

# 大样本下基于空间向量聚类的非概率证据融合方法研究

曹黎侠, 王嘉鑫, 毛春妮

(西安工业大学 基础学院, 陕西 西安 710021)

**摘要:** 针对大样本情况下的冲突证据融合问题, 综合考虑证据之间的支持关系和证据本身的不确定性, 提出了一种基于空间向量 Jousselme 距离聚类的证据融合新方法, 用于处理大样本情况下的证据融合问题, 并降低证据冲突带来的影响。基于逐步融合的理念, 先通过聚类分析计算出证据间的距离对证据进行分类, 再采用 Murphy 方法对同一类证据进行融合, 然后对初步融合后的新证据运用 D-S(Dempster-Shafer) 组合规则实现排序和决策; 最后将该方法用于高校在校心理社会性发展问题的研究。实证分析表明, 新方法有效可行, 实证结果可为心理社会性发展预测、评估等研究提供参考。

**关键词:** 证据冲突; 大样本; 向量空间聚类; 心理社会性发展

**中图分类号:** O213.2; O29 **文献标志码:** A **文章编号:** 1001-8735(2025)03-0221-08

**DOI:**10.3969/j.issn.1001-8735.2025.03.001

复杂系统是人工智能领域研究的热点问题之一, 由于系统的不确定性、动态性、多源性、高维性等特征, 使得复杂系统的优化、数据处理、信息融合等成为人工智能领域迫切需要解决的问题和社会关注的焦点<sup>[1-3]</sup>。信息融合作为处理矛盾数据和缺失信息的技术, 受到广泛重视。作为一种信息融合策略, 证据理论尤为适用于充满不确定因素的情景中, 它能够有效构建和评估信息的模糊性与不准确性, 区分“未知”和“不确定”信息, 增强信息之间的互补关系, 提高决策的准确性<sup>[4-5]</sup>。但也存在高度矛盾的证据冲突可能会得到无效结果的现象。

鉴于此, 一些学者采用加权平均法对冲突信息进行平均分配, 再采用 D-S 组合规则对加权证据进行融合, 使融合信息集中于不确定性测度小的属性<sup>[6]</sup>, 但这种方法没有考虑证据之间的关联度等特征, 可能导致结果误判。另一些学者针对 D-S 组合规则在证据高度冲突时产生失效、反直觉的结果, 研究出两种方法: 其一是对 D-S 组合规则的修改; 其二是对证据主体进行预处理<sup>[7-8]</sup>。还有学者研究表明, 当前融合信息的种类越来越多元化, 但没有一个通用且被统一认可的信息融合方法, 可以彻底解决不确定性和高度冲突的数据融合问题<sup>[9]</sup>。而且目前的 D-S 组合规则更适用于中、小样本, 处理大样本时计算步骤烦琐, 而得到的结果并不理想<sup>[10]</sup>。因此, 本文提出一种基于证据相似性的分类方法来改进证据组合规则, 在大样本情况下, 提高证据融合的准确性并减少计算量。

## 1 理论基础

### 1.1 信息融合

信息融合是一种形式框架, 常用的方法有加权平均法、贝叶斯推理、证据理论、卡尔曼滤波、模糊理论和神经网络等。目前主要是基于信息融合层次进行分类, 分为数据层融合方法、特征层融合方法以及决策层融合方法, 每个方法又可运用数学和统计学的不同方法。

收稿日期: 2024-11-14

基金项目: 陕西省科技厅软科学研究资助项目“‘一带一路’生态共同体建设中的陕西作用与策略研究”(2022KMR056); 西安工业大学研究生教育教学改革研究资助项目“教育新质生产力催生研究生心理社会性发展增长点的培育分析”(XAGDYJ240108)。

作者简介: 曹黎侠(1972—), 女, 教授, 博士, 主要从事应用统计、复杂网络、不确定性数学理论等研究, E-mail: caolx\_8@163.com。

### 1.2 证据理论

设  $\Theta$  是一个识别框架,由系统的所有假设命题或所有可能的决策或判断结论构成,表示为  $\Theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_n\}^{[11]}$ 。

基本概率分配(BPA)是定义在识别框架  $\Theta$  上的一个函数  $m(2^\Theta \rightarrow [0, 1])$ ,称为 mass 函数,并且满足  $m(\emptyset) = 0$  且  $\sum_{A \subseteq \Theta} m(A) = 1, m(A) > 0$  的  $A$  称为焦点。

信度函数(belief function),在识别框架  $\Theta$  上基于 BPA  $m$  的函数

$$Bel(A) = \sum_{B \subseteq A} m(B)。$$

似然度函数(plausibility function),在识别框架  $\Theta$  上基于 BPA  $m$  的函数

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B)。$$

对于  $\forall A \subseteq \Theta$ ,识别框架  $\Theta$  上的有限个 mass 函数  $m_1, m_2, \dots, m_n$  的 D-S 合成规则为

$$(m_1 \oplus m_2 \oplus \dots \oplus m_n)(A) = \frac{1}{K} \sum_{A_1 \cap A_2 \cap \dots \cap A_n = A} m_1(A_1) \cdot m_2(A_2) \cdot \dots \cdot m_n(A_n), \tag{1}$$

其中,

$$K = \sum_{A_1 \cap A_2 \cap \dots \cap A_n \neq \emptyset} m_1(A_1) \cdot m_2(A_2) \cdot \dots \cdot m_n(A_n) = 1 - \sum_{A_1 \cap A_2 \cap \dots \cap A_n = \emptyset} m_1(B) \cdot m_2(C) \cdot \dots \cdot m_n(A_n)。 \tag{2}$$

冲突系数  $K$  越大,表明各证据间的矛盾越尖锐,融合后会导致错误结论。

## 2 基于空间向量聚类的改进证据理论方法

### 2.1 大样本情况下信息融合的改进方向

信息融合过程中最常用的方法是加和乘的规则,当证据之间冲突比较大,加性规则可以有效地将冲突信息进行平均分配以降低影响;当证据之间冲突较小,乘性规则可使融合信息向不确定性小的属性靠拢。大样本情况下指标维数多,为统计分析建模带来不便,并且变量之间会有一定的联系。因此,选择变量之间的相似关系将其聚合为若干类,以减少与证据之间的冲突。

鉴于此,本文提出改进的 D-S 证据理论方法,基本原理是通过分类加权多次融合的方式,有效减少因证据之间冲突对融合结果的影响。首先将同类证据聚合为一个整体,利用加性规则对所含信息初步融合,再基于 Jousselme 距离确定分类后新证据之间的权重,运用乘性规则对初步融合的证据加权二次融合,最后将二次融合结果代入组合规则中再融合。

### 2.2 改进的证据融合方法

大样本情况下信息融合的核心方法有聚类分析、Murphy 方法及 Jousselme 距离。本文针对当前这 3 种方法存在的弊端,提出一种改进的分类加权的新方法。

#### 2.2.1 改进方法的设计思路

(1)空间向量的聚类。本文分别运用 R 型和 Q 型聚类分析<sup>[12]</sup>,将空间类似向量组成多个类,具体步骤如下。

Step1 运用肘部法则,确定分类数  $k$ 。肘部法则是一种用于确定在  $k$  均值聚类算法中质心数的技术。肘部法的核心思想是分类数  $k$  越大,样本划分越精细,每个类的聚合程度逐渐提高,误差平方和(SSE)逐渐变小。利用连续迭代确定  $k$  值,从  $k=1$  到  $k=n$ ( $n$  是根据要求选择的超参数),随着  $k$  的增加,SSE 会随之降低,当下降幅度明显趋缓时,定为  $k$  值。利用公式(3)得到最小化点到聚类中心的距离  $D$ :

$$D = \sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2)。 \tag{3}$$

Step2 分别运用 R 型和 Q 型聚类分析,确定分类并比较聚类结果。根据 R 型聚类法,首先确定每个样本属于每个簇的隶属度,然后运用式(4)计算每个样本到每个簇中心的距离  $r$ ,再根据当前的隶

属度重新计算每个样本属于每个簇数的隶属度,最后根据当前隶属度重新计算每个簇的中心。重复上述操作,一直到隶属度不再发生变化或达到最大的迭代次数。

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

根据 Q 型聚类法,首先构造  $n$  个类,每个类中只包含一个样本点,每一类的平台高度为 0,然后运用式(5)计算  $n$  个样本点两两之间的距离  $d$ ,再将距离最近的两类合并为新类,并以这两类间的距离值作为新的平台高度,最后再次计算新类与当前各类的距离,若类的个数已经等于 1,则分为一类,否则继续上一步,直到分为  $n$  类。

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

(2)运用 Murphy 方法对分类后的证据初步融合。Murthy 在平均信度优先情况下,提出了式(6)处理证据冲突:

$$\bar{m}(A) = \frac{1}{n} \sum_{i=1}^n m_i(A), \forall A \subseteq \Theta \quad (6)$$

据式(6),Murphy 对 D-S 组合规则进行  $(n-1)$  次组合计算得到融合结果。

(3)基于 Jousselme 距离解决证据冲突的方法。Jousselme 等<sup>[13]</sup>从空间向量的角度提出在向量空间中测量证据距离的方法,利用证据距离来描述两种不同来源证据之间的冲突,具体步骤如下。

Step1 确定证据间的相似性。

$$d(m_1, m_2) = \left[ \frac{1}{2} (m_1 - m_2) D (m_1 - m_2)^T \right]^{\frac{1}{2}} \quad (7)$$

其中  $m_j$  为行向量,  $D$  为  $2^n \times 2^n$  的矩阵,其元素定义为  $D(i, j) = |A_i^1 \cap A_j^2| / |A_i^1 \cup A_j^2|$ ,  $A_i^k$  表示第  $k$  个证据  $m_k$  的第  $i$  个子集;  $|A|$  表示集合中元素个数。

Step2 计算每个证据的可信度。经过 Step1 计算出证据距离,构建出证据相似度以及证据可靠度,见式(8)–(10):

$$\text{si}(m_i, m_j) = 1 - d(m_i, m_j), \quad (8)$$

$$\text{sup}(m_i) = \sum_{j=1, j \neq i}^n \text{si}(m_i, m_j), \quad (9)$$

$$\text{cred}(m_i) = \frac{\text{sup}(m_i)}{\sum_{j=1}^n \text{sup}(m_j)} \quad (10)$$

其中  $\text{si}(m_i, m_j)$  表示证据  $m_i$  与  $m_j$  之间的相似度,  $\text{sup}(m_i)$  表示其他证据对证据  $m_i$  的支持度,  $\text{cred}(m_i)$  表示为证据  $m_i$  支持度的归一化值,表示证据  $m_i$  的可信度,可以作为证据  $m_i$  的权重。

Step3 将计算出的权重对证据进行加权平均。对相互独立的证据  $m_j$ ,利用可信度  $\text{cred}(m_j)$  加权处理,得到平均证据,即  $A_r$  基本概率的加权平均值

$$\bar{m}_w(A_r) = \sum_{j=1}^n \text{cred}(m_j) m_j(A_r) \quad (11)$$

Step4 利用组合规则,即公式(1)对加权平均证据  $\bar{m}_w(A_r)$  合成  $(n-1)$  次。

### 2.2.2 改进方法的算法步骤

依据以上设计思路,改进方法的算法步骤如下。

Step1 基于证据的相似度对证据进行分类。首先通过肘部法则确认分类簇数  $k$ ,然后利用 R 型和 Q 型聚类法对证据分类,记为  $T_{\theta i}$ ,其中  $\theta = R, Q$  融合的方法,  $i = 1, 2, \dots, n$ ,  $n$  为分类数。

Step2 对同类型证据进行融合。采用 Murphy 方法对集合  $T_{\theta_i}$  中的证据  $m_{j_i}(j_i=1,2,\dots,n_i)$  进行数据融合,将融合后的证据记为  $\bar{m}_{\theta_i}$ ,表示分类后该类型  $\theta_i(i=1,2,\dots,n)$  证据的融合结果。

Step3 对不同类型的证据进行加权。通过 Jousselme 距离度量证据之间的距离<sup>[14-15]</sup>,确定证据  $m_{j_i}$  的信度,证据  $\bar{m}_{\theta_i}$  的可信度表示集合  $T_{\theta_i}$  中的所有证据  $m_{j_i}(j_i=1,2,\dots,n_i)$  的信度总和,即

$$\text{cred}(\bar{m}_{\theta_i}) = \sum_{j_i=1}^{n_i} \text{cred}(m_{j_i}),$$

其中  $m_{j_i}$  为分类后  $\theta_i$  的证据,  $\text{cred}(\bar{m}_{\theta_i})$  为证据  $\bar{m}_{\theta_i}$  的权重  $w_{\theta_i}$ 。利用可信度  $\text{cred}(\bar{m}_{\theta_i})$  对证据  $\bar{m}_{\theta_i}$  加权处理得到加权平均证据,即  $A_r$  基本概率的分类加权平均值为

$$\bar{m}_{cw}(A_r) = \sum_{i=1}^n \text{cred}(m_{\theta_i}) m_{\theta_i}.$$

Step4 利用证据理论组合规则对加权平均证据  $\bar{m}_{cw}(A_r)$  合成  $(s-1)$  次,  $s(1 \leq s \leq n)$  为证据分类个数。

### 3 新证据融合方法的应用

#### 3.1 高校在校心理社会性发展研究设计

3.1.1 调查问卷设计 在校大学生虽然拥有良好的教育条件、生活环境,具有自信、开阔的眼界以及独立的个性,但是在面对生活中所出现的复杂关系和情况时,可能会对他们的自信心造成一定的打击,甚至还会对其信念、认知产生强烈的冲击<sup>[16]</sup>。特别是近几年全球经济形式持续低迷,给在校大学生造成不小的心理压力<sup>[17]</sup>。因此,本文对在校心理社会性发展进行调查,并做简单评估。

本研究数据主要来源于问卷调查。依据影响在校心理社会性发展的主要因素设计了 23 个题目,题目内容可以分为情景认知、自我认同、人际认知、亲社会情感、人际交往能力、自我控制、社会性发展以及环境适应八个不同的维度。

3.1.2 问卷数据的整理 发放问卷 323 份,有效问卷 303 份。问卷内容是基于  $n$  个证据  $m$  个命题,数据处理时需要将选项进行分类,按照问卷评价中的赋分情况进行数据统计。为确保问卷赋分的有效性,本文选择专家评判法<sup>[18]</sup>,邀请 5 位专家、老师对问卷赋分情况进行评估,并分为“合理”“较合理”“一般”“较不合理”“不合理”5 个评价等级。

问卷中的 1—7 题是用于确认在校身份基本情况,8—23 题用于在校心理社会性发展研究,但第 8 题数据无法直接用于研究,在总结时作为因素之一。按照问卷评价中的赋分进行数据统计,将赋分后的同一分值的选项百分比进行相加,且选用每个题目所有选项归一化后的值(问卷和数据略)。

#### 3.2 基于聚类分析的改进证据理论方法应用

3.2.1 传统证据理论的应用 根据整理后的问卷数据,将在校心理社会性发展按照健康等级划分为优、良、一般、不良,即识别框架为  $\Theta = \{r_1, r_2, r_3, r_4\}$ ,其中  $r_1$  表示优,  $r_2$  表示良,  $r_3$  表示一般,  $r_4$  表示不良,15 个题目作为证据  $m_i(i=1,2,\dots,15)$ ,是定义在识别框架  $\Theta$  中的 15 个独立的 BPA。假设两个及以上元素 mass 函数值均为 0,即  $r_1, r_2, r_3, r_4$  相互独立。

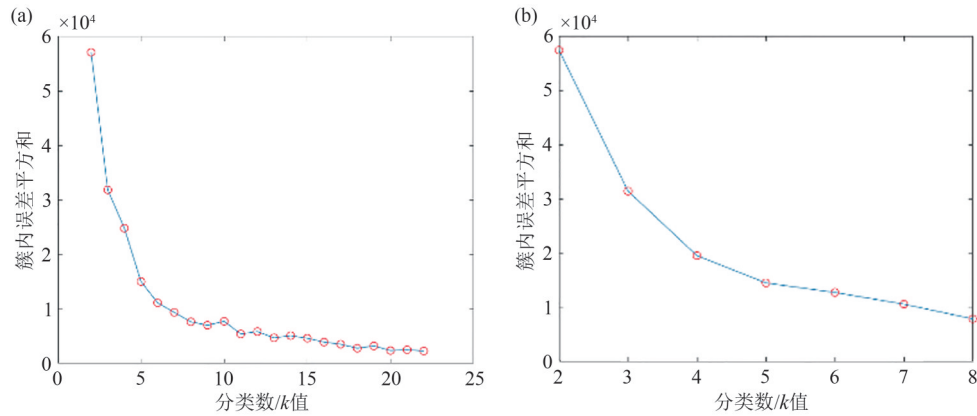
根据传统的证据理论组合规则,将数据代入 D-S 合成规则式(1)中,将所有证据进行融合,使用 MATLAB 运行出融合结果为

$$r_1 = 0.00000297, r_2 = 0.99999702, r_3 = 0.00000001, r_4 = 0. \tag{12}$$

传统方法融合结果与问卷收集结果有较大出入,即与直觉相悖。

#### 3.2.2 改进后的证据理论模型

Step1 利用肘部法确认分类的簇数。不同  $k$  值的聚类偏差图如图 1 所示。根据拐点判断出簇数为 6 和 4,由于分类数越大分类越精确,但计算量增大,而分类数过小会导致分类误差大,综合考虑选择簇数为 6。



注：(a)为  $k=22$ ，(b)为  $k=8$ 。

图 1 不同  $k$  值的聚类偏差图

Fig. 1 The different  $k$ -values clustering deviation chart

Step2 分类并融合分类数据。

R 型聚类法：①通过 R 型聚类将 15 个题目分为 6 类，第 1 类有 8 题，第 2 类有 3 题，第 3 类有 4 题，第 4 类有 5 题，第 5 类有 6 题，第 6 类有 7 题。

②根据分类，利用 Murphy 方法再次进行计算。

第一类、第二类的平均值为

$$\bar{m}_{RM1} = 0.2165 \quad 0.6890 \quad 0.0718 \quad 0.0227,$$

$$\bar{m}_{RM2} = 0.4346 \quad 0.2101 \quad 0.2706 \quad 0.0847.$$

再用组合规则式(1)对第一类融合 7 次、第二类融合 2 次，得到

$$\bar{m}_{RMZ1} = 0.000096 \quad 0.9999 \quad 0.000000012 \quad 0.0000000000014,$$

$$\bar{m}_{RMZ2} = 0.7343 \quad 0.0830 \quad 0.1773 \quad 0.0054.$$

整理新的分类数据，见表 1。

表 1 R 型聚类 Murphy 方法的新证据

Tab. 1 The new evidence of R-type clustering Murphy method

类别	4	3	2	1
第一类	0.000096	0.9999	0.000000012	0.0000000000014
第二类	0.7343	0.0830	0.1773	0.0054
第三类	0.0099	0.0990	0.6040	0.2871
第四类	0.0660	0.5413	0.3927	0
第五类	0.3267	0.6238	0.0264	0.0231
第六类	0.0924	0.2310	0.4819	0.1947

③利用 Jousselme 距离函数计算新的分类数据。

第一类的权重

$$\omega_{R11} = 0.1267, \omega_{R12} = 0.1259, \omega_{R13} = 0.1233,$$

$$\omega_{R14} = 0.1269, \omega_{R15} = 0.1172, \omega_{R16} = 0.1279,$$

$$\omega_{R17} = 0.1259, \omega_{R18} = 0.1263.$$

加权平均后得到

$$m_{RJ1} = 0.2171 \quad 0.6896 \quad 0.0708 \quad 0.0225.$$

将第一类加权平均后得到的结果融合 7 次，得到融合结果

$$m_{RJD1} = 0.000096 \quad 0.9999 \quad 0.000000012 \quad 0.0000000000013.$$

第二类的权重  $\omega_{R21} = 0.3344, \omega_{R22} = 0.3592, \omega_{R23} = 0.3064$ 。

加权平均后得到

$$m_{RJ2} = 0.4373 \quad 0.2167 \quad 0.2613 \quad 0.0847。$$

将第二类加权平均后得到的结果融合 2 次,得到融合结果

$$m_{RJD2} = 0.7450 \quad 0.0907 \quad 0.1589 \quad 0.0054。$$

整理新的分类数据,见表 2。

表 2 R 型聚类加权的新证据

Tab. 2 The new evidence of R-type cluster weighting

类别	4	3	2	1
第一类	0.000 096	0.999 9	0.000 000 012	0.000 000 000 001 3
第二类	0.745 0	0.090 7	0.158 9	0.005 4
第三类	0.009 9	0.099 0	0.604 0	0.287 1
第四类	0.066 0	0.541 3	0.392 7	0
第五类	0.326 7	0.623 8	0.026 4	0.023 1
第六类	0.092 4	0.231 0	0.481 9	0.194 7

Q 型聚类法:同样的步骤,通过 Q 型聚类将 15 个题目分为 6 类,利用 Murphy 方法得到新的分类数据,见表 3。

表 3 Q 型聚类 Murphy 方法的新证据

Tab. 3 The new evidence of Q-type clustering Murphy method

类别	4	3	2	1
第一类	0.000 4	0.999 4	0.000 1	0.000 000 003 8
第二类	0.020 4	0.979 4	0.000 006 8	0.000 000 10
第三类	0.006 8	0.071 1	0.770 4	0.151 7
第四类	0.534 7	0.227 7	0.049 5	0.188 1
第五类	0.432 4	0.326 7	0.211 2	0.029 7
第六类	0.336 7	0.075 9	0.551 2	0.036 2

利用 Joussemle 距离函数得到新的分类数据,见表 4。

表 4 Q 型聚类加权下新的证据

Tab. 4 The new evidence of Q-type cluster weighting

类别	4	3	2	1
第一类	0.000 4	0.999 4	0.000 074	0.000 000 004 3
第二类	0.020 8	0.979 1	0.000 006 9	0.000 000 099
第三类	0.006 8	0.071 1	0.770 4	0.151 7
第四类	0.534 7	0.227 7	0.049 5	0.188 1
第五类	0.432 4	0.326 7	0.211 2	0.029 7
第六类	0.336 7	0.075 9	0.551 2	0.036 2

由表 1—4 可见,Murphy 方法和 Joussemle 距离函数加权法得到的结果相差较小,表明两种聚类方法所得到的分类结果都很合理,因此,Murphy 方法可用于类内数据融合。

Step3 利用 Joussemle 距离函数确认类间权重。

①对 R 型分类后再将通过 Murphy 方法所得到的数据(表 1)加权,得到类间的权重

$$\omega_{R1} = 0.1294, \omega_{R2} = 0.1290, \omega_{R3} = 0.1612。$$

$$\omega_{R4} = 0.2005, \omega_{R5} = 0.1839, \omega_{R6} = 0.1960。$$

证据修正后的 BPA:

$$\bar{m}_{CW1} = 0.1977 \quad 0.4246 \quad 0.2983 \quad 0.0894。$$

②对 Q 型分类后再将通过 Murphy 方法所得到的数据(表 2)加权,得到类间的权重

$$w_{Q1} = 0.1508, w_{Q2} = 0.1556, w_{Q3} = 0.1334。$$

$$w_{Q4} = 0.1773, w_{Q5} = 0.2063, w_{Q6} = 0.1766。$$

证据修正后的 BPA

$$\bar{m}_{CW2} = 0.2476 \quad 0.4338 \quad 0.2525 \quad 0.0661。$$

Step4 结果的合成。

①将在 R 型聚类下修正后的 BPA 合成 5 次得

$$r_{R1} = 0.0066, r_{R2} = 0.8867, r_{R3} = 0.1066, r_{R4} = 0.000077。 \quad (13)$$

冲突系数  $k_R = 0.0066$ 。

②将在 Q 型聚类下修正后的 BPA 合成 5 次得

$$r_{Q1} = 0.0322, r_{Q2} = 0.9315, r_{Q3} = 0.0362, r_{Q4} = 0.000010。 \quad (14)$$

冲突系数  $k_Q = 0.0071$ 。

### 3.3 模型结果分析

通过数据模拟的结果可知,经典组合规则融合方法,与问卷数据相差过大,一定程度上背离了人的直觉判断;用于在校心理社会性发展研究的影响,与现实状况有差异。改进后的信息融合方法所得出的融合结果与问卷数据结果更加符合。

根据问卷数据以及所计算出的冲突系数,满足  $k_R < k_Q$ ,则 R 型聚类更适合作为分类方法,与实际相符合。

## 4 结论

本文针对大样本下证据冲突问题,基于空间向量聚类的思想,提出一种改进的证据融合方法。实证分析表明,本文提出的证据融合方法是科学可行的,以下为主要研究结论。

(1)构建的改进信息融合模型,在大样本下减少了传统 D-S 理论的计算量。针对 D-S 证据理论在处理证据冲突时的局限性,在聚类分析、Murphy 方法和 Jousselme 距离函数的基础上,提出的分类加权方法使证据修正权重更加合理,有更高的稳定性。

(2)提出的证据融合新方法,在大样本下解决证据冲突具有一定的实用性。案例分析表明,所提方法能够明显减少冲突证据的影响,且整合后的值更加集中和精确,提高了处理冲突证据时的逻辑性。

(3)基于证据理论的信息融合方法扩展了证据理论的范畴。该方法在复杂系统领域,能够解决样本容量大、样本数据尺度大、数据高维性复杂系统数据的驱动优化问题。

(4)为在校心理社会性发展研究提供一定参考。根据实证分析结果,在校中 0.66% 的学生心理社会性发展状况为优,88.67% 的学生状况为良,仍有 10% 左右的学生状况处于一般或不良,此数据可以作为教育工作者从事相关研究的参考。

### 参考文献:

- [1] 熊山. 基于证据理论的不完整信息融合研究[D]. 重庆:西南大学, 2022.
- [2] LIU Z. A new sine similarity measure based on evidence theory for conflict management[J]. Communications in Statistics-Theory and Methods, 2024: 1-17.
- [3] 边福音. 基于证据理论的信息融合方法研究[D]. 沈阳:东北大学, 2017.
- [4] KLEIN J W, GREENAWAY K H, BASTIAN B. Identity fusion is associated with outgroup trust and social exploration: Evidence for the fusion-secure base hypothesis[J]. British Journal of Social Psychology, 2024, 63(3): 1184-1206.
- [5] 徐选华, 黄丽. 基于复杂网络的大群体应急决策专家意见与信任信息融合方法及应用[J]. 数据分析与知识发现,

- 2022, 6(S1): 348-363.
- [6] ZHANG F X, ZHANG L M, LIU Z Y. An improved Dempster-Shafer evidence theory based on the Chebyshev distance and its application in rock burst prewarnings[J]. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 2024, 10(1): 04023055.
- [7] XIAO F Y. Generalized belief function in complex evidence theory[J]. Journal of Intelligent & Fuzzy Systems, 2020, 38(4): 3665-3673.
- [8] 刘晓光, 胡学钢. D-S 证据理论在决策支持系统中的应用[J]. 计算机系统应用, 2010, 19(10): 112-116.
- [9] 李明. 一种冲突证据的分类加权融合方法[J]. 西安邮电大学学报, 2023, 28(6): 74-81.
- [10] 董煜, 张友鹏. 基于聚类赋权的冲突证据组合方法[J]. 通信学报, 2023, 44(3): 157-163.
- [11] 邓展. 基于 D-S 证据理论的不确定信息决策方法研究[D]. 南京: 南京理工大学, 2022.
- [12] 孙永河, 张梦迪. 基于 R 型聚类的层次型 DEMATEL 分析方法[J]. 数学的实践与认识, 2019, 49(6): 42-51.
- [13] JOUSSELME A, GRENIER D, BOSSÉ É. A new distance between two bodies of evidence[J]. Information Fusion, 2001, 2(2): 91-101.
- [14] 张雅媛, 孙力帆, 郑国强. 基于改进 Jusselme 证据距离的多传感器决策融合方法[J]. 仪表技术与传感器, 2019(7): 82-87.
- [15] 王俊, 金勇进. 概率-非概率调查样本的整合推断问题研究: 核匹配方法[J]. 统计与信息论坛, 2024, 39(10): 3-12.
- [16] 刘峻源, 杨影. 新时代大学生心理健康现状及对策[J]. 吉林省教育学院学报, 2023, 39(12): 23-27.
- [17] 张嘉玲, 何铭颖. 当代大学生就业心态现状及对策研究[J]. 科教导刊, 2023(30): 142-144.
- [18] 戴硕. 基于模糊综合评判法的 XX 工厂节能技术改造项目投资效果评价研究[J]. 中国经贸导刊, 2024(8): 32-34.

## Non-probabilistic Evidence Fusion Method Based on Spatial Vector Clustering for Large-sample Scenarios

CAO Lixia, WANG Jiaxin, MAO Chunni

(School of Science Xi'an Technological University, Xi'an 710021, China)

**Abstract:** To address the issue of conflicting evidence fusion in large-sample scenarios, this study proposed a novel evidence fusion method based on spatial vector Jusselme distance clustering. The method comprehensively considered the supportive relationships between evidence and the uncertainty of the evidence itself, aiming to reduce the impact of evidence conflicts. Guided by a stepwise fusion approach, this study first calculated inter-evidence distances through cluster analysis for evidence classification. Murphy's method was then used to fuse evidence within the same category, and then the D-S combination rule was applied to the newly fused evidence for ordering and decision-making. Finally, the method was employed to study psychosocial development issues among university students. Empirical analysis demonstrated the effectiveness and feasibility of the proposed method, and the results can provide reference for predicting and assessing psychosocial development.

**Key words:** evidence conflict; large sample; vector space clustering; psychosocial development

【责任编辑 闫立华】