

Измерение расстояния между объектами с пропусками при помощи метода попарного сравнения

Бишук Антон Юрьевич

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Москва
2021 г

Цель

Предложить метод, способный определять расстояние между объектами путем попарного сравнения объектов между собой

Решаемая проблема

Описания объектов могут содержать пропущенные значений некоторых показателей. Это означает то, что мы не знаем значение признака для текущего объекта, однако всё же необходимо объекты сравнить между собой.

Методы решения

Представим ряд подходов, применение которых, позволяет сравнивать объекты между собой, даже если некоторые значения неизвестны.

- 1 Заполнить все пропуски одним специальным символом, считать уровень доверия.
- 2 Заполнить значение средним по группе признаков (среднее арифметическое/медиана/мода).

Методы сравнения

- 1 Косинусное расстояние,
- 2 Поэлементное сравнение.

Данный подход сфокусирован на том, чтобы присвоить эксперту некоторые величины, которые будут характеризовать наше доверие к его рейтингу.

- 1 Заменяем все пропуски подобранным специальным символом и далее убираем эти признаки и рассмотрения.
- 2 Считаем косинусное расстояние между текущим вектором и всеми остальными и выбирает ту пару, расстояние между которыми минимально.
- 3 При поэлементном сравнении: для каждой пары если элементы совпадают, то увеличиваем счетчик на 1 и в конце нормируем это на длину объекта.
- 4 Заполняем пропуски теми значениями, которые стоят у самого похожего на рассматриваемый объект. В случае если среди самых похожих также нет этого значения, то заменяем его на специальный символ группы, который выбран из определенной логики.
- 5 Пропуски в признаках не являющихся рейтингом заполним согласно графику, полученному в осях (признак, средний рейтинг продукта)
- 6 Присваиваем объекту степень доверия.
- 7 Ранжируем оценки согласно средним по продуктам в случае рейтинга, по значению угла наклона (признак, средний рейтинг) (если необходимо перевести признак в другие координаты) и по уровню доверия.

Сравнение рейтингов

