[J]. CAAI transactions on intelligent systems, 2016, 11(3): 287–293.

[7] LIU Yong, KANG Xiangping, MIAO Duoqian, et al. A knowledge acquisition method based on concept lattice and inclusion degree for ordered information systems[J]. International journal of machine learning and cybernetics, 2019, 10(11): 3245–3261.

[8] 张功亮, 陈钰, 周茜, 等. 基于领域本体的信息语义相关检索 [J]. 计算机工程, 2011, 37(20): 33–35, 38.
ZHANG Gongliang, CHEN Yu, ZHOU Xi, et al. Information semantic relativity retrieval based on domain ontology[J]. Computer engineering, 2011, 37(20): 33–35, 38.

[9] 刘保相, 孟肖丽. 基于关联分析的气象云图识别问题研究 [J]. 智能系统学报, 2014, 9(5): 595–601.
LIU Baoxiang, MENG Xiaoli. The study on nephogram recognition based on relational analysis[J]. CAAI transactions on intelligent systems, 2014, 9(5): 595–601.

[10] 陈湘, 吴跃. 基于概念格挖掘 GIS 中的关联规则 [J]. 计算机应用, 2011, 31(3): 686–689.
CHEN Xiang, WU Yue. Mining association rules of geographic information system based on concept lattice[J]. Journal of computer applications, 2011, 31(3): 686–689.

[11] WU Weizhi, LEUNG Y, MI Jusheng. Granular computing and knowledge reduction in formal contexts[J]. IEEE transactions on knowledge and data engineering, 2009, 21(10): 1461–1474.

[12] MI Jusheng, WU Weizhi, ZHANG Wenxiu. Approaches

to knowledge reduction based on variable precision rough set model[J]. Information sciences, 2004, 159(3−4): 255–272.

[13] MI Jusheng, LEUNG Y, WU Weizhi. Approaches to attribute reduction in concept lattices induced by axialities[J]. Knowledge-based systems, 2010, 23(6): 504–511.

[14] CHEN Jinkun, LI Jinjin. An application of rough sets to graph theory[J]. Information sciences, 2012, 201: 114–127.

[15] CHEN Jinkun, MI Jusheng, XIE Bin, et al. A fast attribute reduction method for large formal decision contexts[J]. International journal of approximate reasoning, 2019, 106: 1–17.

[16] SHAO Mingwen, LEUNG Y, WU Weizhi. Rule acquisition and complexity reduction in formal decision contexts[J]. International journal of approximate reasoning, 2014, 55(1): 259–274.

[17] LI Jinhai, MEI Changlin, LV Yuejin. Knowledge reduction in decision formal contexts[J]. Knowledge-based systems, 2011, 24(5): 709–715.

[18] LI Jinhai, MEI Changlin, LV Yuejin. Knowledge reduction in real decision formal contexts[J]. Information sciences, 2012, 189: 191–207.

[19] LI Jinhai, MEI Changlin, WANG Junhong, et al. Rule-preserved object compression in formal decision contexts using concept lattices[J]. Knowledge-based systems, 2014, 71: 435–445.

[20] QI Jianjun, WEI Ling, YAO Yiyu. Three-way formal concept analysis[C]//Proceedings of the 9th International Conference on Rough Sets and Knowledge Technology. Shanghai: Springer, 2014: 732−741.

[21] REN Ruisi, WEI Ling. The attribute reductions of three-way concept lattices[J]. Knowledge-based systems, 2016, 99: 92–102.

[22] WEI Ling, LIU Lin, QI Jianjun, et al. Rules acquisition of formal decision contexts based on three-way concept lattices[J]. Information sciences, 2020, 516: 529–544.

[23] LI Junyu, WANG Xia, WU Weizhi, et al. Attribute reduction in inconsistent formal decision contexts based on congruence relations[J]. International journal of machine learning and cybernetics, 2017, 8(1): 81–94.

[24] 张继福, 张素兰, 郑链. 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.
ZHANG Jifu, ZHANG Sulan, ZHENG Lian. Weighted concept lattice and incremental construction[J]. Pattern recognition and artificial intelligence, 2005, 18(2): 171–176.

[25] 张素兰, 郭平, 张继福. 基于信息熵和偏差的加权概念格内涵权值获取 [J]. 北京理工大学学报, 2011, 31(1): 59–63.
ZHANG Sulan, GUO Ping, ZHANG Jifu. Intension weight value acquisition of weighted concept lattice based on information entropy and deviance[J]. Transactions of Beijing Institute of Technology, 2011, 31(1): 59–63.

作者简介:

张晓鹤, 博士研究生, 主要研究方向为概念格、关联规则挖掘。

陈德刚, 教授, 博士生导师, 主要研究方向为机器学习、数据挖掘。完成自然科学基金面上项目 3 项、数学天元基金 1 项, 参加 973 课题 1 项。发表学术论文 150 余篇。

米据生, 教授, 博士生导师, 主要研究方向为粗糙集、粒计算、概念格、数据挖掘与近似推理。主持国家自然科学基金项目 3 项, 教育部博士点基金项目 1 项。获得省级自然科学奖 3 项, 发表学术论文 130 余篇。