

Технология коррекции и обработки парных сравнений

Пшеничный Д.О.
Двоенко С.Д.
Тульский государственный
университет

ММРО-2019 г.



Представление экспериментальных данных

- В современном ИАД данные могут быть представлены матрицами парных сравнений.
- Парные сравнения являют собой либо «похожесть» элементов друг на друга, либо их «различие».
- В общем случае пространство признаков исследуемых элементов не задаётся явно.
- Элементы анализируемого множества могут быть помещены в некоторое метрическое пространство. «Различие» между ними является расстоянием между соответствующими векторами.
- В этом случае «похожесть», как неотрицательная величина, обратная по смыслу «различию» между такими элементами, будет являться функцией близости, т.е. скалярными произведениями между векторами в положительном квадранте этого пространства.

Представление экспериментальных данных

- Известно, что для корректного погружения исследуемых объектов в метрическое пространство матрица скалярных произведений между соответствующими векторами не должна содержать отрицательных собственных чисел.
- В общем случае матрица скалярных произведений, получаемая из экспериментов, может не обладать таким свойством.
- Результаты, получаемые методами ИАД на матрицах парных сравнений, в этом случае, строго говоря, не являются математически корректными.
- Мы предлагаем технологию для восстановления метрической корректности произвольных матриц парных сравнений.

Математическое обоснование

- Критерий Сильвестра: квадратичная форма матрицы является положительно полуопределённой тогда и только тогда, когда значения всех главных миноров матрицы являются неотрицательными числами.
- Собственные значения матрицы положительно полуопределённой квадратичной формы являются неотрицательными числами.
- Матрица неопределённой формы может содержать собственные значения разных знаков.
- Следствие из закона инерции Сильвестра: количество отрицательных собственных значений в квадратной матрице совпадает с количеством знакоперемен в последовательности её главных миноров.

Технология коррекции

- Технология коррекции произвольной матрицы парных сравнений состоит из следующих шагов.
 1. Приведение произвольной матрицы парных сравнений к матрице нормированных скалярных произведений.
 2. Локализация отрицательных собственных чисел.
 3. **Коррекция матрицы нормированных скалярных произведений с локализованными собственными числами.**
 4. Восстановление исходного вида матрицы.

Коррекция матрицы нормированных скалярных произведений

- Пусть имеется матрица $S(n, n)$ нормированных скалярных произведений, соответствующих «похожести» элементов друг на друга.
- В этом случае соответствующие векторы в признаковом пространстве являются нормированными, и их концы размещены на гиперсфере единичного радиуса.
- Все элементы матрицы $S(n, n)$: $s_{ij} = s_{ji}$, $-1 < s_{ij} < 1$, $s_{ii} = 1$.
- Все элементы матрицы – это скалярные произведения между соответствующими векторами, т.е. косинусы углов между ними.

Коррекция матрицы нормированных скалярных произведений

- В квадратной матрице $S(n, n)$ существует последовательность S_1, S_2, \dots, S_n главных миноров.
- $S_1 = (s_{11}) = 1$
- $S_2 = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} = \begin{pmatrix} 1 & s_{12} \\ s_{12} & 1 \end{pmatrix}$
- $S_3 = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix} = \begin{pmatrix} 1 & s_{12} & s_{13} \\ s_{12} & 1 & s_{23} \\ s_{13} & s_{23} & 1 \end{pmatrix}$
- $S_n = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \dots & \dots & \ddots & \dots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{pmatrix} = \begin{pmatrix} 1 & s_{12} & \dots & s_{1n} \\ s_{12} & 1 & \dots & s_{2n} \\ \dots & \dots & \ddots & \dots \\ s_{1n} & s_{2n} & \dots & 1 \end{pmatrix}$

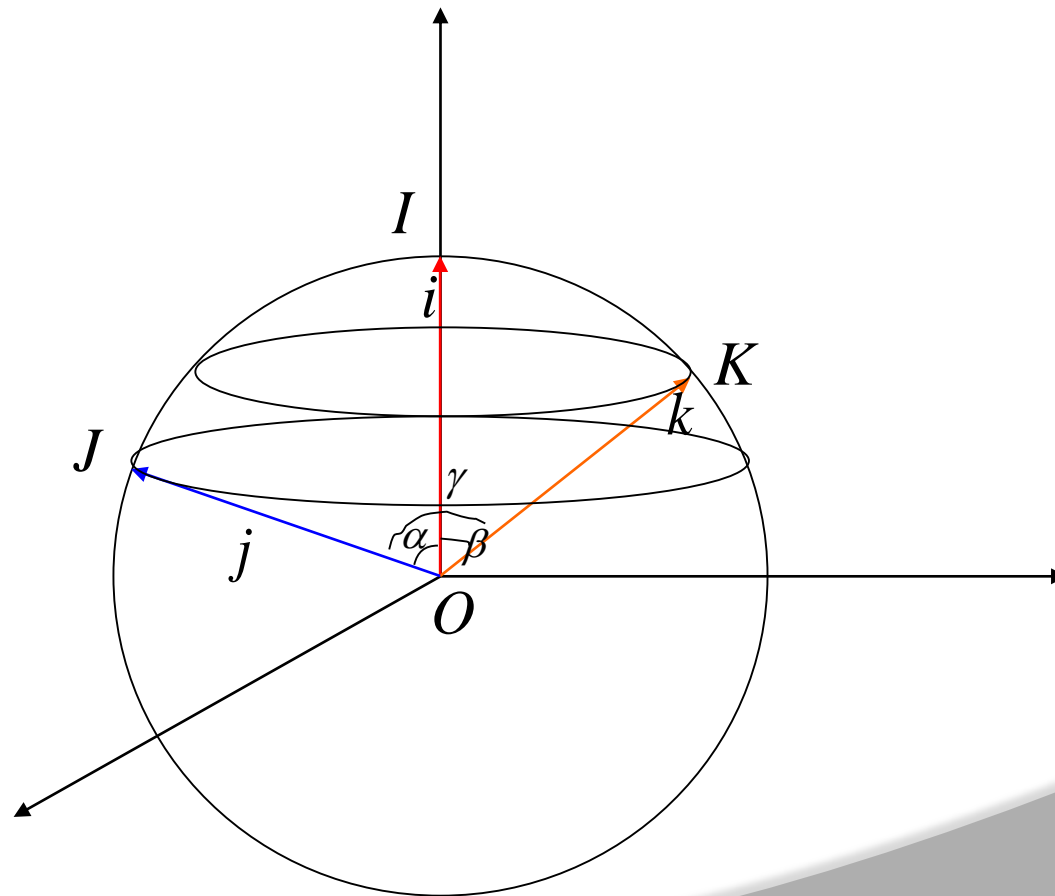
Коррекция матрицы нормированных скалярных произведений

- Рассматривается последовательность главных миноров матрицы нормированных скалярных произведений, вычисляются их значения.
- Предлагается скорректировать последовательность главных миноров так, чтобы все значения миноров оказались положительными.
- Для этого предлагается последовательно скорректировать элементы в последних строках и столбцах каждого встреченного отрицательного минора. Значения полученных миноров будут положительными. Значения предыдущих миноров не изменятся.
- **Теорема.** При возникновении метрического нарушения множество элементов не может определить множество гиперсфер в специальном координатном пространстве.

Доказательство

- Пусть x_i – это i -я координата в некотором n -мерном пространстве.
- Все возможные положения первого вектора – единичная n -мерная гиперсфера, т.е. $x_1^2 + \dots + x_n^2 = 1$
- Все возможные положения второго вектора – та же самая гиперсфера. При этом он имеет близость s_{12} с первым вектором.
- По теореме косинусов расстояние между концами первого и второго векторов $d_{12} = \sqrt{2 - 2s_{12}}$, т.е. все возможные положения второго вектора относительно первого: $x_1^2 + \dots + (x_n - 1)^2 = 2 - 2s_{12}$.

Доказательство



Доказательство

- Из этих двух ограничений получаются все возможные положения второго вектора: $x_1^2 + \dots + x_{n-1}^2 = 1 - s_{12}^2$.
- Т.к. значение первого главного минора $S_1 = 1$, а второго $-S_2 = 1 - s_{12}^2$, полученное выражение имеет следующий вид: $x_1^2 + \dots + x_{n-1}^2 = S_2/S_1$.
- Все возможные положения второго вектора определяют $(n - 1)$ -мерную гиперсферу с центром в начале координат и радиусом $\sqrt{S_2/S_1}$.

Доказательство

- Корректные положения третьего вектора оказываются следующими.
- Относительно первого вектора: $x_1^2 + \dots + x_{n-1}^2 = 1 - s_{13}^2$.
- Относительно второго вектора: $x_1^2 + \dots + \left(x_{n-1} - \sqrt{1 - s_{12}^2}\right)^2 + (x_n - s_{12})^2 = 2 - 2s_{23}$.
- Объединение этих двух выражений даёт: $x_1^2 + \dots + x_{n-2}^2 = (1 + 2s_{12}s_{13}s_{23} - s_{12}^2 - s_{13}^2 - s_{23}^2)/(1 - s_{12}^2)$.
- Замена числителя и знаменателя в правой части даёт: $x_1^2 + \dots + x_{n-2}^2 = S_3/S_2$.

Доказательство

- Все возможные положения третьего вектора определяют $(n - 2)$ -мерную гиперсферу с центром в начале координат и радиусом $\sqrt{S_3/S_2}$.
- Таким образом, для k -го вектора корректные позиции – это $(n - k + 1)$ -мерная гиперсфера радиуса $\sqrt{S_k/S_{k-1}}$.
- В специальное n -мерное пространство поочерёдно помещаются векторы для соответствующих элементов.

Доказательство

- Пусть первый вектор расположен вдоль n -й координатной оси. Его координаты $(0, \dots, 0, 1)$.
- В общем случае k -й вектор определяется координатами

$$\begin{aligned}c_1 &= \dots = c_{n-k} = 0; \\c_{n-k+1} &= \sqrt{S_k/S_{k-1}}; \\c_{n-k+t} &= \left(\left((S_k)_{k-t+2}^{k-t+1} \right) \dots \right)_k^{k-1} / \sqrt{S_k/S_{k-1}}, \quad t = 2, \dots, k-1; \\c_n &= S_{1k},\end{aligned}$$

где запись вида $(S_k)_j^i$ обозначает k -й минор матрицы S , из которого вычеркнуты i -я строка и j -й столбец; координаты c_i с индексами вне диапазона $1 \leq i \leq n$ не существуют, а $S_0 = 1$.

- Если очередной главный минор оказывается отрицательным, то возникает метрическое нарушение, т.к. радиус очередной гиперсферы и координата c_{n-k+t} имеют комплексное значение ■.

Доказательство вкратце

- Векторы помещаются в специальное координатное пространство.
- Координаты вектора соответствуют некоторому соотношению между минорами матрицы нормированных скалярных произведений.
- Если некоторый главный минор оказывается отрицательным, то возникает метрическое нарушение, т.к. соответствующий вектор не может быть помещён в такое пространство.

Следствия из теоремы

- Из теоремы имеются следствия:
- Последовательность главных миноров матрицы нормированных скалярных произведений является невозрастающей по модулю, начиная с единицы.
- Если значения всех главных миноров положительны, то метрические нарушения отсутствуют.
- Значение минора S_k совпадает со значением минора S_{k-1} тогда и только тогда, когда парные сравнения k -го элемента со всеми предыдущими $k - 1$ элементами равны 0.

Коррекция матрицы нормированных скалярных произведений

- Просмотр главных миноров матрицы нормированных скалярных произведений $S(n, n)$, пока не встретится очередной минор со значением $S_k < 0, k = 1, \dots, n$.
- Корректировка элементов последней строки и столбца этого минора, так чтобы значение скорректированного минора оказалось положительным.
- Продолжение просмотра главных миноров и корректировка всех отрицательных миноров.

Коррекция матрицы нормированных скалярных произведений

- Оптимальная коррекция минора: корректировка некоторых элементов s_{kp} и s_{pk} последних строки и столбца минора S_k , где p – это элементы множества $P \subseteq I = \{1, \dots, k-1\}$.
- Требуется решить задачу условной оптимизации:

$$\begin{cases} \sum_{p \in P} (s_{pk} - x_p)^2 \rightarrow \min \\ \sum_{p \in P} \sum_{q \in P} x_p x_q r_{pq} = C \end{cases}$$

где $x_p = s_{kp} = s_{pk}$; r_{pq} – элемент матрицы

$R(k-1, k-1) = S(k-1, k-1)^{-1}$; $C = 1 - \tau$;

$\tau = S'_k / S_{k-1}$ – доля от значения детерминанта предыдущего главного минора; S'_k – новое значение скорректированного минора.

Коррекция матрицы нормированных скалярных произведений

- По методу множителей Лагранжа задача оптимизации сводится к решению системы уравнений

$$\left\{ \begin{array}{l} \lambda \sum_{i \in P} x_i r_{ip} + \sum_{i \in I \setminus P} s_{ki} r_{ip} = s_{kp} - x_p, p \in P \\ \sum_{i \in P} \sum_{j \in P} x_i x_j r_{ij} + \sum_{i \in P} \sum_{j \in I \setminus P} x_i s_{jk} r_{ij} + \\ \sum_{i \in I \setminus P} \sum_{j \in P} s_{ki} x_j r_{ij} + \sum_{i \in I \setminus P} \sum_{j \in I \setminus P} s_{ki} s_{jk} r_{ij} = C. \end{array} \right.$$

Коррекция матрицы нормированных скалярных произведений

- Число уравнений в этой системе зависит от мощности множества P .
- Система может быть решена численным методом, например, итерационным методом Ньютона.

Технология коррекции

1. Приведение произвольной матрицы парных сравнений к матрице нормированных скалярных произведений.
2. **Локализация отрицательных собственных чисел.**
3. Коррекция матрицы нормированных скалярных произведений с локализованными собственными числами.
4. Восстановление исходного вида матрицы.

Локализация отрицательных собственных значений

- В общем случае каждая коррекция влечёт «шлейф» дополнительных коррекций.
- Для наименьшего количества коррекций элементов множества их следует расположить в таком порядке, чтобы знакоперемены в последовательности главных миноров оказались как можно ближе к концу.
- Одновременная перестановка двух строк и столбцов квадратной матрицы не изменяет её собственных значений.

Локализация отрицательных собственных значений

- Пусть в матрице $S(n, n)$ содержатся v отрицательных собственных значений.
- При идеальной перестановке элементов множества первый отрицательный главный минор будет S_{n-v+1} .
- Знаки последующих $v - 1$ миноров будут чередоваться.
- Метрику нарушают не более v объектов.

Локализация отрицательных собственных значений

- Элементам множества соответствует матрица нормированных скалярных произведений $S(n, n)$.
- Требуется найти такой элемент, при удалении которого детерминант полученной матрицы $S(n - 1, n - 1)$ сменит знак и будет максимальным по модулю.
- Этот элемент займёт n -е место в ранжировке.
- Если удаление ни одного из элементов не приводит к смене знака полученного минора, n -е место займёт элемент, при удалении которого детерминант окажется максимальным по модулю.
- Повторение процедуры поиска элемента множества для получившейся матрицы размером $(n - 1, n - 1)$.

Локализация отрицательных собственных значений

- В результате нарушающие метрику элементы множества окажутся в конце последовательности.
- Количество дополнительных нарушений и количество коррекций уменьшается.

Метрические коррекции и обусловленность

- Число обусловленности $Cond(S(n, n))$ матрицы $S(n, n)$ определяет степень её вырожденности.
- Хорошо обусловленная матрица имеет малое число обусловленности.
- $Cond(S(n, n)) = \|S(n, n)\| \cdot \|S(n, n)^{-1}\|$.
- $\|S(n, n)\|$ – норма матрицы, может быть определена как, например, $\|S(n, n)\| = \max|\lambda|$.
- Тогда $Cond(S(n, n)) = \lambda_1/\lambda_n, \lambda_{max} = \lambda_1 > \dots > \lambda_n = \lambda_{min} > 0$.
- Все собственные числа считаются положительными, т.к. речь идёт о скорректированной положительно определённой матрице.

Метрические коррекции и обусловленность

- Собственные значения матрицы не связаны напрямую со значением её детерминанта.
- Пусть $S_k < 0, k \leq n$ – значение первого отрицательного минора $S(k, k)$ в последовательности главных миноров.
- Наилучшая с точки зрения минимального отклонения скорректированных элементов от исходных коррекция минора $S(k, k)$ происходит при $S'_k = 0$.
- В этом случае $Cond(S'(k, k)) = \infty$.
- Наилучшая с точки зрения максимального числа обусловленности скорректированной матрицы коррекция происходит при $S'_k = S_{k-1}$.
- В этом случае $Cond(S'(k, k)) = Cond(S'(k-1, k-1))$. Однако все скорректированные элементы последних строки и столбца становятся нулевыми.

Метрические коррекции и обусловленность

- Значение $S'_k = C$ скорректированного минора – параметр оптимизации.
- Требуется совместить два противоположных условия: минимум отклонения элементов и минимальное число обусловленности.

Метрические коррекции и обусловленность

- Пусть нормированные близости – это коэффициенты корреляции в статистическом смысле.
- Нулевая гипотеза об отсутствии корреляции между соответствующими вариационными рядами проверяется по t -распределению Стьюдента $T = x_p \sqrt{(m-2)/(1-x_p^2)}$, $p = 1, \dots, k$, $m-2$ – число степеней свободы (m – размер выборки).
- Нулевая гипотеза проверяется для модифицированных элементов последней строки текущего минора $S(k, k)$ по условию $T > t(\alpha, m-2)$, где $t(\alpha, m-2)$ – критическая точка распределения Стьюдента при уровне значимости α , например, $\alpha = 0.01$.
- Выборка и её размер m заранее неизвестны. Значимая близость при статистически достаточном размере выборки $m \geq 122$ равна 0.208; при среднем размере выборки $m \geq 62$ – 0.2948; при недостаточном размере выборки $m \geq 32$ – 0.4097.

Метрические коррекции и обусловленность

- Значимый уровень парной близости $x(\alpha, m) = \sqrt{t^2/(t^2 - m + 2)}$.
- Для каждого $0 < \tau \leq 1$, где $\tau = S'_k/S_{k-1}$ – доля от значения детерминанта предыдущего главного минора, определяется количество n_1 значимых парных близостей $x_p > x(\alpha, m)$, $p = 1, \dots, k - 1$.
- Оптимальное число обусловленности соответствует пороговому уровню коррекции τ , начиная с которого число n_1 значимых значений парных близостей $x_p > x(\alpha, m)$, $p = 1, \dots, k - 1$ резко падает.

Технология коррекции

1. **Приведение произвольной матрицы парных сравнений к матрице нормированных скалярных произведений.**
2. Локализация отрицательных собственных чисел.
3. Коррекция матрицы нормированных скалярных произведений с локализованными собственными числами.
4. **Восстановление исходного вида матрицы.**

Коррекция матриц близостей и расстояний

- В этом случае специальное координатное пространство не может быть определено, как в случае с нормированной матрицей скалярных произведений.
- Ненормированная матрица $\tilde{S}(n, n)$ перед коррекцией нормируется преобразованием $s_{ij} = \tilde{s}_{ij} / \sqrt{\tilde{s}_{ii}\tilde{s}_{jj}}$.
- Нормированная матрица корректируется.
- Скорректированная матрица денормируется обратно преобразованием $\tilde{s}'_{ij} = s'_{ij} \sqrt{\tilde{s}_{ii}\tilde{s}_{jj}}$.

Коррекция матриц близостей и расстояний

- Пусть $D(n, n)$ – матрица расстояний между элементами множества.
- Преобразование в матрицу нормированных скалярных произведений требует определения начала координат.
- Расстояние от каждого объекта до начала координат $d_{0i}^2 = \frac{1}{n} \sum_{p=1}^n d_{ip}^2 - \Delta$, $\Delta = \sigma^2$, где σ^2 – дисперсия элементов множества.
- В этом случае начало координат совпадёт с центром тяжести элементов множества.
- При $\Delta = 0$, т.е. $d_{0i}^2 = \frac{1}{n} \sum_{p=1}^n d_{ip}^2$ начало координат выносится за выпуклую оболочку множества.
- В этом случае нормированные скалярные произведения будут являть собой близости между элементами.

Коррекция матриц близостей и расстояний

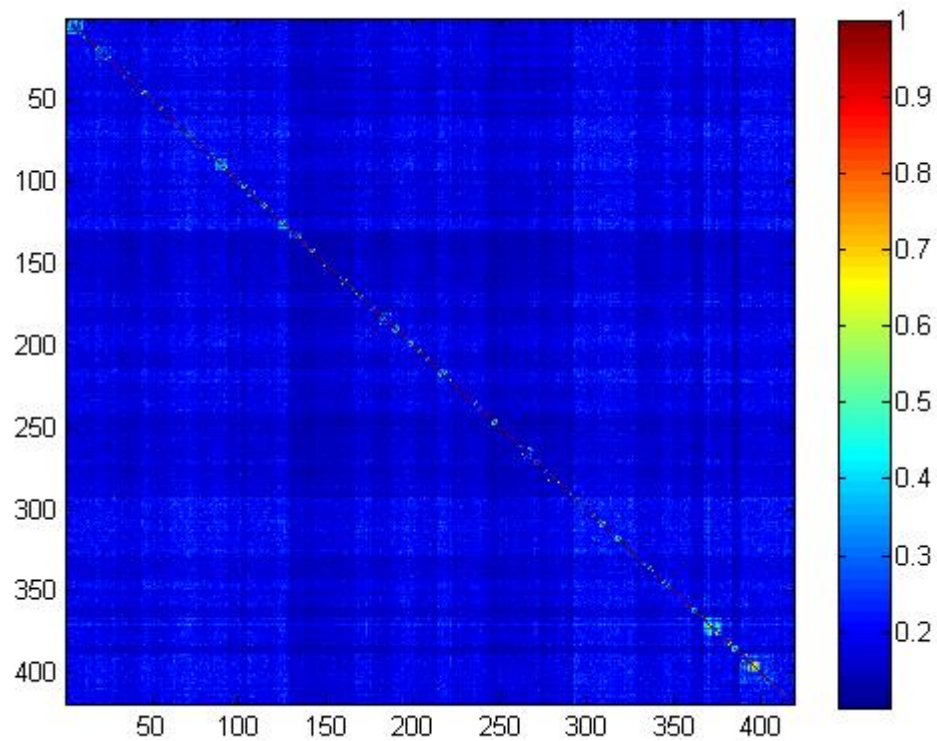
- Относительно полученного начала координат вычисляется матрица нормированных скалярных произведений.
- Она корректируется.
- Скорректированная матрица преобразуется обратно в матрицу расстояний $D'(n, n)$, $d'_{ij} = s'_{ii} + s'_{jj} - 2s'_{ij}$.

Эксперимент

- Нормированная матрица скалярных произведений $S(420,420)$ между 420 белковыми последовательностями.
- В матрице 413 положительных, 5 отрицательных и 2 нулевых собственных числа.
- В матрице две пары одинаковых строк и столбцов, которые приводят к появлению нулевых собственных чисел. Удаляя их, получаем матрицу $S(418,418)$.

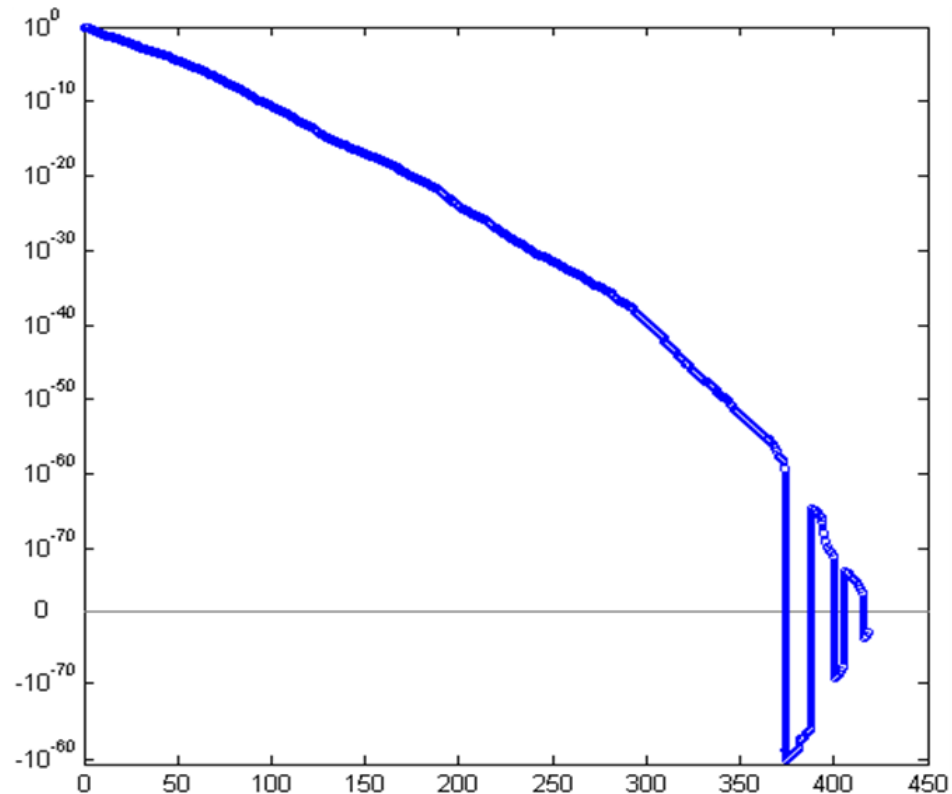
Эксперимент

- Матрица может быть показана в виде такой тепловой карты:



Эксперимент

- Последовательность значений детерминантов главных миноров имеет следующий вид:

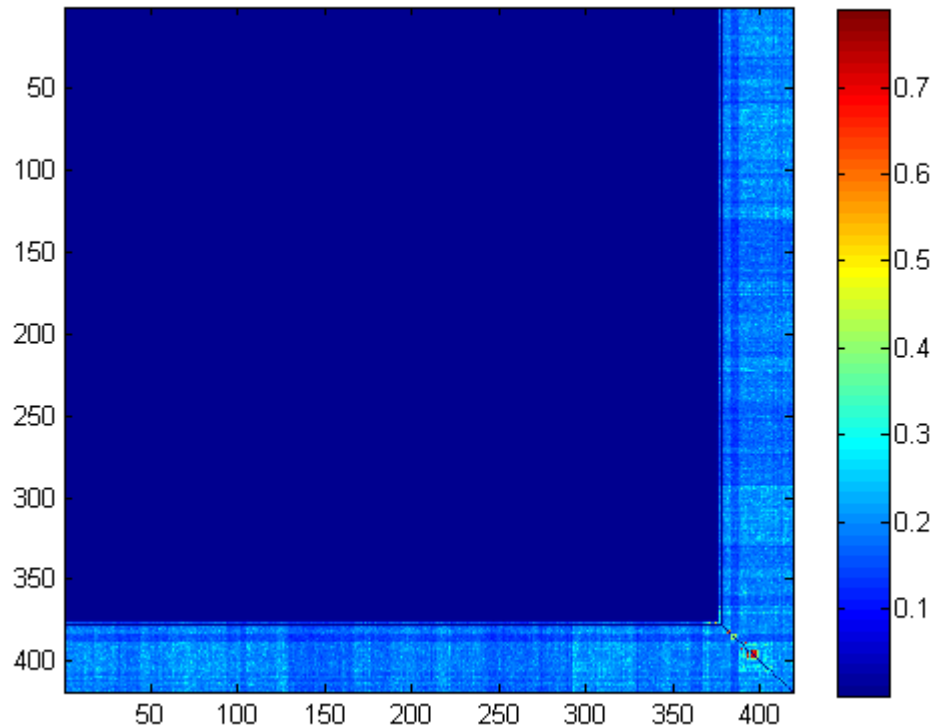


Эксперимент

- В исходной последовательности элементов множества первый отрицательный минор встречался на 375-й позиции.
- При коррекции матрицы с 375-го минора изменение миноров влечёт «шлейф» дополнительных коррекций.
- При заданном уровне коррекции $\tau = 0.3$ потребовалось скорректировать 43 минора.

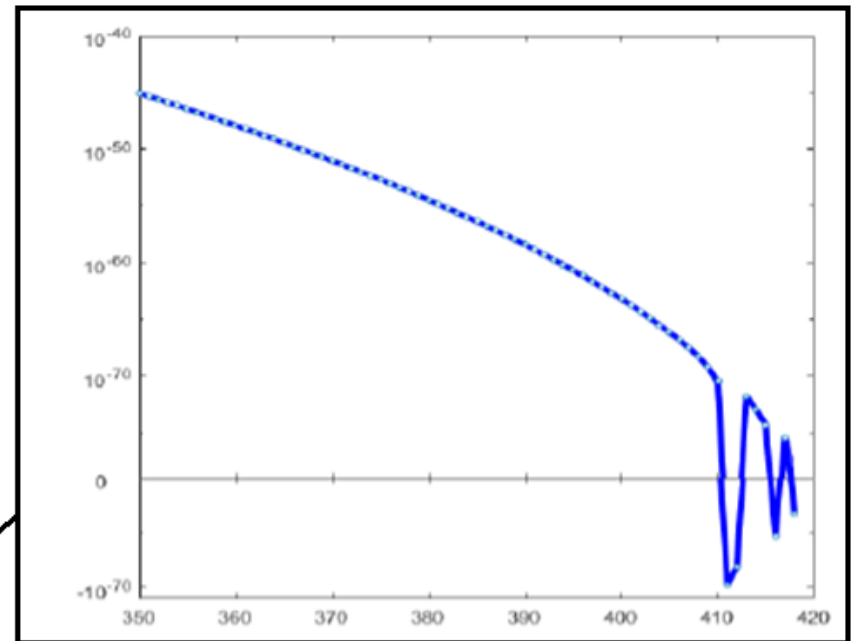
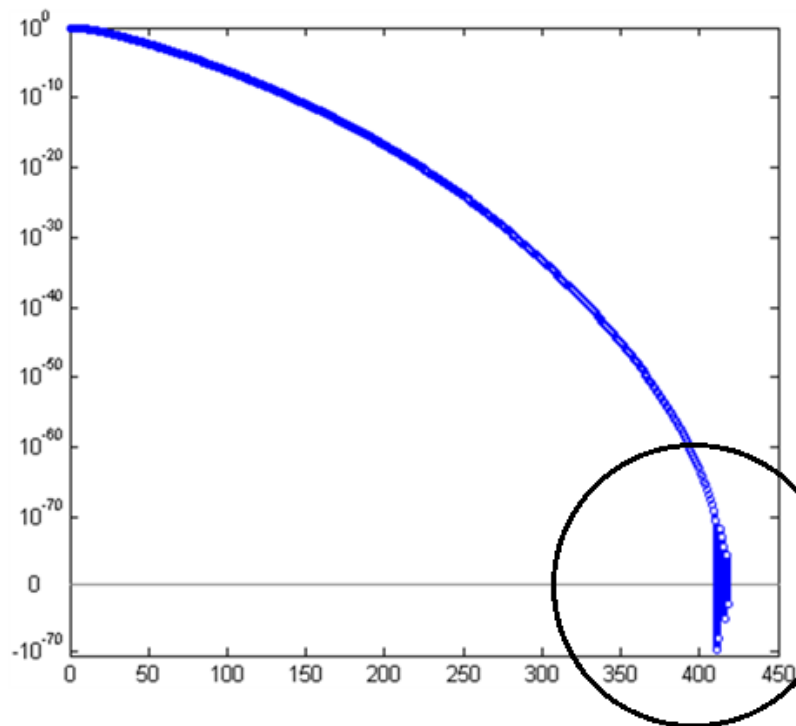
Эксперимент

- Разница между скорректированной матрицей и исходной может быть показана такой тепловой картой:



Эксперимент

- В матрице близостей, соответствующей оптимальной перестановке, последовательность значений детерминантов главных миноров имеет следующий вид:

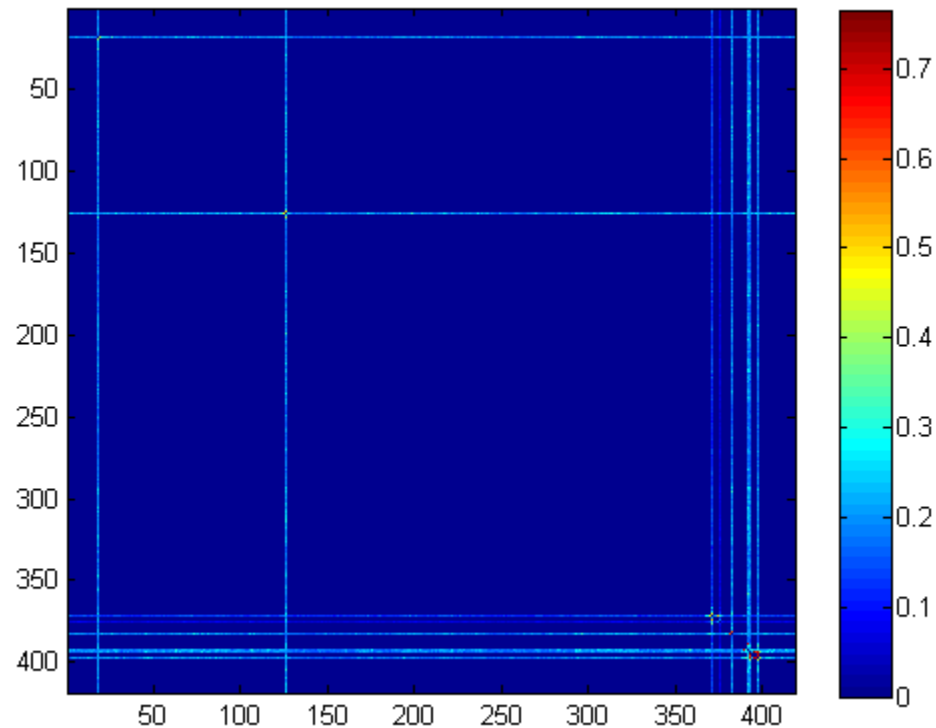


Эксперимент

- Первый отрицательный главный минор – на 411-й позиции.
- Интервал знакоперемен немного превышает число отрицательных собственных чисел (5) и состоит из 8 миноров.
- При этой последовательности элементов были скорректированы 7 миноров.

Эксперимент

- В этом случае разница между скорректированной матрицей и исходной может быть показана такой тепловой картой:



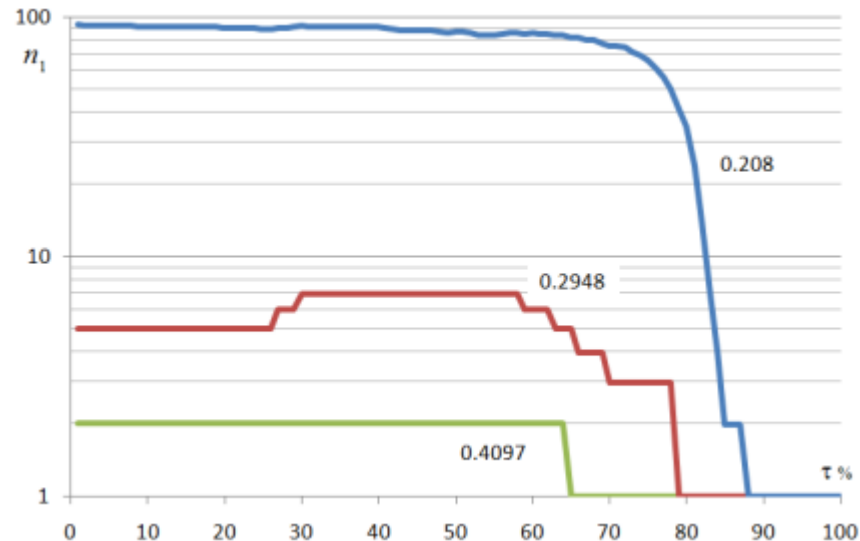
Эксперимент

Размер минора	$\tau = 0.01$	$\tau = 0.3$		$\tau = 0.6$				
	<i>Cond</i>	<i>Cond</i>	отно- шение <i>Cond</i>	<i>Cond</i>	<i>D</i>	λ_{\min}	λ_{\max}	отно- шение <i>Cond</i>
411	383759.99	22110.24	17.36	20972.27	0.201	0.0036	75.697	18.30
412	83425.31	25926.10	3.22	21025.49	0.372	0.0036	75.889	3.97
413	279173.09	26320.00	10.61	21086.91	0.234	0.0036	76.100	13.24
414	113028.51	27741.61	4.07	21144.72	0.399	0.0036	76.308	5.35
416	49729.21	44612.87	1.11	30039.22	0.489	0.0026	76.767	1.66
417	295262.92	44746.38	6.60	30123.56	0.144	0.0026	76.982	9.80
418	79583.51	44914.97	1.77	30120.85	0.297	0.0026	77.204	2.63

- Результаты коррекции миноров в оптимальной последовательности в зависимости от выбранного τ .

Эксперимент

- Изменение числа значимых близостей для определения оптимального числа обусловленности скорректированного минора:



- При статистическом подходе следует определить $\tau = 0.6$, т.к. для выборок всех размеров графики резко падают начиная с этого уровня.
- Аналогичные графики получены при коррекции последующих миноров. Для коррекции так же следует определить $\tau = 0.6$.

Выводы

- Технология коррекции метрических нарушений в матрицах парных сравнений предназначена для исправления экспериментальных матриц парных сравнений и обеспечивает улучшение обусловленности скорректированных матриц.
- Это позволяет избежать применения теоретически сложных вычислительных алгоритмов для обработки плохо обусловленных матриц и математически корректно применить алгоритмы машинного обучения при решении задач распознавания, группировки, кластер-анализа и т.д.
- Предложенная технология может служить основанием для создания новых алгоритмов.
- В частности, процедура локализации позволяет получить множество наименее коррелирующих признаков, которые могут являться начальным решением в задаче факторного анализа.

**Спасибо за
внимание!**