D型概率决策形式背景下的规则获取

赵 凡 魏 玲
(西北大学数学学院 西安 710127)

摘 要 基于不确定性决策问题,提出一种D型概率决策形式背景,并针对D型概率决策形式背景定义"△"算子,获得概率形式概念,构造相应的概念格。又定义了D型概率决策形式背景的协调性,在协调的背景上进行规则获取。进一步,剔除冗余规则,简化规则集。最后,给出概率概念格生成及规则获取算法,以便于计算机的实现。

关键词 不确定性决策,D型概率决策形式背景,规则获取,冗余规则
中图法分类号 O29,TP18 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.08.047

Rule Acquisition of D-type Probabilistic Decision Formal Context

ZHAO Fan WEI Ling
(School of Mathematics, Northwest University, Xi'an 710127, China)

Abstract Based on uncertainty decision problems, D-type probabilistic decision formal context was proposed. On this probabilistic decision formal context, a new operator "△" was defined, the probabilistic concepts were obtained and the corresponding concept lattice was constructed. Then we defined the consistence of D-type probability decision formal context and studied the rule acquisition on the consistent D-type probability decision formal context. Furthermore, by eliminating the redundant rules, the rules were simplified. Finally, the algorithms for generating the probability concept lattice and acquiring the decision rules were presented.

Keywords Uncertainty decision, D-type probabilistic decision formal context, Rule acquisition, Redundant rule

1 引言

随着科技的飞速发展,大数据时代已经到来,如何更加方便、快速地从海量数据中筛选出有用的知识,一直是人们努力的方向。知识发现是从数据集中识别正确、新颖、有潜在应用价值以及最终可为人们理解的模式的过程[1]。形式概念分析由德国数学家Wille R.教授[2]于1982年提出,是一种重要、有效的知识发现方法,它的两个重要内容是形式背景和形式概念。形式背景表征了对象和属性之间的二元关系,基于该二元关系,可以生成形式概念。由一个形式背景的所有形式概念和它们之间的泛化-特化关系能够得到一个完备格,称为概念格。到目前为止,国内外学者对形式概念分析理论已经开展了多角度、深层次的研究,主要有概念格的构造[3-4]、概念格模型的推广[5]、概念格的约简[6-8]、概念格的规则提取[9-10]以及概念格的应用[11-12]等。

除了研究一般意义上的形式背景,张文修、魏玲提出用概念格理论研究决策形式背景[6]。李金海等研究了基于概念格的决策形式背景属性约简及规则提取问题[9]。此外,也有各种扩展决策形式背景不断被提出,例如:模糊决策形式背景、实值决策形式背景、不完备决策形式背景、随机决策形式背景等[13-16]。在不同的决策形式背景下,概念格构造、规则提取、属性约简都是研究的热点问题[17-18]。

在实际生活中,对象与决策属性之间的关系有时会带有一定的不确定性,例如在医学中,医生给病人诊病时是根据病人的临床表现来判断病人是否患了某种疾病,而临床表现与疾病的发生之间并不是完全确定的关系。如果利用概率表示这种不确定性,那么这样的决策形式背景就是本文提出的D型概率决策形式背景。本文借助概念格理论知识,在D型概率决策形式背景上定义"△"算子,获得概率形式概念,并构造相应的概念格。又利用包含度的概念,定义D型概率决策形式背景的协调性,在协调的背景上获取决策规则。考虑到在一般实际应用中,概率概念格规模较大,获取的规则较多,可能会产生冗余规则,因此通过剔除冗余规则,使规则集更加简洁、紧凑。最后,本文给出了概率概念格生成及决策规则获取算法,以便于计算机实现。

2 相关概念

2.1 形式概念分析相关概念

定义1[19] 称(U,A,I)为一个形式背景,其中$U=\{x_1,x_2,\cdots,x_n\}$为对象集,称每个$x_i(i\leqslant n)$为一个对象;非空集合$A=\{a_1,a_2,\cdots,a_m\}$为属性集,称每个$a_j(j\leqslant m)$为一个属性;I为U到A上的二元关系,$I\subseteq U\otimes A$。

到稿日期:2016-07-07 返修日期:2016-09-28 本文受国家自然科学基金(11371014,11071281)资助。
赵 凡(1992-),女,硕士生,主要研究方向为形式概念分析、粗糙集,E-mail:18220569448@163.com;魏 玲(1972-),女,教授,博士生导师,主要研究方向为形式概念分析、粗糙集,E-mail:wl@nwu.edu.cn(通信作者)。

在形式背景(U,A,I)中，若$(x,a)\in I$，则称对象x具有属性a，记为xIa，用 1 表示xIa，用 0 表示$x\overline{I}a$。这样，形式背景就可以表示为只有 0 和 1 的表格。

定义 2[19]　设(U,A,I)为一个形式背景，在对象集$X\subseteq U$和属性集$B\subseteq A$上分别定义运算：

$$X^*=\{a\mid a\in A,\forall x\in X,(x,a)\in I\}$$

$$B^*=\{x\mid x\in U,\forall a\in B,(x,a)\in I\}$$

其中，X^*表示X中所有对象共同具有的属性集合，B^*表示具有B中所有属性的对象集合。

根据定义 2 可得形式概念的定义。

定义 3[19]　设在形式背景(U,A,I)上，如果一个二元组(X,B)满足$X^*=B$且$X=B^*$，则称(X,B)是一个形式概念，简称概念。其中，X称为概念的外延，B称为概念的内涵。

用$L(U,A,I)$表示形式背景(U,A,I)的全体概念，$(X_1,B_1),(X_2,B_2)\in L(U,A,I)$，记$(X_1,B_1)\leqslant(X_2,B_2)\Leftrightarrow X_1\subseteq X_2(\Leftrightarrow B_1\supseteq B_2)$，则“$\leqslant$”是$L(U,A,I)$上的偏序关系。并且$(X_1,B_1)\wedge(X_2,B_2)=(X_1\cap X_2,(B_1\cup B_2)^{**})$，$(X_1,B_1)\vee(X_2,B_2)=((X_1\cup X_2)^{**},B_1\cap B_2)$也是概念，从而$L(U,A,I)$是完备格，也是概念格。

为了与后文的概率概念格区别，本文称上述概念格为经典概念格。

当属性集中包含条件属性和决策属性两类属性时，称这样的形式背景为决策形式背景。

定义 4[20]　设(U,A,I)与(U,C,J)是两个形式背景，有相同的论域，则称(U,A,I,C,J)为决策形式背景，称$L(U,A,I)$为条件概念格，称$L(U,C,J)$为决策概念格。

定义 5[20]　设$L(U,A_1,I_1)$和$L(U,A_2,I_2)$是两个概念格，如果$\forall(X_1,B_1)\in L(U,A_1,I_1)$，$\exists(X_2,B_2)\in L(U,A_2,I_2)$，使得$X_1=X_2$，则称$L(U,A_1,I_1)$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$L(U,A_2,I_2)\leqslant L(U,A_1,I_1)$，那么两个概念格同构，记作$L(U,A_1,I_1)\cong L(U,A_2,I_2)$。

2.2　包含度

在某些情况下，集合之间不再是包含关系，而是满足某种包含程度，称这种包含程度为包含度。

定义 6[21]　设U为有限论域，对于任意$X,Y\subseteq U$，称$D(Y|X)$为包含度，若满足以下条件：

(1)$0\leqslant D(Y|X)\leqslant 1$；

(2)当$X\subseteq Y$时，$D(Y|X)=1$；

(3)当$X\subseteq Y\subseteq Z$时，$D(X|Z)\leqslant D(X|Y)$。

显然，$D(Y|X)=\dfrac{|X\cap Y|}{|X|}$是一个包含度。

3　D 型概率决策形式背景

在实际生活中，有时会出现对象与条件属性之间的关系是确定的，但与决策属性之间的关系带有一定的不确定性，比如，医生根据病人有“流鼻涕”这一症状判断病人患感冒的概率为 0.7，而并不能确定一定患感冒。如果利用概率表示这种不确定性，那么称这样的决策形式背景是 D 型概率决策形式背景。本节针对 D 型概率决策形式背景定义“△”算子，获得概率形式概念，并构造相应的概念格。

3.1　D 型概率决策形式背景定义

定义 7　称(U,A,I,D,J)是一个 D 型概率决策形式背景，其中$U=\{x_1,x_2,\cdots,x_n\}$为对象集，称每个$x_i(i\leqslant n)$为一个对象；$A=\{a_1,a_2,\cdots,a_m\}$为条件属性集，称每个$a_j(j\leqslant m)$为一个条件属性；I为U到A上的二元关系，若$(x,a)\in I$，则称对象x具有属性a，用 1 表示，否则，称对象x不具有属性a，用 0 表示；$D=\{d_1,d_2,\cdots,d_k\}$为决策属性集，称每个$d_l(l\leqslant k)$为一个决策属性；J为U到D上的关系集，$J=\{d_l:U\to[0,1]\}(l\leqslant k)$，$d_l(x_i)=p\in[0,1]$表示对象$x_i$具有决策属性$d_l$的概率为$p$。

例 1　设(U,A,I,D,J)是一个 D 型概率决策形式背景，如表 1 所列。其中，对象集$U=\{x_1,x_2,x_3,x_4\}$表示 4 位病患，条件属性集$A=\{a_1,a_2,a_3\}$表示 3 种临床症状，决策属性集$D=\{d_1,d_2\}$表示 2 种疾病（各决策属性并非全部可能的决策结果，且无法保证相互之间的独立性，因此每行概率之和不一定为 1）。

表 1　D 型概率决策形式背景(U,A,I,D,J)

U	a_1	a_2	a_3	d_1	d_2
x_1	1	1	0	0.7	0.2
x_2	0	1	1	0.1	0.6
x_3	1	0	1	0.5	0.9
x_4	0	1	0	0.1	0.3

其中，$d_1(x_1)=0.7$表示病患x_1患疾病d_1的概率为 0.7。

3.2　概率形式概念

在实际生活中存在这种情况：有些事情出现的结果只有达到一定程度时，才能够引起人们的重视。例如在医学中，如果根据食欲增大这一症状判断病人患甲亢的概率为 0.8，则病人需要及时进行相关的检查与治疗；如果根据该症状判断患甲亢的概率仅为 0.1，则基本可以排除患此病。因此，不妨设一个参考值α，只关注决策值超过参数α的情况。

根据定义 7，定义“△”运算如下。

定义 8　设(U,A,I,D,J)是一个 D 型概率决策形式背景，$X\subseteq U,B\subseteq A,C\subseteq D,\alpha\in[0,1]$是参数。在对象集$X$和属性集$B\dot{\cup}C\subseteq A\dot{\cup}D$上分别定义“△”运算：

$$X^{\triangle}=\{a\dot{\cup}\{d_l\}\mid a\dot{\cup}\{d_l\}\in A\dot{\cup}D,\forall x\in X,(x,a)\in I\text{ 且 }d_l(x)\geqslant\alpha\}$$

$$(B\dot{\cup}C)^{\triangle}=\{x\mid x\in U,\forall a\dot{\cup}\{d_l\}\in B\dot{\cup}C,(x,a)\in I\}$$

定义 9　设(U,A,I,D,J)是一个 D 型概率决策形式背景，对一个二元组$(X,B\dot{\cup}C)$，如果满足$X^{\triangle}=B\dot{\cup}C$且$(B\dot{\cup}C)^{\triangle}=X$，则称$(X,B\dot{\cup}C)$是一个概率形式概念，简称概率概念。$X$称为概率概念的外延，$B\dot{\cup}C$称为概率概念的内涵，其中$B$称为条件内涵，$C$称为决策内涵。

记$L(U,A,I,D,J)$表示形式背景(U,A,I,D,J)上的全体概率概念。

性质 1　设(U,A,I,D,J)是一个 D 型概率决策形式背景，$\forall X,X_1,X_2\subseteq U,B\dot{\cup}C,B_1\dot{\cup}C_1,B_2\dot{\cup}C_2\subseteq A\dot{\cup}D$，以下性质成立：

(1) $X_1\subseteq X_2\Rightarrow X_2^{\triangle}\subseteq X_1^{\triangle}$, $B_1\dot{\cup}C_1\subseteq B_2\dot{\cup}C_2\Rightarrow(B_2\dot{\cup}C_2)^{\triangle}\subseteq(B_1\dot{\cup}C_1)^{\triangle}$;

(2) $X\subseteq X^{\triangle\triangle}$, $B\dot{\cup}C\subseteq(B\dot{\cup}C)^{\triangle\triangle}$;

(3) $X^{\triangle}=X^{\triangle\triangle\triangle}$, $(B\dot{\cup}C)^{\triangle}=(B\dot{\cup}C)^{\triangle\triangle\triangle}$;

(4) $X\subseteq(B\dot{\cup}C)^{\triangle}\Leftrightarrow B\dot{\cup}C\subseteq X^{\triangle}$;

(5) $(X_1\cup X_2)^{\triangle}=X_1^{\triangle}\cap X_2^{\triangle}$, $(C_1\cup C_2)^{\triangle}=C_1^{\triangle}\cap C_2^{\triangle}$;

(6) $(X_1\cap X_2)^{\triangle}\supseteq X_1^{\triangle}\cup X_2^{\triangle}$, $(C_1\cap C_2)^{\triangle}\supseteq C_1^{\triangle}\cup C_2^{\triangle}$;

(7) $(X^{\triangle\triangle},X^{\triangle})$ 和 $((B\dot{\cup}C)^{\triangle},(B\dot{\cup}C)^{\triangle\triangle})$ 都是概率概念。

证明:(1)与(2)显然成立。

由(1)和(2)可以证明 $X^{\triangle\triangle\triangle}\subseteq X^{\triangle}$,再令 $X^{\triangle}$ 代替 X,则由(2)可以证明 $X^{\triangle}\subseteq X^{\triangle\triangle\triangle}$,于是 $X^{\triangle}=X^{\triangle\triangle\triangle}$,(3)即证。同理可证(4)-(7)。

例 2(续例 1) 不妨取 $\alpha=0.5$,由定义 8 和定义 9 有:$\{x_2,x_3\}^{\triangle}=\{a\dot{\cup}\{d_l\}\mid a\dot{\cup}\{d_l\}\in A\dot{\cup}D,\forall x\in\{x_2,x_3\},(x,a)\in I$ 且 $d_l(x)\geqslant\alpha\}=\{a_3\dot{\cup}d_2\}$, $\{a_3\dot{\cup}d_2\}^{\triangle}=\{x\mid x\in U,\forall a\dot{\cup}\{d_l\}\in\{a_3\dot{\cup}d_2\},(x,a)\in I\}=\{x_2,x_3\}$,于是 $(x_2x_3,a_3\dot{\cup}d_2)$ 是一个概率概念。所有概率概念为 $L(U,A,I,D,J)=\{(U,\emptyset\dot{\cup}\emptyset),(x_1,a_1a_2\dot{\cup}d_1),(x_2,a_2a_3\dot{\cup}d_2),(x_3,a_1a_3\dot{\cup}d_1d_2),(x_1x_3,a_1\dot{\cup}d_1),(x_2x_3,a_3\dot{\cup}d_2),(x_1x_2x_4,a_2\dot{\cup}\emptyset),(\emptyset,A\dot{\cup}D)\}$。

3.3 概率概念格

定义 10 设 (U,A,I,D,J) 是一个 D 型概率决策形式背景,$X_1,X_2\subseteq U$, $B_1\dot{\cup}C_1,B_2\dot{\cup}C_2\subseteq A\dot{\cup}D$,如果 $(X_1,B_1\dot{\cup}C_1)\leqslant(X_2,B_2\dot{\cup}C_2)\Leftrightarrow X_1\subseteq X_2(\Leftrightarrow B_1\dot{\cup}C_1\supseteq B_2\dot{\cup}C_2)$,则"$\leqslant$"是 $L(U,A,I,D,J)$ 上的概率偏序关系。

定义 11 设 $(X_1,B_1\dot{\cup}C_1),(X_2,B_2\dot{\cup}C_2)$ 是 D 型概率决策形式背景 (U,A,I,D,J) 上的两个概率概念,如果 $(X_1,B_1\dot{\cup}C_1)\leqslant(X_2,B_2\dot{\cup}C_2)$,且二者之间不存在其他概率概念 $(X_3,B_3\dot{\cup}C_3)$,满足 $(X_1,B_1\dot{\cup}C_1)\leqslant(X_3,B_3\dot{\cup}C_3)\leqslant(X_2,B_2\dot{\cup}C_2)$,则称 $(X_1,B_1\dot{\cup}C_1)$ 是 $(X_2,B_2\dot{\cup}C_2)$ 的子节点,$(X_2,B_2\dot{\cup}C_2)$ 是 $(X_1,B_1\dot{\cup}C_1)$ 的父节点。

定理 1 若 $(X_1,B_1\dot{\cup}C_1),(X_2,B_2\dot{\cup}C_2)$ 是 D 型概率决策形式背景 (U,A,I,D,J) 上的两个概率概念,则 $(X_1,B_1\dot{\cup}C_1)\wedge(X_2,B_2\dot{\cup}C_2)=(X_1\cap X_2,((B_1\dot{\cup}C_1)\cup(B_2\dot{\cup}C_2))^{\triangle\triangle})$ 和 $(X_1,B_1\dot{\cup}C_1)\vee(X_2,B_2\dot{\cup}C_2)=((X_1\cup X_2)^{\triangle\triangle},(B_1\dot{\cup}C_1)\cap(B_2\dot{\cup}C_2))$ 也是概率概念,从而 $L(U,A,I,D,J)$ 是完备格,也是概率概念格。

证明:由 $(X_1,B_1\dot{\cup}C_1),(X_2,B_2\dot{\cup}C_2)$ 是概率概念,有 $X_1^{\triangle}=B_1\dot{\cup}C_1$, $X_2^{\triangle}=B_2\dot{\cup}C_2$, $(B_1\dot{\cup}C_1)^{\triangle}=X_1$, $(B_2\dot{\cup}C_2)^{\triangle}=X_2$。又由性质 1(5)有,$(X_1\cap X_2)^{\triangle}=((B_1\dot{\cup}C_1)^{\triangle}\cap(B_2\dot{\cup}C_2)^{\triangle})^{\triangle}$, $((B_1\dot{\cup}C_1)\cup(B_2\dot{\cup}C_2))^{\triangle\triangle\triangle}=((B_1\dot{\cup}C_1)\cup(B_2\dot{\cup}C_2))^{\triangle}=(B_1\dot{\cup}C_1)^{\triangle}\cap(B_2\dot{\cup}C_2)^{\triangle}=X_1\cap X_2$。因此 $(X_1\cap X_2,((B_1\dot{\cup}C_1)\cup(B_2\dot{\cup}C_2))^{\triangle\triangle})$ 是概率概念。同理可证,$((X_1\cup X_2)^{\triangle\triangle},(B_1\dot{\cup}C_1)\cap(B_2\dot{\cup}C_2))$ 也是概率概念。

又 $(U,\emptyset\dot{\cup}C),(\emptyset,A\dot{\cup}D)\in L(U,A,I,D,J)$,其中 C 可以是 D 的任意子集,包括 D 和 $\emptyset$,因此 $L(U,A,I,D,J)$ 是完备格。

基于概率概念格的定义,进一步研究发现概率概念格 $L(U,A,I,D,J)$ 与经典概念格 $L(U,A,I)$ 之间存在同构关系。

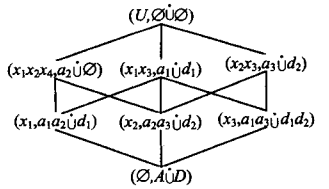
定理 2 设 (U,A,I,D,J) 是一个 D 型概率决策形式背景,(U,A,I) 是条件形式背景,则:

$L(U,A,I,D,J)\cong L(U,A,I)$

证明:设映射 $f:L(U,A,I)\to L(U,A,I,D,J)$,表示 $\forall(X,B)\in L(U,A,I)$,有 $f(X,B)=(X,B\dot{\cup}C)$, $C=\{d_l\mid d_l(X)\geqslant\alpha\}$。则 $\forall(X_1,B_1),(X_2,B_2)\in L(U,A,I)$,一方面,$f((X_1,B_1)\vee(X_2,B_2))=f((X_1\cup X_2)^{**},B_1\cap B_2)=((X_1\cup X_2)^{**},(B_1\cap B_2)\dot{\cup}C)$;另一方面,$f(X_1,B_1)\vee f(X_2,B_2)=(X_1,B_1\dot{\cup}C_1)\vee(X_2,B_2\dot{\cup}C_2)=((X_1\cup X_2)^{\triangle\triangle},(B_1\dot{\cup}C_1)\cap(B_2\dot{\cup}C_2))$。

又根据定义 2 和定义 8,有:$X^{\triangle}=X^{*}\cup\{d_l(x)\geqslant\alpha\}$, $X^{\triangle\triangle}=(X^{*}\cup\{d_l(x)\geqslant\alpha\})^{\triangle}=X^{**}$,因此 $(X_1\cup X_2)^{\triangle\triangle}=(X_1\cup X_2)^{**}$,于是 $f((X_1,B_1)\vee(X_2,B_2))=f(X_1,B_1)\vee f(X_2,B_2)$。同理可验证 $\wedge$ 运算,故 f 是同态映射。显然 f 是双射,因此 f 是同构映射。

例 3(续例 1) 概率概念格 $L(U,A,I,D,J)=\{(U,\emptyset\dot{\cup}\emptyset),(x_1,a_1a_2\dot{\cup}d_1),(x_2,a_2a_3\dot{\cup}d_2),(x_3,a_1a_3\dot{\cup}d_1d_2),(x_1x_3,a_1\dot{\cup}d_1),(x_2x_3,a_3\dot{\cup}d_2),(x_1x_2x_4,a_2\dot{\cup}\emptyset),(\emptyset,A\dot{\cup}D)\}$,经典概念格 $L(U,A,I)=\{(U,\emptyset),(x_1,a_1a_2),(x_2,a_2a_3),(x_3,a_1a_3),(x_1x_3,a_1),(x_2x_3,a_3),(x_1x_2x_4,a_2),(\emptyset,A)\}$。相应的格图如图 1 和图 2 所示。



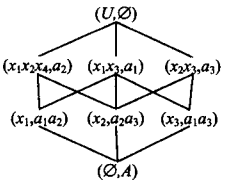


图 1 概率概念格 $L(U,A,I,D,J)$ 图 2 经典概念格 $L(U,A,I)$

由图 1 和图 2 可知,显然,$L(U,A,I,D,J)\cong L(U,A,I)$。

4 协调 D 型概率决策形式背景下的规则获取

本节利用包含度定义了 D 型概率决策形式背景的协调性,在协调的 D 型概率决策形式背景下获取决策规则。进一步,考虑到规则集中可能会存在冗余规则,通过剔除冗余获得非冗余规则集,使规则集更加简洁、紧凑。

4.1 α 协调的 D 型概率决策背景

首先,根据对象在决策属性上的取值大小对对象进行分类。

将决策 $d_l(l\leqslant k)$ 的值域 $[0,1]$ 做如下划分：$\{[0,\alpha_1]_{d_l},(\alpha_1,\alpha_2]_{d_l},\cdots,(\alpha_t,\alpha_{t+1}]_{d_l},\cdots,(\alpha_n,1]_{d_l}\}$。基于这种划分，定义 U 上关于决策 d_l 的等价关系 $R_{d_l}=\{(x_i,x_j)\mid d_l(x_i),d_l(x_j)\in(\alpha_t,\alpha_{t+1}]_{d_l}\}$。$\forall C\subseteq D$，记 $R_C=\bigcap\limits_{d_l\in C}R_{d_l}$，显然 R_C 也是 U 上的等价关系。

记 $L_o(U,A,I)=L(U,A,I)\backslash\{(U,\varnothing),(\varnothing,A)\}$，$L_o(U,A,I,D,J)=L(U,A,I,D,J)\backslash\{(U,\varnothing\dot\cup C),(\varnothing,A\dot\cup D)\}$，$C\subseteq D$。

定义 12　设 (U,A,I,D,J) 是一个 D 型概率决策形式背景，$C\subseteq D,\alpha\in[0,1]$。对于 $L_o(U,A,I)$ 中任意一个外延 X，在决策划分 U/R_C 中存在一个等价类 Y，使 $D(Y\mid X)=\dfrac{|X\cap Y|}{|X|}\geqslant\alpha$，则称 (U,A,I,D,J) 是 α 协调的 D 型概率决策背景。

例 4(续例 1)　根据定义 2 和定义 3 得，$L_o(U,A,I)=\{(x_1,a_1a_2),(x_2,a_2a_3),(x_3,a_1a_3),(x_1x_3,a_1),(x_2x_3,a_3),(x_1x_2x_4,a_2)\}$。

对决策 $d_l(l\leqslant2)$ 的值域 $[0,1]$ 做如下划分，d_1：$\{[0,0.3],(0.3,0.5],(0.5,0.8],(0.8,1]\}$，$d_2$：$\{[0,0.3],(0.3,0.4],(0.4,0.7],(0.7,1]\}$。相应的决策划分为：$U/R_{d_1}=\{x_1,x_2x_4,x_3\}$，$U/R_{d_2}=\{x_1x_4,x_2,x_3\}$，$U/R_D=U/R_{d_1}\cap U/R_{d_2}=\{x_1,x_2,x_3,x_4\}$。

取 $\alpha=0.5$，对 $L_o(U,A,I)$ 的每个元素 (X,B) 计算包含度：$D(x_1\mid x_1)=1$，$D(x_2x_4\mid x_2)=1$，$D(x_3\mid x_3)=1$，$D(x_1\mid x_1x_3)=0.5$，$D(x_2\mid x_2x_3)=0.5$，$D(x_2x_4\mid x_1x_2x_4)=0.67$。

显然，包含度都大于或等于 0.5。因此，(U,A,I,D,J) 是一个 α 协调的 D 型概率决策背景。

4.2　α 协调 D 型概率决策背景下的规则获取

在张文修等人[6,20]提出决策形式背景后，许多学者基于决策形式背景研究规则的提取。魏玲[6]解决了强协调决策形式背景中的规则提取问题。

如 3.2 节所述，生活中有这种情况：只有当某种结果出现的概率超过参数 α 时，这个结果才会引起人们的重视。从乐观角度出发，人们会更倾向于关注出现这个结果的最小可能性是多大。例如，医生根据临床症状判断病人患甲亢的概率已经超过 α，如果病患是乐观者，则会更关心自己患甲亢的最小概率是多少。本节从乐观心态出发，提出如何获取概率决策规则。

定义 13　在 α 协调的 D 型概率决策背景 (U,A,I,D,J) 中，$\forall(X,B\dot\cup C)\in L(U,A,I,D,J)$，$D_0=\{d_l(p_l)\}$，$p_l=\min\{d_l(x)\mid x\in X\}$，$\{d_l\}=C$，称：

"If B, then D_0"

是概率决策规则，记作"$B\to D_0$"。

记 $\Omega(A)=\{B\to D_0\mid(X,B\dot\cup C)\in L(U,A,I,D,J)\}$ 是所有概率决策规则的集合。

在如上定义中，$\{d_l\}=C$。由概率概念定义知 $C=\{d_l\mid d_l(X)\geqslant\alpha\}$，即只保留决策值 $d_l(X)\geqslant\alpha$ 的决策属性。决策结果 d_l 的概率值 p_l 是在大于或等于参数 α 的前提下，取概率的最小值。

需要提及的是，既然可以从乐观心态入手，同样也可以从悲观心态入手，即关注出现某个结果的最大可能性，则在获取决策规则时，$p_l=\max\{d_l(x)\mid x\in X\}$。这两种规则获取方法源于两种不同心态，仅是概率值取法不同，其余皆相同。因此本文仅使用最小概率方法提取规则，最大概率方法与此类似。

例 5(续例 1)　概率概念格 $L(U,A,I,D,J)$ 见例 2，$\alpha=0.5$，则 $L_o(U,A,I,D,J)=\{(x_1,a_1a_2\dot\cup d_1),(x_2,a_2a_3\dot\cup d_2),(x_3,a_1a_3\dot\cup d_1d_2),(x_1x_3,a_1\dot\cup d_1),(x_2x_3,a_3\dot\cup d_2),(x_1x_2x_4,a_2\dot\cup\varnothing)\}$。

对于 $L_o(U,A,I,D,J)$ 中的所有概念，逐一计算 D_0。

$(x_1,a_1a_2\dot\cup d_1)$：$p=\min d_1(x_1)=0.7$，$D_0=\{d_1(0.7)\}$；

$(x_2,a_2a_3\dot\cup d_2)$：$p=\min d_2(x_2)=0.6$，$D_0=\{d_2(0.6)\}$；

$(x_3,a_1a_3\dot\cup d_1d_2)$：$p_1=\min d_1(x_3)=0.5$，$p_2=\min d_2(x_3)=0.9$，$D_0=\{d_1(0.5),d_2(0.9)\}$；

$(x_1x_3,a_1\dot\cup d_1)$：$p=\min\{d_1(x_1),d_1(x_3)\}=\min\{0.7,0.5\}=0.5$，$D_0=\{d_1(0.5)\}$；

$(x_2x_3,a_3\dot\cup d_2)$：$p=\min\{d_2(x_2),d_2(x_3)\}=\min\{0.6,0.9\}=0.6$，$D_0=\{d_2(0.6)\}$；

$(x_1x_2x_4,a_2\dot\cup\varnothing)$：$D_0=\varnothing$。

因此，共得到 5 条概率决策规则：

r_1：$a_1a_2\to\{d_1(0.7)\}$；

r_2：$a_2a_3\to\{d_2(0.6)\}$；

r_3：$a_1a_3\to\{d_1(0.5),d_2(0.9)\}$；

r_4：$a_1\to\{d_1(0.5)\}$；

r_5：$a_3\to\{d_2(0.6)\}$。

分析以上 5 条规则，得出以下结论：

(1)当病人具有症状 a_1 时，患疾病 d_1 的最小概率为 0.5，因此要警惕是否患疾病 d_1；

(2)当病人具有症状 a_3 时，患疾病 d_2 的最小概率为 0.6，因此要警惕是否患疾病 d_2；

(3)症状 a_2 不能作为判断是否患有疾病 d_1 和 d_2 的重要依据。

在例 5 中共得到 5 条决策规则，比较规则 r_2 和 r_5 发现，这两条规则有相同的决策结果，但规则 r_5 的属性比 r_2 少，即规则 r_5 用较少的属性得到了与 r_2 相同的决策结果。换言之，规则 r_2 能够被更简洁的规则 r_5 替代，称这种能被替代的决策规则为冗余规则，其定义如下。

定义 14　设 (U,A,I,D,J) 是一个 α 协调的 D 型概率决策背景，规则集为 $\Omega(A)$。如果 $\forall B'\to D_0'\in\Omega(A)$，$\exists B''\to D_0''\in\Omega(A)\backslash\{B'\to D_0'\}$，都有 $B''\subseteq B'$ 且 $D_0'=D_0''$，则称决策规则 $B''\to D_0''$ 能够蕴含决策规则 $B'\to D_0'$，记这种蕴含关系为 $B''\to D_0''\Rightarrow B'\to D_0'$，并称决策规则 $B'\to D_0'$ 是 $\Omega(A)$ 中的冗余规则。记 $\Omega(A)$ 中所有非冗余决策规则组成的集合为 $\Omega^*(A)$。

显然，由例 5 有 $\Omega^*(A)=\{r_1,r_3,r_4,r_5\}$。

在 α 协调的 D 型概率决策背景中，冗余规则都能被非冗余规则蕴含。非冗余规则相比冗余规则来说，前者能用较少的属性得到与冗余规则相同的决策结果。因此，可以通过剔除冗余规则和简化规则集，更加快速、省时地获得信息。

5 概率概念格生成、规则获取算法

根据第 3 节、第 4 节的定义及定理，设计概率概念格生成、规则获取算法，如算法 1、算法 2 所示。

算法 1 概率概念格生成算法

输入：D 型概率决策形式背景(U,A,I,D,J)

输出：该 D 型概率决策形式背景的概率概念格 L(U,A,I,D,J)

step1 找到所有概率概念

step1.1 输入 D 型概率决策形式背景(U,A,I,D,J)

step1.2 令 $M=\rho(U)$，$N=|M|$

step1.3 令 $i=0,j=1,k=1,L(U,A,I,D,J)=\varnothing$

step1.4 令 $X_i=M_k,B_j \dot{\cup} C_j=X_i^{\triangle}$

step1.5 i++，令 $X_i=(B_j \dot{\cup} C_j)^{\triangle}$

step1.6 如果 $X_i=X_{i-1}$，则$(X_i,B_j \dot{\cup} C_j)$是概率概念，令 $L(U,A,I,D,J)=L(U,A,I,D,J)\cup\{(X_i,B_j \dot{\cup} C_j)\}$，转 step1.9；否则，转 step1.7

step1.7 j++，令 $B_j \dot{\cup} C_j=X_i^{\triangle}$

step1.8 如果 $B_j \dot{\cup} C_j=B_{j-1} \dot{\cup} C_{j-1}$，则$(X_i,B_j \dot{\cup} C_j)$是概率概念，令 $L(U,A,I,D,J)=L(U,A,I,D,J)\cup\{(X_i,B_j \dot{\cup} C_j)\}$，转 step1.9；否则，转 step1.5

step1.9 如果 $k\geqslant N$，停止，输出 L(U,A,I,D,J)；否则 k++，j++，i++，转 step1.4

step2 根据定义 11 构造根节点和末梢节点，再确定其他概念的前驱和后继关系，构造概率概念格 L(U,A,I,D,J)

算法 2 规则获取算法

输入：α 协调的 D 型概率决策背景(U,A,I,D,J)的概率概念格 L(U,A,I,D,J)

输出：该 D 型概率决策形式背景所有规则的集合 $\Omega(A)$

step1 输入 α 协调的 D 型概率决策背景(U,A,I,D,J)的概率概念格 L(U,A,I,D,J)

step2 令 $L_0(U,A,I,D,J)=L(U,A,I,D,J)\backslash\{(U,\varnothing \dot{\cup} C),(\varnothing,A \dot{\cup} D)\}$

step3 令 $l=L_0(U,A,I,D,J)$，$Z=|l|$

step4 令 $k=1,\Omega(A)=\varnothing$

step5 根据定义 13，对 l_k 计算 D_0

step6 根据定义 13，得到规则“$B\rightarrow D_0$”，令 $\Omega(A)=\Omega(A)\cup\{B\rightarrow D_0\}$

step7 如果 $k\geqslant Z$，停止，输出 $\Omega(A)$；否则 k++，转 step5

结束语 本文针对生活中存在决策属性取值的不确定性这一现象，提出 D 型概率决策形式背景，获得概率形式概念，构造概率概念格，并在 α 协调的 D 型概率决策背景上进行规则获取。考虑到实际应用中存在大量冗余规则，进而通过剔除冗余规则，使规则集更加简洁、紧凑。最后，本文给出了概率概念格生成、决策规则获取的算法，以利于计算机实现。在本文的研究基础上，还可以通过寻找辨识矩阵，定义保持规则集不变的属性约简。除此之外，参数 α 如何取值也是有待解决的问题。

参 考 文 献

[1] FAYYAD U,PIATETSKY-SHAPIOR M,SMYTH G. From Data Mining to Knowledge Discovery in databases[C]// Proc of Menlo Park. California: AAAI Press,1996:1-35.

[2] WILLE R. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts[M]. Heidelberg: Springer Netherlands,1982:445-470.

[3] HO T B. An approach to concept formation based on formal concept analysis[J]. IEICE Transactions on Information and Systems,1995,78(5):553-559.

[4] KUZNETSOV S O. A fast algorithm for computing all intersection of objects in a finite semilattice[J]. Automatic Documentation and Mathematical Linguistics,1993,27(5):11-12.

[5] GANTER B,WILLE R. Conceptual scaling [C]// Proc of Applications of Combinatorics and Graph Theory to the Biological and Social Science. New York: Springer-Verlag,1989:139-167.

[6] WEI L. Reduction Theory and Approach to Rough Set and Concept Lattice [D]. Xi'an: Xi'an Jiaotong University,2005. (in Chinese)
魏玲. 粗糙集与概念格约简理论与方法[D]. 西安：西安交通大学，2005.

[7] LI T J,WU W Z,LIU J. Attribute reduction in Decision Formal Context based on Rough Set[J]. Computer Science,2009,36(8A):48-61. (in Chinese)
李同军，吴伟志，刘军. 基于粗糙集的决策形式背景属性约简[J]. 计算机科学，2009，36(8A)：48-61.

[8] LI J J,ZHANG Y L,WU W Z,et al. Attribute Reduction for Formal Concept and Consistent Decision Formal Concept and Concept Lattice Generation[J]. Chinese Journal of Computers,2014,37(8):1768-1774. (in Chinese)
李进金，张燕兰，吴伟志，等. 形式背景与协调决策形式背景属性约简与概念格生成[J]. 计算机学报，2014，37(8)：1768-1774.

[9] LI J H,LV Y J. Attribute reduction and Rules Extraction in Decision Formal Context based on Concept Lattice[J]. Mathematics in Practice and Theory,2009,39(7):182-188. (in Chinese)
李金海，吕跃进. 基于概念格的决策形式背景属性约简及规则提取[J]. 数学的实践与认识，2009，39(7)：182-188.

[10] LI T. Rule Acquisition on Decision Formal Context [D]. Xi'an: Northwest University,2013. (in Chinese)
李涛. 决策形式背景的知识获取[D]. 西安：西北大学，2013.

[11] YANG S Q,DING S L,CAI S Z,et al. An algorithm of constructing concept lattices for CAT with cognitive diagnosis[J]. Knowledge-Based Systems,2008,21(8):852-855.

[12] COLE R,EKLUND P W. Scalability in formal concept analysis [J]. Computational Intelligence,2000,15(1):11-27.

[13] WU W Z,LEUNG Y,MI J S. Granular computing and knowledge and Data Engineering[J]. IEEE Transactions on Knowledge & Data Engineering,2008,21(10):1461-1474.

[14] LI J H,MEI C L,LV Y J. Knowledge reduction in real decision formal contexts[J]. Information Sciences,2012,189(7):191-207.

[15] LI J H,MEI C L,LV Y J. Incomplete decision contexts: Approximate concept construction,rule acquisition and knowledge reduction[J]. International Journal of Approximate Reasoning,

2013,54(1):149-165.

[16] LIU B X,LI Y. Construction Principles and Algorithms of Concept Lattice Generated by Random Decision Formal Context[J]. Computer Science,2013,40(6A):90-92. (in Chinese)
刘保相,李言. 随机决策形式背景下的概念格构建原理与算法[J]. 计算机科学,2013,40(6A):90-92.

[17] WEI L,Q J J,ZHANG W X. Attribute Reduction of Concept Lattice in Decision Formal Context[J]. Science in China (Series E):Information Science,2008,38(2):195-208. (in Chinese)
魏玲,祁建军,张文修. 决策形式背景的概念格属性约简[J]. 中国科学 E 辑:信息科学,2008,38(2):195-208.

[18] LI J H. Rule Acquisition Oriented Reduction Methods for Concept Lattices and Their Implementation Algorithms [D]. Xi'an:Xi'an Jiaotong University,2012. (in Chinese)
李金海. 面向规则提取的概念格约简方法及其算法实现[D]. 西安:西安交通大学,2012.

[19] GANTER B,WILLE R. Formal Concept Aanlysis [M]. Mathematical Foundations. New York:Springer-Verlag,1999.

[20] 张文修,仇国芳. 基于粗糙集的不确定决策[M]. 北京:清华大学出版社,2005:185-205.

[21] 张文修,梁怡. 不确定性推理原理[M]. 西安:西安交通大学出版社,1996:56-76.

(上接第 273 页)

参 考 文 献

[1] REN L,DU Y,MA S. Visual Analytics Toward Big Data[J]. Journal of Software,2014,25(9):1909-1936. (in Chinese)
任磊,杜一,马帅. 大数据可视分析综述[J]. 软件学报,2014,25(9):1909-1936.

[2] XU J,WANG G Y,YU H. Review of Big Data Processing Based on Granular Compu-ting[J]. Chinese Journal of Computers,2015,38(8):1497-1517. (in Chinese)
徐计,王国胤,于洪. 基于粒计算的大数据处理[J]. 计算机学报,2015,38(8):1497-1517.

[3] ZHAO X J,YANG C M,LI B. A Topic Evolution Minning Algorithm of News Text Based on Feature Evolving[J]. Chinese Journal of Computers,2014(4):819-832. (in Chinese)
赵旭剑,杨春明,李波. 一种基于特征演变的新闻话题演化挖掘方法[J]. 计算机学报,2014(4):819-832.

[4] CUI K,ZHOU B,JIA Y. LDA-based Model for Online Topic Evolution Mining[J]. Computer Science,2010,37(11):156-193. (in Chinese)
崔凯,周斌,贾焰. 一种基于 LDA 的在线主题演化挖掘模型[J]. 计算机科学,2010,37(11):156-193.

[5] HU Y L,BAI L,ZHANG W M. Modeling and Analyzing Topic Evolution[J]. Acta Automatica Sinica,2012,38(10):1690-1697. (in Chinese)
胡艳丽,白亮,张维明. 一种话题演化建模与分析方法[J]. 自动化学报,2012,38(10):1690-1697.

[6] FANG Y,HUANG H Y,XIN X. Topic Evolutionary Analysis for Dynamic Topic Number[J]. Journal of Chinese Information Processing,2014,28(3):142-149. (in Chinese)
方莹,黄河燕,辛欣. 面向动态主题数的话题演化分析[J]. 中文信息学报,2014,28(3):142-149.

[7] XU W,ZHAO B,JI G L. Microblog Topic Evolution Algorithm Based on Retweeti-ng Relationship[J]. Computer Science,2016,43(2):79-100. (in Chinese)
徐伟,赵斌,吉根林. 基于转发关系的微博话题演化算法[J]. 计算机科学,2016,43(2):79-100.

[8] JAYASHRI M,CHITRA P. Topic Clustering and Topic Evolution Based On Temporal Parameters[C]// International Conference on Recent Trends in Information Technology. Chennai,India:IEEE,2012:559-564.

[9] JENSEN S,LIU X Z,YU Y G. Generation of topic evolution trees from heterogeneous bibliographic networks[J]. Journal of Informetrics,2016,4(2):606-621.

[10] JO Y,HOPCROFT J E,LAGOZE C. The Web of Topics:Discovering the Topology of Topic Evolution in a Corpus[C]// WWW 2011-Session:Spatio-Temporal Analysis. Hyderabad,India:ACM,2011:257-266.

[11] ZHAO A H,LIU P U,ZHENG Y. Subtopic Division in News Topic Based on Latent Dirichlet Allocation[J]. Journal of Chinese Computer Systems,2013,34(4):732-737. (in Chinese)
赵爱华,刘培玉,郑燕. 基于 LDA 的新闻话题子话题划分方法[J]. 小型微型计算机系统,2013,34(4):732-737.

[12] DING Z Y,ZHOU B,JIA Y. Detecting Spammers with a Bidirectional Vote Algorithm Based on Statistical Features in Microblogs[J]. Journal of Computer Research and Development,2013,50(11):2336-2348. (in Chinese)
丁兆云,周斌,贾焰. 微博中基于统计特征与双向投票的垃圾用户发现[J]. 计算机研究与发展,2013,50(11):2336-2348.

[13] CAI G Y,PENG L B,WANG Y. Topic Detection and Evolution Analysis on Microblog[C]// International Federation for Information Processing. Trondheim,Norway:2014:67-77.

[14] ZAHO B,XU W,JI G L. Discovering Topic Evolution Topology in a Microblog Corpus [C]// Third International Conference on Advanced Cloud and Big Data. YangZhou,JiangSu,China:CBD,2016:7-14.

[15] BLEI D M,NG A Y,JORDAN M I. Latent Dirichlet Allocation [J]. The Journal of Machine Learning Research,2003,3(3):993-1022.

[16] CAO J P,WANG H,XIA Y Q. Bi-path Evolution Model for Online Topic Model Based on LDA[J]. Acta Automatica Sinica,2014,40(12):2877-2886. (in Chinese)
曹建平,王晖,夏友清. 基于 LDA 的双通道在线主题演化模型[J]. 自动化学报,2014,40(12):2877-2886.