一种基于数据和专家知识构建模糊数的混合方法

Diego García-Zamora*[a]、José Rui Figueira[b]、Miguel Couceiro[c]

[a]数学系，哈恩大学，*23071* 哈恩，西班牙[b]*CEGIST*，*Instituto Superior Técnico*，里斯本大学，葡萄牙[c]*INESC-ID*，*Instituto Superior Técnico*，里斯本大学，葡萄牙

arXiv:2602.01192v1 [math.GM] 1 Feb 2026

摘要

本文提出了一种混合社会技术方法，用于从数值数据构建模糊数，同时通过交互式卡片法（DoC）流程融入专家知识。该方法扩展了现有的卡片法隶属函数构建框架，引入了一个基于凸模糊 k-Means 的数据驱动流程，其中每个计算步骤都会生成中间输出，并将其转换为基于卡牌的结构，供专家进行验证和调整。所提出的方法确保了实证证据与专家语义之间的可解释性、适应性和一致性。

关键词： 模糊数、卡片法、混合建模、数据驱动的隶属度构建、凸模糊 k-Means

1. 引言

构建恰当的隶属函数对于模糊模型的可靠性和可解释性至关重要（Dombi，1990）。过去几十年来，确定隶属函数的研究主要沿着两条路线展开（Schwaab等人，2015；Bilgiç &Türkşen，2000）：数据驱动方法和专家驱动方法。

数据驱动方法依赖于从数值数据集中提取信息（Yadav与Yadav，2015）。这些技术通常通过聚类（Dubois，1980；Dalleau等人，2020；Khairuddin等人，2021；Preud'Homme等人，2021）、神经网络（Ross，2010；George与Santra，2020）、基于直方图的识别（Dubois与Prade，1983；Soelistijanto，2022）、统计建模（Majumder等人，2007；Porebski与Straszecka，2016）、遗传算法等元启发式算法（Suryana等人，2009；Khairuddin等人，2022）、粒子群优化（Fang等人，2008；Li等人，2019）或通用优化方案（Wu与Mendel，2014；Bhattacharyya与Mukherjee，2020）来优化隶属参数，以反映统计规律。这些研究成果已广泛应用于分类、控制、图像处理和模式识别问题（Schwaab等人，2015；Medasani等人，1998；Miliauskaite与Kalibatiene，

通讯作者地址：哈恩大学数学系，邮编23071，西班牙哈恩电子邮箱：dgzamora@ujaen.es（Diego García-Zamora），figueira@tecnico.ulisboa.pt（José Rui Figueira），miguel.couceiro@inesc-id.pt（Miguel Couceiro）

2020；Hasuike等人，2015）。然而，这些方法的稳健性完全取决于可用数据的质量和数量，因此在观测值稀缺或异质的情境中，其适用性受到限制（Schwaab等人，2015）。

专家驱动方法则采取相反的方向，根据人类判断而非实证证据构建隶属函数。该领域包括经典的语言建模框架（Martínez等人，2015），其中通过直观程序 elicitation 偏好，例如直接评分（Norwich与Turksen，1984）、投票调查（Hersh与Caramazza，1976；Ahmad Shukri与Isa，2021；Jain与Khare，2010；Nguyen与Kreinovich，2014）、区间估计（Wang等人，1986；Ukhobotov与Krasil’nikova，2017）、逆向评分（Turksen，1991）、举例说明（Norwich与Turksen，1984）或两两比较（Nieto-Morote与Ruz-Vila，2023）。这些人工方法能够捕捉人类认知中的主观细微差异，但往往耗时较长，而且可能需要分析人员的协助，以避免非技术专家承受认知负担（Sancho-Royo与Verdegay，1999）。

专家驱动路线中较新的方法尝试通过为隶属函数指定预定义形状（通常为三角形或梯形）来简化信息 elicitation 过程（Wu与Mendel，2014；Miliauskaite与Kalibatiene，2020）。具有代表性的例子包括二元组模糊语言建模（Herrera与Martinez，2000年）和犹豫模糊语言术语集（Rodriguez等人，2012年），它们在文献中仍具有重大影响。然而，这类预定义结构隐含地假设决策者之间具有共同的语义理解，忽视了不同个体对定性术语的解释差异（Fan与Liu，2010；Ye等人，2023）。在群体决策场景中，这一局限性更加突出：信息可能采用不同的尺度表达（Ye等人，2023；Chen与Li，2023），对语言具有不同的语义感知（Jiang等人，2022；Liang等人，2021），且粒度层级也各不相同（Fan与Liu，2010；Li等人，2022）。

在这一背景下，如何构建既能反映经验模式、又具备人类可解释性的隶属函数，仍存在研究空白（Dombi与Jónás，2022年；Medvediev等人，2020年；Kara与Kocken，2021年；Wang等人，2022年）。现有方法大多仅在初步阶段征询专家意见，无法让专家随着模型演进而主动参与隶属函数的共同构建（Chakraborty，2001年）。近期研究强调，需要建立社会技术框架，让决策者与分析人员通过迭代反馈和协商，共同构建不确定性表示（Corrente等人，2021年）。

近期由 García-Zamora 等人（2024）提出的基于卡牌的隶属函数 （DoC-MF）方法，为通过专家引 elicitation 构建模糊数提供了一种结构化且可解释的方法。该方法使决策者能够借助直观的基于卡牌的表示表达定性差异，从而在无需直接进行数值判断的情况下，有效衔接人类推理与不确定性建模。

然而，原始 DoC-MF 框架主要针对专家驱动的情境设计，在这类情境中，必须仅通过与决策者互动来构建隶属函数。在许多实际应用中，已有相关数值数据集可用于不确定性的建模。仅依赖专家评估可能会导致偏差，或产生与实证证据不一致的表述

结合实证证据。

为解决这些局限性，我们提出一种混合社会技术方法，将数据驱动的洞见与 DoC-MF 方法的可解释性和灵活性相结合。本研究的主要贡献有三点：

- 我们提出一种新颖的数据驱动流程，从以频数表形式给出的数值观测值中提取初步模糊信息。
- 我们定义了一种机制，将中间计算结果转化为 DoC-MF 方法中的基于卡牌的结构，从而使专家能够在每个阶段进行交互、验证和完善。
- 我们提供了一套统一的隶属函数构造过程，确保经验模式与专家语义之间保持一致。

为支撑数据驱动部分，我们还开发了经典模糊k-均值聚类算法的一个改进版本，称为凸 ***FKM*** （C-FKM）。通过施加凸性约束，该方法确保计算过程中产生的所有簇都对应于有效的模糊数，并与 DoC 引导协议兼容。

因此，最终的混合方法既保持了透明度和可解释性，又能适应异质或非标准的数据分布。通过融合经验数据与专家直觉来构造隶属函数，这种混合方法能够应对多个高风险领域中的关键挑战。例如，在预测医学中，它可以协调定量临床指标与从业者细致入微的定性判断，从而改进诊断模型。同样，在工业过程控制和自主系统中，这种融合使自动控制器能够继承人类操作人员的韧性和“常识”，即使环境条件偏离历史模式，也能确保系统稳定运行。此外，在环境风险评估或金融预测等领域，将数学模型建立在主观专业知识基础上的能力，有助于更忠实地反映现实，克服纯数据驱动模型的局限性；这类模型可能缺乏背景信息，或无法考虑罕见但影响重大的事件。本文聚焦于一项基于真实教育表现数据的数值研究，该研究展示了我们的方法能够生成语义上有意义且与底层证据相符的隶属函数。

本文其余部分安排如下。第 2 节介绍文献中的基本概念，以及理解本文所需的符号。在第 3 节中，我们将建立理论基础，扩展经典的模糊 k-均值方法，使其兼容 DoC-MF 引导方法。随后，第 4 节详细介绍我们基于数据与专家互动、用于获取模糊数的混合方法。在第 5 节中，我们通过一个具体情境展示混合流程的完整示例；第 6 节则提供一些数值比较，以说明该方法面对不同数据分布时的表现。最后，第 7 节总结全文。

2. 预备知识

本节回顾全文所需的基本术语和符号。我们还将重新介绍“卡片组方法”和“模糊 k-均值聚类”，它们将构成我们所提出的模糊数构造方法的基础。

2.1. 模糊集与模糊数

在 $\mathbb{R}$ 上的模糊集是一个映射 $A:\ \mathbb{R}\to[0,1]$。对于 $\alpha\in]0,\ 1]$，α-截集是 A 的子集 $A^{\alpha}=\{x\in\mathbb{R}:\ \ A(x)\geq\alpha\}$，而 A^{0} 表示 A 支撑集的拓扑闭包（Wang 等人（2012））。注意，A^{0}（我们将其称为 A 的支撑集）刻画了实数轴上的点：这些点在某种程度上表现出对模糊集 A 的隶属性；而 A^{1}（即 A 的核）表示 $\mathbb{R}$ 中我们能够完全确定其属于 A 的值。

模糊数是满足以下条件的模糊集 A：（i）$A^{1}\neq\emptyset$（正规性）；（ii）对于任意 $\alpha\in[0,1]$，A^{α} 都是区间（凸性）；以及（iii）A^{0} 有界。模糊数表示某个值$x\in\mathbb{R}$ 被认为与相应不确定性概念相容的程度。为了使这种表示在决策情境中具有意义，隶属函数必须满足若干结构性质，以确保可解释性和内部一致性。等价地，隶属函数在核的左侧递增，并在其右侧递减。这种左右单峰行为反映了如下假设：不确定性在两端最高，而在由核表示的中心概念附近最低。

最后，若一族 $k\in\mathbb{N}$ 模糊数 $A_{1},\ldots,A_{k}$ 满足：对于所有 $x\in[a,b]$，$\sum_{j=1}^{k}A_{j}(x)=1$，则称其为定义域 $[a,b]$ 上的模糊划分。这个概念在本研究中至关重要，因为它确保所有隶属函数共同表示整个定义域且不存在歧义，使每个值都能表达其对不同概念的隶属关系，同时保持决策的全局可解释性和一致性。

2.2. 卡片组方法

DoC方法最初是一种偏好引导技术，旨在借助简单且符合认知直觉的表示方式，帮助专家表达偏好和定性差异（Corrente 等，2021年）。决策者不使用数值尺度，而是通过操作一组相同的卡片来表示有序类别之间的相对差异。在两个参考点之间插入的空白卡片数量，反映了人们感知到的差异强度：中间放置的卡片（单位）越多，人们认为这些概念之间的距离就越大。例如，假设有三个有序对象 $l_{1}\prec l_{2}\prec l_{3}$，需要对它们的表现进行评估。假设决策者确定，第一和第二个对象之间的偏好强度差异可用四个单位表示，而第二和第三个对象之间的偏好差异可用六个单位表示，即：

$$l_{1}\quad[4]\quad l_{2}\quad[6]\quad l_{3}$$

然后，分析人员按比例为这些对象赋予一个 $[0,1]$ 尺度上的数值：

$$v(l_1) = 0 \quad v(l_2) = 0.4 \quad v(l_3) = 1.$$

这种方法避免要求专家直接给出数值判断，同时保留了序数与关系信息。DoC方法的一项关键性质是，只要卡片数量足够多，就能逼近任意一组数值，正如定理 1 所述。

定理 1 (Dutta等人（2025）)。 设$\mathbf{x} = (x_0, \ldots, x_n) \in [0,1]^{n+1}$，且 $n > 1$，是满足以下条件的 n 元组

$$0 = x_0 < x_1 < \cdots < x_n = 1,$$

并令 $m \in \mathbb{N}$ 为 s 使得 $\lfloor 10^m x_{i-1} \rfloor < \lfloor 10^m x_i \rfloor$, $i = 1, \ldots, n$. 定义有理逼近

$$r_i = \frac{\lfloor 10^m x_i \rfloor}{10^m}, \quad i = 0, \ldots, n,$$

其中 $N = 10^m$ 为单位总数。于是，*DoC*方法可以用精度为 10^{-m}的方式表示有序元组 x，所用卡片序列由 N 张卡片组成，并划分为 n组连续单位，这些单位组通过递归方式确定：

$$\begin{aligned} c_1 &= N r_1 = \lfloor 10^m x_1 \rfloor, \\ c_i &= N r_i - \sum_{j=1}^{i-1} c_j = \lfloor 10^m x_i \rfloor - \sum_{j=1}^{i-1} c_j, \qquad i = 2, \ldots, n. \end{aligned}$$

整数 c_1、.....、 c_n 精确规定了应在 r_{i-1} 与 r_i 之间放置的单位数量，并满足

$$r_i = \frac{1}{N} \sum_{j=1}^{i} c_j, \qquad i = 1, \ldots, n.$$

因此，由这些递归计数构建的 *DoC*表示以 10^{-m} 的精度近似原始元组 x。

在近期的研究中，DoC方法已扩展用于构建模糊数 García-Zamora 等人（2024），以获得给定尺度的语言术语的模糊表示。为此，采用一种分三步实施的社会技术方法。首先，通过在尺度的不同层级之间放置空白卡片，确定模糊数代表点的价值标度。每张卡片对应一个离散单位，插入卡片的累积结构可以自然地转换为 $[0,1]$ 中的数值。随后，通过询问决策者关于相应模糊概念的缺失或完全置信范围，推断支撑集和核的边界。最后，再次使用 DoC方法细化隶属函数的左侧和右侧。

这种基于DoC的诱导过程具有多项优势。首先，它避免了三角形或梯形轮廓等任意的函数假设，使隶属曲线能够从专家推理中自然形成

。其次，卡片表示法使决策者能够逐步修订并说明建模选择，从而促进多利益相关方情境下的沟通与协商。此外，该方法在不确定性建模与人类判断之间建立了可追溯性；对于结果能否被接受取决于其可解释性的应用而言，这一点至关重要。

2.3. 模糊 k 均值聚类

由于所提出的混合方法中的若干组成部分依赖于将经验数据划分为不同程度的组，因此我们在此回顾经典 FKM（也称为模糊 c 均值）算法，并介绍本文通篇使用的符号。

设 $D = \{x_1, \ldots, x_n\}$ 为一组实值观测值。给定固定数量 $k \geq 2$ 的簇，FKM 的目标是确定簇中心（原型） v_1、……、$v_k \in \mathbb{R}$，以及隶属度 $u_{ij} \in [0,1]$，以表示 x_i 属于簇 j 的程度，并使得对每个数据点 x_i，$i = 1$、……、n，均满足 $\sum_{j=1}^{k} u_{ij} = 1$，且隶属度是软性的，而非硬性的。经典目标函数 J： $[a,b]^{nk} \times [a,b]^k$ 定义为

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij}^{m} \, \|x_i - v_j\|^2,$$

其中，$U =$(${u_{ij}}$)为隶属度矩阵，$V =$(${v_1}$、……、${v_k}$)为簇中心元组，$m > 1$是标准模糊化参数（用于控制模糊度）；更多背景信息请参见 Bezdek等（1984）。典型取值为 $m = 1.5$ 或 $m = 2$。在隶属度约束下最小化 J，可得众所周知的更新方程:

$$u_{ij} = \left(\sum_{\ell=1}^{k} \left(\frac{\|x_i - v_j\|}{\|x_i - v_\ell\|} \right)^{\frac{2}{m-1}} \right)^{-1}, \quad \text{where} \quad v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} \, x_i}{\sum_{i=1}^{n} u_{ij}^{m}}, \quad j = 1, \ldots, k. \tag{1}$$

在经典方法中，这两个方程会迭代应用，直至收敛（Bezdek等（1984）），通常以 $\|V^{(t+1)} - V^{(t)}\|$ 低于容差阈值作为停止条件。

3. 凸模糊 k-均值

本节介绍 C-FKM，这是对经典 FKM 的一种改进，专门设计为使算法生成的每个类别在构造上都是一个模糊数。该方法在每个质心周围施加局部（凸）支撑，并强制模糊划分在数值轴上相邻:每个观测值只能隶属于两个簇，即其所在位置被这两个簇的中心点夹住。此局部性约束可产生具有紧支撑和非空核的隶属函数，这对于混合式构造具有可解释性的隶属函数以表示数值尺度至关重要。

下面，我们将描述该模型，给出在新约束下计算中心和隶属度的闭式公式，呈现一个收敛性结果，最后给出一种用于计算 C-FKM 解的简单迭代算法。

设 $D = \{x_1, \ldots, x_n\} \subset [a, b] \subset \mathbb{R}$ 为数据集，其中 $[a, b]$ 是从数据或外部来源中选取的边界。固定簇的数量 $k \geq 2$ 和模糊化参数 $m > 1$。记 $V = (v_1, \ldots\ldots, v_k)$ 为有序质心，其中 $a < v_1 < v_2 < \cdots < v_k < b$。对于每个观测值 x_i，我们将左侧括定索引 $j(i)$ 定义为满足以下条件的唯一索引：$v_{j(i)} \leq x_i < v_{j(i)+1}$，并约定 $v_0 := a$ 以及 $v_{k+1} := b$。因此，x_i 位于区间 $[v_{j(i)}, v_{j(i)+1})$ 内。记 $U = (u_{ij}) \in \mathcal{M}_{n\times k}([0, 1])$ 为满足标准单纯形约束的隶属度矩阵

$$u_{ij} \geq 0, \qquad \sum_{j=1}^{k} u_{ij} = 1, \qquad i = 1, \ldots, n.$$

C-FKM 的关键邻接约束是，对于每个 i，我们只允许其对两个相邻簇具有非零隶属度：

$$u_{ij} = 0 \quad \text{for all } j \notin \{j(i), j(i) + 1\}.$$

最后，与经典 FKM 一样，我们考虑目标函数 $J\colon \mathcal{M}_{n\times k}([0, 1]) \times [a, b]^k$

$$J(U, V) = \sum_{i=1}^{n}\sum_{j=1}^{k} u_{ij}^m (x_i - v_j)^2, \quad \forall\, U \in \mathcal{M}_{n\times k}([0, 1]), V \in [a, b]^k$$

因此，C-FKM 问题是在上述单纯形约束和邻接约束下对 J 进行约束极小化：

$$\begin{aligned} \min_{\substack{U=(u_{ij}) \\ V=(v_1,\ldots,v_k)}} \quad & J(U, V) = \sum_{i=1}^{n}\sum_{j=1}^{k} u_{ij}^m (x_i - v_j)^2 \\ \text{subject to} \quad & u_{ij} \geq 0, \qquad \sum_{j=1}^{k} u_{ij} = 1, \qquad i = 1, \ldots, n, \\ & u_{ij} = 0 \quad \text{for all } j \notin \{j(i), j(i) + 1\}, \\ & a < v_1 < v_2 < \cdots < v_k < b. \end{aligned}$$

请注意，对每个 x_i 只能属于两个相邻簇这一限制，是确保支撑集凸且连续，并保证输出为模糊数的结构性改变，正如下文将证明的那样。不过，遵循经典 FKM 的结构，我们先介绍两个定理，它们给出了算法两个交替步骤中的闭式更新：(i) 给定隶属度计算中心点；(ii) 给定中心点计算隶属度。

命题1 (中心更新**)**。设 $U = (u_{ij}) \in \mathcal{M}_{n\times k}([0, 1])$ 是满足单纯形约束和邻接约束的任意可行隶属度矩阵。那么，使 $J_U\colon [a, b]^k \to \mathbb{R}$达到极小值的极小化解 $V^\star$，即，

在 J 中固定 U 后得到的函数 $V \to J(U,V)$，是唯一的，且其表达式为

$$v_j^\star = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m}, \qquad j = 1, \ldots, k.$$

证明。对于固定的 $U \in \mathcal{M}_{n\times k}([0,1])$，目标函数为

$$J_U(V) = \sum_{j=1}^{k}\sum_{i=1}^{n} u_{ij}^m (x_i - v_j)^2, \text{ for } V \in [a,b]^k.$$

因此，当梯度 ∇J_U 为零时，可找到唯一的全局最小值，即

$$\nabla J_U = [\frac{\partial}{\partial v_1} J_U(V), \ldots, \frac{\partial}{\partial v_k} J_U(V)] = \mathbf{0} = [0, \ldots, 0]. \tag{2}$$

由于 $\frac{\partial}{\partial v_j} J_U(V) = -2\sum_{i=1}^{n} u_{ij}^m (x_i - v_j)$，因此只有当 $v_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m}$，式 (2) 才成立，证明至此完成。 □

现在，我们给出一种在已知簇中心的情况下计算隶属度的方法。

定理 2 (隶属度更新)。考虑固定的质心 $V = (v_1, \ldots, v_k)$，其中 $a < v_1 < \cdots < v_k < b$。对于每个 $i \in \{1, \ldots, n\}$，令 $j = j(i)$ 为左侧括定索引，使得 $v_j \le x_i < v_{j+1}$。在对 $\ell \notin \{j, j+1\}$ 满足 $u_{i\ell} = 0$ 的邻接约束下，满足 $\sum_\ell u_{i\ell} = 1$ 和 $u_{i\ell} \ge 0$ 的 $J_V(U) = \sum_{j=1}^{k} u_{ij}^m d_{ij}$ 的唯一最小值 $U^\star \in \mathcal{M}_{n\times k}([0,1])$ 为 $d_{ij} = (x_i - v_j)^2 \; \forall\, i = 1, \ldots, n, \; j = 1, \ldots, k$，其定义为

1. $u_{i,j} = \left(1 + \left(\frac{d_{i,j}}{d_{i,j+1}}\right)^{1/(m-1)}\right)^{-1}$, *and* $u_{i,j+1} = \left(1 + \left(\frac{d_{i,j+1}}{d_{i,j}}\right)^{1/(m-1)}\right)^{-1}$, *whenever* $d_{i,j} > 0$ *and* $d_{i,j+1} > 0$,
2. $u_{i,\,j} = 1$ 和 $u_{i,\,j+1} = 0$，当 $d_{i,\,j} = 0$ 和 $d_{i,\,j+1} > 0$ 时，以及
3. $u_{i,\,j+1} = 1$ 和 $u_{i,\,j} = 0$，当 $d_{i,\,j+1} = 0$ 和 $d_{i,\,j} > 0$ 时。

证明。给定邻接约束，对于固定的 $i = 1, \ldots, n$，目标函数可简化为

$$J_i(u_{i,j}, u_{i,j+1}) = u_{i,j}^m d_{i,j} + u_{i,j+1}^m d_{i,j+1},$$

满足约束条件 $u_{i,j} + u_{i,j+1} = 1$, $u_{i,j}, u_{i,j+1} \ge 0$。因此，我们可以消去 $u_{i,j+1} = 1 - u_{i,j}$，并在 $t \in [0,1]$ 上最小化一维凸函数 $f(t) = t^m d_{i,j} + (1-t)^m d_{i,j+1}$。对于非退化情形 $d_{i,j}, d_{i,j+1} > 0$，一阶条件 $f'(t) = 0$ 给出

$$m t^{m-1} d_{i,j} - m(1-t)^{m-1} d_{i,j+1} = 0$$

因此

$$\Big(\frac{t}{1-t}\Big)^{m-1} = \frac{d_{i,j+1}}{d_{i,j}} \quad \Longrightarrow \quad \frac{t}{1-t} = \Big(\frac{d_{i,j+1}}{d_{i,j}}\Big)^{1/(m-1)}.$$

对 t 求解，恰好得到陈述中的表达式。其中一个距离为零的退化情形，可由如下观察得出：$J_i \geq 0$ 以及任一零距离项都会迫使目标函数通过将该零距离簇赋予全部权重而达到最小。

□

定理 3. 设 $V = (v_1, \ldots, v_k)$ 为一组有序质心，设 $U = (u_{ij})$ 为定理 *2* 中得到的隶属度矩阵，模糊化因子为 $m > 1$，并设 V_j 表示簇 $j = 1, \ldots, k$ 的隶属函数，该隶属函数由在考虑质心 V 的情况下对 x 上的值 u_{ij} 进行插值得到。则对每一个 $j = 1, \ldots, k$，函数 V_j 满足

- 正规性 $(V_j(v_j) = 1)$，
- 其支撑集包含于 $[v_{j-1}, v_{j+1}]$，
- 它在左区间 $[v_{j-1}, v_j]$ 上单调递增，并在右区间 $[v_j, v_{j+1}]$上单调递减。

因此，每个 V_j 都是一个模糊数，而集合 $\{V_1, \ldots, V_k\}$ 是 $[a, b]$的一个模糊划分。

证明。 正规性和支撑集条件可直接由模型构造和邻接约束得出。事实上，根据构造，对于恰好位于 v_j的数据点，在最小化过程中，唯一的最优分配（见定理 2）会将完整隶属度赋予簇 j，因此 $V_j(v_j) = 1$ （正规性）。邻接约束规定，只有当 $x \in [v_{j-1}, v_{j+1}]$ 时，数据点 x 才会对簇 j 具有正隶属度，因此 V_j 的支撑集包含于紧区间 $[v_{j-1}, v_{j+1}]$中。

接下来要证明单调性断言。固定一个指标 j，先考虑右侧区间 $[v_j, v_{j+1}]$。考虑一个指标 $i = 1, \ldots, n$，满足 $v_j < x_i < x_{i+1} < v_{j+1}$。在这种情况下，$x \to (x - v_j)^2$ 和 $x \to (x - v_{j+1})^2$ 的单调性意味着 $d_{i,\,j} d_{i+1,\,j+1} \leq d_{i+1,\,j} d_{i,\,j+1}$。如果假设所有这些值都为正，则可得

$$\begin{aligned}
\frac{d_{i+1,j}}{d_{i+1,j+1}} \geq \frac{d_{i,j}}{d_{i,j+1}} &\iff 1 + \Big(\frac{d_{i+1,j}}{d_{i+1,j+1}}\Big)^{1/(m-1)} \geq 1 + \Big(\frac{d_{i,j}}{d_{i,j+1}}\Big)^{1/(m-1)} \\
&\iff \Big(1 + \Big(\frac{d_{i,j}}{d_{i,j+1}}\Big)^{1/(m-1)}\Big)^{-1} \geq \Big(1 + \Big(\frac{d_{i+1,j}}{d_{i+1,j+1}}\Big)^{1/(m-1)}\Big)^{-1} \\
&\iff u_{ij} \geq u_{(i+1)j},
\end{aligned}$$

这体现了 V_j 在其右侧的单调性。如果其中一些值为零，考虑到某些隶属度将等于 1，单调性仍然成立。为节省篇幅，这里省略完整的讨论。左侧 $[v_j, v_{j+1}]$ 的论证类似。

结合正规性、紧支撑以及已确立的两侧单调性可知，每个 V_j 都是一个正规、凸的隶属函数，其支撑集为紧区间，而这

正是模糊数的定义。最后，由于对于每个 x_i，邻接约束确保各簇之间的隶属度之和为一，族 $\{\mu_j\}$ 构成了 $[a,b]$ 的模糊划分。证明完毕。

□

以下定理证明了我们在实践中采用的交替优化方法是合理的：先用命题 2 更新隶属度，再用命题1更新中心点，并重复此过程直至收敛。

定理 4 (收敛到局部最小值**)**。 令 $\{(U^{(t)}, V_{(t)})\}_{t\geq 1}$ 为 ***C-FKM*** 交替更新所产生的序列（其初始质心向量为 V^0）：

$$U^{(t+1)} \leftarrow \arg\min_U J(U, V^{(t)}) \quad \textit{(using Proposition 2)},$$

$$V^{(t+1)} \leftarrow \arg\min_V J(U^{(t+1)}, V) \quad \textit{(using Proposition 1)}.$$

那么序列　　目标函数值 $J(U_{(t)}\, V_{(t)})$ 单调不增，并收敛到有限极限　　.

证明。 注意，每一步更新都在固定其他变量的同时，对变量的一个子集进行最小化。因此，

$$J(U^{(t+1)}, V^{(t)}) \leq J(U^{(t)}, V^{(t)}), \qquad J(U^{(t+1)}, V^{(t+1)}) \leq J(U^{(t+1)}, V^{(t)}).$$

因此，目标函数值序列单调不增且以下界 0 为界，由单调收敛定理 Spivak (2008)，该序列必定收敛。 □

请注意，C-FKM 算法会交替执行刚才推导出的闭式更新。下面给出简洁的伪代码，可用于实现。

1. 初始化：选择 k 初始质心 $v_1^{(0)} < \cdots < v_k^{(0)}$（例如，在 a 和 b 之间取等间距的值，或使用百分位数）。设置 $t \leftarrow 0$。

2. 重复：

   (a) 对每个 $i = 1,\ldots, n$，相对于 $V_{(t)}$计算 $j(i)$（左侧括定索引）。计算两个距离 $d_{i,j(i)}$ 和 $d_{i,j(i)+1}$，并根据命题 2 更新隶属度 $u_i^{(t+1)}{}_{i}{}_{,j()}$ $u_i^{(t+1)}{}_{i}{}_{,j()+1}$。将所有其他 $u_{i\ell}^{(t+1)} = 0$设为相应值。(b) 根据命题1，使用 $U^{(t+1)}$更新每个质心 $v_j^{(t+1)}$。(c) 如果 $\|V_{(t+1)} - V_{(t)}\|$（或 J的相应下降量）低于容差水平 $\tau > 0$，则停止；否则设置 $t \leftarrow t + 1$。

这里回顾一下 C-FKM 算法的主要优点。一方面，邻接约束确保每个隶属函数 V_j 在相邻质心之间的紧区间之外为零，并且是凸函数。因此，所得的 V_j 是一个模糊数。另一方面，该算法生成分段线性模糊数，可将其纳入 DoC-MF 框架。在这一意义上，所有步骤都可以直接转化为卡片，并展示给决策者进行最终完善，正如我们将在下一节所展示的那样。

4. 混合c　　模糊数的构建，整合数据与专家知识

本节介绍一种混合方法，通过 DoC-MF 方法将专家知识融入数值数据，从而构建模糊数。所提出的流程遵循 DoC-MF 方法的相同三步结构（值尺度构建、核与支撑集的识别，以及左侧与右侧的定义），但在每一步中都加入了基于 C-FKM 的初步数据驱动流程。在每个阶段，数据驱动计算所产生的结果都会转化为基于卡牌的表示，以便决策者在方法进入下一阶段之前调整、验证或完善隶属函数。下文将详细介绍各个阶段。

4.1. 步骤*1*：值尺度构建

混合程序的第一阶段旨在构建一个初始价值标度，使其既反映数据的经验结构，又对参与该过程的决策者保持可解释性。遵循García-Zamora 等人（2024）在DoC-MF 方法中引入的社会技术方法的精神，本阶段将纯粹的数据驱动部分与基于专家的交互式细化相结合。因此，数据驱动部分从数据集中提取具有代表性的参考点，而专家部分则通过DoC方法对这些参考点进行解释和调整。

给定数据集 $D = \{x_1, \ldots, x_n\}$，我们首先应用C-FKM（见第 3 节），将数据划分为 k 个重叠组。本步骤仅需要算法产生的聚类质心，记为$v_1, \ldots, v_k$。这些质心作为数据分布的代表，捕捉数值观测值中的密度区域与模式。为确保该方法后续阶段的一致性，我们对质心进行排序，使得

$$a < v_1 < v_2 < \cdots < v_k < b,$$

其中，边界 $[a, b]$ 可以从数据中获得，也可以由决策者给出。定义（$k + 2$）元组

$$v = (a, v_1, v_2, \ldots, v_k, b).$$

随后，根据定理 1，将有序元组 v 转换为 DoC 结构。该定理保证，在选取适当的精度级别 10^{-m}后，可以将相邻数值之间的距离转换为实际卡片，以表示原始数值之间的强度差异。具体而言，对于所选的 m，我们计算有理逼近$r_i = \lfloor 10^m v_i \rfloor / 10^m$ 以及卡片数量的递归序列

$$c_1 = \lfloor 10^m v_1 \rfloor, \quad \text{and} \quad c_i = \lfloor 10^m v_i \rfloor - \sum_{j=1}^{i-1} c_j, \quad \text{for } i = 2, \ldots, k+1,$$

其中，每个数量 c_i 对应标签 v_{i-1} 和 v_i之间的单位数。随后，这些单位通过卡片链进行直观呈现，将质心之间的间距

以及数据集的几何结构转化为一种易于理解的可视化模型，供决策者方便地查看和操作。

卡片链生成后，会呈现给决策者，以便他们调整参考值之间的间距。这是混合方法的关键环节：即便数据表明质心之间存在特定距离，相邻取值之间被感知到的语义差异也可能不同于纯粹的数值差异。通过在层级之间插入或移除空白卡片，决策者可以增强或减弱尺度特定区域的区分能力。经专家细化的卡片配置将取代原始的数据驱动结构，并成为后续流程步骤所采用的最终价值标度。

示例1. 为说明这一机制，考虑一个缩放至 $[0,1]$ 的数据集，并假设*FKM*识别出三个质心，分别位于 $v_1 = 0.18$、 $v_2 = 0.43$、 $v_3 = 0.72$。计入端点（0 和1）后，我们得到有序元组 $x = (0, 0.18, 0.43, 0.72, 1)$。假设所需精度为 10^{-2}，从而可使用 $m = 2$ 和 $N = 10^2 = 100$ 个单位。应用定理 *1*，我们计算：

$$
\begin{aligned}
c_1 &= \lfloor 100 \cdot 0.18 \rfloor = 18, \\
c_2 &= \lfloor 100 \cdot 0.43 \rfloor - 18 = 25, \\
c_3 &= \lfloor 100 \cdot 0.72 \rfloor - (18 + 25) = 29, \\
c_4 &= \lfloor 100 \cdot 1.00 \rfloor - (18 + 25 + 29) = 28.
\end{aligned}
$$

因此，数据驱动的卡片链包含：

$$
18 \ \textit{cards between } 0 \ \textit{and } 0.18, \qquad 25 \ \textit{cards between } 0.18 \ \textit{and } 0.43,
$$

$$
29 \ \textit{cards between } 0.43 \ \textit{and } 0.72, \qquad 28 \ \textit{cards between } 0.72 \ \textit{and } 1.
$$

面对这种结构，假设一位决策者可能会认为，0.18 与0.43 之间的差异，相对于这些层级之间感知到的语义跃变而言太小了（例如，这些数值对应不同的语言标签时）。因此，决策者可能会在第二个区间中插入 *5* 张额外的空白卡片，使其总数从 *25* 增至 *30*。相反，决策者可能觉得最后一个区间（即0.72 与1之间的区间）跨度过大，于是移除 *4* 张卡片，将其数量从 *28* 减至 *24*。调整后的卡片链取代原卡片链，同时编码质心中包含的经验信息，以及决策者对数值间距的语义判断。

这一 refined 价值标度将成为后续混合方法步骤的基础，用于识别核与支撑集，并构建隶属函数的左侧和右侧。

4.2. 步骤 *2*：核与支撑集的识别

在步骤 1 中经决策者验证质心后，下一步的目标是进一步分析与各类别相关的模糊数的核与支撑集。这

需要将已验证质心转换为完整的隶属函数，确定其核与支撑集，并允许决策者通过卡片组表示法进行交互，以细化所得区间。

令 $V = (v_1, \ldots, v_k)$ 表示已验证质心。根据定理 3，与质心 v_j 相关联的簇的隶属函数可通过以下表达式更新

$$u_{ij} = V_j(x_i) = \begin{cases} \left(1 + \left(\dfrac{(x_i - v_j)^2}{(x_i - v_{j-1})^2}\right)^{\frac{1}{m-1}}\right)^{-1}, & \text{if } v_{j-1} \le x_i < v_j, \\ \left(1 + \left(\dfrac{(x_i - v_j)^2}{(x_i - v_{j+1})^2}\right)^{\frac{1}{m-1}}\right)^{-1}, & \text{if } v_j \le x_i < v_{j+1}, \\ 1, & \text{if } j = 0 \text{ and } x_i \le v_0, \text{ or } j = k \text{ and } x_i \ge v_k, \text{ or } x_i = v_j \\ 0, & \text{otherwise,} \end{cases}$$

请注意，这保证了：(i) 在 v_j处的正规性；(ii) 在 $[v_{j-1}, v_{j+1}]$上的紧支撑；以及 (iii) 质心两侧各自的单调性（见定理 3）。为了获得可解释的模糊数，我们将每个类别 V_j 的核定义为

$$\mathrm{Core}(V_j) = \{x \in D : V_j(x) \ge 1 - \tau\}, \qquad \text{with } \tau \in [0, 1] \text{ typically set to } 0.01.$$

由于 μ_j 在 V_j两侧均单调，核始终是一个区间 $\mathrm{Core}(V_j) = [c_j, c_j]$。由此得到的这些边界

$$a = \underline{c}_1 < \overline{c}_1 < \underline{c}_2 < \overline{c}_2 < \ldots < \underline{c}_k < \overline{c}_k = b$$

根据定理 1 转换为卡片组单位。随后将所得卡片展示给决策者，他们可以修改卡片以扩大或缩小核。请注意，此调整遵守不同模糊数的核不可重叠这一约束。此外，前面步骤中得到的已验证质心必须位于核内。为确保这一点，可以将已验证质心纳入先前的数值链，并允许专家在质心与核的边界之间放置卡片。决策者提供修订后的核后

$$\mathrm{Core}(V_j) = [\underline{c}'_j, \overline{c}'_j], \qquad j = 1, \ldots, k,$$

我们重新计算隶属函数，以纳入修正后的结构信息。更新后的隶属度遵循以下规则：

- 若对于某个 j，有 $x_i \in \mathrm{Core}(V_j)$，则对于所有 $\ell \ne j$，$V_j(x_i) = 1$ 且 $V_\ell(x_i) = 0$。
- 如果 x_i 不属于任何一个核，那么它位于两个核之间。我们用 $j(i)$表示 x_i左侧簇的索引，并根据命题 2 应用双聚类更新规则，使用

相邻核边界作为质心：

$$V_{j(i)}(x_i) = \left(1 + \left(\frac{(x_i - \overline{c}'_{j(i)})^2}{(x_i - \underline{c}'_{j(i)+1})^2}\right)^{\frac{1}{m-1}}\right)^{-1}, \qquad V_{j(i)+1}(x_i) = 1 - V_{j(i)}(x_i),$$

其他所有隶属度均设为零。

随后将更新后的核传递给隶属度更新算法。核内的点隶属度为 1；位于两个相邻核之间的点，则根据双簇规则重新计算其隶属度。由此得到新的模糊划分，充分融合了数据驱动的结构与决策者的语义调整。

这里需要指出，模糊划分的单纯形约束保证了可以直接根据相邻核计算支撑集。因此，在这种方法中，也可以通过确定支撑集来等效地确定核。从这个意义上说，如果决策者觉得这样更方便，也可以依据支撑集边界的信息来完成这一步。请记住，关于核的问题是哪些取值能让你完全确定其属于该类别？，而关于支撑集的问题是哪些取值范围内的值完全不属于该类别？。

示例 2. 假设在我们流程的第一步之后，已验证质心为 $V = (0.20,\ 0.50,\ 0.80)$。应用更新规则后，*C-FKM* 隶属度得到如下核（使用$\mathrm{tol} = 0.99$）

$$\mathrm{Core}(V_1) = [0.18, 0.22], \qquad \mathrm{Core}(V_2) = [0.48, 0.52], \qquad \mathrm{Core}(V_3) = [0.78, 0.82].$$

这六个数值被转换为卡片组单位（见定理 *1*），决策者随后检查生成的卡片。假设他们认为中间核过窄，于是调整卡片数量。逆向执行这一过程后，我们得到

$$\mathrm{Core}(V_2) = [0.46, 0.54].$$

4.3. 第 *3* 步：微调左侧和右侧

一旦决策者验证了所有模糊数的核和/或支撑集，最终的细化阶段便着重于调整其核外部的隶属函数形状。在此步骤中，我们分析每个模糊数的单调分支，并确定有意义的“置信水平”，以表明当我们从核向外移动时，隶属度应以多快或多慢的速度下降。随后，这些置信水平将转换为卡片法单位，以便决策者能够直接对其进行操作和细化。我们首先选择要细化其形状的模糊数 V_j。给定其经过验证的支撑集和核，$\mathsf{Supp}(V_j) = [s_j^l, s_j^r]$，$\mathsf{Core}(V_j) = [c_j^l, c_j^r]$，我们分别关注其左侧和右侧：

$$\text{LHS interval: } [\,s_j^\ell, c_j^\ell\,], \qquad \text{RHS interval: } [\,c_j^r, s_j^r\,].$$

假设经过目视检查，决策者决定调整 V_j 的一侧。因此，我们提取位于区间 $[s_j^l, c_j^l]$ 或 $[c_j^r, s_j^r]$ 内的所有数据点 x_i 及其对应的隶属度值 $V_j(x_i)$。这些值构成隶属度的单调（递增或递减）序列。为了分析并构建该单调分支的形状，我们使用第 3 节中介绍的 C-FKM 算法，对相应区间内的隶属度值 $\{V_j(x_i)\}$ 进行聚簇。令 k_{side} 表示为细化所选的簇数量。对序列 $\{V_j(x_i)\}$ 应用 C-FKM 后，得到

$$\{c_1 < c_2 < \cdots < c_{k_{\text{side}}}\},$$

一组严格递增的质心值，每个值都代表一个“置信水平”，在该水平处，隶属度分支的斜率会发生变化。

这些质心可以解释为隶属度层级，在这些层级处，决策者可能希望引入语义区分：轻微下降、中度下降、快速下降等区域。在此阶段，决策者可以使用定理 1，通过卡片法调整这些质心。

请注意，由于分支严格单调，每个隶属度阈值 c_ℓ 在模糊数的对应区间内恰好对应一个点。因此，对于每个调整后的质心 c_ℓ，我们通过线性插值计算其对应的 x 坐标

$$x_\ell \;=\; \text{interp}(c_\ell;\, \mu_j(x_i), x_i),$$

即，满足 $V_j(x) = c_\ell$ 的唯一 x 值。这给出了断点

$$s_j^\ell = x_0 < x_1 < \cdots < x_{k_{\text{side}}} = c_j^\ell$$

沿对应一侧排列。将有序元组 (x_1， x_2，.....， $x_{k_{\text{side}}}$) 根据定理 1 转换为卡片组单位。由此得到一系列离散单位（卡片），表示相邻断点之间的距离，每张卡片对应隶属函数下降过程中的一个基本“步骤”。

决策者完成卡片调整后，使用累积卡片规则将修改后的卡片分布映射回 x 轴，从而得到更新后的断点 $\tilde{x}_1$、.....、$\tilde{x}_{k_{\text{side}}}$。随后，将所选一侧的精化隶属函数重构为一条经过这些点的分段线性函数

$$(\tilde{x}_1, c_1),\ \ldots,\ (\tilde{x}_{k_{\text{side}}}, c_{k_{\text{side}}}),$$

所得隶属函数仍然是正规的、凸的，且支撑集位于同一区间内。此精化步骤最终综合了 C-FKM 生成的数据驱动结构与决策者表达的语义调整。

例 3. 考虑经过步骤 *2* 得到的模糊数 V_2，其核与支撑集已通过验证

$$\text{Core}(A) = [0.46,\, 0.54], \qquad \text{Supp}(A) = [0.40,\, 0.60].$$

我们着重细化区间 $[0.40, 0.46]$ 上的左侧。假设检查后，决策者对步骤 *2* 中应用 ***C-FKM*** 得出的数值不满意。接着，我们使用 $k_{side} = 3$ 个簇对左侧隶属度值应用 ***C-FKM***。由此得到三个递增的质心隶属度水平：

$$c_1 \approx 0.05, \qquad c_2 \approx 0.58, \qquad c_3 \approx 0.91.$$

这些质心代表左侧分支上不同的“置信水平”。在此阶段，决策者可以看到这些层级对应的卡片数量，即0与 c_1之间有 *5* 个单位， c_1 与 c_2之间有 *52* 个单位， c_2 与 c_3之间有 *34* 个单位， c_3 与 1 之间有 *9* 个单位。假设他/她同意这些数量，并且不修改任何卡片。

由于 V_2 在左侧严格递增，因此每个 c_ℓ 都对应一个通过线性插值求得的唯一 x值。利用数据集中相邻值的隶属度，我们得到

$$x_{c_1} \approx 0.405, \qquad x_{c_2} \approx 0.441, \qquad x_{c_3} \approx 0.455.$$

因此，断点约为

$$(0.40,\ 0.405,\ 0.441,\ 0.455,\ 0.46).$$

使用精度为 $m = 3$ （即 $N = 1000$ 张卡片）的定理 *1*，我们计算出用于展示决策者所得结果的卡片数量，分别为 *93*、*591*、*232* 和 *84* 个单位。随后，决策者查看卡片结构并指出：

> “ x 的值与 c_1 和 c_2 的值之间差异过大。请将 *100* 张卡片从第二个区间移至第一个区间，使过渡更加平滑。”

按照此指示调整卡片分布后，使用累积卡片规则将结果重新转换为数值，以得到新的 x–断点

$$(\tilde{x}_1 = 0.4115, \tilde{x}_2 = 0.4410, \tilde{x}_3 = 0.4549),$$

并通过对这些点进行分段线性插值，重建隶属函数的左侧

$$(0.40, 0),\ (\tilde{x}_1, c_1),\ (\tilde{x}_2, c_2),\ (\tilde{x}_3, c_3),\ (0.46, 1).$$

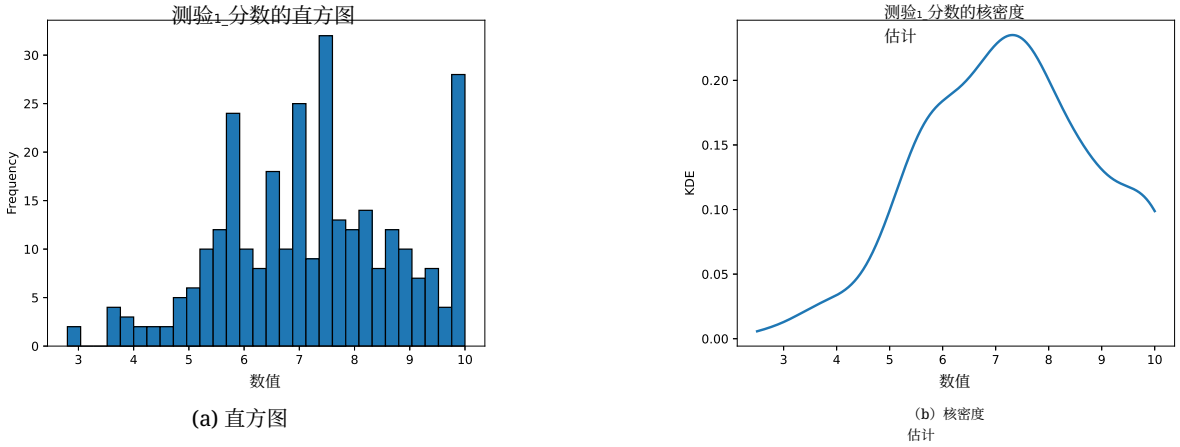
5. 数值示例

为说明所提出的混合方法，我们将其应用于从 Kaggle 获取的真实数据集。[1] 该数据集包含学业表现变量、人口统计描述项，以及与学生学习成果相关的行为指标。在本次数值示例中，我们

[1]https://www.kaggle.com/datasets/muhammadkhubaibahmad/student-表现-and-聚类-dataset

仅关注 quiz1_marks 变量，该变量记录每名学生在第一次平时测验中获得的分数。该变量特别适合我们的研究目的，因为其分布反映了学生表现的差异：表现较差的学生集中在尺度的低端，而相当一部分学生取得了中等和较高分数。

原始数据可能存在缺失值和离群值。因此，必须删除所选变量中包含NaN值的行。图1展示了所得直方图以及相应的核密度估计（KDE）。其形状呈轻度右偏，中等分数高度集中，并有一条较长的尾部延伸至高表现区域。该分布自然地促使构建五个描述低、中、高成就的模糊类。



(a) 直方图

（b）核密度估计

图1：测验1_分数数据的可视化

首先，我们采用第 3 节所述的 C-FKM 算法，参数为 $k = 5$ 和$m = 2$，并将中心点初始化为 $[a, b]$内均匀分布的值，即 $v_j = a + (b - a)\frac{j}{k+1}$，$j = 1$，2，3，4，5。需要说明的是，在我们的方法中，根据质心的固定顺序，每个点 x_i 只允许在两个相邻簇中具有非零隶属度。这保证了所得隶属函数在聚类阶段即为模糊数（正规、凸且具有紧支撑集）。对测验1_分数运行 C-FKM 后得到的质心如下：

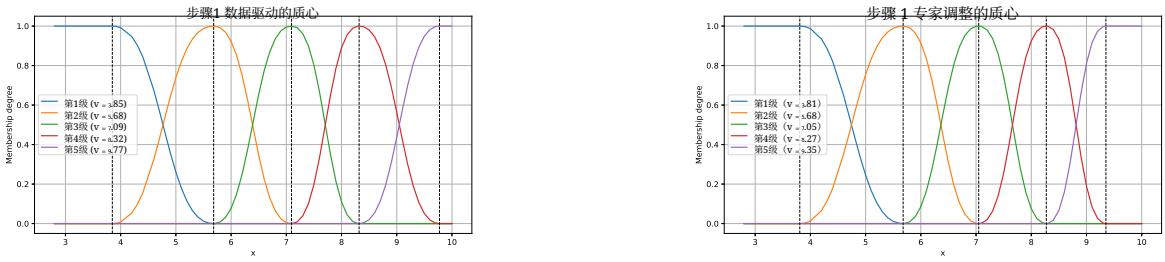
$$v_1 = 3.849, \quad v_2 = 5.683, \quad v_3 = 7.093 \quad v_4 = 8.318 \quad v_5 = 9.774$$

这些值是五个表现等级的数据驱动代表点。相应的隶属函数及质心如图2所示。这些质心被转换为卡牌单位，并展示给决策者进行验证。所得数值为

$$a = 2.8\ [14]\ v_1\ [26]\ v_2\ [19]\ v_3\ [17]\ v_4\ [20]\ v_5\ [4]\ b = 10$$

其中上限 a　　下边界 $[a, b]$ 已设为数据集的边界。鉴于此　　，

决策者认为最高绩效等级过高，因此将 v_4 与 v_5 之间的 5 张卡片移至 v_5 与 b 之间的区间。计算后得到的质心向量为 $v = (3.808, 5.68, 7.048, 8.272, 9.352)$



(a) 数据驱动　　(b) 专家调整

图2：从 C-FKM 获得的隶属函数

给定 C-FKM 隶属函数，每个模糊数的核定义为数据集中 $V_j(x) \geq 1 - \tau$ 且满足 $\tau = 0.01$ 的数值区间。支撑集定义为 $\mu_j(x) > 0$ 的最小区间，并且由于模型约束，可以直接根据核计算得到。因此，各核为 $[2.8, 3.9]$、$[5.6, 5.8]$、$[7.0, 7.1]$、$[8.2, 8.3]$ 和 $[9.3, 10.0]$，所得的下界和上界链，即 2。8、3。9、5。6、5。8、7、7。1、8。2、8。3、9。3、10 再次使用定理 1 转换为单位，并呈现给决策者进行调整：

$$c_1^- [15]\, c_1^+ [23]\, c_2^- [3]\, c_2^+ [17]\, c_3^- [1]\, c_3^+ [15]\, c_4^- [2]\, c_4^+ [14]\, c_5^- [10]\, c_5^+$$

检查这些数值后，决策者决定按如下方式更新单位：

$$c_1^- [14]\, c_1^+ [19]\, c_2^- [7]\, c_2^+ [14]\, c_3^- [5]\, c_3^+ [12]\, c_4^- [5]\, c_4^+ [14]\, c_5^- [10]\, c_5^+$$

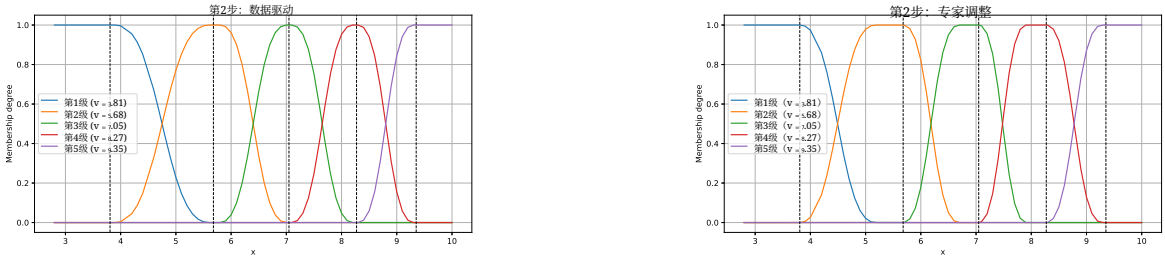
由此得到经验证的核的如下链：

$$[2.8, 3.808]\, [5.176, 5.68]\, [6.688, 7.048]\, [7.912, 8.272]\, [9.28, 10]$$

图3展示了专家干预前后核与支撑集的对比。请注意，经验证质心位于各自的核内。

我们向决策者展示新版隶属函数后，他确认对所有隶属度都满意，但第一核与第二核之间的变化情况除外。

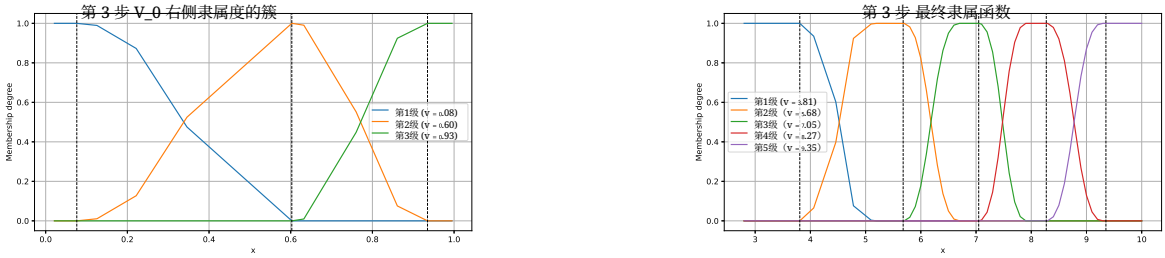
对于这一侧的取值，将位于核右侧的 V_0 隶属度进行聚类，以确定三个代表点，即0.08,0.6和0.93（见图4）。决策者认为这些隶属度已足够具有代表性，因此我们继续插值，并在 x 轴上求出相应的值（即相应 α–截集的上边界），分别为4.86,4.58和4.06。决策者认为这些值并不能代表隶属度的正确断点。因此，我们将它们表示为卡片，并在修改



（a）数据驱动　　（b）专家驱动

图3：验证核与支撑集后的隶属函数

各层级之间的单位后，最终得到经过验证的断点4.06、4.46和4.78。完成此步骤后，即得到最终隶属度，如图4所示。



(a) 隶属度聚类　　(b) 最终隶属函数

图4：流程结束时得到的隶属函数

6. 数据分布对模糊不确定性剖面的影响

图5、图6和图7展示了将该方法应用于三个具有不同分布特性的合成数据集时所得结果的比较：一个对称数据集、一个偏斜数据集和一个多峰数据集。由于这些数据分布不符合正态分布，本节将初始化质心设为百分位数 $100 \cdot j/(k+1)$，其中 $j = 1, \ldots\ldots k$。这项比较分析旨在评估混合方法如何适应数据中存在的不同不确定性结构，并考察所得模糊数是否恰当地反映了这些差异。

图 5 显示，第一个数据集可以用以尺度中点附近为中心的钟形分布很好地表示，而第二个数据集则明显右偏，大多数数值集中在较低区域。第三个数据集呈现出三个彼此分离的峰，表明基础总体中存在多个潜在子群。这些定性特征构成了构建模糊数的基础。

当模糊类的数量设为 $k = 3$ （图 6）时，由对称数据集得到的隶属函数呈现出均衡的形状，具有平滑过渡，且质心几乎等距。这种排列反映出不确定性均匀分布在某个中心概念周围的情况，定义域中没有任何区域在频率或相关性方面占据主导地位。然而，在偏斜的情况下，形状会明显变得不对称：最低类别形成了一个窄核心，并紧密

贴合数据的集中分布，而上层类别则将其支撑集广泛延伸至证据有限的区域。这种行为准确地体现了高值较为罕见、因而应与较高不确定性相关联这一概念。对于多峰数据集，该方法能够识别出与局部密度谷相吻合的不同过渡区域，从而揭示出单模态模型会掩盖的结构。

将类别数量增加到 $k = 5$（图7）可提高表示的粒度，并进一步凸显分布特征的影响。在对称数据集中，新增的模糊数排列有序、间距均匀，在不改变划分固有对称性的前提下增强了可解释性。相比之下，在偏斜和多模态数据集中，新增类别往往集中在数据密度最高区域，而在尾部或模态簇之间留下证据稀少的宽区间。形状变得更加复杂，但仍与底层观测值保持有意义的关联：当数据表明存在清晰边界时，隶属度会发生突变；而在较模糊或代表性不足的区域，则会出现缓慢过渡。

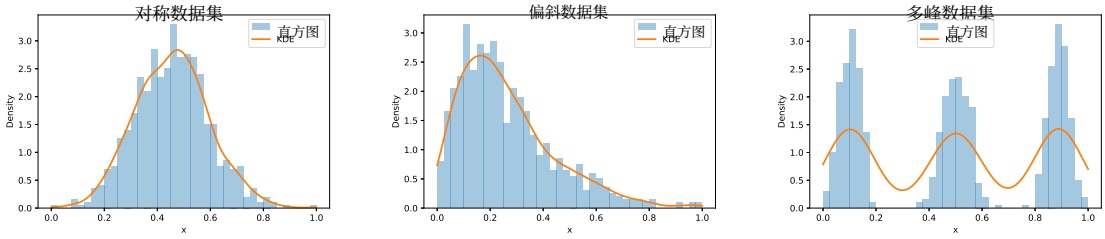
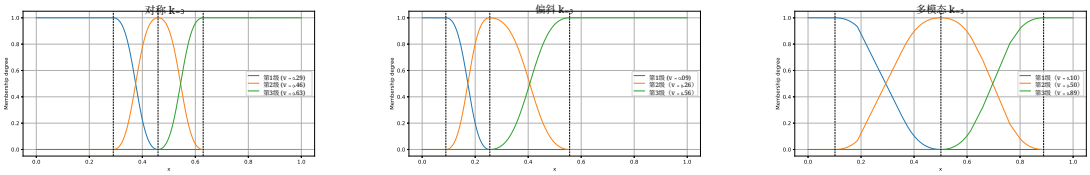
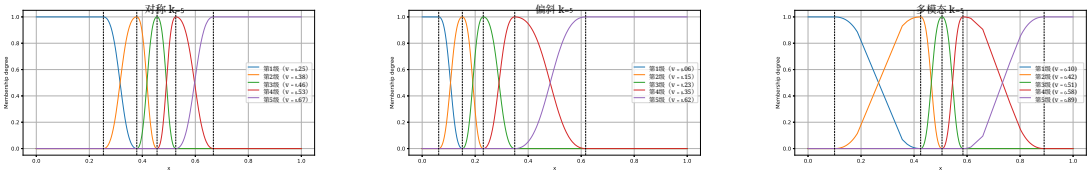


图5：数据的直方图和核密度估计



$k = 3$的模糊类



$k = 5$的模糊类

7. 结论

本研究提出了一种混合社会技术方法，通过 DoC方法将数值数据与专家引 elicitation 相结合，以构建模糊数。所提出的方

法通过引入一种基于经典模糊 k-means 聚类算法的新型凸版本的数据驱动流程，扩展了 DoC-MF 框架，确保计算阶段生成的每个模糊集在构造上都是模糊数。在每个步骤中，数值输出都会转换为卡片式单位，使领域专家能够借助透明且具有认知意义的表示方式，验证并完善所得隶属函数的形状、核、支撑集和斜率。

合成案例研究表明，该方法能够稳定地适应不同的分布结构，例如对称性、偏度和多峰性，同时不会强加僵化的函数形式，也不会牺牲可解释性。基于真实教育数据开展的数值实验进一步展示了该方法捕捉异质性表现特征的能力，同时保留了决策支持所需的语义含义。

该方法在纯数据驱动建模（可能缺乏可解释性）与纯专家驱动构建（可能无法反映实证证据）之间实现了有效的折中。由此得到的隶属函数忠实地表征了定量信息和定性判断，为下游决策过程提供了稳健的基础。

当前研究为未来研究开辟了新的方向。首先，我们将探索把其他距离度量纳入 C-FKM，以便更好地处理序数型、类别型或混合数据类型。其次，可通过将 C-FKM 扩展至多维情境，进一步完善该方法的第三个步骤，从而同时优化隶属函数的两侧，并捕捉更复杂的不确定性形状。最后，我们计划将该方法整合到完整的多准则决策分析框架中，评估其对现实决策问题中排序稳健性和偏好学习任务的影响。

利益冲突声明

作者声明不存在任何需要披露的利益冲突。

生成式人工智能及人工智能辅助技术在写作过程中的使用声明

在准备本文期间，作者使用了 Gemini 3（Google）和 Grammarly（均于2025年12月访问）来改善稿件的语言表达。使用这些工具后，作者根据需要审阅并编辑了内容，并对出版物的内容承担全部责任。

致谢

José Rui Figueira 获得葡萄牙资金资助，该资金由 FCT——科学技术基金会通过项目 UID/97/2025（CEGIST）提供。Diego García-Zamora 获得西班牙科学、创新与大学部流动资助 CAS24/00249 的资助，该资助支持他在 Instituto Superior Técnico、Universidad de Lisboa 进行研究访问。Diego García-Zamora 还感谢 CEGIST 在其研究访问期间提供的所有帮助。

参考文献

Ahmad Shukri, F. A., & Isa, Z.（2021）。基于专家判断的曼达尼型决策系统用于风险评估。*Mathematical Problems in Engineering*，*2021*年，6652419。Bezdek, J. C., Ehrlich, R., & Full, W.（1984）。FCM：模糊C均值聚类算法。*Com-puters & Geosciences*，*10*，191–203。URL：https://www.sciencedirect.com/science/article/ pii/0098300484900207。doi：https://doi.org/10.1016/0098-3004(84)90020-7。Bhattacharyya, R., & Mukherjee, S.（2020）。通过非线性回归评估模糊隶属函数：一种算法方法。*Fuzzy Information and Engineering*，*12*，412–434。Bilgiç, T., & Türkşen, I. B.（2000）。隶属函数的测量：理论与实证研究。载于 D. Dubois 与 H. Prade（编），*Fundamentals of Fuzzy Sets*（第195–227页）。波士顿，马萨诸塞州：Springer US。Chakraborty, D.（2001）。评价项目中语言专家意见模糊性的结构量化。《模糊集与系统》，*119*，171–186。doi：10.1016/S0165-0114(99) 00044-5。Chen, J., & Li, X.（2023）。基于异构信息的医生排名：考虑患者情绪强度的新评分函数。*Expert Systems With Applications*，*219*。Corrente, S., Figueira, J. R., & Greco, S.（2021）。纸牌法中的成对比较表与多准则决策辅助。*European Journal of Operational Research*，*291*，738–756。Dalleau, K., Couceiro, M., & Smaïl-Tabbone, M.（2020）。无监督极端随机树：一种计算异构数据相似性的随机方法。*Int. J. Data Sci. Anal.*，*9*，447–459。Dombi, J.（1990）。作为评价的隶属函数。《模糊集与系统》，*35*，1–21。Dombi, J., & Jónás, T.（2022）。使用中间语气算子构造隶属函数系统。《模糊集与系统》，*451*，206–227。Dubois, D., & Prade, H.（1983）。不公平硬币与必然性测度：迈向直方图的可能性解释。《模糊集与系统》，*10*，15–20。doi：https://doi.org/10.1016/ S0165-0114(83)80099-2。Dubois, D. J.（1980）。*Fuzzy sets and systems: theory and applications* 第144卷。Academic press。Dutta, B., García-Zamora, D., Figueira, J. R., & Martínez, L.（2025）。构建区间二型模糊隶属函数：一种基于卡片法的协同构建方法。Fan, Z.-P., & Liu, Y.（2010）。一种基于多粒度不确定语言信息的群体决策方法。*Expert Systems With Applications*，*37*，4000–4008。

Fang, G., Kwok, N. M., & Ha, Q. (2008). 使用粒子群优化自动调节模糊隶属函数。载于*2008*年*IEEE*亚太计算智能与工业应用研讨会（第324–328页）。第2卷。doi:10.1109/PACIIA.2008.105。García-Zamora, D., Dutta, B., Figueira, J. R., & Martínez, L. (2024). 运用纸牌法构建可解释模糊集以支持决策。*European Journal of Operational Research*, *319*, 246–262。George, S., & Santra, A. K. (2020). 一种改进的结合Takagi-Sugeno模糊模型的长短期记忆网络，用于考虑异常交通状况的交通速度预测。*ComputationalIntelligence*, *36*, 964–993。doi:10.1111/coin.12291。Hasuike, T., Katagiri, H., & Tsubaki, H. (2015). 一种基于统计学和信息论构造适当分段线性隶属函数的算法。《计算机科学进展》, *60*, 994–1003。Herrera, F., & Martinez, L. (2000). 一种用于词计算的二元组模糊语言表示模型。*IEEE*模糊系统汇刊, *8*, 746–752。doi:10.1109/91.890332。Hersh, H. M., & Caramazza, A. (1976). 一种针对自然语言中修饰语和模糊性的模糊集方法。*Journal of Experimental Psychology: General*, *105*, 254–276。Jain, S., & Khare, M. (2010). 用于城市机动车尾气排放建模的模糊隶属函数构造。*Environmental Monitoring and Assessment*, *167*, 691–699。doi:10. 1007/s10661-009-1085-4。Jiang, L., Liu, H., Ma, Y., & Li, Y. (2022). 从灵活语言偏好关系中推导语言信息的个性化个体语义。*Information Fusion*, *81*, 154–170。Kara, N., & Kocken, H. G. (2021). 一种采用双曲隶属函数、带混合约束的多目标固体运输问题模糊方法。*Cybernetics and InformationTechnologies*, *21*, 158–167。doi:10.2478/cait-2021-0049。Khairuddin, S. H., Hasan, M. H., Akhir, E. A. P., & Hashmani, M. A. (2022). 使用遗传调谐生成二型梯形模糊隶属函数。*CMC-Computers Materials &Continua*, *71*, 717–734。Khairuddin, S. H., Hasan, M. H., Hashmani, M. A., & Azam, M. H. (2021). 生成基于聚类的区间二型模糊三角形和梯形隶属函数：结构化文献综述。*Symmetry*, *13*, 239。doi:10.3390/sym13020239。Li, L., Liu, Y., Tu, Y., Zhou, X., & Lev, B. (2022). 一种基于多粒度比例犹豫模糊语言术语集的水资源风险评价群体TODIM方法。*Group Decision and Negotiation*, *31*, 913–944。

Li, Z., Zhou, Y., & Bao, R. (2019). 一种基于优化模糊词袋模型的图像分类方法。*Traitement Du Signal*, *36*, 239–244。doi:10.18280/ts.360306。Liang, H., Li, C.-C., Dong, Y., & Herrera, F. (2021). 基于个性化个体语义的语言意见动态。*IEEE*模糊系统汇刊，*29*，2453–2466。Majumder, D. D., Bhattacharyya, R., & Mukherjee, S. (2007). 隶属函数的评价与提取方法——综述及一种新方法。载于*2007*年国际计算理论与应用会议（*ICCTA'07*）（第277–281页）。doi:10.1109/ICCTA.2007.86。Martínez, L., Rodriguez, R. M., & Herrera, F. (2015). 二元语义模型。施普林格国际出版公司。Medasani, S., Kim, J., & Krishnapuram, R. (1998). 模式识别中隶属函数生成技术概述。《国际近似推理杂志》，*19*，391–417。Medvediev, I., Muzylyov, D., Shramenko, N., Nosko, P., Eliseyev, P., & Ivanov, V. (2020). 运用模糊逻辑设计逻辑语言模型，以计算农业货物物流中卡车的必要性。*Acta Logistica*，*7*，155–166。doi:10.22306/al.v7i3.165。Miliauskaite, J., & Kalibatiene, D. (2020). 一型隶属函数构造的通用框架：服务质量规划案例研究。《国际模糊系统杂志》，*22*，504–521。Nguyen, H. T., & Kreinovich, V. (2014). 如何在计算机系统中完整表示有关不精确属性的专家信息：随机集、模糊集及其他方法综述。《国际一般系统杂志》，*43*，586–609。doi:10.1080/03081079.2014.896354。Nieto-Morote, A., & Ruz-Vila, F. (2023). 模糊语言偏好模型中两两比较的术语集语义。*Entropy*，*25*，722。doi:10.3390/e25050722。Norwich, A., & Turksen, I. (1984). 一种隶属度测量模型及其经验实施的结果。《模糊集与系统》，*12*，1–25。doi:10.1016/ 0165-0114(84)90047-2。Porebski, S., & Straszecka, E. (2016). 模糊焦元的隶属函数。《控制科学档案》，*26*，395–427。doi:10.1515/acsc-2016-0022。Preud' Homme, G., Duarte, K., Dalleau, K., Lacomblez, C., Bresso, E., Smaïl-Tabbone, M., Couceiro, M., Devignes, M.-D., Kobayashi, M., Huttin, O., Ferreira, J. P., Zannad, F., Rossignol, P., & Girerd, N. (2021). 异构数据聚类方法的直接比较：仿真驱动基准测试。《科学报告》，*11*，4202。doi:10.1038/s41598-021-83340-8。Rodriguez, R. M., Martinez, L., & Herrera, F. (2012). 用于决策的犹豫模糊语言术语集。*IEEE*模糊系统汇刊，*20*，109–119。Ross, T. (2010). 《工程应用中的模糊逻辑》。（第3版）。约翰·威立父子出版公司。

Sancho-Royo, A., & Verdegay, J. L. (1999年). 隶属函数的构造方法。《国际智能系统杂志》，*14*，1213–1230。Schwaab, A. A., Nassar, S. M., & de Freitas Filho, P. J. (2015). 自动生成一型和区间二型模糊隶属函数的方法。《计算机科学杂志》，*11*，976–987。doi:10.3844/jcssp.2015.976.987。Soelistijanto, B. (2022年). 机会移动网络中模糊路由方案最优隶属函数的构造。*IEEE Access*，*10*，128498–128513。Spivak, M. (2008). 微积分。（第4版）。得克萨斯州休斯敦：Publish or Perish。Suryana, N., Panessai, I., & Shamsuddin, S. M. (2009). 遗传算法与模糊逻辑控制器隶属函数的设计。载于世界自然与生物启发计算大会论文集（第1753–1758页）。doi:10.1109/NABIC.2009.5393629。Turksen, I. (1991年). 隶属函数的测量及其获取。《模糊集与系统》，*40*，5–38。doi:10.1016/0165-0114(91)90045-R。Ukhobotov, V. I., & Krasil’nikova, E. S. (2017年). 模糊性测量方法。载于*2017*年第二届乌拉尔国际测量会议（*Uralcon*）（第391–396页）。南乌拉尔国立大学。第二届乌拉尔国际测量会议（UralCon），俄罗斯车里雅宾斯克，2017年10月16–19日。Wang, P., Liu, X., & Sanchez, E. (1986年). 集值统计及其在地震工程中的应用。《模糊集与系统》，*18*，347–356。Wang, W., Song, Q., Li, Y., & Zhang, N. (2022年). 基于双模糊逻辑控制器的电动汽车碳化硅功率逆变器主动热控制策略。《*IET*电力应用》，*16*，190–205。doi:10.1049/elp2.12146。Wang, X., Ruan, D., & Kerre, E. E. (2012年). 模糊数学——基本问题 第55卷模糊性与软计算研究。海德堡：Physica-Verlag。doi:10.1007/ 978-3-7908-1856-7。Wu, D., & Mendel, J. M. (2014年). 让实用的区间二型模糊逻辑系统设计变得简单。载于*2014*年*IEEE*模糊系统国际会议（*FUZZ-IEEE*）（第800–807页）。Yadav, H. B., & Yadav, D. K. (2015). 软件度量指标隶属函数的构造。载于P. Samuel（编），国际信息与通信技术会议论文集，*ICICT 2014*（第933–940页）。《计算机科学进展》第46卷。Ye, J., Sun, B., Chu, X., Zhan, J., Bao, Q., & Cai, J. (2023). 多源异构模糊决策系统上的一种新型多样化属性群决策方法及其在痛风诊断中的应用。*IEEE*模糊系统汇刊，*31*，1780–1794。