Метрическая медиана Кемени и кластеризация ранжирований

Двоенко С.Д.
Тульский государственный университет

O & ML: Семинар по оптимизации, машинному обучению искусственному интеллекту (СПбГУ)
20.03.2025

52 слайда

Метрическая медиана Кемени и кластеризация ранжирований

План

1. Проблематика обработки парных сравнений и задачи обработки
2. Метрическая медиана Кемени
3. Заключение
4. Публикации

1. Проблематика обработки парных сравнений и задачи обработки

1.1 Структурные объекты и отношения
1.2 Характеристики структур
1.3 Погружение структур
1.4 Перебор вариантов
1.5 Преодоление дискретности
1.6 Задача погружения в метрическое пространство
1.7 Задача погружения отношений
1.8 Задача разработки алгоритмов

1.1 Структурные объекты и отношения

- В современном интеллектуальном анализе данных (ИАД) часто изучаются сложные структурные объекты (структуры)
- С одной стороны, – это реальные объекты (изображения, белковые структуры и т.п.)
- С другой стороны, – это результаты измерений в разных шкалах. Такие измерения позволяют представить отношения между объектами реального мира

1.1 Структурные объекты и отношения

- Отношения между объектами реального мира часто изучаются на основе взаимодействия между их парами
- Формально – это бинарные отношения $P \subseteq A \times A$ на декартовом произведении всех пар из $A=\{a_1,\ldots a_N\}$
- Часто рассматриваются упорядочения альтернатив (ранжирования), например, при экспертном оценивании проектов

1.1 Структурные объекты и отношения

- Таким образом, структуры определенного вида являются *дискретными* объектами (бинарные отношения, в т.ч. ранжирования, разбиения)
- Часто их представляют матрицами отношений $M_P(N,N)$ с элементами

$$m_{ij} = \begin{cases} 1, (a_i, a_j) \in P \\ 0, (a_i, a_j) \notin P \end{cases} \qquad a_k \in A, k = 1, \ldots N$$

1.2 Характеристики структур

- Признаки позволяют представить объект точкой (вектором) в абстр. многомерном пространстве
- Признаки позволяют применить методы интеллектуального анализа данных (ИАД)
- Часто сложно (или даже невозможно) предложить систему характеристик (признаков) для описания структур
- Поэтому применяют непосредственные парные сравнения (различие или сходство) структур

1.3 Погружение структур

- С одной стороны, важна содержательная интерпретация различия или сходства
- С другой стороны, хочется, чтобы различие было расстоянием (евклидовым)
- Если это так, то множество дискретных объектов (структур) считается погруженным в многомерное признаковое пространство, *как бы* в результате измерения их характеристик (нам недоступных)

1.3 Погружение структур

- Ни характеристики, ни их число неизвестны
- Предполагается, что измерения признаков были выполнены, но потом - утрачены
- Пусть различия – это *расстояния* в метрическом пространстве
- Тогда неотрицательная функция сходства характеризует близость объектов и является *скалярным произведением* в одном и том же квадранте гипотетического пространства

1.3 Погружение структур

- Дискретность структур приводит к их погружению в континуальное метрическое пространство в виде множества изолированных точек
- Между ними нет других экземпляров таких структур
- Их или вообще не бывает, или их надо создавать специально

1.4 Перебор вариантов

- Дискретность структур и изолированность множества представляющих их точек в метрическом пространстве требует применения переборных алгоритмов поиска решений в общем случае
- Основная проблема (при прочих равных) – снижение трудоемкости перебора

1.4 Перебор вариантов

- Континуальность пространства позволяет конструировать (вычислять) новые объекты как новые точки в нем
- Новые объекты обычно обладают *экстремальными* свойствами (средние, ближайшие точки и т.п.)
- Это означает *отказ от перебора* и применение в общем случае алгоритмов градиентного типа

1.5 Преодоление дискретности

- Проблема рассматривается в задачах:
  - *построение сверхтранзитивных отношений* (Миркин Б.Г.)
  - *метризация отношений* (Литвак Б.Г.)
  - *согласование измерений в разнотипных шкалах* (Загоруйко Н.Г., Лбов Г.С.)
- Обычно методы решения этих задач интегрированы в задачи обработки
- Целесообразно выделить и обособить *три* актуальные задачи

1.6 Задача погружения в метрическое пространство

- Погружение множества в метрическое пространство, построение функций расстояния и похожести
  - ✓ Pekalska E., Duin R.P.W. (2005). The Dissimilarity Representation for Pattern Recognition. Foundations and Applications. Singapore: World Scientific.
  - ✓ Bognar J. (1974). Indefinite Inner Product Spaces. N.Y.: Springer-Verlag.
  - ✓ Kemeny, J., Snell, J. (1963). Mathematical Models in the Social Sciences. Blaisdell, NY.
- *Технология метрической коррекции данных в виде парных сравнений для применения известных и разработки новых алгоритмов машинного обучения*

1.7 Задача погружения отношений

- Погружение бинарных отношений (ранжирования), установление соответствия между дискретными объектами и их континуальными окрестностями
  - ✓ Ногин В.Д. (2016). Сужение множества Парето: аксиоматический подход. – М.: Физматлит.
  - ✓ Кузьмин В.Б. (1982). Построение групповых решений в пространствах четких и нечетких бинарных отношений. – М.: Наука.
  - ✓ Миркин Б.Г. (1972). Проблема группового выбора. – М.: Наука.

- *Метрическая медиана Кемени*

1.8 Задача разработки алгоритмов

- Разработка методов и алгоритмов анализа структурных объектов
  - ✓ Ногин В.Д. (2016). Сужение множества Парето: аксиоматический подход. – М.: Физматлит.
  - ✓ Ларичев О.И., Мошкович Е.М. (1996). Качественные методы принятия решений. – М.: Наука.
  - ✓ Миркин Б.Г. (1976). Анализ качественных признаков. – М.: Статистика.

- *Известные и новые алгоритмы машинного обучения для парных сравнений: k-means, k-meanless, Форель, ОРГ Б.Н. Козинца*

2.Метрическая медиана Кемени

2.1 Построение классической МК
2.2 Алгоритм МК
2.3 Новый алгоритм МК
2.4 Метрическая медиана Кемени
2.5 Модификация матрицы потерь
2.6 Кластеризация ранжирований
2.7 Пример
2.8 Взвешенная МК
2.9 Эксперименты

2.1 Построение классической МК

- Ранжирования $P_1,\dots P_n$ альтернатив (проектов) $A=\{a_1,\dots a_N\}$ представлены матрицами отношений $M\,(N,N)$ с элементами

$$m_{ij} = \begin{cases} 1, & a_i \succ a_j \\ 0, & a_i \sim a_j \\ -1, & a_i \prec a_j \end{cases}$$

- Расстояние между ранжированиями P_u и P_v:

$$d(P_u, P_v) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}|m_{ij}^u - m_{ij}^v|$$

2.1 Построение классической МК

- Медиана Кемени *P** строится как ранжирование, наименее удаленное от всех остальных

$$P^* = \arg\min_P \sum_{u=1}^{n} d(P, P_u)$$

- Пусть

$$\sum_{u=1}^{n} d(P,P_u) = \frac{1}{2}\sum_{u=1}^{n}\sum_{i=1}^{N}\sum_{j=1}^{N} | m_{ij} - m_{ij}^u | = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{u=1}^{n} | m_{ij} - m_{ij}^u |\right) =$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{u=1}^{n} d_{ij}(P,P_u)\right) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} q_{ij}$$

2.1 Построение классической МК

- Медиана P^* строится по матрице потерь $Q(N,N)$:

$$q_{ij} = \sum_{u=1}^{n} d_{ij}(P, P_u) \quad d_{ij}(P, P_u) = \begin{cases} 0, & m_{ij}^u = 1 \\ 1, & m_{ij}^u = 0 \\ 2, & m_{ij}^u = -1 \end{cases}$$

- Это штрафы экспертов на наши предположения о парах альтернатив $a_i \succ a_j$ (m_{ij}=1) в неизвестном ранжировании P :

$$P^* = \arg\min_P \sum_{u=1}^{n} d(P, P_u) \,.$$

2.2 Алгоритм МК

1. Вычислить суммы по строкам в матрице $Q(N,N)$
2. Поместить очередную альтернативу с минимальной суммой на очередное место в ранжировании (если их несколько с одной суммой, то взять все), получив ранжирование P_I
3. Просмотреть P_I в обратном порядке: если расположение пары альтернатив *не* соответствует матрице потерь (при $a_i \succ a_j$ *не* выполнено $q_{ij} \leq q_{ji}$), то переставить альтернативы местами: $a_j \succ a_i$
4. Получено ранжирование P_{II} – это медиана Кемени

(Литвак Б.Г. (1982) Экспертная информация. Методы получения и анализа. М.: Радио и связь.)

2.3 Новый алгоритм МК

- Для «длинных» ранжирований часто сложно определить предпочтения на парах объектов после их перестановки в P_{II}
- *Новый алгоритм медианы Кемени*:

1. Построить ранжирование P_I, строки и столбцы матрицы потерь Q_I упорядочить в соответствии с P_I
2. Просмотреть P_I в обратном порядке, выполнив, если нужно, перестановки в парах альтернатив и разместить их без указания предпочтений между ними, получив перестановку R и матрицу Q_R, строки и столбцы которой упорядочены в соответствии с R
3. В полученной перестановке R поочередно удалить альтернативы из Q_R, получив ранжирование P_{II}. Если есть несколько альтернатив с одинаковой суммой по строке, то удалять их одновременно, указав как неразличимые в P_{II}

2.4 Метрическая медиана Кемени

- Пусть ранжирования $P_1,\ldots P_n$ погружены в метрическое пространство без нарушений и представлены парными расстояниями $D(n,n)$
- МК P^* представлена своими расстояниями до остальных ранжирований $d(P^*, P_i)$, $i=1,\ldots n$
- Центр «тяжести» P_0 также представлен своими расстояниями до остальных ранжирований

$$d^2(P_0, P_i) = d_{0i}^2 = \frac{1}{n}\sum_{p=1}^{n} d_{ip}^2 - \frac{1}{2n^2}\sum_{p=1}^{n}\sum_{q=1}^{n} d_{pq}^2, \; i = 1, \ldots n$$

2.4 Метрическая медиана Кемени

- Расстояния $d(P^*, P_i)$, i=1,…n и $d(P_o, P_i)$, i=1,…n *отличаются*, хотя должно быть $P^*=P_o$
- ЦТ P_o в виде ранжирования является ММК
- ММК P_0^* – это МК P^*, у которой расстояния скорректированы и совпадают с расстояниями ЦТ P_o до остальных ранжирований
- ММК строится по модифицированной матрице потерь $Q(N,N)$, где эксперты «по нашей просьбе» немного изменили свои штрафы на наши предположения

2.4 Метрическая медиана Кемени

- Обычно ММК содержит меньше неразличимых альтернатив, чем МК, т.к. «менее дискретна»
- Содержательно ММК не отличается от МК, лишь уточняя взаимные положения некоторых ранее неразличимых альтернатив
- Формально ММК как бинарное отношение в этом случае отличается от МК
- Модификация $Q(N,N)$ выполняется итерационно, пока групповое бинарное отношение P_0^* для ЦТ P_o изменяется

2.5 Модификация матрицы потерь

Рассмотрим ранжирования P_u, P_0 и P^*, где $\delta = d(P_0, P_u) - d(P^*, P_u) \neq 0$. Пусть $\delta > 0$. Для компенсации такой разницы в расстояниях, необходимо равномерно распределить значение $\delta > 0$ по ненулевым элементам матрицы отношений M_{P_u}, сформировав новую матрицу отношений с элементами:

$$m_{ij}^u = \begin{cases} +1 + 2\delta/k\,, & a_i \succ a_j \\ 0\,, & a_i \sim a_j \\ -1 - 2\delta/k\,, & a_i \prec a_j \end{cases}$$

где $k = N^2 - N - N_0$ – общее число ненулевых элементов без главной диагонали, N_0 – число нулевых недиагональных элементов $m_{ij}^u = 0$. Очевидно, что такая матрица отношений не изменяет ранжирования эксперта P_u.

В этом случае при $m_{ij} = 1$ предположение $a_i \succ a_j$ в неизвестном ранжировании P штрафуется экспертом с формированием частичных расстояний:

$$d_{ij}(P, P_u) = \begin{cases} 0 + 2\delta/k\,, & m_{ij}^u = +1 + 2\delta/k \\ 1\,, & m_{ij}^u = 0 \\ 2 + 2\delta/k\,, & m_{ij}^u = -1 - 2\delta/k \end{cases}$$

2.5 Модификация матрицы потерь

Пусть $\delta < 0$. Для компенсации такой разницы в расстояниях также необходимо равномерно распределить значение $\delta < 0$ по ненулевым элементам матрицы отношений M_{P_u}. Кроме того, множество изменяемых элементов $k' = k - \Delta k$ дополнительно уменьшается за счет Δk совпадающих элементов $m_{ij}^u = m_{ij}^*$ в матрицах отношений, представляющих ранжирование эксперта P_u и медиану P^*. Действительно, в этом случае $d_{ij}(P^*, P_u) = 0$ и уменьшить его невозможно. Добавление любой отрицательной величины к значению m_{ij}^u в этом случае будет означать изменение ранжирования эксперта, при котором расстояние между ранжированиями P_u и P^* только возрастет.

2.5 Модификация матрицы потерь

Если величина $-1 < 2\delta/k' < 0$, то можно, не изменяя ранжирования P_u, сформировать новую матрицу отношений M_{P_u} с элементами:

$$m_{ij}^u = \begin{cases} +1 - |2\delta/k'|\,, & a_i \succ a_j \\ 0\,, & a_i \sim a_j \\ -1 + |2\delta/k'|\,, & a_i \prec a_j \\ m_{ij}^*\,, & m_{ij}^u = m_{ij}^*\,, \end{cases}$$

где при $m_{ij} = 1$ предположение $a_i \succ a_j$ в неизвестном ранжировании P штрафуется экспертом с формированием частичных расстояний:

$$d_{ij}(P, P_u) = \begin{cases} 0, & m_{ij}^u = 1 \\ 1, & m_{ij}^u = 0 \\ 2, & m_{ij}^u = -1 \\ 0 + |2\delta/k'|, & m_{ij}^u = +1 - |2\delta/k'| \\ 2 - |2\delta/k'|, & m_{ij}^u = -1 + |2\delta/k'|\,. \end{cases}$$

2.6 Кластеризация ранжирований

- Появление неразличимых альтернатив в МК и ММК часто вызвано не просто «дискретностью» группового отношения, но и разногласиями во мнениях, часто при большом числе альтернатив
- Согласованность экспертных мнений можно оценить
- Низкая согласованность мнений экспертов обычно указывает на появление групп, что приводит к задаче *кластеризации* для их выделения

2.6 Кластеризация ранжирований

- Пусть ранжирования $P_1,\dots P_n$ корректно погружены в метрическое пространство и представлены расстояниями $D(n,n)$
- Алгоритм *k-means* строит разбиение на K кластеров с центрами P_0^k, $k=1,\dots K$, представленными своими расстояниями до остальных $i=1,\dots n$ объектов (ранжирований):

$$d^2(P_0^k, P_i) = \frac{1}{n_k}\sum_{p=1}^{n_k} d^2(P_i, P_p) - \frac{1}{2n_k^2}\sum_{p=1}^{n_k}\sum_{q=1}^{n_k} d^2(P_p, P_q)$$

2.6 Кластеризация ранжирований

- В полученной кластеризации для ЦТ каждого кластера P_0^k, $k = 1,\dots K$ строятся их ММК $P_0^{*(k)}$
- Окончательное согласование индивидуальных мнений достигается методами формирования консенсуса (метод Дельфи и т.п.) при проведении экспертизы

2.7 Пример: данные

- Известные организации Consumer Reports, J.D.Power, Auto Express, TUV Rheinland и др. публикуют ежегодные результаты тестирования различной продукции, например легковых автомобилей. Обычно это частично пересекающиеся списки моделей автомобилей.
- Представлен сборный список автомобилей премиум-класса 2015-16 годов в списках этих организаций. Для номеров альтернатив используется код из двух цифр, чтобы отличать их от позиций в ранжировании
- Таким образом, рейтинг автомобилей в данном списке не является абсолютным.
- Эксперты (указанные выше организации) перечислены в произвольном порядке

2.7 Рейтинги автомобилей

Код	Модель	Ранжирования экспертов				P^*	P_0^*
01	Porsche	2	4	1	1	1	1
02	Lexus	1	1	6	6	2	2
03	Toyota	4	3	2	7	3	3
04	Buick	3	2	5	8	4	4
05	Chevrolet	6	7	4	4	5	5
06	Linkoln	8	5	7	5	6	6
07	BMW	12	12	3	3	11	12
08	GMC	5	8	8	11	7	7
09	Infiniti	11	11	9	2	8.5	9
10	M-Benz	10	6	11	9	8.5	8
11	Audi	9	10	10	10	11	11
12	Acura	7	9	12	12	11	10

2.7 Матрица расстояний

$$
\begin{array}{c}
1 \\ 2 \\ 3 \\ 4 \\ P* \\ P_0
\end{array}
\begin{pmatrix}
0 & 20 & 46 & 68 & 22 & 29.228 \\
20 & 0 & 50 & 64 & 22 & 28.605 \\
46 & 50 & 0 & 34 & 28 & 22.633 \\
68 & 64 & 34 & 0 & 46 & 39.221 \\
22 & 22 & 28 & 46 & 0 & 6.02 \\
29.228 & 28.605 & 22.633 & 39.221 & 6.02 & 0
\end{pmatrix}
$$

2.7 Коррекция штрафов экспертов

P^*	P_0	Различие δ	Число изменяемых элементов k	Среднее $2\delta / k$
22 22 28 46	29.228 28.605 22.633 39.221	+7.228 +6.605 -5.367 -6.779	132 132 32 50	+0.10951 +0.10008 -0.33544 -0.27118

2.7 Матрицы потерь

$$
\begin{array}{c}
01\\ 02\\ 03\\ 04\\ 05\\ 06\\ 07\\ 08\\ 09\\ 10\\ 11\\ 12
\end{array}
\begin{pmatrix}
0 & 4 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
4 & 0 & 2 & 2 & 4 & 2 & 4 & 0 & 2 & 0 & 0 & 0\\
6 & 6 & 0 & 4 & 2 & 2 & 2 & 0 & 2 & 0 & 0 & 0\\
6 & 6 & 4 & 0 & 4 & 2 & 4 & 0 & 2 & 0 & 0 & 0\\
8 & 4 & 6 & 4 & 0 & 2 & 4 & 2 & 2 & 2 & 0 & 0\\
8 & 6 & 6 & 6 & 6 & 0 & 4 & 2 & 2 & 0 & 0 & 2\\
8 & 4 & 6 & 4 & 4 & 4 & 0 & 4 & 6 & 4 & 4 & 4\\
8 & 8 & 8 & 8 & 6 & 6 & 4 & 0 & 2 & 4 & 2 & 0\\
8 & 6 & 6 & 6 & 6 & 6 & 2 & 6 & 0 & 4 & 4 & 4\\
8 & 8 & 8 & 8 & 6 & 8 & 4 & 4 & 4 & 0 & 4 & 2\\
8 & 8 & 8 & 8 & 8 & 8 & 4 & 6 & 4 & 4 & 0 & 4\\
8 & 8 & 8 & 8 & 8 & 6 & 4 & 8 & 4 & 6 & 4 & 0
\end{pmatrix}
$$

$$
\begin{array}{c}
01\\ 02\\ 03\\ 04\\ 05\\ 06\\ 07\\ 08\\ 09\\ 10\\ 11\\ 12
\end{array}
\begin{pmatrix}
0 & 4.21 & 2.21 & 2.21 & 0.21 & 0.21 & 0.21 & 0.21 & 0.21 & 0.21 & 0.21 & 0.21\\
4.21 & 0 & 1.87 & 1.87 & 3.60 & 1.94 & 3.60 & 0.21 & 1.94 & 0.21 & 0.21 & 0.21\\
6.21 & 6.55 & 0 & 4.21 & 1.94 & 1.94 & 1.94 & 0.21 & 1.94 & 0.21 & 0.21 & 0.21\\
6.21 & 6.55 & 4.21 & 0 & 3.60 & 1.94 & 3.60 & 0.21 & 1.94 & 0.21 & 0.21 & 0.21\\
8.21 & 4.82 & 6.48 & 4.82 & 0 & 2.21 & 3.60 & 2.21 & 1.94 & 2.21 & 0.21 & 0.21\\
8.21 & 6.48 & 6.48 & 6.48 & 6.21 & 0 & 3.60 & 2.21 & 1.94 & 0.21 & 0.21 & 2.21\\
8.21 & 4.82 & 6.48 & 4.82 & 4.82 & 4.82 & 0 & 4.82 & 6.55 & 4.82 & 4.82 & 4.82\\
8.21 & 8.21 & 8.21 & 8.21 & 6.21 & 6.21 & 3.60 & 0 & 1.94 & 3.94 & 1.94 & 0.21\\
8.21 & 6.48 & 6.48 & 6.48 & 6.48 & 6.48 & 1.87 & 6.48 & 0 & 4.82 & 4.22 & 4.21\\
8.21 & 8.21 & 8.21 & 8.21 & 6.21 & 8.21 & 3.60 & 4.48 & 3.60 & 0 & 3.87 & 2.21\\
8.21 & 8.21 & 8.21 & 8.21 & 8.21 & 8.21 & 3.60 & 6.48 & 4.21 & 4.55 & 0 & 4.82\\
8.21 & 8.21 & 8.21 & 8.21 & 8.21 & 6.21 & 3.60 & 8.21 & 4.21 & 6.21 & 3.60 & 0
\end{pmatrix}
$$

МК **ММК**

2.7 Ранжирования

ранжи-рования \ место	1	2	3	4	5	6	7	8	9	10	11	12
P^*	01	$\succ$ 02	$\succ$ 03	$\succ$ 04	$\succ$ 05	$\succ$ 06	$\succ$ 08	$\succ$ 10	$\sim$ 09	$\succ$ 12	$\sim$ 11	$\sim$ 07
P_0^*	01	$\succ$ 02	$\succ$ 03	$\succ$ 04	$\succ$ 05	$\succ$ 06	$\succ$ 08	$\succ$ 10	$\succ$ 09	$\succ$ 12	$\succ$ 11	$\succ$ 07
P_1	02	$\succ$ 01	$\succ$ 04	$\succ$ 03	$\succ$ 08	$\succ$ 05	$\succ$ 12	$\succ$ 06	$\succ$ 11	$\succ$ **10**	$\succ$ **09**	$\succ$ **07**
P_2	02	$\succ$ 04	$\succ$ 03	$\succ$ 01	$\succ$ 06	$\succ$ **10**	$\succ$ 05	$\succ$ 08	$\succ$ 12	$\succ$ 11	$\succ$ **09**	$\succ$ **07**
P_3	01	$\succ$ 03	$\succ$ **07**	$\succ$ 05	$\succ$ 04	$\succ$ 02	$\succ$ 06	$\succ$ 08	$\succ$ **09**	$\succ$ 11	$\succ$ **10**	$\succ$ 12
P_4	01	$\succ$ **09**	$\succ$ **07**	$\succ$ 05	$\succ$ 06	$\succ$ 02	$\succ$ 03	$\succ$ 04	$\succ$ **10**	$\succ$ 11	$\succ$ 08	$\succ$ 12

2.7 Кластеризация ранжирований

P_1	0	20	**46**	**68**
P_2	20	0	**50**	**64**
P_3	**46**	**50**	0	34
P_4	**68**	**64**	34	0

2.7 Кластеризация на основе ММК

- Построение ММК применимо для измерений в менее мощных шкалах (ранжирования)
- ММК позволяет кластеризовать ранжирования обычным *k-means* : после его остановки центры кластеров определяются как ММК

- *Недостаток*: необходимо помнить коррекции штрафов экспертов, т.к. без них будет снова получена обычная МК (устранение – взвешенная МК)

2.8 Взвешенная МК

- Назначим веса экспертов: $w_u \geq 0, u = 1,\dots n$, $\sum_{u=1}^{n} w_u = 1$
- Взвешенная матрица потерь $Q(N,N) = \sum_{u=1}^{n} w_u Q_u(N,N)$ состоит из элементов $q_{ij} = \sum_{u=1}^{n} w_u d_{ij}(P,P_u)$.
- Взвешенные частичные расстояния при условии $m_{ij} = 1$ определены как

$$\tilde{d}_{ij}(P,P_u) = w_u d_{ij}(P,P_u) = w_u \,| m_{ij} - m_{ij}^u | = \begin{cases} 0, m_{ij}^u = 1 \\ w_u, m_{ij}^u = 0 \\ 2w_u, m_{ij}^u = -1. \end{cases}$$

- Линейная комбинация $P_1,\dots P_n$ определяет внутри выпуклой оболочки множества $P_1,\dots P_n$ ранжирование P^* как взвешенную медиану Кемени (ВМК).

2.8 Взвешенная МК

- Ранжировка эксперта: $P_i = P^*;\ w_i = 1, w_{j \neq i} = 0,\ j = 1, \dots n.$
- МК: $P^*;\ w_i = 1,\ i = 1, \dots n$.
- МК: $P^*;\ w_i = \frac{1}{n},\ i = 1, \dots n.$
- МК: $P^*;\ w_i = const > 0,\ i = 1, \dots n$.
- Задача условной оптимизации поиска весов $w_1, \dots w_n$ для $\delta_u = d(P_0, P_u) - d(P^*, P_u), u = 1, \dots n$:

$$\sum_{u=1}^{n} \delta_u^2 \to \min, \sum_{u=1}^{n} w_u = 1,\ w_u \geq 0, u = 1, \dots n.$$

- **Решение**: основано на методе Гаусса-Зайделя покоординатного спуска, где вес каждого эксперта определяет «координатную» ось.

2.8 Взвешенная МК

- Свойства ВМК:
  - ВМК - это ММК P_0^*, т.к. это медиана P^* с такими же расстояниями, как у центра P_o .
  - В итоге, ранжирование P_0^* представлено своей матрицей отношений M_0^* с расстояниями до других ранжирований, представленных взвешенными исходными матрицами $M_u, u = 1, \ldots n$ в противоположность модифицированным матрицам для поиска ММК.

2.9 Эксперименты: данные

- **Данные**: автомобили, представленные в 2015-16 гг. в разных пересекающихся списках от Consumer reports, J.D. Power, Auto Express и TUV Rheiland в ежегодных обзорах результатов тестирования:

Number	Model	Expert rankings				P^*	P_0^*	$P_{0.6}^*$
01	Porsche	2	4	1	1	1	1	1
02	Lexus	1	1	6	6	2	2	2
03	Toyota	4	3	2	7	3	3	4
04	Buick	3	2	5	8	4	4	3
05	Chevrolet	6	7	4	4	5	5	5
06	Lincoln	8	5	7	5	6	6	6
07	BMW	12	12	3	3	11	12	12
08	GMC	5	8	8	11	7	7	7
09	Infiniti	11	11	9	2	8.5	9	11
10	M-Benz	10	6	11	9	8.5	8	8
11	Audi	9	10	10	10	11	11	10
12	Acura	7	9	12	12	11	10	9

ММК и ВМК представляют групповое отношение без неразличимых альтернатив

2.9 Расстояния

- Расстояния между ранжированиями $P_1,\ldots P_4$,МК P^*, ММК P_0^* и центром P_0 :

$$\begin{array}{c} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P^* \\ P_0 \\ P_0^* \end{array}\left(\begin{array}{cccc:ccc} \boxed{0} & \boxed{20} & 46 & 68 & 22 & 29.23 & 18 \\ \boxed{20} & \boxed{0} & 50 & 64 & 22 & 28.61 & 18 \\ 46 & 50 & \boxed{0} & \boxed{34} & 28 & 22.63 & 32 \\ 68 & 64 & \boxed{34} & \boxed{0} & 46 & 39.22 & 50 \\ \hdashline 22 & 22 & 28 & 46 & 0 & 6.02 & 4 \\ 29.23 & 28.61 & 22.63 & 39.22 & 6.02 & 0 & 10.6 \\ 18 & 18 & 32 & 50 & 4 & 10.6 & 0 \end{array}\right)$$

- Коррекция расстояний для ММК:

P^*	P_0	δ	$k(k')$	$2\delta / k$
22	29.228	+7.228	132	+0.10951
22	28.605	+6.605	132	+0.10008
28	22.633	-5.367	32	-0.33544
46	39.221	-6.779	50	-0.27118

2.9 Ранжирования

Place	1	2	3	4	5	6	7	8	9	10	11	12
P^*	01 $\succ$	02 $\succ$	03 $\succ$	04 $\succ$	05 $\succ$	06 $\succ$	08 $\succ$	10 $\sim$	09 $\succ$	12 $\sim$	11 $\sim$	07
P_0^*	01 $\succ$	02 $\succ$	03 $\succ$	04 $\succ$	05 $\succ$	06 $\succ$	08 $\succ$	10 $\succ$	09 $\succ$	12 $\succ$	11 $\succ$	07
P_1	02 $\succ$	01 $\succ$	04 $\succ$	03 $\succ$	08 $\succ$	05 $\succ$	12 $\succ$	06 $\succ$	11 $\succ$	10 $\succ$	09 $\succ$	07
P_2	02 $\succ$	04 $\succ$	03 $\succ$	01 $\succ$	06 $\succ$	10 $\succ$	05 $\succ$	08 $\succ$	12 $\succ$	11 $\succ$	09 $\succ$	07
P_3	01 $\succ$	03 $\succ$	07 $\succ$	05 $\succ$	04 $\succ$	02 $\succ$	06 $\succ$	08 $\succ$	09 $\succ$	11 $\succ$	10 $\succ$	12
P_4	01 $\succ$	09 $\succ$	07 $\succ$	05 $\succ$	06 $\succ$	02 $\succ$	03 $\succ$	04 $\succ$	10 $\succ$	11 $\succ$	08 $\succ$	12

- Легко увидеть, что кластеры различаются характерными предпочтениями, показанными на слайде
- В МК соответствующие альтернативы неразличимы в парах
- В ММК воспроизводятся характерные предпочтения 1-го кластера, все альтернативы различимы

2.9 Типы МК

Place	1	2	3	4	5	6	7	8	9	10	11	12
P^*	01 $\succ$	02 $\succ$	03 $\succ$	04 $\succ$	05 $\succ$	06 $\succ$	08 $\succ$	10 $\sim$	09 $\succ$	12 $\sim$	11 $\sim$	07
P_0^*	01 $\succ$	02 $\succ$	03 $\succ$	04 $\succ$	05 $\succ$	06 $\succ$	08 $\succ$	10 $\succ$	09 $\succ$	12 $\succ$	11 $\succ$	07
$P^*_{0.6}$	01 $\succ$	02 $\succ$	04 $\succ$	03 $\succ$	05 $\succ$	06 $\succ$	08 $\succ$	10 $\succ$	12 $\succ$	11 $\succ$	09 $\succ$	07
$P^*_{0.4}$	01 $\succ$	03 $\succ$	05 $\succ$	02 $\succ$	04 $\succ$	06 $\succ$	08 $\succ$	09 $\succ$	07 $\succ$	10 $\succ$	11 $\succ$	12
P^{*1}	02 $\succ$	04 $\succ$	01 $\sim$	03 $\succ$	05 $\sim$	06 $\sim$	08 $\succ$	12 $\succ$	10 $\sim$	11 $\succ$	09 $\succ$	07

- Для наглядности альтернативам каждого кластера назначались одинаковые веса
- Легко увидеть, что ВМК1 $P^*_{0.6}$ с весами (0.6, 0.4) воспроизводит предпочтения 1-го кластера что соответствует ММК
- В то же время ВМК2 $P^*_{0.4}$ с весами (0.4, 0.6) воспроизводит предпочтения 2-го кластера
- МК только для 1-го кластера также содержит неразличимые альтернативы

2.9 Матрицы потерь

Q_{MK}

01	0	4	2	2	0	0	0	0	0	0	0	0
02	4	0	2	2	4	2	4	0	2	0	0	0
03	6	6	0	4	2	2	2	0	2	0	0	0
04	6	6	4	0	4	2						
05	8	4	6	4	0	2						
06	8	6	6	6	6	0						
07	8	4	6	4	4	4						
08	8	8	8	8	6	6						
09	8	6	6	6	6	6						
10	8	8	8	8	6	8						
11	8	8	8	8	8	8						
12	8	8	8	8	8	6						

Q_{MMK}

01	0	4.21	2.21	2.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
02	4.21	0	1.87	1.87	3.60	1.94	3.60	0.21	1.94	0.21	0.21	0.21
03	6.21	6.55	0	4.21	1.94	1.94	1.94	0.21	1.94	0.21	0.21	0.21
04	6.21	6.55	4.21	0	3.60	1.94	3.60	0.21	1.94	0.21	0.21	0.21
05	8.21	4.82	6.48	4.82	0	2.21	3.60	2.21	1.94	2.21	0.21	0.21
06	8.21	6.48	6.48	6.48	6.21	0	3.60	2.21	1.94	0.21	0.21	2.21
07	8.21	4.82	6.48	4.82	4.82	4.82	0	4.8				
08	8.21	8.21	8.21	8.21	6.21	6.21	3.60					
09	8.21	6.48	6.48	6.48	6.48	6.48	1.87	6.4				
10	8.21	8.21	8.21	8.21	6.21	8.21	3.60	4.4				
11	8.21	8.21	8.21	8.21	8.21	8.21	3.60	6.4				
12	8.21	8.21	8.21	8.21	8.21	6.21	3.60	8.2				

Q_{BMK1}

01	0	4.8	2.4	2.4	0	0	0	0	0	0	0	0
02	3.6	0	1.6	1.6	3.2	1.6	3.2	0	1.6	0	0	0
03	5.6	6.4	0	4.8	1.6	1.6	1.6	0	1.6	0	0	0
04	5.6	6.4	3.2	0	3.2	1.6	3.2	0	1.6	0	0	0
05	8	4.8	6.4	4.8	0	2.4	3.2	2.4	1.6	2.4	0	0
06	8	6.4	6.4	6.4	5.6	0	3.2	2.4	1.6	0	0	2.4
07	8	4.2	6.4	4.8	4.8	4.8	0	4.8	6.4	4.8	4.8	4.8
08	8	8	8	5.6	5.6	5.6	3.2	0	1.6	4	1.6	0
09	8	6.4	6.4	6.4	6.4	6.4	1.6	6.4	0	4.8	4.8	4.8
10	8	8	8	8	5.6	8	3.2	4	3.2	0	4	2.4
11	8	8	8	8	8	8	3.2	6.4	3.2	4	0	4.8
12	8	8	8	8	8	5.6	3.2	8	3.2	5.6	3.2	0

2.9 Все ранжирования вместе

Place	1	2	3	4	5	6	7	8	9	10	11	12
P^*	01 ≻	02 ≻	03 ≻	04 ≻	05 ≻	06 ≻	08 ≻	10 ∼	09 ≻	12 ∼	11 ∼	07
P_0^*	01 ≻	02 ≻	03 ≻	04 ≻	05 ≻	06 ≻	08 ≻	10 ≻	09 ≻	12 ≻	11 ≻	07
P_1	02 ≻	01 ≻	04 ≻	03 ≻	08 ≻	05 ≻	12 ≻	06 ≻	11 ≻	10 ≻	09 ≻	07
P_2	02 ≻	04 ≻	03 ≻	01 ≻	06 ≻	10 ≻	05 ≻	08 ≻	12 ≻	11 ≻	09 ≻	07
P_3	01 ≻	03 ≻	07 ≻	05 ≻	04 ≻	02 ≻	06 ≻	08 ≻	09 ≻	11 ≻	10 ≻	12
P_4	01 ≻	09 ≻	07 ≻	05 ≻	06 ≻	02 ≻	03 ≻	04 ≻	10 ≻	11 ≻	08 ≻	12

Place	1	2	3	4	5	6	7	8	9	10	11	12
P^*	01 ≻	02 ≻	03 ≻	04 ≻	05 ≻	06 ≻	08 ≻	10 ∼	09 ≻	12 ∼	11 ∼	07
P_0^*	01 ≻	02 ≻	03 ≻	04 ≻	05 ≻	06 ≻	08 ≻	10 ≻	09 ≻	12 ≻	11 ≻	07
$P^*_{0.6}$	01 ≻	02 ≻	04 ≻	03 ≻	05 ≻	06 ≻	08 ≻	10 ≻	12 ≻	11 ≻	09 ≻	07
$P^*_{0.4}$	01 ≻	03 ≻	05 ≻	02 ≻	04 ≻	06 ≻	08 ≻	09 ≻	07 ≻	10 ≻	11 ≻	12
P^{*1}	02 ≻	04 ≻	01 ∼	03 ≻	05 ∼	06 ∼	08 ≻	12 ≻	10 ∼	11 ≻	09 ≻	07

2.9 Результаты

- ВМК1 для весов (0.6, 0.4) содержит предпочтения $10 \succ 09$ и $12 \succ 11$, где 07 находится на последней позиции в (P_1,P_2)
- ВМК2 для весов (0.4, 0.6) содержит противоположные предпочтения $09 \succ 10$ и $11 \succ 12$ с 07 перед ними как в (P_3,P_4)
- В обеих ВМК отсутствуют неразличимые альтернативы в противовес МК P^*
- ВМК лучше различает альтернативы, чем МК
- ММК и ВМК1 содержат одинаковые предпочтения $10 \succ 09$ и $12 \succ 11$, где 07 находится в последней позиции как в (P_1,P_2).
- Группа (P_1,P_2) более компактна. Т.к. МК ближе к ней, то ВМК1 более предпочтительна для ЛПР

3. Заключение

- Как бинарное отношение ММК не обязательно совпадает с МК
- ММК является метрически корректным центром множества ранжирований
- ММК является базисом для решения задач группового выбора как задач кластеризации, где консенсус ищется как подходящая ВМК для разных групп экспертов как кластеров
- Практический подход к агрегированию рангов при конфликтах основан на модификации матриц потерь в алгоритме Кемени.

3. Заключение

- Практический результат заключается в следующем: не следует опираться только на одно из конфликтующих мнений, лучше скомбинировать в некоторой пропорции предпочтительное мнение с другими конфликтующими мнениями, чтобы сделать его лучше и исключить неразличимые альтернативы

4. Публикации

1. Dvoenko S., Kurbakov M., Tran X.M. (**2022**) On ranking of image quality metrics. ICABDE 2021. LNDECT, vol. 124. Pp. 421-433. Springer Nature. //doi.org/10.1007/978-3-030-97610-1_33
2. Dvoenko S.D. (**2021**) A developing of the Kemeny median: new types and algorithms. Pattern Recognition and Image Analysis. 31(3): 376–380. Springer Nature. //doi.org/10.1134/S105466182103007X
3. Dvoenko S.D., Pshenichny D.O. (**2021**) Rank aggregation based on new types of the Kemeny's median. Pattern Recognition and Image Analysis. 31(2): 185-196. Springer Nature. //doi.org/10.1134/S1054661821020061
4. Dvoenko S., Pshenichny D. (**2021**). On new Kemeny's medians. ICPR 2021 Workshops. LNCS, vol. 12665. Pp. 97–102. Springer Nature. //doi.org/10.1007/978-3-030-68821-9_9
5. Dvoenko S.D., Pshenichny D.O. (**2019**) On a metric Kemeny's median. Intelligent Data Processing. IDP 2016. CCIS, vol. 794. Pp 44–57. Springer Cham. //doi.org/10.1007/978-3-030-35400-8_4

Двоенко С.Д. Дополнение к докладу. Об эвристическом алгоритме подбора весов линейной комбинации матриц отношений в задаче построения медианы Кемени.

Задача. Пусть ранжирования представлены матрицей расстояний $D(n,n)$, ЦТ P_0 представлен своими расстояниями $d(P_0,P_u)$, $u=1,\dots n$, МК P^* представлена своими расстояниями $d(P^*,P_u)$, $u=1,\dots n$. Найдем различия $\delta_u = d(P_0,P_u) - d(P^*,P_u)$, $u=1,\dots n$. Решить задачу $\sum_{u=1}^{n}\delta_u^2 \to \min$ при ограничениях $\sum_{u=1}^{n} w_u = 1$, $w_u \ge 0, u=1,\dots n$.

Рассмотрим процедуру определения весов индивидуальных матриц штрафов $Q_u(N,N)$ экспертов $u=1,\dots n$. Данная процедура основана на схеме алгоритма покоординатного спуска Гаусса-Зайделя, где каждый эксперт определяет условную координатную ось.

Будем считать аналогом покоординатного варьирования изменение веса индивидуальной матрицы штрафов эксперта $Q_u(N,N)$ на интервале от 0 до 1. На начальном шаге индивидуальные веса всех n экспертов изменяются от $1/(n$-$1)$, …0,…$1/(n$-$1)$, когда матрица штрафов $Q_u(N,N)$ полностью исключена, а веса остальных n-1 матриц индивидуальных штрафов одинаковы, до весов 0,…1,…0, когда учитывается только матрица штрафов $Q_u(N,N)$, а остальные исключены.

Каждый шаг заканчивается после поочередного варьирования веса каждого эксперта и выбора того из них u^*, варьирование веса штрафов которого обеспечило минимальное суммарное отклонение $\sum_{u=1}^{n}\delta_u^2 \to \min$ по разнице в расстояниях ЦТ P_0 и промежуточной ММК $P_0^{u^*}$ до остальных ранжирований.

На очередном шаге варьируется вес матрицы $Q_u(N,N)$ на интервале $0 \le p \le 1$. Он также определяет ее нормированный вес $w_u = p$. Остальным матрицам штрафов соответствуют найденные к данному шагу постоянные веса $q_i, i=1,\dots n$, где $i \ne u$, сумма которых постоянна $V = \sum_{i=1}^{n} q_i, i \ne u$

Нормированные веса матриц штрафов остальных ранжирований изменяются на интервале от $w_i = q_i / V$ до $w_i = 0$, принимая значения $w_i = q_i(1-p)/V$ внутри интервала варьирования. Исходные веса $p = q_u$, $q_i, i=1,\dots n$, $i \ne u$ соответствуют внутренней точке интервала варьирования $w_i = q_i, i=1,\dots n$.

На очередном шаге варьирования определяется линейная комбинация матриц штрафов $Q(N,N) = \sum_{u=1}^{n} w_u Q_u(N,N)$, строится МК алгоритмом Кемени, и определяются отклонения $\delta_u, u=1,\dots n$.

Решение данной задачи, вообще говоря, по своей формулировке не гарантирует достижения нулевой разницы в расстояниях ЦТ P_0 и ММК P_0^* до остальных ранжирований. Поэтому, если это так, то предлагается, приняв найденную ММК за исходную ММК P_0’, модифицировать матрицы отношений экспертов, получив окончательное ранжирование, представляющее ЦТ P_0 как ММК P_0^*. Ожидается, что итоговое ранжирование уже не изменится.

Следует отметить, что этот эвристический алгоритм задуман как общая схема для подбора весов линейной комбинации чего бы то ни было как дискретных объектов в подходящем представлении. Естественно, с эвристическими дополнениями по необходимости, например, как выше в задаче о медиане [1]. Например, эта схема была применена в задаче обработки двумерных массивов (растровые изображения с 4- или 8-связной решеткой соседства) и трехмерных массивов (кубы пространственного представления толщи пород отраженными сигналами в геосейсмических испытаниях с 6- или 14-связной решеткой соседства, а вообще-то – для любых решеток с произвольным числом связей их элементов) [2].

Напомним, что в решетке нет направления (некаузальная структура, в отличие от сигнала, распространяющегося вдоль некоторой оси). Тогда такой граф, как некаузальную структуру, следует заменить набором каузальных структур – древовидных графов соседства. Для каждого графа соседства обработка организуется последовательно: от терминалов к корню (например, вычисляются т.н. фильтрационные вероятности отнесения отсчета к определенному классу принадлежности как протяженной однородной в некотором смысле области), а потом - от корня к терминалам (пересчет в т.н. интерполяционные вероятности). Но деревья получены удалением некоторых ребер решетки соседства, что, естественно, ведет к потере информации о соседстве. Поэтому важно сформировать много разных деревьев так, чтобы покрыть все множество связей решетки соседства, какой бы она ни была. Очевидно, что разные деревья соседства своими связями отражают разные особенности свойств каждой из текстурных областей. Поэтому их надо скомбинировать оптимальным образом.

Следует также заметить, что эта задача была обобщена на т.н. «полное множество» ациклических графов соседства, где они комбинируются естественным образом без необходимости поиска весов [2].

[1] Dvoenko, S.D. A Developing of the Kemeny Median: New Types and Algorithms. *Pattern Recognit. Image Anal.* **31**, 376–380 (2021). https://doi.org/10.1134/S105466182103007X

[2] Dvoenko, S.D. Recognition of dependent objects based on acyclic Markov models. *Pattern Recognit. Image Anal.* **22**, 28–38 (2012). https://doi.org/10.1134/S1054661812010130