Pure Mathematics 理论数学, 2020, 10(5), 393-403
Published Online May 2020 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2020.105049

Hans汉斯

Research on Textual Formal Context Classification Model Based on Three-Way Decision

Hua Mao*, Siqi Liang, Zhenyu Wu

School of Mathematics and Information Science, Hebei University, Baoding Hebei
Email: *mh@hbu.edu.cn

Received: Apr. 6th, 2020; accepted: Apr. 23rd, 2020; published: Apr. 30th, 2020

Abstract

As an effective model for data processing, the three-way decision has been widely applied to granular computing, concept lattices, and fuzzy mathematical theory since it was proposed. In particular, the combination with granular computing is a practical method for dealing with big data problems. In this paper, a new composite function is established by triangular fuzzy numbers, and the composite data is used to transform the text-type data appearing in the formal context. Then, using the ideas of the three-way decisions and thresholds in fuzzy mathematics, the transformed interval can be effectively enlarged or reduced. Finally, after repeated divisions, an interval division suitable for the actual situation is obtained. According to the textual data partitioning model proposed in this paper, a corresponding algorithm is proposed, and an example is used to prove the feasibility and applicability of the algorithm.

Keywords

Three-Way Decision, Textual Formal Context, Triangle Fuzzy Number, λ-Cut Set

基于三支决策的文本形式背景分类模型的研究

毛　华*，梁思齐，武振宇

河北大学，数学与信息科学学院，河北　保定
Email: *mh@hbu.edu.cn

收稿日期：2020年4月6日；录用日期：2020年4月23日；发布日期：2020年4月30日

*通讯作者。

文章引用: 毛华, 梁思齐, 武振宇. 基于三支决策的文本形式背景分类模型的研究[J]. 理论数学, 2020, 10(5): 393-403.
DOI: 10.12677/pm.2020.105049

摘　要

三支决策作为一种数据处理的有效模型，自提出以来，被广泛扩展应用到粒计算、概念格及模糊数学理论。特别是与粒计算的结合是处理大数据问题的一种实用方法。本文通过三角模糊数建立一个新的复合函数，并通过这个复合函数转化形式背景中出现的文本型数据。然后，利用三支决策的思想以及模糊数学中的阈值，可以将转化后的区间有效地放大或者缩小。最后，经过反复的划分，得到一种适合实际情况的区间划分。根据本文提出的文本型数据的划分模型，提出对应的算法，并用一个实例证明算法的可行性及适用性。

关键词

三支决策，文本型形式背景，三角模糊函数，λ-截集

1. 引言

三支决策(3WD)是 Yao [1]在 2009 年提出。由于三支决策思想是一种符合人类认知和思维模式的处理问题的方式，一经提出就在医学、工程、管理、信息等领域得到了广泛的应用[2] [3] [4]。近几年来在与粒计算结合处理多粒度信息系统中不同粒度下的信息方面，受到国内外学者的广泛关注，并取得一些重要成果[5] [6]。将三支决策的思想与文本分析相结合，能够在信息不足或者描述不明确的情况下，将有争议的部分延迟处理，有效避免了二支决策中由于信息不足引起的盲目决策问题。三支决策的最初的思想就是将整体分为三部分，不同的部分采取不同的处理方法。目前仍有一些研究人员认为这三部分必须相互独立，现在 Yao [7] [8]提出了一种新的 Three-way decision+ (3WD+)模型，发现三部分之间存在着内在联系，并不完全独立，三个部分之间可以有交叉，在一定条件下甚至可以相互转化。这个扩展极大促进了三支决策的发展。

1965 年 Zaded [9]的文章标志着模糊数学的诞生。随着计算机的快速发展，需要处理的数据规模越来越大。一些学者将模糊数学与粒计算、三支决策等理论结合，并取得一些成果。比如折延宏等[10]在多粒度决策信息系统中采用三支的方法，将多粒度决策系统的信息分层，然后讨论粒度协调性，以解决对象更新的问题；魏玲等[11]将三支决策与粒计算、形式概念分析结合，提出三支概念分析理论，并得到成功的应用。毛华等[12] [13]在三支概念基础上，提出一种新的概念格的分类，并研究了模糊概念格表达与测量，取得一些成果。

在一个多粒度信息系统中，不同粒度下的评价标准在很多时候并不相同，那么单个粒度中的形式背景必然不同。形式背景中属性值通常有数值型、区间型、文本型等。同类型的评价标准仅仅需要消除量纲的影响就可相互转化，如区间转区间、数值转数值这些都是相对容易的，相关方法也比较多。但是区间转数值、文本转数值等不同类型的评价标准之间的转换不仅涉及量纲，因此难度大大提升。尤其是相对于数值和区间型数据，文本型数据还具有主观性，因此，在转化为数值或区间形式时的难点会更多。同时，文本型数据又是生产、适合以及科研中最常见的一种数据形式之一，所以，文本值的数值化很有研究价值。

虽然关于数值型形式背景分类、统一化归的方法，现已有不少成果[14] [15] [16]。但是关于文本型的形式背景数值化方法，目前的研究成果相对较少，影响了含有文本型形式背景的应用与发展。为了解决这一问题，本文提供一种方法，既可以用于文本型形式背景详细地划分，从而得到一种数值化的结果，也可以用于在一个多粒度信息系统中，对于不同粒度，由于标度不同而导致的数值型和文本型等不同类型的形式背景，可以通过化归到同一度量标准的方法来处理这些不同类型的度量标准的情况。

然而在实际应用中，更多地会遇到无法准确判定的情况。比如在网络购物的评价过程中，在对关键词抓取之后，对于评价的系统评分的这个过程，就是文本的数值化过程。不同的用户的评判标准是不同的，所以，在此文本型数据的数值化过程就需要选择一个合适的阈值。本文结合三支与模糊数学的隶属度原则，将多粒度信息系统中出现的文本型粒度提取出来，然后通过语言标度分割和不同的阈值的选取，将原本分割后形成的三部分进一步地划分，从而达到符合用户需求的情况。本文提出的三支模型，无论是在适用性上，还是在应用范围上，都比传统的语言标度分割方法更具优势，在划分上也会更加精确。

2. 基础知识

本节将本文所需要用到的定义依次地介绍。

首先，需要介绍形式背景和对象与属性之间的关系。

定义 2.1 [11]形式背景 K 是一个三元组 $K=(G,M,I)$，其中 K 为所有对象的集合，M 为所有属性的集合，$I\subseteq G\times M$ 为 G 和 M 中元素之间的关系集合。对于 $g\in G,m\in M$，$(g,m)\in I$ 记为 gIm，表示“对象 g 拥有属性 m”。

对象的属性通常有一个统一的度量标准，但是在多粒度信息系统中，不同的粒度度量标准有所差异。属性值的表示形式通常有数值 $x,x\in R$，区间 $[a,b],a,b\in R$，以及文本 w，w 为文字语言，本文中将所有的数值数据和区间数据统称为数值型数据，非数值型数据称为文本型数据。尤其是涉及到文本型数据的度量的时候，应当将文本型转化为更容易量化的数值型。

定义 2.2 [11] 一个文本型形式背景 $(G,M,\tilde{W},\tilde{I})$。其中 G 是对象集，M 是属性集，$\tilde{W}$ 是文本型属性值的集合，$\tilde{I}$ 是它们之间的一个三元关系，定义为：$\tilde{I}\subseteq G\times M\times\tilde{W}$，当且仅当 $\forall g\in G,m\in M$ 有且仅有一个文本值 $\tilde{w}\in\tilde{W}$ 满足 $(g,m,\tilde{w})\in\tilde{I}$。

对于不同类型的形式背景，通常有不同的处理方法。对文本型形式背景，如何把文本值转化为数值或者区间，就需要用到三角模糊函数和语言标度分割原理。其中在文本值转化为三角模糊函数中，迫切需要定义左边 NEG 和右边 POS 的起始域的边界值分别记为 w^{M_L},w^{M_R}，这将在下面的定义中完成。

定义 2.3 [11] 将对象 $g_i\in G$ 在属性 $m_j\in M$ 下的属性值记为 w_{ij}，$w_{ij}\in\tilde{W}$。属性值 $w_{ij}\in\tilde{W}$ 所对应的状态集合记为 $X=\{P,N,B\}$。当属性值 $0\le w_{ij}\le w^{M_L}$ 时属于 N 即 NEG；当属性值 $w^{M_R}\le w_{ij}\le w^{U}$ 时属于 P，即 POS，中间状态属于 B 即 BET。

其次，在将转化文本值的过程中，还需充分了解隶属度方面的知识。

定义 2.4 [9] 给出映射 $\mu_A:X\to[0,1]$，$x\mapsto\mu_A(x)$，则称 μ_A 确定一个 X 的模糊子集 $\underset{\sim}{A}$。μ_A 称为 $\underset{\sim}{A}$ 的隶属函数，$\mu_A(x)$ 称为 x 对 $\underset{\sim}{A}$ 的隶属度。

隶属函数，是用于表征模糊集合的数学工具。为了描述元素 u 对 U 上的一个模糊集合的隶属关系，由于这种关系的不分明性，它将用从区间[0,1]中所取的数值代替 0，1 这两值来描述，表示元素属于某模糊集合的“真实程度”。

定义 2.5 [9] 给定一个论域 U，那么 U 到单位区间[0,1]的一个映射 $\mu_A:U\mapsto[0,1]$ 称为 U 上的一个模糊集，或 U 的一个模糊子集。

模糊集合这一概念的出现使得数学的思维和方法可以用于处理模糊性现象，从而构成了模糊集合论，

即模糊数学的基础。

为了将文本值表示成为数值形式，引入三角模糊数的定义如下：

定义 2.6 [9] 所谓给定论域 U 上的一个模糊集，是指对任何 $x\in U$，都有一个数 $\mu(x)\in[0,1]$ 与之对应，$\mu(x)$ 称为 x 对 U 的隶属度，μ 称为 x 的隶属函数。

设 s 和 u 分别为模糊数的下限和上限，m 为可能性最大的值，那么模糊数用 (s,m,u) 表示。

关于阈值及最大隶属度的相关定义如下：

定义 2.7 [9] 设 $\underset{\sim}{A}\in\mathcal{F}(X)$，对任意 $\lambda\in[0,1]$，记 $(\underset{\sim}{A})_\lambda \triangleq \underset{\sim}{A}_\lambda \triangleq \{x|\underset{\sim}{A}(x)\geq\lambda\}$ 称 A_λ 为 $\underset{\sim}{A}$ 的 λ-截集，λ 称为置信水平。又记 $(\underset{\sim}{A})_\lambda \triangleq \underset{\sim}{A}_\lambda \triangleq \{x|\underset{\sim}{A}(x)\geq\lambda\}$，称 A_λ 为 $\underset{\sim}{A}$ 的 λ-强截集。

定义 2.8 [9] 最大隶属度原则：设 $\underset{\sim}{A}^{(1)},\underset{\sim}{A}^{(2)},\cdots,\underset{\sim}{A}^{(n)}\in\mathcal{F}(X)$ 为 n 个标准模型，$x_0\in X$，如果存在 $i=\{1,2,3,\cdots,n\}$，使得隶属度 $\underset{\sim}{A}^{(i)}(x_0)=\max\left\{\underset{\sim}{A}^{(1)}(x_0),\underset{\sim}{A}^{(2)}(x_0),\cdots,\underset{\sim}{A}^{(n)}(x_0)\right\}$，则称 x_0 相对隶属于 $\underset{\sim}{A}^{(i)}$。

3. 新的文本型形式背景转换模型

对于含有文本属性值的形式背景，本节将给出一种文本值数值化的新模型。该模型采用三支决策思想与模糊数学相结合的方式，选取不同的阈值反复地进行循环，最终得到符合用户要求的三部分。

3.1. 定义

当在一个多粒度信息系统中遇到语言描述的属性值时，通常采用三角模糊函数依语言标度分割后，一个对象同时属于一个属性下的一到两栏，这样就造成非此即彼的情况。实际情况中，文本描述是比较模糊的，有些无法用数值准确地界定。基于文献[9]中给出的三角模糊函数最后得到的每个对象最多隶属于两栏中，而改进后的三角模糊函数的复合函数定义如下：

定义 3.1 设存在四个三角模糊数 $\tilde{w}_1=(s_1,m_1,u_1)$，$\tilde{w}_2=(s_2,m_2,u_2)$，$\tilde{w}_3=(s_3,m_3,u_3)$，$\tilde{w}_4=(s_4,m_4,u_4)$。任给 $x\in[0,1]$，分别对应的隶属度如下

$$\tilde{w}(x)=\begin{cases}x/w^M, & 0\leq x\leq w^{M_L}\\ \left(w^M-x\right)/\left(w^M-w^{M_L}\right), & w^{M_L}\leq x\leq w^M\\ \left(w^{M_R}-x\right)/\left(w^{M_R}-w^{M_L}\right), & w^M\leq x\leq w^{M_R}\\ \left(w^U-x\right)/\left(w^U-w^{M_R}\right), & w^{M_R}\leq x\leq w^U\end{cases}$$

则称为复合函数。

此处，1) w^{M_L} =“C 标准的最右边的点”；2) w^M =“B 标准的中点”；3) w^{M_R} =“A 标准的最左边的点”；4) $w^U=1$。

从定义 3.1 中可以看出，同一事件在一个评价标准下，会有三种不同的评价，假设这三种评价为 A，B，C。例如，在学生考试评价体系中，A 为优秀，B 为及格，C 为不及格。

3.2. 模型建立的思想

首先，对文本型数据描述的属性值，用三角模糊数依次表示为数值形式，得到一个对应于原形式背景的数值形式背景。

其次，在此数值形式背景下，利用改进后的三角模糊函数，对此文本值进行处理，得到一个对应于此形式背景中的属性值之几何坐标表示。其中横坐标表示：所有三角模糊数所在的[0,1]区间；纵坐标表示：三角模糊数区间上的点在相应的文本值的隶属度。

最后，在上面得到的几何坐标表示下，对于横坐标从左向右依次完成如下工作：

1) 从左到右第一个隶属度为 1 的点的横坐标记为 a，直至到一个隶属度为 1 的点的横坐标记为 b。$[0,a]=NEG$，$[b,1]=POS$，$(a,b)=BET$。

2) 在此几何表示中利用阈值和隶属度，并结合三支决策的思想，将横坐标的[0,1]区间分为三部分，记为 NEG,BET,POS 之后，在此三部分上讨论提供原形式背景的用户对背景内容划分的界定要求，达到用户对原形式背景中全体对象进行综合评价的要求。

以上从“首先”至“最后”所有步骤，实际上是复合函数的代数计算过程，通过这些步骤可得到一个基于用户需求的数值显示背景的三支划分。

三支决策中，通常划分的三部分，可以每部分都有对象，也可以其中某一部分为空。例如，如果顾客对某一项的评价全部为满意，那么差的那部分就为空。故在用模糊函数对文本型形式背景进行划分的时候，不必拘泥于三部分一定要含有相应的区间或元。当三支的其中一部分是空集的时候，可在下一步划分的时候，将其他部分划过去，也可将其看作三支的一个特殊情况。

3.3. 模型实现的算法过程

对于文本型形式背景三角模糊数的复合函数的算法如下：

算法 1 基于 λ 的三支划分

输入：文本值 W_1,W_2,W_3,W_4，三角模糊函数 $\tilde{w}(x)$，阈值 λ，对象 a_i，属性 b_j，$i=1,2,\cdots,n$，$j=1,2,\cdots,m$；

输出： $NEG,POS,BET,NEG_1,POS_1,BET_1$；

Step0：初始化 $NEG_1=\varnothing,BET_1=\varnothing,POS_1=\varnothing$，$i=1,j=1$；

Step1：将文本值表示成三角模糊数的形式：$\tilde{w}_1=(s_1,m_1,u_1)$，$\tilde{w}_2=(s_2,m_2,u_2)$，$\tilde{w}_3=(s_3,m_3,u_3)$，$\tilde{w}_4=(s_4,m_4,u_4)$，其中 $0\le s_1,s_2,s_3,s_4\le 1$，$0\le m_1,m_2,m_3,m_4\le 1$，$0\le u_1,u_2,u_3,u_4\le 1$；

Step2：开始计算符合函数中各点的坐标，将文本值在三角模糊函数上的第一个隶属度为 1 的点的横坐标记为 a，最后一个隶属度为 1 的点的横坐标记为 b，即分别是 m_1 和 m_4，则 $(a,b)\subseteq BET_1$，$[0,a]\subseteq NEG_1$，$[b,1]\subseteq POS_1$；

Step3：记 W_1 与 W_2 的交点为 (c,y_1)，W_3 与 W_4 的交点为 (d,y_2)；记 $y=\lambda$ 与三角模糊函数上的在点 (c,y_1) 的左右两边的交点横坐标记为 c_1 和 c_2；在点 (d,y_2) 的左右两边的交点的横坐标记为 d_1 和 d_2；

Step3.1：当 $0\le\lambda\le\min\{y_1,y_2\}$ 时，$c_1=c_1',c_2=c_2',d_1=d_1',d_2=d_2'$；在左边 $[c_1',c]$，隶属度 $\tilde{A}^{(1)}(W_1)>\tilde{A}^{(1)}(W_2)>\lambda$，则 $[c_1',c]\subseteq NEG_2$；$[c_2',c]$ 中，因 $\tilde{A}^{(1)}(W_1)<\tilde{A}^{(1)}(W_2)$，故 $[0,c_2']\not\subset NEG_2$。右边，$[d,d_2']$ 中，有 $\tilde{A}^{(1)}(W_4)>\tilde{A}^{(1)}(W_3)>\lambda$，故 $[d,d_2']\subseteq POS_2$；$[d_1',d]$ 中，因 $\tilde{A}^{(1)}(W_4)<\tilde{A}^{(1)}(W_3)$，故 $[d_1',d]\not\subset POS_2$。

故 $0\le\lambda\le\min\{y_1,y_2\}$ 时，$[0,c]\subseteq NEG_2$，$[d_2',1]\subseteq POS_2$，$(c,d_2')\subseteq BET_2$；

Step3.2：当 $\max\{y_1,y_2\}\le\lambda\le 1$ 时，$c_1=c_1'',c_2=c_2'',d_1=d_1'',d_2=d_2''$；在左边 $[c_1'',c]$，隶属度 $\tilde{A}^{(2)}(W_2)<\tilde{A}^{(2)}(W_1)<\lambda$，则 $[c_1'',c]\not\subset NEG_2$，故 $[0,c_1'']\subseteq NEG_2$。右边 $[d,d_2'']$ 中，$\tilde{A}^{(2)}(W_3)<\tilde{A}^{(2)}(W_4)<\lambda$，故 $[d,d_2'']\not\subset POS_2$。

故 $\max\{y_1,y_2\}\le\lambda\le 1$ 时，$[0,c_1'']\subseteq NEG_2$，$[c_1'',d_2'']\subseteq BET_2$，$[d_2'',1]\subseteq POS_2$；

Step3.3：当 $y_1<\lambda<y_2$ 时，$c_1=c_1''',c_2=c_2''',d_1=d_1''',d_2=d_2'''$；在左边 $[c_1''',c]$，中隶属度 $\tilde{A}^{(3)}(W_2)<\tilde{A}^{(3)}(W_1)<\lambda$，故 $[c_1''',c]\not\subset NEG_2$；右边 $[d,d_2''']$，中隶属度 $\lambda<\tilde{A}^{(3)}(W_3)<\tilde{A}^{(3)}(W_4)$，故 $[d,d_2''']\subseteq POS_2$；在 $[d_1''',d]$ 中，隶属度 $\lambda<\tilde{A}^{(3)}(W_4)<\tilde{A}^{(3)}(W_3)$，故 $[d_1''',d]\not\subset POS_2$。

故 $y_1<\lambda<y_2$ 时，$[0,c_1''']\subseteq NEG_2$，$[c_1''',d]\subseteq BET_2$，$[d,1]\subseteq POS_2$；

Step3.4：当 $y_2<\lambda<y_1$ 时，$c_1=c_1^4,c_2=c_2^4,d_1=d_1^4,d_2=d_2^4$；在左边 $[c_1^4,c]$，中隶属度 $\lambda<\tilde{A}^{(4)}(W_2)<\tilde{A}^{(4)}(W_1)$，故 $[c_1^4,c]\subseteq NEG_2$；在 $[c,c_2^4]$ 中，隶属度 $\lambda<\tilde{A}^{(1)}(\tilde{w}_1)<\tilde{A}^{(2)}(\tilde{w}_2)$，故 $[c,c_2^4]\not\subset NEG_2$；左边 $[d,d_2^4]$ 中，隶属度 $\tilde{A}^{(4)}(W_1)<\tilde{A}^{(4)}(W_2)<\lambda$，故 $[d,d_2^4]\not\subset POS_2$。

故 $y_2<\lambda<y_1$时，$[0,c]\subseteq NEG_2$，$\left[c,d_2^4\right]\subseteq BET_2$，$\left[d_2^4,1\right]\subseteq POS_2$；

Step4： 当$i<n$时，$i=i+1$；否则转 Step5；

Step5： 当 $j<m$时，$j=j+1$；否则转 Step6；

Step6： 输出 $NEG=NEG_2, BET=BET_2, POS=POS_2, NEG_1, BET_1, POS_1$。

算法 2 最优三支划分算法

输入： 阈值$\{\lambda_1,\lambda_2,\cdots,\lambda_t\}$，其中$t=1,2,\cdots,s$，$\Delta=|y_1-y_2|/10$，$\lambda_{t+1}=\lambda_t+\Delta$，$NEG'$和$NEG''$其中$NEG'\subseteq NEG''\subseteq NEG_1$；

输出： NEG, BET, POS；

Step0： 初始化当$t=1$；

Step1： 调取算法 1；

Step2： 当$|NEG'|\le|NEG|$或者$|NEG''|\ge|NEG|$时，算法结束；否则$t=t+1$，转 Step2；

Step3： 输出 NEG, BET, POS。

注：NEG'选取的是含于算法 1 中所得到的NEG_1，选取的方式是按用户对NEG这一部分的需求。

在划分过程中，用户可得到NEG, BET, POS三部分。实际上，这三部分是由用户对其中一部分限定了范围而得到的。

比如用户需要“优秀”、“合格”和“不合格”三部分，但是对其中某一部分的范围有所要求。对于不同的要求都可以用上述算法得到。

当对“优秀”这部分要求的范围为 0.3 时，可以将算法 2 中的NEG'替换成为POS'，并且$|POS'|=0.3$，即$POS'=[0.7,1]$。选取不同的阈值λ，对算法 1 中的NEG_1, BET_1, POS_1，再次划分直至得到符合用户需求的最佳划分。当对POS的要求不同时，选取的阈值也不同，则最后出来的三部分自然不同。

当用户需要得到中间部分即“合格”这部分时，同样可以用以上算法，将NEG'替换为BET'通过选取不同的阈值得到符合需求的最佳划分。

3.4. 算法分析

当用户需要得到中间部分即“合格”这部分时，同样可以用以上算法，将NEG'替换为BET'通过选取不同的阈值得到符合需求的最佳划分。本节分三部分完成：一是算法的正确性、二是算法的复杂度分析、三是与已有成果的比较。

3.4.1. 算法的正确性

算法主要思想是，通过取不同的阈值对得到的三部分，不停地进行收缩或者放大以便达到用户需要。下面分别对算法 1 和算法 2 加以分析。

算法 1 是将文本值用三角模糊数表示出来，然后对三角模糊数进行初步的分层得到NEG_1，BET_1和POS_1。然后选取一个阈值λ再次进行分层。根据λ取值不同分别对应四种不同的处理方式，在对应的情况下依据隶属度原则再次进行分层。所以算法 1 在有限步内可以结束。

定理 3.1 当给出一个初始阈值λ对已知三角模糊数属性值分层，一定存在一个梯度Δ可以在有限步内使$NEG'\subseteq NEG$或者$NEG\subseteq NEG''$。

定理 3.1 是对算法 2 的总结。要说明算法 2 能在有限步内结束，即证明定理 3.1 的正确性。

证明：当$\lambda=\dfrac{|y_1-y_2|}{2}$时，对应的是算法 1 中的第一种情况，此时输出的$NEG\supseteq NEG_1$。在算法 2 中阈值$\lambda$取值逐步上升，当$\lambda$取值在$y_1$，$y_2$中间时，对应的算法 1 中的第三、四情况，此时依据算法NEG的区间会逐渐减小，直至λ取值增至大于y_1和y_2，此时NEG区间仍然在缩小。故随着阈值的逐渐增大，

NEG 的区间逐渐在减小，由于算法 2 中的限定条件为$\left|NEG'\right|\le\left|NEG\right|$或者$\left|NEG''\right|\ge\left|NEG\right|$，故最后一定存在一个满意阈值 λ (该阈值为所有满足要求的最大阈值)使 $NEG\subseteq NEG''$，或者存在一个满意阈值 λ (该阈值为所有满足要求的最小阈值)使 $NEG'\subseteq NEG$。当 λ 的第一个取值较大进行缩小时，同上理可以得到一个满意阈值，将文本值划分为需要的三部分。

之后，算法 2 根据对 NEG，BET 或者 POS 的输出要求，选择对 λ 梯度放大或者缩小，再次调用算法 1 直至得到符合最终要求的三部分。

3.4.2. 算法的分析

第二，算法的复杂度分析：

在算法的正确性分析可知，算法 1 的复杂度为 $O(nm)$，算法 2 的复杂度为：$O(s)$，其中 $\max\{s\}=\left[\frac{10}{|y_1-y_2|}\right]$。

故算法整体的复杂度为 $O(mns)$。

第三，与已有成果的比较：

在多粒度信息系统中，处理文本值的常用方法实际上是一个二支划分的过程。而本文将三支决策的方法引入文本值数值化的划分中。与二支方法相比，上述算法模型主要有以下突出优势：

通过三支的思想，解决了传统方法在文本值数值化过程中，直接将区间一分为二的做法。从隶属度的角度，将区间分为三部分，得到一种更符合实际情况的划分。并建立一个复合函数，将三角模糊数表示的文本值用几何形式直观的表示出来。通过阈值的变化，能过扩大或者缩小划分出的某一区间，能够根据需求，灵活的选择不同的阈值，从而扩大应用范围。

3.5. 实例

下面给出一个实例说明算法的实际操作步骤。

例 1：在某购物网站的售后评价体系中顾客可分别对物流服务 a_1、货物质量 a_2、卖家态度 a_3、售后服务 a_4 四项评价，四项评价档次分别有“满意、一般、待改进、较差”，顾客可根据实际情况对此次购物体验做出相应的评价。评价提交后平台的系统会对此商品做出一个整体的评价，但平台出于长效发展的考虑，在整体的服务质量都不高的情况下，可将评价标准降低，即扩大 POS 这部分的范围或者缩小 NEG 的范围；当平台出于可持续发展的考虑应该使 BET 的部分在 NEG，BET 和 POS 三部分占比最大。故用户(即此例中的购物平台)可以根据自身的需要调整阈值将评价分为三部分。

下面为某平台提供的近几位购买某一产品的顾客的评价结果，如表 1。

从表 1 中可以看出，评价为“一般”和“满意”的较多。为了提高平台的货物质量 a_2，在将其区间化的过程中可适当扩大 NEG 部分的范围，以达到促进发展的目的。

第一步，将文本值转化为相应的三角模糊数：W = {较差，一般，待改进，满意}。其中：“较差” = W_1 = (0,0.2,0.4)，“待改进” = W_2 = (0.3,0.5,0.7)，“一般” = W_3 = (0.4,0.6,0.8)，“满意” = W_4 = (0.6,0.8,1)。

根据模糊数的定义将文本型背景转化为数值型(见表 2)。

第二步，根据语言标度分割原理和给定的三角模糊数，可以得出如图 1 的三角模糊数的图示。对于 $[a,b]$ 区间内，在横坐标轴上，从左向右，令 $x/w^M=\left(w^M-x\right)/\left(w^M-w^{M_L}\right)$。得到 $x=c$ 确定点 (c,y_1)，其中 $y_1=\tilde{w}(c)$；

令 $\left(w^{M_R}-x\right)/\left(w^{M_R}-w^{M_L}\right)=\left(w^U-x\right)/\left(w^U-w^{M_R}\right)$，得到 $x=d$。

确定点 (d,y_2)，其中 $y_2=\tilde{w}(d)$。且 $y_1\ne y_2$。

Table 1. Original formal context
表 1. 原形式背景

顾客评价	a_1	a_2	a_3	a_4
b_1	满意	一般	一般	满意
b_2	一般	满意	满意	一般
b_3	满意	满意	一般	一般
b_4	较差	待改进	一般	待改进

Table 2. Numeric formal context
表 2. 数值形式背景

顾客评价	a_1	a_2	a_3	a_4
b_1	[0.6, 0.8, 1]	[0.4, 0.6, 0.8]	[0.4, 0.6, 0.8]	[0.6, 0.8, 1]
b_2	[0.4, 0.6, 0.8]	[0.6, 0.8, 1]	[0.6, 0.8, 1]	[0.4, 0.6, 0.8]
b_3	[0.6, 0.8, 1]	[0.6, 0.8, 1]	[0.4, 0.6, 0.8]	[0.4, 0.6, 0.8]
b_4	[0, 0.2, 0.4]	[0.3, 0.5, 0.7]	[0.4, 0.6, 0.8]	[0.3, 0.5, 0.7]

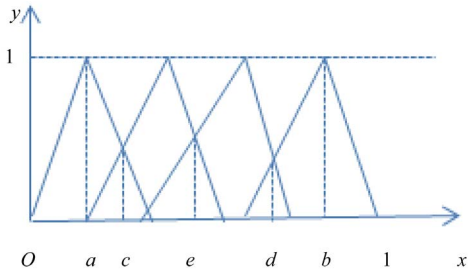

Figure 1. Collective representation of attribute values in **Table 2**
图 1. **表 2** 中属性值的集合表示

另外， $O=0, a=w^{M_L}, e=w^{M}, b=w^{M_R}, 1=w^{U}$ 。

从图 1 中可以看出，改进后的三角模糊函数使一个对象可以同时归结到一个属性下的一到三栏中甚至四栏中。可根据语言标度分割原理，将三角模糊函数的图画出来，当它在某一栏的隶属度明显高于其他栏的时候，如果在图的左边则记为 *NEG*，如果在图右边则记为 *POS*；当同时属于两栏或三栏时，记为 *BET*，然后对 *BET* 这一部分进行分析细化，再根据实际情况给出合理的划分。

第三步，在图 1 中，将第一个拐点的横坐标记为 a，最后一个拐点的横坐标记为 b，函数图像的第一个交点的横坐标记为 c，纵坐标记为 y_1， $W_3=$“一般”和 $W_2=$“待改进”，三角模糊函数的交点为横坐标记为 e；将最后一个交点的横坐标记为 d，纵坐标记为 y_2。由函数图像可明显地看出，当 $x\in[0,a]$ 时，该区间在“较差”这一属性下的隶属度明显高于其他，故可记 $[0,a]\subseteq NEG$；同理可记 $[b,1]\subseteq POS$.而 a 与 b 之间的部分可暂记为 *BET*，即 $(a,b)\subseteq BET$ 。

第四步，根据实际情况，需要给出一个阈值 $\lambda(0\le\lambda\le1)$。

当 λ 取值较小，即 $0\le\lambda\le\min\{y_1,y_2\}$ 时，如图 2 所示。

记 $y=\lambda$ ，在三角模糊函数图像除 $[0,a]$ 区间上的第一个交点的横坐标记为 c'。显然在左边 $[a,c]$ 中，在 $\tilde{w}_1$ 这一三角模糊函数上的值明显比 $\tilde{w}_2$ 上的值隶属度要大的多，故 $[a,c]\subseteq NEG$；在右边，记 $y=\lambda$ ，在

三角函数图像除$[d,b]$区间上的最后一个交点的横坐标记为d，同理$[d,b]\subseteq POS$。

综上，当$0\le\lambda\le\min\{y_1,y_2\}$时，$[0,c]\subseteq NEG$，$[d,1]\subseteq POS$，$(c,d)\subseteq BET$。

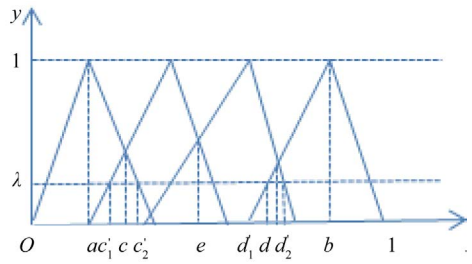

Figure 2. Graphical of triangular fuzzy number when threshold is small
图 2. 阈值较小时三角模糊数图示

当λ取值较大时，即$\max\{y_1,y_2\}\le\lambda\le1$时，如图 3 所示。

记$y=\lambda$，在三角模糊函数图像除$[0,a]$区间上的第一个交点的横坐标记为c''，与除$[b,1]$外的最后一个交点的横坐标记为d''。在左边，显然函数在$[a,c'']$区间上的取值大于λ，故可将$[a,c'']$并入NEG中；在右边，同理将$[d'',d]$并入POS中。至于中间的$[c'',c]$和$[d,d'']$这两部分，暂时还放在BET中。

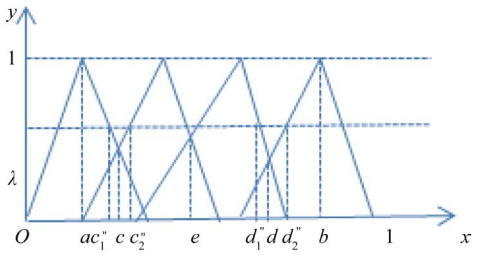

Figure 3. Graphical of triangular fuzzy number when threshold is big
图 3. 阈值较大时三角模糊数图示

综上，当$\max\{y_1,y_2\}\le\lambda\le1$时，$[0,c'']\subseteq NEG$，$[d'',1]\subseteq POS$，$(c'',d'')\subseteq BET$。

当$\min\{y_1,y_2\}<\lambda<\max\{y_1,y_2\}$时，此时又分为两种情况$y_1<y_2$和$y_1>y_2$，分别如下图 4 和图 5 所示：

同上记$y=\lambda$，与三角模糊函数图像的交点的横坐标分别为c''' (除$[0,a]$区间上的第一个交点的横坐标)和d''' (除$[b,1]$外的最后一个交点的横坐标)。

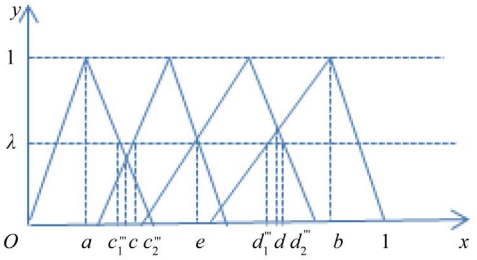

Figure 4. Graphical of triangular fuzzy number ($y_1<y_2$)
图 4. 三角模糊数图示($y_1<y_2$)

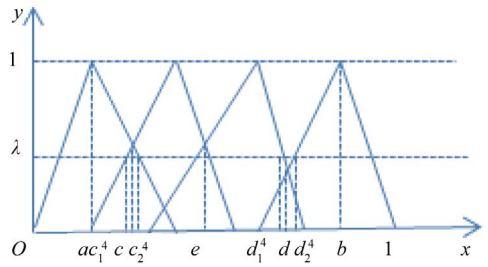

Figure 5. Graphical of triangular fuzzy number ($y_1 > y_2$)
图 5. 三角模糊数图示($y_1 > y_2$)

当 $y_1 < y_2$ 即 $y_1 < \lambda < y_2$ 时，由图 4 可以看出，在 $x \in [c''', c]$ 中时，$\tilde{w}_1$ 在三角模糊函数上的隶属度大于在 $\tilde{w}_2$ 上的隶属度，故 $[c''', c] \subseteq NEG$。在右边，$x \in [d, d''']$ 中，显然 $\tilde{w}_4$ 在三角模糊函数上的隶属度大于 $\tilde{w}_3$，将 $y = \lambda$ 与 $\tilde{w}_4$ 的另一个交点记为 d_1，当 $x \in [d_1, d]$ 时，虽然隶属度上 $\tilde{w}_4$ 比 $\tilde{w}_3$ 小，但是在 $\tilde{w}_4$ 的隶属度大于 λ，所以仍然有 $[d_1, d] \subseteq POS$。

综上，当 $y_1 < \lambda < y_2$ 时，$[0, c] \subseteq NEG$，$[d_1, 1] \subseteq POS$，$(c, d_1) \subseteq BET$。

当 $y_1 > y_2$，即 $y_2 < \lambda < y_1$ 时，由图 5 可以看出，在 $[c_2''', c]$ 中，$\tilde{w}_1$ 在三角模糊函数上的隶属度大于在 $\tilde{w}_2$ 上的隶属度，必然有 $[0, c] \subseteq NEG$；将 $\tilde{w}_1$ 与 $y = \lambda$ 的另一个交点的横坐标记为 c_2，在 $[c, c_2]$ 中，虽然 $\tilde{w}_1$ 在三角模糊函数上的隶属度小于在 $\tilde{w}_2$ 上的隶属度，但是大于 λ，所以 $[0, c_2] \subseteq NEG$。在右边同图 4 的分析可知 $[d, 1] \subseteq POS$。

综上可得，当 $y_2 < \lambda < y_1$ 时，$[0, c_2] \subseteq NEG$，$[d, 1] \subseteq POS$，$[c_2, d] \subseteq BET$。

第五步，选取新的 λ 重复第四步，直至 $|NEG| \geq 0.4$ 停止；

第六步，得到最终的 $NEG = [0, 0.4]$，$BET = [0.4, 0.7]$，$POS = [0.7, 1]$。

上述步骤是将所有文本型属性值转化为用三角模糊函数表述的区间形式，然后对属性 a_2 中区间型属性值的三支划分。可以容易地看出，此形式背景中 $i = 1, 2, 3, 4$，$j = 1, 2, 3, 4$，此计算中选取了一个阈值 λ，以及 $j = 2$ 的情况，故 $s = 1, m = 4, n = 1$，故此处的算法复杂度为 $O(mns) = O(4)$。

4. 结语

当一个多粒度信息系统中存在不同的度量标准时，为了方便决策，需要将所有的属性值转化在一个统一的度量标准下。而文本型数据的度量标准一直是量纲统一化的重点和难点，本文提出的这种方法可以用于文本型数据在数值化过程中的取值困难的问题。本文用三角模糊函数表示文本值，并定义了一个三角模糊数的复合函数，使得文本值从简单的模糊区间形式转化为直观的几何形式。从几何表示中可以看出，文本值在区间上的每个点对应的隶属度，继而利用阈值与三支决策的思想，将文本值对应的区间进行放大或缩小，最后得到用户所需的划分结果。

本文采用三支决策的思想，避免了与传统的语言标度划分方法在划分过于绝对的弊端。用三角模糊数表示文本值为后续区间的缩放提供了方便，定义的复合函数将区间形式的文本值表示转化成了几何形式，这种表现方式更加的直观。并且本文提供的模型，能够使用户根据需要合理的变换阈值，以得到所需结果。这种文本型形式背景的数值转化方法，在当今互联网的迅猛发展和企业管理员工评价机制上具有实际且广泛的应用。

由于本文处理的文本值的个数为 4 个，这是最常见的情况。但当文本值个数大于 4 个时，仍有很大的研究价值，需进一步探讨。

基金项目

基金项目：国家自然科学基金(No.61572011)、河北省自然科学基金(No.A2018201117)、河北大学研究生创新资助项目(No.hbu2019ss030)。

参考文献

[1] Yao, Y.Y., Wen, P., *et al*. (2009) Three-Way Decision: An Interpretation of Rules in Rough Set Theory. *Rough Sets and Knowledge Technology Proceedings*, **5589**, 642-649. https://doi.org/10.1007/978-3-642-02962-2_81

[2] 黄智力, 罗键. 属性值为三角模糊数的决策对象可能度关系模型[J]. 控制与决策, 2018, 33(11): 1931-1940.

[3] 谢莹. 上海市公交夜宵线评估与调整[J]. 交通与港航, 2016, 12(6): 46-50.

[4] 苗夺谦, 张清华, 等. 从人类智能到机器实现模型——粒计算理论与方法[J]. 智能系统学报, 2016, 11(6): 743-757.

[5] 姚一豫, 祁建军, 等. 基于三支决策的形式概念分析、粗糙集与粒计算[J]. 西北大学学报(自然科学版), 2018, 48(4): 477-487.

[6] 方宇, 闵帆, 等. 序贯三支决策的代价敏感分类方法[J]. 南京大学学报(自然科学), 2018, 54(1): 148-156.

[7] Hu, M.J. and Yao, Y.Y. (2019) Structured Approximations as a Basis for Three-Way Decisions in Rough Set Theory. *Knowledge-Based Systems*, **165**, 92-109. https://doi.org/10.1016/j.knosys.2018.11.022

[8] Van Vught, C.L., *et al*. (2018) Comparison of Corneal Model for Peripheral Vision Analysis. *Acta Ophthalmologica*, **96**, 36.

[9] Zadeh, L.A. (1987) Fuzzy Sets and Applications. John Wiley & Sons, New York.

[10] She, Y.H., Li, J.H., *et al*. (2015) A Local Approach to Rule Induction in Multi-Scale Decision Tables. *Knowledge-Based Systems*, **89**, 398-410. https://doi.org/10.1016/j.knosys.2015.07.020

[11] 魏玲. 三支决策与粒计算[J]. 西北大学学报(自然科学版), 2018, 48(4): 477.

[12] Mao, H. and Lin, G.M. (2017) Interval Neutrosophic Fuzzy Concept Lattice Representation and Interval-Similarity Measure. *Journal of Intelligent & Fuzzy Systems*, **33**, 957-967. https://doi.org/10.3233/JIFS-162272

[13] Mao, H., Zhao, S.F., *et al*. (2018) Relationships between Three-Way Concepts and Classical Concepts. *Journal of Intelligent & Fuzzy Systems*, **35**, 1063-1075. https://doi.org/10.3233/JIFS-17530

[14] Mao, H. (2017) Classification Lattices Are Geometric for Complete Atomistic Lattices. *Open Math*, **15**, 959-973. https://doi.org/10.1515/math-2017-0078

[15] Mao, H. (2012) Complete Atomistic Lattices Are Classification Lattices Lationships. *Algebra Universails*, **68**, 293-294. https://doi.org/10.1007/s00012-012-0200-5

[16] Mao, H. (2016) Characterizations of Atomistic Complete Finite Lattice Relative to Geometric Ones. *Miskolc Mathematical Notes*, **677**, 421-440. https://doi.org/10.18514/MMN.2016.677