

Построение интегрального индикатора на основе Парето расслоения и определение критической массы экспертов

Вареник Наталия Викторовна

Московский физико-технический институт
Физтех-школа прикладной математики и информатики
Кафедра интеллектуальных систем

Москва, 2021

Задача

Требуется построить рейтинг продуктов питания. Продукт питания описан набором числовых признаков и экспертных оценок.

Требования к модели построения рейтинга

- Устойчива при добавлении новых объектов.
- Работает с пропусками в данных.
- Обеспечивает полный порядок на объектах.

Решение

Использование метода Парето расслоения на экспертных оценках продуктов.

Постановка задачи

Дано

$\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^n$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$ - признаковое описание i -го объекта, где $x_{ij} \in L_j$ - элемент частично упорядоченного множества $\forall j = \overline{1, m}$.

Статистические предположения о данных

- Независимость и сравнимость объектов.
- Мультимодальность (с конечным количеством максимумов) распределения оценок экспертов.

Задача

Построить интегральный индикатор рейтинга со свойствами:

- Похожие объекты рядом по интегральному индикатору.
- Чем больше тем лучше.
- Устойчив во времени, ожидаем и очевиден.

Парето расслоение: определение

Отношение недоминирования

Объект a не доминируем, если

$$\nexists b_i : b_{ij} \geq a_j \quad \forall j.$$

Если два объекта не доминируют друг друга, то они лежат в одном классе эквивалентности.

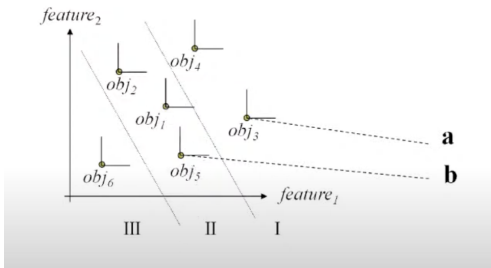


Иллюстрация Парето расслоения.

Алгоритм

- 1 Для набора точек строится POF (недоминируемые точки).
- 2 Производится исключение точек, образующих POF.
- 3 Построение POF повторяется для оставшихся точек.

Предположения метода

Признаки допускают введение отношения частичного порядка.

В предлагаемом решении выполнено, так как в качестве признаков используются экспертные оценки.

Свойства

- Устойчив относительно монотонного неубывающего преобразования признаков.
- Неустойчив относительно вращения базиса.

Полученный интегральный индикатор

- *Заполнение пропусков* - медиана экспертных оценок продукта.
- Продукты рассматриваются в пространстве экспертных оценок. Несравнимость продуктов возникает в случае, если все эксперты поставили этим продуктам одинаковые оценки, в нашем случае это исключено по правилам проставления оценок.
- Поиск недоминируемого продукта соответствует нахождению продукта с максимальной суммой координат в пространстве экспертов. *Интегральный индикатор:*

$$R_i = - \sum_{j=1}^J \tilde{r}_{ji},$$

R_i - интегральный индикатор i -го объекта,

\tilde{r}_j – нормированный вектор оценок j -го эксперта с заполненными пропусками.

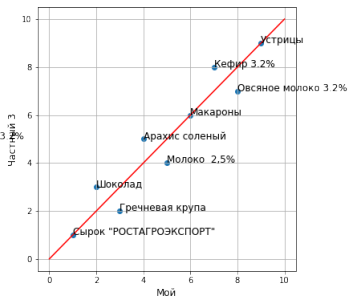
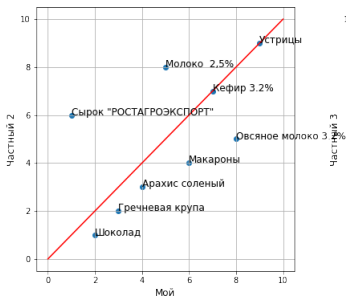
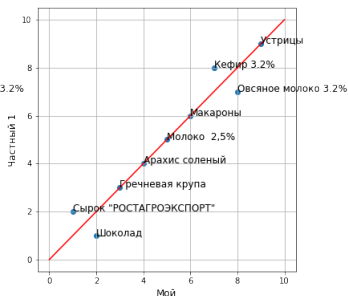
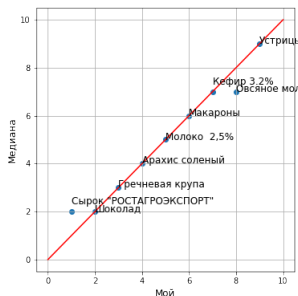
Взвешивание экспертов

В предложенном решении не предполагается, так как Парето расслоение инвариантно к линейным изменениям признаков ($w > 0$), а признаки это экспертные оценки.

Сведение к пространству продуктов

Если рассматривать экспертов в пространстве продуктов, то нахождение недоминируемого продукта соответствует выбору оси с максимальной суммой ее значений.

Сравнение полученного рейтинга с другими



Отказ от принятия решения: метод 1

- На отложенной выборке строится зависимость расстояния Левенштейна полученного ответа до ответа на выборке без пропусков от количества пропусков в выборке.
- Отказ производится при превышении выбранного порога на количество пропусков.
- Критерий применимости: выборка из того же распределения, что и отложенная.



Отказ от принятия решения: метод 2

- Предлагается использовать вспомогательный метод ранжирования, слабо чувствительный к наличию пропусков в данных.
- Отказ производится в том случае, если расхождение между ответами (расстояние Левенштейна) больше эмпирически выбранного порога.

Определение критической массы экспертов

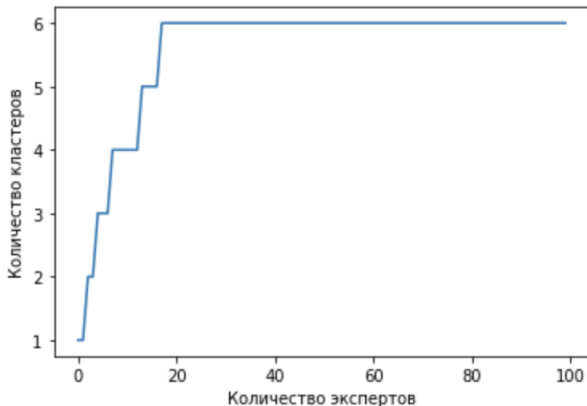
Гипотеза

При кластеризации экспертных оценок в пространстве объектов начиная с некоторого объема количество кластеров перестанет увеличиваться. Что соответствует согласованности экспертов в случае их адекватности.

Данные

- Из всех перестановок K объектов случайным образом выбирается N штук. Они соответствуют определенными типам мнений экспертов.
- Данные сгенерированы путем наложения шума на выбранные перестановки с добавлением случайного веса на вхождение зашумлённой перестановки в выборку.
- Часть значений в данных пропущена, для выполнения требований работы с пропусками.

Определение критической массы экспертов: эксперимент



Для графика сгенерированы данные с 5 типами мнений экспертов, 6 кластер образовался в результате наличия пропусков в рейтингах.