



# Mathematical Modeling and Empirical Analysis of Multi-source Teaching Evaluation Data

Yingcan Wang, Jianxin Zhao\*

Naval Submarine Academy, Qingdao, China

## Email address:

wycdsgswhh@163.com (Yingcan Wang), jxzhao@bit.edu.cn (Jianxin Zhao)

## To cite this article:

Yingcan Wang, Jianxin Zhao. (2025). Mathematical Modeling and Empirical Analysis of Multi-source Teaching Evaluation Data. *Science Innovation*, 13(2), 16-21. <https://doi.org/10.11648/j.si.20251302.12>

**Received:** 25 March 2025; **Accepted:** 16 April 2025; **Published:** 29 April 2025

**Abstract:** To address the inconsistency between student evaluation satisfaction data and supervisory evaluation conclusions in university teaching assessments, this study introduces structured data such as class size and course intensity to establish a "normalization-entropy weight-TOPSIS" evaluation methodology. Using a semester's teaching evaluation dataset from the Naval Submarine Academy as an empirical testbed, we implemented a comparative analysis of four distinct data preprocessing strategies: 1) RobustScaler enhanced with Sigmoid function transformation, 2) conventional RobustScaler, 3) Z-score standardization, and 4) Min-Max normalization. The experimental design rigorously evaluated each method's capacity to harmonize student feedback with expert evaluations through correlation analysis and distribution pattern verification. The empirical validation demonstrates that when applying the proposed evaluation framework to datasets processed by the Sigmoid-enhanced RobustScaler method, the resulting assessment scores achieved the highest correlation coefficient with supervisory ratings. Compared with traditional methods relying solely on student satisfaction rates, when applied within the proposed evaluation framework, this approach improved the correlation coefficient between the resulting assessment scores and supervisory ratings by 19.0%, which means generated assessment scores demonstrating superior alignment with supervisory ratings. The novel evaluation approach demonstrated superior performance in this practical application scenario, effectively resolving the contradiction inherent in single-source evaluation systems that exclusively utilize student feedback.

**Keywords:** Multi-source Teaching Evaluation Data, Teaching Evaluation Indicator System, Entropy Weight Method, TOPSIS Method

---

## 多源评教数据的数学建模与实证分析

王应灿, 赵建昕\*

海军潜艇学院, 青岛, 中国

## 邮箱

wycdsgswhh@163.com (王应灿), jxzhao@bit.edu.cn (赵建昕)

**摘要:** 针对高校评教中存在的学生评教满意度数据与督导评教数据之间结论不一致的问题, 通过引入班级规模、课程强度等其他的结构化数据, 建立了“归一化-熵权-TOPSIS”评价方法。以海军潜艇学院某学期的评教数据为例, 横向对比结合Sigmoid函数的RobustScaler方法、RobustScaler方法、Z-score标准化方法、最大最小化方法在数据集上的表现, 通过实例数据对评价方法进行了验证, 结果显示, 经过结合Sigmoid函数的RobustScaler方法处理的数据应用此评价方法得到的评教得分数据与督导评分的相关系数最大, 相较于较传统方法, 即只利用学生满意率评教的方式, 相关系数提升19.0%, 在此实例问题中表现较好, 有效解决了单一利用学生评教满意度数据开展评教的矛盾问题。

**关键词:** 多源评教数据, 教学评价指标体系, 熵权法, TOPSIS方法

---

## 1. 引言

当前高校教学评价体系中，学生主观评价与客观教学质量之间的系统性偏差已成为制约教育质量诊断精度的关键瓶颈。研究表明，学生因认知水平差异（如低年级知识储备不足）或评价标准模糊化（如将“课程难度”误解为教学缺陷），导致主观评分与专家评审结果的相关性不足[1]，这暴露了结果性评价数据（如成果评价、学生反馈），而忽视过程评价数据对教学质量评价影响的结构性缺陷。

为突破此局限，学界开始探索基于数学建模的客观评价方法：熵权法因其在指标权重分配中规避主观偏好的特性[2]，成为重构评价体系的理论突破口。然而，现有研究仍存在三大短板：其一，单一熵权法易受指标共线性干扰，难以兼顾评价结果的区分度（TOPSIS 方法优势）与非线性关系捕捉（神经网络优势）；其二，静态权重模型无法适应教学场景的动态变化，例如职业教育与通识课程对“实践能力”指标的需求权重差异可达0.35[3]；其三，数据集未能很好区分教师间、同教师不同课程间各教学指标的差异，导致不能精准找出学生主观评价与客观教学质量之间存在系统性偏差的原因，无法构建出平衡的教学质量评价体系[4]。

为此，本文以“教师-课程”绑定后的信息为数据集，对教员及其授课题目进行分析，提出“归一化-熵权-TOPSIS”三级融合框架：首先，利用改进熵权法（利用 Sigmoid函数预处理数据抑制共线性）生成指标权重；其次，通过 TOPSIS方法量化教学水平与理想解的贴近度，进而得到教学质量评分；最后，对比多种处理数据方式，得到最适合数据集的评价指标体系。

## 2. 数据特征分析与预处理

### 2.1. 评教数据特征分析

评教结果涉及诸多指标，其中既有结构化的量化数据，也有非结构化的文本数据。对结构化数据，要对其做标准化和归一化处理；对非结构化数据，要通过数据处理的方式将数据结构化。

### 2.2. 评教数据质量评估与清洗

#### 2.2.1. 数据信度评估

为评估数据的一致性和稳定性，确保测量工具的可靠性，可以对数据进行重测信度和内部一致性信度检验。

#### 2.2.2. 数据效度评估

为评估数据准确性，即测量工具是否真实反映目标概念，对数据进行内容效度 CVI 和结构效度检验。

#### 2.2.3. 数据正态性评估

为评估检验数据是否符合正态分布，确保参数统计方法的有效性，对数据进行正态性评估。

### 2.3. 评教数据的标准化与归一化处理

评教数据的标准化方式有很多，比较常见且有效的有最大最小法、Z-score标准化方法、RobustScaler标准化方法以及通过函数压缩映射的方法。

#### 2.3.1. 最大最小化方法

对数据分布比较分散的数据集，使用最大最小化方法将数据线性映射到指定区间，保留原始分布。数据处理公式为

$$X'_{ij} = \frac{X_{ij} - \min\{X_{\square j}\}}{\max\{X_{\square j}\} - \min\{X_{\square j}\}} \quad (1)$$

其中， $i$  表示第  $i$  条数据， $j$  表示第  $j$  个指标。

#### 2.3.2. Z-score标准化方法

Z-Score 标准化（标准差标准化）是一种将数据转换为均值为 0、标准差为 1 的标准正态分布数据处理方法。

与最大最小化方法不同，Z-Score 通过数据的统计特性（均值和方差）进行缩放，对异常值更加鲁棒，适用于大多数数值型特征。数据处理公式为

$$X'_{ij} = \frac{X_{ij} - \text{mean}\{X_{\square j}\}}{\text{STD}\{X_{\square j}\}} \quad (2)$$

其中， $i$  表示第  $i$  条数据， $j$  表示第  $j$  个指标。

#### 2.3.3. RobustScaler标准化方法

对于具有极端值的数据集，使用样本均值作为中心点处理数据，会出现异常值影响结果准确性的弊端。若同时数据集还不具备正态性，就更不适合用 Z-Score 标准化方法。RobustScaler 是一种基于中位数和四分位数的数据标准化方法，专为处理含异常值或非正态分布数据设计。与 Z-Score 和最大最小化方法不同，它通过消除极端值的影响，使标准化过程更具鲁棒性。数据处理公式为

$$X'_{ij} = \frac{X_{ij} - \text{median}\{X_{\square j}\}}{TQ\{X_{\square j}\} - FQ\{X_{\square j}\}} \quad (3)$$

其中， $TQ\{X_{\square j}\}$  表示第三四分位数， $FQ\{X_{\square j}\}$  表示第一四分位数， $\text{median}\{X_{\square j}\}$  表示中位数。

#### 2.3.4. Sigmoid函数数据处理法

评教数据中往往会出现一些对指标增量不敏感的指标，比如授课规模，随授课规模增大到一定程度，其对课程的影响不大。为突出这一特点，将数据用 Sigmoid 函数压缩。数据处理公式为

$$X'_{ij} = \frac{1}{1 + e^{\frac{\text{mean}\{X_{\square j}\} - X_{ij}}{\text{STD}\{X_{\square j}\}}}} \quad (4)$$

其中， $\text{mean}\{X_{\square j}\}$  表示样本均值， $\text{STD}\{X_{\square j}\}$  表示样本标准差。

### 3. 基于熵权法和TOPSIS方法的数学建模方法

#### 3.1. 熵权法和TOPSIS的方法适应度

##### 3.1.1. 熵权法适应度

(1) 客观性强: 客观赋权, 避免人为干扰

(a) 数据驱动

权重完全由数据离散程度决定, 避免了根据客观数据进行主观打分的不稳定性[5]。例如, 若某指标(如“授课人数”)数据差异显著(标准差大), 则熵值低、权重高, 反映其对评价结果影响更大; 而若某指标(如“课程时长”)数据差异不显著(标准差小), 则熵值低、权重高, 反映其对评价结果影响小。而非人为决定某指标的重要度。

(b) 抗偏性强

若存在“老好人”式统一高分或低分(数据离散度低), 熵权法会自动降低该指标权重, 减少无效评分对结果的影响[6]。

(2) 适应性强: 动态适应不同数据集

权重随数据分布变化自动调整[7]。例如, “课程时长”数据差异较小(熵值高), 则权重降低; 而“授课人数”数据差异大(熵值低), 权重提升, 体现评价重点动态迁移。

(3) 理论严谨: 数学严谨, 兼容性强

标准化处理消除量纲差异[8](如将0-100分和0-10次互动统一为0-1区间), 支持多类型指标(效益型/成本型)直接计算。

(4) 可解释性高: 结果可解释性高

通过权重值直观展示各指标重要性[9]。例如, 若“课堂互动”权重为0.4, 说明其对教师综合评价的贡献占比达40%, 便于管理者针对性优化教学策略。

##### 3.1.2. TOPSIS方法的突出优势

(1) 直观反映方案优劣

几何意义明确: 通过计算与正/负理想解的欧氏距离, 将复杂多指标问题转化为空间中的“接近度”排序[10]。例如, 教师A的接近度0.8表示其综合表现距离“最优教师”更近, 距离“最差教师”更远。

全局对比: 避免局部最优陷阱。例如, 某教师“教学态度”满分但“内容深度”极低, TOPSIS会因远离正理想解而排名靠后, 防止单一指标过高掩盖短板。

(2) 灵活处理矛盾指标

支持效益型(越大越好, 如学生成绩)与成本型(越小越好, 如作业批改延迟天数)指标共存[11]。例如, 在评教中自动区分“互动频次(效益型)”和“课程难度(需适中, 需自定义理想解)”。

通过自定义正/负理想解适应复杂需求。例如, 若“学生成绩提升率”的理想解不是最大值而是预设目标值(如20%), TOPSIS可灵活调整计算逻辑。

(3) 结果稳健, 敏感性可控

接近度的取值范围为[0,1], 排序结果对极端值不敏感[12]。例如, 某教师因一项指标异常高分, 但其综合接近度可能仍低于均衡型教师, 避免排名失真。

##### 3.1.3. 熵权-TOPSIS融合的协同优势

(1) “客观赋权+科学排序”双保险

熵权法解决“指标重要性如何量化”, TOPSIS解决“方案优劣如何对比”, 两者互补[13]。

(2) 支持动态敏感性分析

通过调整熵权法输入数据(如增加最新学期评分), 实时更新权重与排名。

#### 3.2. 评教指标权重分配模型

根据熵权法[14]的理论, 设计三个步骤, 计算得到每个指标对应的权重。

(1) 生成指标出现概率

根据处理后的数据  $X'_{ij} (i=1,2,...,m; j=1,2,...,n)$ , 通过计算每条数据的每个指标, 与总数据的这一指标和的比值, 生成第  $i$  条数据(共  $n$  条)的第  $j$  个指标(共  $m$  个)

$$P_{ij} = \frac{X'_{ij}}{\sum_{i=1}^n X'_{ij}} \quad \text{出现的概率}$$

(2) 生成各指标信息熵

通过每条数据中各指标的出现概率, 整合得出  $m$  个指标的信息熵

$$E_j = \frac{\sum_{i=1}^n P_{ij} \ln P_{ij}}{\ln n} \quad (5)$$

(3) 生成各指标权重

利用生成的各指标信息熵, 计算各指标权重

$$W_j = \frac{1 - E_j}{m - \sum_{i=1}^m E_j} \quad (6)$$

#### 3.3. 依托权重分配的评教得分模型

应用熵权法得到的每个指标的权重, 通过TOPSIS方法给出最后评分[15]。根据TOPSIS方法的理论, 设计三个步骤, 得到每条数据的最终评分。

(1) 生成正、负理想解

根据处理后的数据, 找到最优点和最劣点分别是

$$X_{\max} = (\max\{X'_{1j}\}, \dots, \max\{X'_{mj}\}) \quad (7)$$

$$X_{\min} = (\min\{X'_{1j}\}, \dots, \min\{X'_{mj}\}) \quad (8)$$

(2) 生成与正、负理想解间距离

根据各指标的权重, 分别计算每条数据与最优点和最劣点之间的距离, 作为其与正、负理想解间距离, 即每条数据与理想解的距离可如下求解。

与正理想解间距离为:

$$X_i^+ = \sqrt{W_1(X'_{i1} - \max\{X'_{1j}\})^2 + \dots + W_m(X'_{im} - \max\{X'_{mj}\})^2} \quad (9)$$

与负理想解间距离为:

$$X_i^- = \sqrt{W_1(X'_{i1} - \min\{X'_{1j}\})^2 + \dots + W_m(X'_{im} - \min\{X'_{mj}\})^2} \quad (10)$$

(3) 通过与正、负理想解间距离生成最终评分

通过一个函数计算第  $i$  条数据的最终评分  $S_i$ :



$$S_i = \frac{X_i^-}{X_i^+ + X_i^-} \quad (11)$$

## 4. 评教数据的实证分析与模型验证

### 4.1. 数据来源与结构分析

数据源自海军潜艇学院某学期共276条的评教数据。

将教师-课程绑定作为一条数据，梳理出每条数据的客观指标和主观指标。其中，客观指标有：课程名称、授课学时、授课规模、课程强度（授课学时/课程天数）、开课单位、学生层次、教师职称；主观指标有：学生满意率、教学督导分数。根据问题描述，以上因素均可以影响教学质量的评价。

将指标按其数据是否结构化，分为结构化指标和非结构化指标，其中，非结构化指标有：课程名称、开课单位、学生层次、教师职称；结构化指标有：授课学时、授课规模、课程强度、学生满意率、教学督导分数。

本文仅考虑结构化指标对结果的影响，根据数据的含义，最能反应教学质量的分数是教学督导分数。因此为了获得更加准确的评价系统，将教学督导分数作为监督集、其他结构化指标作为评价指标，构建评价指标体系，通过训练得到更准确的评教方式。

### 4.2. 数据质量评估

由于实际评教数据来源于真实教学过程，其真实有效性毋庸置疑，只需对评教数据进行正态性进行检验，判断数据的质量。

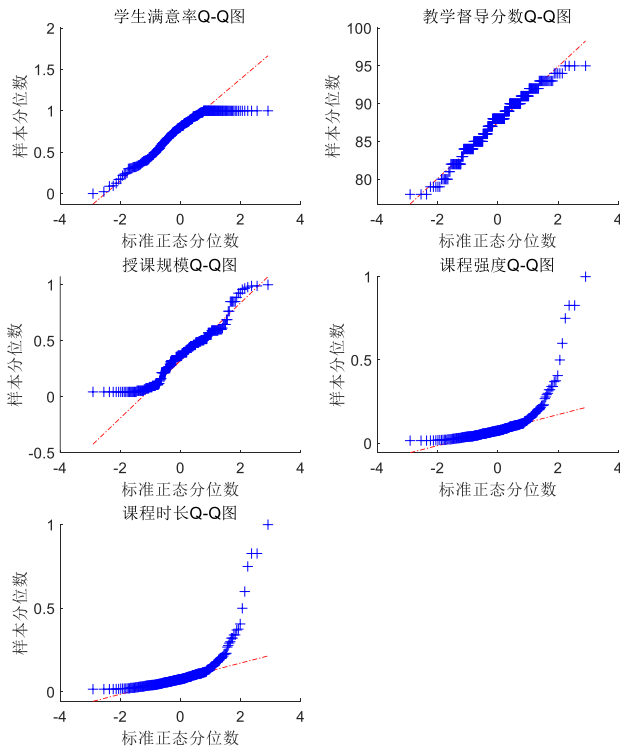


图1 正态分布检验Q-Q图。

通过Kolmogorov-Smirnov检验，发现学生满意率、授课学时、授课规模、课程强度均不服从正态分布，如图2。

分开来看，教学督导评分服从正态分布，但数据存在极端值现象；学生满意率上限为1，这将可能取得更高值的数据压缩在了1这一点，导致在很多数据集中在1这一点，使整个数据集存在很多极端值，进而失去了正态性；而课程时长、课程强度、授课规模是人为构造的数据分布，其本来的分布特征因排课等人为统筹设计而被改变，自然不符合正态性。

为使数据具有正态性，对教学督导评分和学生满意率，理论上可以使用RobustScaler数据标准化方法减小极端值的影响；对授课规模和课程时长因素，二者对督导评分指标的存在边际效用递减现象，尝试对授课规模和课程时长使用Sigmoid函数处理。

### 4.3. 数据的描述性统计

因本文仅考虑结构化指标对结果的影响，故为研究课程时长、班级人数等客观因素与学员评教优秀率、督导评分等主观因素之间的关系，对评教数据进行描述性统计。统计数据显示，班级优秀率分数最高为100分，最低为0分。标准差为24.71，差异巨大；督导评分最高为95，最低为78，标准差为3.74。这表明主观分数中督导评分具有更小的波动性。下面对每个结构化指标进行具体分析：

#### 4.3.1. 结构化指标频数分析

各结构化指标对应的频数如图1：

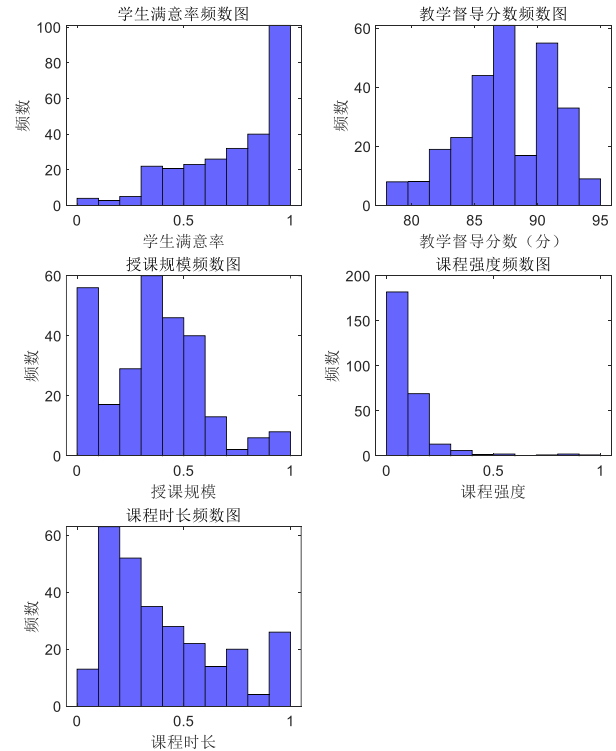


图2 结构化指标频数图。

通过频数图，能更直观看到各指标的分布有自己的特征，分布最密集的地方出现在数据集的不同位置。而欲判断各指标是否能同时取到最密集的地方，进而判断各指标间的关系，需要对各指标的相关性进行研究。

#### 4.3.2. 结构化指标相关性分析

对各结构化指标之间的相关性进行分析，如表1。

表1 各指标相关性系数表。

	学生满意率	督导分数	授课规模	课程强度	课程时长
1	1	0.195	-0.419	-0.072	0.013
2	0.195	1	0.058	-0.055	0.1008
3	-0.419	0.058	1	-0.052	0.1192
4	-0.072	-0.055	-0.052	1	-0.4827
5	0.013	0.1008	0.1192	-0.4827	1

通过相关性系数表,发现督导分数与授课规模存在明显的负相关性,认为授课规模对督导分数的影响显著;同时学生满意度与督导分数相关性明显不足,认为学生满意度与督导分数关系不大。

据此,可以得到结论:

- a. 授课规模与教学质量相关性大;
- b. 单纯依靠学生满意度不能准确衡量教学质量。

因此,构建模型的目的是提高模型最终评价分数与督导分数的相关性,提高评价方法的合理性。同时,用授课规模所占模型权重的大小验证模型的合理性,保证模型结果与原始数据的相关性一致。

4.4. 模型验证

为解决只靠学生评价对教师授课质量排序的问题,应用之前所说的模型方法,在评教实例数据集上,尝试对教学评价体系进行研究,以找到一个适合此评教实例。

之后,对比得分与督导评分之间的相关性,衡量得分能反应真实情况的程度。

4.4.1. 模型实验组设置

根据实例数据分析,尝试使用RobustScaler标准化方法、最大最小化方法、Z-score标准化方法对数据进行处理,并对比模型结果。同时,由于授课规模和课程时长对督导评分指标的存在边际效用递减现象,尝试对授课规模和课程时长使用Sigmoid函数处理。

根据相关性分析,发现学生满意率、课程强度和课程时长与督导分数正相关,督导分数与授课规模负相关。由此可知,正向指标为学生满意率、课程强度和课程时长,负向指标为授课规模。

据此,构建教学评价指标体系模型,将方法如下标号。

- 方法1: 只使用RobustScaler方法处理数据;
- 方法2: 只使用Z-score标准化方法处理数据;
- 方法3: 只使用最大最小化方法处理数据;
- 方法4: RobustScaler方法和Sigmoid函数处理数据。

4.4.2. 模型结果

根据熵权法,得到评分各指标的权重如表2:

表2 各方法对应权重表。

	方法1	方法2	方法3	方法4
授课规模	0.317	0.317	0.4833	0.313
学生满意率	0.427	0.427	0.084	0.536
课程强度	0.146	0.146	0.133	0.013
课程时长	0.110	0.110	0.300	0.138

基于此权重和各方法对应处理后的数据集,应用TOPSIS方法,得到最终得分。各方法最终得分与督导分数之间的相关性如表3:

表3 各方法最终评分与督导评分相关系数表。

	督导分数	方法1	方法2	方法3	方法4
督导分数	1	0.208	0.131	-0.134	0.232
方法1	0.208	1	0.891	-0.035	0.678
方法2	0.131	0.891	1	0.121	0.382
方法3	-0.134	-0.035	0.121	1	-0.173
方法4	0.232	0.678	0.382	-0.173	1

5. 结论

根据模型,对比RobustScaler方法、Z-score标准化方法和最大最小值方法处理数据的结果,得到以下结论:

- a. 模型有效提高了评价方法的合理性。根据模型结果,实例数据集更适合使用通过使用RobustScaler方法和Sigmoid函数处理,最后得分与督导分数的相关性比只使用RobustScaler方法时提高了11.5%,比学生满意率与督导分数之间的相关性提高了19.0%,认为在衡量教学质量方面,此模型比学生满意率的表现有较大提升,给出了更有效的评价方法。
- b. 模型符合原始数据集的数据特征,有较强可信度。根据模型结果,使用了RobustScaler方法和Sigmoid函数处理的模型,其授课规模所占比重大,符合原始数据中督导分数和授课规模相关性大这一特点。
- c. 模型兼顾学生评价和专家评价给出最终教学评价得分。根据模型结果,使用了RobustScaler方法和Sigmoid函数处理的模型,其学生满意率所占权重较大,说明模型在兼顾学生满意率的基础上,提高了评价准确度,符合教学评价应该结合学生评价和专家评价两部分的朴素认知。

致谢

本文为基金项目《多源评教数据的数学建模与实证分析研究》的阶段性成果之一。

参考文献

[1] 张珊珊. 《基于数据分析的高校教育教学效果评价与优化研究》. 黑龙江科学 16, 期 01 (2025年): 58-60+65.

[2] 李虹毓, 吴晓琴, 赵艳霞, 梁少英, 李运和汪凯. 《高职学生在线学习行为评价指标体系的构建》. 科教文汇, 2024年. CNKI.

[3] 刘剑娥. 《中职计算机专业增值性教学评价体系的构建策略》. 信息系统工程, 期 01 (2025年): 161-64.

[4] 熊晶晶, 和雷培梁. 《高校教师教学能力评价体系建构——基于全国103所高校的实证研究》. 泉州师范学院学报, 2024年. CNKI.

[5] 龚宇霞, 吕越. 基于熵权法的绿色建筑节能潜力评价方法研究 [J]. 绿色建筑, 2025, 17(02): 15-18+24.

[6] 权家乐, 惠永昌, 汤恒仁, 等. 基于相对熵组合权重的战斗机作战能力评估 [J]. 工程数学学报, 2025, 42(02): 379-387.

- [7] 方晓峰, 王亚林, 杨喆. 基于层次熵权法的军校研究生综合素质评价研究 [J]. 高教学刊, 2025, 11(06): 100-103.
- [8] 李根, 白祥. 基于熵权法的克拉玛依市乌尔禾特色小镇综合效益评价 [J]. 中南农业科技, 2025, 46(02): 152-156.
- [9] 王军, 杨庭, 王溪, 等. 基于熵权法的电力运检装备质量评价方法 [J]. 微型电脑应用, 2025, 41(02): 240-243.
- [10] 孙妮妮, 孙怀波, 李俊林. 基于AHP-TOPSIS模型的“数学建模”课程思政的成效分析 [J]. 湖北理工学院学报, 2025, 41(02): 73-77.
- [11] 李木洲, 曾思鑫. 新高考科目改革的区域评价差异与推进策略——基于熵权TOPSIS模型的分析 [J]. 中国高教研究, 2025, (03): 59-67.  
<https://doi.org/10.16298/j.cnki.1004-3667.2025.03.08>
- [12] 李洪英, 卢凯莉, 王亚楠. 基于概率语言术语集改进TOPSIS法的社区应急治理韧性评价研究 [J/OL]. 数学的实践与认识, 1-12 [2025-04-15].  
<https://doi.org/10.20266/j.math.24-0713>
- [13] 王芷怡, 王贵强, 冯国会, 等. 基于熵权-TOPSIS法供热系统运行效果综合评价与案例分析[J]. 建筑节能(中英文), 2025, 53(03): 76-82.
- [14] 程启月. 评测指标权重确定的结构熵权法 [J]. 系统工程理论与实践, 2010, 30(07): 1225-1228.
- [15] 陈雷, 王延章. 基于熵权系数与TOPSIS集成评价决策方法的研究 [J]. 控制与决策, 2003, (04): 456-459.