

贝叶斯修正的 AHP—熵值法在教育测评中的应用*

沈光辉^{1,2}, 叶彤¹, 孙小坚³, 范涌峰⁴

(1. 江苏大学教师教育学院, 镇江 212013; 2. 江苏大学教育测评循证研究中心, 镇江 212013;
3. 西南大学数学与统计学院, 重庆 400715; 4. 西南大学教育学部, 重庆 400715)

摘要: 指标权重确定方法在教育测评中起到关键作用, 将直接影响测评结果。层次分析法 (Analytic Hierarchy Process, AHP) 和熵值法是教育测评中最为常用的权重确定方法, 但分别存在主观性较强和样本依赖性强的问题。此外, 常用的主客观综合赋权法忽视了专家水平对指标权重的影响。而贝叶斯修正的 AHP—熵值法的权重计算方法则通过利用专家自身权重对 AHP 法所得权重进行贝叶斯修正, 然后与熵值法结合以得到组合权重。研究在综合分析贝叶斯修正的 AHP—熵值法的优势及计算方法基础上, 以内地西藏班教育质量测评为例, 通过对多种方法进行比较, 发现该方法计算出的结果可行、可操作且有效。

关键词: 贝叶斯; 层次分析法; 熵值法; 教育测评; 组合赋权

中图分类号: B841.2

文献标识码: A

文章编号: 1003–5184(2023)03–0269–09

2018 年 9 月习近平总书记在全国教育大会上强调: 要扭转不科学的教育评价导向, 克服“五唯”, 从根本上解决教育评价指挥棒问题。2020 年 10 月, 中共中央国务院印发《深化新时代教育评价改革总体方案》, 明确提出要改进结果评价, 强化过程评价, 探索增值评价, 健全综合评价。教育评价是新时代教育改革和发展最关键的领域, 也是近二十年我国教育改革和发展的热点和难点问题。建立科学有效的教育测评模型是新时代教育评价改革的重要议题, 是教育评价功能得以发挥的重要保证。为此, 近年来不少学者围绕我国各级各类教育质量、教育管理等方面问题建立了相关教育测评模型, 并形成了比较成熟的教育测评模型构建范式。其中教育测评模型构建的一般步骤为: 首先确定教育测评模型构建的价值取向, 然后明确教育测评对象的操作定义, 接着构建教育测评指标体系, 最后确定教育测评指标权重并生成教育测评模型。其中指标权重的确定是测评模型的关键一环, 将直接影响测评结果 (范涌峰, 宋乃庆, 2019)。在教育测评模型中常用的权重计算方法有主观赋权法 (AHP 法、德尔菲法等)、客观赋权法 (熵值法、主成分分析法、因子分析法、变异系数法等) 和主客观综合赋权法 (综合运用

主观赋权法和客观赋权法)。这些方法各有优势, 但是在使用中仍存在一些不足, 而贝叶斯修正的 AHP—熵值法对于弥补权重计算中的这些不足有重要价值, 在教育测评中具有重要的应用前景。下文将围绕当前教育测评中常用指标权重计算方法及主要局限、贝叶斯修正的 AHP—熵值法权重计算、贝叶斯修正的 AHP—熵值法的应用以及多种方法比较展开。需要说明的是文中提到的 AHP、加权 AHP、贝叶斯修正的 AHP、AHP—熵值法、贝叶斯修正的 AHP—熵值法中涉及的 AHP 均为层次分析法。

1 当前教育测评中常用指标权重计算方法及主要局限

AHP 是教育测评中最常用的主观赋权法之一。AHP 的基本原理就是把所要研究的复杂问题看作一个大系统, 通过对系统的多个因素的分析, 划出各因素间相互联系的有序层次; 再请专家对每一层次各因素进行较为客观的判断后, 相应给出相对重要性的定量表示; 进而建立数学模型, 计算出每一层次全部因素的相对重要性的权值, 并加以排序; 最后根据排序结果进行规划决策和选择解决问题的措施 (丁家玲, 叶金华, 2003)。AHP 法计算权重取决于专家的判断, 专家的权威性直接影响到权重大小, 如

* 基金项目: 国家社会科学基金重大项目“国家基础教育质量监测与评价体系研究” (19ZDA359), 2022 年度教育部人文社会科学研究青年项目“中小学复杂问题解决能力及其影响机制的追踪研究” (22YJC880065), 2023 年度教育部人文社会科学研究西部和边疆地区项目“中小学教师评价素养发展指数构建及应用研究” (23XJC880004), 2020 年度重庆市博士后特别资助项目“STEAM 教育的学习动机测评和 CAT 系统的实现”。

果选择的专家比较权威,计算出的权重结果将比较接近实际情况,该方法易于操作,调研专家人数不多,主要通过问卷的形式进行数据调研,然后通过 R 语言、Matlab、Yaahp 等专业软件直接计算结果。其中重要的一步就是构造专家打分判断矩阵,判断矩阵既可以计算指标权重,也可以对专家打分的一致性进行检验,保证了各个专家打分的逻辑性。教育研究者普遍使用该方法计算指标权重,如内地民族班教育质量测评模型(宋乃庆等,2019),采用 AHP 法构造专家判断矩阵,对少量未通过一致性检验的判断矩阵进行科学修正,使用 R 语言计算出最终指标权重;大学生就业创业教育评价指标体系(刘海滨,杨颖秀,陈雷,2012),组建了 50 人专家团队,以专家组的形式确定出统一的判断矩阵,通过使用 Matlab 计算出最终指标权重;STEM 教育质量评价指标体系(李艳燕等,2020),通过对 9 位 STEM 教学经验丰富、拥有丰硕研究成果的专家进行调研,构造判断矩阵,通过 Yaahp 软件使用 AHP 法计算出权重,最后对每位专家的指标权重取均值得到最终权重。然而,AHP 法在确定权重时存在以下不足:虽然用到了数学公式进行计算,但评价结果依然有较强的主观性;通常只计算一、二级指标权重,观测点的权重由于不易获得而常取等值,主要原因在于观测点的判断矩阵较多,专家填答起来比较困难。

在客观赋权法上,有不少教育研究者通过熵值法进行指标权重计算。熵值法计算指标权重比较客观,通过数据驱动计算指标权重。熵值法主要利用信息论中的信息熵进行权重计算,信息熵是系统无序程度的度量,信息是系统有序程度的度量,二者绝对值相等,符号相反,信息熵越小则信息量越大、指标权重越大(郭显光,1998)。如应用型高校实践教学质量评价指标体系(郑谦等,2016)、全日制专业硕士教育评价指标体系(胡恩华,陈沛然,顾桂芳,2015)、中国高等教育省域财政压力指数(罗志敏,马浚锋,2020)等均通过熵值法计算指标权重。然而,熵值法忽略了指标本身的重要程度,有时确定的指标权重与预期相差很远,另外受样本的影响很大,计算结果不稳定(郭显光,1994)。

此外,主客观综合赋权法在测评模型中受到广泛推广,主客观结合法通常将一种主观权重计算方法与一种客观权重计算方法进行组合求综合权重。它克服了主、客观两类评价方法的缺陷,同时吸收了两类评价方法的优点,对评价对象作了更为全面、合

理的评价(曾宪报,1997)。组合评价方法可分为评价权重的组合和评价结论的组合两个方向(彭张林等,2016),其中评价权重的组合又分为区分主客观权重的组合和不区分主客观权重的组合两类,评价结论的组合又分为序值组合和评价值组合两类。目前常用的区分主客观权重的评价权重的组合方法有融入德尔菲法的因子分析(张乔木,2017)、变异系数-AHP 组合赋权(张玉玲,迟国泰,祝志川,2011)、AHP 法与熵值法相结合(沈光辉,高鑫,宋乃庆,2022)等,这些方法都很好的克服了评价结果过于主观或者过于客观的问题,但是这些方法忽视了专家水平对指标权重的影响。不同专家会因职称、专业领域、对评价选题的熟悉程度等方面不同而存在水平差异,由此得出的权重差异较大,所以取平均的做法并不合适。

为缓解以上方法的不足,本研究尝试在 AHP-熵值法的基础上将专家自身的权重信息纳入考虑,换言之,研究采用贝叶斯修正的 AHP-熵值法来计算一、二级指标权重及观测点权重,以获得更加科学、合理的测评结果。该方法首先分别用贝叶斯修正的 AHP 法和熵值法计算权重,然后将两者进行组合,分别采用线性加权和最小二乘法对主、客观方法的权重进行分配,最后通过比较总残差平方和选取最接近多种原始方法结果的组合方法。

2 贝叶斯修正的 AHP-熵值法权重计算

2.1 贝叶斯修正的 AHP 法权重计算

贝叶斯修正的 AHP 法在 AHP 法的基础上加入了专家自身权重,利用贝叶斯思想对专家权重进行修正,然后将修正后的专家权重与 AHP 法所得的指标权重进行加权求出最终结果(江浩,江兵,2010)。贝叶斯修正的 AHP 法计算权重具体步骤如下:

(1) 计算专家初始权重

根据经验对 n 个专家 (p_1, p_2, \dots, p_n) 给出一个合理的先验分值 $S(p_1), S(p_2), \dots, S(p_n)$, 对先验分值进行标准化处理得到初始专家权重 $S'(p_1), S'(p_2), \dots, S'(p_n)$ 。这里的先验分值可以通过专家职称、判断依据、对研究内容的熟悉程度等方面获得,如专家职称教授、副教授、讲师分别对应分值 1、0.8、0.6;专家对指标重要程度判断依据由实践经验、理论分析、国内外同行了解、直观感觉四部分组成,实践经验方面填写大、中、小分别对应分值 0.5、0.4、0.3,理论分析方面填写大、中、小分别对应分值 0.3、0.2、0.1,国内外同行了解填写大、中、小分值均

为 0.1, 直观感觉填写大、中、小分值均为 0.1; 专家对研究内容的熟悉程度非常熟悉、比较熟悉、一般、比较不熟悉、非常不熟悉分别对应分值 1、0.8、0.6、0.4、0.2 (Shen, 2023; 沈光辉, 2023)。

(2) 计算各级指标的初步加权 AHP 权重

n 个专家对 m 个指标 (f_1, f_2, \dots, f_m) 进行两两比较, 得出判断矩阵 A , 对通过一致性检验的判断矩阵进行归一化处理得到指标权重 $S'(f_j/p_i), i = 1, 2, \dots, n; j = 1, 2, \dots, m$ 。

由全概率公式可以求得指标的初步加权 AHP 权重 $S'(f_j)$:

$$S'(f_j) = \sum_{i=1}^n S'(p_i) S'(f_j/p_i), i = 1, 2, \dots, n; j = 1, 2, \dots, m \quad (1)$$

(3) 修正专家权重和加权 AHP 权重

根据贝叶斯公式对专家权重进行修正, 得到后验专家权重 $S'(p_i/f_j)$

$$S'(p_i/f_j) = \frac{S'(p_i) S'(f_j/p_i)}{S'(f_j)}, i = 1, 2, \dots, n; j = 1, 2, \dots, m \quad (2)$$

根据后验专家权重 $S'(p_i/f_j)$ 以及 AHP 法计算的指标权重 $S'(f_j/p_i)$ 计算指标的后验加权 AHP 权重, 即贝叶斯修正的 AHP 法所得权重 $U(f_j)$:

$$U(f_j) = \sum_{i=1}^n Z(p_i/f_j) Z(f_j/p_i), i = 1, 2, \dots, n; j = 1, 2, \dots, m \quad (3)$$

2.2 熵值法权重计算

(1) 原始数据矩阵归一化

q 个学生在 m 个指标上的得分矩阵为 \bar{A} , \bar{a}_{ij} 表示第 i 个学生在第 j 个指标上的得分, 其中 $i = 1, 2, \dots, q; j = 1, 2, \dots, m$ 。

对得分矩阵进行规范化处理得到标准矩阵 A' , A' 中各项元素:

$$a'_{ij} = \frac{\bar{a}_{ij} - \min(\bar{a}_{ij})}{\max(\bar{a}_{ij}) - \min(\bar{a}_{ij})} \quad (\text{正向指标}) \text{ 或者}$$

$$a'_{ij} = \frac{\min(\bar{a}_{ij}) - \bar{a}_{ij}}{\max(\bar{a}_{ij}) - \min(\bar{a}_{ij})} \quad (\text{负向指标})$$

(2) 计算熵值

计算第 j 项指标下第 i 个学生提供信息的特征比重为 p_{ij}

$$p_{ij} = \frac{a'_{ij}}{\sum_{i=1}^q a'_{ij}} \quad (4)$$

指标 j 的熵值 e_j

$$e_j = -k \sum_{i=1}^q p_{ij} \ln p_{ij} \quad (5)$$

其中 k 与样本量 q 有关。当指标各个观测值都取等时, 变异程度最小, 信息量最小, 熵值最大为 1, 即 $e_j = 1, p_{ij} = \frac{1}{q}$, 此时 $k = \frac{1}{\ln q}$ 。将 $k = \frac{1}{\ln q}$ 带入公式(5), 则

$$e_j = -\frac{1}{\ln q} \sum_{i=1}^q p_{ij} \ln p_{ij} \quad (6)$$

(3) 计算熵权

各指标的客观权重 $V = (v_1, v_2, \dots, v_m)$, 其中 v_j

$$v_j = \frac{1 - e_j}{\sum_{j=1}^m (1 - e_j)} \quad (7)$$

2.3 贝叶斯修正的 AHP—熵值法权重计算

贝叶斯修正的 AHP—熵值法综合考虑了贝叶斯修正的 AHP 法和熵值法的权重, 确定主客观方法权重比例的方式通常有两种: 第一主观判断两类方法权重, 如根据专家意见, 将两者按照一定的比例进行线性加权, 即线性加权组合评价方法; 第二客观判断两类方法权重, 如基于最小二乘的思想确定权重, 即最小二乘意义下的组合评价方法。

(1) 线性加权组合评价方法

$$O = \alpha U + (1 - \alpha) V \quad (8)$$

其中 U 表示贝叶斯修正的 AHP 法计算得到的指标权重, V 表示熵值法计算得到的指标权重, α 为比例系数, $\alpha \in (0, 1)$, 在赋权时需对主、客观权重所占比例进行权衡, 根据专家建议确定 (李芳等, 2020)。

(2) 最小二乘意义下的组合评价方法

采用毛定祥 (2002) 提出的一种最小二乘意义下主客观评价一致的组合评价方法, 具体如下。

令贝叶斯修正的 AHP 法计算得到的指标权重为

$$U = (u_1, u_2, \dots, u_m)^T$$

熵值法计算得到的指标权重为

$$V = (v_1, v_2, \dots, v_m)^T$$

最小二乘意义下的组合评价指标权重为

$$W = (w_1, w_2, \dots, w_m)^T$$

q 个测评对象在 m 个指标上标准化后的数据阵为 $Z = (Z_{ij})_{q \times m}$, Z 矩阵也是上文熵值法中 \bar{A} 矩阵标准化处理的结果。第 i 个评价对象的评价值为

$$g_i = \sum_{j=1}^m w_j z_{ij}, (i = 1, 2, \dots, q) \quad (9)$$

对所有评价对象的所有指标而言,主客观赋权下的评价值的偏差应当越小越好,为此建立如下最小二乘法优化组合评价模型。

$$\min H(w) =$$

$$\sum_{i=1}^q \sum_{j=1}^m \{ [(u_j - w_j)z_{ij}]^2 + [(v_j - w_j)z_{ij}]^2 \} \quad (10)$$

约束条件为

$$\sum_{j=1}^m w_j = 1, w_j \geq 0 (j = 1, 2, \dots, m) \quad (11)$$

构造 Lagrange 函数

$$L =$$

$$\sum_{i=1}^q \sum_{j=1}^m \{ [(u_j - w_j)z_{ij}]^2 + [(v_j - w_j)z_{ij}]^2 \} + 4\lambda (\sum_{j=1}^m w_j - 1) \quad (12)$$

$$\text{令 } \frac{\partial L}{\partial w_j} = - \sum_{i=1}^q 2(u_j + v_j - 2w_j)z_{ij}^2 + 4\lambda = 0 \quad (j = 1, 2, \dots, m) \quad (13)$$

$$\frac{\partial L}{\partial \lambda} = 4(\sum_{j=1}^m w_j - 1) = 0 \quad (14)$$

用矩阵表示为

$$\begin{bmatrix} A & e \\ e^T & 0 \end{bmatrix} \cdot \begin{bmatrix} W \\ \lambda \end{bmatrix} = \begin{bmatrix} B \\ 1 \end{bmatrix} \quad (15)$$

其中 A 为 $m \times m$ 对角阵, e 、 W 、 B 均为 $m \times 1$ 的向量。

$$A = \text{diag} \left[\sum_{i=1}^q z_{i1}^2, \sum_{i=1}^q z_{i2}^2, \dots, \sum_{i=1}^q z_{im}^2 \right] \quad (16)$$

$$e = [1, 1, \dots, 1]^T \quad (17)$$

$$W = [w_1, w_2, \dots, w_m]^T \quad (18)$$

$$B =$$

$$\left[\sum_{i=1}^q \frac{1}{2}(u_1 + v_1)z_{i1}^2, \sum_{i=1}^q \frac{1}{2}(u_2 + v_2)z_{i2}^2, \dots, \sum_{i=1}^q \frac{1}{2}(u_m + v_m)z_{im}^2 \right]^T \quad (19)$$

解上述矩阵方程得到

$$W = A^{-1} \cdot \left[B + \frac{1 - e^T A^{-1} B}{e^T A^{-1} e} \cdot e \right] \quad (20)$$

(3) 两种贝叶斯修正的 AHP-熵值法结果比较

不论是基于专家赋权还是基于最小二乘思想的主客观组合赋权法,都是一种主客观结合的折中思想,权重结果介于主观权重和客观权重之间,均具有可行性,两者之间选择一种方案即可。当专家不好确定主客观权重配比时,可考虑用最小二乘法思想

计算组合权重。若专家可以给出很好的判断,且要在专家赋权法与最小二乘思想赋权法之间择优选择时,可借鉴王一任(2012)的研究思路,认为原始多种评价法的结果与组合评价法的结果之间虽不尽相同,但应非常相近,最佳的组合方法可选定为最接近多种原始方法结果的组合法。为此,通过比较两个组合方法与贝叶斯修正的 AHP 法和熵值法之间学生得分总的残差平方和 来确定最优方案,当最小时方案最优,具体计算公式如下。

由于专家赋权组合法得到的 q 个学生的得分向量为 $O = (o_1, o_2, \dots, o_q)^T$,最小二乘思想组合法得到的学生得分向量为 $W = (w_1, w_2, \dots, w_q)^T$,贝叶斯修正的 AHP 法计算得到的学生得分向量为 $U = (u_1, u_2, \dots, u_q)^T$,熵值法得到的学生得分向量为 $V = (v_1, v_2, \dots, v_q)^T$ 。

则有

$$SSE_O = \sum_{i=1}^q (o_i - u_i)^2 + \sum_{i=1}^q (o_i - v_i)^2, (i = 1, 2, \dots, q) \quad (21)$$

$$SSE_W = \sum_{i=1}^q (w_i - u_i)^2 + \sum_{i=1}^q (w_i - v_i)^2, (i = 1, 2, \dots, q) \quad (22)$$

最佳组合方法对应的总的残差平方和

$$SSE' = \min(SSE_O, SSE_W) \quad (23)$$

3 贝叶斯修正的 AHP-熵值法的应用

以沈光辉(2019)构建的内地民族班教育质量测评模型为例,对贝叶斯修正的 AHP-熵值法进行应用。内地民族班教育质量测评模型包含品德与劳动、学业发展、身心发展、审美与兴趣、服务西藏 5 个一级指标和理想信念等 17 个二级指标,并根据一、二级指标设计了 27 道测试题。在用 AHP 法计算一、二级指标时调研了 6 位相关领域比较权威的专家,最终 5 位专家的数据有效。学生测试问卷以 A 地西藏中学高中 9 个班 350 名学生为调查对象,共发放验证调查问卷 350 份,回收有效问卷 302 份,有效率为 86.29%。

3.1 贝叶斯修正的 AHP 法结果

根据 5 位专家的判断矩阵,通过 AHP 法计算出一、二级指标的权重,结合实际情况给出五位专家的先验权重为(0.30, 0.25, 0.15, 0.15, 0.15)。根据公式(1)~(3)计算出贝叶斯修正的 AHP 法的权重,具体结果见表 1。

表 1 贝叶斯修正的 AHP 法权重结果

一级指标	一级指标权重	二级指标	二级指标权重	题目编号	题目权重
品德与劳动	0.22	理想信念	0.30	1	0.022
				2	0.022
				3	0.022
		人格品质	0.29	4	0.032
				5	0.032
				6	0.042
		劳动情感	0.19	7	0.024
				8	0.024
		劳动技能	0.22	9	0.031
				10	0.031
学业发展	0.23	知识技能态度	0.27	11	0.053
				12	0.048
		汉语水平	0.23	13	0.033
		藏语水平	0.21	14	0.033
		实践创新	0.29	15	0.037
				16	0.037
身心发展	0.24	身体健康	0.31	17	0.094
		心理健康	0.39	18	0.072
		身心调控	0.30	19	0.074
		审美素质	0.49	20	0.020
				21	0.020
审美与兴趣	0.15	兴趣与求知欲	0.27	22	0.036
				23	0.058
		爱好特长	0.24	24	0.038
		民族团结	0.36	25	0.021
服务西藏	0.16	返藏意愿	0.24	26	0.021
				27	0.021
		民族文化遗产	0.40		

3.2 熵值法结果 测试题权重,具体结果如表 2 所示。

根据公式(4)~(7)计算出一、二级指标权重及

表 2 熵值法权重结果

一级指标	二级指标	题目编号	e_j	一级指标权重	二级指标权重	题目权重
品德与劳动	理想信念	1	0.992	0.248	0.093	0.019
		2	0.976			0.055
		3	0.992			0.019
	人格品质	4	0.991		0.034	0.020
		5	0.994			0.014
		6	0.983			0.039
	劳动情感	7	0.972		0.082	0.064
		8	0.992			0.018
	知识技能态度	9	0.975		0.106	0.057
		10	0.979			0.049
学业发展	汉语水平	11	0.986	0.223	0.032	0.032
	藏语水平	12	0.991		0.021	0.021
	实践创新	13	0.988		0.064	0.028
		14	0.984			0.036

续表 2

一级指标	二级指标	题目编号	e_j	一级指标权重	二级指标权重	题目权重
身心发展	身体健康	15	0.985	0.164	0.082	0.035
		16	0.980			0.047
	心理健康	17	0.983		0.039	0.039
		18	0.981			0.043
审美与兴趣	身心调控	19	0.974	0.130	0.060	0.060
		20	0.990			0.023
	兴趣与求知欲	21	0.988		0.051	0.028
		22	0.992			0.019
服务西藏	爱好特长	23	0.980	0.234	0.046	0.046
		24	0.988			0.027
	民族团结	25	0.982		0.161	0.042
		26	0.972			0.064
	返藏意愿	27	0.976			0.055

3.3 综合权重结果

将基于线性加权组合的贝叶斯修正的 AHP - 熵值法记为“组合权重 1”,在确定主客观配比时,专家一致认为该项研究中主观赋值相对重要,通过整合专家对主客观权重占比的意见,认为主观权重占

60% 比较合适,根据公式(8)取 $\alpha = 0.6$ 进行权重计算;最小二乘意义下的贝叶斯修正的 AHP - 熵值法记为“组合权重 2”,根据公式(9) - (20)计算权重。两种组合法所得结果如表 3 所示。

表 3 一、二级指标及测试题综合权重

一级指标	1	2	3	4	5				
组合权重 1	0.23	0.23	0.21	0.14	0.19				
组合权重 2	0.24	0.23	0.20	0.14	0.20				
二级指标	1	2	3	4	5	6	7	8	9
组合权重 1	0.077	0.052	0.041	0.061	0.080	0.045	0.038	0.066	0.077
组合权重 2	0.080	0.049	0.041	0.065	0.084	0.043	0.035	0.065	0.078
二级指标	10	11	12	13	14	15	16	17	
组合权重 1	0.072	0.060	0.068	0.045	0.029	0.053	0.034	0.103	
组合权重 2	0.066	0.057	0.067	0.046	0.027	0.052	0.033	0.113	
测试题	1	2	3	4	5	6	7	8	9
组合权重 1	0.021	0.035	0.021	0.027	0.025	0.041	0.040	0.021	0.042
组合权重 2	0.021	0.039	0.021	0.026	0.023	0.041	0.044	0.021	0.044
测试题	10	11	12	13	14	15	16	17	18
组合权重 1	0.038	0.045	0.038	0.031	0.034	0.036	0.041	0.072	0.060
组合权重 2	0.040	0.043	0.035	0.031	0.034	0.036	0.042	0.066	0.057
测试题	19	20	21	22	23	24	25	26	27
组合权重 1	0.068	0.021	0.023	0.029	0.053	0.034	0.029	0.039	0.035
组合权重 2	0.067	0.022	0.024	0.027	0.052	0.033	0.031	0.043	0.038

注:表 3 中一、二级指标及测试题编号与表 2 中前三列名称由上至下相对应。

3.4 结果比较

(1) 贝叶斯修正的 AHP - 熵值法与 AHP - 熵值法测评结果比较

为比较贝叶斯修正的 AHP - 熵值法与 AHP - 熵值法测评权重优劣,首先求出 AHP、加权贝叶斯修正的 AHP、熵值法、基于线性加权的贝叶斯修正

的 AHP - 熵值法、基于最小二乘的贝叶斯修正的 AHP - 熵值法、基于线性加权的 AHP - 熵值法、基于最小二乘的 AHP - 熵值法 7 种权重结果,然后分别求出 7 种权重结果与 302 名学生在 27 个测试题上得分的加权平均分(满分为 5 分),最后运用公式(21) - (23)分别在线性加权方式和最小二乘思想

两种方法下,计算贝叶斯修正的 AHP—熵值法和 AHP—熵值法与两组主、客观单独测评结果之间的残差平方和,通过比较残差平方和大小来判断方法的适用性,具体结果如表 4 所示。发现在线性加权方式下,贝叶斯修正的 AHP—熵值法与两组主、客观单独测评结果之间的残差平方和分别为 9.9729 和 9.8796,要小于 AHP—熵值法与两组主、客观单独测评结果之间的残差平方和 23.3294、23.5816,

说明贝叶斯修正的 AHP—熵值法测评的结果更加处于主客观之间;同样在最小二乘思想下,贝叶斯修正的 AHP—熵值法与两组主、客观单独测评结果之间的残差平方和分别为 13.9199 和 14.3790,要小于 AHP—熵值法与两组主、客观单独测评结果之间的残差平方和 54.6522 和 52.4674。综上,从该案例来看贝叶斯修正的 AHP—熵值法测评结果更佳,更加接近主、客观原始计算结果。

表 4 贝叶斯修正的 AHP—熵值法与 AHP—熵值法测评结果比较

比较指标	基于线性加权的 B—C	基于线性加权的 A—C	基于最小二乘的 B—C	基于最小二乘的 A—C
与 A、C 之间的残差平方和	9.9729	23.3294	13.9199	54.6522
与 B、C 之间的残差平方和	9.8796	23.5816	14.3790	52.4674

注:AHP、贝叶斯修正的 AHP、熵值法分别用 A、B、C 表示

(2)两种贝叶斯修正的 AHP—熵值法比较

两种贝叶斯修正的 AHP—熵值法所得的一、二级指标及测试题权重的等级相关系数 R_s 在 0.01 的显著性水平下相关性显著,且相关系数 R_s 分别为 1、0.9951、0.9817,表明两种贝叶斯修正的 AHP—熵值法计算的权重排序具有较高的一致性。从表 4 可以看出,基于线性加权的贝叶斯修正的 AHP—熵值法所得学生得分的总残差平方和 SSE_0 为 9.8796,基于最小二乘思想的贝叶斯修正的 AHP—熵值法所得权重的总残差平方和 SSE_w 为 14.3790,说明在该案例中基于线性加权的贝叶斯修正的 AHP—熵值法效果更佳。

(3)主、客观方法及贝叶斯修正的 AHP—熵值法之间的比较

采用 AHP、加权 AHP、贝叶斯修正的 AHP、熵值法、基于线性加权的贝叶斯修正的 AHP—熵值法(贝叶斯修正的 AHP—熵值法 1)和最小二乘意义下的贝叶斯修正的 AHP—熵值法(贝叶斯修正的 AHP—熵值法 2)六种方法计算一、二级指标及测试题权重并进行比较,详细情况见图 1—3。

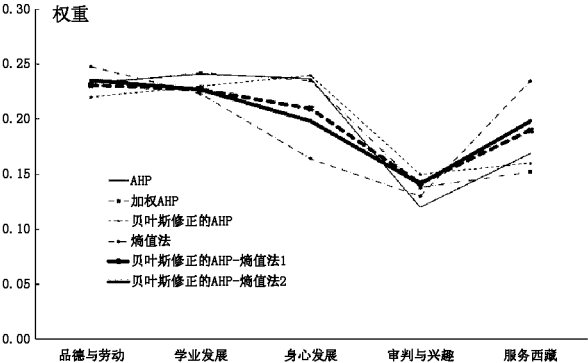


图 1 一级指标权重比较

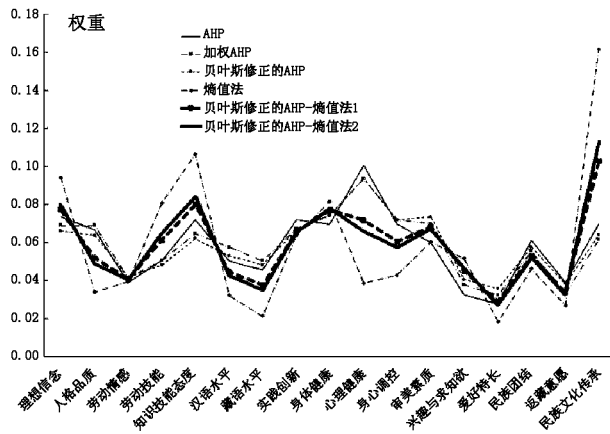


图 2 二级指标权重比较

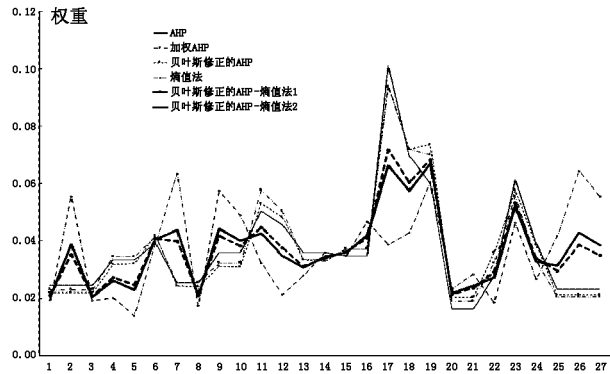


图 3 测试题指标权重比较

从图 1~3 可以看出,AHP、加权 AHP 和贝叶斯修正的 AHP 计算的一、二级指标和测试题权重变化不大,与该案例中给出的专家先验权重差异不大有关,贝叶斯修正起到微调权重的作用;熵值法计算的一、二级指标权重和测试题权重与三类 AHP 法计算的权重有较大不同,一级指标中身心发展和服务西藏的权重差异较大,二级指标中人格品质、劳动技能、知识技能态度、心理健康、民族文化传承的权

重差异较大,测试题中题目 2、5、7、9、11、12、17、18、25、26、27 的权重差异较大;贝叶斯修正的 AHP - 熵值法综合了主观权重和客观权重信息,其权重值处于三类 AHP 法和熵值法之间,平衡了部分指标或试题主客观权重差异较大的问题。

4 讨论

指标权重计算是教育测评中的重点和难点,主观和客观权重计算方法均被广泛使用,然而主观赋权受专家水平约束,客观赋权受样本数据约束。同时,主客观相结合赋权的方式也受到了不少学者的青睐,但是目前这类方法普遍忽视了专家自身的权重。为此,提出贝叶斯修正的 AHP - 熵值法进行指标权重计算,该方法采用主客观相结合的方式将 AHP 法和熵值法进行结合,同时纳入专家自身权重,并基于数据对专家权重进行了贝叶斯修正,最后计算综合权重。通过实例分析,发现该方法具有可行性、可操作和有效性。为更好的进行权重方法选择,研究对贝叶斯修正 AHP - 熵值法进行如下讨论。

(1) 贝叶斯修正的 AHP - 熵值法与传统方法比较及其选择。与传统教育测评权重计算方法相比,该模型可以克服专家水平差异、主观性或客观性强、观测点权重难获得等不足。其中贝叶斯修正的思想有助于削弱专家先验权重的主观性,使得专家的权重更加合理、客观。虽然贝叶斯修正的 AHP - 熵值法平衡了主客观评价法,考虑了专家本身存在的权重差异等,但随着纳入因素的增多,要求的条件也会更加苛刻。因此,当专家信息全面,主客观数据均可获得时,可采用贝叶斯修正的 AHP - 熵值法;当无法获得更加全面的专家信息时,可考虑使用常规的主客观结合法计算指标权重;当无法同时获得主客观数据时,单一的主观或客观评价方法依旧可以使用。

(2) 两种贝叶斯修正的 AHP - 熵值法的优缺点及其选择。基于线性加权的贝叶斯修正的 AHP - 熵值法的优点在于对主客观方法进行组合时计算方法简单,只需要计算简单的加权平均数,但是难点在于要注意权衡主客观权重配比,需充分考虑专家的权威性,以专家团队达成共识为宜。此外,该方法需要给出两次专家咨询结果,这在一定程度上也加重了结果的主观性。最小二乘意义下的贝叶斯修正的 AHP - 熵值法优点在于只需要用到一次专家主观赋权数据,但是缺点是对主客观方法进行组合时相比

线性加权的方式要更为复杂。两种方法选择思路:在专家咨询数据不易获得的情况下,采用最小二乘意义下的贝叶斯修正的 AHP - 熵值法;在专家咨询数据容易获得的情况下,可以计算出两种贝叶斯修正的 AHP - 熵值法测评结果,并通过与主、客观单独测评结果之间的残差平方和进行比较,选择残差平方和最小的方法。

(3) 贝叶斯修正的 AHP - 熵值法的应用前景。当前,主客观综合评价法已成为权重计算方法的主流趋势。第一,该方法在教育学和心理学领域有很好的应用前景。教育学和心理学领域开展了大量量表开发与模型构建工作,未来测评需求大,可以将该方法引入进来。为新时代教育评价改革服务,为落实中共中央、国务院《深化新时代教育评价改革总体方案》提出的改进结果评价,强化过程评价,探索增值评价,健全综合评价服务。第二,该方法可拓展应用领域广泛,不论是在教育学、心理学、管理学、经济学、艺术学等人文社会科学领域,还是在医学、计算机、环境科学等自然科学领域,都可以使用该方法进行测评指标权重计算。

参考文献

- 丁家玲,叶金华.(2003).层次分析法和模糊综合评判在教师课堂教学质量评价中的应用.武汉大学学报(社会科学版),56(2),241-245.
- 范涌峰,宋乃庆.(2019).大数据时代的教育测评模型及其范式构建.中国社会科学,40(12),139-155.
- 郭显光.(1994).熵值法及其在综合评价中的应用.财贸研究,15(6),56-60.
- 郭显光.(1998).改进的熵值法及其在经济效益评价中的应用.系统工程理论与实践,18(12),99-103.
- 胡恩华,陈沛然,顾桂芳.(2015).全日制专业硕士教育评价指标体系的构建.研究生教育研究,30(1),75-79.
- 江浩,江兵.(2010).专家权重的贝叶斯修正法.统计与决策,26(8),161-162.
- 李芳,何思俊,支锦亦,王超,向泽锐.(2020).基于AHP-熵权法的高速列车乘客车厢设计满意度评价.机械设计,37(2),121-125.
- 李艳燕,董笑男,李新,张媛.(2020).STEM教育质量评价指标体系构建.现代远程教育研究,32(2),48-55.
- 刘海滨,杨颖秀,陈雷.(2012).基于AHP的大学生就业创业教育评价指标体系构建.东北师大学报(哲学社会科学版),62(6),227-232.
- 罗志敏,马浚锋.(2020).中国高等教育省域财政压力指数的研制与测评.教育与经济,36(3),39-49.
- 毛定祥.(2002).一种最小二乘意义下主客观评价一致的组

- 合评价方法. *中国管理科学*, 10(5), 96 – 98.
- 彭张林, 张强, 王素凤, 杨善林. (2016). 基于评价结论的二次组合评价方法研究. *中国管理科学*, 24(9), 156 – 164.
- 沈光辉. (2019). 内地民族班教育质量测评模型研究 (硕士学位论文). 西南大学, 重庆.
- 沈光辉. (2023). 初中生创客素养测评模型研究 (博士学位论文). 西南大学, 重庆.
- 沈光辉, 高鑫, 宋乃庆. (2022). 基于创客教育视角的初中生创新能力测评模型构建与应用. *中国电化教育*, 43(4), 71 – 78.
- 宋乃庆, 沈光辉, 裴昌根, 陈婷. (2019). 内地民族班教育质量测评模型的构建探析——以内地西藏班为例. *西藏大学学报 (社会科学版)*, 34(1), 209 – 215.
- 王一任. (2012). 综合评价方法若干问题研究及其医学应用 (博士学位论文). 中南大学, 长沙.
- 曾宪报. (1997). 关于组合评价法的事前事后检验. *统计研究*, 14(6), 56 – 58.
- 张乔木. (2017). 基于德尔菲调查和聚类分析的关键共性技术预见研究——以山西省新材料行业为例. *科技管理研究*, 37(13), 121 – 124.
- 张玉玲, 迟国泰, 祝志川. (2011). 基于变异系数 – AHP 的经济评价模型及中国十五期间实证研究. *管理评论*, 23(1), 3 – 13.
- 郑谦, 汪伟忠, 赵伟峰, 胡月英. (2016). 应用型高校实践教学质量评价指标体系研究. *高教探索*, 32(12), 36 – 40.
- Shen, G. (2023). *The evaluation of chinese middle school students' maker literacy*. Niagara Falls: Untested Ideas Research Center.

The Application of Bayesian Modified AHP – Entropy Method in Educational Evaluation

Shen Guanghui^{1,2}, Ye Tong¹, Sun Xiaojian³, Fan Yongfeng⁴

(1. School of Teacher Education, Jiangsu University, Zhenjiang 212013;

2. The Evidence – based Research Center for Educational Assessment, Jiangsu University, Zhenjiang 212013;

3. School of Mathematics and Statistics, Southwest University, Chongqing 400715;

4. Faculty of Education, Southwest University, Chongqing 400715)

Abstract: Index weight determination method plays a key role in education evaluation, which will directly affect the evaluation results. Analytic hierarchy process (AHP) and entropy method are the most commonly used methods, but they have the prominent problems of strong subjectivity and high sample dependence. In addition, the commonly used subjective and objective comprehensive weighting method ignores the influence of expert level on index weight. The weight calculation method of Bayesian modified AHP – entropy firstly modifies the weight obtained by AHP method by using the expert's own weight, and then calculates the combined weight with entropy method. Based on the comprehensive analysis of the advantages and calculation methods of Bayesian modified AHP – entropy, taking the education quality evaluation of Tibetan classes in Chinese hinterland as an example, through the comparison of various methods, it is found that the results calculated by this method are feasible, operable and effective.

Key words: Bayes; analytic hierarchy process; entropy method; education evaluation; combination weights