Линейные и нелинейные критерии: нормализация и свертка критериев

1. Нормализация и свертка линейных критериев
2. Нормализация и свертка нелинейных критериев

Нормализация и свертка линейных критериев

Основные определения:

Объекты (рейтинга / ранжирования)

- **Объекты, который можно сравнить в группе**

Критерии

- **Свойства объекта рейтинга, которые могут быть использованы для дифференциации объектов в группе**
- **Критерий – признак, на основании которого производится оценка, определение или классификация чего-либо; мерило; показатель эффективности, качества**

Показатели (индикаторы)

- **Значения, связанные с объектом рейтинга - монотонные функции от критерия**

Критерии

Большее значение лучше (положительный критерий)

Меньшее значение лучше (отрицательный критерий)

Примеры:

- ***ВВП на душу населения - положительный критерий***
- ***Внешний долг, уровень безработицы - отрицательный критерий***

Основные определения:

- Нормализация – установление нормы, образца; приведение к норме, к нормальному состоянию, к единой мере, масштабу.

- Свертка – определение комплексного свойства объекта по ряду признаков.

Ранги

- Места объектов в упорядоченном списке в соответствии с показателями рейтингов

Определения

N - количество объектов, $i = 1;N$

M - число критериев, $j = 1;M$

x_{ij} - значение j-го критерия для i-го объекта

R_{ij} - частный показатель (индекс) j-го критерия i-го объекта

R_i - общий показатель (индекс) i-го объекта

r_{ij} - частный ранг i-го объекта по с j-му критерию

r_i - общий ранг i-го объекта

Типы алгоритмов нормализации

<u>1-й тип </u> **– нормализация к лучшему значению (максимальному при положительном критерии)**

$$R_{ij} = C \cdot \frac{x_{ij}}{x_j^{max}}$$

R_{ij} – нормализованное значение x_{ij},
C – коэффициент масштабирования
x_j^{max} – максимальное из значений критериев (по j-му критерию) по конкретному варианту

Типы алгоритмов нормализации

1-й тип – нормализация к лучшему значению (минимальному при отрицательном критерии)

$$\mathbf{R_{ij} = C * \frac{x_j^{min}}{x_{ij}}}$$ **или** $$R_{ij} = C \cdot \left(1 - \frac{x_{ij} - x_j^{min}}{x_j^{max}}\right)$$

R_{ij} **– нормализованное значение** x_{ij},
C **– коэффициент масштабирования**
x_j^{min} **– минимальное значение переменной по j-му критерию**
x_j^{max} **– максимальное значение переменной по j-му критерию**

недостатком алгоритма является то, что он существенно зависит от максимально возможного уровня критериев, определяемого условием задачи. Предпочтение автоматически отдается критерию с максимальной величиной.

Вместо максимального значения в знаменателе может быть использован супремум критериев – максимально возможное, а не фактическое значение. При этом алгоритм нормализации является более справедливым к критериям с меньшей величиной и не зависит от их масштаба. Однако нередко супремумом является бесконечность и его применение невозможно.

Минусы:

- диапазон:

$$\left[C\frac{x_j^{min}}{x_j^{max}};\, C \right]$$

- «назначение скрытого веса» (перераспределение весов аутсайдерам)
- в редких случаях возможны отрицательные значения для левой границы

"Назначение скрытого веса"

Для положительного критерия:

$$R_{ij}^{\text{Best}} = R_{ij}^{\text{Basic}}\left(1 - a_j\right) + a_j C$$

$$\Delta R_{ij} = R_{ij}^{\text{Best}} - R_{ij}^{\text{Basic}} = a_j\left(C - R_{ij}^{\text{Basic}}\right)$$

где:

$$a_j = \frac{x_j^{min}}{x_j^{max}}, a_j \in [0;1)$$

- R_{ij}^{Best} зависит от диапазона критерия и значения критерия i-го объекта
- R^{Best} всегда не меньше, чем R^{Basic}

2-й тип

$$R_{ij} = C\frac{x_{ij}}{x_{max} - x_{min}}$$

x_{min}- **минимальное из значений критериев по конкретному варианту**

При использовании этого алгоритма требуется специальная проверка условий инвариантности к началу координат и масштабам измерения (по крайней мере, в некоторых случаях) и наблюдается зависимость от пределов измерения критериев. При этом граничные значения (максимальное и минимальное) служат источниками погрешностей, например, от округления, вычисления и т.д., превышающими в большинстве случаев погрешности других значений диапазона измерения критериев.

3-*й тип* – с учетом характеристик приоритета критериев

$$R_{ij} = \frac{\alpha_j x_{ij}}{x_{max}}$$

α_j **– мера важности критерия**

Хотя учет приоритета из-за соображений ясности аргументации лучше всего задавать после нормализации

4-*й тип* – с учетом максимизации-минимизации критериев (базовая нормализация)

- **Нормализация (положительный критерий)**

$$R_{ij} = C \cdot \frac{x_{ij} - x_j^{min}}{x_j^{max} - x_j^{min}}$$

R_{ij} – норм , частный показатель

C – коэффициент масштабирования

x_j^{min} – минимальное значение переменной по j-му критерию

x_j^{max} – максимальное значение переменной по j-му критерию

4-*й тип* - **Нормализация (отрицательный критерий)**

$$R_{ij} = C \cdot \left(1 - \frac{x_{ij} - x_j^{min}}{x_j^{max} - x_j^{min}}\right)$$

Замечание

Возможно наложение штрафа на значения вне установленного x_j^{min} (нормативного значения)

В отличие от первых 3-х алгоритмов, четвертый вносит более здравый смысл в процедуру нормализации и подтверждает ее целесообразность. Важно не просто перейти к условным единицам измерения критериев, но и ориентировать желаемое направление их изменения, поскольку, как правило, одни критерии необходимо свести к минимуму, а другие – к максимуму.

По этой формуле можно нормализовать не только критерии, но и непосредственно физические величины (температура, содержание веществ и др.)

В результате этой операции все критерии приводятся к единой шкале измерения в пределах [0;1] и к единой безразмерной единице измерения, что обеспечивает их беспрепятственное и объективное сопоставление.

5-*й тип* –нормализация к среднему значению

- **Нормализация (положительный критерий)**

$$R_{ij} = C \cdot \frac{x_{ij}}{x_j^{avg}}$$

R_{ij} – **нормализованное значение** x_{ij}, **частный показатель**

C – **коэффициент масштабирования**

x_j^{avg} – **среднее значение переменной по j-му критерию**

5-*й тип* –нормализация к среднему значению

- **Нормализация (отрицательный критерий)**

$$R_{ij} = C \cdot \frac{x_j^{min} + x_j^{max} - x_{ij}}{x_j^{avg}}$$

R_{ij} **– нормализованное значение *x_{ij}*, частный показатель**

C **– коэффициент масштабирования**

x_j^{avg} – **среднее значение переменной по j-му критерию**

x_j^{min} **– минимальное значение переменной по *j*-му критерию**

x_j^{max} **– максимальное значение переменной по *j*-му критерию**

Нормализация к среднему

Плюсы:

Легко определить "средний" объект (R_{ij}≈1)

Минусы:

Диапазон

$$\left[C\frac{x_j^{min}}{x_j^{avg}}; C\frac{x_j^{max}}{x_j^{avg}} \right]$$

может быть непредсказуемо широк / узок

-"Назначение скрытых весов" при агрегировании критериев

- Возможны отрицательные значения для левой границы

- Может отсутствовать "средний" объект

- Асимметричные левая и правая границы

Агрегирование / свертка

В основе многокритериальной оптимизации часто лежит принцип аддитивности критериев, то есть использование интегрального показателя, наиболее известным из которых является линейная свертка

$$R_i = \sum_{j=1}^{M} w_j \cdot R_{ij}$$

R_i - общий (*обобщенный*) показатель (индекс) i-го объекта

R_{ij} - частный показатель (индекс) j-го критерия i-го объекта

w_j - вес j-го критерия, сумма всех весов объекта = 1

$$r_i = \mathbf{rank}\ R_i$$

r_i - общий ранг i-го объекта

Свертывание критериев требует приведения оценок с различными оценочными шкалами и единицами измерения к сопоставимому виду.

Для этого количественные оценки во многих случаях нормализуются – приводятся к стандартной шкале, как правило, от 0 до 1, крайние значения которой соответствуют худшему (0) и лучшему (1) из рассматриваемых значений критерия либо могут задаваться экспертным путем.

Масштабирование

$$R_i^* = C \cdot \frac{R_i - R^{min}}{R^{max} - R^{min}}$$

R_i^*- масштабированное значение общего показателя (индекса) *i*-го объекта

R^{min} - минимальное значение среди общих показателей (индексов)

R^{max} - максимальное значение среди общих показателей (индексов)

Свойства:

- Используется только для «удобства»
- Не меняет порядок рангов
- Меняет относительные различия в сравнении объектов

Недостатки линейных показателей

1. **Взаимозаменяемость частных индикаторов (без учета весов)**
2. **Нелинейная шкала (почти все показатели нелинейны, сводим к линейным)**
3. **Изменение диапазона индикаторов (возможны изменения с течением времени, различные выборки) – необходимость установления незыблемых минимальных и максимальных значений**
4. **Парадокс сильного лидера (стратегия следования за лидером)**
5. **Противоречивые результаты в подгруппах**

Пример странового рейтинга

1. **ВВП по ППС в 2016 г., млрд.долл**
2. **Расходы на ИиР, млрд. долл**
3. **Доля расходов на ИиР от ВВП в 2016 г., %**
4. **Внешний долг, % от ВВП**
5. **индекс Джинни**

Страна	ВВП по ППС в 2016 г., млрд.долл	Расходы на ИиР, млрд. долл	Доля расходов на ИиР от ВВП в 2016 г., %	Внешний долг, % от ВВП	индекс Джинни
США	18559	514	2,77	105	41,1
Китай	20015	396,3	1,98	47	42,2
Япония	4913	166,6	3,39	238	32,1
Германия	3741	109,3	2,92	64	30,1
Индия	8409	71,5	0,85	71	35,2
Франция	2657	60,1	2,26	97	33,1
Россия	3396	50,9	1,5	16	41,6
Великобритания	2558	45,5	1,78	88	32,6
Бразилия	3072	37,2	1,21	83	51,5
Канада	1646	29,5	1,79	90	33,7
Австралия	1167	27,9	2,39	41	34,9

Страна	ВВП по ППС в 2016 г., млрд.долл	Доля расходов на ИиР от ВВП в 2016 г., %	Внешний долг, % от ВВП
США	**18559**	**2,77**	**105**
Китай	**20015**	**1,98**	**47**
Япония	**4913**	**3,39**	**238**
Германия	**3741**	**2,92**	**64**
Индия	**8409**	**0,85**	**71**
Франция	**2657**	**2,26**	**97**
Россия	**3396**	**1,5**	**16**
Великобритания	**2558**	**1,78**	**88**
Бразилия	**3072**	**1,21**	**83**
Канада	**1646**	**1,79**	**90**
Австралия	**1167**	**2,39**	**41**
min	***1167***	***0,85***	***16***
max	***20015***	***3,39***	***238***

Страна	ВВП по ППС в 2016 г., млрд.долл	Доля расходов на ИиР от ВВП в 2016 г., %	Внешний долг, % от ВВП
США	92,28	75,59	59,91
Китай	100,00	44,49	86,04
Япония	19,87	100,00	0,00
Германия	13,66	81,50	78,38
Индия	38,42	0,00	75,23
Франция	7,91	55,51	63,51
Россия	11,83	25,59	100,00
Великобритания	7,38	36,61	67,57
Бразилия	10,11	14,17	69,82
Канада	2,54	37,01	66,67
Австралия	0,00	60,63	88,74
min	*0*	*0*	*0*
max	*100*	*100*	*100*

Агрегирование / свертка

Страна	Index	Rank
США	**75,93**	**2**
Китай	**76,84**	**1**
Япония	**39,96**	**7**
Германия	**57,84**	**3**
Индия	**37,88**	**8**
Франция	**42,31**	**6**
Россия	**45,81**	**5**
Великобритания	**37,19**	**9**
Бразилия	**31,37**	**11**
Канада	**35,41**	**10**
Австралия	**49,79**	**4**
min	***31,37***	
max	***76,84***	

Масштабирование

Страна	Index	Rank
США	**97,99**	**2**
Китай	**100,00**	**1**
Япония	**18,89**	**7**
Германия	**58,22**	**3**
Индия	**14,33**	**8**
Франция	**24,07**	**6**
Россия	**31,75**	**5**
Великобритания	**12,80**	**9**
Бразилия	**0,00**	**11**
Канада	**8,88**	**10**
Австралия	**40,51**	**4**
min	***0,00***	
max	***100,00***	

Страна	ВВП по ППС в 2016 г., млрд.долл	Доля расходов на ИиР от ВВП в 2016 г., %	Внешний долг, % от ВВП	Index	Rank
США	92,73	81,71	62,61	79,01	2
Китай	100,00	58,41	86,97	81,79	1
Япония	24,55	100,00	6,72	43,76	9
Германия	18,69	86,14	79,83	61,55	3
Индия	42,01	25,07	76,89	47,99	7
Франция	13,28	66,67	65,97	48,64	6
Россия	16,97	44,25	100,00	53,74	5
Великобритания	12,78	52,51	69,75	45,01	8
Бразилия	15,35	35,69	71,85	40,96	11
Канада	8,22	52,80	68,91	43,31	10
Австралия	5,83	70,50	89,50	55,28	4
min	*5,83*	*25,07*	*6,72*	*40,96*	
max	*100*	*100*	*100*	*81,79*	

Страна	ВВП по ППС в 2016 г., млрд.долл	Доля расходов на ИиР от ВВП в 2016 г., %	Внешний долг, % от ВВП
США	**18559**	**2,77**	**105**
Китай	**20015**	**1,98**	**47**
Япония	**4913**	**3,39**	**238**
Германия	**3741**	**2,92**	**64**
Индия	**8409**	**0,85**	**71**
Франция	**2657**	**2,26**	**97**
Россия	**3396**	**1,5**	**16**
Великобритания	**2558**	**1,78**	**88**
Бразилия	**3072**	**1,21**	**83**
Канада	**1646**	**1,79**	**90**
Австралия	**1167**	**2,39**	**41**
min	***1167***	***0,85***	***16***
max	***20015***	***3,39***	***238***
avg	***6375,73***	***2,08***	***85,45***

Нормализация к среднему

Страна	ВВП по ППС в 2016 г., млрд.долл	Доля расходов на ИиР от ВВП в 2016 г., %	Внешний долг, % от ВВП	Index	Rank
США	**291,09**	**133,17**	**174,36**	**199,62**	***2***
Китай	**313,92**	**95,19**	**242,23**	**217,17**	***1***
Япония	**77,06**	**162,98**	**18,72**	**86,35**	***11***
Германия	**58,68**	**140,38**	**222,34**	**140,55**	***3***
Индия	**131,89**	**40,87**	**214,15**	**128,99**	***5***
Франция	**41,67**	**108,65**	**183,72**	**111,41**	***7***
Россия	**53,26**	**72,12**	**278,51**	**134,67**	***4***
Великобритания	**40,12**	**85,58**	**194,26**	**106,70**	***8***
Бразилия	**48,18**	**58,17**	**200,11**	**102,19**	***9***
Канада	**25,82**	**86,06**	**191,91**	**101,31**	***10***
Австралия	**18,30**	**114,90**	**249,26**	**127,55**	***6***
min	***18,30***	***40,87***	***18,72***	***86,35***	
max	***313,92***	***162,98***	***278,51***	***217,17***	

2. Нормализация и свертка нелинейных критериев

Диффернцирующая сила (дискриминативность) индикаторов – способность индикатора генерировать различая между объектами (тест)

Например, показатель ВВП и ВВП на душу населения

Способы преобразования критериев:

- Логарифмическое преобразование x ’ = log (x)
- Экспоненциальное x’ = exp (x) (описание развития событий, например, смена технологий)
- Персентиль (или перцентиль или процентиль) x’ = percentile (x) — методика измерения в статистике, которая показывает процент значений измеряемой метрики, который находится ниже значения персентиля. Например, если говорить о времени ответа системы, 99й персентиль на отметке 100 миллисекунд говорит о том, что 99% измеряемых запросов выполнились за 100 миллисекунд и менее.
- И др.

Процедура формального теста однородности:

1. **Рассчитываются данные квантилей (0%, 10%, … 100%)**
2. **Определяются разности между соседними квантилями (они должны быть распределены равномерно)**
3. **Протестировать распределение разностей (например, t-тест Стьюдента, χ2-Пирсона…)**

Нелинейное агрегирование (свертка):

1. Среднее арифметическое взвешенное
2. Среднее геометрическое (также часто используется в ранжировании экономических объектов)
3. Среднее экспоненциальное
4. Среднее логарифмическое