

基于模糊集合的评审决策信息融合

王璐^{1,2}, 李根梓³, 戴亚飞¹, 付雪峰^{1,*}

¹国家自然科学基金委员会, 北京 100085

²苏州大学功能纳米与软物质研究院, 江苏 苏州 215123

³雁栖湖基础制造技术研究院(北京)有限公司, 北京 101407

On Information Fusion by Fuzzy Sets for Decision Making in Peer Review

Lu Wang^{1,2}, Genzi Li³, Yafei Dai¹, Xuefeng Fu^{1,*}

¹ National Natural Science Foundation of China, Beijing 100085, China.

² Institute of Functional Nano & Soft Materials (FUNSOM), Soochow University, Suzhou 215123, Jiangsu Province, China.

³ Yanqi Lake Institute of Basic Manufacturing Technology (Beijing) Co., Ltd., Beijing 101407, China.

*Corresponding author. Email: fuxf@nsfc.gov.cn.

表 S1 项目评审模型示意
Table S1 Project review model.

项目 \ 评委	P_1	P_2	P_3	P_{n-1}	P_n
G_1	S_{11}	S_{12}	S_{13}	$S_{1(n-1)}$	S_{1n}
G_2	S_{21}	S_{22}	S_{23}	$S_{2(n-1)}$	S_{2n}
...
...
G_{m-1}	$S_{(m-1)1}$	$S_{(m-1)2}$	$S_{(m-1)3}$	$S_{(m-1)(n-1)}$	$S_{(m-1)n}$
G_m	S_{m1}	S_{m2}	S_{m3}	$S_{m(n-1)}$	S_{mn}

表 S2 符号和名词术语列表
Table S2 Terms and Symbols.

序号	名称	定义
1	项目 G_i	项目包中有 m 个项目： $i \in [1, m]$ ，即表 S1 有 m 行
2	评委 P_j	负责评审此包项目的 n 个评委： $j \in [1, n]$ ，即表 S1 有 n 列
3	评分 S_{ij}	P_j 对 G_i 的评分，即表 S1 中的单元格内容
4	项目评分集合 $S(G_i n)$	项目 G_i 所得到的 n 个评分的集合，即表 S1 中的行
5	评分集合平均分 μ_s	项目评分集合 $S(G_i n)$ 中所有评分 S_{ij} 的平均分
6	评分与均分的差 Δ_i	S_{ij} 与 μ_s 的差值
7	评委打分集合 $S(P_j m)$	评委 P_j 对 m 个项目评分的集合，即表 S1 中的列
8	评委最高分 $\text{Max}[S(P_j m)]$	评委 P_j 在 m 个项目中打出的最高分，即表 S1 每列中的最高分
9	评委最低分 $\text{Min}[S(P_j m)]$	评委 P_j 在 m 个项目中打出的最低分，即表 S1 每列中的最低分
10	归一化评分 NS_{ij}	用 Min-Max Normalization 方法 ^[1] 以评委 P_j 为索引(即以表 1 中的列为索引)对 S_{ij} 进行归一化处理： $NS_{ij} = \frac{s_{ij} - \text{Min}[S(R_j m)]}{\text{Max}[S(R_j m)] - \text{Min}[S(R_j m)]}$ ，将评分转化为归一化评分 NS_{ij} 。通过归一化处理，可以消除评委 P_j 对于 m 个项目整体的“手松”或“手紧”，实现各评委间评分尺度的统一
11	归一化项目评分集合 $NS(G_i n)$	以项目 G_i 为索引(即以表 S1 中的行为索引)，对归一化后的评分进行归集，得到归一化项目评分集合 $NS(G_i n)$
12	置换标准偏差 $\sigma(j i)$	对项目 G_i 的归一化评分集合 $NS(G_i n)$ ，采用留一法交叉验证，对每个 NS_{ij} 计算不包含其本身的 $n-1$ 个数值的标准偏差 $\sigma(j i)$ ，称为 NS_{ij} 的置换标准偏差；对 $NS(G_i n)$ 中的 n 个数值均进行上述操作，可得到 n 个 $\sigma(j i)$ ，构成项目 G_i 的置换标准偏差集合
13	置换标准偏差集合的平均值 μ_σ	项目 G_i 的置换标准偏差集合中所包含数值的平均值
14	置换标准偏差集合的标准偏差 σ_σ	项目 G_i 的置换标准偏差集合中所包含数值的标准偏差
15	奈尔统计量 R_j	使用奈尔(Nair)检验法对项目 G_i 的置换标准偏差集合中的 $\sigma(j i)$ 进行奈尔统计量计算
16	奈尔临界值 R_c	可查表：《中华人民共和国国家标准 GB/T 4883-2008 数据的统计处理和解释正态样本离群值的判断和处理》附表 A.1 奈尔(Nair)检验的临界值表；一般取检出水平 $\alpha = 0.05$
17	最小奈尔统计量 R_{\min}	项目 G_i 的置换标准偏差集合的奈尔统计量 R_j 中的最小值
18	模糊参数 w_{ij}	取值为 0 到 1，表示一个对象对于模糊集合的隶属程度
19	共识评分 ES_{ij}	通过对评分 S_{ij} 进行转换而获得
20	共识评分均值 μ_{ES}	项目 G_i 所对应的共识评分 ES_{ij} 的算术平均值

1 留一法交叉验证

留一法交叉验证(Leave-one-out Cross-validation, 简称 LOO)通常用来进行模型精度估计^[2,3]。假定一个数据集包含 a 个样本，LOO 方法一般包括如下步骤：1) 取出一个样本 k ；2) 用剩余 $a-1$ 个样本构造模型，并用此

模型预测样本 k 对应的值；3) 比较样本 k 的预测值和实际值，获得模型在 k 点的精度估计；4) 对 a 个样本逐一执行上述步骤，获得模型的整体精度估计^[4]。

在本文研究的问题中，给定某项目的一组评分，通过求其标准偏差，我们可以了解这组评分整体的分散程度。但每个评分对整体分散程度的影响是未知的。我们借鉴 LOO 方法的思想，通过计算某个分数被取出后剩余分数的标准偏差（该被取出分数的置换标准偏差），来获取每个评分对整体分散程度的影响。我们针对归一化评分进行计算，具体方法如下：1) 取出第 j 个归一化评分 NS_{ij} ；2) 对 $NS(G_i|n)$ 中其余的 $n - 1$ 个数值求标准偏差 $\sigma(j|i)$ ，称为 NS_{ij} 的置换标准偏差；3) 对 $NS(G_i|n)$ 中的 n 个数值均进行上述操作，可得到 n 个 $\sigma(j|i)$ ，构成项目 G_i 的置换标准偏差集合。将这 n 个 $\sigma(j|i)$ 进行排序，较小的 $\sigma(j|i)$ 值表示不包含 NS_{ij} 的其余 $n - 1$ 个分数的分散程度较小，或对应的 NS_{ij} 对 n 个分数整体分散程度的影响较大；而较大的 $\sigma(j|i)$ 值表示不包含 NS_{ij} 的其余 $n - 1$ 个分数的分散程度较大，或对应的 NS_{ij} 对 n 个分数整体分散程度的影响较小。根据 NS_{ij} 和 S_{ij} 的映射关系，通过计算 $\sigma(j|i)$ ，就可以量化评估每个 S_{ij} 对项目评分整体分散程度的影响。

2 奈尔检验

在一组平行测定中，若有个别数据与平均值差别较大，则把此数据视为离群值^[5,6]。类似地，一组项目评分的质量也可以通过离群值(Outlier)的识别进行评估。离群值的判定方法有很多种，其中，奈尔检验法的适用样本量为 $3 \leq n \leq 100$ ，比较适合项目评审中一组专家数量的一般情况。同时，项目评分经过统计量转换后可视为符合正态分布，且总体标准差为已知，满足奈尔检验法的适用条件。

奈尔检验法(Nair Test)首先计算每个样本的奈尔统计量，然后逐个将其与奈尔临界值 R_c (查表获得，一般设检出水平 $\alpha = 0.05$) 进行比较和判定：当样本的奈尔统计量大于奈尔临界值时，判定该样本为离群值^[5,6]。在本研究中，可能存在的离群值对应的是对分数整体分散程度有较大影响的 NS_{ij} ，或较小的 $\sigma(j|i)$ 值。因此，依据奈尔检验法中的下侧情形，按照式(1)对置换标准偏差 $\sigma(j|i)$ 进行奈尔统计量 R_j 的计算：

$$R_j = [\mu_\sigma - \sigma(j|i)] / \sigma_\sigma \quad (1)$$

其中， μ_σ 是置换标准偏差集合的平均值， σ_σ 是置换标准偏差集合的标准偏差。由式(1)可知，当 $\sigma(j|i) \leq \mu_\sigma$ 时， $R_j \geq 0$ ；当 $\sigma(j|i)$ 足够小时，会存在 $R_j > R_c$ ；当 $\sigma(j|i) > \mu_\sigma$ 时， $R_j < 0 < R_c$ 。因此，可以将 R_j 作为 S_{ij} 对项目评分整体分散程度影响的量化评估指标。

项目 G_i 的置换标准偏差集合中的 $\sigma(j|i)$ 及该集合的标准偏差 σ_σ 综合起来可反映专家对 G_i 的共识程度以及 APPS 方法的适用范围：1) 当 $\sigma(j|i)$ 均较大，且 σ_σ 较大时，表明专家对 G_i 的评价较为分散且差异程度较大，此时应考虑可能为非共识类项目，不适用于 APPS 方法；2) 当 $\sigma(j|i)$ 均较大，但 σ_σ 较小时，表明专家对 G_i 的评价较为分散但差异程度较小，适用 APPS 方法；3) 当 $\sigma(j|i)$ 部分较小，但 σ_σ 较大时，表明部分专家对 G_i 的评价与其余专家的差异程度较大，适用 APPS 方法；4) 当 $\sigma(j|i)$ 均较小，且 σ_σ 较小时，表明专家对 G_i 的共识程度较高，适用 APPS 方法；5) 特别地，当 $\sigma_\sigma = 0$ 时，则 $\sigma(j|i)$ 均相等且为 0，说明专家对某项目的价值预期完全达成了共识，都给出了相同的评分，无评审偏差需用 APPS 方法来处理。综上所述，APPS 适用于第 2)、第 3) 和第 4) 类情形。

3 一致性函数

结合奈尔检验的离群值判定方法，构造一致性函数如下：

$$w_{ij} = \begin{cases} 1, & R_j \geq R_c \\ \frac{R_j - R_{min}}{R_c - R_{min}}, & R_j < R_c \end{cases} \quad (2)$$

其中， R_j 是置换标准偏差 $\sigma(j|i)$ 的奈尔统计量，由式(1)计算获得， R_j 与 S_{ij} 是一一映射； R_c 是奈尔临界值，可查表获得^[6]； R_{min} 是置换标准偏差集合的奈尔统计量 R_j 中的最小值。

由式(2)可知，当 $R_j \geq R_c$ 时，评分 S_{ij} 所对应的 w_{ij} 为 1，即 S_{ij} 完全隶属于评分所对应的模糊集合；当 $R_j < R_c$ 时， S_{ij} 所对应的 w_{ij} 小于 1，即 S_{ij} 以小于 1 的一致性隶属于评分所对应的模糊集合。特别是当 $R_j = R_{min}$ 时，可知 $w_{ij} = 0$ ，即 S_{ij} 完全不隶属于评分所对应的模糊集合。通过式(2)，我们可对每个评分 S_{ij} 都赋予一个取值为 0 到 1 的模糊参数 w_{ij} ，从而实现评审结果的模糊集合表达。

评分 S_{ij} 的模糊参数 w_{ij} 还可以作为评委 P_j 对项目 G_i 的评审贡献的量化评价指标。当 w_{ij} 趋近于1时,表示评审专家可靠地识别了项目的真实质量,并且用与之足够接近的质量预期进行了较为准确的表达;反之,当 w_{ij} 趋近于0时,表示评审专家给出了不准确的质量预期,或者给出了不准确的表达。通过计算评委 P_j 所对应 w_{ij} 的均值和标准差等,就可以评估评委 P_j 的评审贡献。对于足够多的 w_{ij} ,其均值是收敛的,其标准差是稳定的,可以用来构建 P_j 的评审行为模型,并可根据新获得的 w_{ij} 进行持续更新。对照这个评审行为模型,通过分析评委 P_j 在本次评审中的全部 w_{ij} ,就能够对 P_j 在本次评审中的贡献给出综合评价;通过分析评委 P_j 在多次评审中的所有 w_{ij} ,就能够对评委 P_j 在多次评审中的长期贡献给出总体评价。上述评价结果可以作为评审专家贡献的参考依据之一。

4 共识函数

在实现评审结果模糊集合表达的基础上,我们还需将其转化为单一分值,以便完成项目排序等后续工作。我们将一组评分以提高共识程度为目标而进行收敛的过程构造为共识函数如下:

$$ES_{ij} = S_{ij} - (1 - w_{ij})\Delta_i \quad (3)$$

其中, ES_{ij} 表示共识评分, S_{ij} 是原始评分, w_{ij} 是由式(2)获得的该评分的模糊参数, Δ_i 表示该原始评分 S_{ij} 与原始平均分 μ_s 的差:

$$\Delta_i = S_{ij} - \mu_s \quad (4)$$

由式(4)可知,当 S_{ij} 大于或小于 μ_s 时,对应的 Δ_i 分别大于0或小于0。又由式(3)可知,通过对 S_{ij} 进行一定程度的 Δ_i 量值的增(当 Δ_i 小于0时)或减(当 Δ_i 大于0时),可实现 S_{ij} 向 μ_s 的初始收敛。此外,由上述式(3)和式(4)可知,当 S_{ij} 等于 μ_s 时,对应的 Δ_i 为0,则 $ES_{ij} = S_{ij}$,表示该评分 S_{ij} 就是收敛的初始目标。

对于一组给定的评分,收敛目标是确定但未知的。将初始平均分作为初始的收敛目标是一种合理的选择,可通过迭代实现向实际收敛目标的逼近。通过共识函数,我们可将原始评分 S_{ij} 转换为共识评分 ES_{ij} 。将获得的一组共识评分求平均值 μ_{ES} 。可以用新获得的 ES_{ij} 和 μ_{ES} 来重新计算 Δ_i ,实现收敛目标的更新:

$$\Delta_i = ES_{ij} - \mu_{ES} \quad (5)$$

通过式(3)和式(5)的多次迭代,就可以实现 μ_{ES} 的稳定,从而逼近实际收敛目标,得到一组专家对一个项目评分模糊集合的单一分值表达。

除了 Δ_i 的量值以外,评分 S_{ij} 收敛的程度还取决于 $1 - w_{ij}$ 的值。可分为以下三种情况:1)当 S_{ij} 的模糊参数 w_{ij} 等于1时,说明 S_{ij} 完全隶属于该评分对应的模糊集合,而此时 $1 - w_{ij} = 0$,表示 S_{ij} 保持原值,不需要进行收敛;2)当 S_{ij} 的模糊参数 w_{ij} 等于0时,说明 S_{ij} 完全不隶属于该评分对应的模糊集合,而此时 $1 - w_{ij} = 1$,表示 S_{ij} 需要完全收敛到 μ_s ;3)当 S_{ij} 的模糊参数 w_{ij} 介于0和1之间时,说明 S_{ij} 以小于1的一致性隶属于该评分对应的模糊集合,而此时 $0 < 1 - w_{ij} < 1$,表示 S_{ij} 需要以 $1 - w_{ij}$ 的程度向 μ_s 进行收敛。

5 APPS 算法的具体步骤

1) **步骤 1:** n 个评委给 m 个项目打分,共得到 $m \times n$ 个评分 S_{ij} ,构造项目评分集合 $S(G_i|n)$ 和评委打分集合 $S(P_j|m)$;求 $S(G_i|n)$ 的均值 μ_s ,令 $\Delta_i = S_{ij} - \mu_s$;

2) **步骤 2:**对1个评委 P_j 分别计算评委最高分 $Max[S(P_j|m)]$ 和评委最低分 $Min[S(P_j|m)]$,然后计算该评委 P_j 的 m 个归一化评分 NS_{ij} ;

3) **步骤 3:**对 n 个评委分别进行步骤2的计算,然后构造每个项目的归一化项目评分集合 $NS(G_i|n)$;

4) **步骤 4:**用leave-one-out (LOO)方法计算项目 G_i 所得 n 个评分的置换标准偏差 $\sigma(f|i)$,进而构造项目 G_i 的置换标准偏差集合(注意,此处运算应采用LOO对应的 $n - 1$ 个数值的平均值);

5) **步骤 5:**分别计算置换标准偏差集合的平均值 μ_σ 和置换标准偏差集合的标准偏差 σ_σ ,然后计算奈尔统计量 R_j ;

6) **步骤 6:**通过一致性函数计算模糊参数 w_{ij} :对于 $R_j \geq R_c$,令 $w_{ij} = 1$;对于 $R_j < R_c$,令 $w_{ij} = \frac{R_j - R_{\min}}{R_c - R_{\min}}$;构建项目 G_i 的评分模糊集合;

7) **步骤 7:**通过共识函数 $ES_{ij} = S_{ij} - (1 - w_{ij})\Delta_i$ 将原始评分转换为共识评分,构造项目 G_i 的共识评分集合,

计算项目 G_i 的共识评分均值 μ_{ES} ；本步骤通过图3所示的收敛过程来实现；

8) 进行 **M 循环**，对下一个项目执行上述 4-7 步的操作，直到 m 个项目全部完成；

9) **步骤 8**：获得 m 个 μ_{ES} ，分别作为对应的 m 个项目的有效分，可作为项目排序等后续工作的依据；

10) **步骤 9**：获得 $m \times n$ 个 w_{ij} ，分别按照对应评委 P_j 进行归集，每个评委对应 m 个 w_{ij} ，统计处理后作为专家贡献统计的参考依据之一。

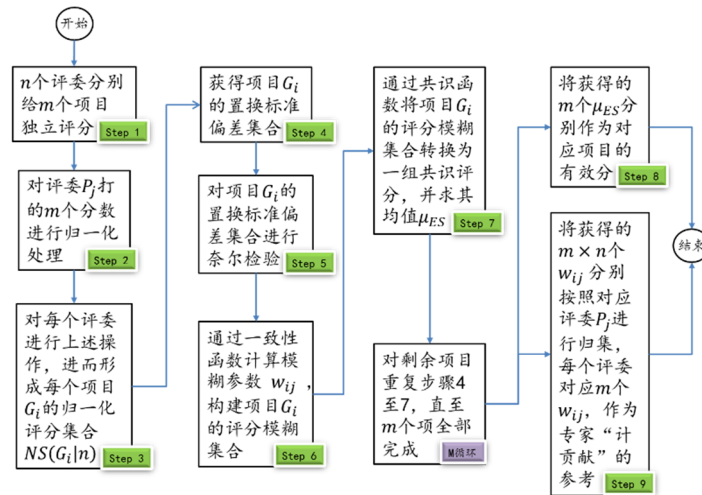


图 S1 APPS 算法流程图

Fig. S1 Flowchart of APPS method.

References

- (1) Chakraborty, S.; YehC, H. A Simulation based Comparative Study of Normalization Procedures in Multiattribute Decision Making. Proceedings of the 6th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, Greece, World Scientific and Engineering Academy and Society (WSEAS), 2007.
- (2) Vehtari, A.; Gelman, A.; Gabry, J. *Statist. Comput.* **2017**, 27 (5), 1413. doi: 10.1007/s11222-016-9696-4
- (3) Geisser, S. *J. Am. Statist. Association* **1975**, 70 (350), 320. doi: 10.1080/01621459.1975.10479865
- (4) Li, G.; Aute, V.; Azarm, S. *Struct. Multidiscip. Optim.* **2010**, 40 (1-6), 137. doi: 10.1007/s00158-009-0395-z
- (5) ISO 16269-4:2010, Statistical Interpretation of Data-Part 4: Detection and Treatment of Outliers.
- (6) GB/T 4883-2008, 数据的统计处理和解释正态样本离群值的判断和处理.
GB/T 4883-2008, Statistical Interpretation of Data - Detection and Treatment of Outliers in the Normal Sample.