评价决策类问题

应用场景：对含有多项指标的不同项目之间进行评价与决策
视频教程：1-1-1 评价决策类-层次分析法模型精讲_哔哩哔哩_bilibili

1、层次分析法

1）问题的层次化

> 目标层：只有一个元素，代表分析问题的预定目标或理想结果
> 准则层：包含为实现目标所涉及的中间环节
> 方案层：包含为实现目标可供选择的各种措施、决策方案等

2） 基本步骤

- 建立层次结构模型
- 构造个层次中的判断矩阵：对指标的重要性两两比较，矩阵中元素 a_{ij} 表示第 i 项指标相比第 j 项指标的重要程度，具体标度与含义如下：

标度	含义
1	同等重要
3	稍微重要
5	明显重要
7	强烈重要
9	极端重要
2、4、6、8	上述相邻判断的中间值

> 显然，a_{ij} 与 a_{ji} 为倒数，$a_{ii}=1$

- 一致性检验 :对判断矩阵的构建中两两比较，需满足判断关系的*传递性*，即

$$a_{ij}=a_{ik}*a_{kj}$$

  计算一致性检验的步骤：

> 1. 计算一致性指标
>
> $$CI = \frac{\lambda_{max}}{n-1}$$
>
> 2. 查找对应的平均随机一致性指标 RI
> 3. 计算一致性比例
>
> $$CR = \frac{CI}{RI}$$
>
> CR < 0.1 时，认为判断矩阵为一致矩阵

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46

- 求权重

> 算术平均法：
>
> > 1. 判断矩阵归一化（按列归一）
> > 2. 按行求和，得到的元素除以 n 即为权重向量
> >   几何平均法：
> > 3. 判断矩阵按行相乘得到新的列向量
> > 4. 新矩阵的各元素开 n 次方
> > 5. 归一化
> >   特征值法：
> > 6. 求出矩阵的最大特征值以及对应的特征向量
> > 7. 对求出的特征向量进行归一化

- 求最终评价

层次分析发的 python 实现

使用 numpy 实现

2、Topsis 法（优劣解距离法）

基本概念：

> 理想解：取备选方案中各项数据的最好值得到的解
> 负理想解：取备选方案各项数据中的最差值得到的解

基本步骤：

1. 将原始矩阵正向化（将所有指标转换为极大型指标，即指标对应数值越大该项指标越优）
   对于指标，可以根据需求分为以下几类

指标名称	指标特点	转换公式
极大型指标	越大越好	无
极小型指标	越小越好	$x' = x_{max} - x$
中间型指标	越接近某个值越好	对一组中间型序列$\{x_i\}$，最优值为 x_{best}，$M = \max \lVert x_i - x_{best} \rVert, x_i' = 1 - \frac{\lVert x_i - x_{best} \rVert}{M}$
区间型指标	落在某个区间最好	对一组区间型序列$\{x_i\}$，最优区间为 $[a, b]$，$M = \max \lVert a - \min\{x_i\}, \max\{x_i\} - b \rVert, x_i' = \begin{cases} 1 - \frac{a - x_i}{M}, & x_i < a \\ 1, & a \le x_i \le b \\ 1 - \frac{x_i - b}{M}, & x_i > b \end{cases}$

2. 正向矩阵标准化

> 对标准化矩阵 Z 中的每个元素：
>
> $$Z_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}^2}$$

> 标准化后，*可以*对不同指标附上权重，确定权重的方法有层次分析法、熵权法、Delphi 法、对数最小二乘法等
> 3. 计算得分并归一化
>
> $$S_i = \frac{D_i^-}{D_i^+ + D_i^-}$$
>
> 其中，S_i 为得分，D_i^+ 为评价对象与最大值的距离，D_i^- 为评价对象与最小值的距离
> 定义最大值 $Z^+ = \{\max\{z_{11}, z_{21}, \ldots z_{n1}\}, \ldots, \max\{z_{1n}, z_{2n}, \ldots z_{nn}\}\}$
> 同理可以定义最小值 Z^-
> 此时即可定义第 i 个评价对象与最大值之间的距离 $D_i^+ = \sqrt{\sum_{j=1}^{m}\left(Z_j^+ - z_{ij}\right)^2}$ 与 D_i^-
> 计算得分后进行归一化处理得到 S_i

3、熵权法

> 指标离散程度越大，反映的信息量就越多，权重就越大
> 若某项指标的值全部相等，则该指标在综合评价中不起作用

1. 数据标准化
2. 计算概率矩阵 P（即归一化处理）
3. 计算熵权

$$e_j = -\frac{1}{\ln n}\sum_{i=1}^{n} p_{ij}\ln p_{ij}\ (\text{j = 1,2,...,m}) \quad d_j = 1 - e_j \quad W_j = \frac{d_j}{\sum_{j=1}^{m} d_j}$$

> 由 e_j 定义易知，当 $p_{1j} = p_{2j} = \ldots = p_{nj} = \frac{1}{n}$ 时，$e_j = 1$，代表信息熵最大，信息效用值 d_j 最小
> 对信息效用值 d_j 进行归一化处理即可得到熵权 W_j

4、模糊综合评价

> 对于无法精确评价的*模糊性概念*进行评价的模型进行评价

模糊集合与隶属函数的一般概念

模糊集合

- 用来描述模糊性概念
- *元素可以同时存在于对立的两个模糊集合中*
- 使用*隶属函数*进行刻画

隶属函数

$$\mu_A: \quad X \to [0,1] \quad x \to \mu_A(x)$$

称 μ_A 为 A 的隶属函数，$\mu_A(x)$ 为 x 对模糊集 A 的隶属度，其中，当 $\mu_A(x) = 0.5$ 时认为最具有模糊性

模糊集合的表示方法

当论域 X 为有限集时，记 $X = \{x_1, x_2, \ldots, x_n\}$

1. Zadeh 表示法：

$$A = \sum_{i=1}^{n} \frac{\mu_A(x_i)}{x_i}$$

2. 序偶表示法：

$$A = \{(x_1, \mu_A(x_1)), (x_2, \mu_A(x_2)), \ldots, (x_n, \mu_A(x_n))\}$$

3. 向量表示法：

$$A=(\mu_A(x_1),\mu_A(x_2),\ldots,\mu_A(x_n))$$

隶属函数的确定方法

1. 模糊统计法：使用统计手段，隶属频率定义隶属度
2. 借助已有的客观尺度：例，用恩格尔系数来衡量模糊集合“小康家庭”
3. 指派法（比赛常用）：根据主观意愿指派一个隶属函数
4. 其他方法：如德尔菲法、二元对比排序法、综合加权法等...

单层次模糊评价问题概述

- 引入三个集合：因素集 U、评语集 V、权重集 A
- 确定由 U 到 V 的模糊综合判断矩阵

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & \cdots & \cdots & r_{nm} \end{bmatrix}$$

其中，r_{ij} 代表指标 u_i 对评语 v_j 的隶属度

- 进行矩阵合成运算

$$B=AOR$$

$$b_j=(a_1\bullet r_{1j})+(a_2\bullet r_{2j})+\ldots+(a_n\bullet r_{nj})$$

其中，$\bullet$ 为实数乘法运算，+为 $\oplus$ 运算，$a\oplus b=\min(1,a+b)$
取数值最大的评语为综合评判结果

多层次模糊评价问题

若因素集 U 中的因素众多，可以将其按属性分隔为 s 个不重合的子集，对每个子集 U_i 分别求出评价矩阵并分别作出综合决策，得到一级评价向量 $B_i=A_iOR_i$
再将每个 U_i 视为一个元素，再进行一次单层次模糊综合评价，其中，按照不同 U_i 反映属性的重要程度得到二级权重 A
此时二级模糊综合评价模型：$B=AOR$

> 可以视因素 U 的复杂情况划分为三级或更高级模型

5、灰色关联分析

> 根据序列曲线几何形状的相似程度判断其联系是否紧密。曲线越接近，相应序列之间的关联度就越大
> 优点：对样本量的多少与样本有无规律都适用、计算量小
> 缺点：主观性较强，部分指标最优值难以确定（不建议在美赛中使用）

关联分析步骤

- 母序列（参考序列）：能反应系统行为特征的数据序列，记为

$$x_0 = \{x_{01}, x_{02}, \ldots, x_{0m}\}^T$$

- 子序列（比较序列）：由影响系统行为的因素组成的数据序列，记为

$$X_{nm} = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

- 数据预处理：对每个元素进行标准化（每个元素除以对应指标的均值）
- 计算灰色关联系数：计算子序列各个指标与母序列的关联系数

$$a = \min_i \min_k |x_0(k) - x_i(k)| \quad b = \max_i \max_k |x_0(k) - x_i(k)|$$

构造：

$$\xi_i(k) = y(x_0(k), x_i(k)) = \frac{a + \rho b}{|x_0(k) - x_i(k)| + \rho b}$$

其中 ρ 为分辨系数，一般取 0.5
- 对每个指标的关联系数分别求平均值，得到的就是灰色关联度，其值越高则影响力越大

主成分分析法

> 对变量进行降维，降低复杂性

假设有 n 个样本，p 个指标，构成 n x p 的样本矩阵 x

1. 进行标准化处理：标准化数据 $X_{ij} = \frac{x_{ij} - \overline{x}_j}{S_j}$

$$S_j : \text{标准差} \quad x_j : \text{均值}$$

2. 计算标准化样本的协方差矩阵/样本相关系数矩阵 R（numpy 指令实现即可）
3. 计算 R 的特征值 $\lambda_1, \ldots, \lambda_p$ 与特征向量 $a_1, \ldots, a_p$，并按降序排列
4. 计算主成分贡献率 $\alpha_i = \frac{\lambda_i}{\sum_{k=1}^{p} \lambda_k}$，累计贡献率 $\sum G = \frac{\sum_{k=1}^{i} \lambda_k}{\sum_{k=1}^{p} \lambda_k}$
5. 一般取累计贡献率超过 80%的特征值为对应的第一、第二、...、第 m 个主成分，第 i 个主成分 $F_i = a_{1i}X_1 + \ldots + a_{pi}X_p$
6. 根据系数分析主成分的意义：对某个主成分而言，指标前的系数越大，代表该指标对于该主成分的影响越大
7. 利用主成分结果进行后续的分析
   - 主成分得分
   - 聚类分析

- 回归分析