第11期　　电　子　学　报　　Vol. 50　No.11
2022年11月　　ACTA ELECTRONICA SINICA　　Nov.　2022

旁观者视角下粒的多层次描述

智慧来[1]，张　丽[1]，李金海[2]

（1. 河南理工大学计算机科学与技术学院，河南焦作 454003；2. 昆明理工大学数据科学研究中心，云南昆明 650500）

摘　要： 不确定性在大数据中普遍存在.现有的研究极少从外部视角研究数据的不确定性，特别是模糊数据，经常无意识地假定获得的原始数据是完全真实可靠的，不存在任何偏差.然而，鉴于数据获取的主体普遍存在着不确定性，在大多数情况下这种做法是不可取的.为此，本文基于模糊概念格提出旁观者视角下粒的多层次描述.该方法基本思想是通过单个元素隶属度找出模糊形式背景中所有的原子粒，利用原子粒的自由组合来判定一个目标粒是否为可定义粒，并由原子粒的描述获取可定义粒的描述.接着，借助粗糙集上下近似思想给出一种不可定义粒的近似描述方法，它的主要思想是用可定义粒近似刻画不可定义粒.其中，查全率和查准率是为不可定义粒选择最合适近似描述的两个重要评价指标，而近似描述的质量则通过描述精度进行度量.最后，通过商品评价的实验分析说明了所提多层次数据分析方法的必要性与可行性，同时表明模糊概念格的真值越丰富得到的多层次粒描述结果越好.

关键词： 不确定性；模糊集；粒计算；概念格；粒描述；模糊概念；多层次数据分析

中图分类号： TP182；O236　　**文献标识码：** A　　**文章编号：** 0372-2112(2022)11-2568-07

电子学报URL:http://www. ejournal. org. cn　　**DOI:**10.12263/DZXB.20211164

Multi-Level Description of Granules From an Outsider's Perspective

ZHI Hui-lai[1]，ZHANG Li[1]，LI Jin-hai[2]

（1. *School of Computer Science and Technology*，*Henan Polytechnic University*，*Jiaozuo*，*Henan* 454003，*China*；

2. *Data Science Research Center*，*Kunming University of Science and Technology*，*Kunming*，*Yunnan* 650500，*China*）

Abstract： Uncertainty is pervasive in big data. In the existing researches, the uncertainty of data was seldom concerned from an external perspective, especially for the fuzzy data, and it was often unconsciously assumed that the original obtained data is completely truthful and reliable without any bias. However, due to the ubiquitous uncertainty of the subject in data acquisition, this kind of assumptions are not reasonable in most cases. In order to solve this problem, a multi-level description method of granules is proposed from an outsider's perspective based on fuzzy concept lattice. The basic idea of this method is to find out all the atomic granules in the fuzzy formal context by using the degree of membership of single element, and further determine whether a target granule is a definable granule by a free combination of the atomic granules, which can be used to obtain the description of a definable granule by the descriptions of atomic granules. After that, an approximate method to describe indefinable granules is proposed based on the ideas of the lower and upper approximations which are from the rough set theory, and the main idea of this method is to describe the indefinable granules approximately by means of definable granules, where recall and precision are two important evaluation criteria to select the most appropriate approximate descriptions for the indefinable granules, while the quality of the approximate descriptions is measured by description accuracy. Finally, the necessity and feasibility of the proposed multi-level data analysis methods are illustrated by an experimental analysis of commodity evaluation, and at the same time, the bigger the number of membership degrees in the fuzzy concept lattice, the better the obtained results of multi-level granule description.

Key words： uncertainty；fuzzy set；granular computing；concept lattice；granule description；fuzzy concept；multi-level data analysis

收稿日期：2021-08-28；修回日期：2021-11-25；责任编辑：覃怀银

基金项目：国家自然科学基金（No.11971211）；河南省高校基本科研业务费专项（No.NSFRF210318）

1 引言

不确定性在大数据中普遍存在[1]. 模糊集是20世纪60年代Zadeh教授提出的一种数学理论[2],广泛应用于不确定性建模和分析的各个领域. 在实际应用中引入模糊集,可以建立各种不同类型的模糊信息系统.

在现有的研究中,经常无意识地假定获得的原始数据是完全真实可靠的,不存在任何偏差. 然而,在大多数情况下,这种做法是不合理的. 例如,在刑事侦查过程中对证人进行调查取证时,证人的年龄、教育程度、职业、身体状况、性格等因素都会影响办案人员对其证词的采信程度. 又如,在科学实验中,观测仪器的精度以及实验人员的观测方式与方法都会增加实验结果的不确定性. 此外,数据分析的目的不同,数据分析所要求的精确程度也会有所差异. 再如,在调查民众对某项城市规划的意见时,采用完全信任、部分信任和不信任来表示对调查结果的采信程度是比较经济且有效的一种方式[3,4]. 但在有些情况下,比如国家之间的外交谈判,这种分析依然是不够的, 或者说是不充分的.

粒计算是以信息粒为基本认知单元分析和解决复杂问题的一种新方法[5,6],对粒进行恰当地刻画并赋予一定的语义解释是基于粒计算实现可解释人工智能的关键[7]. Pedrycz[8]引入了一种新的粒描述符以及隶属度函数研究粒的描述,以提高其可解释性. Zhi 等[9]研究了粒的共性属性刻画,并讨论了可定义粒的最简描述以及不可定义粒的近似描述. 为了分析对象的独有属性,智慧来等[10]又提出了基于必然属性的粒描述方法. Li 等[11]从知识表示的角度研究了粒的可定义性及其描述. Zhu 等人[12]使用最优化思想寻找粒的最优刻画. 苗夺谦等人[13]基于等价关系用集合运算的手段探讨了粒描述. 然而,现有方法的不足之处在于,并非以旁观者的视角进行粒描述,从而忽略了旁观者视角下粒的不确定性.

为了解决上述问题,本文提出一种旁观者视角下粒的多层次描述方法. 注意到形式概念分析通过外延和内涵二元视角研究数据之间的关联,在某种意义上为讨论对象赋予了一定的语义解释[14]. 在此基础上模糊形式概念分析是以模糊概念作为分析工具处理模糊数据[15,16],截至目前研究者们已建立若干模糊概念格模型[17~21]. 其中,L-模糊概念格模型使用偏序真值集L刻画数据的不确定性程度,因其完美的数学性质获得了广泛关注[15~21]. 鉴于此,本文采用L-模糊概念格模型实现粒的多层次描述. 具体地,提出模糊原子粒及其描述,借助模糊原子粒的描述实现可定义模糊粒的刻画,利用可定义模糊粒进一步逼近不可定义模糊粒,并通过商务智能应用说明粒的多层次描述方法的优势.

2 基础知识

本节介绍若干基本概念,详见文献[19~21].

2.1 L-模糊概念格与L-模糊三支概念格

$(L,\wedge,\vee,\otimes,\rightarrow,0,1)$是一个完备剩余格,其中$(L,\wedge,\vee,0,1)$是包含最小元0和最大元1的完全格,$(L,\otimes)$是一个可交换幺半群,$\otimes$和$\rightarrow$满足:$a\otimes b\le c$当且仅当$b\le a\rightarrow c$.

X上的一个模糊集S是从X到L的一个映射,并记X上的全体模糊集为L^X. 类似地,X与Y之间的模糊二元关系I是$X\times Y$到L的一个映射. 在模糊集理论中,模糊集S可记作$S=\left\{\frac{S(x)}{x}\middle| x\in X\right\}$.

令$S_1,S_2\in L^X$为X上的模糊集. 记$S_1\subseteq S_2$当且仅当$\forall x\in X, S_1(x)\le S_2(x)$,并定义$S_1$到$S_2$的距离为$\delta(S_1,S_2)=\frac{\sum_{x\in X}(S_2(x)-S_1(x))}{|X|}$. 此外,$S_1,S_2$的交运算与并运算分别定义为

$$S_1\cap S_2=\left\{\frac{S_1(x)\wedge S_2(x)}{x}\middle| x\in X\right\} \tag{1}$$

$$S_1\cup S_2=\left\{\frac{S_1(x)\vee S_2(x)}{x}\middle| x\in X\right\} \tag{2}$$

在一个模糊形式背景$K=(X,Y,I,L)$中,属性集Y通过模糊二元关系I刻画了对象集X. 其中,$I(x,y)$表示对象x拥有属性y的真值度.

对于$A\in L^X$和$B\in L^Y$,分别定义算子$\uparrow: L^X\rightarrow L^Y$和$\downarrow: L^Y\rightarrow L^X$为

$$A^{\uparrow}(y)=\bigwedge_{x\in X}(A(x)\rightarrow I(x,y)) \tag{3}$$

$$B^{\downarrow}(x)=\bigwedge_{y\in Y}(B(y)\rightarrow I(x,y)) \tag{4}$$

在本文中,蕴涵$\rightarrow$采用标准Lukasiewicz运算,即对于$a,b\in L$, $a\rightarrow b=\min\{1-a+b,1\}$.

若$A^{\uparrow}=B$, $B^{\downarrow}=A$,则称(A,B)是K的一个L-模糊概念,并称A和B分别为(A,B)的外延与内涵.

K的所有L-模糊概念依据外延或内涵的包含关系构成一个完备格,称为L-模糊概念格,记为$\bar{L}(K)$.

为了进一步引入K的L-模糊三支概念,分别定义算子$\nearrow: L^X\rightarrow L^Y$和$\searrow: L^Y\rightarrow L^X$为

$$A^{\nearrow}(y)=\bigwedge_{x\in X}(A(x)\rightarrow I^c(x,y)) \tag{5}$$

$$B^{\searrow}(x)=\bigwedge_{y\in Y}(B(y)\rightarrow I^c(x,y)) \tag{6}$$

其中$I^c(x,y)=1-I(x,y)$.

对于$A\in L^X$和$B,C\in L^Y$,分别定义算子$\Uparrow: L^X\rightarrow L^Y\times L^Y$和$\Downarrow: L^Y\times L^Y\rightarrow L^X$为

$$A^{\Uparrow}(y)=\left(A^{\uparrow}(y),A^{\nearrow}(y)\right) \tag{7}$$

$$(B,C)^{\Downarrow}(x)=B^{\downarrow}(x)\cap C^{\searrow}(x) \tag{8}$$

若$A^{\Uparrow}=(B,C)$且$(B,C)^{\Downarrow}=A$，则称$(A,(B,C))$是K的一个L-模糊三支概念. 记L-模糊三支概念格为$\tilde{L}(K)$.

定理1[19] 设$\{(A_t,(B_t,C_t))|t\in T\}\subseteq\tilde{L}(K)$，$T$是指标集，则

$$\bigcup_{t\in T}(A_t,(B_t,C_t))=\left(\left(\bigcup_{t\in T}A_t\right)^{\Uparrow\Downarrow},\bigcap_{t\in T}(B_t,C_t)\right) \tag{9}$$

$$\bigcap_{t\in T}(A_t,(B_t,C_t))=\left(\bigcap_{t\in T}A_t,\left(\bigcup_{t\in T}(B_t,C_t)\right)^{\Downarrow\Uparrow}\right) \tag{10}$$

2.2 多层次数据分析与模糊粒描述

为了方便，用$I(x,y)=0$表示对象x一定不具有属性y；用$I(x,y)=0.5$表示对象x是否具有属性y是不确定的；用$I(x,y)=1$表示对象x一定具有属性y. 那么，从旁观者的角度，对x的信任程度表述如下：(1)若完全不信任x，则用$\frac{0}{x}$表示；(2)若部分信任x，则用$\frac{0.5}{x}$表示；(3)若完全信任x，则用$\frac{1}{x}$表示.

在这种情形下，如何以旁观者的视角通过x的陈述来获得有关y的合理性的描述？这是一个非常重要的问题.

若将上述$x\in X$与$y\in Y$扩展为模糊集$A\in L^X$和$B\in L^Y$，三值集$\{0,0.5,1\}$扩展为标准Lukasiewicz多值集，则称对A、B及两者之间二元关系的研究为多层次数据分析.

从信息粒的角度，对A与B之间二元关系的分析又可视为用B描述A. 因此，亦称此分析过程为模糊粒的多层次描述.

定义1 设$K=(X,Y,I,L)$为一个模糊形式背景，$A\in L^X,B\in L^Y$. 对于$\frac{t_i}{x_i}\in A$，若$\left(\frac{t_i}{x_i}\right)^{\uparrow}\supseteq B$或$\left(\frac{t_i}{x_i}\right)^{\nearrow}\supseteq B$成立，则称$A$具有$B$，记做$A\rhd B$.

若β是一个模糊属性的合取式，则用$\|\beta\|$表示β中包含的单个模糊属性的集合. 如$\beta=\frac{0.5}{y_1}\wedge\frac{1}{y_2}\wedge\neg\frac{0.5}{y_3}$，有$\|\beta\|=\left\{\frac{0.5}{y_1},\frac{1}{y_2},\neg\frac{0.5}{y_3}\right\}$.

定义2 给定一个模糊属性的合取式β，称

$$m(\beta)=\left\{a\in A|a\rhd\|\beta\|,A\in L^X\right\} \tag{11}$$

是β的外延. 进一步，若$m(\beta)=A$，则称β是A的一个描述.

定义3 设$K=(X,Y,I,L)$为一个模糊形式背景，$A\in L^X$. 若β是A的一个描述，且不存在A的另一个描述γ使得$\|\gamma\|\subseteq\|\beta\|$，则称$\beta$是$A$的一个极小描述. 进一步，若$\beta$是$A$的一个极小描述，并满足：

（1）若A存在另一个极小描述γ_1，则$\|\gamma_1\|$中的单个模糊属性个数大于$\|\beta\|$中的单个模糊属性个数；

（2）若A存在另一个极小描述γ_2使得$\|\gamma_2\|$和$\|\beta\|$中的单个模糊属性个数相等，则$m(\wedge p_i)\subseteq m(\wedge q_j)$，其中$p_i\in\|\beta\|-\|\gamma_2\|$，$q_j\in\|\gamma_2\|-\|\beta\|$，则称$\beta$是$A$的一个最简描述.

对于一个模糊粒，其最理想的刻画就是找到其最简描述. 这也是本文将要重点讨论的一个问题.

3 可定义模糊粒的多层次描述

本节提出模糊原子粒，由模糊原子粒的组合获得可定义模糊粒的描述.

3.1 模糊原子粒及其描述

定义4 设$K=(X,Y,I,L)$为一个模糊形式背景，$\left\{\frac{t}{y}\right\}\in L^Y$. 当$t\neq0$时，称$\left(\frac{t}{y}\right)^{\downarrow}$或$\left(\frac{t}{y}\right)^{\searrow}$为$y$在$t$值下诱导的模糊原子粒，简称原子粒，并记全体原子粒组成的集合为

$$M=\left\{\left(\frac{t}{y}\right)^{\downarrow}\middle|\left\{\frac{t}{y}\right\}\in L^Y,t\neq0\right\}\cup\left\{\left(\frac{t}{y}\right)^{\searrow}\middle|\left\{\frac{t}{y}\right\}\in L^Y,t\neq0\right\}.$$

对于一个模糊原子粒A，由定义可知一定存在$\left\{\frac{t}{y}\right\}\in L^Y$使得$\left(\frac{t}{y}\right)^{\downarrow}=A$或$\left(\frac{t}{y}\right)^{\searrow}=A$，因此$A$有最简描述$\frac{t}{y}$或$\neg\frac{t}{y}$.

例1 表1是一个模糊形式背景$K=(X,Y,I,L)$，$L=\{0,0.5,1\}$，则K的正属性诱导的原子粒及其最简描述如表2所示.

表1 模糊形式背景K

X	y_1	y_2	y_3	y_4	y_5
x_1	0.5	0.5	1	0.5	0
x_2	1	0.5	0.5	1	0.5
x_3	0	1	0.5	0.5	1

3.2 可定义模糊粒及其描述

判定一个可定义模糊粒的基本思想是：若一个模糊粒可以表示为若干模糊原子粒的交组合，则这个模糊粒是可定义的. 再结合定理1不难得到以下定理.

定理2 设$K=(X,Y,I,L)$为一个模糊形式背景，$A\in L^X$，M为全体模糊原子粒的集合. 若存在$\{A_t\in M|t\in T\}$使得$\bigcap_{t\in T}A_t=A$，则A是可定义的，且有描述

表2　K的原子粒和最简描述

模糊原子粒	最简描述	模糊原子粒	最简描述
$\left\{\frac{1}{x_1},\frac{1}{x_2},\frac{0.5}{x_3}\right\}$	$\frac{0.5}{y_1},\neg\frac{0.5}{y_2},\neg\frac{0.5}{y_5}$	$\left\{\frac{1}{x_1},\frac{0.5}{x_2},\frac{0.5}{x_3}\right\}$	$\frac{1}{y_3}$
$\left\{\frac{0.5}{x_1},\frac{1}{x_2}\right\}$	$\frac{1}{y_1}$	$\left\{\frac{0.5}{x_1},\frac{1}{x_2},\frac{0.5}{x_3}\right\}$	$\frac{1}{y_4}$
$\left\{\frac{1}{x_1},\frac{1}{x_2},\frac{1}{x_3}\right\}$	$\frac{0.5}{y_2},\frac{0.5}{y_3},\frac{0.5}{y_4}$	$\left\{\frac{0.5}{x_1},\frac{1}{x_2},\frac{1}{x_3}\right\}$	$\frac{0.5}{y_5},\neg\frac{0.5}{y_3}$
$\left\{\frac{0.5}{x_1},\frac{0.5}{x_2},\frac{1}{x_3}\right\}$	$\frac{1}{y_2}$	$\left\{\frac{0.5}{x_2},\frac{1}{x_3}\right\}$	$\frac{1}{y_5}$

$\underset{t\in T}{\wedge}d(A_t)$，其中$d(A_t)$是$A_t$的一个最简描述.

上述定理得到的描述并不一定是最简的，原因是这些描述之间可能存在信息重复. 为此，需要讨论满足该条件的最小原子粒的集合.

定义5　设G_1与G_2是两个模糊集合族. 若对于任意的$S_i\in G_1$，存在$S_j\in G_2$使得$S_i\subseteq S_j$，则称G_1小于G_2，记作$G_1\le G_2$.

由粒的最简描述的定义，可得以下推论.

推论1　设$K=(X,Y,I,L)$为一个模糊形式背景，$A\in L^X$，M为全体模糊原子粒的集合. 若存在$\{A_t\in M|t\in T\}$使得$\bigcap_{t\in T}A_t=A$，且不存在$\{A_s\in M|s\in T'\}\le\{A_t\in M|t\in T\}$使得$\bigcap_{s\in T'}A_s=A$，则$A$是可定义的，且有最简描述$\underset{t\in T}{\wedge}d(A_t)$，其中$d(A_t)$是$A_t$的一个最简描述.

为了找到满足推论1的模糊原子粒集合，下面定义模糊集合族的约简运算.

定义6　令$G=\{S_i|i\in T\}$是一个模糊集合族，$T=\{1,2,\cdots,n\}$是指标集，$T'\subseteq T$. 若$\bigcap_{j\in T'}S_j=\bigcap_{i\in T}S_i$，且对于任意的$k\in T'$有$\bigcap_{j\in T'-\{k\}}S_j\ne\bigcap_{i\in T}S_i$，则称$G'=\{S_j|j\in T'\}$是$G$的一个约简，并记$G$的全体约简为$\mathrm{red}(G)$.

基于上述讨论，算法1给出了一种可定义模糊粒的判定和求解最简描述的方法.

算法1　可定义模糊粒的判定与最简描述

输入：模糊形式背景$K=(X,Y,I,L)$，$A\in L^X$.
输出：判定A是否是可定义的，若是则进一步输出其最简描述.
(1)令$M=\varnothing$；
(2)找出所有最小的$\left(\left(\frac{t}{y}\right)^\theta,\frac{t}{y}\right)$使得$\left(\frac{t}{y}\right)^\theta\supseteq A$，并将$\left(\frac{t}{y_i}\right)^\theta$加入$M$，其中$\theta$代表算子$\downarrow$或$\searrow$；
(3)若$A\ne\bigcap_{m\in M}m$，则A是不可定义的，算法结束；
(4)计算$\mathrm{red}(M)$，结果记为$M_i(i\in T)$，其中T是指标集；
(5)返回A的最简描述$\underset{r\in M_i}{\wedge}d(r)$，其中$d(r)$是$r$的一个最简描述.

例2　表1中的模糊形式背景K的非空可定义模糊粒及其最简描述如表3所示.

表3　K的可定义模糊粒和最简描述

可定义模糊粒	最简描述	可定义模糊粒	最简描述
$\left\{\frac{1}{x_1},\frac{1}{x_2},\frac{1}{x_3}\right\}$	$\frac{0.5}{y_2},\frac{0.5}{y_3},\frac{0.5}{y_4}$	$\left\{\frac{0.5}{x_2},\frac{0.5}{x_3}\right\}$	$\neg\frac{1}{y_3}$
$\left\{\frac{0.5}{x_1},\frac{1}{x_2},\frac{1}{x_3}\right\}$	$\frac{0.5}{y_5},\neg\frac{0.5}{y_3}$	$\left\{\frac{1}{x_1},\frac{0.5}{x_2}\right\}$	$\neg\frac{1}{y_5}$
$\left\{\frac{1}{x_1},\frac{1}{x_2},\frac{0.5}{x_3}\right\}$	$\frac{0.5}{y_1},\neg\frac{0.5}{y_2},\neg\frac{0.5}{y_5}$	$\left\{\frac{0.5}{x_1},\frac{1}{x_3}\right\}$	$\neg\frac{1}{y_1}$
$\left\{\frac{1}{x_1},\frac{0.5}{x_2},\frac{1}{x_3}\right\}$	$\neg\frac{0.5}{y_1},\neg\frac{0.5}{y_4}$	$\left\{\frac{0.5}{x_1},\frac{0.5}{x_3}\right\}$	$\neg\frac{1}{y_4}$
$\left\{\frac{0.5}{x_1},\frac{0.5}{x_2},\frac{1}{x_3}\right\}$	$\frac{1}{y_2}$	$\left\{\frac{0.5}{x_1},\frac{0.5}{x_2}\right\}$	$\neg\frac{1}{y_2}$
$\left\{\frac{0.5}{x_1},\frac{1}{x_2},\frac{0.5}{x_3}\right\}$	$\frac{1}{y_4}$	$\left\{\frac{0.5}{x_1}\right\}$	$\neg\frac{1}{y_2}\wedge\neg\frac{1}{y_4}$
$\left\{\frac{1}{x_1},\frac{0.5}{x_2},\frac{0.5}{x_3}\right\}$	$\frac{1}{y_3}$	$\left\{\frac{0.5}{x_2}\right\}$	$\neg\frac{1}{y_2}\wedge\neg\frac{1}{y_3}$
$\left\{\frac{0.5}{x_1},\frac{1}{x_2}\right\}$	$\frac{1}{y_1}$	$\left\{\frac{0.5}{x_3}\right\}$	$\neg\frac{1}{y_3}\wedge\neg\frac{1}{y_4}$
$\left\{\frac{0.5}{x_2},\frac{1}{x_3}\right\}$	$\frac{1}{y_5}$	$\left\{\frac{0.5}{x_1},\frac{0.5}{x_2},\frac{0.5}{x_3}\right\}$	$\frac{1}{y_2}\wedge\frac{1}{y_3}$,
$\left\{\frac{1}{x_3}\right\}$	$\neg\frac{1}{y_1}\wedge\neg\frac{1}{y_5}$		$\frac{1}{y_2}\wedge\frac{1}{y_4}$, $\frac{1}{y_3}\wedge\frac{1}{y_4}$

在下文中，用$\bar{L}_E(K)$与$\bar{L}_I(K)$分别表示$\bar{L}(K)$中全体L-模糊概念的外延与内涵构成的集合，用$\tilde{L}_E(K)$、$\tilde{L}_I^+(K)$与$\tilde{L}_I^-(K)$分别表示$\tilde{L}(K)$中全体L-三支模糊概念的外延、内涵第一元与内涵第二元构成的集合，记K的补背景为K^c. 在此基础上，不难得到以下定理.

定理3　设$K=(X,Y,I,L)$为一个模糊形式背景，则以下命题成立：

(1) $\bar{L}_E(K)\subseteq\tilde{L}_E(K)$；
(2) $\bar{L}_E(K^c)\subseteq\tilde{L}_E(K)$；
(3) $\bar{L}_I(K)=\tilde{L}_I^+(K)$；
(4) $\bar{L}_I(K^c)=\tilde{L}_I^-(K)$.

由定理3可得$\bar{L}_E(K)\bigcup\bar{L}_E(K^c)\subseteq\tilde{L}_E(K)$，从而在理论上证明了同时使用正属性和负属性比单独使用正属性或负属性能够描述更多的模糊粒.

4　不可定义模糊粒的多层次描述

类似于经典形式背景的情况[22]，本文将一对可定义模糊粒的描述作为不可定义模糊粒的近似描述.

定义7　设$K=(X,Y,I,L)$为一个模糊形式背景，$A\in L^X$是一个不可定义模糊粒. 若$A_u\supseteq A$且A_u是可定

义模糊粒，但不存在另一个可定义模糊粒 $A_{u'}$ 使得 $A_u \supseteq A_{u'} \supseteq A$，则称 A_u 是 A 的上近似模糊粒. 若 $A_l \subseteq A$ 且 A_l 是可定义模糊粒，但不存在另一个可定义模糊粒 $A_{l'}$ 使得 $A_l \subseteq A_{l'} \subseteq A$，则称 A_l 是 A 的下近似模糊粒.

在下文中，用 $d(A_u)$ 与 $d(A_l)$ 分别代表 A_u 与 A_l 的一个描述. 此外，令 $t(A)=\{C|C\subseteq A\}$ 表示 A 中包含的全体模糊粒，并用 $|t(A)|$ 表示集合 $t(A)$ 的元素个数.

定义 8 设 $K=(X,Y,I,L)$ 为一个模糊形式背景，$A\in L^X$，β 是 A 的一个描述，则 β 对于 A 的查全率和查准率分别定义为

$$\mu(\beta,A)=\frac{|t(m(\beta))\cap t(A)|}{|t(A)|} \quad (12)$$

$$\nu(\beta,A)=\frac{|t(m(\beta))\cap t(A)|}{|t(m(\beta))|} \quad (13)$$

查全率和查准率是反映目标模糊粒的描述效果的两个评价指标，通常其值越大越好，但一般很难同时达到最大值.

对于不可定义模糊粒 A，至少存在以下三种描述方式：

（1）用 $d(A_u)$ 作为 A 的近似描述，有 $\mu(d(A_u),A)=1$，$\nu(d(A_u),A)<1$.

（2）用 $d(A_l)$ 作为 A 的近似描述，有 $\mu(d(A_l),A)<1$，$\nu(d(A_l),A)=1$.

（3）用区间 $[d(A_l),d(A_u)]$ 作为 A 的近似描述，定义描述精度 $\tau=\dfrac{|t(A_l)|}{|t(A_u)|}$.

例 3 考虑表 1 中的模糊形式背景 K 的一个不可定义模糊粒 $A=\left\{\dfrac{1}{x_2},\dfrac{0.5}{x_3}\right\}$，可得上近似模糊粒 $A_u=\left\{\dfrac{0.5}{x_1},\dfrac{1}{x_2},\dfrac{0.5}{x_3}\right\}$，下近似模糊粒 $A_l=\left\{\dfrac{0.5}{x_2},\dfrac{0.5}{x_3}\right\}$，故 A 有以下三种描述方式：

（1）用 $d(A_u)=\dfrac{1}{y_4}$ 作为 A 的近似描述，有 $\mu(d(A_u),A)=1$，$\nu(d(A_u),A)=0.5$.

（2）用 $d(A_l)=\neg\dfrac{1}{y_3}$ 作为 A 的近似描述，有 $\mu(d(A_l),A)=0.67$，$\nu(d(A_l),A)=1$.

（3）用区间 $[d(A_l),d(A_u)]=[\neg\dfrac{1}{y_3},\dfrac{1}{y_4}]$ 作为 A 的近似描述，其描述精度 $\tau=\dfrac{|t(A_l)|}{|t(A_u)|}=0.33$.

5 商务智能应用与实验分析

商品评价数据可形式化表示为一个模糊形式背景 $K=(X,Y,I,L)$，其中 X 表示消费者集合，Y 表示商品内在价值属性集合，I 表示消费者对商品内在价值属性的评价，L 代表分析层次.

给定 $A\in L^X$，则消费者 A 对商品评价的一致性定义为

$$c(A)=\frac{\sum_{y\in Y}A^{\uparrow}(y)+\sum_{y\in Y}A^{\nearrow}(y)}{2|Y|} \quad (14)$$

进一步，设 $0<t_1<t_2<1$. 若 $c(A)\le t_1$，则称 A 对商品的评价是不一致的；若 $t_1<c(A)<t_2$，则称 A 对商品的评价是可能一致的；若 $c(A)\ge t_2$，则称 A 对商品的评价是一致的.

对于 $A\in L^X$，若存在 L-模糊三支概念 $(A,(B,C))$，则 $c(A)=\dfrac{\sum_{l_i/y_i\in B}l_i+\sum_{l_j/y_j\in C}l_j}{2|Y|}$. 若不存在 L-模糊三支概念的外延等于 A，则从外延大于 A 的 L-模糊三支概念中找出最小的那个概念，利用这个概念的内涵计算 $c(A)$.

下面通过实验来说明多层次数据分析的有效性，其流程设计如下：

步骤 1 创建模糊形式背景 K，其中对象和属性间二元模糊关系的取值来源于 $\{0,0.5,1\}$. 设置 t_1 和 t_2 的值，且满足 $0<t_1<t_2<1$.

步骤 2 令 $L_1=\{0,0.5,1\}$. 首先，建立 L-模糊三支概念格，记作 $\tilde{L}(K,L_1)$. 找出其所有外延可能为一致状态的概念，并记其数量为 N_1，计算这些概念占全体概念的比例 $r_1=N_1/|\tilde{L}(K,L_1)|$. 其次，找出 $\tilde{L}(K,L_1)$ 中所有外延为不一致状态的概念，并记这些概念的外延全体为 $\tilde{L}_{E'}(K,L_1)$. 对于每一个 $A\in\tilde{L}_{E'}(K,L_1)$，在 $\tilde{L}(K,L_1)$ 中找出外延大于 A 且是一致状态的概念中最小的那个概念，并记此概念外延为 $\bar{A}$. 计算平均改变量 $\mathrm{AVE}_1=\dfrac{\sum_{A\in\tilde{L}_{E'}(K,L_1)}\delta(A,\bar{A})}{|\tilde{L}_{E'}(K,L_1)|}$.

步骤 3 令 $L_2=\{0,0.25,0.5,0.75,1\}$. 同理，首先建立 L-模糊三支概念格，记作 $\tilde{L}(K,L_2)$，找出其所有外延为可能一致状态的概念，并记其数量为 N_2，计算这些概念占全体概念的比例 $r_2=N_2/|\tilde{L}(K,L_2)|$. 其次，记 $\tilde{L}(K,L_2)$ 中所有外延为不一致状态的概念外延全体为 $\tilde{L}_{E'}(K,L_2)$；对于每一个 $A\in\tilde{L}_{E'}(K,L_2)$，在 $\tilde{L}(K,L_2)$ 中找出外延大于 A 且是一致状态的概念中最小的那个概念，并记此概念

外延为 $\bar{\bar{A}}$. 计算平均改变量 $\mathrm{AVE}_2 = \dfrac{\sum_{A \in \tilde{L}_{E'}(K, L_2)} \delta(A, \bar{\bar{A}})}{|\tilde{L}_{E'}(K, L_2)|}$.

步骤 4　构建新的模糊形式背景,重复上述步骤以获得充分多的分析数据.

在实验中,随机生成了 12 个数据集,每三个数据集为一组. 数据集#1、#2 和#3 包含 5 个对象和 50 个属性,数据集#4、#5 和#6 包含 6 个对象和 60 个属性,数据集#7、#8 和#9 包含 7 个对象和 70 个属性,数据集#10、#11 和#12 包含 8 个对象和 80 个属性. 此外,设置 $t_1 = 0.35$, $t_2 = 0.65$.

实验结果如表 4 和表 5 所示,其中表 5 中的 $d = (\mathrm{AVE}_1 - \mathrm{AVE}_2)/\mathrm{AVE}_1$ 表示减小的比例. 实验结果表明提高分析层次能够增大可能一致性的中间状态以延迟决策,并可以减小对象集从不一致状态到一致状态的改变量以降低决策风险.

表 4　提高分析层次前后外延可能不一致的概念数量占比统计

数据集	$\lvert\tilde{L}(K, L_1)\rvert$	N_1	r_1/%	$\lvert\tilde{L}(K, L_2)\rvert$	N_2	r_2/%
#1	208	89	42.79	2567	1208	47.06
#2	203	87	42.86	2507	1291	51.50
#3	221	88	39.82	2791	1263	45.25
#4	621	217	34.94	12735	4867	38.22
#5	542	218	40.22	10788	4640	43.01
#6	614	205	33.39	12663	4695	37.08
#7	1271	500	39.34	38656	16027	41.46
#8	1462	446	30.51	44377	15226	34.31
#9	1387	488	35.18	43665	16259	37.24
#10	3544	1094	30.87	171016	56816	33.22
#11	3327	955	28.70	152613	49380	32.36
#12	3637	1084	29.80	177459	57503	32.40

表 5　提高分析层次前后状态改变量统计

数据集	AVE_1	AVE_2	d/%
#1	0.3378	0.3122	7.58
#2	0.2967	0.2913	1.82
#3	0.3396	0.3216	5.30
#4	0.3340	0.3134	6.17
#5	0.3291	0.3062	6.96
#6	0.3466	0.3217	7.18
#7	0.3197	0.2881	9.88
#8	0.3297	0.2987	9.40
#9	0.3318	0.2977	10.28
#10	0.3205	0.2797	12.73
#11	0.3151	0.2798	11.20
#12	0.3172	0.2803	11.63

6　总结

本文提出从旁观者的视角深入思考数据来源的不确定性对数据分析带来的影响,并采用 L-模糊概念辅助实现粒的多层次描述. 本文的贡献有两个方面:(1)利用 L 模糊概念实现了双重不确定性的粒描述;(2)L 真值集实现了粒描述的多层次刻画,真值越丰富,多层次粒描述的效果越好. 若从数学的角度出发,将 L 从有限离散集扩展到[0, 1]连续区间,进而扩展到实数集,此时的数学模型如何建立,其语义如何解释,都是十分重要且有趣的课题. 此外,如何利用多种模糊概念格模型[17,19]提出复杂数据的多层次分析方法也是值得进一步探讨的问题.

参考文献

[1] 梁吉业, 钱宇华, 李德玉, 等. 大数据挖掘的粒计算理论与方法[J]. 中国科学: 信息科学, 2015, 45(11): 1355-1369.
LIANG Ji-ye, QIAN Yu-hua, LI De-yu, et al. Theory and method of granular computing for big data mining[J]. Science China: Information Sciences, 2015, 45(11): 1355-1369. (in Chinese)

[2] ZADEH L A. Fuzzy sets[J]. Information Control, 1965, 8(3): 338-353.

[3] YAO Y Y. Three-way decision and granular computing[J]. International Journal of Approximate Reasoning, 2018, 103: 107-123.

[4] LIU D, YANG X, LI T R. Three-way decisions: beyond rough sets and granular computing[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(5): 989-1002.

[5] ZADEH L A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. Fuzzy Sets and Systems, 1997, 90(2): 111-127.

[6] 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.
XU Ji, WANG Guo-yin, YU Hong. Review of big data processing based on granular computing[J]. Chinese Journal of Computers, 2015, 38(8): 1497-1517. (in Chinese)

[7] PEDRYCZ W. Interpretable Artificial Intelligence: A Perspective of Granular Computing[M]. Berlin: Springer-Verlag, 2021: 333-369.

[8] PEDRYCZ W. From numeric to granular description and interpretation of information granules[J]. Fundamenta Informaticae, 2013, 127(1-4): 399-412.

[9] ZHI H L, LI J H. Granule description based on formal concept analysis[J]. Knowledge-Based Systems, 2016, 104: 62-73.

[10] 智慧来, 李金海. 基于必然属性分析的粒描述[J]. 计算机学报, 2018, 41(12): 2702-2719.
ZHI Hui-lai, LI Jin-hai. Granule description based on necessary attribute analysis[J]. Chinese Journal of Computers, 2018, 41(12): 2702-2719. (in Chinese)

[11] LI J H, LIU Z M. Granule description in knowledge granularity and representation[J]. Knowledge-Based Systems, 2020, 203: 106160.

[12] ZHU X B, PDERYCZ W, LI Z W. Granular data description: Designing ellipsoidal information granules[J]. IEEE Transactions on Cybernetics, 2017, 47(12): 4475-4484.

[13] 苗夺谦, 徐菲菲, 姚一豫, 等. 粒计算的集合论描述[J]. 计算机学报, 2012, 35(2): 351-363.
MIAO Duo-qian, XU Fei-fei, YAO Yi-yu, et al. Set-theoretic formulation of granular computing[J]. Chinese Journal of Computers, 2012, 35(2): 351-363. (in Chinese)

[14] GANTER B, WILLE R. Formal Concept Analysis: Mathematical Foundation[M]. Berlin: Springer-Verlag, 1999: 25-36.

[15] BURUSCO JUANDEABURRE A, FUENTES-GONZALEZ R. The study of the L-fuzzy concept lattice[J]. Mathware and Soft Computing, 1994, 1(3): 209-218.

[16] BELOHLAVEK R, DVORAK J, OUTRATA J. Fast factorization by similarity in formal concept analysis of data with fuzzy attributes[J]. Journal of Computer and System Sciences, 2007, 73(6): 1012-1022.

[17] 刘宗田, 强宇, 周文, 等. 一种模糊概念格模型及其渐进式构造算法[J]. 计算机学报, 2007, 30(2): 184-188.
LIU Zong-tian, QIANG Yu, ZHOU Wen, et al. A fuzzy concept lattice model and its incremental construction algorithm[J]. Chinese Journal of Computers, 2007, 30(2): 184-188. (in Chinese)

[18] 邹丽, 冯凯华, 刘新. 语言值直觉模糊概念格及其应用[J]. 计算机研究与发展, 2018, 55(8): 1726-1734.
ZOU Li, FENG Kai-hua, LIU Xin. Linguistic-valued intuitionistic fuzzy concept lattice and its application[J]. Journal of Computer Research and Development, 2018, 55(8): 1726-1734. (in Chinese)

[19] HE X L, WEI L, SHE Y H. L-fuzzy concept analysis for three-way decisions: basic definitions and fuzzy inference mechanisms[J]. International Journal of Machine Learning and Cybernetics, 2018, 9(11): 1857-1867.

[20] ZHANG Z. Constructing L-fuzzy concept lattices without fuzzy Galois closure operation[J]. Fuzzy Sets and Systems, 2018, 333: 71-86.

[21] BARTL E, KONECNY J. L-concept analysis with positive and negative attributes[J]. Information Sciences, 2016, 360: 96-111.

[22] YAO Y Y. Rough-set concept analysis: Interpreting RS-definable concepts based on ideas from formal concept analysis[J]. Information Sciences, 2016, 346-347: 442-462.

作者简介

智慧来 男,1981年3月出生于河南省洛阳市. 2010年毕业于上海大学计算机工程与科学学院. 现为河南理工大学计算机科学与技术学院副教授、硕士生导师. 主要研究方向为形式概念分析、粗糙集、粒计算等.
E-mail: zhihuilai@126.com

张 丽 女,1995年4月出生于河南省焦作市. 现为河南理工大学计算机科学与技术学院硕士研究生. 主要研究方向为形式概念分析、粒计算等.
E-mail: liizhang13@163.com

李金海(通讯作者) 男,1984年1月出生于江西省上饶市. 2012年毕业于西安交通大学数学与统计学院. 现为昆明理工大学数据科学研究中心教授、博士生导师. 主要研究方向为概念格、粗糙集、模糊集、粒计算、认知计算等.
E-mail: jhlixjtu@163.com