ПОРОГОВОЕ АГРЕГИРОВАНИЕ ВЕРОЯТНОСТНЫХ РАНЖИРОВОК

А.Е. Лепский

Международный центр анализа и выбора решений, НИУ ВШЭ, Москва

XIV Всероссийское совещание по проблемам управления (ВСПУ-2024),
17–20 июня, 2024, ИПУ РАН, Москва

Задача ранжирования альтернатив

Задача. Требуется ранжировать альтернативы, каждая из которых оценивается по ряду критериев.
Популярный метод решения — агрегирование критериальных векторов, соответствующих каждой альтернативе, в ранги этих альтернатив [Подиновский 2023].

Зачастую такие правила должны иметь **некомпенсаторный характер**, т.е. низкие оценки по одним критериям не могут быть компенсированы высокими оценками по другим критериям.

Примеры: выбор товаров по ряду критериев, отбор кандидатов на должность и т. д.

Одним из некомпенсторных методов агрегирования предпочтений является **пороговое правило агрегирования, ПА** [Алескеров и Якуба 2003, Алескеров и др 2007].

План презентации

Классическое ПА: все альтернативы оцениваются в m-градационной шкале.

Оценки в альтернативах могут быть **неточными**. В этом случае возникает задача **обобщения ПА**.

Характер неточностей данных (например, экспертных оценок): вероятностный, интервальный, нечеткий и т. д.

План презентации

- Классическая постановка ППА;
- Задача ПА с неточными данными;
- Задача ПА с вероятностными данными
  - Случайная мощность множества оценок;
  - Свойства случайной мощности;
  - Лексикографическое ранжирование случайных мощностей;
- Численный пример;
- Заключение.

Классическая постановка ПА

[Алескеров и Якуба 2003, Алескеров и др 2007]

Пусть X — множество альтернатив-векторов $\mathbf{x} = (x_1, ..., x_n)$, где $x_i \in \{1, 2, 3\}$. Требуется построить оператор агрегирования $\varphi_n = \varphi : X \to \mathbb{R}$, удовлетворяющий условиям:

1) Парето-доминирование: если $\mathbf{x}, \mathbf{y} \in X$ и $x_i \geq y_i \quad \forall i$, $\exists s: \ x_s > y_s$, то $\varphi(\mathbf{x}) > \varphi(\mathbf{y})$;
2) попарная компенсируемость критериев: если $\mathbf{x}, \mathbf{y} \in X$ и $v_k(\mathbf{x}) = v_k(\mathbf{y})$ $k = 1, 2$, то $\varphi(\mathbf{x}) = \varphi(\mathbf{y})$, где $v_k(\mathbf{x}) = |\{i : x_i = k\}|$ — число оценок k в альтернативе $\mathbf{x}$, $k = 1, 2, 3$;
3) пороговая некомпенсируемость: $\varphi(\underbrace{2, \ldots, 2}_{n}) > \varphi(\mathbf{x})$ $\forall \mathbf{x} \in X$: $\exists s: \ x_s = 1$;
4) аксиома редукции: если $\forall \mathbf{x}, \mathbf{y} \in X$ $\exists s: \ x_s = y_s$, то $\varphi_n(\mathbf{x}) > \varphi_n(\mathbf{y}) \Leftrightarrow \varphi_{n-1}(\mathbf{x}_{-s}) > \varphi_{n-1}(\mathbf{y}_{-s})$, где $\mathbf{x}_{-s} = (x_1, \ldots, x_{s-1}, x_{s+1}, \ldots, x_n)$.

Показано, что указанным аксиомам удовлетворяет **лексикографическое правило** агрегирования:

$$\varphi(\mathbf{x}) > \varphi(\mathbf{y}) \Leftrightarrow$$

$$v_1(\mathbf{x}) < v_1(\mathbf{y})$$

$$\text{или } \exists j \in \{1, 2\} : v_k(\mathbf{x}) = v_k(\mathbf{y}) \forall k \leq j \text{ и } v_{k+1}(\mathbf{x}) < v_{k+1}(\mathbf{y}).$$

В [Aleskerov et al 2010] задача ПА была обобщена на случай m-градационных шкал, $m \geq 3$.

Задача ПА с неточными данными

Два основных подхода к построению обобщений ПА на случай неточных данных:

- построение аксиоматики и правила агрегирования на основе этой новой аксиоматики;
- построение аналога вектора мощности оценок для неточных данных.

Аналог вектора мощности оценок для неточных данных:

- построение вектора скалярных мощностей множества неточных оценок;
- построение неточной функции мощности множества неточных оценок.

Скалярный подход — используется неотрицательная (скалярная) функция, определенная на множестве неточных оценок и обобщающая понятие мощности. Например, в [Lepskiy 2023] построена скалярная функция мощности для множества нечетких экспертных оценок.

Неточная функция мощности множества неточных оценок определенного рана строится в рамках рассматриваемой модели неточности. Например, в [Lepskiy 2024] рассматривалась нечеткая мощность для множества нечетких оценок.

Задача ПА с вероятностными данными

Пусть каждая альтернатива представлена n-мерным вектором $\widetilde{\mathbf{x}} = (\widetilde{x_1}, \ldots, \widetilde{x_n})$, где оценка $\widetilde{x}_i$ характеризуется тройкой неотрицательных чисел $\widetilde{x}_i = (p_{iL}, p_{iM}, p_{iH})$, $p_{iL} + p_{iM} + p_{iH} = 1$, $i = 1, \ldots, n$. Числа p_{iL}, p_{iM} и p_{iH} — субъективные вероятности, с которыми i-й эксперт оценивает принадлежность альтернативы классам низких оценок L, средних оценок M и высоких оценок H соответственно.

Относительно альтернативы $\widetilde{\mathbf{x}}$ имеем матрицу

$$\mathbf{P}\left(\widetilde{\mathbf{x}}\right) = \left(p_{iS}\right)_{i=1;\, S=L,M,H}^{n}.$$

Случайная мощность множества оценок

Рассмотрим СВ $C_S(\widetilde{\mathbf{x}})$, принимающую значения $k \in \{0, \ldots, n\}$ и $P\{C_S(\widetilde{\mathbf{x}}) = k\}$ — вероятность того, что среди оценок n экспертов будет ровно k оценок класса $S \in \{L, M, H\}$.

СВ $C_S(\widetilde{\mathbf{x}})$ — случайные мощности множества оценок класса $S \in \{L, M, H\}$.

Пусть $I_k = \{(i_1, \ldots, i_k) : 1 \leq i_1 < \ldots < i_k \leq n\}$. Тогда $P\{C_S(\widetilde{\mathbf{x}}) = k\} = q_k(\mathbf{p}_S)$, где $\mathbf{p}_S = (p_{1S}, \ldots, p_{nS})$ и

$$q_0(\boldsymbol{\alpha}) = (1 - \alpha_1) \ldots (1 - \alpha_n),$$

$$q_k(\boldsymbol{\alpha}) = \sum_{(i_1, \ldots, i_k) \in I_k} \alpha_{i_1} \ldots \alpha_{i_k} (1 - \alpha_{j_1}) \ldots (1 - \alpha_{j_{n-k}}), \quad (1)$$

где $\{j_1, \ldots, j_{n-k}\} = \{1, \ldots, n\} \backslash \{i_1, \ldots, i_k\}$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in [0, 1]^n$.

Свойства случайной мощности

Пусть $Q(\boldsymbol{\alpha})$ — СВ, имеющая распределение (1).

Отметим некоторые его свойства:

1) если $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \{0,1\}^n$, то $q_k(\boldsymbol{\alpha}) = \begin{cases} 1, & k = \sum_{i=1}^n \alpha_i, \\ 0, & k \neq \sum_{i=1}^n \alpha_i, \end{cases}$ $k = 0, \ldots, n$;
2) если $\alpha_1 = \ldots = \alpha_n = p \in (0,1)$, то $q_k = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}$, $k = 0, \ldots, n$ — биномиальное распределение;
3) $E[Q(\boldsymbol{\alpha})] = \sum_{k=1}^n \alpha_k$;
4) $\sigma^2[Q(\boldsymbol{\alpha})] = \sum_{k=1}^n \alpha_k(1-\alpha_k)$;
5) распределение $\mathbf{Q}(\boldsymbol{\alpha}) = (q_0(\boldsymbol{\alpha}), \ldots, q_n(\boldsymbol{\alpha}))$ — унимодальное, т. е. $\exists r, s \in \{0, \ldots, n\}$, $r \leq s$: последовательности $(q_0(\boldsymbol{\alpha}), \ldots, q_r(\boldsymbol{\alpha}))$ — неубывающая, $(q_s(\boldsymbol{\alpha}), \ldots, q_n(\boldsymbol{\alpha}))$ — невозрастающая, $q_r(\boldsymbol{\alpha}) = \ldots = q_s(\boldsymbol{\alpha})$.

Свойство 1) означает, что в случае вырожденных (неслучайных) оценок распределение $\mathbf{Q}(\boldsymbol{\alpha}) = (q_0(\boldsymbol{\alpha}), \ldots, q_n(\boldsymbol{\alpha}))$ также становится вырожденным и $q_k(\boldsymbol{\alpha}) = 1$ для индекса k, равного мощности множества (неслучайных) оценок.

Из свойства 3) следует, что средние значения случайных величин $C_S\left(\widetilde{\mathbf{x}}\right)$ будут равны $E\left[C_S\left(\widetilde{\mathbf{x}}\right)\right] = \sum_{i=1}^{n} p_{iS}$, $S \in \{L, M, H\}$.

Лексикографическое ранжирование случайных мощностей оценок

Для применения лексикографического правила ранжирования множества векторных вероятностных альтернатив $\{\widetilde{\mathbf{x}}\}$ относительно функции вероятностной мощности $\mathbf{C}(\widetilde{\mathbf{x}}) = (C_L(\widetilde{\mathbf{x}}), C_M(\widetilde{\mathbf{x}}), C_H(\widetilde{\mathbf{x}}))$ необходимо использовать некоторое правило сравнения вероятностных распределений [Лепский 2017].

Например, если использовать сравнение по среднему значению $E[\cdot]$, то

$$\varphi(\widetilde{\mathbf{x}}) > \varphi(\widetilde{\mathbf{y}}) \Leftrightarrow E[C_L(\widetilde{\mathbf{x}})] < E[C_L(\widetilde{\mathbf{y}})]$$

$$\text{или}\;\; E[C_L(\widetilde{\mathbf{x}})] = E[C_L(\widetilde{\mathbf{y}})],\;\; E[C_M(\widetilde{\mathbf{x}})] < E[C_M(\widetilde{\mathbf{y}})]$$

$$\text{или}\;\; E[C_L(\widetilde{\mathbf{x}})] = E[C_L(\widetilde{\mathbf{y}})],\;\; E[C_M(\widetilde{\mathbf{x}})] = E[C_M(\widetilde{\mathbf{y}})],$$

$$E[C_H(\widetilde{\mathbf{x}})] < E[C_H(\widetilde{\mathbf{y}})].$$

Численный пример

Пусть заданы 5 вероятностных оценок двух альтернатив: векторы $\widetilde{\mathbf{x}} = (\widetilde{x_1}, \ldots, \widetilde{x_5})$ и $\widetilde{\mathbf{y}} = (\widetilde{y_1}, \ldots, \widetilde{y_5})$.

$\widetilde{\mathbf{x}}$	$\widetilde{\mathbf{y}}$
$\widetilde{x_1} = (0.3,\ 0.6,\ 0.1)$	$\widetilde{y_1} = (0.2,\ 0.3,\ 0.5)$
$\widetilde{x_2} = (0.4,\ 0.5,\ 0.1)$	$\widetilde{y_2} = (0.2,\ 0.7,\ 0.1)$
$\widetilde{x_3} = (0.6,\ 0.2,\ 0.2)$	$\widetilde{y_3} = (0.8,\ 0.2,\ 0)$
$\widetilde{x_4} = (0.7,\ 0.2,\ 0.1)$	$\widetilde{y_4} = (0.1,\ 0.3,\ 0.6)$
$\widetilde{x_5} = (0.3,\ 0.3,\ 0.4)$	$\widetilde{y_5} = (0.1,\ 0.9,\ 0)$

Результаты вычислений распределений $\mathbf{C}(\widetilde{\mathbf{x}}) = (C_L(\widetilde{\mathbf{x}}), C_M(\widetilde{\mathbf{x}}), C_H(\widetilde{\mathbf{x}}))$ и $\mathbf{C}(\widetilde{\mathbf{y}}) = (C_L(\widetilde{\mathbf{y}}), C_M(\widetilde{\mathbf{y}}), C_H(\widetilde{\mathbf{y}}))$ по формуле (1) и их математические ожидания:

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$E[\cdot]$
$P\{C_L(\widetilde{\mathbf{x}})=k\}$	0.03	0.19	0.36	0.3	0.11	0.01	2.3
$P\{C_M(\widetilde{\mathbf{x}})=k\}$	0.09	0.31	0.37	0.19	0.04	0	1.8
$P\{C_H(\widetilde{\mathbf{x}})=k\}$	0.35	0.44	0.18	0.03	0	0	0.9
$P\{C_L(\widetilde{\mathbf{y}})=k\}$	0.1	0.49	0.32	0.08	0.01	0	1.4
$P\{C_M(\widetilde{\mathbf{y}})=k\}$	0.01	0.15	0.4	0.33	0.1	0.01	2.4
$P\{C_H(\widetilde{\mathbf{y}})=k\}$	0.18	0.47	0.32	0.03	0	0	1.2

Тогда, применяя сравнение по среднему значению, получим, что

$$\varphi\left(\widetilde{\mathbf{x}}\right) < \varphi\left(\widetilde{\mathbf{y}}\right),$$

так как

$$E\left[C_L\left(\widetilde{\mathbf{x}}\right)\right] > E\left[C_L\left(\widetilde{\mathbf{y}}\right)\right].$$

Заключение

1. рассмотрена задача обобщения порогового правила агрегирования трехградационных ранжировок на случай, когда оценки альтернатив по ряду критериев заданы в виде дискретных вероятностных распределений;
2. построена случайная мощность множества вероятностных оценок заданного ранга;
3. исследованы некоторые свойства случайной мощности.

Список источников

- Алескеров Ф.Т., Якуба В.И. Об одном методе агрегирования ранжировок специального вида // Тез. докл. II Междунар. конф. по пробл. упр. М.: Институт проблем управления им. В.А. Трапезникова РАН, 2003. С. 116.
- Алескеров Ф.Т., Юзбашев Д.А., Якуба В.И. Пороговое агрегирование трехградационных ранжировок // Автоматика и телемеханика. 2007. № 1. С. 147–152.
- Лепский А.Е. Стохастическое и нечеткое упорядочивание методом минимальных преобразований // Автоматика и Телемеханика. 2017. № 1. С. 59–79.
- Подиновский В.В. Многокритериальные задачи принятия решений: теория и методы анализа. М.: Юрайт, 2023.
- Aleskerov F., Chistyakov V., Kalyagin V. Social threshold aggregations // Social Choice Welfare. 2010. Vol. 35. P. 627–646.
- Lepskiy, A. Fuzzy Threshold Aggregation // In: Kahraman C. et al. (eds), Intelligent and Fuzzy Systems. INFUS 2023. LNNS. 2023. Cham: Springer. Vol. 758. P. 69–76.
- Lepskiy, A. Threshold Aggregation of Fuzzy Data Using Fuzzy Cardinalities of a Set of Fuzzy Estimates // In: Kahraman C. et al. (eds), Intelligent and Fuzzy Systems. INFUS 2024. LNNS. 2024. Cham: Springer. (in print)

Спасибо за внимание

alex.lepskiy@gmail.com

alepskiy@hse.ru

https://www.hse.ru/en/org/persons/10586209