

# 六元语言值直觉模糊的金融决策推理方法

邹 丽<sup>1,2</sup> 王 颖<sup>1</sup> 谭雪微<sup>1</sup>

(辽宁师范大学计算机与信息技术学院 大连 116081)<sup>1</sup>

(南京大学计算机软件新技术国家重点实验室 南京 210093)<sup>2</sup>

**摘 要** 为了处理金融决策中的不确定性信息,采用直觉格值的方法来表述语言值。基于语言真值直觉格值系统,改进了一个个人金融决策辅助系统模型。该系统的推理方法是直觉模糊逻辑近似推理方法进行扩展,通过使用相似度的方法来处理模糊问题,实现了一种较为理想的不确定性推理方法。最后给出了一个实例,结果表明所提出的方法可以灵活有效地处理金融决策问题。

**关键词** 直觉格值,相似度,语言值因子,真值程度

**中图法分类号** TP182 **文献标识码** A

## Financial Decision-making Reasoning Method of 6-element Linguistic Truth-valued Intuitionistic Fuzzy System

ZOU Li<sup>1,2</sup> WANG Ying<sup>1</sup> TAN Xue-wei<sup>1</sup>

(School of Computer and Information Technology, Liaoning Normal University, Dalian 116081, China)<sup>1</sup>

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)<sup>2</sup>

**Abstract** In order to process the information with uncertainty in the financial decision-making, the present work uses the intuitionistic lattice implication algebra which can deal with the linguistic truth-valued. Based on the linguistic truth-valued intuitionistic fuzzy logic, a new model of personal financial decision auxiliary system was established. Reasoning mechanism of the system is the extension of the intuitionistic fuzzy logic and approximate reasoning method with the similarity method. To deal with the fuzzy problem, an uncertainty reasoning method was achieved. An example was given, which illustrates that the proposed method is flexible and effective when handling the financial decision-making problem.

**Keywords** Intuitionistic lattice valued, Similarity, Linguistic factors, Truth degree

### 1 引言

在对不确定性信息的表示和处理上,经典逻辑以及带一否定逻辑已不能满足信息处理的需求。对此,研究者们提出了直觉模糊集理论<sup>[1]</sup>。与经典的模糊集理论相比,直觉模糊集增加了非隶属度和犹豫度,这样就可以相对全面地刻画客观世界的模糊性本质。

知识的表示以及处理等是知识工程的基础。在知识表示方面,许多研究者提出了一些不同的研究思想与方法。年志刚等提出了混合知识表示和神经网络知识表示,并做了推理研究<sup>[2]</sup>;张传宏、傅秀芬等提出了产生式结合关系数据库的方法表示知识<sup>[3]</sup>,把专家系统理论知识应用于成绩分析系统中;郑宏亮等人提出了一种基于十元格蕴涵代数的知识表示方法<sup>[4]</sup>。在直觉模糊推理方面,雷英杰、王宝树等人基于模糊推理提出了直觉模糊近似推理<sup>[5]</sup>;夏佩伦等人对基于直觉模糊相似度量度的近似推理方法做了许多研究<sup>[6-8]</sup>;林娟等人基于集合的上下近似给出了一种粗糙集间的相似度量方法<sup>[9]</sup>;范成

礼等将直觉模糊粗糙集应用于多属性决策问题,提出了基于改进的直觉模糊粗糙集相似度的多属性决策方法<sup>[10]</sup>。

然而,用数值运算处理带有不确定性信息问题的模糊集理论具有局限性,因此,出现了基于文字计算的语言值信息的处理方法。为了建立一种既能处理不可比信息,又能同时处理命题的支持证据和反对证据的不确定性自动推理系统,邹丽等学者在格值逻辑系统的框架下建立了  $2n$  元语言真值直觉模糊代数,以及基于  $2n$  元语言真值直觉模糊格的直觉模糊命题逻辑系统,基于该逻辑代数系统做了大量的有关于不确定性推理与自动推理方法的研究<sup>[11,12]</sup>。刘德山等研究了语言真值直觉模糊命题逻辑系统的性质<sup>[13]</sup>,得到了其公理及推理规则语言真值直觉格值知识表示,使现实生活中一些不确定或犹豫的知识在计算机中能表示、运行,并推理出相应的结果,更加精确地进行推理、更加方便地解决许多生活中的问题。

本文基于语言真值直觉格值系统,研究了在一个金融投资决策具体实例中的语言值模糊知识表示及推理,最后改进了一个小型的金融决策系统。

本文受国家自然科学基金(61105059, 61175055, 61173100)资助。

邹 丽 女,博士,副教授,主要研究领域为多值逻辑与不确定性推理、智能信息处理, E-mail: zoulicn@163.com; 王 颖 (1990-), 女, 硕士生, 主要研究领域为多值逻辑与不确定性推理、智能信息处理; 谭雪微 (1990-), 女, 硕士生, 主要研究领域为多值逻辑与不确定性推理、智能信息处理。

## 2 六元语言真值直觉模糊代数

### 2.1 直觉模糊集

在直觉模糊集中, Atanassov<sup>[1]</sup>采用隶属度和非隶属度表示对象隶属于某一集合的程度。形式地, 直觉模糊集定义为:

**定义 1<sup>[14]</sup>** 设  $X$  是一个给定论域, 则  $X$  上的一个直觉模糊集  $A$  为  $A = \{ \langle x, u(x), v(x) \rangle \mid x \in X \}$ , 其中  $u(x): X[0, 1]$  和  $v(x): X[0, 1]$  分别代表  $A$  的隶属函数  $u(x)$  和非隶属函数  $v(x)$ , 且对于  $A$  上的所有  $x \in X, 0 \leq u(x) + v(x) \leq 1$  成立。

在直觉模糊集  $A$  中,  $\pi A(x) = 1 - uA(x) - vA(x) (x \in U)$  称为  $x$  隶属于  $A$  的怀疑度。在 Zadeh 的模糊集中, 若  $uA(x)$  是  $x$  隶属于  $A$  的隶属度, 则  $1 - uA(x)$  是非隶属度, 即  $uA(x) + 1 - uA(x) = 1$ 。从这个角度来看, 直觉模糊集是模糊集的推广。

**定义 2<sup>[14]</sup>** 设  $A$  和  $B$  是给定论域  $X$  上的直觉模糊子集, 即  $A = \{ \langle x, uA(x), vA(x) \rangle \mid x \in U \}$  和  $B = \{ \langle x, uB(x), vB(x) \rangle \mid x \in U \}$ , 则有

1)  $A \cap B = \{ \langle x, uA(x) \wedge uB(x), vA(x) \vee vB(x) \rangle \mid x \in X \}$ ;

2)  $A \cup B = \{ \langle x, uA(x) \vee uB(x), vA(x) \wedge vB(x) \rangle \mid x \in X \}$ ;

3)  $A' = \{ \langle x, uA(x), vA(x) \rangle \mid x \in U \}$ 。

论域  $U$  上的所有直觉模糊集的全体记为  $IFS(U)$ , 且直觉模糊集之间有如下序关系:  $A, B \in IFS(U), A \leq B$ , 当且仅当  $x \in U, uA(x) \leq uB(x)$  且  $vA(x) \geq vB(x)$ ; 自然地,  $A = B$  当且仅当  $A \leq B$  且  $B \leq A$ 。

### 2.2 语言真值直觉模糊格

$LI_{2n} = (VI_{2n}, \cup, \cap)$  是语言真值直觉模糊格<sup>[13]</sup>。图 1 可以理解为:  $\{ (h_1, f), (h_2, f), \dots, (h_n, f) \}$  构成的直线为横轴,  $\{ (h_1, t), (h_2, t), \dots, (h_n, t) \}$  构成的直线为纵轴, 则  $LI_{2n}$  的结构图是该坐标系中的图。

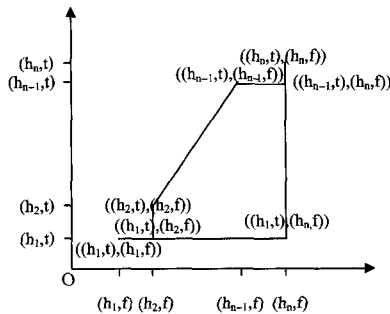


图 1 语言真值直觉模糊格  $LI_{2n}$  的结构图

**定义 3<sup>[14]</sup>** 对任意  $(h_i, t), (h_j, f) \in L_{V(n \times 2)}, ((h_i, t), (h_j, f))$  是一个语言真值直觉模糊对当且仅当  $i \leq j$ 。

**定理 1<sup>[14]</sup>** 对任意  $((h_i, t), (h_j, f)), ((h_k, t), (h_l, f)) \in LI_{2n}, ((h_i, t), (h_j, f)) \leq ((h_k, t), (h_l, f))$  当且仅当  $(h_i, t) \leq (h_k, t)$ , 且  $(h_j, f) \geq (h_l, f)$  当且仅当  $i \leq k$  且  $j \leq l$ 。

**定义 4<sup>[14]</sup>** 对任意  $((h_i, t), (h_j, f)), ((h_k, t), (h_l, f)) \in LI_{2n}, ((h_i, t), (h_j, f)) \leq ((h_k, t), (h_l, f))$  当且仅当  $i \leq k$  且  $j \leq l$  有

1)  $((h_i, t), (h_j, f)) \rightarrow ((h_k, t), (h_l, f)) = ((h_{\min(n, n-i+k, n-j+l)}, t), (h_{\min(n, n-i+l)}, f))$ ;

2)  $((h_i, t), (h_j, f)) \cup ((h_k, t), (h_l, f)) = ((h_{\max(i, k)}, t), (h_{\max(j, l)}, f))$ ;

3)  $((h_i, t), (h_j, f)) \cap ((h_k, t), (h_l, f)) = ((h_{\min(i, k)}, t), (h_{\min(j, l)}, f))$ ;

4)  $((h_i, t), (h_j, f))' = ((h_{n-j+1}, t), (h_{n-i+1}, f))$ 。

**定义 5<sup>[14]</sup>** 在六元语言真值格蕴涵代数  $L_{V(3 \times 2)} = \{ A, B, C, J, K, S \}$  中, 对任意  $((h_i, t), (h_j, f)) \in L_{V(3 \times 2)}, ((h_i, t), (h_j, f))$  称为一个六元语言真值直觉模糊对, 若  $((h_i, t), (h_j, f))$  满足  $(h_i, t)' \geq (h_j, f)$ , 其中, 运算“ $'$ ”为  $L_{V(3 \times 2)}$  中的逆序对合。

例 1 在六元语言真值格蕴涵代数  $L_{V(3 \times 2)}$  中(见图 2), 对于  $(B, J)$ , 有  $B' = K$ , 在图 2 中  $K \geq J$ , 因此,  $(B, J)$  是一个六元语言真值直觉模糊对。同理可得,  $(A, J), (B, K), (C, J), (C, K), (C, S)$  也是一个六元语言真值直觉模糊对, 因此可得出六元语言真值格蕴涵代数  $L_{V(3 \times 2)}$  一共有 6 个语言真值直觉模糊对。

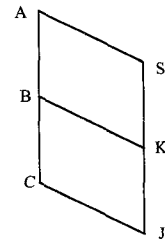


图 2  $L_{V(3 \times 2)}$  的结构图

**推论 1<sup>[14]</sup>** 语言真值直觉模糊格  $LI_6 = (VI_6, \cup, \cap)$  是一个有界分配格(见图 3), 其中,  $((h_3, t), (h_3, f)) = (A, J)$  和  $((h_1, t), (h_1, f)) = (C, S)$  分别为  $LI_6$  的最大元和最小元。

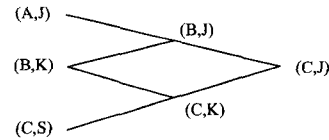


图 3 语言真值直觉模糊格  $LI_6$  的结构图

## 3 基于语言真值直觉格值的金融决策系统

下面以一个简单的金融投资辅助决策实际问题为例, 来说明如何运用六元语言真值直觉格值表达处理模糊知识。这个辅助决策程序的作用是根据用户的收入和目前的银行存款, 帮助用户决定是把钱存入银行还是投资到股票市场, 一些投资者也许会把他们的钱用在这两方面。依据以下原则<sup>[15]</sup>:

(1) 投资者如果存款少, 则不管他们的收入有多少都应该存入银行;

(2) 投资者如果存款多且拥有高收入, 有可能考虑股票投资;

(3) 投资者如果有较多存款且收入不低, 则把多余的钱大部分用于买股票, 少部分用于存款;

(4) 投资者如果有较少存款且收入不低, 则把多余的钱大部分用于存款, 少部分用于买股票。

注: 如果投资者的收入低或收入很少, 则视其为无多余的钱。

谓词符号采用大写字母, 函词符号采用小写字母。以个人为论域, 对于任意  $X$  属于论域, 引入必要的函词和谓词。规定:

$MUCH(\text{savings}(X))$ : 表示  $X$  的存款;

$MUCH(\text{income}(X))$ : 表示  $X$  的收入;

INVESTMENT( $X$ , stocks):表示用于买股票;

INVESTMENT( $X$ , savings):表示用于存款;

MORE(mf stocks( $X$ ), mf savings( $X$ )):表示用于股票的钱多于用于存款的钱;

MORE(mf savings( $X$ ), mf stocks( $X$ )):表示用于存款的钱多于买股票的钱。

(1)  $X$  的存款多的真值表示:  $((h_3, T), (h_3, F))$  MUCH (savings( $X$ ));

(2)  $X$  的存款较多的真值表示:  $((h_2, T), (h_3, F))$  MUCH (savings( $X$ ));

(3)  $X$  的存款较少的真值表示:  $((h_1, T), (h_2, F))$  MUCH (savings( $X$ ));

(4)  $X$  的存款少的真值表示:  $((h_1, T), (h_1, F))$  MUCH (savings( $X$ ));

(5)  $X$  的收入多的真值表示:  $((h_3, T), (h_3, F))$  MUCH (income( $X$ ));

(6)  $X$  的收入较多的真值表示:  $((h_2, T), (h_3, F))$  MUCH (income( $X$ ));

(7)  $X$  的收入较少的真值表示:  $((h_1, T), (h_2, F))$  MUCH (income( $X$ ));

(8)  $X$  的收入很少的真值表示:  $((h_1, T), (h_1, F))$  MUCH (income( $X$ ))。

将不同的投资策略用逻辑表达式来表示, 规则(1)可表示为:

$((h_1, T), (h_1, F))$  MUCH (savings( $X$ ))  $\rightarrow$  INVESTMENT( $X$ , savings)。

同样, 规则(2)可表示为:

$((h_3, T), (h_3, F))$  MUCH (savings( $X$ ))  $\wedge$   $((h_3, T), (h_3, F))$  MUCH (income( $X$ ))  $\rightarrow$  INVESTMENT( $X$ , stocks)。

规则(3)可表示为:

$((h_2, T), (h_3, F))$  MUCH (savings( $X$ ))  $\wedge$   $((h_2, T), (h_3, F))$  MUCH (income( $X$ ))  $\rightarrow$  INVESTMENT( $X$ , stocks)  $\wedge$  INVESTMENT( $X$ , savings)  $\wedge$  MORE(mf stocks( $X$ ), mf savings( $X$ ))。

规则(4)可表示为:

$((h_1, T), (h_2, F))$  MUCH (savings( $X$ ))  $\wedge$   $((h_2, T), (h_3, F))$  MUCH (income( $X$ ))  $\rightarrow$  INVESTMENT( $X$ , stocks)  $\wedge$  INVESTMENT( $X$ , savings)  $\wedge$  MORE(mf savings( $X$ ), mf stocks( $X$ ))。

### 3.1 基于语言真值直觉格值的金融决策知识推理

#### 3.1.1 语言真值直觉模糊相似性测度的公理化定义

**定义 6** 在语言真值模糊格  $LI_{2n}$  中, 对于任意直觉格值  $A = ((h_i, t), (h_j, f))$ ,  $B = ((h_k, t), (h_l, f)) \in LI_{2n}$ , 距离  $d(A, B) = |i - k| + |j - l|$ , 相似度  $S(A, B) = \begin{cases} ((h_{|n-d|}, t), (h_n, f)), & 0 \leq d \leq n-1 \\ ((h_1, t), (h_{|2n-d-1|}, f)), & n-1 < d \leq 2(n-1) \end{cases}$

**性质 1** 在语言真值直觉模糊格上(如图 3 所示), 语言真值直觉模糊格值  $A = ((h_i, t), (h_j, f))$ ,  $B = ((h_k, t), (h_l, f)) \in LI_{2n}$  有如下性质:

(1)  $((h_1, t), (h_1, f)) \leq S(A, B) \leq ((h_n, t), (h_n, f))$ ;

(2) 若  $(h_i, t) = (h_k, t)$ ,  $(h_j, f) = (h_l, f)$ , 则  $S(A, B) = ((h_n, t), (h_n, f))$  当且仅当  $A = B$ ;

(3)  $S(A, B) = S(B, A)$ ;

(4)  $S(A, B) = ((h_1, t), (h_1, f))$  当且仅当  $A = ((h_1, t), (h_1, f))$ ,  $B = ((h_n, t), (h_n, f))$  或  $A = ((h_n, t), (h_n, f))$ ,  $B = ((h_1, t), (h_1, f))$ ;

(5) 若  $A \leq B \leq C$ , 则  $S(A, C) \leq \min[S(A, B), S(B, C)]$ 。

证明: (1) 根据定义 6, 相似度

$$S(A, B) = \begin{cases} ((h_{|n-d|}, t), (h_n, f)), & 0 \leq d \leq n-1 \\ ((h_1, t), (h_{|2n-d-1|}, f)), & n-1 < d \leq 2(n-1) \end{cases}$$

当  $d=0$  时,  $S(A, B)$  最大, 为  $((h_n, t), (h_n, f))$ ; 当  $d=2(n-1)$  时,  $S(A, B)$  最小, 为  $((h_1, t), (h_1, f))$ 。

(2) 当  $A=B$ , 有  $(h_i, t) = (h_k, t)$ ,  $(h_j, f) = (h_l, f)$  时, 距离  $d(A, B) = |i - k| + |j - l| = 0$ 。根据定义 6,  $S(A, B) = ((h_n, t), (h_n, f))$ ; 当  $S(A, B) = ((h_n, t), (h_n, f))$  时, 距离  $d(A, B) = |i - k| + |j - l| = 0$ , 有  $(h_i, t) = (h_k, t)$ ,  $(h_j, f) = (h_l, f)$  即  $A=B$ 。

(3) 根据定义 6, 距离  $d(A, B) = |i - k| + |j - l|$ , 而  $d(B, A) = |k - i| + |l - j|$ , 则  $d(A, B) = d(B, A)$ , 相似度  $S(A, B) = \begin{cases} ((h_{|n-d|}, t), (h_n, f)), & 0 \leq d \leq n-1 \\ ((h_1, t), (h_{|2n-d-1|}, f)), & n-1 < d \leq 2(n-1) \end{cases}$   $d$  相同, 所以  $S(A, B) = S(B, A)$ 。

(4)  $((h_1, t), (h_1, f))$  和  $((h_n, t), (h_n, f))$  分别为  $LI_{2n}$  中的最小元和最大元, 距离最大, 所以相似度最小。

(5) 对任意直觉格值  $A = ((h_i, t), (h_j, f))$ ,  $B = ((h_k, t), (h_l, f))$ ,  $C = ((h_m, t), (h_n, f)) \in LI_{2n}$ , 当  $A \leq B \leq C$  时, 则有  $i + j \leq k + l \leq m + n$ ,  $d(A, B) = |i - k| + |j - l| = |i + j - (k + l)|$ ,  $d(A, C) = |i + j - (m + n)|$ ,  $d(B, C) = |k + l - (m + n)|$ ,  $d(A, C) \geq d(A, B)$ ,  $d(A, C) \geq d(B, C)$ 。由于距离越大, 相似度越小, 因此有  $S(A, C) \leq S(A, B)$ ,  $S(A, C) \leq S(B, C)$ , 即  $S(A, C) \leq \min[S(A, B), S(B, C)]$ 。

#### 3.1.2 语言真值直觉格值近似推理方法

首先量化观测事实与规则前件的接近程度, 然后利用模式匹配选择哪些规则是否被激活, 其次利用基于相似度量构建的修正函数, 在推理阶段修正规则的后件得到推理结果<sup>[16]</sup>。在新提出的方法中, 相似度由语言真值直觉模糊对的距离得到, 应用领域为基于规则的推理系统。

例如一条简单的规则  $R: P \rightarrow Q$ , 其基本推理模型为:

规则:  $P \rightarrow Q$

输入:  $P'$

输出:  $Q'$

其中  $P$  与  $P'$  是论域  $X$  上的语言值直觉模糊对,  $Q$  与  $Q'$  是论域  $Y$  上的语言值直觉模糊对,  $X, Y \in LI_{2n}$ 。则语言真值直觉格值近似推理方法:

(1) 计算观测事实  $P$  (即推理的输入部分) 与规则前件  $P'$  的相似度  $S(P, P')$ 。

(2) 对每条规则预先设定阈值  $\tau_i$ , 比较  $S(P, P')$  与  $\tau_i$  的大小, 如果  $S(P, P') \geq \tau_i$ , 则应的规则被激活。同一观测事实可能有一条或多条规则被激活。

(3) 基于修正函数的推理规则输出。

如果规则被激活, 则以  $S(P, P')$  为参数构造修正函数  $V = f(S(P, P'))$  (赋值函数), 并用  $V$  来修正被激活规则后件  $Q$  的隶属度  $SQ$ , 如此可得到近似推理的结果  $Q'$ , 即  $VQ$ 。

值得注意的是:  $S(P, P')$ 、 $V$  与  $\tau_i$  的选择是与应用领域密切相关的, 因此系统进行实际运作之前需要进行实例研究, 而

且对不同相似度和修正函数的实验绝对有必要。如果出现令人不满意的结果,可以选择不同的  $S(P, P')$ 、 $V$  与  $\tau_i$ 。进一步的工作仍需对相似度和修正函数进行研究。

(4)对多条推理规则结果的集结。

如果有多条规则被同时激活,需要对每条规则推出的结论进行集结。

### 3.1.3 金融决策算法

**定义 7** 设  $\alpha_T$  为表 1 中对于模糊谓词 MUCH 被视为最真的调查数据,  $\epsilon_T$  为此数据的弹性幅度。  $\alpha_F$  为对于模糊谓词  $\sim$ MUCH 被视为最真的调查数据,  $\epsilon_F$  为此数据的弹性幅度。对于任意调查数据  $x$ ,  $V(M(x))$  为  $x$  对于模糊谓词 MUCH 的真值程度赋值函数,使得

$$V(M(x)) = \begin{cases} ((h_1, T), (h_2, F)), & x \leq \alpha_F + \epsilon_F \\ ((h_1, T), (h_3, F)), & \alpha_F + \epsilon_F < x < \alpha_T - \epsilon_T \\ ((h_2, T), (h_3, F)), & x \geq \alpha_T - \epsilon_T \end{cases}$$

表 1 收入和存款调查结果

地区	高收入(元/月) (±500)	低收入(元/月) (±100)	存款多(万元) (±2)	存款少(万元) (±1)
江苏	11000 以上	1340 以下	16.0 以上	8.2 以下
上海	14400 以上	2000 以下	21.0 以上	10.0 以下
安徽	5000 以上	920 以下	10.0 以上	5.6 以下
山东	7000 以上	1100 以下	12.4 以上	6.8 以下

**例 2** 对于高收入观念来说,上海对应综合数据最高,为 14400,我们视为对于“MUCH”最真,500 为其弹性幅度  $\epsilon_T$ ;对于低收入观念,安徽综合调查数据最低,为 920,我们视为对于“ $\sim$  MUCH”最真,100 为其弹性幅度  $\epsilon_F$ 。因而,对于收入来讲:

$$V(M(x)) = \begin{cases} ((h_1, T), (h_2, F)), & x \leq 1020 \\ ((h_1, T), (h_3, F)), & 1020 < x < 13900 \\ ((h_2, T), (h_3, F)), & x \geq 13900 \end{cases}$$

根据以上的推理和分析,给出六元语言真值直觉格值的金融决策算法如下:

- 1)根据用户收入与存款计算相应的语言真值直觉模糊对  $((h_i, t), (h_j, f))$ 。
- 2)根据语言值直觉模糊相似度计算事实  $((h_i, t), (h_j, f))$  与每一条规则前件  $((h_k, t), (h_l, f))$  的相似度  $S_x$ 。
- 3)比较相似度  $S_x$  与阈值  $\tau_i = ((h_m, t), (h_n, f))$  的大小,当  $S_x = \tau_i$  时,则该规则被激活,选择执行被激活的规则。
- 4)对多条推理规则结果的集结。如果只有一条规则被激活,则选择该规则为最优方案;如果有多条规则被同时激活,需要对每条规则推出的结论进行集结。

### 3.2 具体实现

将这些逻辑公式转化为模糊产生式规则。

**例 3** 假设现在有一人在江苏无锡生活,月薪 5000 元,家庭银行存款 12 万元,问其投资方案如何。根据对应真值程度函数,可分别得出:

$$V(\text{MUCH}(\text{savings}(X))) = V(\text{MUCH}(\text{savings}(12))) = ((h_1, T), (h_3, F))$$

$$V(\text{MUCH}(\text{income}(X))) = V(\text{MUCH}(\text{income}(5000))) = ((h_1, T), (h_3, F))$$

$$V(\text{MUCH}(\text{savings}(12))) \wedge V(\text{MUCH}(\text{income}(5000))) = ((h_1, T), (h_3, F))$$

首先规定阈值  $\tau = ((h_3, T), (h_3, F))$ ,事实与规则前件的距

离为 0,即事实与规则前件完全匹配。根据金融决策系统模型:

规则 1:  $((h_1, T), (h_1, F)) \text{MUCH}(\text{savings}(X)) \rightarrow \text{INVESTMENT}(X, \text{savings})$ 。

前提条件的匹配程度:  $d_1 = 2, S_1 = ((h_{13-21}, T), (h_3, F)) < \tau$ ,说明此人存款较多,所以规则无法执行。

规则 2:  $((h_3, T), (h_3, F)) \text{MUCH}(\text{savings}(X)) \wedge ((h_3, T), (h_3, F)) \text{MUCH}(\text{income}(X)) \rightarrow \text{INVESTMENT}(X, \text{stocks})$ 。

前提条件的匹配程度:  $d_2 = 2, S_2 < \tau$ ,说明此人不属于多存款、高收入,所以规则无法执行。

规则 3:  $((h_2, T), (h_3, F)) \text{MUCH}(\text{savings}(X)) \wedge ((h_2, T), (h_3, F)) \text{MUCH}(\text{income}(X)) \rightarrow \text{INVESTMENT}(X, \text{stocks}) \wedge \text{INVESTMENT}(X, \text{savings}) \wedge \text{MORE}(\text{mf stocks}(X), \text{mf savings}(X))$ 。

前提条件的匹配程度:  $d_3 = 1, S_3 = ((h_{13-11}, T), (h_3, F)) < \tau$ ,说明此人不属于较多存款、较高收入,所以规则无法执行。

规则 4:  $((h_1, T), (h_2, F)) \text{MUCH}(\text{savings}(X)) \wedge ((h_2, T), (h_3, F)) \text{MUCH}(\text{income}(X)) \rightarrow \text{INVESTMENT}(X, \text{stocks}) \wedge \text{INVESTMENT}(X, \text{savings}) \wedge \text{MORE}(\text{mf savings}(X), \text{mf stocks}(X))$ 。

前提条件的匹配程度:  $d_4 = 0, S_4 = ((h_{13-01}, T), (h_3, F)) = \tau$ ,说明此人属于较少存款、较高收入,满足前提条件,所以执行规则 4。

综上,此人可采取规则 4 进行理财。

在基于六元语言值直觉模糊逻辑的推理中,个体之间可以出现不可比的现象,这符合人类的直觉思维。该例子属于一个具体实例,说明语言真值直觉模糊对在金融投资系统中的用法。由于其考虑了正反两方面信息,知识表示更全面,因此决策结果会更有效。

**结束语** 基于直觉模糊逻辑,本文提出一种利用相似度的语言值直觉模糊推理方法,给出了相关的性质和推理方法。进一步改进一个小型金融决策辅助系统,以具体算例验证了该方法的有效性。

知识表示方法是多种多样的,语言真值直觉格值知识表示的应用前景非常广泛。本文基于语言真值直觉格值,给出一种拓展了的知识表示方法,并讨论其应用,其很多性质及推理方法有待于进一步深入研究。

### 参考文献

- [1] Atanassov K. Intuitionistic fuzzy sets [J]. Fuzzy Sets and Systems, 1986, 20(1): 87-96
- [2] 年志刚,梁式,麻芳兰,等. 知识表示方法研究与应用[J]. 计算机应用研究, 2007, 24(5): 235-236
- [3] 张传宏,傅秀芬,张俊芝. 知识与知识库在成绩分析专家系统中的应用[J]. 信息技术, 2009(8): 43-49
- [4] 郑宏亮,徐本强,邹丽. 一种基于十元格蕴涵代数的知识表示方法[J]. 计算机应用与软件, 2013, 30(1): 37-40
- [5] 雷英杰,王宝树,路艳丽. 基于直觉模糊逻辑的近似推理方法[J]. 控制与决策, 2006, 21(3): 306-310
- [6] 夏佩伦. 不确定性推理方法研究[J]. 火力与指挥控制, 2010, 35(11): 88-91
- [7] 海新,雷英杰. 基于加权相似性测度的直觉模糊近似推理[J]. 计

[8] 雷阳, 雷英杰, 冯有钱, 等. 基于直觉模糊推理的目标识别方法[J]. 控制与决策, 2011, 26: 1163-1174

[9] 林娟, 米据生, 解滨. 粗糙集两种相似性度量[J]. 计算机科学, 2015, 42(6): 97-100

[10] 范成礼, 邢清华, 邹志刚, 等. 基于直觉模糊粗糙集相似度的多属性决策方法[J]. 计算机工程与应用, 2014, 50(7): 121-124

[11] 张云霞, 崔晓松, 邹丽. 一种基于十八元语言值模糊相似矩阵的聚类方法[J]. 山东大学学报, 2013, 43(1): 1-7

[12] 邹丽, 谭雪微, 张云霞. 语言真值直觉模糊逻辑的知识推理[J].

[13] Liu De-shan, Yin Ming'e, Zou Li. Reasoning rules of linguistic truth-valued intuitionistic fuzzy propositional logic system[J]. CEA, 2011, 47(33): 62-64

[14] 邹丽. 基于语言真值格蕴涵代数的格值命题逻辑及其归结自动推理研究[D]. 成都: 西南交通大学, 2010

[15] 徐江, 潘正华. 金融投资决策中的模糊知识及其不同否定的表示与推理[J]. 计算机应用与软件, 2011, 28(3): 38-40

[16] 田野, 陈东锋, 雷英杰. 基于直觉模糊相似度量的近似推理方法[J]. 空军工程大学学报(自然科学版), 2007, 8(6): 81-82

(上接第 57 页)

其中,  $k$  为数据集中的准确的聚类数目,  $P_j$  表示数据集中第  $j$  个标记簇。因此,  $k$  和  $P_j$  都是预先就知道的。聚类熵是一个 0 到 1 的值, 0 表示微簇  $C_i$  由一个  $P_j$  完整包含了; 1 则表示微簇  $C_i$  均匀地包括了所有的人工标记簇, 这种情况就是一个很差的结果。因此, 聚类熵的值应该越接近于 0 越好。不同聚类算法的聚类熵值如表 3 所列。

表 3 不同聚类算法的聚类熵值

数据集	各聚类算法及相应的聚类熵值		
	K-Modes	K-Prototypes	APDA
zoo	0.161	0.15	0.0705
Heart Disease	0.283	0.22	0.201

从表 3 可以看出, 3 个聚类算法的聚类熵值有所不同, 本文提出的 APDA 算法的聚类熵值相比 K-Modes 和 K-Prototypes 算法在 zoo 数据集上提升了 50% 以上, 在 Heart Disease 数据集上提升了 10% 以上, 因此, APAD 算法更好地完成了数据聚类的任务。由于 K-Modes 和 K-Prototypes 算法在不同程度上忽略了分类属性值的整体分布情况, 使得前两个算法的聚类结果没有本文的算法好。

另一方面, 仿真实验对比了不同算法的运行时间。由于 K-Prototypes 算法的时间复杂度为  $O(nkt)$ , 其聚类速度较快, 能够快速处理大型数据集, 而且在执行效率方面, 相对于许多目前的混合属性数据聚类算法有优势。因此, 本文的算法选择 K-Prototypes 算法进行对比。两个算法的执行时间如表 4 所列。

表 4 不同算法的执行时间对比(单位: s)

数据集	算法	
	K-Prototypes	APDA
zoo	2	1.5
Heart Disease	5.8	5.7

从表 4 中的数据可以看出, 对于 zoo 数据集, APDA 算法在执行效率方面比 K-Prototypes 算法稍快, 但是在 Heart Disease 数据集上两者的效率几乎一致, 其稍稍好于 K-Prototypes 算法, 这是由于 K-Prototypes 算法需要随机地选择  $k$  个数据点作为初始中心点, 这个过程对于聚类算法的正确率和效率方面都有一定的影响。

**结束语** 本文提出了一种基于属性距离的混合属性近邻传播聚类算法, 它改进了传统混合属性数据点之间的距离公式, 消除了虚拟中心点对聚类结果带来的影响, 提高了聚类分辨率, 同时将近邻传播聚类算法由原来仅适用于数值属性聚类的数据集, 扩展到了适用于混合属性数据集的聚类问题中。将本文算法与经典的混合属性聚类算法进行仿真对比, 对不同聚类算法的聚类熵这一指标值进行比较, 证明了基于维度

属性距离的混合属性近邻传播算法确实较传统的混合属性聚类算法在聚类效果方面有了一定的提升。

### 参 考 文 献

[1] Tan P N, Steinbach M, Kumar V. 数据挖掘导论[M]. 范明, 范宏建, 等译. 北京: 人民邮电出版社, 2011

[2] Kaufman L, Rousseeuw P J. Finding Groups in Data: An Introduction to Cluster Analysis [M]. New York: John Wiley & Sons, 1990

[3] 黄德才, 沈仙桥, 陆亿红. 混合属性数据流的二重 k 近邻聚类算法[J]. 计算机科学, 2013, 40(10): 226-230

[4] Huang Zhe-xue. Clustering Large Data Sets with Mixed Numeric and Categorical Values[C]// Proceedings of PAKDD'97. Singapore, World Scientific, 1997: 21-35

[5] Chatzis S P. A Fuzzy C-Means-Type Algorithm for Clustering of Data with Mixed Numeric and Categorical Attributes Employing a Probabilistic Dissimilarity Functional[J]. Expert Systems with Applications, 2011, 38(7): 8684-8689

[6] 白天, 冀进朝, 何加亮, 等. 混合属性数据聚类的新方法[J]. 吉林大学学报(工学版), 2013, 43(1): 130-134

[7] Frey B J, Dueck D. Clustering by passing messages between data points[J]. Science, 2007, 315(5814): 972-976

[8] Qian Y, Yao F, Jia S. Band selection for hyperspectral imagery using affinity propagation[J]. IET Computer Vision, 2009, 3(4): 213-222

[9] Li G, Guo L, Liu T. Grouping of brain MR images via affinity propagation[C]// IEEE International Symposium on Circuits and Systems, 2009(ISCAS 2009). IEEE, 2009: 2425-2428

[10] Dueck D, Frey B J, Jovic N, et al. Constructing treatment portfolios using affinity propagation[M]// Research in Computational Molecular Biology. Springer Berlin Heidelberg, 2008: 360-371

[11] Sumedha M L, Weigt M. Unsupervised and semi-supervised clustering by message passing: soft-constraint affinity propagation[J]. The European Physical Journal B-Condensed Matter and Complex Systems, 2008, 66(1): 125-135

[12] 刘晓楠, 尹美娟, 李明涛, 等. 面向大规模数据的分层近邻传播聚类算法[J]. 计算机科学, 2014, 41(3): 185-188

[13] Furtlehner C, Sebag M, Zhang X. Scaling analysis of affinity propagation[J]. Physical Review E, 2010, 81(6): 066102

[14] Zhang X, Furtlehner C, Sebag M. Data streaming with affinity propagation[M]// Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2008: 628-643

[15] 张建朋, 陈福才, 李邵梅, 等. 基于密度与近邻传播的数据流聚类算法[J]. 自动化学报, 2014, 40(2): 277-288

[16] 王开军, 张军英, 李丹, 等. 自适应仿射传播聚类[J]. 自动化学报, 2007, 33(12): 1242-1246