

Задача ранжирования объектов на основе экспертных оценок и показателей

Мокров Петр

Московский Физико-Технический Институт

Математические Методы Прогнозирования

07.04.2020

Постановка задачи

Заданы множества объектов $Y = \{y_i\}_{i=1}^N$, каждый объект y_i описывается набором показателей (признаков) $x_{i,1}, x_{i,2}, \dots, x_{i,K}$. Кроме того, имеется M экспертов, каждый из которых ранжирует (определяет частичный порядок) на множестве объектов.

Задача

Данные: Набор продуктов питания, описываемые следующими характеристиками: калорийность, цена 100 г., белки, жиры, углеводы. Также даны рейтинги нескольких экспертов.

Цель: Построить интегральный индикатор (ранжирование) исследуемых объектов, который был удовлетворял требованию потенциального заказчика, сравнить полученный индикатор с базовым индикатором

Исследуемая проблема

Построение рейтинга предпочтения одних экспертов другим

Метод решения задачи: Предобработка

Будем интерпретировать данные (показатели и оценки экспертов) как дискретные вероятностные распределения (на множестве объектов):

- 1 *Ранжирование эксперта*: будем назначать вероятностную массу обратно пропорционально рангу, данному экспертом
- 2 *Показатель*: предполагаем выполнение свойства *tbtb*, заменяем значения показателя на ранг в соответствии с величиной показателя и действуем аналогично предыдущему случаю

Работа с пропусками Простым и естественным способом заполнения пропусков (в контексте используемых вероятностных распределений) является использование фиксированной вероятностной массы, назначаемой пропуску (значение зависит от задачи)

Определение

Пусть $x_1, x_2, \dots, x_{n_x}; y_1, y_2, \dots, y_{n_y} \in (\mathbb{R}^n, \|\cdot\|)$, на точках x_1, \dots, x_{n_x} и y_1, \dots, y_{n_y} заданы соответственно дискретные распределения \mathbb{P}_x и \mathbb{P}_y . Тогда расстояние вассерштейна между ними (здесь $M \in \mathbb{R}_+^{n_x \times n_y} : M_{i,j} = \|x_i - y_j\|$):

$$W(\mathbb{P}_x, \mathbb{P}_y) = \min_{\gamma \in \mathbb{R}_+^{n_x \times n_y}} \sum_{i,j} \gamma_{i,j} M_{i,j}$$

Определение

Пусть на точках $\{x_j^i\}_{j=1}^{n_i} \subset \mathbb{R}^n$ заданы дискретные распределения \mathbb{P}_i , $\{x_j^0\}_{j=1}^{n_0} \subset \mathbb{R}^n$. Барицентром распределений \mathbb{P}_i на $\{x_j^0\}_{j=1}^{n_0}$ называется $(\alpha_i \geq 0)$:

$$\mathbb{P}_m = \arg \min_{\mathbb{P}: \text{supp}(\mathbb{P}) = x^0} \sum \alpha_i W(\mathbb{P}, \mathbb{P}_i)$$

Идея

Решение основной задачи. В качестве интегрального индикатора мы будем использовать барицентр вероятностных распределений, заданных показателями и оценками экспертов. В качестве supp всех распределений будем рассматривать эквидистантно расположенные точки в \mathbb{R}^{N-1} .

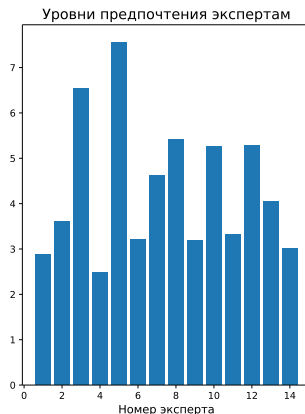
Ранжирование экспертов. В качестве меры предпочтительности одного эксперта другому мы рассматриваем $\frac{1}{d_i}$, $d_i = W(\mathbb{P}_m, \mathbb{P}_i)$

Информация к размышлению

- Использовать свободу в выборе точек $\{x_j^i\}_{j=1}^{n_i}$. (Например, интерпретировать расстояние до точки как "уверенность" в рейтинговой отметке)
- Использовать outlier detection методы для построения метрики доверия экспертам

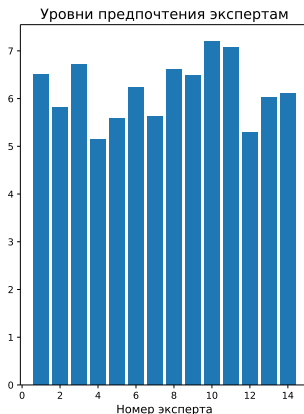
Практика: Построение интегральных индикаторов

- Константа заполнения пропусков, используемая в эксперименте = 0
- Код доступен в репозитории <https://github.com/PetrMokrov/Ratings>



Практика: Построение интегральных индикаторов

- Константа заполнения пропусков, используемая в эксперименте $= \frac{1}{10} = \frac{1}{N}$ (N - количество экспертов)



Практика: Базовый интегральный индикатор

В основе базового интегрального индикатора лежит предположение о том, что близкие (в смысле ранжирования экспертами/показателями) объекты должны иметь близкие значения интегрального индикатора. Напомним определение:

Определение

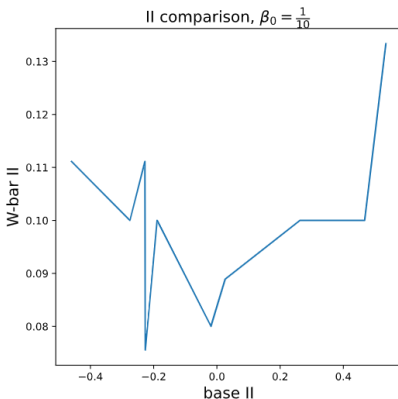
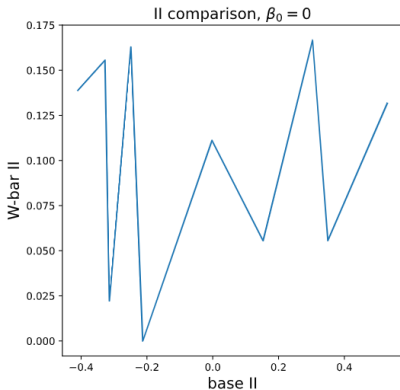
Пусть даны две выборки x_1, \dots, x_n и y_1, \dots, y_n . Тогда коэффициентом корреляции Кендалла τ называется:

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j)$$

Построим матрицу $M \in \mathbb{R}^{N \times N}$ попарных корреляций между объектами, рассматривая ранги экспертов/показателей в качестве выборок, ассоциированных с объектом. Тогда в качестве интегрального индикатора будем использовать первую главную компоненту этой матрицы.

Практика: Сравнение с базовым индикатором

- β_0 есть константа заполнения пропусков
- Есть подозрение, что используемый базовый интегральный индикатор работает не очень хорошо



Спасибо за внимание!