

# 一种具有相依关系的二维云推理方法及其在预测中的应用

刘德浩<sup>1</sup> 王倩<sup>2</sup>

(四川大学数学学院 成都 610065)<sup>1</sup> (倍特期货 成都 611731)<sup>2</sup>

**摘要** 基于云模型的定性知识推理通过构造规则发生器实现不确定性推理。然而传统的二维云推理方法并未考虑条件云之间的相互影响,故此对传统方法进行改进,通过条件云云滴的乘积作用,构造相应的规则发生器,提出一种基于相依条件云的二维云推理方法。将该方法应用于美国 GDP 预测,预测结果表明该方法的效果明显优于传统方法,进而表明该方法可行且有效,并在一定程度上弥补了传统方法的不足。

**关键词** 不确定性推理,相依条件云,规则发生器,知识预测

中图分类号 TP18 文献标识码 A

## Uncertainty Reasoning Based on Related Planar Cloud and Application in Prediction

LIU De-hao<sup>1</sup> WANG Qian<sup>2</sup>

(College of Mathematics, Sichuan University, Chengdu 610065, China)<sup>1</sup> (Brilliant Futures, Chengdu 611731, China)<sup>2</sup>

**Abstract** Qualitative reasoning based on cloud model implements uncertainty reasoning by constructing cloud regulation generators. But the traditional two-dimensional cloud reasoning methods do not take notice of the interaction between the condition clouds, as an improvement on the traditional method, a new approach to planar cloud reasoning method based on related condition clouds were suggested in this paper, which was followed by the focus on the product of cloud droplets and the construction of corresponded regulation generators. Then this new approach was applied in the prediction of GDP of America. The prediction results show that the new approach outperforms the traditional one, which indicates that the new method is feasible and effective, and to some extent remedies the shortcomings of traditional way.

**Keywords** Uncertainty reasoning, Related conditional cloud, Regulation generator, Prediction of information

### 1 引言

自然语言中普遍存在着不确定性,这些不确定性主要表现为模糊性与随机性。李德毅院士在充分考虑随机性和模糊性的关联关系的基础上,提出一种实现定性概念与其定量表示之间转换的模型——云模型,用来描述自然语言中概念的不确定性。并在云模型的基础上,构造规则发生器,实现不确定性推理。

基于云模型的不确定性推理方法,尤其是基于二维云的不确定性推理方法发展迅速,并且已被广泛应用于人工智能、数据挖掘、知识发现、可靠性评估、预测、信任评估等领域<sup>[1-6]</sup>。

但是,目前基于多维云的不确定性推理方法都没有考虑条件云之间可能存在的相依关系,例如通过身高、体重来研究肥胖。身高体重之间具有统计相关关系,因为通常身高越高,体重越重。这种情况下,传统的云推理方法不考虑条件云之间的相互影响,就不甚合理,因而将之应用于预测等方面,效果也可能不甚理想。

本文正是在充分考虑条件云相互影响的基础上,构造一种新的基于相依条件云的双条件单规则发生器,实现二维云推理,并将之应用于 GDP 预测。结果表明这种具有相依关系的二维云推理方法相比于传统方法具有更好的预测效果。这也表明本文方法是一种可行而且有效的二维云推理方法,其

也是对云模型理论与应用的一种补充发展。

### 2 云模型与云推理

#### 2.1 云模型概念

**定义 1** 设  $U$  是一个用精确数值表示的定量论域,  $C$  是  $U$  上的定性概念,若  $x \in U$  是概念  $C$  的一次随机实现,  $x$  对  $C$  的确定度  $\mu(x) \in [0, 1]$  是有稳定倾向的随机数:

$$\mu: U \rightarrow [0, 1]$$
$$x \rightarrow \mu(x), \forall x \in U$$

则  $x$  在论域  $U$  上的分布称为云,每一个  $x$  称为一个云滴<sup>[7]</sup>。

在上面定义中提及的随机实现是概率意义下的实现,即定量数值  $x$  的取值具有随机性,  $x$  对  $C$  的确定度  $\mu(x)$  是模糊集意义下的隶属度,同时又具有概率意义下的分布。这些都体现了模糊性和随机性的关联性。

云模型用 3 个参数,即期望  $E_x$ 、熵  $E_n$  和超熵  $H_e$  来反映定性概念的整体特征。其中,期望  $E_x$  是概念在论域中的中心值,最能代表这个定性概念;熵  $E_n$  表示定性概念的不确定性度量,由概念的模糊性和随机性共同决定,反映了论域中可被这个概念所接受的数值范围;超熵  $H_e$  是熵  $E_n$  的不确定性度量,可用来描述云的厚度。

另一方面,正态云模型由于具有良好的数学性质和一定的普适性而成为研究重点,其应用也最为广泛<sup>[8]</sup>。

刘德浩(1988—),男,博士,主要研究方向为应用数学, E-mail:ldh200612@126.com;王倩(1989—),女,硕士,主要研究方向为时间序列分析、人工智能。

定义2 设  $C$  是论域  $U$  上的定性概念,若  $x \in U$  是概念  $C$  的一次随机实现,且  $x$  还满足  $x \sim N(Ex, En'^2)$ , 其中,  $En' \sim N(En, He^2)$ , 且  $x$  对  $C$  的确定度  $\mu(x)$  满足

$$\mu = e^{-\frac{(x-Ex)^2}{2(En')^2}}$$

则  $x$  在论域  $U$  上的分布称为正态云。

## 2.2 云发生器

云发生器一般是指用计算机实现的一种特定算法,分为两种:正向云发生器与逆向云发生器。正向云发生器是根据云模型的参数模拟生成云滴;逆向云发生器则是通过云滴估计云模型的参数。

### 算法1 一维云正向发生器算法

输入:定性概念的数字特征  $(Ex, En, He)$  和云滴数  $n$ 。

输出: $n$  个云滴的定量值  $x$  及其确定度  $\mu$ 。

(1) 生成以  $En$  为期望值、 $He^2$  为方差的正态随机数  $En'_i$ ;

(2) 生成以  $Ex$  为期望值、 $En'^2_i$  为方差的正态随机数  $x_i$ ;

(3) 计算  $\mu_i = e^{-\frac{(x_i-Ex)^2}{2(En'_i)^2}}$ ;

(4) 具有确定度  $\mu_i$  的  $x_i$  成为数域中的一个云滴;

(5) 重复步骤(1)~步骤(4),直至产生要求的  $n$  个云滴为止。

### 算法2 无需确定度信息的逆向云发生器算法

输入:样本  $\{x_i\}_{i=1}^n$ 。

输出:定性概念的数字特征  $(Ex, En, He)$ 。

(1) 根据  $x_i$  计算样本均值  $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$ , 样本方差  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$ , 取  $Ex = \bar{X}$ ;

(2) 计算  $En, En = \sqrt{\frac{\pi}{2} \frac{1}{n} \sum_{i=1}^n |x_i - Ex|}$ ;

(3) 计算  $He, He = \sqrt{S^2 - En^2}$ 。

二维云模型的云发生器算法与一维的算法是类似的,此处不再赘述,可参见文献[9]。

## 2.3 传统云规则发生器

不确定性推理中,一般用条件和规则来表达不确定性知识。设一个单条件单规则表达为:

if  $A$  then  $B$

其中,  $A, B$  分别对应于论域  $U_1$  与  $U_2$  上的定性概念  $C_1$  和  $C_2$ 。如果给定前件论域  $U_1$  中的一个点  $a$ , 前件云发生器是指通过云发生器生成这个点  $a$  属于概念  $C_1$  的确定度;若给定一确定度  $\mu \in [0, 1]$ , 后件云发生器是指通过云发生器生成论域  $U_2$  中概念  $C_2$  上满足这个确定度的云滴。前件云发生器和后件云发生器组合构成云规则发生器。前件云可以是一维或者多维的,后件云通常是一维的。

另一方面,很多时候人们并不预先知道表达定性概念的云的数字特征,而只知道定性概念的一些定量数值,也就是代表定性概念的一个样本。例如为了得到描述“身高”的云模型,人们一般先通过随机抽样测量得到  $n$  个人身高的具体数值,然后通过逆向云发生器算法计算出“身高”的数字特征。

因此构建云规则发生器时,如果不能事先确定知道定性概念的数字特征,可以不把定性概念的数字特征作为输入,而把代表这个定性概念的一个样本作为输入。这种情形下的云规则发生器算法如算法3所示。

### 算法3 单条件单规则发生器

输入:代表前件定性概念  $A$  的一个样本  $\{x_{A_i}\}_{i=1}^n$ ;

前件论域中的一个特定值  $x_A$ ;

代表后件定性概念  $B$  的一个样本  $\{x_{B_i}\}_{i=1}^m$ 。

输出:后件论域中的云滴  $x_B$  及其确定度  $\mu$ 。

(1) 根据算法2,利用样本  $\{x_{A_i}\}_{i=1}^n$  与  $\{x_{B_i}\}_{i=1}^m$  分别计算  $A$  与  $B$  的数字特征  $(Ex_A, En_A, He_A)$  与  $(Ex_B, En_B, He_B)$ ;

(2) 产生以  $En_A$  为期望,  $He_A^2$  为方差的正态随机数  $En'_A$ ;

(3) 计算确定度  $\mu = e^{-\frac{(x_A - Ex_A)^2}{2(En'_A)^2}}$ ;

(4) 产生以  $En_B$  为期望,  $He_B^2$  为方差的正态随机数  $En'_B$ ;

(5) 如果输入值激活规则前件的上升沿,则规则后件也选择上升沿,反之亦然,即

$$x_B = \begin{cases} Ex_B + En'_B \sqrt{-2 \ln \mu}, & x_A \geq Ex_A \\ Ex_B - En'_B \sqrt{-2 \ln \mu}, & x_A < Ex_A \end{cases}$$

值得指出的是,单条件单规则发生器算法保证了推理过程中不确定性的传递,因为当前件论域中某一特定的输入值激活前件云时,产生的确定度  $\mu$  是随机的,而该确定度又作为后件云的输入,随机产生一个云滴  $(x_B, \mu)$ 。

复杂的定性规则可以用多维云对象来表达,因此复杂的不确定性推理也可以通过基于多维云的规则发生器表达。一个  $n$  条件单规则发生器中,由于输入值可能激活规则前件的上升沿或者下降沿,因此规则后件的输出可能有  $2^n$  种情况,故一个多条件单规则发生器需要对激活规则详细讨论。一种传统双条件单规则发生器算法如算法4所示[8]。

### 算法4 传统双条件单规则发生器

输入:分别代表二维前件定性概念  $A_1$  与  $A_2$  的一个样本  $\{x_{A1_i}\}_{i=1}^n$  与

$\{x_{A2_i}\}_{i=1}^n$ ;

前件论域中的一个特定值  $(x_{A1}, x_{A2})$ ;

代表后件定性概念  $B$  的一个样本  $\{x_{B_i}\}_{i=1}^m$ 。

输出:后件论域中的云滴  $x_B$  及其确定度  $\mu$ 。

(1) 根据算法2分别利用样本  $\{x_{A1_i}\}_{i=1}^n, \{x_{A2_i}\}_{i=1}^n$  与  $\{x_{B_i}\}_{i=1}^m$  计算  $A_1, A_2$  与  $B$  的数字特征,即:

$(Ex_{A1}, Ex_{A2}, En_{A1}, En_{A2}, He_{A1}, He_{A2})$  与  $(Ex_B, En_B, He_B)$ ;

(2) 产生以  $En_{A1}$  为期望,  $He_{A1}^2$  为方差的正态随机数  $En'_{A1}$ , 以  $En_{A2}$  为期望,  $He_{A2}^2$  为方差的正态随机数  $En'_{A2}$ ;

(3) 计算确定度  $\mu = e^{-\left[\frac{(x_{A1} - Ex_{A1})^2}{2(En'_{A1})^2} + \frac{(x_{A2} - Ex_{A2})^2}{2(En'_{A2})^2}\right]}$ ;

(4) 产生以  $En_B$  为期望,  $He_B^2$  为方差的正态随机数  $En'_B$ ;

(5) 输入值激活规则:

(5.1) 若  $x_{A1} \leq Ex_{A1}, x_{A2} \leq Ex_{A2}$ , 则

$$x_B = Ex_B - En'_B \sqrt{-2 \ln \mu};$$

(5.2) 若  $x_{A1} > Ex_{A1}, x_{A2} > Ex_{A2}$ , 则

$$x_B = Ex_B + En'_B \sqrt{-2 \ln \mu};$$

(5.3) 若  $x_{A1} \leq Ex_{A1}, x_{A2} > Ex_{A2}$ , 则令

$$\mu_1 = e^{-\frac{(x_{A1} - Ex_{A1})^2}{2(En'_{A1})^2}}, x_{B1} = Ex_B - En'_B \sqrt{-2 \ln \mu_1}$$

$$\mu_2 = e^{-\frac{(x_{A2} - Ex_{A2})^2}{2(En'_{A2})^2}}, x_{B2} = Ex_B + En'_B \sqrt{-2 \ln \mu_2}$$

并取  $x_B = (x_{B1} + x_{B2})/2$ ;

(5.4) 若  $x_{A1} > Ex_{A1}, x_{A2} \leq Ex_{A2}$ , 则令

$$\mu_1 = e^{-\frac{(x_{A1} - Ex_{A1})^2}{2(En'_{A1})^2}}, x_{B1} = Ex_B + En'_B \sqrt{-2 \ln \mu_1}$$

$$\mu_2 = e^{-\frac{(x_{A2} - Ex_{A2})^2}{2(En'_{A2})^2}}, x_{B2} = Ex_B - En'_B \sqrt{-2 \ln \mu_2}$$

并取  $x_B = (x_{B1} + x_{B2})/2$ 。

## 3 一种基于相依条件云的双条件单规则发生器

如果前件中的两个定性概念之间并不独立,则传统双条

件单规则发生器算法(即算法4)有不合理之处:考虑极端情况,假设算法4中前件的两个输入样本 $\{x_{A1_i}\}_{i=1}^n$ 与 $\{x_{A2_i}\}_{i=1}^n$ 的相关系数为1(甚至完全一样),则由 $A_1, A_2$ 合起来对 $B$ 进行推理的结果与由 $A_1, A_2$ 分别对 $B$ 进行推理的结果应当是一致的。换句话说,“if  $A_1, A_2$  then  $B$ ”应当等同于“if  $A_1$  then  $B$ ”,也等同于“if  $A_2$  then  $B$ ”。

但依据算法4,“if  $A_1, A_2$  then  $B$ ”与“if  $A_1$  then  $B$ ”或“if  $A_2$  then  $B$ ”并不等同,因为此时根据算法4的第(3)步算出的确定度会比真实情况偏小。

下面通过一个例子来说明传统双条件单规则发生器算法在条件云不独立时的不足之处。

设 $A$ 是参数为 $(0, 1, 0.1)$ 的一个云模型,依据算法1生成 $A$ 的一个长度为 $n$ 的样本 $\{x_{A_i}\}_{i=1}^n$ ,其中 $n=10000$ 。再设 $\{x_{B_i}\}$ 与 $\{x_{C_i}\}$ 分别是云 $B$ 与 $C$ 的样本,并且满足当 $1 \leq i \leq n$ 时

$$x_{B_i} = 0.2x_{A_i} + 5$$

$$x_{C_i} = 2x_{A_i} + 3$$

则可以利用云 $A, B$ 来预测云 $C$ 。

给定前件论域中的输入值 $x_A=0.5$ ,则由云 $B, C$ 与 $A$ 之间的关系知,与之对应的 $x_B=5.1, x_C=4$ 。下面先根据单条件单规则发生器算法(即算法3),分别用云 $A$ 输入0.5与云 $B$ 输入5.1预测 $C$ ;再根据双条件单规则发生器算法(即算法4),用云 $A, B$ 以 $(0.5, 5.1)$ 为输入值预测云 $C$ ,并将预测结果与真实值 $x_C=4$ 作比较。由于每一次的预测结果都不相同,为相互之间作比较,预测1000次,取中位数作为最终的预测值。预测结果与相对误差如表1所列。

表1 预测结果及相对误差

前件云	A	B	A, B
预测结果	3.9865	3.8283	4.7090
相对误差	0.0034	0.0429	0.1772

由表1可见,当前件云具有相互影响时,传统双条件单规则发生器算法的预测效果甚至远远不如单条件单规则算法。

为了弥补传统双条件单规则发生器算法处理基于相依条件云的云规则推理的不足之处,本文考虑对相互影响的两个云的云滴进行乘积作用,得到一个混合条件云,然后利用该混合条件云构造新的云规则发生器进行云推理。

这种具有交互作用的双条件单规则发生器的具体算法如算法5所示。

算法5 基于相依条件云的双条件单规则发生器算法  
输入:分别代表二维前件定性概念 $A_1$ 与 $A_2$ 的一个样本 $\{x_{A1_i}\}_{i=1}^n$ 与

$\{x_{A2_i}\}_{i=1}^n$ ;

前件论域中的一个特定值 $(x_{A1}, x_{A2})$ ;

代表后件定性概念 $B$ 的一个样本 $\{x_{B_i}\}_{i=1}^m$ 。

输出:后件论域中的云滴 $x_B$ 及其确定度 $\mu$ 。

(1) 构造一个新样本 $z_i = x_{A1} \times x_{A2}, i=1, 2, \dots, n$ , 并令 $z_0 = x_{A1} \times x_{A2}$ 。

(2) 根据算法2,由样本 $\{z_i\}_{i=1}^n$ 计算其代表的云 $Z$ 的数字特征 $(E_{X_Z}, E_{N_Z}, He_Z)$ 。

由样本 $\{x_{B_i}\}_{i=1}^m$ 计算 $B$ 的数字特征 $(E_{X_B}, E_{N_B}, He_B)$ ;

由样本 $\{x_{A1_i}\}_{i=1}^n$ 计算 $A_1$ 的数字特征 $(E_{X_{A1}}, E_{N_{A1}}, He_{A1})$ ;

由样本 $\{x_{A2_i}\}_{i=1}^n$ 计算 $A_2$ 的数字特征 $(E_{X_{A2}}, E_{N_{A2}}, He_{A2})$ ;

(3) 根据算法2,生成正态随机数:

$$E_{N_B'} \sim N(E_{N_B}, He_B^2), E_{N_Z'} \sim N(E_{N_Z}, He_Z^2)$$

$$E_{N_{A1}'} \sim N(E_{N_{A1}}, He_{A1}^2), E_{N_{A2}'} \sim N(E_{N_{A2}}, He_{A2}^2)$$

(4) 计算确定度

$$\mu_Z = e^{-\frac{(x_0 - E_{X_Z})^2}{2(E_{N_Z'})^2}}, \mu_{A1} = e^{-\frac{(x_{A1} - E_{X_{A1}})^2}{2(E_{N_{A1}'}^2)}, \mu_{A2} = e^{-\frac{(x_{A2} - E_{X_{A2}})^2}{2(E_{N_{A2}'}^2)}$$

(5) 输入值激活规则:

(5.1) 若 $x_{A1} \leq E_{X_{A1}}, x_{A2} \leq E_{X_{A2}}$ , 则

$$x_B = E_{X_B} - E_{N_B'} \sqrt{-2 \ln \mu_Z};$$

(5.2) 若 $x_{A1} > E_{X_{A1}}, x_{A2} > E_{X_{A2}}$ , 则

$$x_B = E_{X_B} + E_{N_B'} \sqrt{-2 \ln \mu_Z};$$

(5.3) 若 $x_{A1} \leq E_{X_{A1}}, x_{A2} > E_{X_{A2}}$ , 则令

$$x_{B1} = E_{X_B} - E_{N_B'} \sqrt{-2 \ln \mu_{A1}}, x_{B2} = E_{X_B} + E_{N_B'} \sqrt{-2 \ln \mu_{A2}}$$

并取 $x_{B1}, x_{B2}$ 的加权平均作为输出值:

$$x_B = \frac{|x_{A1} - E_{X_{A1}}| x_{B2} + |x_{A2} - E_{X_{A2}}| x_{B1}}{|x_{A1} - E_{X_{A1}}| + |x_{A2} - E_{X_{A2}}|}$$

(5.4) 若 $x_{A1} > E_{X_{A1}}, x_{A2} \leq E_{X_{A2}}$ , 则令

$$x_{B1} = E_{X_B} + E_{N_B'} \sqrt{-2 \ln \mu_{A1}}, x_{B2} = E_{X_B} - E_{N_B'} \sqrt{-2 \ln \mu_{A2}}$$

并取 $x_{B1}, x_{B2}$ 的加权平均作为输出值:

$$x_B = \frac{|x_{A1} - E_{X_{A1}}| x_{B2} + |x_{A2} - E_{X_{A2}}| x_{B1}}{|x_{A1} - E_{X_{A1}}| + |x_{A2} - E_{X_{A2}}|}$$

值得指出的是,这种具有交互作用云推理方法不仅考虑到了条件云之间的相互影响,也在一定程度上实现了对条件云的降维。

#### 4 基于相依条件云的二维云推理在预测中的应用

在数据挖掘中存在大量预测问题,如通过一个或几个属性值推测某个新属性的值、通过历史数据预测未来数据等。下面将本文提出的新的二维云推理方法应用于预测GDP。

分析美国从1961年1月到1984年12月每季度的出口额(Exports, EXP)、国民收入(Per capita Income, PCI)与国内生产总值(Gross domestic product, GDP)的数据发现,这三者之间有较高的相关性,因此用EXP与PCI来预测GDP。再注意到EXP, PCI与GDP都逐季度增加,故考虑通过EXP与PCI的增量来预测GDP的增量,进而预测GDP。

设 $\{e_i\}_{i=1}^{96}, \{p_i\}_{i=1}^{96}, \{g_i\}_{i=1}^{96}$ 分别是EXP, PCI与GDP的从1961年到1984年共96个季度的现实。则

$$\Delta e_i = e_{i+1} - e_i, \Delta p_i = p_{i+1} - p_i, \Delta g_i = g_{i+1} - g_i$$

就是EXP, PCI与GDP的增量。于是若预测出了 $\Delta g_n$ , 则由 $g_{n+1} = g_n + \Delta g_n$ 得到下一时刻GDP的预测值 $g_{n+1}$ 。

为比较分析本文提出的具有相依关系的云推理方法与传统方法之间的差异,下面一方面根据单条件单规则发生器算法,用EXP与PCI分别单独预测GDP;另一方面也分别根据传统双条件单规则发生器算法与本文提出的算法5,将EXP和PCI结合起来预测GDP。

下面以基于相依条件云的双条件单规则发生器算法(即算法5)为例说明预测GDP的具体步骤:

设要预测第 $n+1$ 个季度的GDP,即 $g_{n+1}$ , 则

输入:  $\{e_i\}_{i=1}^{n+1}, \{p_i\}_{i=1}^{n+1}, \{g_i\}_{i=1}^n$ 。

输出:  $g_{n+1}$ 。

(1) 计算增量 $\Delta e_i = e_{i+1} - e_i, \Delta p_i = p_{i+1} - p_i, i=1, 2, \dots, n$ 与 $\Delta g_i = g_{i+1} - g_i, i=1, 2, \dots, n-1$ ;

(2) 将 $\{\Delta e_i\}_{i=1}^{n-1}, \{\Delta p_i\}_{i=1}^{n-1}$ 作为前件定性概念的输入样本,  $\{\Delta g_i\}_{i=1}^{n-1}$ 作为后件定性概念的输入样本,  $(\Delta e_n, \Delta p_n)$ 作为前件论域的特定输入值,根据算法5计算得到 $\Delta g_n$ 的下一个预测值;

(3) 重复步骤(2) $m$ 次,得到 $m$ 个 $\Delta g_n$ 的预测值,取这些预测值的中位数作为 $\Delta g_n$ 的最终预测值 $\overline{\Delta g_n}$ ;

(4) 由 $g_{n+1} = g_n + \overline{\Delta g_n}$ 计算得到 $g_{n+1}$ 。

根据上述预测步骤, 并取  $m=1000$ , 预测 1984 年的 4 个季度的 GDP, 预测结果与相对误差如表 2 所列。

表 2 1984 年 4 个季度的 GDP 预测结果的相对误差

时间	算法 3 误差 (出口额)	算法 3 误差 (人均收入)	算法 4 误差	算法 5 误差
1984-I	0.0092	0.0038	0.0052	0.0010
1984-II	0.0052	0.0060	0.0079	0.0035
1984-III	0.0014	0.0103	0.0120	0.0079
1984-IV	0.0017	0.0032	0.0055	0.0029
2 范数	0.0108	0.0130	0.0162	0.0092

由表 2 可见, 本文提出的基于相依条件云的二维云推理方法(算法 5)的预测效果优于传统的双条件单规则(算法 4)、单条件单规则(算法 3)发生器算法。这也表明本文提出算法处理具有相依关系二维云推理是可行、有效的, 这也弥补了传统双条件单规则推理的不足。

### 参考文献

[1] 邸凯昌, 李德毅, 李德仁. 云理论及其在空间数据挖掘和知识发

现中的应用[J]. 中国图像图形学报, 1999, 4(11): 930-935  
 [2] 杨朝晖, 李德毅. 二维云模型及其在预测中的应用[J]. 计算机学报, 1998, 21(11): 961-969  
 [3] 李德毅, 杜鹃. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005: 335-361  
 [4] 陈昊, 李兵. 云推理方法及其在预测中的应用[J]. 计算机科学, 2011, 38(7): 209-224  
 [5] 蒋嵘, 李德毅, 陈晖. 基于云模型的时间序列预测[J]. 解放军理工大学学报, 2000, 1(5): 13-18  
 [6] 张仕斌, 许春香. 基于云模型的信任评估方法研究[J]. 计算机学报, 2013, 36(2): 422-431  
 [7] 李德毅, 史雪梅, 孟海军. 隶属云和隶属云发生器[J]. 计算机研究和发展, 1995, 32(6): 15-20  
 [8] 李德毅, 刘常昱. 论正态云模型的普适性[J]. 中国工程科学, 2004, 6(8): 28-33  
 [9] Chen Hao, Li Bing. Approach to uncertain reasoning based on cloud model[J]. Journal of Chinese Computer Systems, 2011, 32(12): 2449-2455

(上接第 82 页)

下对文献[10]中的 IEARA 算法和本文中的 newS 算法进行了分析比较(见表 2)。选用 CPU: AMD Athlon64(2800+), 1.0GB 内存配置的计算机进行仿真实验。

表 1 测试数据集信息

序号	数据集名称	数据集来源	样本容量	C	完备
1	Breast cancer wisconsin	UCI	699	9	否
2	Primary-tumor	UCI	339	17	否
3	Audiology-standardized	UCI	200	69	否
4	Crx-local	ROSE	690	15	否
5	House-votes-84data	ROSE	435	16	否
6	Lsd265	ROSE	265	35	是

表 2 基于容差关系的属性约简算法比较(ms)

序号	IEARA 算法 约简结果	IEARA 算法 约简时间	newS 约简结果	newS 约简时间	IEARA- newS
1	5	25781	5	22937	2844
2	16	44110	16	38000	6110
3	16	72359	16	62110	10249
4	13	68453	13	55484	12969
5	15	46578	15	38204	8374
6	9	30078	9	25312	4766

将两种算法约简时间进行绘图比较, 可以更加直观地比较两个算法的效率关系, 如图 1 所示。

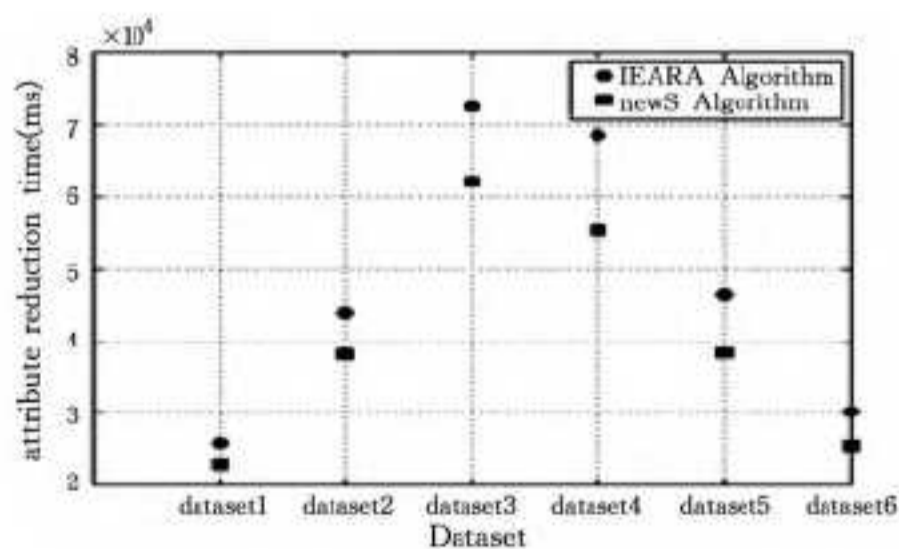


图 1 IEARA 算法与 newS 算法比较

从图中可以看出, newS 算法的约简时间总是低于 IEARA 算法的约简时间。

结束语 本文研究基于容差关系下的不完备信息系统属性约简算法。基于属性约简子集  $redu$  与属性全集  $CAttr$  之

间的关系, 将属性子集  $redu$  的相似类的近似度作为对属性子集整体结构的协调权重纳入模糊熵的结构中, 综合考虑属性子集  $redu$  与  $CAttr-redu$  之间必然具有的内在的联系, 定义了一种新的知识熵, 提出了一种新的应用近似模糊熵的不完备信息系统属性约简算法。在 ROSE 和 UCI data 数据集上进行仿真实验, 实验结果表明本文提出的 newS 算法不仅是可行的, 而且具有较高的属性约简时间效率。

### 参考文献

[1] 戴逸翔, 王雪, 李宣平, 等. 面向生物信息感知网络稀疏脑电测量的模糊粗糙情绪识别[J]. 仪器仪表学报, 2014, 35(8): 1693-1698  
 [2] 韩利强, 陈泽华, 曹长青, 等. TEP 故障诊断方法研究[J]. 计算机应用与软件, 2014, 31(7): 82-85  
 [3] 马文萍, 黄媛媛, 李豪, 等. 基于粗糙集与差分免疫模糊聚类算法的图像分割[J]. 软件学报, 2014, 25(11): 2675-2689  
 [4] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2008  
 [5] 曾晓辉, 文展. 不完备信息系统的属性约简算法[J]. 计算机工程, 2009, 35(24): 185-187  
 [6] 滕书华, 周石琳, 孙即祥, 等. 基于条件熵的不完备信息系统属性约简算法[J]. 国防科技大学学报, 2010, 32(1): 90-94  
 [7] 周志平, 刘付显. 一种改进的模糊粗糙集知识约简方法[J]. 计算机工程与应用, 2012, 48(18): 132-135  
 [8] 滕书华, 鲁敏, 杨阿锋, 等. 基于一般二元关系的粗糙集加权不确定性度量[J]. 计算机学报, 2014, 37(3): 649-665  
 [9] 韦碧鹏, 吕跃进, 李金海.  $\alpha$  优势关系下粗糙集模型的属性约简[J]. 智能系统学报, 2014, 9(2): 251-258  
 [10] 付昂, 王国胤, 胡军. 基于信息熵的不完备信息系统属性约简算法[J]. 重庆邮电大学学报, 2008, 20(5): 586-592  
 [11] 李秀红, 史开泉. 一种基于知识粒度的不完备信息系统的属性约简算法[J]. 计算机科学, 2006, 33(11): 169-170, 199  
 [12] 王青海. 互信息的序决策信息系统属性约简研究[J]. 计算机工程与设计, 2012, 37(7): 2822-2826