

Векторные метрики близости объекта оценки и объектов - аналогов

Слуцкий А.А., к.т.н.,
Заседание МСЭО 25.07.2025

Что такое метрики?

Если опустить математические обозначения и ограничиться вербальным изложением, в математическом анализе метрика - расстояние, а точнее - функция, определённая для любых двух элементов x, y из множества X , удовлетворяющая условиям :

1. расстояние между точками всегда положительно или равно нулю, кроме случаев, когда они совпадают;
2. расстояние от одной точки до другой такое же, как и расстояние от второй точки до первой (симметрия);
3. прямое расстояние между двумя точками не превышает расстояние через третью точку (неравенство треугольника - длина любой из сторон треугольника не больше суммы длин двух других его сторон).

Множество X с введённой на нём метрикой, называется метрическим пространством. Существенно, что на одном и том же множестве X могут рассматриваться разные метрики, при этом получаются различные метрические пространства, что является абсолютно нормальным.

Таким образом, с позиций математического анализа предлагаемое далее использование ряда общепринятых метрик представляет собой рассмотрение множества, состоящего из объекта оценки и объектов - аналогов в разных метрических пространствах.

Обратим внимание на то, что предлагаемый подход основан на строгой базе научно - математической теории, а, стало быть, обеспечивает выполнение требования необходимости исследуемых причинно - следственных связей - одним из двух, наряду с достаточностью - требований к достоверности результата умозаключения.

С профессиональной оценочной точки зрения это означает использование разных методов оценки в рамках одного - сравнительного - подхода к оценке.

Векторные метрики близости

Предлагается использование в практике оценки следующих векторных метрик «близости» объекта оценки и объекта - аналога.

Метрики семейства метрики Минковского	Иные
1. Манхэттен	5. Косинусное сходство
2. Канберра	
3. Евклидова	6. <u>Махаланобиса</u>
4. Чебышева	

Метод балансировки матриц - тоже по смыслу векторный - сюда не включён, но имеется в виду, как седьмой метод количественной эндогенной корректировки цен объектов - аналогов.

Качественные методы эндогенной корректировки цен объектов - аналогов представляют собой иной класс методов, направленных на достижение той же цели. Их надо рассматривать отдельно.

Метрики семейства метрики Минковского

Формулы приведены для одного - j - того - объекта - аналога, имеющего, как и объект оценки i ценообразующих факторов (элементов сравнения).

Общий случай.

Расстояние Минковского или метрика Минковского (МетМин) - это метрика в нормированном векторном пространстве.

В «переводе на оценочный язык»

Метрики семейства метрики Минковского (2)

$$\text{МетМин} = \left(\sum_{i=1}^n |\text{ЦОФ}_{iоо} - \text{ЦОФ}_{iоа}|^p \right)^{1/p}$$

где

ЦОФ_{іоо} - нормализованное значение *i* - того ценообразующего фактора объекта оценки;

ЦОФ_{іоа} - нормализованное значение *i* - того ценообразующего фактора объекта аналога.

p — целое число.

Метрики семейства метрики Минковского (3)

Для $p \geq 1$, метрика Минковского является метрикой в силу неравенства Минковского.

Обычно расстояние Минковского используется при p равном 1 или 2. При этом:

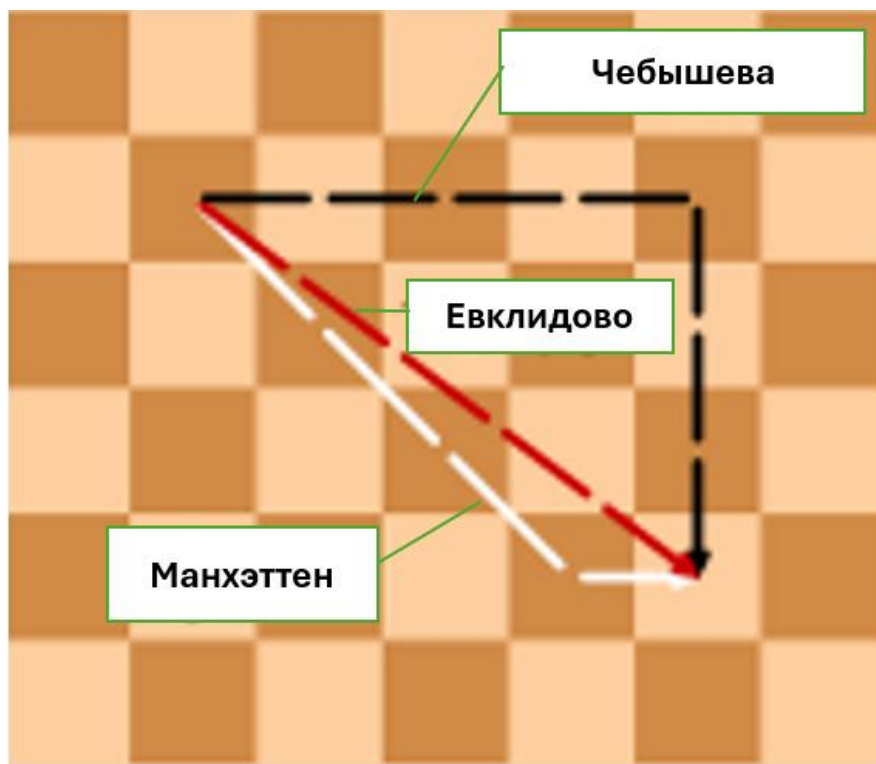
- ▶ при $p = 1$ метрика Минковского соответствует метрике Манхэттен
- ▶ при $p = 2$ метрика Минковского соответствует евклидовой метрике.

Кроме того, в предельном случае, когда p стремится к бесконечности, метрика Минковского соответствует метрике Чебышева.

Таким образом, вместо использования метрики Минковского в общем виде лучше использовать три отдельные метрики, имеющие внятное толкование.

Метрики семейства метрики Минковского (4)

Соотношение метрик на шахматной доске



Метрика Манхэттен

Представляет собой сумму модулей разностей координат точек в n-мерном вещественном векторном пространстве с заданной системой координат, или сумму длин проекций такого отрезка, который соединяет точки на оси координат.

$$\text{МетМан} = \sum_{i=1}^n |\text{ЦОФ}_{i00} - \text{ЦОФ}_{i0a}|$$

Метрика Канберра

Взвешенная версия метрики Манхэттен:

$$\text{МетМан} = \sum_{i=1}^n \frac{|\text{ЦОФ}_{i00} - \text{ЦОФ}_{i0a}|}{|\text{ЦОФ}_{i00}| + |\text{ЦОФ}_{i0a}|}$$

Евклидова метрика

Метрика в евклидовом пространстве — расстояние между двумя точками евклидова пространства, вычисляемое по теореме Пифагора.

$$\text{МетЕвкл} = \sqrt{\sum_{i=1}^n (\text{ЦОФ}_{i00} - \text{ЦОФ}_{i0a})^2}$$

Метрика Чебышева

Метрика на векторном пространстве, задаваемая как максимум модуля разности компонент векторов.

$$\text{МетЧeб} = \max_{i=1}^n |\text{ЦОФ}_{i00} - \text{ЦОФ}_{i0a}|$$

Иные метрики. Косинусное сходство

Мера сходства между двумя ненулевыми векторами, заданными в пространстве со скалярным произведением — косинус угла между векторами, то есть скалярное произведение векторов, делённое на произведение их длин

$$\text{МетКос} = \frac{\sum_{i=1}^n \text{ЦОФ}_{i00} \times \text{ЦОФ}_{i0a}}{\sqrt{\sum_{i=1}^n (\text{ЦОФ}_{i00})^2} \times \sqrt{\sum_{i=1}^n (\text{ЦОФ}_{i0a})^2}}$$

Недостатки приведённых метрик

Недостатком всех указанных метрик является то, что они не учитывают возможную корреляцию между ценообразующими факторами. От этого недостатка свободна метрика Махаланобиса.

Иные метрики. Метрика Махаланобиса

Мера расстояния между векторами случайных величин, обобщающая понятие евклидова расстояния. Позволяет определять сходство неизвестной и известной выборки. Оно отличается от евклидовой метрики тем, что учитывает корреляции между переменными и инвариантно к масштабу.

Формально, расстояние Махаланобиса от многомерного вектора $x = (x_1, x_2, x_3, \dots, x_N)^T$ до множества со средним значением $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_N)^T$ и матрицей ковариации S определяется следующим образом

$$\text{МетМах} = \sqrt{(\text{ЦОФ}_{i00} - \text{ЦОФ}_{i0a})^T x (S)^{-1} (\text{ЦОФ}_{i00} - \text{ЦОФ}_{i0a})}$$

Иные метрики. Метрика Махаланобиса (2)

По причине наличия в расчёте метрики Махаланобиса матрицы ковариации, эта метрика учитывает возможную корреляцию между ценообразующими факторами. Однако, в отличие от предыдущих метрик, она основана на предположении о том, что точки множества сферически распределены вокруг центра масс (то есть равномерно по всем измерениям). Если же распределение явно не сферическое (например, эллипсоидальное), то было бы естественным учитывать в вероятности принадлежности не только расстояние до центра масс, но и направление на него.

Однако, как представляется, указанное направление учитывается метрикой косинуса. Поэтому одновременное использование метрик и их сопоставление может обеспечить дополнительное обоснование результата.

Нормализация значений ценообразующих факторов

Поскольку значения ценообразующих факторов имеют разную размерность, могут сильно различаться по порядкам величин и т.д. до вычисления метрик их надо нормализовать - привести к безразмерным единицам и одному интервалу изменения.

При этом нормализация улучшает производительность алгоритмов, основанных на расстояниях, таких как k-ближайших соседей .

Необходимо подчеркнуть, что в силу отсутствия нормальности распределения данных в отличие от нормализации, процедура стандартизации данных в данном случае не применима.

Нормализация осуществляется путём несложного минимаксного масштабирования.

Нормализация значений ценообразующих факторов (2)

$$\text{ЦОФ}_{\text{норм}i} = \frac{\text{ЦОФ}_i - \text{ЦОФ}_{\text{мин}i}}{\text{ЦОФ}_{\text{макс}i} - \text{ЦОФ}_{\text{мин}i}}$$

где

ЦОФ_{норм} - нормализованное значение *i* - того ценообразующего фактора;

ЦОФ - исходное, не нормализованное значение *i* - того ценообразующего фактора;

ЦОФ_{мин} - минимальное не нормализованное значение *i* - того ценообразующего фактора из всех значений этого фактора у объекта оценки и объектов - аналогов;

ЦОФ_{макс} - максимальное не нормализованное значение *i* - того ценообразующего фактора из всех значений этого фактора у объекта оценки и объектов - аналогов.

В результате минимаксного масштабирования факторы утрачивают размерность, а интервалы их изменения становятся равными и заключёнными в пределы от 0 до 1.

Нормализация значений ценообразующих факторов (3). Примеры

Пример для пяти аналогов и объекта оценки. Абсолютные значения

	Значения ценообразующих факторов			
	абсолютные		нормализованные безразмерные	
	Площадь, кв.м.	Этаж	Площадь	Этаж
Аналог 1	50	1	0,00	0,00
Аналог 2	63	3	0,13	0,29
Аналог 3	76	5	0,25	0,57
Аналог 4	87	6	0,36	0,71
Аналог 5	102	8	0,51	1,00
Объект оценки	70	4	0,20	0,43
Мин	50	1		
Макс	102	8		
Макс - Мин	52	7		

Нормализация значений ценообразующих факторов (4). Примеры

Пример оцифровывания качественных характеристик

Качественная характеристика состояния отделки	Ранговая характеристика	Нормализованные безразмерные характеристики
Без отделки	1	0,00
Черновая отделка	2	0,25
Требуется ремонт	3	0,50
Рабочее	4	0,75
После ремонта / Новое	5	1,00
Мин	1	
Макс	5	
Макс - Мин	4	

Аналогично оцифровываются бинарные (есть / нет) характеристики. Нормализованные значения будут либо 1, либо 0.

Определение веса цены объекта - аналога в рыночной стоимости объекта оценки

Поскольку с увеличением многомерного расстояния соответствующие метрики возрастают, что должно приводить к снижению веса соответствующего объекта - аналога в рыночной стоимости объекта оценки, для определения соответствующего веса необходимо использовать обратные значения соответствующих метрик.

Соответственно, вес j - того аналога в рыночной стоимости объекта оценки определяется следующим образом:

$$\text{Вес}_j = \frac{\frac{1}{\text{Мет. } j}}{\sum_{j=1}^k \frac{1}{\text{Мет. } j}}$$

Определение веса цены объекта - аналога в рыночной стоимости объекта оценки (2)

При желании увеличить вес «далёких» объектов аналогов, рекомендуется применять квадраты значений соответствующих метрик:

$$\text{Вес}_j = \frac{\frac{1}{(\text{Мет}_j)^2}}{\sum_{j=1}^k \frac{1}{(\text{Мет}_j)^2}}$$

Однако, целесообразность такой меры пока не ясна.

Определение рыночной стоимости

Рыночная стоимость объекта оценки определяется как сумма взвешенных цен объектов - аналогов

$$PC_{oo} = \sum_{j=0}^k \text{Вес}_j \times \text{Цена}_j$$

где k - количество объектов - аналогов

Необходимость обеспечения однородности используемых данных.

Никакие метрики не дадут достоверного результата в случае, если используемые цены «аналогов» не представляют собой набор однородных данных, включают в себя выбросы, т.е. «аналоги» не имеют единого ценообразования, т.е. фактически аналогами не являются.

Поэтому необходимо установить ограничение по предварительному обеспечению однородности цен аналогов. Например - установление двух одновременных требований:

- ▶ Максимальная по модулю относительная к среднему или иной мере центральной тенденции ошибка - 35%, что соответствует максимальному отношению максимального и минимального значений 2,07 раза (при симметричных ошибках в минус и в плюс, что маловероятно);
- ▶ Средняя по модулю относительная к среднему или иной мере центральной тенденции ошибка - 15%.

Заключение

Никакие метрики не дадут достоверного результата в случае, если

- ▶ на надлежащем уровне не выполнен анализ рынка
- ▶ на основании анализа рынка не определены ценообразующие факторы и они подставляются в метрики наугад;
- ▶ на основании рынка обоснованно не отобраны объекты - аналоги.

При невыполнении этих условий замена «справочников» на метрики не даст ничего, поскольку на тенденциозно подобранных аналогах получить заранее заданный результат, конечно, сложнее, но по - прежнему возможно.