考虑混合指派偏好增量提取的多准则分类方法

姜继存[1]，朱建军[1†]，刘小弟[2]

(1. 南京航空航天大学 经济与管理学院，南京 211106；
2. 安徽工业大学 微电子与数据科学学院，安徽 马鞍山 243032)

摘　要: 多准则分类 (MCS) 问题在实际的社会生活与工业工程中普遍存在, 随着决策问题的日益复杂, 决策信息呈现出多元性与异构性特征, 单一偏好结构的 MCS 方法在很多场景下无法满足现实决策需求. 此外, 复杂的决策环境导致部分目标分类不确定, 大大降低了结果的可靠性. 为此, 研究面向混合指派偏好增量提取的多准则分类问题. 首先, 在增量偏好提取过程中分析决策者混合指派偏好特征, 以此构建分类模型分布空间; 然后, 利用随机多准则可接受性分析思想 (SMAA), 建立面向混合指派偏好的分类不确定性度量方法; 接着, 构建考虑分类不确定性度量的混合指派偏好增量提取方法, 主动学习决策者的混合指派偏好以逐步减小分类模型不确定性, 从而保证 MCS 结果稳定性; 最后, 以 MBA 项目分类问题表明所提出方法的可行性和有效性.

关键词: 多准则分类；混合指派偏好；分类不确定性；增量提取；随机多准则可接受性分析；代表模型

中图分类号: C934　　**文献标志码:** A

DOI: 10.13195/j.kzyjc.2026.0419

引用格式: 姜继存, 朱建军, 刘小弟. 考虑混合指派偏好增量提取的多准则分类方法 [J]. 控制与决策.

Multiple criteria sorting method considering incremental elicitation with hybrid assignment preferences

JIANG Ji-cun[1]，*ZHU Jian-jun*[1†]，*LIU Xiao-di*[2]

(1. School of Economics and Management，Nanjing University of Aeronautics and Astronautics，Nanjing 211106，China； 2. School of Microelectronics and Data Science， Anhui University of Technology， Ma'anshan 243032，China)

Abstract: Multiple criteria sorting (MCS) problems are prevalent in real-world social life and industrial engineering. As decision-making issues become increasingly complex, decision information exhibits characteristics of diversity and heterogeneity, which makes single-preference-structured MCS methods insufficient to meet practical decision-making needs in many scenarios. Additionally, the complex decision-making environment leads to uncertainties in partial objective sorting, significantly reducing the result reliability. Hence, this paper investigates multiple criteria sorting problems with hybrid assignment preferences. During the process of incremental preference elicitation, the hybrid assignment characteristics of decision-maker are analysed to construct distribution space of sorting model. Then, sorting uncertainty measure methods under hybrid assignment preferences are advanced by stochastic multi-criteria acceptability analysis (SMAA). Based on the uncertainty metric, a hybrid incremental preference elicitation method is developed to actively learn the hybrid assignment preferences of decision-maker, thereby progressively reducing the uncertainty of sorting model and ensuring the stability of MCS results. Finally, a case study on MBA sorting is presented to demonstrate the feasibility and effectiveness of the proposed method.

Keywords: multiple criteria sorting； hybrid assignment preference； sorting uncertainty； incremental elicitation； stochastic multi-criteria acceptability analysis；representative model

0　引　言

多准则分类 (MCS) 旨在根据决策方案准则性能划分其所属类别的过程, 在社会工业生产和管理实践中应用非常广泛[1-2]. 基于价值函数的多准则决策

收稿日期: 2026-05-08；**录用日期:** 2026-09-06.
基金项目: 国家自然科学基金项目 (72071106, 72571135).
责任编委: 刘宝碇.
†通信作者. E-mail: zhujianjun@nuaa.edu.cn.

分类方法具有强大的可解释性和直观性, 受到了国内外学者的广泛关注[3-4]. 该方法通过提取决策者偏好获得价值函数, 主要包含偏好批量提取和增量提取两种方式[5-6]. 偏好增量提取允许决策者逐步提供指派偏好, 通过算法迭代持续优化分类模型[7]. 相比于偏好批量提取, 偏好增量提取可以有效减少决策者的整体判断, 降低认知负担, 提高分类结果稳定性, 已经成为重要的分类方法[1]. 如文献 [8] 提出了一种基于遗憾最小化的增量偏好提取方法, 能够减小 MCS 中准则权重的不确定性, 文献 [9] 提出一种基于 Choquet 积分的增量提取方法, 能较好考虑准则的交互作用, 文献 [10] 针对带有类别基数限制的 MCS 问题, 开发了一种交互式增量提取方法, 能够检验迭代过程中的指派偏好不一致, 文献 [11] 提出了一种考虑迭代交互的启发式增量提取方法, 能较好减小分类模型的不确定性影响, 文献 [12] 通过增量提取过程主动学习决策者关于准则的非单调认知, 能较大程度上兼容决策者的非一致指派偏好. 上述研究为解决 MCS 问题作出了积极贡献, 但现有方法要求决策者提供单一结构的确定性指派偏好.

由于决策习惯、决策压力等因素导致决策者认知往往具有不确定性和多元性, 面向单一结构、确定性偏好的 MCS 问题在很多场景下难以满足现实决策需求, 包括确定型指派偏好、不确定型指派偏好和成对比较型指派偏好等多类型混合指派偏好在日常决策中普遍存在. 实际上, 混合偏好信息在排序决策[13-14] 与共识决策[15-16] 中已得到较为系统的研究, 然而, MCS 问题中的混合指派偏好与上述问题存在本质差异, 排序决策旨在获得方案的完全序, 混合偏好信息 (如“方案 a 优于 b”) 直接用于推断方案的优劣关系, MCS 目标是将方案划分至预定义的有序类别, 混合指派偏好直接或间接反映方案与类别边界的相对位置; 共识决策中的混合偏好主要关注不同决策者的评估差异, 重点是冲突消解与共识达成, MCS 则聚焦于决策者关于方案指派的不确定性和多元性, 核心在于如何构建兼容且稳定的分类模型[17]. 文献调研表明, 一些分类研究中也涉及了混合偏好, 如文献 [5] 和文献 [18] 等, 但是这些方法属于批量提取范畴, 且较少考虑 MCS 的结果稳定性.

混合指派偏好增量提取研究主要包括以下难点: 1) 确定型、不确定型和成对比较型指派偏好具有不同的指派形式, 如何构建兼容所有异构偏好的分类模型是首要挑战. 2) 混合指派偏好的引入导致分类模型分布空间更加复杂, 分类不确定性呈现出多样化. 对此, 本文提出一种考虑混合偏好增量提取的 MCS 方法, 主要贡献如下:

1) 针对决策者的混合指派偏好, 构建了面向混合指派偏好的分类模型分布空间;

2) 利用随机多准则可接受性分析 (SMAA) 思想, 提出了基于混合指派偏好的多元分类不确定性度量方法;

3) 构建了交互问答式增量偏好提取过程, 通过策略迭代快速减小分类模型不确定性并获得稳定的 MCS 结果.

1 基础概念

记 $A=\{a_1,\ldots,a_m\}$为方案集, $M=\{1,\ldots,m\}$为方案数量集, $G=\{g_1,\ldots,g_n\}$为评估准则, $N=\{1,\ldots,n\}$为准则数量集, $V=(v_{ij})_{m\times n}$为决策信息, v_{ij}表示a_i关于g_j的性能水平. MCS 需要将$a_i\in A$划分到预定类别$C=\{C_1,\ldots,C_p\}$, $P=\{1,\ldots,p\}$为类别数量集. 其中, $C_q\succ C_{q-1}, q=2,\ldots,p$. 设置分类阈值$T=\{t_0,\ldots,t_p\}$满足

$$t_q\geqslant t_{q-1}+\varepsilon,\ q=1,\ldots,p,\ t_0=0,$$
$$t_p=1+\varepsilon,\ p\geqslant 2.$$

其中: ε为极小正数, 可将$t_q>t_{q-1}$转化为$t_q\geqslant t_{q-1}+\varepsilon$[17].

基于阈值的 MCS 方法通过价值函数聚合a_i边际价值, 对比分类阈值以实现分类.

$$U(a_i)=\sum_{j=1}^{n}u_j(v_{ij}),i\in M. \tag{1}$$

其中, $U(\cdot)$表示价值函数, $U(a_i)\in[0,1]$表示a_i综合价值. $u_j(v_{ij})$表示v_{ij}边际价值. $u_j(\cdot)$为边际价值函数. 如果$U(a_i)\in[t_{q-1},t_q)$, 则方案a_i属于类别C_q.

由于任意连续函数都可以通过连续线性函数进行逼近, 分段线性函数常常作为边际价值函数的基本刻画[19]. 设 $S_j=[L_j,R_j]$为准则 g_j论域, $L_j=\min\limits_{i\in M}\{v_{ij}\}, R_j=\max\limits_{i\in M}\{v_{ij}\}(j\in N)$. S_j可划分为κ_j个长度相等子区间$[x_j^0,x_j^1],\ldots,[x_j^{\kappa_j-1},x_j^{\kappa_j}]$. 其中, $x_j^k=L_j+\dfrac{k}{\kappa_j}(R_j-L_j), k=0,1,\ldots,\kappa_j$为特征点, 控制边际价值函数的形式和分布. 边际价值函数满足以下归一化条件[20].

$$u_j(L_j)=0,\ \sum_{j=1}^{n}u_j(R_j)=1,\ j\in N. \tag{2}$$

通过线性插值确定性能水平v_{ij}的边际价值

$$u_j(v_{ij}) = u_j(x_j^k) + \frac{v_{ij}-x_j^k}{x_j^{k+1}-x_j^k} \times (u_j(x_j^{k+1}) - u_j(x_j^k)),\ v_{ij} \in [x_j^k, x_j^{k+1}]. \tag{3}$$

综上所述, 价值函数与分类阈值应满足式 (4) 的基本条件, 构成类别的序关系约束和边际价值函数的单调性约束.

$$E_{Base}\begin{cases} U(a_i) = \sum\limits_{j=1}^{n} u_j(v_{ij}),\ i \in M; \\ u_j(v_{ij}) = u_j(x_j^k) + \dfrac{v_{ij}-x_j^k}{x_j^{k+1}-x_j^k} \times \\ \qquad (u_j(x_j^{k+1}) - u_j(x_j^k)); \\ v_{ij} \in [x_j^k, x_j^{k+1}],\ j \in N,\ i \in M; \\ t_i - t_{i-1} \geqslant \varepsilon,\ i \in P,\ t_0 = 0, \\ \quad t_p = 1-\varepsilon; \\ \sum\limits_{j=1}^{n} u_j(R_j) = 1,\ u_j(L_j) = 0,\ j \in N; \\ u_j(x_j^{k+1}) \geqslant u_j(x_j^k), \\ \quad k = 0, \ldots, \kappa_j - 1,\ j \in N. \end{cases} \tag{4}$$

2　建模与分析方法

依托基于阈值的 MCS 框架兼容混合指派偏好, 研究和构造分类模型分布空间, 通过 SMAA 方法量化分类不确定性, 逐步压缩分类模型分布空间以获得稳定的 MCS 结果和分类代表模型.

2.1　混合指派偏好下的分类模型分布空间构建

增量偏好提取是由决策分析师主导的交互问答式提取过程, 通过计算方案属于各类别的概率分布度量分类不确定性, 主动选择信息量最大的方案作为下一轮迭代的提问对象, 这一过程需要统计和分析所有的可行分类模型, 因此, 构建分类模型分布空间是实现增量偏好提取的重要前提. 混合指派偏好下的分类模型分布空间实际上是建立兼容的指派约束, 将多元结构的偏好嵌入同一模型框架之中, 使决策者可以用异构形式表达其指派偏好. 本文分类模型分布空间的构建过程分析如下.

确定型指派偏好. 若决策者可以提供方案的确定型类别, 记为

$$X^C = \{a_i \to C_{[a_i]} | a_i \in A^C, C_{[a_i]} \in C\}.$$

其中 A^C 为确定型指派的参考方案集, 参考文献 [12] 的方法, 根据分类阈值与价值函数唯一划分方案分类, 即约束集 E_{X^C}.

$$E_{X^C}\begin{cases} t_{[a_i]-1} \leqslant U(a_i) \leqslant t_{[a_i]} - \varepsilon, \\ [a_i] \in P; a_i \in A^C \end{cases} \tag{5}$$

其中, $[a_i]$ 表示 $a_i \in A^C$ 的类别序号.

不确定型指派偏好. 若方案类别无法唯一确定, 则记

$$X^U = \{a_i \to [C_{[La_i]}, C_{[Ua_i]}] | a_i \in A^U, C_{[La_i]}, C_{[Ua_i]} \in C\}$$

为不确定型指派偏好. 其中 A^U 为不确定型指派的参考方案集, 参考文献 [5] 的方法, 通过分类阈值和价值函数刻画方案的分类区间, 得到约束集 E_{X^U}.

$$E_{X^U}\begin{cases} t_{[La_i]-1} \leqslant U(a_i) \leqslant t_{[Ua_i]} - \varepsilon, \\ [La_i], [Ua_i] \in P; a_i \in A^U \end{cases} \tag{6}$$

其中, $[La_i]$ 和 $[Ua_i]$ 分别表示 $a_i \in A^U$ 的下界和上界类别序号.

成对比较型指派偏好. 当决策者只能给出方案的比较关系而无法确定具体类别时, 本文构建 $X^P = \{(a_i \rhd a_j, N_{ij}) | a_i, a_j \in A^P, N_{ij} \in \{0,\ 1,\ \ldots, p-1\}\}$, 刻画决策者的成对比较型指派偏好. 其中, $(a_i \rhd a_j, N_{ij})$ 表示 a_i 优于 a_j 的类别为 N_{ij}, A^P 为成对比较的参考方案集. 该偏好广泛存在于人才评级、学术评审等实际 MCS 场景中, 决策者只需比较方案间的类别差异, 可以大幅降低判断难度[1,5,21]. 由于决策者仅提供两个方案关于类别的差距数值, 同一组差距数值可能对应多种不同的方案类别组合, 无法直接采用价值函数和分类阈值直接识别. 为使分类模型分布空间可以同时兼容异构指派偏好, 本文通过指示变量和混合整数线性约束将成对比较指派转化为可嵌入价值函数的分析框架, 形成兼容三种异构偏好的分类模型分布空间.

对于成对比较型指派偏好 $(a_i \rhd a_j, N_{ij})$, 方案的具体类别未知, 但其分类相互关联. 如果可以预先给定 a_i 或 a_j 的类别, 则可以根据 N_{ij} 唯一确定另一方案的类别. 以 a_i 为例, 为了将 a_i 的类别归属转化为可计算的数学表达, 引入 p 个 0-1 变量 $\eta_{iq}, q = 1, \ldots, p$. 其中, $\eta_{iq} = 1$ 表示方案 a_i 属于类别 C_q, $\eta_{iq} = 0$ 表示 a_i 不属于 C_q. 为确保每个方案有且仅有唯一分类, 必须满足类别唯一性约束 $\sum\limits_{q=1}^{p} \eta_{iq} = 1$. 在此基础上, 可构建类别序号表达式 $\sum\limits_{q=1}^{p} q\eta_{iq}$. 由于只有其真实类别对应的 $\eta_{iq} = 1$, $\sum\limits_{q=1}^{p} q\eta_{iq}$ 的计算结果实际上等于 a_i 的类别序号 q. 显然, 若 $\eta_{iq} = 1$, 则 $a_i \to C_q$, a_i 的综合价值应当隶属于 C_q 的分类区间, 即 $t_{q-1} \leqslant U(a_i) \leqslant t_q - \varepsilon$. 同理, 若 $\eta_{iq} = 0$, 则 $t_{q-1} \leqslant U(a_i) \leqslant t_q - \varepsilon$ 不成立. 为此, 构造

$t_{q-1}-M(1-\eta_{iq})\leqslant U(a_i)\leqslant t_q-\varepsilon+M(1-\eta_{iq})$. 其中: M为充分大的正数, 保证a_i的价值仅隶属于一组相邻的分类阈值. 上述构造同样适用于方案a_j. 基于上述分析, $\sum\limits_{q=1}^{p}q\eta_{iq}$和$\sum\limits_{q=1}^{p}q\eta_{jq}$分别为$a_i$和$a_j$的所属类别序号, 又因$a_i$的类别比$a_j$高出$N_{ij}$, 可得

$$\sum_{q=1}^{p}q\eta_{iq}-\sum_{q=1}^{p}q\eta_{jq}=N_{ij}.$$

综上所述, 成对比较型指派偏好的分布空间可以表征成如下形式:

$$E_{X^P}\begin{cases}t_{q-1}-M(1-\eta_{iq})\leqslant U(a_i)\leqslant\\ \quad t_q-\varepsilon+M(1-\eta_{iq}), q\in P; a_i\in A^P\\ \sum\limits_{q=1}^{p}q\eta_{iq}-\sum\limits_{q=1}^{p}q\eta_{jq}=N_{ij},\\ a_i,a_j\in A^P;(a_i\rhd a_j,N_{ij})\\ \sum\limits_{q=1}^{p}\eta_{iq}=1, a_i\in A^P\\ \eta_{iq}\in\{0,1\}, q\in P; a_i\in A^P\end{cases}\tag{7}$$

E_{X^P}通过 0-1 指示变量和混合整数线性约束, 将原本无法直接嵌入价值函数框架的成对比较型指派偏好转化为价值函数与分类阈值的代数约束, 为统一建模混合指派偏好提供了条件.

构建混合指派偏好的分类模型分布空间, 需要确保价值函数与分类阈值能够复现决策者的混合指派偏好. 为此, 以基础模型约束E_{Base}为底层框架, 嵌入混合指派偏好的约束条件式 (5)-(7), 联立构成总约束集E_{Space}, 即所有可行分类模型的必要条件.

$$E_{Space}\{E_{Base},E_{X^C},E_{X^U},E_{X^P}.\tag{8}$$

如果E_{Space}非空, 意味着存在分类模型可以复现混合指派偏好, 所有满足E_{Space}的边际价值函数和分类阈值构成分类模型分布空间M_{Space}. 如果E_{Space}为空集, 说明初始的混合指派偏好存在矛盾, 需要进行一致性调整才能开展方案分类. 为此, 亟需建立有效的偏好冲突检测与消解机制. 本文基于混合指派偏好引入正向与负向误差变量$\varrho_i^{C+},\varrho_i^{C-},\varrho_i^{U+},\varrho_i^{U-},\varrho_i^{P+},\varrho_i^{P-}\geqslant 0$. 其中, $\varrho_i^{C+},\varrho_i^{C-},a_i\in A^C$为确定型指派偏好的正向与负向误差. $\varrho_i^{U+},\varrho_i^{U-},a_i\in A^U$为不确定型指派偏好的正向与负向误差. $\varrho_i^{P+},\varrho_i^{P-},a_i\in A^C$为成对比较型指派偏好的正向与负向误差. 由此建立考虑分类误差的混合指派偏好约束集$\tilde{E}_{Space}$, 根据E_{Space}决策变量的有界性, $\tilde{E}_{Space}$必然非空, 具体证明见定理 1.

$$\tilde{E}_{Space}\begin{cases}E_{Base}\\ t_{[a_i]-1}-\varrho_i^{C+}\leqslant U(a_i)\leqslant\\ \quad t_{[a_i]}-\varepsilon+\varrho_i^{C-},[a_i]\in P; a_i\in A^C\\ t_{[La_i]-1}-\varrho_i^{U+}\leqslant U(a_i)\leqslant t_{[Ua_i]}-\varepsilon+\\ \quad \varrho_i^{U-},[La_i],[Ua_i]\in P; a_i\in A^U\\ t_{q-1}-M(1-\eta_{iq})-\varrho_i^{P+}\leqslant U(a_i)\leqslant\\ \quad t_q-\varepsilon+M(1-\eta_{iq})+\varrho_i^{P-},\\ q\in P; a_i\in A^P\\ \sum\limits_{q=1}^{p}q\eta_{iq}-\sum\limits_{q=1}^{p}q\eta_{jq}=N_{ij},\\ a_i,a_j\in A^P;(a_i\rhd a_j,N_{ij})\\ \sum\limits_{q=1}^{p}\eta_{iq}=1,\eta_{iq}\in\{0,1\},\\ q\in P; a_i\in A^P\\ \varrho_i^{C+},\varrho_i^{C-},\varrho_i^{U+},\varrho_i^{U-},\varrho_i^{P+},\varrho_i^{P-}\geqslant 0\end{cases}\tag{9}$$

定理 1 $\tilde{E}_{Space}$必然非空.

证明过程略.

当$\varrho_i^{C+},\varrho_i^{C-},\varrho_i^{U+},\varrho_i^{U-},\varrho_i^{P+},\varrho_i^{P-}=0$时, 意味着$E_{Space}$成立. 因此, 可以通过误差最小化检验并量化偏好冲突. 由此, 构建混合整数线性规划 M1.

$$min\sum_{a_i\in A^C}(\varrho_i^{C+}+\varrho_i^{C-})+\sum_{a_i\in A^U}(\varrho_i^{U+}+\varrho_i^{U-})+\sum_{a_i\in A^P}(\varrho_i^{P+}+\varrho_i^{P-}).\tag{M1}$$

$$\text{s.t. }\tilde{E}_{Space}\tag{10}$$

根据上述分析, 如果 M1 的最优目标函数值为 0, 则E_{Space}非空, 决策者的初始混合指派偏好具有一致性. 否则, 决策者需要对指派偏好进行一致性调整. 此时可以将 M1 的分类误差反馈至决策者, 制定调整策略.

2.2 基于分类模型分布空间的分类不确定性度量

混合指派偏好分类模型分布空间中, 不同的分类模型可能将同一方案划分至多个类别, 导致分类具有不确定性. 当分类不确定性较高时, 方案的类别归属难以判定, 决策结果稳定性与可信度随之降低. 因此, 只有准确刻画分类模型的不确定性程度, 才能判断当前决策结果是否满足稳定性要求, 进而决定是否需要通过增量偏好提取压缩分类模型分布空间.

在分类模型分布空间M_{Space}中, 由于边际价值函数和分类阈值均为连续变量, 可行模型构成高维连续空间中的可行域, 通常包含无穷多元素. 因此, 无法通过枚举法直接计算每个方案的分类概率, 这为分类不确定性度量带来了挑战[11]. 现有的增量偏

好提取过程中, 常用的方法是利用 SMAA 获得足够的分类模型, 由此计算方案关于各个类别的分类概率[11]. SMAA 由 Lahdelma 等首次提出, 旨在处理决策问题中的准则不确定性问题[22]. 随后, Tervonen 与 Lahdelma 详细阐述了 SMAA 在各个领域的拓展应用, 显著增强了该方法的适用性[23]. 近年来, SMAA 与不同决策模型相融合, 成为处理决策不确定性的有效工具. 例如, Corrente 等[24] 将 SMAA 与 PROMETHEE 方法相结合, 基于权重不确定性评价结果的稳定性. Kadziński 等[11] 利用 SMAA 基于确定型指派偏好评估待分类方案的分类不确定性, 进而通过交互式增量提取保证 MCS 的稳定性. Ilbahar 等[25] 通过 SMAA 分析模糊信息和相关参数的不确定性. Pereira 等[26] 将 SMAA 与 TODIM 方法相结合分析决策中的理性不确定性, 通过蒙特卡洛模拟构建了一套鲁棒分类体系. 上述研究需要在参数空间内进行随机抽样, 由此统计决策的可接受性概率. 其中, 抽样方法主要包含两种: 第一, 在整个参数空间随机生成样本, 不满足约束则剔除; 第二, 首先生成一组满足条件的样本, 通过添加随机噪声生成新样本. 然而, 如果可行域的参数空间较小, 两种方法可能生成较多的无效样本, 降低计算效率. 这一问题本质上缺乏考虑随机抽样的边界条件. 例如, 假设某一准则的可行域为 [0.2,0.3], 如果在准则的原参数空间 [0,1] 进行随机抽样, 将导致大量样本无法满足约束条件. 因此, 本文利用优化模型构造可行域的外接超立方体, 通过边界约束提高 SMAA 抽样效率.

首先, 在E_{Space}的基础上分别计算分类模型参数的下限和上限, 获得每个参数的可行取值区间. 随后, 将随机抽样范围限制在这些区间构成的超立方体内, 超立方体的每个维度对应一个分类模型参数. 由于超立方体内部不再包含复杂的约束条件, 抽样过程只需在各个区间内生成随机数, 即可快速检验生成模型是否满足原约束E_{Space}. 具体而言, 即在E_{Space}中求解分类模型相关参数$u_j(x_j^k), j \in N; k = 0,1,\ldots,\kappa_j$和$t_i, i = 0,\ldots,p$的最大化与最小化问题 M2, 确定参数的可行取值区间.

$$\begin{aligned} &min/\max u_j(x_j^k)(j \in N; k = 0,\ldots,\kappa_j);\\ &min/\max t_i(i = 0,1,\ldots,p) \end{aligned} \tag{M2}$$

$$\text{s.t. } E_{Space} \tag{11}$$

设 M2 的最优解为$\min u_j(x_j^k) = u_j(x_j^k)^-$, $\max u_j(x_j^k) = u_j(x_j^k)^+$, $\min t_i = t_i^-, \max t_i = t_i^+$. 这些参数区间共同构成了$M_{Space}$的外接超立方体边界, 限定 SMAA 的抽样范围, 由此提高抽样效率. 如图 1 所示, 分类模型分布空间通过 M2 的最优解限制在一个超立方体内部, 所有分类模型参数分布其中, 超立方体的每一维度对应分类模型的参数取值. 随后, 在参数区间内抽取足够的分类模型, 为量化分类不确定性提供依据, 步骤包括:

1) 在$[u_j(x_j^k)^-, u_j(x_j^k)^+]$和$[t_i^-, t_i^+]$中利用均匀分布随机抽样生成一组满足$E_{Base}$的随机数;

2) 验证生成的随机数是否满足约束条件 (5)-(7);

3) 若满足则输出随机数构成的分类模型; 否则不予输出. 重复上述步骤直至获得H个分类模型. 记$U_k, T_k, k = 1, \ldots, H$为第$k$组价值函数和分类阈值, 组成第$k$个分类模型. 对于$U_k$和$T_k = \{t_{k0}, \ldots, t_{k(p-1)}, t_{kp}\}, k = 1, \ldots, H$, 通过以下规则获得所有分类模型关于方案的指派偏好$J(a_i) = [J_{kl}^{a_i}]_{H\times p}$.

$$J_{kl}^{a_i} = \begin{cases} 1, & t_{k(l-1)} \leqslant U_k(a_i) < t_{kl} \\ 0, & \text{其他} \end{cases} \tag{12}$$

其中, $J_{kl}^{a_i} = 1$表示第k个分类模型认为$a_i \to C_l$. 由此, 根据$J(a_i)$获得分类概率分布矩阵$CPD = [P(a_i \to C_q)]_{p\times m}$. 其中,

$$\begin{aligned} &P(a_i \to C_q) = \frac{\sum_{k=1}^{H} J_{kq}^{a_i}}{H}, q \in P; a_i \in A\\ &P(a_i \to C_q) \geqslant 0, \sum_{q=1}^{p} P(a_i \to C_q) = 1. \end{aligned} \tag{13}$$

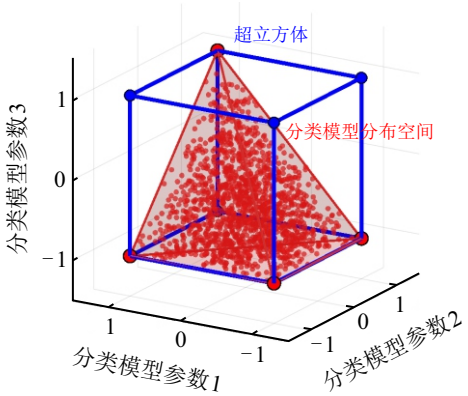


图 1　超立方体示意

考虑到分类概率的多种分布特征, 单一分类不确定性度量可能无法满足实际的决策需求. 为此, 本文通过信息熵、最小置信水平和边际置信水平定义分类不确定性度量, 分别表征分类概率的整体特征、极值特征和边际特征. 其中, 信息熵度量UN_i^{EN}综合考虑所有类别的概率分布, 反映分类不确定性的整体水平; 最小置信水平UN_i^{LC}关注最大分类概率, 反映决策者对方案最可能分类的把握程度; 边际置信

水平UN_i^{MC}关注最大分类概率与次最大分类概率的差距, 反映最可能类别与次可能类别的边际竞争程度. 此外, 在实际 MCS 中, 不同类别的误分类成本可能差异显著 (如“不合格”与“合格”的误判代价远高于“良好”与“优秀”), 忽略类别风险可能导致高风险类别的不确定性未被充分重视, 造成重大的决策损失[27]. 为此, 本文定义考虑类别权重的分类不确定性度量.

定义 1 记类别C_q的权重为λ_q, 满足$\lambda_q \geqslant 0$, $\sum\limits_{q=1}^{p}\lambda_q = 1$. 考虑类别权重的分类概率定义为

$$R(a_i \to C_q) = \frac{\lambda_q P(a_i \to C_q)}{\sum\limits_{q=1}^{p}\lambda_q P(a_i \to C_q)} \tag{14}$$

其中, $R(a_i \to C_q)$表示方案a_i考虑类别权重的分类概率. 不同于仅依据分类概率分布进行度量的方法, 定义 1 通过类别权重综合考虑分类概率与误分类风险. 当某一类别的误分类风险较高时, 赋予更大的权重, 保证对应的分类概率在后续不确定性度量中占据更大的作用. 由此, 如果两个方案具有相同的概率分布, 高风险类别对应的不确定性可以获得高度重视, 避免高风险类别被平均化处理. 当各类别权重相等时, 定义 1 退化为传统分类概率.

定义 2 对于方案 a_i 的分类概率$R(a_i \to C_q)$, $q = 1, \ldots, p$, 其信息熵度量 UN_i^{EN} 定义为

$$UN_i^{EN} = \frac{\sum\limits_{q=1}^{p} R(a_i \to C_q) log(R(a_i \to C_q))}{-\log p}. \tag{15}$$

当分类概率均匀分布于各个类别时, 信息熵取最大值 1, 表示分类不确定性最大; 当某一类别的分类概率为 1、其余为 0 时, 信息熵取最小值 0, 表示分类完全确定.

定义 3 记$\max\limits_{q}\{R(a_i \to C_q)\}$为$a_i$最高分类概率, 则最小置信水平$UN_i^{LC}$定义为

$$UN_i^{LC} = 1 - \max_{q}\{R(a_i \to C_q)\}. \tag{16}$$

最高分类概率越低, 表示决策者对方案归属的把握越小, 分类不确定性越高. 最高分类概率为$1/p$时, UN_i^{LC}达到最大值 $1 - 1/p$; 最高分类概率为 1 时, $UN_i^{LC} = 0$, 方案分类唯一确定.

定义 4 记$R(a_i \to C_q)_1$和$R(a_i \to C_q)_2$分别为a_i分类概率的最大值和次最大值, 则边际置信水平UN_i^{MC}定义为

$$UN_i^{MC} = 1 - (R(a_i \to C_q)_1 - R(a_i \to C_q)_2). \tag{17}$$

最大概率与次最大概率相差较大时, 类别归属更加明确, 不确定性较低; 若两者接近, 方案对于两个类别的归属模糊, 不确定性较高. 如果$R(a_i \to C_q)_1 = R(a_i \to C_q)_2$, 有$UN_i^{MC} = 1$, 难以区分$a_i$的类别归属; 当$R(a_i \to C_q)_1 = 1, R(a_i \to C_q)_2 = 0$时, $UN_i^{MC} = 0$, a_i唯一划分至$R(a_i \to C_q)_1$对应的类别.

定理 2 不确定性度量的三个指标满足有界性 $0 \leqslant UN_i^{EN}, UN_i^{MC} \leqslant 1, 0 \leqslant UN_i^{LC} \leqslant 1 - \frac{1}{p}$.

证明过程略.

UN_i^{EN}反映分类概率的整体分散程度, 适用于判断方案分类的整体模糊性, 但是无法判断分类概率是否存在竞争关系. UN_i^{MC}恰好弥补这一不足, 通过最可能分类与次可能分类的概率差距, 捕捉分类的竞争程度. UN_i^{LC}关注最大分类概率的绝对水平, 直接反映决策者对于最可能类别的把握程度. 三者结合使用, 能够全面捕捉分类概率的分布特征, 为混合指派偏好下的增量问题选择提供多种依据.

记 a_i分类概率为 $(R(a_i \to C_1), \ldots, R(a_i \to C_p))$, 若$R(a_i \to C_1) = \ldots = R(a_i \to C_p)$, 很难判断$a_i$分类, 分类不确定性最大, 代入可得$UN_i^{EN} = UN_i^{MC} = 1, UN_i^{LC} = 1 - \frac{1}{p}$. 当$R(a_i \to C_j) = 1$, $R(a_i \to C_k) = 0, k \neq j$时, 可以唯一确定$a_i$分类, 分类不确定性最小, 此时 $UN_i^{EN} = UN_i^{MC} = UN_i^{LC} = 0$. 因此, 上述度量可以直观且合理地反映分类不确定性.

2.3 考虑类别密度的分类代表模型确定

为提高决策效率, 增量偏好提取往往遵循最大信息量原则, 方案不确定性越大则信息量越大, 决策者通过判断该方案指派保证增量偏好提取的每次迭代, 可以最大程度降低待分类方案的不确定性[12]. 本文选择最大信息量方案a_{max}作为混合指派偏好的增量问题, 即

$$IM_Y(a_{max}) = \arg\{\max_{a_i \in A}\{UN_i^Y\}\}. \tag{18}$$

其中, $Y \in \{EN, LC, MC\}$分别表示通过信息熵、最小置信水平和边际置信水平度量分类不确定性. 在每一次迭代中, 决策分析师根据当前方案分类不确定性最大化原则, 筛选出信息量最大的方案 a_{max}. 之后, 询问决策者“您认为a_{max}如何分类”, 决策者根据认知自由选择以下形式之一提供指派偏好, 确定型指派偏好: 指定 a_{max} 所属的单一类别; 不确定型指派偏好: 指定 a_{max} 的类别区间; 成对比较型指派偏好: 将 a_{max} 与其它方案进行比较, 给出两者的

类别差距.

通过增量偏好提取, 可以在 MCS 中逐步挖掘决策者的混合指派偏好, 减小分类模型不确定性. 当 MCS 不确定性足够小时, 即可终止混合增量偏好提取过程, 由此设定以下终止条件:

$$min\{\max_{q\in P}\{R(a_i \to C_q)|a_i \in A\}\} \geqslant \beta. \tag{19}$$

其中, $\beta \in [0,1]$表示决策分析师设定的不确定性可接受阈值. 此外, 考虑到决策者认知的有限性, 在实际的 MCS 问题中无法持续提供增量偏好, 因此, 增量提取的整体迭代次数也可作为终止条件[12]. 当混合增量偏好提取过程满足终止条件时, 可以根据最终的分类概率分布获得稳定的 MCS 结果:

$$\begin{aligned} &C_{[a_i]} = \arg\{max_{q\in P}\{R(a_i \to C_q)\}\}, \\ &a_i \in A; C_{[a_i]} \in C. \end{aligned} \tag{20}$$

经过上述混合增量偏好提取过程, 分类模型分布空间逐渐缩小, 分类模型不确定性逐渐减小. 然而, 压缩后的M_{Space}中仍然存在多个可行的分类模型, 均能复现 MCS 结果. 为从这些模型中确定一个代表模型用于解释分类机理、支撑管理决策, 本文提出分类代表模型的选择方法. 现有方法大多通过最大化类别之间的最小差异ε计算 MCS 代表模型[17]. 然而, 此类方法主要关注分类阈值的分布差异, 忽视了类别内的方案分布特征, 可能导致分类阈值与方案分布不匹配. 例如, 某类别间隔过大却仅包含极少数方案, 而另一类别间隔过小却包含大量方案. 为解决上述问题, 本文定义类别密度以及类别密度分布方差.

定义 5　设$T = \{t_0, \ldots, t_p\}$为分类阈值, 类别C_q中方案数为n_q, 则类别密度定义为

$$\rho_q = \frac{n_q}{t_q - t_{q-1}}. \tag{21}$$

其中, ρ_q表示单位分类区间中方案的分布密集程度. 由此, 进一步定义类别密度分布方差为

$$\sigma(\rho) = \frac{1}{p}\sum_{q=1}^{p}(\rho_q - \frac{1}{p}\sum_{q=1}^{p}\rho_q)^2. \tag{22}$$

$\sigma(\rho)$主要考虑类别中方案的实际分布, 保证单位分类区间内的方案分布更加均衡. 当某一类别包含较多方案时, 其对应分类区间将适当扩大. 如果类别中方案较少, 分类区间相应缩小. $\sigma(\rho)$越小, 意味着$\rho_q, q = 1, \ldots, p$更加接近, 由此在类别密度均衡性的基础上保证类别区间与类别内的方案分布适配.

基于上述思想, 本文构建模型 M3 选择分类代表模型. 在增量偏好提取结束后, 从分类模型分布空间中选出一组可复现 MCS 结果的价值函数与分类阈值, 使其可以最小化类别密度分布方差, 从而确保分类阈值的均衡性和类别密度分布的合理性.

$$\min\sigma(\rho) \tag{M3}$$

$$\text{s.t. } E_{Base} \tag{23}$$

$$t_{[a_i]-1} \leqslant U(a_i) \leqslant t_{[a_i]} - \varepsilon, a_i \in A \tag{24}$$

其中, 约束 (23) 刻画分类阈值和价值函数的有序性和单调性. 约束 (24) 保证方案的最终分类符合增量偏好提取结果. 模型 M3 为非线性规划. 求解模型 M3, 可以获得分类阈值$t_i, i = 0, 1, \ldots, p$和价值函数$U(\cdot)$, 二者共同构成 MCS 的代表模型, 进而可以定量分析决策者的偏好认知.

计算复杂度主要源于 SMAA 的随机抽样、混合整数线性规划问题 M2 与非线性规划问题 M3. 其中, 单次随机抽样包括$n\kappa_j$个特征点与$p-1$个分类阈值. 根据 Tervonen 与 Lahdelma 的研究[23], H个分类模型的计算复杂度为$O(Hn\kappa_j(p-1))$.M2 包含$p-1+n\kappa_j$个连续变量与$p|A^P|$个整数变量.M3 包含$p-1+n\kappa_j$个连续变量. 决策规模较小时, 可以通过 Cplex、Gurobi、Lingo 等软件快速求解. 若决策规模较大, 可以通过智能优化算法提高求解效率. 图 2 为本文方法的流程图.

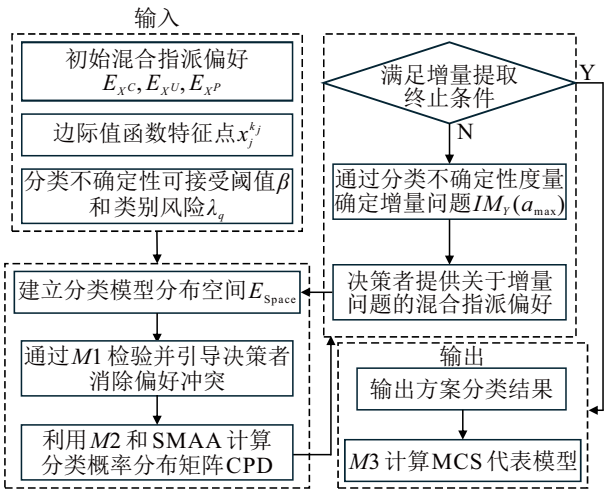


图 2　本文方法流程

3　算例分析

3.1　算例背景

MBA 项目分级评估可以帮助院校明确差异化优势、优化教育资源配置, 推动商科教育质量整体提升, 文献 [28] 研究了 27 所大学的 MBA 项目分类问题 (C_1: 较差、C_2: 一般、C_3: 优秀、C_4: 非常优秀), 各类别权重相等为 0.25, 通过《金融时报》年报数据 (http://rankings.ft.com/rankings/mba/rankings.html) 设定了三个效益型指标 (g_1: 校友职业发展、g_2: 多样

性、g_3: 创意生成). 参考文献 [28], 这些项目分类依次为(4, 1, 1, 1, 3, 3, 3, 2, 1, 3, 2, 3, 4, 1, 1, 2, 2, 2, 1, 1, 1, 2, 2, 1, 2, 1, 3).

本节采用信息熵度量分类不确定性, 通过上述MCS 问题说明本文方法计算过程.

为了直观展示本文方法过程, 随机抽取 5 个方案作为确定型指派偏好, 结合对应类别获得$X^C = \{a_1 \to C_4, a_2 \to C_1, a_6 \to C_3, a_{23} \to C_2, a_{27} \to C_3\}$. 抽取 1 个方案, 通过引入相邻类别误差作为不确定型指派偏好$X^U = \{a_4 \in [C_1, C_2]\}$. 抽取 2 个方案, 根据分类差异作为成对比较型指派偏好$X^P = \{(a_{13} \rhd a_{22},\ 2)\}$. 决策分析师设定边际价值函数的特征点数量为 2, 终止条件中不确定性可接受阈值$\beta = 1$

3.2 过程分析

由于 M1 的最优目标函数值为 0, 混合指派偏好具有一致性. 在第一次迭代中, 非参考方案a_8, a_{18}和成对比较参考方案a_{13}, a_{22}具有唯一分类. 由此, 第一次迭代消除了$(a_{13} \rhd a_{22}, 2)$分类不确定性. 根据 M2, 可以在分类模型分布空间中计算边际价值函数的形状参数和分类阈值取值范围, 分别为$x_1^0 = x_2^0 = x_3^0 = 0$、$x_1^1 \in [0, 0.278]$、$x_2^1 \in [0.713, 1]$、$x_3^1 \in [0, 0.187]$、$t_1 \in [0.097, 0.255]$、$t_2 \in [0.241, 0.472]$和$t_3 \in [0.500, 0.831]$. 根据$EN$策略, $a_{10} \in A^N$具有最大信息量 0.499, 说明第一次迭代中a_{10}分类不确定性最大, 决策分析师询问决策者关于a_{10}的指派偏好. 假设决策者无法准确判断, 选择提供不确定型指派偏好$a_{10} \to [C_3, C_4]$. 将$a_{10} \to [C_3, C_4]$与当前X^U合并, 检查分类模型分布空间非空, 更新不确定型指派偏好. 持续迭代, 直至获得满足终止条件的 MCS 结果.

本文方法经过 8 次迭代, 分类不确定性逐渐减小直至为 0, 获得稳定的决策结果, 所有方案类别唯一确定. 方案具体分类结果见表 1. 其中, a_1, a_{10}, a_{13}的 MBA 项目评价为“C_4: 非常优秀”, 其课程设置与教育理念可以作为优秀模板进行推广, 促进商科教育发展.

表 1 MCS 结果

类别	C_1	C_2	C_3	C_4
方案	$a_2, a_3, a_4, a_9, a_{14}, a_{15}, a_{19}, a_{20}, a_{21}, a_{24}, a_{26}$	$a_8, a_{11}, a_{17}, a_{18}, a_{22}, a_{23}$	$a_5, a_6, a_7, a_{12}, a_{16}, a_{25}, a_{27}$	a_1, a_{10}, a_{13}

为了解释分类机理、支撑管理决策, 根据MCS 结果通过 M3 获得 MCS 的代表模型. 其中, 边际价值函数的形状参数分别为$x_1^0 = x_2^0 = x_3^0 = 0$、$x_1^1 = 0.0768$、$x_2^1 = 0.8687$和$x_3^1 = 0.0545$. 分类阈值分别为$t_1 = 0.2247$、$t_2 = 0.3495$和$t_3 = 0.7012$.

3.3 结果与比较分析

3.3.1 关于特征点的数量

特征点用以控制边际价值函数的形式和分布, 随着其数量增加, 分类模型的灵活性逐渐提高. 本节假设所有准则的边际价值函数具有相同数量且分布均匀的特征点, 记为$\kappa_j = k \in \{2, 3, 4\}$. 此外, 为了保证分析的有效性, 假设决策者的混合增量指派偏好与表 B2 一致, 信息量度量策略为 EN. 由此通过 M3 获得 MCS 的综合价值和分类阈值, 结果见图 3.

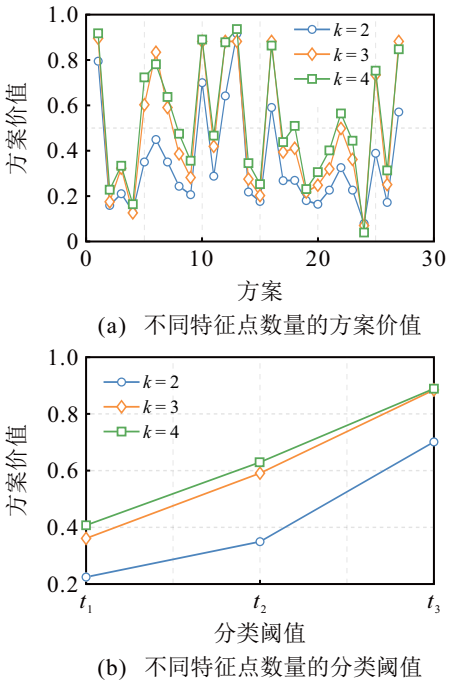


图 3 不同特征点数量下的 MCS 结果

通过图 3 可知, 随着边际价值函数的特征点数量改变, 方案的综合价值和分类阈值存在差异. 显然, 相比于$k = 2$, $k = 3$和$k = 4$关于方案的综合价值和分类阈值具有更高的相似性. 根据各个准则的边际价值函数 (图 4), $k = 3$和$k = 4$关于所有准则的边际价值函数同样具有明显的相似性. $k = 2$和$k = 3$时, 方案性能的平均差异为 0.1427, $k = 2$和$k = 4$的差异为 0.1817, $k = 3$和$k = 4$的差异为 0.0504. 此外, 通过计算$k \in \{5, \ldots, 10\}$, 方案的综合

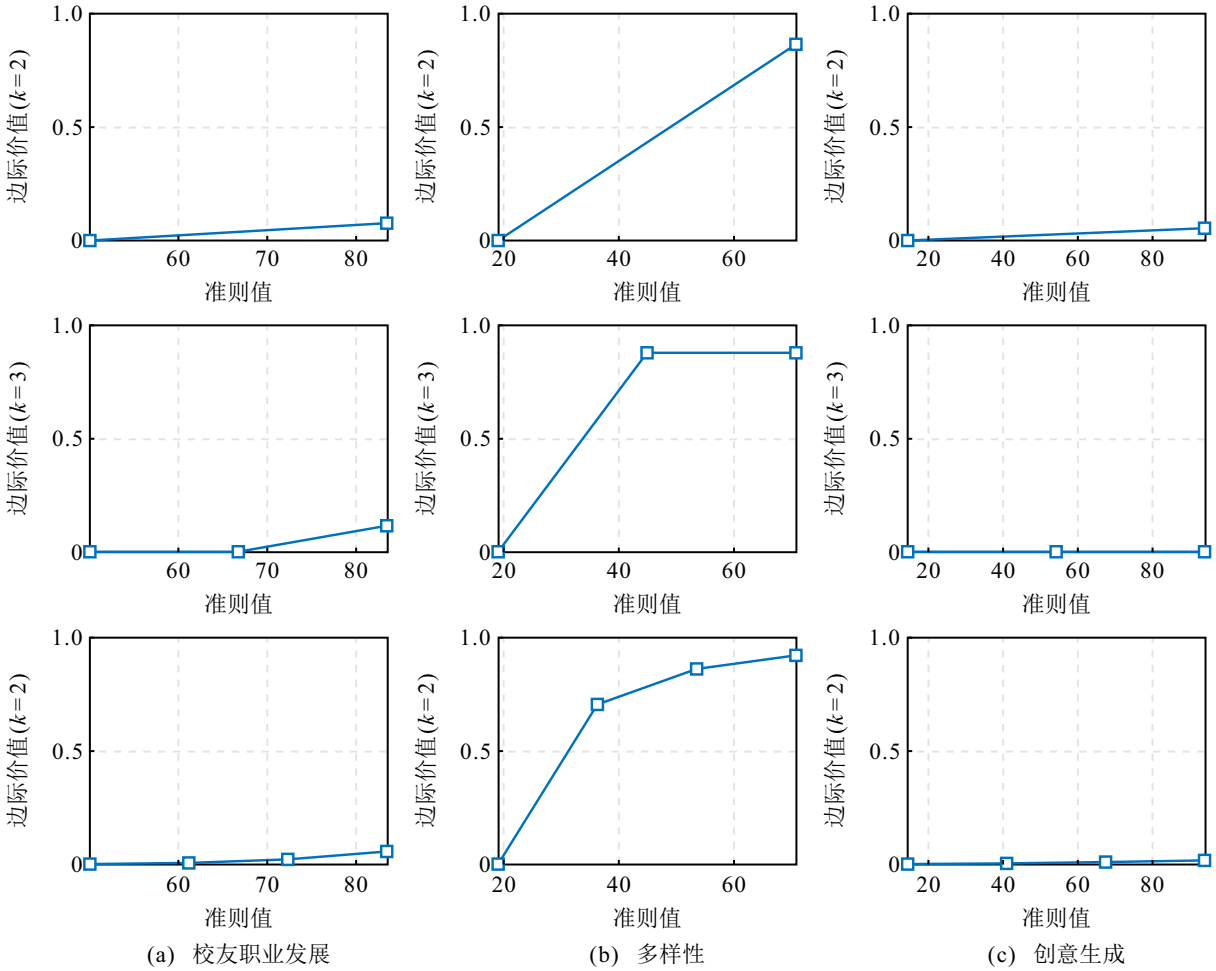


图 4 准则关于不同特征点的边际价值函数

价值和分类阈值差异都接近于 0. 上述结果表明, 随着特征点数量增加, 分类结果趋于稳定, 反映出本文方法具有稳定性.

3.3.2 不同信息量度量策略的结果

为了讨论信息量度量策略对于 MCS 的影响, 当决策分析师询问某一方案的分类时, 假设决策者根据分类概率分布矩阵中方案的最大分类概率所属类别进行回答, 如果存在多个最大概率, 则通过不确定型指派偏好进行回复. 不确定性可接受阈值保持 $\beta = 1$. 迭代过程见表 2.

根据表 2, 三种策略的迭代次数相同. 然而, 增量问题的提取顺序不同, 分别为

$$EN: a_{10} \to a_{25} \to a_5 \to a_{21} \to a_{12} \to a_{17} \to a_3$$

$$LC: a_{10} \to a_{25} \to a_{12} \to a_{21} \to a_5 \to a_{17} \to a_3$$

$$MC: a_{10} \to a_{25} \to a_{12} \to a_5 \to a_{21} \to a_{17} \to a_3$$

上述结果说明a_5, a_{12}和a_{21}分类不确定性相近, 不同的信息量度量导致其不确定性排序存在差异. 为了讨论不同策略在多场景中的适用性, 参照算例的初始指派偏好抽样方法, 分别随机抽取 7 个确定型指派偏好、7 个不确定型指派偏好和 7 个成对比较型指派偏好. 针对三种情景, 分别利用 EN、LC 和 MC 策略进行混合增量偏好提取. 为保证对比的有效性, 增量提取的方案指派与其真实类别相同. 上述试验重复 100 次, 图 5 为不同情景下三种策略的迭代次数.

根据图 5, 对于确定型指派偏好,LC 策略的平均迭代次数最少, 为 6.15 次. 对于不确定型指派偏好, EN 策略的平均迭代次数最少, 为 11.02 次. 对于成对比较型指派偏好,LC 策略的平均迭代次数最少, 为 8.53 次. 由于 EN 策略能够根据方案对于所有类别的分类概率捕捉其分类不确定性, 适用于整体分

表 2 不同信息量度量的增量偏好提取过程

迭代	最大信息量		
	EN	LC	MC
1	0.499	0.476	0.946
2	0.500	0.496	0.966
3	0.509	0.473	0.853
4	0.498	0.443	0.793
5	0.499	0.293	0.920
6	0.380	0.220	0.440
7	0.361	0.196	0.393
8	0	0	0

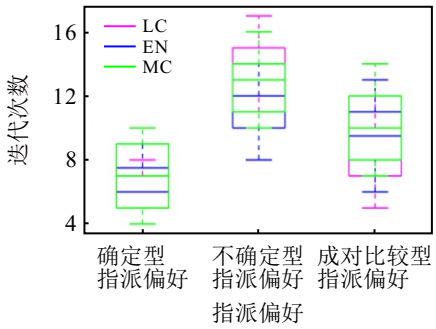


图 5 不同信息量度量和指派偏好的迭代次数

类模糊的 MCS 问题. 例如, 不确定型指派中一个方案可能隶属于多个类别,EN 策略对于不确定性的整体刻画与此类情境相适配, 迭代次数最少. 由于 LC 策略能够在所有方案的最高分类概率中直接识别置信度最低的方案, 该策略适用于需要保证每个方案都具有较高分类置信度的决策场景, 如医疗诊断、安全评估等高风险决策问题. 由于此类问题的高风险特征, 方案分类具有更加明确的约束, 其最大分类概率普遍较高. MC 策略适用于相邻类别难以区分的决策情境, 其中, 最可能类别与次可能类别之间存在竞争, 如企业信用评级中的相邻等级难以区分. 该策略通过计算分类概率的边际差异捕捉不确定性最大的方案.

在实际的 MCS 问题中, 决策初期由于分类信息较少, 方案的整体分类不确定性较高. 随着增量提取持续进行, 分类模型分布空间逐渐缩小, 方案的分类逐渐收敛于部分类别. 由此, 增量提取初期可以通过 EN 策略快速减小方案关于整体类别的不确定性, 之后利用 LC 或 MC 策略聚焦于方案分类的局部特征, 提高迭代效率. 为验证组合策略的计算效率, 基于本节不确定型指派偏好试验, 设置 EN+LC 与 EN+MC 策略. 增量提取前$G \in \{0,1,\ldots,20\}$次迭代采用 EN 策略, 后续采用 LC 或 MC 策略, 每种G取值进行 100 次重复试验, 统计不同G下的平均迭代次数.

根据图 6, EN+LC 与 EN+MC 策略的平均迭代次数随着G的改变呈现出先减后增的趋势, 都在$G=9$时达到最小, 分别为 10.51 和 10.85. 因此, 前

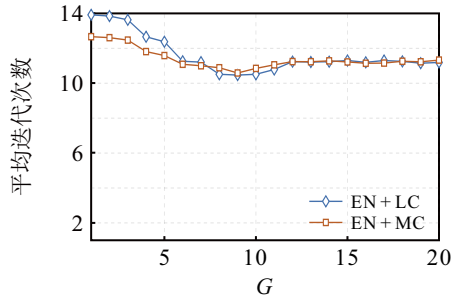


图 6 不同信息量组合策略的平均迭代次数

9 次增量提取过程中方案的整体分类不确定性较大, 需要 EN 策略减小整体分类不确定性. 之后, 方案分类集中于部分类别, LC 和 MC 策略可以提高计算效率.

3.3.3 关于不确定性可接受阈值

β影响方法的迭代终止条件. 理论上, β值越大, 表明决策分析师对于分类结果具有更高的稳定性要求, 为达到这一更严格的标准, 往往需要进行多次迭代, 逐步压缩分类模型分布空间, 提高方案分类的唯一性. 为了具体量化决策分析师的可接受心理对于本文方法的影响, 预设$\beta \in \{0.1,0.2,\ldots,1\}$, 分析方法的迭代次数差异, 结果见图 7.

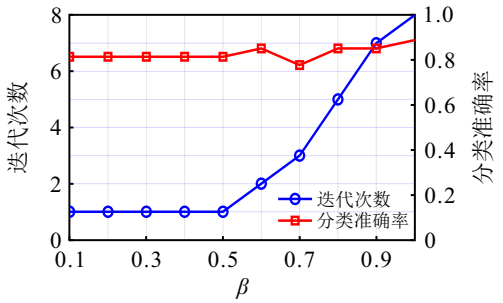
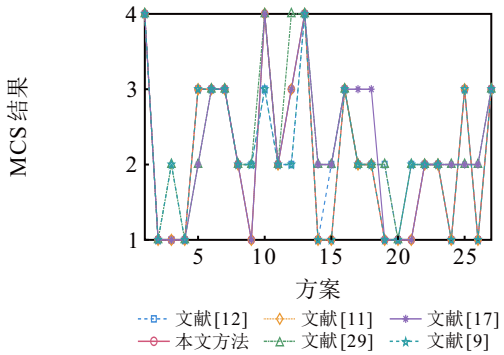


图 7 不同β下本文方法迭代次数与分类准确率

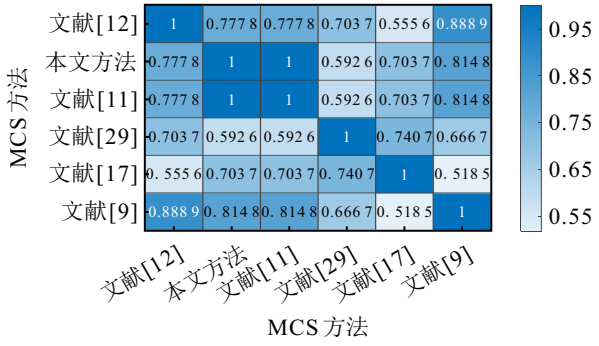
根据图 7, 随着β增大, 迭代次数依次增大. 如果决策分析师设定不确定性可接受阈值β较高, 意味着追求稳定性较高的方案分类, 必须通过多轮增量偏好提取持续压缩分类模型分布空间, 直至所有方案的最大分类概率达到或超过β. 特别地, 当$\beta \leqslant 0.5$时, 仅一次迭代达到本文方法的终止条件. 经过计算, 第一次迭代$min\{\max\limits_{q\in P}\{PRO(a_i \to C_q)|a_i \in A\}\} = 0.5277$, 意味着即使是当前分类不确定性最大的方案, 其对于某一类别的概率也达到或超过β, 该分类结果可以被直接采纳. 分类准确率并非随着β的增大而递增, $\beta = 0.6$的分类准确率为 0.8519, 而$\beta = 0.7$的分类准确率下降为 0.7778. 根据表 B2, 随着β增大, 决策者需要逐步提供更多的混合指派偏好降低分类不确定性. 然而, 实际问题中决策者的知识能力往往有限, 无法完全保证其所有增量偏好与真实分类相符. 当$\beta = 0.7$时, 决策者认为$a_{25} \to C_3$, 而a_{25}的真实分类为$a_{25} \to C_2$. 认知偏差导致分类代表模型偏离真实分类结果, 直至后续迭代过程中决策者不断提供与真实分类相符的偏好, 分类准确率持续上升. 上述现象在实际决策中普遍存在, 说明决策者的认知专业性对于确定合理的分类代表模型至关重要. 因此, 可以在增量提取中建立偏好更新机制, 引导决策者主动纠正认知偏差, 提高 MCS 结果的可靠性.

3.3.4　不同方法的比较分析

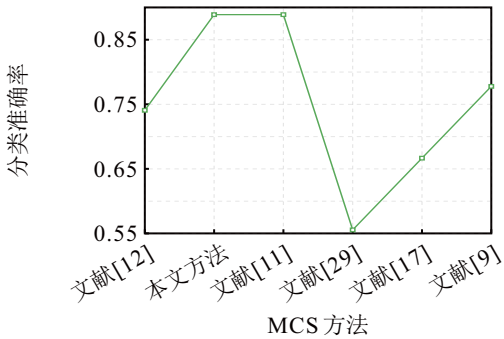
偏好增量提取和批量提取是解决 MCS 的主要方法, 为讨论两类方法的差异性, 本节利用现有的增量提取方法[9,11-12] 和批量提取方法[17,29] 计算本文算例. 为保证比较的有效性和全面性, 首先定义分类相似度与准确率. 然后, 计算各个方法的 MCS 结果、分类相似度和分类准确率, 结果见图 8.



(a)　不同方法的MCS 结果



(b)　不同方法的分类相似度



(c)　不同方法的分类准确率

图 8　不同方法 MCS 结果、相似度和准确率

定义 6　假设$\{h_{11},\ldots,h_{1m}\},h_{1i}\in P$和$\{h_{21},\ldots,h_{2m}\},h_{2i}\in P$分别为方法$\Lambda_1$和$\Lambda_2$的 MCS 结果, 两种方法的分类相似度定义为

$$S(\Lambda_1,\Lambda_2)=\frac{\sum_{i=1}^{m}z_i}{m},z_i=\begin{cases}1, & h_{1i}=h_{2i}\\ 0, & \text{其他}\end{cases}\tag{25}$$

定义 7　假设 MCS 真实分类为$\{\overline{h_1},\ldots,\overline{h_m}\}$, $\overline{h_i}\in P$, 方法Λ_1的 MCS 结果为$\{h_{11},\ldots,h_{1m}\}$, $h_{1i}\in P$, 方法Λ_1的分类准确率定义为

$$Acc(\Lambda_1)=\frac{\sum_{i=1}^{m}f_i}{m},f_i=\begin{cases}1, & h_{1i}=\overline{h_i}\\ 0, & \text{其他}\end{cases}\tag{26}$$

根据图 8, 不同方法的 MCS 结果 (图 8(a))、分类相似度 (图 8(b)) 和分类准确率 (图 8(c)) 存在差异. 根据图 8(a) 与图 8(b), 本文方法与文献 [11] 的分类结果一致, 与文献 [9] 和 [12] 的相似度分别为 81.48% 和 77.78%, 整体处于较高水平. 据此, 尽管不同增量偏好提取方法存在差异, 但本文方法计算结果与主流增量提取方法保持了较高一致性, 验证了其在分类趋势上的合理性. 相比之下, 本文方法与基于批量提取的文献 [29] 和 [17] 的相似度分别为 59.26% 和 70.37%, 差异相对较大. 显然, 是否采用增量偏好提取策略对 MCS 结果具有显著影响, 增量提取在匹配决策者认知过程和动态调整分类模型方面具有独特优势. 根据图 8(c), 本文方法与文献 [11] 的分类准确率最高, 均为 88.89%; 文献 [12]、[29]、[17] 和 [9] 的准确率分别为 74.07%、55.56%、66.67% 和 77.78%. 因此, 基于增量偏好提取的方法在分类准确率上优于基于批量提取的方法, 说明增量提取策略能够通过决策者偏好的更新迭代, 逐步获得精准且稳定的 MCS 结果. 与文献 [11] 相比, 两个方法在准确率上相当, 但本文方法可以兼容多种混合指派偏好, 能够适应复杂的 MCS 场景, 具有更强的普适性.

3.3.5　关于初始偏好数量与偏好构成

本节分析算例中的初始偏好数量以及偏好构成对于本文方法迭代次数和分类准确率的影响. 以表 1 为基础, 记初始偏好数量$S\in\{2,\ldots,10\}$, 随机抽取S个初始偏好$\{a_1,\ldots,a_S\}$及其分类$\{[a_1],\ldots,[a_S]\}$, 作为确定型指派偏好. 对于不确定型指派偏好, 将S个初始偏好的分类随机向上或向下扩展一类. 对于成对比较型指派偏好, 逐一选取$a_i\in\{a_1,\ldots,a_S\},i=1,\ldots,S$, 通过$a_j\in\{a_1,\ldots,a_S\}/a_i$获得成对比较关系. 增量提取终止条件$\beta=1$, 增量提取方案类别与表 1 一致, 分类不确定性度量为 EN. 针对各种初始偏好数量与构成, 计算 100 次试验的平均迭代次数, 结果见图 9.

根据图 9, 三种指派偏好的迭代次数差异显著. 随着初始偏好增加, 由于确定型指派偏好本身具有唯一指派, 平均迭代次数最少, 与其他偏好类型的平均迭代差异为 2.368 6 和 2.0858. $S\in\{2,3,4\}$时, 不确定型指派偏好的迭代次数显著少于成对比较型指派偏好, 但是$S\geqslant 5$时反之. 分析如下: 成对比较偏好的分类不确定性表现为两个方案类别组成的多样性,

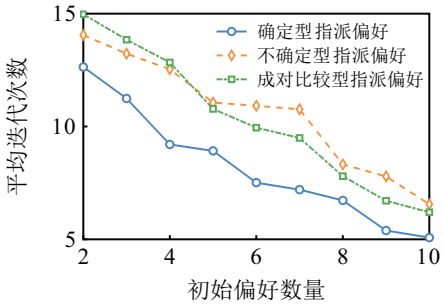


图 9 不同初始偏好数量与构成的迭代次数

不确定型指派偏好的不确定性表现为方案的分类区间, 由于偏好数量较少, 方案之间的相互制约较少, 类别组成的多样性相比于分类区间表现出更强的灵活性, 由此产生更大的分类不确定性, 平均迭代次数更多; 随着初始偏好增加, 成对比较偏好相比于不确定型指派偏好具有更强的相互影响, 大大降低了类别组成的多样性, 分类区间的不确定性显著上升.

3.4 数值试验

3.4.1 计算效率分析

为实现 SMAA 高效抽样, 通过模型 M2 求解所有边际价值函数参数$u_j(x_j^k)$和分类阈值t_i的上下界, 本质是混合整数线性规划问题, 在大规模决策问题中, 准则数或者类别数通常较大,M2 的求解复杂度持续上升. 为分析本文方法的计算效率与扩展性, 本节设计以下模拟试验: 设置准则数$n \in \{10,11,\ldots,30\}$、类别数$p \in \{5,6,\ldots,25\}$, 随机生成$10 \times p$个方案, 相关的准则性能$v_{ij} \in [0,1]$, 随机生成准则权重$w_j \geqslant 0, \sum_{j=1}^{n} w_j = 1$, 根据方案得分$\sum_{j=1}^{n} w_j v_{ij}$保证每一类别包含 10 个方案, 每一类别随机抽取 5 个方案作为确定型指派偏好和不确定型指派偏好, 随机抽取两个类别所有方案作为成对比较型指派偏好. 设置特征点数量为 2, 统计 M2 在不同决策场景的计算时间.

根据图 10, 随着决策规模扩大, M2 计算时间不断增加, 整体表现为指数上升. 在准则数$n \leqslant 30$且类别数$p \leqslant 25$的情况下, 最大求解时间为 746.5776s,

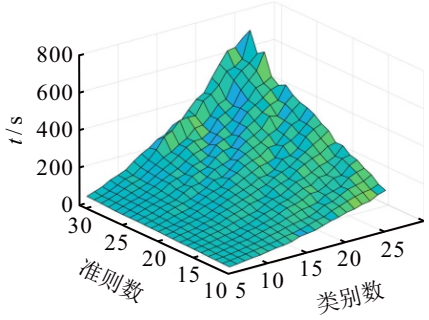
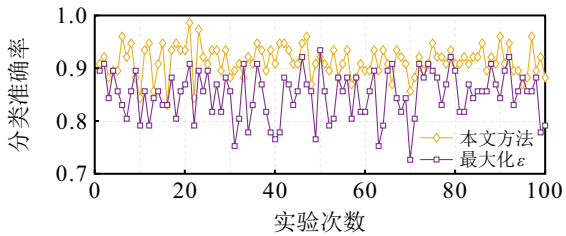


图 10 M2 的计算时间

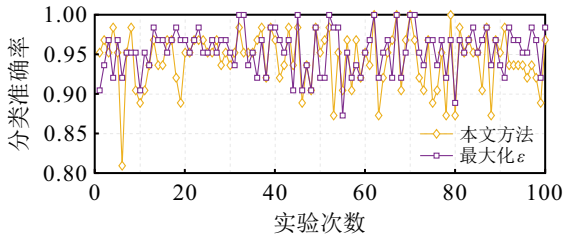
处于可接受范围, 本文方法对于大规模问题具有一定的适配性与扩展性. 此外, 类别数量的增加相比于准则数量的增加对于计算效率的影响更加显著. 以同等规模为例, 当$p = 25$时, 准则数量由 10 变为 30, 求解时间相差 520.7426s. 当$n = 30$时, 类别数量由 5 增加至 25, 求解时间相差 711.2727s. 因此, 在实际的大规模 MCS 问题中, 可以优先对类别数量进行压缩或分层处理, 从而提高本文方法的计算效率.

3.4.2 代表模型分析

为了分析代表模型的可行性与优越性, 本节通过真实数据集计算最大化相邻阈值差异ε与本文方法的分类准确率. 数据集 1(10.24432/C5231X) 包含 5 个准则、4 个类别不同类别的方案分布具有不均衡特征. 数据集 2(https://archive.ics.uci.edu/ml/index.php) 包含 6 个准则、4 个类别, 类别之间的方案分布相似. 为保证比较有效性, 设置特征点数为 2. 针对同一数据集, 各类别随机抽取 70% 方案作为训练集, 剩余 30% 方案作为测试集, 将训练集及其类别代入模型 M3, 分别通过目标函数$\min\sigma(\rho)$和$max\varepsilon$获得价值函数与分类阈值, 进而计算测试集分类, 对比真实分类统计两种方法的分类准确率, 结果见图 11.



(a) 数据集1



(b) 数据集2

图 11 不同方法的分类准确率

根据图 11, 两种方法的分类性能在不同决策场景下差异显著. 对于数据集 1, 本文方法显著优于ε最大化方法, 在 100 次随机试验中, 本文方法在 95% 的情况下分类准确率更高, 两者的平均分类性能相差 6.56%. 对于数据集 2, 本文方法与传统方法性能接近. 在 100 次随机试验中, 本文方法与ε最大化方法分别在 47% 和 53% 的情况下具有更高的分类准确率. 综上所述, 由于本文方法考虑了不同类别的方案分布特征, 可以有效反映类别中的方案数量

对于分类代表模型的影响, 相比于现有方法普适性更优.

4 结 论

本文提出了一种考虑混合指派偏好的增量偏好提取方法, 允许决策者通过更加自由的方式提供关于方案的指派. 该方法可以兼容多种指派偏好的不确定性, 能通过多种策略量化分类不确定性对于MCS 的影响. 对于不同的分类不确定性度量策略, LC 策略适用于确定型指派偏好或成对比较型指派偏好较多的情况. 当不确定型指派偏好较多时, EN 策略展现出更优的适用性. 本文通过确定型、不确定型和成对比较型指派偏好构建统一建模框架. 然而, 实际决策中的偏好形式可能更为复杂, 需要进一步融入分类边界约束、类别基数限制等异构信息[7], 增强方法普适性. 同时, 现实决策往往涉及多专家群体, 如何将当前单决策者框架拓展至群体协同环境值得深入研究. 此外, 决策者的偏好可能随着信息更新而动态演变, 构建具有自适应能力的增量偏好提取机制能够及时响应偏好波动并保证结果稳定性.

参考文献 (References)

[1] Alvarez P A, Ishizaka A, Martínez L. Multiple-criteria decision-making sorting methods: A survey[J]. Expert Systems with Applications, 2021, 183: 115368.

[2] Ben Amor S, Belaid F, Benkraiem R, et al. Multi-criteria classification, sorting, and clustering: A bibliometric review and research agenda[J]. Annals of Operations Research, 2023, 325(2): 771-793.

[3] Wu X L, Liao H C, Zhang C H. Preference disaggregation analysis for sorting problems in the context of group decision-making with uncertain and inconsistent preferences[J]. Information Fusion, 2024, 101: 102014.

[4] Wójcik M, Kadziński M, Ciomek K. Selection of a representative sorting model in a preference disaggregation setting: A review of existing procedures, new proposals, and experimental comparison[J]. Knowledge-Based Systems, 2023, 278: 110871.

[5] 周坤, 巩在武. 解聚范式下考虑准则舍取的多准则分类方法[J]. 系统工程理论与实践, DOI: 10.12011/SETP2024-1740.
(Zhou K, Gong Z W. A multi-criteria sorting methods considering criteria selection within the disaggregation paradigm[J]. Systems Engineering — Theory & Practice, DOI: 10.12011/SETP2024-1740.)

[6] Liu J P, Kadziński M, Liao X W, et al. Data-driven preference learning methods for value-driven multiple criteria sorting with interacting criteria[J]. INFORMS Journal on Computing, 2020, 33(2): 419-835.

[7] Belahcène K, Mousseau V, Ouerdane W, et al. Multiple criteria sorting models and methods — Part I: Survey of the literature[J]. 4OR, 2023, 21(1): 1-46.

[8] Benabbou N, Perny P, Viappiani P. A regret-based preference elicitation approach for sorting with multicriteria reference profiles[C]. From Multicriteria Decision Making to Preference Learning. Paderborn, 2016: hal-01423287.

[9] Benabbou N, Perny P, Viappiani P. Incremental elicitation of Choquet capacities for multicriteria choice, ranking and sorting problems[J]. Artificial Intelligence, 2017, 246: 152-180.

[10] Özpeynirci S, Özpeynirci Ö, Mousseau V. An interactive algorithm for multiple criteria constrained sorting problem[J]. Annals of Operations Research, 2018, 267(1): 447-466.

[11] Kadziński M, Ciomek K. Active learning strategies for interactive elicitation of assignment examples for threshold-based multiple criteria sorting[J]. European Journal of Operational Research, 2021, 293(2): 658-680.

[12] Li Z L, Zhang Z, Pedrycz W. An incremental preference elicitation-based approach to learning potentially non-monotonic preferences in multi-criteria sorting[J]. European Journal of Operational Research, 2025, 323(2): 553-570.

[13] 张发明, 张淋茜, 朱姝琪. 基于前景理论的多阶段混合信息决策方法[J]. 管理学报, 2024, 21(4): 605-615.
(Zhang F M, Zhang L Q, Zhu S Q. A method for multi-stage mixed information decision-making based on prospect theory[J]. Chinese Journal of Management, 2024, 21(4): 605-615.)

[14] 张发明, 张淋茜, 韩江涛. 基于证据理论的多阶段混合型偏好信息群决策方法[J]. 系统工程理论与实践, 2023, 43(12): 3619-3625.
(Zhang F M, Zhang L Q, Han J T. A method for multi-stage group decision making based on evidence theory under hybrid preference information[J]. Systems Engineering — Theory & Practice, 2023, 43(12): 3619-3625.)

[15] Zhang S T, Zhang S J, Zhu F L, et al. Integrating case learning and consensus reaching in multi-criteria group sorting under hybrid assessment information[J]. Engineering Applications of Artificial Intelligence, 2025, 146: 110323.

[16] Libório M P, Ekel P, D'Angelo M F S V, et al. Dealing with heterogeneous information in multi-criteria group decision-making problems: A comprehensive design framework[J]. IEEE Access, 2025, 13: 114444-114459.

[17] Zhang Z, Li Z L, Yu W Y. Lexicographic optimization-based approaches to learning a representative model for multi-criteria sorting with non-monotonic criteria[J]. Computers & Operations Research, 2025, 175: 106917.

[18] 张恒杰, 朱文凤, 董玉成. 异质环境下基于最小冒犯准则的分类共识决策方法研究[J]. 系统工程理论与实践, 2022, 42(5): 1378-1390.
(Zhang H J, Zhu W F, Dong Y C. The classification-based consensus model with minimum violations in

heterogeneous group decision making[J]. Systems Engineering — Theory & Practice, 2022, 42(5): 1378-1390.)

[19] Li Z L, Zhang Z, Pedrycz W. Integrating machine learning models to learn potentially non-monotonic preferences for multi-criteria sorting from large-scale assignment examples[J]. Omega, 2025, 131: 103219.

[20] Zhang S J, Zhang S T, Tian H, et al. Multi-criteria consensus sorting model with flexible linguistic preferences based on fuzzy information granulation from the perspective of preference disaggregation[J]. Engineering Applications of Artificial Intelligence, 2025, 154: 111021.

[21] Kadziński M, Ciomek K. Integrated framework for preference modeling and robustness analysis for outranking-based multiple criteria sorting with ELECTRE and PROMETHEE[J]. Information Sciences, 2016, 352/353: 167-187.

[22] Lahdelma R, Hokkanen J, Salminen P. SMAA - Stochastic multiobjective acceptability analysis[J]. European Journal of Operational Research, 1998, 106(1): 137-143.

[23] Tervonen T, Lahdelma R. Implementing stochastic multicriteria acceptability analysis[J]. European Journal of Operational Research, 2007, 178(2): 500-513.

[24] Corrente S, Greco S, Leonardi F, et al. The hierarchical SMAA-PROMETHEE method applied to assess the sustainability of European cities[J]. Applied Intelligence, 2021, 51(9): 6430-6448.

[25] İlbaş A, Gürdere A, Boran F E. An integrated intuitionistic fuzzy set and stochastic multi-criteria acceptability analysis approach for supplier selection[J]. Neural Computing and Applications, 2023, 35(5): 3937-3953.

[26] Pereira J, Contreras P, Taramasco C, et al. A SMAA-based framework for robustness analysis in multi-criteria facility location under uncertainty[J]. Transportation Research — Part E: Logistics and Transportation Review, 2026, 213: 105002.

[27] De Bock K W, Coussement K, Lessmann S. Cost-sensitive business failure prediction when misclassification costs are uncertain: A heterogeneous ensemble selection approach[J]. European Journal of Operational Research, 2020, 285(2): 612-630.

[28] Kadziński M, Ghaderi M, Dąbrowski M. Contingent preference disaggregation model for multiple criteria sorting problem[J]. European Journal of Operational Research, 2020, 281(2): 369-387.

[29] Kadziński M, Martyn K, Cinelli M, et al. Preference disaggregation method for value-based multi-decision sorting problems with a real-world application in nanotechnology[J]. Knowledge-Based Systems, 2021, 218: 106879.

作者简介

姜继存 (1997−), 男, 博士生, 主要研究方向为决策理论与方法, E-mail: jjc18371411594@163.com;

朱建军 (1976−), 男, 教授, 博士, 主要研究方向为智能决策与风险分析、应急管理决策, E-mail: zhujianjun@nuaa.edu.cn;

刘小弟 (1981−), 男, 教授, 博士, 主要研究方向为模糊决策、多属性决策理论与方法, E-mail: lxy1160@163.com.