

引用格式:张冉,王然,林燕鸿,康彤.直觉模糊集及其在网络舆情中的应用[J].中国传媒大学学报(自然科学版),2023,30(06):19-27.
文章编号:1673-4793(2023)06-0019-09

直觉模糊集及其在网络舆情中的应用

张冉,王然,林燕鸿,康彤*

(中国传媒大学数据科学与智能媒体学院,北京100024)

摘要:网络舆情是社会舆情在互联网空间的映射,是社会舆情的直接反映。由于实际问题的模糊性,模糊理论是一种有效的解决问题的工具。直觉模糊集在传统模糊集的基础上,增加了非隶属度因子,从而可以更充分、更准确地刻画不确定信息。本文从直觉模糊集的角度出发,概述了现有的理论及应用成果,包括直觉模糊集的基本概念、运算等,以及决策领域的信息测度、聚合算子、多属性决策方法。本文重点研究直觉模糊集在网络舆情决策领域的应用价值,探讨了当前的研究现状以及未来的发展方向。

关键词:直觉模糊集;舆情;模糊测度;聚合算子;多属性决策

中图分类号:O159;C934 **文献标识码:**A

Intuition fuzzy set and its application in network public opinion

ZHANG Ran, WANG Ran, LIN Yanhong, KANG Tong*

(School of Data Science and Intelligence Media, Communication University of China, Beijing 100024, China)

Abstract: Online public opinion is the mapping of social public opinion in the internet space and a direct reflection of social public opinion. Due to the ambiguity of practical problems, fuzzy theory is an effective tool for solving problems. Intuition fuzzy sets add non membership factors on the basis of traditional fuzzy sets, which can more fully and accurately characterize uncertain information. In this article an overview of existing theories and application achievements was provided from the perspective of intuition fuzzy sets, including the basic concepts and operations of intuition fuzzy sets, as well as information measures, aggregation operators, and multi-attribute decision-making methods in the decision-making field. The application value of intuition fuzzy sets in the field of online public opinion decision-making was focused on, and the current research status and future development directions were explored.

Keywords: intuition fuzzy set; public opinion; fuzzy measure; aggregation operator; multi-attribute decision-making

1 引言

由于网络传播技术不断发展、网民数量飞速增加,新型媒体赋予广大受众前所未有的传播权力,因此网络舆论的力量不容忽视。网络舆情的研究,主要包括网络舆情的基础理论、外部性特征、管理与引导

以及依托计算机和网络技术建构监测、分析、研判、预警系统等内容^[1]。舆情决策一般是以模型为基础,通过调查问卷、数据统计、专家打分等传统方式以及网络爬虫、文本分析等新型方式获取数据,然后经过一系列的信息处理,实现对舆情事件的排序、选择或者评价等目标。随着网络舆情研究的不断深入,经典数

基金项目:国家自然科学基金面上项目(42074108);中央高校基础研究基金项目(CUC23GG14)

作者简介(*为通讯作者):张冉(1999-),女,硕士研究生,主要从事模糊数学与决策应用研究。Email:rzhang9@cuc.edu.cn;康彤(1967-),男,教授,博士生导师,主要从事科学与工程与信息计算研究。Email:kangtong@cuc.edu.cn

学的方法越来越难以表述模糊的、不精确的信息,于是研究人员开始引入模糊理论来解决有关问题。如刘^[2]基于三角模糊数,构建了完整的网络舆情预警指标体系;李^[3]等人基于48个案例进行模糊集定性比较分析,研究网络舆情的热度生成机理;张^[4]等人基于直觉模糊推理的方法,研究网络舆情监测等级评估。Madhusudhanan^[5]研究新的情感分析算法,利用模糊逻辑提高了对公众在线评论的情绪分类准确性。Sánchez^[6]从舆论挖掘的角度,利用多层次模糊认知图谱评估舆论的质量,对于决策以及社会政治分析非常有价值。

本文总结了近年来直觉模糊集理论的一些研究成果,对相关的方法进行了比较分析,并概述了其应用于网络舆情决策中的研究现状,提出了进一步的建议。

2 直觉模糊集的概念与发展

2.1 基本概念

1965年,美国学者 Zadeh^[7]在传统明晰集的基础上,创立了模糊集(Fuzzy Set, FS),研究非精确的现象,同时提出了相关的数学性质,奠定了模糊数学的基础。

(定义2.1)将论域 X 上的模糊集 F 定义为

$$F = \{ \langle x, \mu_F(x) \rangle \mid x \in X \}$$

其中, $\mu_F(x) \in [0, 1], x \in X$ 叫做隶属度。与普通集合完全由特征函数唯一确定一样,模糊集完全由其隶属函数唯一确定,可以很好地处理模糊性和不确定性的问题。

但是传统模糊集忽略了问题的犹豫性。例如,在投票当中,人们可能投“支持”、“反对”或者“弃权”,后两者的信息无法被表达出来。1986年,Atanassov^[8]提出直觉模糊集(Intuition Fuzzy Set, IFS)的概念,增加了非隶属度,充分反应了决策中的肯定、否定和犹豫的特征,是对传统模糊集的著名拓展之一。

(定义2.2)将论域 X 上的直觉模糊集 A 定义为

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \}$$

其中, $\mu_A(x) \in [0, 1], \nu_A(x) \in [0, 1]$ 分别叫做隶属度和非隶属度,并且满足对 $\forall x \in X, 0 \leq \mu_A(x) + \nu_A(x) \leq 1, \pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$ 称为犹豫度。把 $(\mu_A(x), \nu_A(x))$ 称为直觉模糊数(Intuition Fuzzy Number, IFN)。特别地,当 $\pi_A(x) = 0$ 时,IFS就成为FS,所以传统模糊集可以看作直觉模糊集的特殊情况。直觉模糊集可以传

达更多的信息,比传统模糊集在处理模糊性和不确定性等方面更加灵活实用。

2.2 直觉模糊集的关系与运算

(定义2.3)设 X 是非空集合,直觉模糊集

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \},$$

$$A_1 = \{ \langle x, \mu_{A_1}(x), \nu_{A_1}(x) \rangle \mid x \in X \},$$

$$A_2 = \{ \langle x, \mu_{A_2}(x), \nu_{A_2}(x) \rangle \mid x \in X \},$$
 则^[9]

$$(1) \bar{A} = \{ \langle x, \nu_A(x), \mu_A(x) \rangle \mid x \in X \};$$

$$(2) A_1 \cap A_2 = \{ \langle x, \min \{ \mu_{A_1}(x), \mu_{A_2}(x) \}, \max \{ \nu_{A_1}(x), \nu_{A_2}(x) \} \rangle \mid x \in X \};$$

$$(3) A_1 \cup A_2 = \{ \langle x, \max \{ \mu_{A_1}(x), \mu_{A_2}(x) \}, \min \{ \nu_{A_1}(x), \nu_{A_2}(x) \} \rangle \mid x \in X \};$$

$$(4) A_1 + A_2 = \{ \langle x, \mu_{A_1}(x) + \mu_{A_2}(x) - \mu_{A_1}(x)\mu_{A_2}(x), \nu_{A_1}(x)\nu_{A_2}(x) \rangle \mid x \in X \};$$

$$(5) A_1 \cdot A_2 = \{ \langle x, \mu_{A_1}(x)\mu_{A_2}(x), \nu_{A_1}(x) + \nu_{A_2}(x) - \nu_{A_1}(x)\nu_{A_2}(x) \rangle \mid x \in X \};$$

$$(6) nA = \{ \langle x, 1 - (1 - \mu_A(x))^n, (\nu_A(x))^n \rangle \mid x \in X \};$$

$$(7) A^n = \{ \langle x, (\mu_A(x))^n, 1 - (1 - \nu_A(x))^n \rangle \mid x \in X \}.$$

可见,以上集合运算结果仍是直觉模糊集。类似地,有直觉模糊数的运算法则,与以上集合的运算规则一致,在此省略。

决策模型最后是对直觉模糊数的排序和比较大小的问题,需要借助评分函数和精度函数。

(定义2.4)设 $\alpha_i = (\mu_i, \nu_i), i = 1, 2$ 是两个直觉模糊数, $s(\alpha_i) = \mu_i - \nu_i$ 称为 α_i 的分数函数, $h(\alpha_i) = \mu_i + \nu_i$ 称为 α_i 的精度函数,那么^[10]

$$(1) \text{如果 } s(\alpha_1) < s(\alpha_2), \text{ 则 } \alpha_1 < \alpha_2;$$

$$(2) \text{如果 } s(\alpha_1) > s(\alpha_2), \text{ 则 } \alpha_1 > \alpha_2;$$

$$(3) \text{如果 } s(\alpha_1) = s(\alpha_2), \text{ 则}$$

$$\text{若 } h(\alpha_1) = h(\alpha_2), \text{ 则 } \alpha_1 = \alpha_2;$$

$$\text{若 } h(\alpha_1) < h(\alpha_2), \text{ 则 } \alpha_1 < \alpha_2;$$

$$\text{若 } h(\alpha_1) > h(\alpha_2), \text{ 则 } \alpha_1 > \alpha_2.$$

2.3 直觉模糊集的发展

经过几十年的发展,IFS已应用于解决现实中的许多决策相关的问题,涉及各行各业,具有广阔的应用前景和深刻的社会意义。同时,已有许多学者研究并拓展了直觉模糊集,研究主要有以下方面,而图1表达了这些集合之间的关系:

(1)处理用区间值表示的隶属度和非隶属度,区间

值直觉模糊集(Interval-Valued Intuitionistic Fuzzy Set, IVIFS);

(2)用集合表示的隶属度和非隶属度,集值直觉模糊集(Set-Valued Intuitionistic Fuzzy Set, SVIFS);

(3)将参数理论引入直觉模糊集,直觉模糊软集(Intuitionistic Fuzzy Soft Set, IFSS);

(4)通过与粗糙集融合来处理上下近似,直觉模糊粗糙集(Intuitionistic Fuzzy Rough Set, IFRS);

(5)语言形式的信息结合直觉模糊理论,语言直觉模糊集(Language Intuitionistic Fuzzy Set, LIFS);

(6)n型模糊集尤其是二型模糊集表征多重不确定性,n型直觉模糊集(Type-n Intuitionistic Fuzzy Set, T-n IFS)。

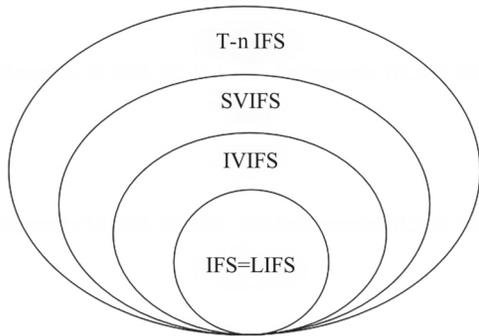


图1 不同类型的直觉模糊集之间的包含关系

3 直觉模糊多属性决策

3.1 问题描述

模糊决策是直觉模糊集最广泛也是最典型的应用。决策(Decision-Making, DM)可以看作是一个根据提供的决策信息从多个方案中进行选择、排序或分类的过程。问题一般包含如下因素:可选方案集合A,

属性集合C,代表属性重要程度的权重。多属性决策(Multi-Attribute Decision Making, MADM)的流程见图2,其中决策矩阵P的元素 c_{ij} 表示第i个方案对第j个属性的隶属程度,用直觉模糊数表示。

决策过程主要包含三个阶段:建模阶段、聚合阶段和选择阶段,其中第二个阶段是核心阶段。因此,人们提出了不同的聚合方法或技术来处理决策者通过直觉模糊集表达的评价信息。下面将对相关的方法展开讨论。

3.2 聚合算子

用于决策的众多聚合算子主要分为两类:一类基于可加测度,即算术均值、几何均值等加权聚合算子;另一类是基于不可加测度的聚合算子,如模糊积分聚合算子。

基于可加测度的聚合算子适用于IFN之间相互独立时。设 $\alpha_i = (\mu_i, \nu_i), i = 1, 2, \dots, n$ 是IFNs的集合, $w = (w_1, w_2, \dots, w_n)^T$ 是 α_i 权重的集合,满足 $\sum_{i=1}^n w_i = 1$ 。 $\sigma(i)$ 是满足 $\sigma(i) > \sigma(i+1), i = 1, 2, \dots, n$ 的有序排列。相关的聚合算子定义如下:

直觉模糊加权平均(Intuitionistic Fuzzy Weighted Average, IFWA)算子

$$IFWA[\alpha_1, \alpha_2, \dots, \alpha_n] = \left(\sum_{i=1}^n w_i \mu_i, \sum_{i=1}^n w_i \nu_i \right),$$

直觉模糊有序加权平均(Intuitionistic Fuzzy Ordered Weighted Average, IFOWA)算子

$$IFOWA[\alpha_1, \alpha_2, \dots, \alpha_n] = \left(\sum_{i=1}^n w_i \mu_{\sigma(i)}, \sum_{i=1}^n w_i \nu_{\sigma(i)} \right),$$

直觉模糊加权几何(Intuitionistic Fuzzy Weighted Geometric, IFWG)算子

$$IFWG(\alpha_1, \alpha_2, \dots, \alpha_n) = \left(\prod_{i=1}^n \mu_i^{w_i}, \prod_{i=1}^n \nu_i^{w_i} \right),$$

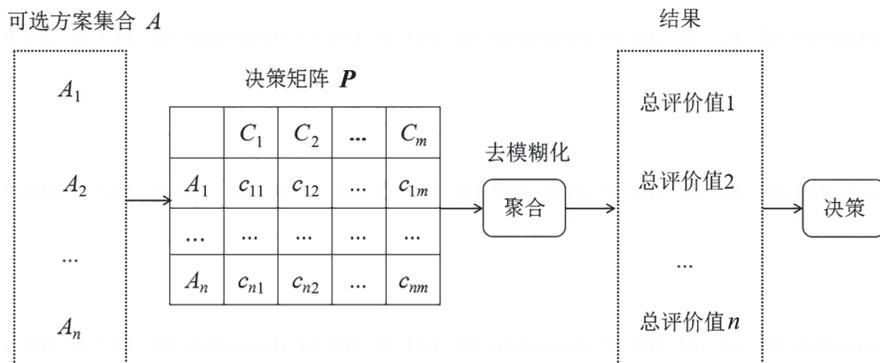


图2 多属性决策流程

直觉模糊有序加权几何(Intuitionistic Fuzzy Ordered Weighted Geometric, IFOWG)算子

$$\text{IFOWG}(\alpha_1, \alpha_2, \dots, \alpha_n) = \left(\prod_{i=1}^n (\mu_{\sigma(i)})^{w_i}, \prod_{i=1}^n (\nu_{\sigma(i)})^{w_i} \right)$$

不同的聚合算子之间是有密切联系的。当各直觉模糊数的权重相等,即 $w = (1/n, 1/n, \dots, 1/n)^T$ 时,IFWA 就成为直觉模糊平均(Intuitionistic Fuzzy Average, IFA)算子。类似地,当 $w = (1, 0, \dots, 0)^T$ 时,IFOWA 就成为最大值算子;当 $w = (0, 0, \dots, 1)^T$ 时,等价于最小值算子;当 $w = (1/n, 1/n, \dots, 1/n)^T$ 时,等价于 IFWA 算子。当 IFWG 算子中的权重相等时,等价于直觉模糊几何(Intuitionistic Fuzzy Geometric, IFG)算子。

此外,不同的聚合算子计算复杂度存在差异。我们只考虑乘法、除法、平方根和幂的次数,加减法的次数由于在计算机中计算速度较快而省略。直觉模糊加权平均算子 IFWA 的计算复杂度为 $O(2n)$,对于 IFOWA 算子,存在排序的过程,因此计算复杂度为 $O(2n+n\log 2n)$ 。IFWG 算子的计算复杂度为 $O(4n-2)$,而 IFOWG 算子的计算复杂度为 $O(4n+n\log 2n-2)$ 。

基于加权平均和基于加权几何平均的算子虽然都是对数据进行加权处理,但二者的侧重点不同。前者强调整体数据的影响,后者突出单个数据的作用。而有序类的算子首先需要给定数据从大到小排序之后,赋予每个位置一定的权重再进行聚合。无论哪种聚合算子,都是存在缺陷的。比如,对于 IFWG 算子,有一个评价中非隶属度为 0,最终的聚合结果总是 $(u, 0)$,出现违背直觉的现象,其他决策者的评价都被忽略不计了。因此,选用的聚合算子可能直接影响模型的决策结果。

$$\text{IFCA}(\alpha_1, \alpha_2, \dots, \alpha_n) = \left(\sum_{i=1}^n \mu_{(i)} (m(A_{(i)}) - m(A_{(i+1)})), \sum_{i=1}^n \nu_{(i)} (m(A_{(i)}) - m(A_{(i+1)})) \right)$$

$$\text{IFCG}(\alpha_1, \alpha_2, \dots, \alpha_n) = \left(\prod_{i=1}^n (\mu_{(i)})^{m(A_{(i)}) - m(A_{(i+1)})}, \prod_{i=1}^n (\nu_{(i)})^{m(A_{(i)}) - m(A_{(i+1)})} \right)$$

其他模糊积分也可以用来做聚合算子,例如 Sugeno 积分、凹积分和泛积分等。研究人员更倾向于使用 Choquet 积分。因为当模糊测度成为一般的可加测度时,Choquet 积分聚合算子便等价于前面讨论过的 IFWA、IFOWA 等。这些聚合算子都有幂等性、有界性、单调性等性质,详细内容及证明参见文献[9]。

当前的许多聚合运算符只是通过推广一些经典的

考虑到 MADM 问题的属性之间可能存在相互联系,而模糊测度具有不可加性质,恰好可以表征属性之间的各种关系。故研究人员也用基于不可加测度(即模糊测度)的模糊积分作为信息聚合的工具。

(定义 3.1) 设 X 是有限的集合, $\mathcal{P}(X)$ 是 X 的所有子集组成的集合,如果集函数 $m: \mathcal{P}(X) \rightarrow [0, 1]$ 满足^[11]

$$(1) m(\emptyset) = 0, m(X) = 1;$$

$$(2) E \in \mathcal{P}(X), F \in \mathcal{P}(X), E \subset F, \text{则 } m(E) \leq m(F).$$

那么称 m 为定义在 $\mathcal{P}(X)$ 上的模糊测度,其有如下性质:

如果 $m(E \cup F) < m(E) + m(F)$,则意味着 E, F 之间有相互削弱的关系,称为模糊测度的次可加性;

如果 $m(E \cup F) > m(E) + m(F)$,则意味着 E, F 之间有相互增强的关系,称为模糊测度的超可加性;

如果 $m(E \cup F) = m(E) + m(F)$,则意味着 E, F 之间不存在相互作用,并且满足可加性,模糊测度就等价于一般意义上的可加测度。

(定义 3.2) 给定模糊测度 m ,离散的 Choquet 积分被定义为^[12]: $Ch_m: [0, 1]^n \rightarrow [0, 1]$,

$$Ch_m(x) = \sum_{i=1}^n (x(\sigma(i)) - x(\sigma(i+1))) m(\{1, \dots, \sigma(i)\})$$

其中 $x(\sigma(n+1)) = 0$, $\sigma: \{0, 1, \dots, n\} \rightarrow \{0, 1, \dots, n\}$ 是满足 $x(\sigma(i)) \geq x(\sigma(j))$ 的排列, $i \geq j$ 。

开发基于上述模糊测度和模糊积分的聚合算子。设 $\alpha_i = (\mu_i, \nu_i)$, $i = 1, 2, \dots, n$ 是 IFNs 的集合, m 是模糊测度。 (\cdot) 表示 X 上的一种排列,使得对所有的 $i = 1, 2, \dots, n$, $\alpha_{(i)} \leq \alpha_{(i+1)}$, $A_{(i)} = (\{i, \dots, n\})$, $A_{(n+1)} = \emptyset$ 在决策中表示属性集合。有如下聚合算子

聚合运算符来定义的,从实际应用的角度来看,可能没有定义。在未来的研究中,不能只局限于加权平均的思想,应该开发更多新的聚合算子。另一方面,利用 Choquet 积分或者其他模糊积分进行聚合运算还没有得到广泛的研究,目前关注更多的是其纯数学的理论性质,相关研究可见文献[13-17]。而模糊测度的不可加性质更符合应用中的现实问题,具有研究的价值。

3.3 信息测度

在决策问题中主要会用到的信息测度包括距离测度、相似测度以及熵测度。他们之间同样既相互区别又相互联系。

3.3.1 距离测度

距离测度是描述两个对象差异的重要工具,学者已经提出了许多形式的直觉模糊距离测度。比如, Szmidt和Kacprzyk^[18]首先定义了IFS之间的四个基本距离度量:Hamming距离、标准化的Hamming距离、欧几里得距离和标准化的欧几里得距离。Chen^[19]等人提出了一种新的基于直角三角形质心的IFS之间的

距离测度,并证明了所提出的一些性质。Shen^[20]找到直角三角形中信息最丰富的点,并定义了新的距离公式。距离测度的定义如下。

(定义3.5)设 A, B, C 是直觉模糊集,如果映射 d 满足下列性质,则称之为距离测度^[20]:

$$(1) 0 \leq d(A, B) \leq 1;$$

$$(2) d(A, B) = 0 \text{ 当且仅当 } A = B;$$

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

$$(4) \text{若 } A \subseteq B \subseteq C, \text{ 那么 } d(A, C) \geq d(A, B) \text{ 以及 } d(A, C) \geq d(B, C)。$$

设 A, B 是论域 $X = (x_1, x_2, \dots, x_n)$ 上的两个直觉模糊集,常用的距离测度有

标准化的汉明距离

$$d_1 = \frac{1}{2n} \sum_{i=1}^n (|\mu_A(x_i) - \mu_B(x_i)| + |v_A(x_i) - v_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|),$$

标准化的欧几里得距离

$$d_2 = \sqrt{\frac{1}{2n} \sum_{i=1}^n ((\mu_A(x_i) - \mu_B(x_i))^2 + (v_A(x_i) - v_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2)},$$

质心距离

$$d_3 = \frac{|2(\mu_a - \mu_\beta) - (v_a - v_\beta)|}{3} \times \left(1 - \frac{\pi_a + \pi_\beta}{2}\right) + \frac{|2(v_a - v_\beta) - (\mu_a - \mu_\beta)|}{3} \times \left(\frac{\pi_a + \pi_\beta}{2}\right),$$

信息距离

$$d_4 = \sqrt{\frac{(\tilde{\mu}_a - \tilde{\mu}_\beta)^2 + (\tilde{v}_a - \tilde{v}_\beta)^2}{2}},$$

这里

$$\tilde{\mu}_a = \mu_a \left(1 + \frac{2}{3} \pi_a (1 + \pi_a)\right), \tilde{\mu}_\beta = \mu_\beta \left(1 + \frac{2}{3} \pi_\beta (1 + \pi_\beta)\right)$$

$$\tilde{v}_a = v_a \left(1 + \frac{2}{3} \pi_a (1 + \pi_a)\right), \tilde{v}_\beta = v_\beta \left(1 + \frac{2}{3} \pi_\beta (1 + \pi_\beta)\right)$$

用同一个数值例子来说明不同距离测度的差异。

(例3.1) 设 $\alpha_1 = \{ \langle x_1, 0.4, 0.5 \rangle, \langle x_2, 0.6, 0 \rangle, \langle x_3, 0.1, 0.5 \rangle \}$, $\alpha_2 = \{ \langle x_1, 0.5, 0 \rangle, \langle x_2, 0.5, 0.2 \rangle, \langle x_3, 0.4, 0.5 \rangle \}$, $\alpha_3 = \{ \langle x_1, 0.3, 0.3 \rangle, \langle x_2, 0.7, 0.1 \rangle, \langle x_3, 0.2, 0.6 \rangle \}$, 应用上述几种距离测度计算出的距离如表1。

表1 距离测度计算结果

	$d(\alpha_1, \alpha_2)$	$d(\alpha_1, \alpha_3)$	$d(\alpha_2, \alpha_3)$
d_1	0.3333	0.2333	0.2333
d_2	0.3317	0.2082	0.2082
d_3	0.1978	0.0306	0.1894
d_4	0.2996	0.0796	0.2389

从表中可以看出,尽管采用的距离公式不同,对三个IFN的距离排序是一致的,即 $d(\alpha_1, \alpha_2) > d(\alpha_2, \alpha_3) \geq d(\alpha_1, \alpha_3)$ 。但是,前两个公式计算出来 (α_1, α_3) 和 (α_2, α_3) 的距离是相等的。如果这是一个分类问题或者模式识别问题,那么利用标准化的欧氏距离或者标准的汉明距离将无法实现目标。事实上, α_2 与 α_3 的犹豫程度更接近,从直觉角度认为 α_2 与 α_3 的距离应该小于 α_1 与 α_3 的距离。所以提出的距离公式除了具有区分度以外,还应该考虑符合直觉性。

相似性测度和距离测度类似。两组数据之间的

距离越大,那么相似性就越小。所以在计算相似度 l 时,常采用 $l = 1 - d$ 的变换方法。但两者还是有区别的。在应用中,如果更重视差异,如异常检测、分类等,常使用距离测度。如果更重视相似程度,如聚类和推荐系统,常用相似测度。

3.3.2 直觉模糊熵测度

Shannon 在 1948 年提出了熵值理论,主要思想是指标的变异程度越大,包含的信息越多,熵值越小,则对应的权重也应该越大。Szmidt^[21]总结出直觉模糊熵(Intuitionistic Fuzzy Entropy, IFE)应满足的准则,成为学者研究新的 IFE 的参照条件。Fu^[22]等人研究多属性群体决策(Multi-Attribute Group Decision-Making, MAGDM)问题,用于网络舆论突发事件应急响应决策,研究了几种已有的 IFE 及其优缺点,进而提出了改进的基于三角函数的直觉模糊熵公式。Yuan 和 Luo^[23]得出结论,权重不仅依赖于熵度量,还依赖于可选方案和属性的数量。Chen^[24]等人提出了基于指数函数的 IFE,结合情绪分析挖掘民意,建立了 MAGDM 信息融合模型。

设 E 为直觉模糊熵, A, B 是 X 上的两个直觉模糊集。熵测度 $E(A)$ 表示映射: $\text{IFS}(X) \rightarrow [0, 1]$, 满足以下条件^[24]:

- (1) $E(A) = 0$, 当且仅当 A 是经典集合;
- (2) $E(A) = 1$, 当且仅当对于 $\forall x_i \in X, \mu_A(x_i) = \nu_A(x_i)$;
- (3) $E(A) = E(A^c)$;
- (4) 如果 $A \leq B$, 那么 $E(A) \leq E(B)$;

Jiang^[25]给出了公理的改进,增加了新的规则:

- (5) 如果 $\pi_A(x) \leq \pi_B(x)$ 并且 $|\mu_A(x) - \nu_A(x)| = |\mu_B(x) - \nu_B(x)|$, 那么 $E(A) \leq E(B)$ 。

下面总结了引用率比较高的熵测度公式,来讨论直觉模糊熵测度。

$$E_1(A) = 1 - \frac{1}{n} \sum_{i=1}^n |\mu_A(x_i) - \nu_A(x_i)|$$

$$E_2(A) = \frac{1}{n} \sum_{i=1}^n \cos \frac{\mu_A(x_i) - \nu_A(x_i)}{2(1 + \pi_A(x_i))} \pi$$

$$E_3(A) = \frac{1}{n} \sum_{i=1}^n \frac{\min \{ \mu_A(x_i), \nu_A(x_i) \} + \pi_A(x_i)}{\max \{ \mu_A(x_i), \nu_A(x_i) \} + \pi_A(x_i)}$$

$$E_4(A) = \frac{1}{n} \sum_{i=1}^n \frac{\mu_A(x_i) \nu_A(x_i) + \pi_A(x_i)}{\mu_A(x_i)^2 + \nu_A(x_i)^2 - \mu_A(x_i) \nu_A(x_i) + \pi_A(x_i)}$$

$$E_5(A) = \frac{1}{n} \sum_{i=1}^n \cos \frac{(\mu_A(x_i) - \nu_A(x_i))(1 - \pi_A(x_i))}{2} \pi$$

(例 3.2) 设直觉模糊数 $A_1 = \langle 0, 0.2 \rangle, A_2 = \langle 0.9, 0 \rangle, A_3 = \langle 0.3, 0.4 \rangle, A_4 = \langle 0.1, 0.2 \rangle, A_5 = \langle 0.3, 0.3 \rangle$, 由以上熵测度公式计算出的值如表 2。

表 2 熵测度计算结果

	A_1	A_2	A_3	A_4	A_5
E_1	0.8000	0.1000	0.9000	0.9000	1.0000
E_2	0.9848	0.2817	0.9927	0.9957	1.0000
E_3	0.8000	0.1000	0.8571	0.8889	1.0000
E_4	0.9524	0.1099	0.9767	0.9863	1.0000
E_5	0.9980	0.2940	0.9939	0.9989	1.0000

从上表可以发现, $E_1(A_3) = E_1(A_4)$, 比较 A_3 和 A_4 , A_4 的模糊程度更大, 故蕴含的信息更大, 模糊熵也应该越大。原因在于模糊熵测度 E_1 中不包含犹豫度函数, 这也是直觉模糊集的三个决定函数之一, 不能被忽略; 前四个公式都可以得出熵测度的大小关系 $E(A_5) > E(A_4) > E(A_3) > E(A_1) > E(A_2)$, 唯独第五个式子中 A_3 和 A_1 的大小顺序颠倒了。而事实上 $A_1 < A_3$, 所以按照公理的要求, 应该满足 $E(A_1) < E(A_3)$; $E(A_5) = 1.0000$, 这表明模糊程度最大的时候, 蕴含的信息也最多, 应重点关注。

另外, Zeng^[26]等讨论了直觉模糊集间的相似测度与直觉模糊熵的关系, 并证明了直觉模糊熵与相似测度可以相互转化。至此, 我们可以得出结论, 在一定条件下, 直觉模糊熵的距离测度、相似测度以及熵测度之间都是可以转化的。选择合适的测度方法, 可以帮助我们精准决策。但是, 目前对于信息测度方法相关的解释仍不足, 例如一些几何解释, 每项操作的含义等, 需要对真正有用性进行明确的理由分析。

4 直觉模糊集在网络舆情决策的应用

4.1 决策模型

决策主要有多准则决策(Multiple Criteria Decision Making, MCDM)、多属性决策, 有时候需要收集多个具有不同知识结构和经验的决策者来进行群体决策, 决策步骤是相似的, 本文讨论一般的多属性决策问题。多属性决策模型的步骤如下:

Step1: 输入直觉模糊决策矩阵

$$P = (p_{ij})_{m \times n} = (\mu_{ij}, \nu_{ij})_{m \times n}。$$

Step2: 数据预处理。数据预处理主要包括: 正向化(将所有指标转化为极大值)、标准化(消除量纲的影响)、归一化(将属性值排列在 $[0, 1]$ 区间上)。

Step3:确定属性的权重。属性权重越大,表明该属性在我们所研究的问题中的重要性越大。

Step4:信息聚合。选用合适的聚合算子,对决策矩阵和属性的权重进行信息融合,得到综合评估结果。

Step5:排序选择,做出决策,输出决策结果。

事实上,各种算法模型的前两步基本是一致的。主要区别在于后面求权重的方法、信息聚合算子以及如何排序。

4.2 应用概况

我们在数据库中搜索近年以来的文章,关键词考虑“直觉模糊集”、“舆情”、“决策”,在表3中汇总了发表在国内外期刊上的相关文章。决策模型中包含多种方法的组合,下面按照决策模型的步骤展开讨论。

首先关注把信息转换为直觉模糊决策矩阵的方法。德尔菲法是一种专家咨询的方法,在模糊决策领域也最常用,如文献[11, 22, 27, 28],由多位具有不同知识结构和经验的专家给出关于隶属度、非隶属度和犹豫度的意见值,得到决策矩阵。需要注意的是,对于群体决策而言,一般满足专家人数在20人及以上。基于概率统计的方法得到模糊化的信息矩阵见文献[4-5],涉及到获取网络数据或政府官方公布的数据,如新闻的点击数量、跟帖数量等,将其转化为直觉模糊数。情感分析等自然语言处理技术也受到舆情研究者的关注,但是从目前的研究来看,数据大都来自于微博,大数据环境下网络舆情的其他平台数据和隐性指标未能完全考虑。在未来研究中,应选取更多的数据源。

确定权重最直观的方法是层次分析法 (Analytic Hierarchy Process, AHP)。需要确定多级指标,依据专家两两比较进行打分,通过一致性检验的结果才可使用。基于熵测度的方法是一种完全客观的方法,从数据本身获取权重,因此不断有新的直觉模糊熵公式出现。将AHP和IFE所得的权重进行线性组合也是一种改进的方法,综合体现了主客观意见。此外,也有学者采用优化模型确定未知权重的大小,如文献[20, 22],需要严格把控所研究问题的约束条件。

不同的模糊集有相应的聚合算子。文献[11, 29]分别基于Copula函数和阿基米德范数定义了概率区间犹豫模糊集和语言直觉模糊集中的新算子,还证明了算子的数学性质。算子中设置可变函数,若取不同的函数表达式就可以得到不同的聚合形式,而不改变决策结果。

模型最后是对结果排序。如果聚合结果是直觉模糊数的形式,可以通过比较评分函数和精度函数直接进行排序择优。若聚合结果是数值的形式,基于格贴近度或理想解的思想进行排序,如优劣距离法 (Technique for Order Preference by Similarity to Ideal Solution, TOPSIS)、折中妥协法 (VISeKriter-ijumska Optimizacija I Kompromisno Resenje, VIKOR) 等方法,设定到正理想解的距离最小、到负理想解的距离最大。理想解的定义也有不同,如选用绝对理想解,信息完全已知即 $IFN = (1, 0)$ 为正理想解,完全未知即 $IFN = (0, 1)$ 为负理想解,或者相对理想解,分别取决策矩阵的最大、最小直觉模糊数。模糊推理适用于定性分析,允许基于信息不准确或不完整情况做出决

表3 近年来直觉模糊理论在舆情决策领域的文献

年份	作者	模糊理论	方法	决策模型	应用
2017	张艳丰等 ^[4]	直觉模糊集	直觉模糊推理、改进的施拉姆大众信息传播模式、AHP、模糊综合评判	—	网络舆情监测预警评估
2019	周小领等 ^[29]	概率区间犹豫模糊集	阿基米德范数、运算法则、有序加权几何算子、	MAGDM	网络舆情预测系统
2019	Xu等 ^[30]	区间直觉模糊集	自然语言处理、模糊关联规则挖掘、模糊知识图谱、信息熵、聚类、评分函数、精度函数	MAGDM	群体突发事件应急方案决策
2020	Zhang等 ^[31]	概率语言术语集	Wasserstein距离测度、动态权重	MAGDM	网络舆情突发事件评估
2021	林玲等 ^[32]	直觉模糊集	决策者风险偏好、Choquet积分、投影法、得分函数	—	网络舆情预警
2021	Roszkowska等 ^[33]	直觉模糊集	TOPSIS、距离测度	MCDM	公众生活质量调查与评估
2022	Liu等 ^[11]	语言直觉模糊集	语言直觉模糊 Choquet-Copula 聚合算子	MAGDM	网络突发舆情事件
2022	Fu等 ^[22]	区间直觉模糊集	直觉模糊熵、距离测度、理想解、偏好建模	MAGDM	多个网络舆情突发事件的优先处理决策
2022	Chen等 ^[24]	直觉模糊集	情绪分析、直觉模糊熵、聚类、群体共识	MAGDM	公众与专家参与的舆情决策
2022	Peng等 ^[28]	语言直觉模糊集	Choquet积分算子、Shapley函数	MADM	雾霾网络舆情风险评估
2023	朱国庆等 ^[27]	直觉模糊集	信息生态系统、模糊规则、直觉模糊数、直觉模糊熵、AHP、效用值方法、德尔菲法、汉明距离	—	森林火灾网络舆情风险预警评估

策,采用“IF-THEN”规则库,设定模糊规则后,推理结果即评价结果。

从表3中可以看出,传统的直觉模糊集及拓展的直觉模糊集在舆情分析领域有广泛的应用价值。目前研究最多的是直觉模糊集,相应的方法都可以扩展到其他模糊集上。每种方法都有自己的特点,决策实践者在实际决策中很难选择合适的方法。因此,需要对这些方法进行充分比较,展示每种方法的优点和缺点。群体决策比单一的决策更加科学,但是表征决策者偏好的研究较少,还需深入研究。此外,目前应用领域集中在应急决策、舆情风险评估、舆情预测等,网络舆情领域还有许多其他问题如舆情引导、舆情传播动力学研究、演化规律等问题,研究较少。我们可以考虑这些问题是否能利用模糊集或者直觉模糊集的方法得以更好地解决。

5 总结

模糊理论的有关研究已经在模糊控制、模式识别、模糊聚类分析、模糊决策、系统理论、信息检索、机器学习等各个方面都取得了广泛的应用。直觉模糊集及其相关的理论为人们表达信息和做出决策提供了一种灵活便捷的方式,近年来受到广泛关注,并引申出了各种决策方法和模型。本文主要介绍了直觉模糊集的基本定义、常用的理论和方法,以及在多属性决策模型中的应用。同时,梳理了其在舆情决策领域的研究现状,展示了所采用的方法、具体应用的问题。通过比较分析当前研究成果,为学者找到最适合的方法或者开发出新方法提供一些参考。

在未来的研究工作中,还可以考虑将其他形式的模糊集用于网络舆情决策,如比直觉模糊集表达能力更强的毕达哥拉斯模糊集、 q 阶正交模糊集等。由于多属性决策问题的一般性,在解决舆情问题时,也可以参考其他领域的决策方法。

参考文献(References):

- [1] 李昌祖,张洪生.网络舆情的概念解析[J].现代传播:中国传媒大学学报,2010(9):139-140.
- [2] 刘毅.基于三角模糊数的网络舆情预警指标体系构建[J].统计与决策,2012(2):12-15.
- [3] 李晚莲,高光涵.突发公共事件网络舆情热度生成机理研究——基于48个案例的模糊集定性比较分析(fsQCA)[J].情报杂志,2020,39(7):94-100.
- [4] 张艳丰,李贺,彭丽徽.基于直觉模糊推理的网络舆情监测预警评估方法研究[J].情报杂志,2017,36(6):111-117.
- [5] Madhusudhanan S, Moorthi M. Optimized fuzzy technique for enhancing sentiment analysis[J]. Cluster Computing, 2019, 22(Suppl 5): 11929-11939.
- [6] Sánchez H, Aguilar J, Terán O, et al. Modeling the process of shaping the public opinion through Multilevel Fuzzy Cognitive Maps[J]. Applied Soft Computing, 2019, 85: 105756.
- [7] Zadeh L A. Fuzzy sets[J]. Information and Control, 1965, 8(3): 338-353.
- [8] Atanassov K T, Stoeva S. Intuitionistic fuzzy sets[J]. Fuzzy Sets and Systems, 1986, 20(1): 87-96.
- [9] 徐泽水.直觉模糊信息集成理论及应用[M].北京:科学出版社,2008.
- [10] Xu Z, Zhao N. Information fusion for intuitionistic fuzzy decision making: an overview[J]. Information Fusion, 2016, 28: 10-23.
- [11] Liu Y, Wei G, Liu H, et al. Group decision making for internet public opinion emergency based upon linguistic intuitionistic fuzzy information[J]. International Journal of Machine Learning and Cybernetics, 2022; 13:579 - 594.
- [12] Choquet G. Theory of capacities[J]. Annales de l'institut Fourier. 1954, 5: 131-295.
- [13] Kang T, Li J. On equivalence of decomposition integrals based on different monotone measures[J]. Fuzzy Sets and Systems, 2023, 457: 142-155.
- [14] Kang T, Yan L, Li J. Coincidences of nonlinear integrals related to ordered pair of non-additive measures[J]. International Journal of Approximate Reasoning, 2023, 152: 124-135.
- [15] Kang T, Yan L, Ye L, et al. Some notes on the pan-integrals of set-valued functions[J]. Fuzzy Sets and Systems, 2024: 108794.
- [16] Yan L, Kang T, Zhang H. Decomposition integrals of set-valued functions based on fuzzy measures[J]. Mathematics, 2023, 11(13): 3013.
- [17] Kang T, Wu D, Li J. Pan-integrals of set-valued functions based on fuzzy measures[J]. Fuzzy Sets and Systems, 2023: 108632.
- [18] Szmjdt E, Kacprzyk J. Distances between intuitionistic fuzzy sets[J]. Fuzzy sets and systems, 2000, 114(3): 505-518.
- [19] Chen S M, Cheng S H, Lan T C. A novel similarity measure between intuitionistic fuzzy sets based on the centroid points of transformed fuzzy numbers with applications to pattern recognition[J]. Information Sciences, 2016, 343: 15-40.
- [20] Shen F, Ma X, Li Z, et al. An extended intuitionistic fuzzy TOPSIS method based on a new distance measure with an application to credit risk evaluation[J]. Information Sciences, 2018, 428: 105-119.

- [21] Szmidi E, Kacprzyk J. Entropy for intuitionistic fuzzy sets [J]. *Fuzzy Sets and Systems*, 2001, 118(3): 467-477.
- [22] Fu S, Xiao Y, Zhou H. Contingency response decision of network public opinion emergencies based on intuitionistic fuzzy entropy and preference information of decision makers[J]. *Scientific Reports*, 2022, 12(1): 3246.
- [23] Yuan J, Luo X. Approach for multi-attribute decision making based on novel intuitionistic fuzzy entropy and evidential reasoning [J]. *Computers & Industrial Engineering*, 2019, 135: 643-654.
- [24] Chen X, Zhang W, Xu X, et al. A public and large-scale expert information fusion method and its application: mining public opinion via sentiment analysis and measuring public dynamic reliability [J]. *Information Fusion*, 2022, 78: 71-85.
- [25] Jiang D, Wang Y. A new entropy and its properties based on the improved axiomatic definition of intuitionistic fuzzy entropy[J]. *Mathematical Problems in Engineering*, 2018, 2018: 1-6.
- [26] Zeng W, Li H. Relationship between similarity measure and entropy of interval valued fuzzy sets [J]. *Fuzzy sets and Systems*, 2006, 157(11): 1477-1484.
- [27] 朱国庆, 陈燕, 郎坤, 等. 信息生态视角下森林火灾网络舆情风险预警研究[J]. *东北大学学报(自然科学版)*, 2023, 44(8): 1208.
- [28] Peng B, Zheng C, Wei G, et al. Risk assessment method on haze networks public opinion based on intuitionistic fuzzy choquet integral [J]. *International Journal of Fuzzy Systems*, 2022, 24(6): 2858-2872.
- [29] 周小颖, 马庆功. 概率犹豫模糊算法及其网络舆情预测模型选择[J]. *计算机工程与应用*, 2019, 55(04): 179-184+192.
- [30] Xu X, Yin X, Chen X. A large-group emergency risk decision method based on data mining of public attribute preferences [J]. *Knowledge-Based Systems*, 2019, 163: 495-509.
- [31] Zhang S, Ma Z, Liu X, et al. An integrated consensus improving strategy based on PL-Wasserstein distance and its application in the evaluation of network public opinion emergencies[J]. *Complexity*, 2020, 2020: 1-24.
- [32] 林玲, 陈福集, 谢加良, 等. 考虑风险偏好的网络舆情预警模型——基于直觉模糊和 Choquet 积分[J]. *情报杂志*, 2021, 40(10): 52-58.
- [33] Roszkowska E, Kusterka-Jefmańska M, Jefmański B. Intuitionistic fuzzy TOPSIS as a method for assessing socioeconomic phenomena on the basis of survey data[J]. *Entropy*, 2021, 23(5): 563.

编辑:赵志军