

非均衡语言信息的计算方法及其在TOPSIS法中的应用

蔡 玫[†], 巩在武, 于小兵

(南京信息工程大学 经济管理学院, 南京 210044)

摘要: 大多用语言描述偏好的决策问题假设语言是均匀、对称分布的, 然而有些问题需要采用非均衡语言. 针对这一问题, 提出一种基于符号化方法的语言计算模型. 首先, 构造一种基于基础语言集合的加权图, 用图形中的点描述非均衡语言; 然后, 定义图形中任意两点间的曼哈顿距离公式, 用于计算非均衡语言的距离; 最后, 将其用于逼近于理想值的排序方法(TOPSIS), 并给出算例. 所提出的方法不仅图像化非均衡语言, 而且在解决TOPSIS问题时比欧氏距离测度更具优越性.

关键词: 非均衡语言集合; 逼近理想解排序法; 计算模型; 距离测度

中图分类号: C934; N945.25

文献标志码: A

A computational approach of unbalanced linguistic terms and its application in TOPSIS

CAI Mei[†], GONG Zai-wu, YU Xiao-bing

(School of Economics and Management, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Many decision problems using linguistic approaches to assess preferences are assumed that the linguistic term set is uniform and symmetrical distributed. However, there exist decision making problems whose assessments are unbalanced linguistic term sets. Therefore, a linguistic computational model based on symbolic models is proposed. Firstly, a weighted graph composed of basic labels is constructed. An unbalanced linguistic label is represented by some vertices in the graph. Then the Manhattan distance measure in the graph of any vertices is defined, with is applied to compute the distance between the two unbalanced linguistic terms. Finally, a decision model is designed to solve the TOPSIS problem. A numerical example is given to illustrate that the proposed method can visualise the unbalanced linguistic term set, and the Manhattan distance is better than the Euclidean distance when solving TOPSIS problems.

Keywords: unbalanced linguistic term set; technique for order preference by similarity to ideal solution(TOPSIS); computational model; distance measure

0 引言

人们在决策过程中提供自己的知识或偏好时会带有一些不确定性. 模糊逻辑和模糊语言为处理此类问题提供了方法^[1-2]. 自 Zadeh^[1] 于 1975 年首次提出语言变量的概念之后, 相关研究随之展开. 在经典的语言计算模型中, 用模糊集合描述语言变量的语义, 用语言变量模拟不精确的概念, 进行模糊推理, 并给出各类集结算子^[3-4]. 大部分的决策问题中出现的评价价值采用的语言信息都是均衡、对称分布的. 然而, 这样的评价价值形式单一, 已不能满足理论的发展和

实用的需要. 非均衡语言^[5] 经常出现在各类决策问题中, 例如 Herrera-Viedma 等^[6] 提出了信息检索系统中的非均衡语言决策问题, Meng 等^[7] 提出了医疗服务质量评价中的非均衡语言决策问题. 非均衡语言的决策方法已成为决策领域的新问题, 然而此类问题的研究并不常见.

语言层次结构(LH)是目前最为常用的解决非均衡语言决策问题的方法^[5, 8-9]. LH 由不同粒度的均衡语言集合组成, 不同粒度的语言集合之间的相互转换是这类方法不可避免的难题. 文献[10-11]提出的转

收稿日期: 2016-02-17; 修回日期: 2016-05-21.

基金项目: 国家自然科学基金项目(71401078, 71503134); 江苏高校哲学社会科学研究基金项目(2014SJB059); 江苏省社会科学基金项目(14EYD006).

作者简介: 蔡玫(1980—), 女, 副教授, 博士, 从事模糊决策分析等研究; 巩在武(1975—), 男, 教授, 博士生导师, 从事模糊决策分析等研究.

[†]通讯作者. E-mail: sanmoon_1980@163.com

化方法最初用于解决多粒度均衡语言间的融合和转化问题,随后在均衡语言的计算基础上各类非均衡语言的计算方法被提出,这些方法^[2, 7, 12]都是利用转换公式将非均衡语言转化成LH的不同层次上的均衡语言信息.虽然在转化过程中,非均衡语言的不同精确度通过均衡语言集合的不同粒度反映出来,但是在信息融合与集结时,往往只利用了均衡语言的位置指数而丢失了精确度,得到的结果不够合理.可以看出,这类方法得到了广大学者的认同,但复杂的转化过程以及转化过程的合理性仍需深入研究.

近几年,一些学者从不同角度提出了非均衡语言信息的表示方法. Dong等^[12]提出了用区间符号比例描述非均衡语言的方法; Wang等^[13]提出了一种数值模拟函数,用来产生均衡或者非均衡语言信息; Dong等^[2]用偏好关系描述了决策中的非均衡语言信息.上述方法都避免了不同层次语言之间的复杂的转化过程,为非均衡语言偏好的决策问题提供了新的思路.但是,相较于成熟的语言层级结构方法,没有系统地分析非均衡语言信息的一些性质,不能提供普遍适用的计算模型和决策方法.同样,在保留非均衡语言精确度信息方面也存在不合理的地方.

本文在现有研究的基础上提出一种非均衡语言计算模型,试图在两方面做出改进:一是在语言信息的表示方面避免复杂的转化过程;二是在语言信息的融合与集结方面尽量保留全面的信息,使决策的结果合理.所提出的方法采用图形描述非均衡语言,利用图形中点的位置反映出非均衡语言的排序大小和精确度;定义点与点之间的几何距离作为决策中偏好集结的基础,用以解决TOPSIS决策问题.

1 预备知识

Herrera等^[14]提出的二元语义模型是对符号化模型的扩展.一般的符号化模型在表示语言集中的一个标签时只用标签的位置指数描述,但是人们的语言信息往往不是一个标签就能表示出来的,因此会造成信息丢失. Herrera等^[14]用一对数值 (s_i, α) 描述语言信息, s_i 是语言集中的基础标签, α 是人们想要表达的语言信息与 s_i 之间的差距.

定义有序标签集. $S = \{s_0, s_1, \dots, s_g\}$ 是一组粒度为 $g+1$ 的有序的语言评价信息集.其中: $s_0 < s_1 < \dots < s_g$, s_i 为语言标签, $g+1$ 为标签集 S 的标签个数.每一个标签联系着一个 $[0, 1]$ 区间上的模糊数,可以用三角隶属函数或者梯形隶属函数表示.

二元语义是指针对某目标或对象、准则给出的评价语言信息由二元组 (s_k, α_k) 表示.其中元素 s_k 、

α_k 含义描述为: 1) s_k 为预先定义好的有序的语言评价信息集 S 中的第 k 个标签; 2) α_k 为符号转移值,且满足 $\alpha \in [-0.5, 0.5)$,表示评价结果与 s_k 的偏差.

采用二元语义表示语言评价信息并进行运算可有效避免语言评价信息集结和运算中出现的信息损失和扭曲,也使语言信息计算结果更为精确.语言信息与二元组之间的转化函数以及二元组与数值之间的转化函数定义如下.

定义1 假设 $S = \{s_0, s_1, \dots, s_g\}$ 是一个语言集合, $\beta \in [0, g]$ 为符号运算结果的数值.二元组与 β 之间的转化函数 Δ 以及反函数 Δ^{-1} 可以用下面的等式表示:

$$\Delta^{-1}: S \times [-0.5, 0.5) \rightarrow [0, g],$$

$$\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta. \quad (1)$$

$$\Delta: [0, g] \rightarrow S \times [-0.5, 0.5),$$

$$\Delta(\beta) = (s_i, \alpha), \begin{cases} s_i, i = \text{round}(\beta), \\ \alpha = \beta - i, \alpha \in [-0.5, 0.5). \end{cases} \quad (2)$$

在决策过程中,大量问题是多粒度语言决策问题^[15],即偏好来自不同粒度的语言集合, LH被广泛用以解决这类问题. LH定义为^[5]: $LH = \bigcup_t l(t, n(t))$.其中: $l(t, n(t))$ 为层次 t 的语言集合, $n(t)$ 为集合中语言标签的数量(粒度).每个层次的语言集合为

$$S^{n(t)} = \{s_0^{n(t)}, s_1^{n(t)}, \dots, s_{n(t)-1}^{n(t)}\}.$$

文献[10-11]给出了基于二元语义的转化公式,将层次 t 上的语言标签转换成 t' 上的语言标签.在决策中,利用转化公式将不同粒度的语言转化成统一形式 $\{(s_0, \alpha_0^i), t(s_1, \alpha_1^i), \dots, (s_T, \alpha_T^i)\}$,然后进行计算.

2 非均衡语言表示模型及距离测度

2.1 非均衡语言的图像化描述

给出非平衡语言集合的定义.

定义2 非平衡语言集合 $S^U = \{s_0^u, s_1^u, \dots, s_n^u\}$ 是一个粒度为 $n+1$ 的有序标签集.每一个标签联系着一个 $[0, 1]$ 区间上的模糊数.

S^U 满足以下条件:

1) S^U 中的标签是非均匀、不对称分布的.

2) f 取值在 $[0, 1]$ 上,是规格化后的偏好的取值,数值越大表示偏好越强.

3) s_i^u 的隶属度函数 $\mu_{s_i^u}(f) \in [0, 1]$ 表示 s_i^u 描述 f 的适合程度,并且 $\max(\mu_{s_i^u}(f)) = 1$.

4) $f_i \in [0, 1]$ 是 s_i^u 的中心取值, $\mu_{s_i^u}(f_i) = 1$,即 f_i 最适合描述 s_i^u 的取值. f_{i-1} 和 f_{i+1} 是 s_i^u 的左、右极限,也是 s_{i-1}^u 和 s_{i+1}^u 的中心取值.越靠近 f_i 的取值越

适合描述 $s_i^u, \mu_{s_i^u}(f)$ 越大.

5) $s_i^u = s_{iL} \cup s_{iC} \cup s_{iR}$. 其中: s_{iL} 为 s_i^u 的左半部分, s_{iC} 为 s_i^u 的中心, s_{iR} 为 s_i^u 的右半部分.

6) $\text{interval}(s_i^u)$ 反映出每个标签的不确定程度, 公式为 $\min\{(f_i - f_{i-1}), (f_{i+1} - f_i)\}$.

定义2关于非均衡语言集合的假设与文献[5]不同之处在于: Herrera等^[5]认为非均衡语言集合满足 $S = S_L \cup S_C \cup S_R$, 即分成左子集、中间子集和右子集. 左子集和右子集都有自己的中心语言标签, 并且每个子集满足关于中心语言标签的均匀、对称分布. 可以看出, 这类非均衡语言集是定义2的一种特殊情况. 定义2的假设放宽了要求, 仅仅需要 s_{i-1}^u 和 s_{i+1}^u 的中心值是 s_i^u 的左极限和右极限. 因此, 定义2具有更广泛的应用空间.

下面给出非均衡语言集合的图像化描述方法, 并给出图 $G_{S_n^*}$ 的定义.

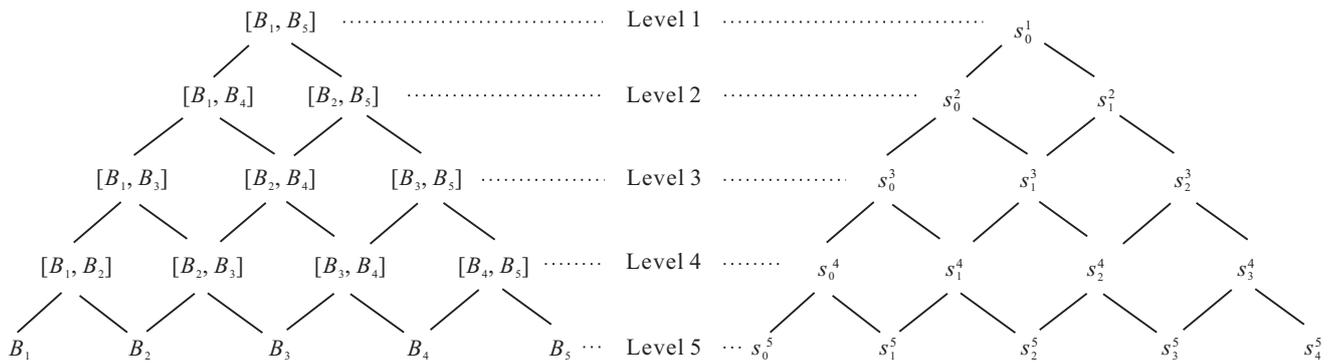


图1 $G_{S_n^*}$ 与LH的对比

图1中,任何一点都可以表示为

$$x_{i,j} = x_{ijL} \cup x_{ijC} \cup x_{ijR}.$$

其中: x_{ijL} 为点 $x_{i,j}$ 的左半部分, x_{ijC} 为点 $x_{i,j}$ 的中心, x_{ijR} 为点 $x_{i,j}$ 的右半部分.

图 $G_{S_n^*}$ 可以被看作是一种特殊的层次结构. 假设有一个与图 $G_{S_n^*}$ 对应的层次结构 $LH = l(1, 1) \cup l(2, 2) \cup l(3, 3) \cup l(4, 4) \cup l(5, 5)$, 可以为LH中的任一个标签 $s_k^{n(t)}$ 找到一个 $G_{S_n^*}$ 中对应的点 $x_{i,j} = [B_i, B_j]$ (见图1), 转换函数 τ_1 定义如下:

$$\tau_1(s_k^{n(t)}) \rightarrow x_{i,j}, n(t) = t,$$

$$i + n - n(t) = j, i + j = n - n(t) + (1 + k) \times 2.$$

非均衡语言集合里的语言标签有两种情形: 1) s_i^u 关于中心值 f_i 对称; 2) s_i^u 关于中心值 f_i 不对称. 如果 s_i^u 关于中心值 f_i 对称, 则可以用转换函数 τ_2 将 s_i^u 转换成来自语言层次结构的一个标签 $s_k^{n(t)}$, 即

$$\tau_2(s_i^u) \rightarrow s_k^{n(t)}, n(t) = 1/\text{interval}(s_i^u) + 1,$$

定义3 假设一个图 $G_{S_n^*}$ 的基础层是 S_n^* . S_n^* 包含元素 B_1, B_2, \dots, B_n , 每个元素是图的最底层的一个点. 图的点集合为

$$V(G_{S_n^*}) = \{x_{i,j} = [B_i, B_j] | B_i, B_j \in S_n^*, i \leq j\},$$

图的边集合为

$$E(G_{S_n^*}) = \{x_{i,j} \sim x_{r,s} | (r = i, s = j + 1) \text{ or } (s = j, r = i + 1)\}.$$

例如, 一个基于 $S_n^* = \{B_1, B_2, \dots, B_5\}$ 的图 $G_{S_n^*}$ (见图1左半边的描述), S_n^* 中的标签 B_i 表示一个语言标签 s_{i-1}^n, s_{i-1}^n 属于基础语言标签集 $S^n = \{s_0^n, s_1^n, \dots, s_{n-1}^n\}$, S^n 满足以下要求: 1) f 取值在 $[0, 1]$ 上, 是规格化后的偏好的取值, 数值越大表示偏好越强; 2) 隶属度函数 $\mu_{s_i^n}(f) \in [0, 1]$ 表示用 s_i^n 描述取值 f 的适合程度, $\max(\mu_{s_i^n}(f)) = 1, s_i^n$ 关于中心取值左右对称.

$$k = f_{ic} \times (n(t) - 1).$$

如果 s_i^u 关于中心值 f_i 不对称, 则需要分别确定它的左、右部分对应的标签.

下面给出算法1以解决转换问题.

算法1

输入 $s_i^u, G_{S_n^*}, S^n = \{s_0^n, s_1^n, \dots, s_{n-1}^n\}$

开始

如果 s_i^u 关于中心值对称

那么

用 $s_k^{n(t)}$ 表示 s_i^u : $\tau_2(s_i^u) \rightarrow s_k^{n(t)}$

用 $x_{j,p}$ 代替 $s_k^{n(t)}$: $\tau_1(s_k^{n(t)}) \rightarrow x_{j,p}$

s_i^u 表示为 $x_{j,p}$

否则 $s_i^u = s_{iL} \cup s_{iC} \cup s_{iR}$

找出中心 $s_{iC} = B_k$, 其中 $k = f_i \times (n - 1) + 1$

确定 $s_k^{n(t)}$, 其中 $s_k^{n(t)}$ 的左半部分与 s_{iL} 相同

用 x_{j_i, p_i} 代替 $s_k^{n(t)}$: $\tau_1(s_k^{n(t)}) \rightarrow x_{j_i, p_i}$

确定 $s_k^{n(t)}$, 其中 $s_k^{n(t)}$ 的右半部分与 s_{iR} 相同
 用 x_{j_r, p_r} 代替 $s_k^{n(t)}$: $\tau_1(s_k^{n(t)}) \rightarrow x_{j_r, p_r}$
 s_i^u 表示为 $s_i^u = x_{j_l, p_l} \cup B_k \cup x_{j_r, p_r}$
 结束

例1 以Herrera等^[5]的成绩评价系统为例, 给出非均衡语言的转化过程.

1) 构建与 S_n^* 相关的图 $G_{S_n^*}$, 即

$$S_n^* = \{B_1, B_2, \dots, B_9\}, S^9 = \{s_0^9, s_1^9, \dots, s_8^9\}.$$

2) 用图 $G_{S_n^*}$ 上点的集合描述 $S = \{F, D, C, B, A\}$, 描述过程见图2, 且有 $F \leftarrow \{x_{1,9}\}, D \leftarrow \{x_{1,9} \cup B_5 \cup x_{3,7}\}, C \leftarrow \{x_{5,9} \cup B_7 \cup B_7\}, B \leftarrow \{B_8\}, A \leftarrow \{B_9\}$.

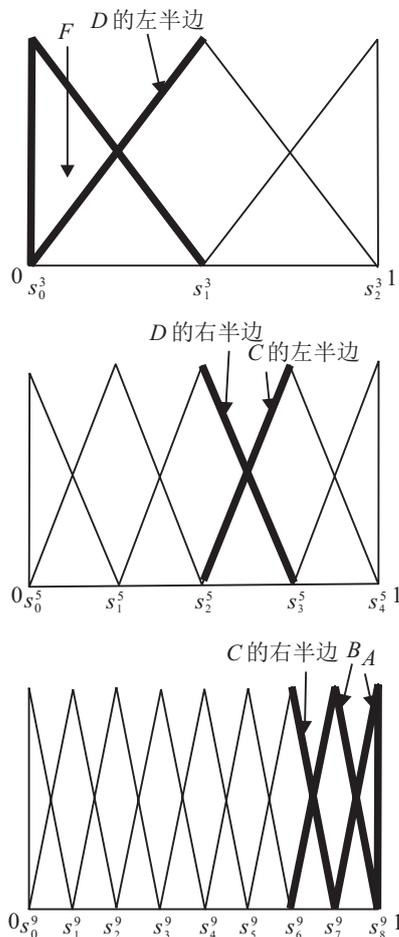


图2 成绩评价系统的转换过程演示

2.2 图中两点间的距离测度

Roselló等^[16]提出了一种用曼哈顿距离测量两个语言之间距离的方法. 假设两个语言变量用图 G_L 中的两个点 $x_{h,k} = [l_h, l_k], x_{h',k'} = [l_{h'}, l_{k'}]$ 表示, 两点间的距离为

$$d([l_h, l_k], [l_{h'}, l_{k'}]) = |h - h'| + |k - k'|. \quad (3)$$

以往的距离公式多为欧氏距离, 即在 m 维空间中两点之间的真实距离. 曼哈顿距离则是两个点在标

准坐标系上的绝对轴距总和.

本文提出的距离测度以文献[16]为基础, 下面给出图 $G_{S_n^*}$ 中任意两点间的距离计算方法. 任意两点的距离计算分两种情况: 第1种是直接采用由一个边相连的两个相邻节点; 第2种是采用不相邻的两点.

针对第1种情况给出图 $G_{S_n^*}$ 中边的距离公式, 即权重函数定义.

定义4 边 $E(G_{S_n^*}) = \{x_{i,j} \sim x_{r,s} | (r = i, s = j+1) \text{ or } (s = j, r = i+1)\}$ 的权重函数 $w(G_{S_n^*}) \rightarrow R^+$ 定义如下:

$$w(x_{i,j} \sim x_{r,s}) = 1 + k \cdot |I(x_{i,j}) - I(x_{r,s})|. \quad (4)$$

其中: $k \in [0, +\infty)$ 为精确度指数, I 为精确度函数.

图 $G_{S_n^*}$ 中的一个点即一个信息粒, 用两个属性描述一个信息粒包含的信息, 即覆盖指数 cov 和专属指数 spe ^[17]. 覆盖指数描述信息粒包含信息量的大小, 专属指数描述信息粒的排序大小. 定义4的距离公式从这两个方面反映出两个信息粒的差别. 由于式(4)中两点是相邻节点, 排序上相差1. 式(4)保证图中边的距离大于等于1, 同时精确度相差越大, 两点的距离越大.

I 是一个非负连续函数, 式(4)通过 I 反映出信息粒的精确度, 定义为

$$I(x_{i,j}) = \ln \frac{1}{\text{cov}(x_{i,j})}, \quad (5)$$

$$\text{cov}(x_{i,j}) = \frac{\text{card}(B_i, B_{i+1}, \dots, B_j)}{n}, \quad (6)$$

其中 $\text{card}(B_i, B_{i+1}, \dots, B_j)$ 为 $(B_i, B_{i+1}, \dots, B_j)$ 的粒度个数.

若两个点 $x_{i,j}$ 和 $x_{h,k}$ 满足 $\text{cov}(x_{i,j}) \leq \text{cov}(x_{h,k})$, 则 $I(x_{i,j}) \geq I(x_{h,k})$, 即 $x_{i,j}$ 的精确度大于 $x_{h,k}$ 的精确度.

k 调节精确度在计算距离时的重要程度, k 越大, $|I(x_{i,j}) - I(x_{r,s})|$ 在计算边权重时的分量越大. 当 $k = 0$ 时, 表示不考虑两个信息粒在精确度上的差距, 即图 $G_{S_n^*}$ 中任何一条边的权重都是1. 文献[16, 18]定义的模糊犹豫语言之间的距离公式(3)是(4)在 $k = 0$ 时的一种特殊情况. 例如, 图1中的 $G_{S_n^*}$ 的边 $x_{1,5} \sim x_{1,4}$ 的距离为

$$1 + k \left| \ln \frac{5}{5} - \ln \frac{5}{4} \right| = 1 + 1.0k,$$

边 $x_{1,4} \sim x_{1,3}$ 的距离为

$$1 + k \left| \ln \frac{5}{4} - \ln \frac{5}{3} \right| = 1 + 1.2k.$$

决策者可以通过 k 调节图的边距离.

下面讨论不相连的两点之间的距离计算问

题. 假设有两个点 v_0 和 v_k , 有几条路都可以从 v_0 到 v_k , 其中最短的一条路就是 v_0 与 v_k 之间的距离. 从 v_0 到 v_k 的路径为

$$\phi(p) = w(v_0 \sim v_1) + w(v_1 \sim v_2) + \dots + w(v_{k-1} \sim v_k).$$

其中: $v_i \sim v_{i+1}$ 为相邻的两个点, $w(v_i \sim v_{i+1})$ 为边 $E(v_i \sim v_{i+1})$ 的权重.

点 v_0 与 v_k 之间的距离为

$$d(v_0 \sim v_k) = \min\{\phi(p) : v_0 \rightsquigarrow v_k\},$$

其中 $p(v_0 \rightsquigarrow v_k)$ 为一条可达的路径. 本文利用改进的 Dijkstra's 算法^[9] 求解最短路径. 假设有一个点 $s \in V$, 点 s 到其他不包含该点集合 \bar{S} 的距离通过以下公式得到:

$$d(s \sim \bar{S}) = \min_{x_{i,j} \in S, x_{r,s} \in \bar{S}} \{d(s, x_{i,j}) + w(x_{i,j} \sim x_{r,s})\}.$$

下面求从 $x_{i,j}$ 到 $x_{r,s}$ 的最短路径.

算法 2

Step 1: 设 $l(s) = 0, l(x_{r,s}) = \infty; x_{r,s} \neq s; S :=$

$\{s\}, \bar{S} := V - \{s\}; u_0 = s, i = 0.$

Step 2: 更新 $l(x_{r,s}), S$ 和 $\bar{S}.$

1) 针对每一个 $x_{r,s} \in \bar{S} \cap \text{neighbor}(u_i)$, 分别采用 $\min\{l(x_{r,s}), d(u_i, x_{i,j}) + w(u_i \sim x_{r,s})\}$ 替代 $l(x_{r,s})$;

2) 选择一个合适的 $u_{i+1} \in \text{neighbor}(u_i)$ 来计算

$$\min_{x_{r,s} \in \bar{S}} l(x_{r,s});$$

3) 设 $S := S \cup \{u_{i+1}\}, \bar{S} := \bar{S} - \{u_{i+1}\}.$

Step 3: $i = i + 1.$

Step 4: 重复 Step 2 和 Step 3, 直到 $\bar{S} = \emptyset$, 停止.

最短路径即为 $x_{i,j} \rightarrow u_0 \rightarrow u_1 \dots u_{n-1} \rightarrow x_{r,s}.$

推论 1 如果图 $G_{S_n^*}$ 所有边的权重都为 1, 则 $x_{i,j}$ 与 $x_{r,s}$ 两点间距离为

$$d([l_i, l_j], [l_r, l_s]) = |i - r| + |j - s|.$$

证明 从 $x_{i,j}$ 到 $x_{r,s}$ 的最短路径可以分成两部分: 第 1 部分从 $x_{i,j}$ 到 $x_{a,j}$; 第 2 部分从 $x_{a,j}$ 到 $x_{r,s}$. $x_{a,j}$ 和 $x_{r,s}$ 位于同一层, $x_{a,j}$ 是介于 $x_{i,j}$ 与 $x_{r,s}$ 之间的纵向的一个点.

假设 $r \geq i, s \geq j$, 如图 3 所示.

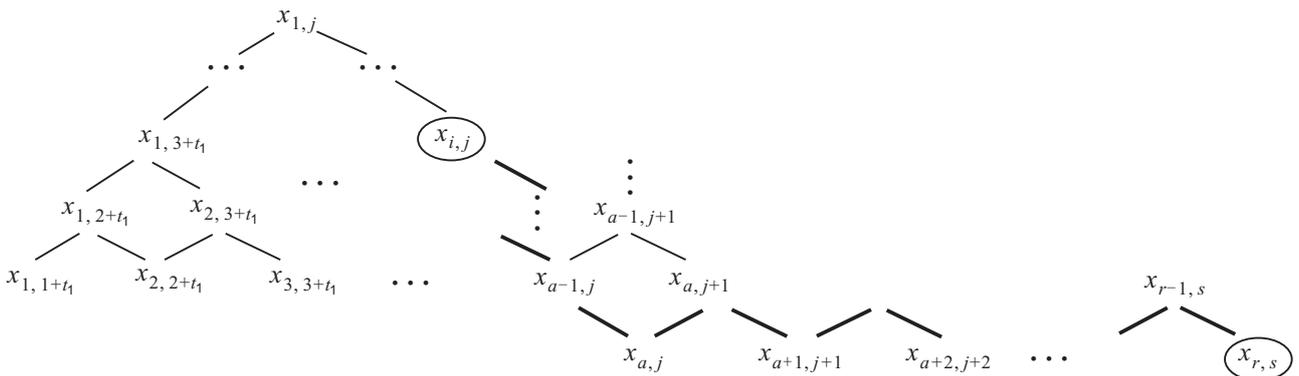


图 3 点 $x_{i,j}$ 和 $x_{r,s}$ 的位置

图 3 中加粗的线就是从 $x_{i,j}$ 到 $x_{r,s}$ 的最短路径, 分两步计算最短路径上的距离.

从 $x_{i,j}$ 到 $x_{a,j}$ 的最短距离, 计算得到

$$d([B_i, B_j], [B_a, B_j]) = |a - i| = a - i;$$

从 $x_{a,j}$ 到 $x_{r,s}$ 的最短距离, 计算得到

$$d([B_a, B_j], [B_r, B_s]) = (s - j) + (r - a);$$

从 $x_{i,j}$ 到 $x_{r,s}$ 的最短距离, 计算得到

$$\begin{aligned} d([B_i, B_j], [B_r, B_s]) &= \\ (a - i) + (s - j) + (r - a) &= \\ (r - i) + (s - j). \end{aligned}$$

其他的情况可以得到相同的结果, 即最后归纳为

$$d([B_i, B_j], [B_r, B_s]) = |i - r| + |j - s|. \quad \square$$

2.3 非均衡语言的距离测度

在上节分析的基础上, 得到非均衡语言标签的距离测度定义如下.

定义 5 假设 $s_i^u, s_l^u \in S^U, s_i^u$ 可以表示为点的集合 $F = x_{j_l, p_l} \cup B_k \cup x_{j_r, p_r}, s_l^u$ 可以表示为点的集合 $\epsilon = x_{j'_l, p'_l} \cup B_{k'} \cup x_{j'_r, p'_r}$, 则 s_i^u 与 s_l^u 之间的距离可以表示为

$$\begin{aligned} d(F, \epsilon) &= \\ \frac{1}{6}d([B_{j_l}, B_{p_l}], [B_{j'_l}, B_{p'_l}]) &+ \frac{4}{6}d(B_k, B_{k'}) + \\ \frac{1}{6}d([B_{j_r}, B_{p_r}], [B_{p'_r}, B_{j'_r}]) &= \\ \frac{1}{6}(|j_l - j'_l| + |p_l - p'_l|) &+ \\ \frac{4}{6}|k - k'| + \frac{1}{6}(|j_r - j'_r| &+ |p_r - p'_r|). \end{aligned}$$

通过算法1将非均衡语言图像化描述为3个点,即代表左半边信息的点 $x_{j_l, p_l}(x_{j_l', p_l'})$ 、代表中心信息的点 $B_k(B_{k'})$ 和代表右半边信息的点 $x_{j_r, p_r}(x_{j_r', p_r'})$.两个非均衡语言之间的距离由这3个点之间的距离加权平均得到.这3部分在计算距离测度时的权重不同:中心点反映的是非均衡语言的专属指数,表示信息粒的排序大小;左右点反映的是覆盖系数,表示信息粒的精确度.已有的多数方法仅仅考虑两个非均衡语言的排序差距,即 $d(B_k, B_{k'})$ 的权重是1.定义5也认为排序差距较重要,为 $d(B_k, B_{k'})$ 分配了 $\frac{4}{6}$ 的权重,为覆盖指数保留了 $\frac{2}{6}$ 的权重(左右点距离各分配了 $\frac{1}{6}$ 的权重),这样的权重分配既与多数距离测度思想一致,又反映出不同精确度的非均衡语言的差异.

2.4 曼哈顿距离与欧氏距离在非均衡语言的距离测度计算上的比较

语言标签的距离测度计算分为两种情况:1)相同语义的语言标签的距离测度计算;2)不同语义的语言标签的距离测度计算.

已有的处理多粒度语言或非均衡语言的距离测度的方法都是采用欧氏距离,忽略了语言标签作为一个信息粒具有专属性和精确性的二维属性.已有的方法只考虑了每个语言标签的专属性,即语言标签在排序大小方面的信息,将两个语言标签之间的距离简单化为一维距离,即排序大小的距离.

相同语义欧氏距离定义如下.

定义6^[20] 假设有有序标签集 $S = \{s_0, s_1, \dots, s_g\}$,二元组 (s_i, α_i) 与 (s_j, α_j) 之间的距离为

$$d((s_i, \alpha_i), (s_j, \alpha_j)) = \Delta|\Delta^{-1}(s_i, \alpha_i) - \Delta^{-1}(s_j, \alpha_j)|. \quad (7)$$

不相同语义的语言标签(即语言标签来自不同粒度语言集合),通常采用统一化方法将其统一为相同语义的语言标签的组合,形如 $\{(s_i, \alpha_i), (s_{i+1}, \alpha_{i+1}), \dots, (s_k, \alpha_k)\}$,然后计算两个不同语义语言标签之间的距离.

定义7^[21] 假设有有序标签集 $S = \{s_0, s_1, \dots, s_g\}$ 是基础语言标签集合,两个语言标签 δ 和 ϵ 转化成 $\{(s_i, \alpha_i), (s_{i+1}, \alpha_{i+1}), \dots, (s_k, \alpha_k)\}, \{(s_j, \alpha_j), (s_{j+1}, \alpha_{j+1}), \dots, (s_p, \alpha_p)\}$. δ 与 ϵ 之间的欧氏距离为

$$cv(\delta) = \frac{\sum_{l=i}^k \Delta^{-1}(s_l) \cdot \alpha_l}{\sum_{l=i}^k \alpha_l}, \quad (8)$$

$$cv(\epsilon) = \frac{\sum_{l=j}^p \Delta^{-1}(s_l) \cdot \alpha_l}{\sum_{l=j}^p \alpha_l}, \quad (9)$$

$$d(\delta, \epsilon) = |cv(\delta) - cv(\epsilon)|. \quad (10)$$

已有的欧氏距离的方法虽然在计算上大大简便了,但是忽略了语言标签在精确度上的差异.一个语言标签的信息应该是二维的,即专属性维度和精确性维度.相同语义的语言标签在精确性程度上相同,所以计算距离测度时暂且忽略精确性维度,对决策的影响不大.但是,多粒度语言或非均衡语言在精确程度上差别很大,是不可或缺的信息,因此计算距离测度时忽略精确性维度是不妥的.

本文提出的可视化图像表示非均衡语言的方法,实质保留了任何语言标签的二维信息,横轴方向是专属维度,纵轴方向是模糊维度,两点之间的曼哈顿距离标明两个点在坐标系上的绝对轴距总和.因此,保留语言标签作为信息粒的全部信息,是对已有距离测度的一个重要改进.

3 基于距离测度的TOPSIS法

3.1 决策步骤

采用理想解法求解多指标决策的一个关键问题是求解最好方案(最差方案)指标值与各候选方案的值之间的距离.通常TOPSIS法中的距离是指欧氏距离.欧氏距离在测量非平衡语言标签之间距离时有不足之处,本文提出基于曼哈顿距离的TOPSIS方法,为解决非平衡语言的多准则决策问题给出新的思路.

假设对于一个多准则决策问题,候选方案集为 $A = \{A_1, A_2, \dots, A_m\}$,每个方案的准则集为 $C = \{C_1, C_2, \dots, C_n\}$.设决策矩阵为 $X = (a_{ij})_{m \times n}$,其中 a_{ij} 表示方案 A_i 的准则 C_j 的评价值. $W = \{w_1, w_2, \dots, w_n\}$ 是准则集对应的权重集,满足 $\sum_{i=1}^n w_i = 1$.具体步骤如下.

Step 1: 将非平衡语言标签 a_{ij} 通过图形中的点描述.

Step 2: 确定如下正理想解和负理想解:

$$A^+ = \{v_1^+, v_2^+, \dots, v_n^+\},$$

$$A^- = \{v_1^-, v_2^-, \dots, v_n^-\}.$$

Step 3: 计算如下距离矩阵:

$$D^+ = [d(a_{ij}, v_i^+)]_{m \times n} = [d_{ij}^+]_{m \times n},$$

$$D^- = [d(a_{ij}, v_i^-)]_{m \times n} = [d_{ij}^-]_{m \times n}.$$

Step 4: 计算方案 A_i 与最优解和最差解的距离 d_i^+ 和 d_i^- , 即

$$d_i^+ = \sum_{j=1}^n d_{ij}^+ \times w_j,$$

$$d_i^- = \sum_{j=1}^n d_{ij}^- \times w_j.$$

Step 5: 计算方案 A_i 与理想解和负理想解的相对贴近度

$$RC_i = \frac{d_i^-}{d_i^- + d_i^+}, \quad i = 1, 2, \dots, m.$$

Step 6: 比较相对贴近度的大小, RC_i 越大, 方案 A_i 越优.

3.2 算 例

本节采用文献[5]的一个算例验证所提出的方法. 假设有两个学生 John Smith 和 Martina Grant 完成 6 门功课的考试, 每门功课在综合评价中的地位一样. 现在需要给出这两名学生的综合评价成绩. 已知学生的各门功课的成绩如表 1 所示, 成绩评价系统采用的评分来自非均衡语言集合 $S^U = \{F, D, C, B, A\}$.

表 1 两名学生的成绩评价

姓名	T_1	T_2	T_3	T_4	T_5	T_6
Smith J(A_1)	D	C	B	C	C	C
Grant M(A_2)	A	D	D	C	B	A

1) 将非均衡语言集合 $S = \{F, D, C, B, A\}$ 用图形中的点描述, 转化结果见图 1.

2) 确定如下正理想解与负理想解:

$$A^+ = \{A, A, A, A, A, A\},$$

$$A^- = \{F, F, F, F, F, F\}.$$

3) 计算距离矩阵.

利用定义 4 中的式 (4) 计算边的权重, 假设 $k = 0$, 可以得到所有边权重都为 1. 图 $G_{S_n^*}$ 中两点的距离公式为

$$d(x_{h,k}, x_{h',k'}) = |h - h'| + |k - k'|,$$

$$D^+ = [d(a_{ij}, A)]_{m \times n} = \begin{bmatrix} 8 & 4 & 2 & 4 & 4 & 4 \\ 0 & 8 & 8 & 5 & 2 & 0 \end{bmatrix},$$

$$D^- = [d(a_{ij}, F)]_{m \times n} = \begin{bmatrix} \frac{10}{3} & \frac{22}{3} & 8 & \frac{22}{3} & \frac{22}{3} & \frac{22}{3} \\ 8 & \frac{10}{2} & \frac{10}{3} & \frac{22}{3} & 8 & 8 \end{bmatrix}.$$

4) 分别计算方案 A_i 与最优解和最差解的距离 d_i^+ 和 d_i^- , 即

$$d_1^+ = \sum_{j=1}^n d(a_{ij}, v_i^+) \times w_j = \frac{13}{3},$$

$$d_2^+ = \sum_{j=1}^n d(a_{ij}, v_i^+) \times w_j = \frac{11}{3},$$

$$d_1^- = \sum_{j=1}^n d(a_{ij}, v_i^-) \times w_j = \frac{56}{9},$$

$$d_2^- = \sum_{j=1}^n d(a_{ij}, v_i^-) \times w_j = \frac{57}{9}.$$

5) 分别计算方案 A_i 与正理想解和负理想解的相对贴近度

$$RC_1 = \frac{d_1^-}{d_1^- + d_1^+} = 0.59,$$

$$RC_2 = \frac{d_2^-}{d_2^- + d_2^+} = 0.63.$$

6) 比较相对贴近度的大小.

$RC_1 < RC_2$, 所以 Martina Grant 综合评价优于 John Smith.

Herrera 等^[5] 利用转换函数将成绩评分转化成以 $S^9 = \{s_0^9, s_1^9, \dots, s_8^9\}$ 为基础标签的统一形式. 在计算总评成绩时, 通过式 (1) 用 β 反映不均衡语言的数值, 保留的只是专属性方面的信息, 丧失了精确度方面的信息. 本文利用曼哈顿距离计算总评成绩, 得到与文献[5]一致的结论. 而 2.4 节的分析说明了本方法更有说服力.

4 结 论

已有的很多语言计算模型都是针对均衡、对称分布的语言集合, 而针对非均衡语言集合的技术研究还较为不足. 本模型从图形的角度研究非均衡化语言信息表示方法, 建立的图形在保留已有的语言层次结构所有信息的基础上增加了点与点之间的联系. 利用曼哈顿距离测量非均衡语言标签之间的距离, 解决了距离测度计算中的信息缺失和理论依据不足的缺陷. 距离测度正是 TOPSIS 决策中评价值与最优值 (最差值) 之间距离计算的关键, 应用于这类决策问题有独特的优势.

图形 $G_{S_n^*}$ 的基础是均衡的语言集合 S^n , 语言层次结构也是建立在称之为基础语言的均衡语言集合之上的, 这一点一直是难以攻克的困难. 今后的研究将试图去掉这个限制, 提出具有更多普适性的非均衡语言计算模型.

参考文献(References)

- [1] Zadeh L A. The concept of a linguistic variable and its application to approximate reasoning-I[J]. *Information Sciences*, 1975, 8(3): 199-249.
- [2] Dong Y C, Li C C, Herrera F. An optimization-based approach to adjusting unbalanced linguistic preference relations to obtain a required consistency level[J]. *Information Sciences*, 2015, 292(1): 27-38.
- [3] Xu Y, Wang H. Approaches based on 2-tuple linguistic power aggregation operators for multiple attribute group decision making under linguistic environment[J]. *Applied Soft Computing*, 2011, 11(5): 3988-3997.
- [4] Xu Z S. Induced uncertain linguistic OWA operators applied to group decision making[J]. *Information Fusion*, 2006, 7(2): 231-238.
- [5] Herrera F, Herrera-Viedma E, Martínez L. A fuzzy linguistic methodology to deal with unbalanced linguistic term sets[J]. *IEEE Trans on Fuzzy Systems*, 2008, 16(2): 354-370.
- [6] Herrera-Viedma E, López Herrera A G. A model of an information retrieval system with unbalanced fuzzy linguistic information[J]. *Int J of Intelligent Systems*, 2007, 22(11): 1197-1214.
- [7] Meng D, Pei Z. On weighted unbalanced linguistic aggregation operators in group decision making[J]. *Information Sciences*, 2013, 223(2): 31-41.
- [8] Herrera-Viedma E, Cabrerizo F J, Pérez I J, et al. Applying linguistic OWA operators in consensus models under unbalanced linguistic information[C]. *Studies in Fuzziness and Soft Computing*. Heidelberg: Springer Berlin, 2011: 167-186.
- [9] Cabrerizo F J, Pérez I J, Herrera-Viedma E. Managing the consensus in group decision making in an unbalanced fuzzy linguistic context with incomplete information[J]. *Knowledge-Based Systems*, 2010, 23(2): 169-181.
- [10] Chen Z, Ben-Arieh D. On the fusion of multi-granularity linguistic label sets in group decision making[J]. *Computers and Industrial Engineering*, 2006, 51(3): 526-541.
- [11] Herrera F, Herrera-Viedma E, Martínez L. A fusion approach for managing multi-granularity linguistic term sets in decision making[J]. *Fuzzy Sets and Systems*, 2000, 114(1): 43-58.
- [12] Dong Y C, Wu Y Z, Zhang H J, et al. Multi-granular unbalanced linguistic distribution assessments with interval symbolic proportions[J]. *Knowledge-Based Systems*, 2015, 82(7): 139-151.
- [13] Wang B L, Liang J Y, Qian Y H. A normalized numerical scaling method for the unbalanced multi-granular linguistic sets[J]. *Int J of Uncertainty Fuzziness and Knowledge-Based Systems*, 2015, 23(2): 221-243.
- [14] Herrera F, Martínez L. A 2-tuple fuzzy linguistic representation model for computing with words[J]. *IEEE Trans on Fuzzy Systems*, 2000, 8(6): 746-752.
- [15] 姜艳萍, 樊治平. 基于不同粒度语言判断矩阵的群决策方法[J]. *系统工程学报*, 2006, 21(3): 249-253. (Jiang Y P, Fan Z P. Approach to group decision making with multi-granularity linguistic comparison matrices[J]. *J of Systems Engineering*, 2006, 21(3): 249-253.)
- [16] Roselló L, Sánchez M, Agell N, et al. Using consensus and distances between generalized multi-attribute linguistic assessments for group decision-making[J]. *Information Fusion*, 2014, 17(5): 83-92.
- [17] Pedrycz W, Song M. Granular fuzzy models: A study in knowledge management in fuzzy modeling[J]. *Int J of Approximate Reasoning*, 2012, 53(7): 1061-1079.
- [18] Falcó E, García-Lapresta J L, Roselló L. Allowing agents to be imprecise: A proposal using multiple linguistic terms[J]. *Information Sciences*, 2014, 258(2): 249-265.
- [19] Xu M H, Liu Y Q, Huang Q L, et al. An improved Dijkstra's shortest path algorithm for sparse network[J]. *Applied Mathematics and Computation*, 2007, 185(1): 247-254.
- [20] 廖貅武, 李垣, 董广茂. 一种处理语言评价信息的多属性群决策方法[J]. *系统工程理论与实践*, 2006, 26(9): 90-98. (Liao X W, Li Y, Dong G M. A multi-attribute group decision-making approach dealing with linguistic assessment information[J]. *Systems Engineering — Theory & Practice*, 2006, 26(9): 90-98.)
- [21] Herrera-Viedma E, Mata F, Martínez L, et al. Measurements of consensus in multi-granular linguistic group decision-making modeling decisions for artificial intelligence[M]. Heidelberg: Springer Berlin, 2004: 194-204.

(责任编辑: 闫妍)