УДК 004.33(035)

Informatics and Mathematical Methods in Simulation
Vol. 7 (2017), No. 4, pp. 307-317

# АНАЛИЗ МЕДИАННЫХ МЕТОДОВ КОНСЕНСУСНОГО АГРЕГИРОВАНИЯ РАНГОВЫХ ПРЕДПОЧТЕНИЙ

**В.А. Болтенков, В.И. Куваева, А.В. Позняк**

Одесский национальный политехнический университет,
просп. Шевченко, 1, Одесса, 65044, Украина; e-mail: vaboltenkov@gmail.com

В работе исследованы медианные методы консенсусного агрегирования многоагентных индивидуальных предпочтений, выполненных в ранговых шкалах. Рассмотрены три медианных ранжирования: медиана Кука-Сейфорда в пространстве позиций альтернатив, медиана Литвака в пространстве векторов предпочтений, медиана Кемени в пространстве парных сравнений. Поскольку все медианные ранжирования являются NP-полными задачами, проведена оценка вычислительной сложности различных вариантов приближенных алгоритмов медианного агрегирования. Введено понятие задач агрегирования большой размерности с числом альтернатив более пяти. Установлено, что для задач рангового агрегирования большой размерности точное решение при минимальных вычислительных затратах обеспечивает эвристический алгоритм построения медианы Кемени, предложенный Б.Г. Литваком. Эвристический алгоритм Литвака реализован в электронных таблицах Microsoft Excel для практической задачи построения системы поддержки принятия решений при выборе абитуриентом ИТ-специальности для обучения в университете. Полученные результаты позволяют рекомендовать эвристический алгоритм Литвака для вычисления медианного консенсуса по Кемени для практического применения в задачах агрегирования рангов большой размерности.

**Ключевые слова:** агрегирование ранговых предпочтений, ранговые шкалы, консенсусное ранжирование, медианные алгоритмы, системы поддержки принятия решений

## Введение

Задача агрегирования ранговых предпочтений берет свое начало в теории социального выбора, первые формулировки которой относятся к концу XVIII века [1]. В последние 30 лет она привлекает внимание исследователей как полезный инструмент для коллективного экспертного оценивания, метапоиска, выбора подобных объектов и других прикладных информационных технологий. Задача агрегирования рангов заключается в объединении индивидуальных предпочтений, построенных в ранговой шкале для некоторого множества альтернатив несколькими независимыми агентами (экспертами или другими информационными объектами, например, поисковыми машинами), в единое коллективное ранжирование в той же шкале, которое принято называть «консенсусом» или «консенсусным ранжированием» [1-2]. Имея консенсусное ранжирование, далее можно отсортировать индивидуальные ранжирования по мере их близости к консенсусному. Особое место в консенсусных ранжированиях занимают ранжирования, основанные на медианных принципах. Эти ранжирования являются аксиоматически обоснованными и математически строгими. Однако в литературе отсутствует систематическое, формализованное и содержательное изложение медианных принципов ранжирования и оценка границ их применимости в задачах с большим числом альтернатив.

*Целью* работы является систематизация и анализ методов консенсусного агрегирования ранговых предпочтений, основанных на медианных принципах, оценка их вычислительной сложности и применимости в задачах с большой размерностью множества альтернатив.

**Основная часть**

Все современные методы поиска рангового консенсуса берут свое начало в теории «справедливого» социального выбора, относящегося к работам Борда и Кондорсе [3]. Все методы агрегирования рангов можно разделить на две категории. Первая категория – так называемые позиционные методы, основанные на работе Борда [4]. Консенсусное ранжирование в них определяется с учетом позиций, которые занимают альтернативы в индивидуальных ранжированиях агентов. Простейшим из них является «счет Борда», согласно которому альтернативы в групповом ранжировании располагаются в соответствии с суммой их мест, набранных в индивидуальных ранжированиях. Вторая категория – методы, основанные на парных сравнениях. Эти методы основаны на интуитивной концепции Кондорсе [5], согласно которой альтернатива выигрывает (победитель по Кондорсе), если большинство экспертов считают ее предпочтительной в парных сравнениях со всеми остальными. Обе категории методов консенсусного агрегирования имеют свои преимущества и недостатки и достаточно подробно изучены в работе [6].

Формализуем задачу рангового агрегирования согласно [3]. Пусть существует множество альтернатив $A=\{A_1, A_2,..., A_n\}$, подлежащих упорядочению группой из $K$ агентов по какому-либо критерию (или ряду критериев). Каждый из агентов $k$, $k=\overline{1,K}$ упорядочивает альтернативы и представляет индивидуальное ранжирование

$$P^k=\{A_{k_1}, A_{k_2},..., A_{k_n}\}. \qquad (1)$$

Предполагается, что каждый из агентов может установить, как строгий порядок на множестве альтернатив, так и слабый порядок, т.е. ввести одинаковые ранги для «неразличимых» альтернатив. Например, для множества $A=\{a_1, a_2, a_3, a_4, a_5\}$ может быть введено ранжирование $< a_1 \succ a_2 \succ a_3 \sim a_4 \succ a_5 >$, где знак $\succ$ означает «предпочтительнее, чем … », знак $\sim$ – «равнозначно с … ».

Альтернативы с одинаковыми рангами в русскоязычной литературе называют альтернативами со «связанными рангами», в англоязычной – ties (т.е. ничьими). Каждое индивидуальное ранжирование (1) может быть представлено в виде $P^k=\{q_1^k, q_2^k,..., q_n^k\}$, где $q_i^k$ – позиция, занимаемая $i$-й альтернативой в ранжировании $k$-го агента.

Групповое консенсусное ранжирование, основанное на концепции расстояния, заключается в определении ранжирования $P^k$, ближайшего по некоторой введенной мере ко всем индивидуальным ранжированиям, т.е.

$$\arg\min_P \sum_{k=1}^{K} d(P^k, P) \to \hat{P}. \qquad (2)$$

Решение задачи (2) называется медианным консенсусным ранжированием или в соответствующем контексте просто медианой. Задача (2) может быть решена как дискретная целочисленная задача оптимизации полным перебором, однако показано [3], что объем перебираемых вариантов стремительно растет с ростом мощности множества альтернатив. Так, для множества из 5 альтернатив со связанными рангами требуется перебор 541 варианта, а для множества из 10 альтернатив – уже 102247563 вариантов.

Рассмотрим детально существующие медианные консенсусные ранжирования.

*Медиана Кука-Сейфорда.* Расстояние между двумя ранжированиями по Куку-Сейфорду определяется в пространстве позиций, занимаемых альтернативами:

$$d(P^{k_1}, P^{k_2}) = \sum_{i=1}^{n} \left| q_i^{k_1} - q_i^{k_2} \right|, \tag{3}$$

где $q_i^k$ – позиция, занимаемая $i$-й альтернативой в ранжировании $k$-го агента.

В [7] показано, что расстояние (3) удовлетворяет системе аксиом меры близости, введенных в этой работе. Расстояние ранжирующего вектора $P$ от всего множества ранжирований агентов $\{P^k\}$ равно $d(P, P^{(k)}) = \sum_{k=1}^{K} d(P, P^k) = \sum_{k=1}^{K} \sum_{i=1}^{n} \left| q_i^k - q_i \right|$, где $q_i$ – позиция альтернативы $A_i$ в ранжировании $P$.

Задача оптимизации по Куку-Сейфорду формулируется в виде: найти такое ранжирование $\hat{P}$, для которого

$$d(\hat{P}, P^{(k)}) = \arg\min_{P} \sum_{k=1}^{K} d(P^k, P) = \arg\min_{P} \sum_{k=1}^{K} \sum_{i=1}^{n} \left| q_i^k - q_i \right|. \tag{4}$$

Если предположить, что альтернатива $A_i$ может занимать $j$-ю позицию в ранжировании $P$ $(j = 1,2,...,n)$, расстояние (4) можно записать в виде

$$d(\hat{P}, P^{(k)}) = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} y_{ij}, \tag{5}$$

где $d_{ij} = \sum_{k=1}^{K} \left| d_i^k - j \right|$, $y_{ij} = \begin{cases} 1, \text{ если } A_i \text{ занимает позицию } j \text{ в ранжировании } P, \\ 0, \text{ в противном случае.} \end{cases}$

Задача минимизации (5) может быть представлена как хорошо известная в дискретном программировании задача о назначениях [8]:

$$\arg\min_{y_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} y_{ij},$$

с ограничениями

$$\forall_{i=1,...,n} \sum_{j=1}^{n} y_{ij} = 1, \tag{6}$$

(одна альтернатива может занимать только одну позицию),

$$\forall_{i=1,...,n} \sum_{i=1}^{n} y_{ij} = 1, \tag{7}$$

(в одной позиции может находиться только одна альтернатива).

*Медиана Литвака.* Б.Г. Литвак предложил еще один способ построения консенсусного агрегированного ранжирования [9], который в англоязычной литературе получил наименование медианы Литвака [10]. Подход Литвака основан на так называемом векторе предпочтений. Для ранжирования $k$-го агента вводится вектор предпочтений $\pi^k = [\pi_1^k, \pi_2^k, \ldots, \pi_n^k]$, где $\pi_i^k$ – число альтернатив, предшествующих *i*-й альтернативе в ранжировании. По определению Литвака для двух векторов предпочтений $\pi^{k_1}$ и $\pi^{k_2}$, сформированных ранжирований $P^{k_1}$ и $P^{k_2}$ соответственно, расстояние между ними определяется как $d(P^{k_1}, P^{k_2}) = \sum_{i=1}^{n} \left|\pi_i^{k_1} - \pi_i^{k_2}\right|$. В [11] доказано, что введенное расстояние удовлетворяет введенным аксиомам расстояния близости.

По определению Литвака расстояние от ранжирования $P$ на полном множестве индивидуальных ранжирований агентов $\{P^k\}$ определяется следующим образом:

$$d(P, P^{(k)}) = \sum_{k=1}^{K} \sum_{i=1}^{n} \left|\pi_i^P - \pi_i^k\right|.$$

Введем $h_i^{k(j)} = \left|\pi_i^{P(j)} - \pi_i^k\right|$, где $i = \overline{1,n}$; $j = \overline{1,n}$; $k = \overline{1,K}$. При таких обозначениях медиана Литвака есть результат задачи оптимизации:

$$\arg\min_{y_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i^{(j)} y_{ij} ,$$

с ограничениями (6)-(7).

*Медиана Кемени.* Расстояние между ранжированиями для медианы Кемени [11] определено с применением матрицы парных сравнений. Для заданного ранжирования *n* альтернатив *k*-м агентом ($k = \overline{1,K}$) $P^k : A_{k_1}, A_{k_2}, \ldots, A_{k_n}$ может быть следующим образом построена матрица парных сравнений:

$$C^k = \begin{bmatrix} c_{11}^k, \ldots, c_{1n}^k \\ \ldots \;\; \ldots \;\; \ldots \\ c_{n1}^k, \ldots, c_{nn}^k \end{bmatrix}, \tag{8}$$

$$c_{il}^k = \begin{cases} 1, \text{ для } A_i \succ A_l \\ 0, \text{ для } A_i \sim A_l \\ -1, \text{ для } A_i \prec A_l \end{cases},$$

где $c_{il} = 0,\ i = \overline{1,n};\ l = \overline{1,n}$.

Для двух заданных ранжирований $P^{k_1}$ и $P^{k_2}$ расстояние между ними равно

$$d(P^{k_1}, P^{k_2}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{l=1}^{n} \left|c_{il}^{k_1} - c_{il}^{k_2}\right|.$$

Расстояние между заданным ранжированием $P$ и набором ранжирований, сформированных агентами, определяется как

$$d = d(P, P^{(k)}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{l=1}^{n} \sum_{k=1}^{K} \left|c_{il}^P - c_{il}^k\right|.$$

Предположим, что в ранжировании $P$ $A_i \succ A_1$, т.е. $c_{il}^P = 1$. Введем коэффициенты $r_{il}$ следующим образом

$$r_{il} = \sum_{k=1}^{K} d_{ij}(P, P^{(k)}) = \sum_{k=1}^{K} \left| c_{il}^k - c_{il}^P \right| = \sum_{k=1}^{K} \left| c_{il}^k - 1 \right|.$$

Они называются коэффициентами потерь, а матрица $R = [r_{il}]$ – матрицей потерь. Предполагается, что $r_{il} = 0$ для всех $i = l$.

Можно показать [11], что

$$d = \sum_{\substack{i=1 \\ (i,j)\in I_p^{(1)}}}^{n} \sum_{j=1}^{n} r_{ij} + \sum_{\substack{i=1 \\ (i,j)\in I_p^{(2)}}}^{n} \sum_{j=1}^{n} \sum_{k=1}^{K} \left| c_{ij}^k \right|,$$

где $I_p^{(1)}$ – множество индексов $(i, j)$, для которых $A_i \succ A_j$ в ранжировании $P$ (или множество индексов, для которых $c_{ij}^P = 1$), $I_p^{(2)}$ – множество индексов $(i, j)$, для которых $A_i \sim A_j$ в ранжировании $P$ (или множество индексов, для которых $c_{ij}^P = 0$). Ранжирование $P^M$, такое, что $P^M (P_1, P_2, ..., P^K) = \arg\min_P d(P, P^K)$, называется медианой Кемени.

Таким образом, известны три консенсусных медианы, которые представляют собой решение задачи поиска ранжирования, минимально удаленного от индивидуальных ранжирований агентов в соответствующих пространствах:

– медиана Кука-Сейфорда в пространстве позиций альтернатив;

– медиана Литвака в пространстве векторов предпочтений;

– медиана Кемени в пространстве парных сравнений.

Остановимся на практически важном вопросе вычислительной сложности построения консенсусного рангового агрегирования на основе медианных методов. Доказано, что вычисление всех медианных агрегированных ранжирований есть NP-полные задачи при $n > 5$ [12,13]. Далее будем называть задачи поиска консенсусного медианного ранжирования с $n > 5$ задачами большой размерности. Для достижения возможности решения задачи консенсусного агрегирования за реальное время в таких задачах был предложен ряд алгоритмов, главным образом основанных на решении задачи целочисленного программирования, к которым приводятся задачи рангового агрегирования. В частности, в [14] доказана теорема о том, что задача отыскания медианы Кемени на множестве индивидуальных предпочтений эквивалентна задаче о назначениях. Показано, что решение задачи консенсусного медианного ранжирования может быть получено венгерским методом или как частный случай транспортной задачи с бинарными переменными с вычислительной сложностью $O(n^3)$. Рассмотрен ряд алгоритмов, позволяющих решить задачу медианного консенсусного ранжирования методом ветвей и границ путем приведения вычислительной сложности к полиномиальной [15]. В [16] отмечено, что всего известно 104 алгоритма вычисления, однако указано, что все они нарушают исходную медианную аксиоматику введением тех или иных предположений и тем самым уже являются приближенными или эвристическими алгоритмами поиска консенсусного медианного ранжирования. В процессе исследования проведен анализ более 40 методов, из них наиболее простым представляется приближенный метод вычисления медианы Кемени, предложенный Литваком [9] и сочетающий в себе точность, отсутствие жестких требований к производительности вычислительных ресурсов и главное – алгоритмическую логичность и простоту.

Эвристический метод расчета медианного консенсуса Кемени по Литваку в виде пошагового алгоритма выглядит следующим образом:

***Шаг 1.*** Построить матрицу потерь $R$, где $r_{ij} = d_{ij}(P_1, P_x) + d_{ij}(P_2, P_x) + ... + d_{ij}(P_n, P_x)$; $P_x$ – ранжирование, у которого $a_{ij}^x = a_{ij\max}^x$, т.е. $a_{ij}^x = 1$. Если в матрицах всех ранжирований стоит 1 или -1, то соответствующий элемент матрицы потерь будет равен 0.

***Шаг 2.*** Определить матрицу $Q^n = E^n R^n E^n$, где

$$E^n = \begin{vmatrix} 0\,1 \cdots 1 \\ 1\,0 \cdots 1 \\ \vdots\ \vdots \ddots \vdots \\ 1\,1 \cdots 0 \end{vmatrix}.$$

***Шаг 3.*** Определить наименьший элемент, исключением элементов из главной диагонали матрицы $Q^n - n = q_{ij\min}^n$.

Определить значения: $v^n = 2K\dfrac{(n-2)(n-3)}{2}$; $d_{ji\min}^{(n)} = d_{ji\min}^{(n)} - v^n$.

Альтернативы, определяющие строку и столбец элемента $q_{ij\min}^n$, обозначаются как $A_{i_n}$ и $A_{i_1}$ соответственно.

***Шаг 4.*** Строки и столбцы, относящиеся к альтернативам $A_{i_n}$ и $A_{i_1}$, удалить из матрицы $R^n$. Результат – новая матрица $R^{n-2}$.

***Шаг 5.*** Определить матрицу $Q^{n-2}$: $Q^{n-2n} = E^{n-2} R^{n-2} E^{n-2}$.

***Шаг 6.*** Определить наименьший элемент исключением элементов из главной диагонали матрицы $Q^{n-2} - n = q_{ij\min}^{n-1}$. Определить значения: $v^n = 2K\dfrac{(n-4)(n-5)}{2}$; $d_{ji\min}^{(n-2)} = d_{ji\min}^{(n-2)} - v^{(n-2)}$.

***Шаг 7.*** Строки и столбцы, относящиеся к альтернативам $A_{i_n}$ и $A_{i_1}$, удалить из матрицы $R^{n-2}$.

Шаги 1 - 4 повторяются, до тех пор, пока не будет получено конечное ранжирование по условию $n-2=0$ или $n-2=1$.

*Пример практической реализации поиска консенсусного ранжирования в задаче большой размерности.* Для разработки серверной части системы поддержки принятия решений (СППР) абитуриентами, поступающими на ИТ-специальности института компьютерных систем Одесского национального политехнического университета, была проведена коллективная экспертная оценка качеств, знаний и умений (далее – необходимых качеств) необходимых абитуриенту для успешного обучения на соответствующей специальности. Клиентская часть выполнена в виде Android-приложения и содержит вопросы для абитуриента, ответы на которые позволяют оценить у него наличие тех или иных качеств, знаний и навыков. В результате функционирования СППР подсказывает абитуриенту рекомендуемую для него ИТ-специальность [17].

Для оценки необходимых качеств для пяти специальностей: 113 – «Прикладная математика», 121 – «Инженерия программного обеспечения», 122 – «Компьютерные науки», 123 – «Компьютерная инженерия», 126 – «Информационные системы и технологии», 151 – «Автоматизация и компьютерно-интегрированные технологии», были сформированы шесть коллективов агентов – экспертов численностью соответственно 8, 15 16, 12, 16, 10 человек из ведущих преподавателей выпускающих кафедр. Каждому из экспертов был предложен список альтернатив, состоящий из 30 требуемых качеств (ТКi), приведенный в таблице 1.

**Таблица 1.**

Список альтернатив (знаний, умений, личных качеств абитуриента)

| № | Качества абитуриента | Обозначение |
|---|---|---|
| 1 | Хорошая школьная подготовка по математике | ТК1 |
| 2 | Хорошая школьная подготовка по физике | ТК2 |
| 3 | Хорошая школьная подготовка по информатике | ТК3 |
| 4 | Успешное участие в олимпиадах по математике | ТК4 |
| 5 | Успешное участие в олимпиадах по физике | ТК5 |
| 6 | Успешное участие в олимпиадах по информатике | ТК6 |
| 7 | Успешное участие в работе МАН | ТК7 |
| 8 | Опыт разработки программного продукта | ТК8 |
| 9 | Практическое владение английским языком | ТК9 |
| 10 | Практическое владение другим иностранным языком | ТК10 |
| 11 | Практическое владение одним из языков программирования | ТК11 |
| 12 | Опыт работы с приложениями под ОС Windows | ТК12 |
| 13 | Опыт работы с приложениями под ОС Android | ТК13 |
| 14 | Опыт работы с приложениями под ОС iOS | ТК14 |
| 15 | Практические знания аппаратной части компьютера | ТК15 |
| 16 | Практические умения по установке ОС на компьютер | ТК16 |
| 17 | Практические умения по настройке домашнего Wi-Fi роутера | ТК17 |
| 18 | Умение работать в команде | ТК18 |
| 19 | Умение быстро найти необходимую информацию в интернете | ТК19 |
| 20 | Умение руководить небольшой группой | ТК20 |
| 21 | Умение решать нестандартные задачи | ТК21 |
| 22 | Умение самостоятельно разобраться в незнакомой теме | ТК22 |
| 23 | Умение убедительно объяснить свою точку зрения | ТК23 |
| 24 | Стрессоустойчивость | ТК24 |
| 25 | Усидчивость | ТК25 |
| 26 | Дисциплинированность | ТК26 |
| 27 | Широкий кругозор | ТК27 |
| 28 | Стремление овладеть новыми знаниями и навыками | ТК28 |
| 29 | Ораторское искусство | ТК29 |
| 30 | Умение слушать | ТК30 |

Для формирования консенсусного агрегирования результатов коллективной экспертизы по специальности 151 – «Автоматизация и компьютерно-интегрированные технологии», в которой участвовали 10 экспертов, применим изложенный выше эвристический алгоритм Литвака для приближенного расчета медианы Кемени.

Проведем предварительно ряд процедур предварительной обработки ранжирований [18]:

1. Стандартизация рангов: индивидуальные предпочтения, содержащие связанные ранги, обрабатываются с помощью процедуры стандартизации рангов для приведения ранговых оценок экспертов к сопоставимому виду и приводятся к измерению в одной и той же шкале для всех экспертов;

2. Оценка коэффициента конкордации $W$: перед формированием групповой оценки необходимо выяснить, насколько непротиворечивы индивидуальные экспертные ранжирования и можно ли их агрегировать в расчете на состоятельную консенсусную оценку:

$$W = \frac{S}{m^2(n^3-n)/12 - m\sum_{j=1}^{m} T_j},$$

где $S$ – сумма квадратов разностей между членами суммарного ранжирования и членами ряда, составленного из средних значений; $m$ – число экспертов; $n$ – число ранжируемых объектов;

$$T_j = \frac{1}{12}\sum_{i=1}^{n}(t_{ij}^3 - t_{ij}),$$

где $t_{ij}$ – число повторений ранга $t$ в $j$-том ряду.

Выполнена оценка статистической значимости коэффициента конкордации по критерию $\chi^2$. Показано, что оценке $W = 0{,}64$ при уровне значимости 0,95 соответствует $\chi^2 = 220.12$, поскольку $\chi^2 > \chi^2_{табл}$, $\chi^2_{табл}$ – табличное значение распределения $\chi^2$ для $(n-2)$ степеней свободы, гипотеза о согласованности мнений всей группы экспертов принимается.

3. Выделение значимых альтернатив: для выявления значимости каждой альтернативы определяются коэффициенты значимости ($K_j$) по формуле $K_j = \frac{mn - S_j}{0{,}5mn(n-1)}$, где $S_j$ – сумма рангов для каждой альтернативы.

Для выделения из $n$ альтернатив наиболее значимых, введем порог значимости коэффициентов $K_i \geq 1/n$ и выделим наиболее значимые альтернативы. В результате вычислений выявлено 10 наиболее значимых альтернатив: ТК1, ТК2, ТК3, ТК28, ТК22, ТК25, ТК23, ТК26, ТК27, ТК11, приведенных в таблице 1. Стандартизованные ранги для них представлены в таблице 2.

**Таблица 2.**

Стандартизованные ранговые оценки для значимых альтернатив

| Эксперты | ТК1 | ТК2 | ТК3 | ТК 28 | ТК 22 | ТК 25 | ТК 23 | ТК 26 | ТК 27 | ТК 11 |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Э1 | 33,0 | 33,0 | 33,0 | 21,0 | 33,0 | 21,0 | 28,5 | 21,0 | 28,5 | 21,0 |
| Э2 | 34,0 | 34,0 | 30,0 | 24,5 | 13,0 | 30,0 | 24,5 | 34,0 | 24,5 | 24,5 |
| Э3 | 34,5 | 34,5 | 32,0 | 26,5 | 15,5 | 26,5 | 26,5 | 26,5 | 21,5 | 26,5 |
| Э4 | 34,5 | 30,0 | 23,5 | 32,5 | 34,5 | 23,5 | 28,0 | 23,5 | 23,5 | 14,5 |
| Э5 | 33,0 | 33,0 | 33,0 | 33,0 | 28,0 | 28,0 | 19,0 | 28,0 | 33,0 | 19,0 |
| Э6 | 33,0 | 33,0 | 33,0 | 33,0 | 26,5 | 26,5 | 26,5 | 26,5 | 33,0 | 15,5 |
| Э7 | 33,0 | 33,0 | 33,0 | 33,0 | 27,0 | 33,0 | 30,0 | 27,0 | 16,5 | 27,0 |
| Э8 | 33,5 | 33,5 | 26,5 | 33,5 | 26,5 | 33,5 | 26,5 | 26,5 | 26,5 | 26,5 |
| Э9 | 29,5 | 29,5 | 29,5 | 29,5 | 29,5 | 21,0 | 29,5 | 21,0 | 16,0 | 29,5 |
| Э10 | 33,5 | 21,5 | 33,5 | 21,5 | 33,5 | 21,5 | 21,5 | 21,5 | 21,5 | 30,5 |

Каждую индивидуальную систему предпочтений надо преобразовать в матрицу парных сравнений, элементы которой вычисляются по соотношению (8).

Далее из матриц парных сравнений рассчитываем элементы матрицы потерь R (табл. 3), как указано на шаге 1 эвристического метода Литвака.

**Таблица 3.**

Матрица потерь *R* для построения медианы Кемени

| | | ТК1 | ТК2 | ТК3 | ТК 28 | ТК 22 | ТК 25 | ТК 23 | ТК 26 | ТК 27 | ТК 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| ТК1 | 1 | 0 | 8 | 6 | 5 | 4 | 2 | 1 | 1 | 2 | 1 |
| ТК2 | 2 | 12 | 0 | 7 | 8 | 6 | 3 | 2 | 2 | 3 | 3 |
| ТК3 | 3 | 14 | 13 | 0 | 8 | 6 | 5 | 4 | 4 | 4 | 2 |
| ТК28 | 4 | 15 | 12 | 12 | 0 | 7 | 7 | 6 | 5 | 6 | 6 |
| ТК22 | 5 | 16 | 14 | 14 | 13 | 0 | 10 | 9 | 8 | 9 | 7 |
| ТК25 | 6 | 18 | 17 | 15 | 13 | 10 | 0 | 9 | 9 | 8 | 6 |
| ТК23 | 7 | 19 | 18 | 16 | 14 | 11 | 11 | 0 | 8 | 8 | 7 |
| ТК26 | 8 | 19 | 18 | 16 | 15 | 12 | 11 | 12 | 0 | 9 | 8 |
| ТК27 | 9 | 18 | 17 | 16 | 14 | 11 | 12 | 12 | 11 | 0 | 8 |
| ТК11 | 10 | 19 | 17 | 18 | 14 | 13 | 14 | 13 | 12 | 10 | 0 |

Матрица потерь обрабатывалась по алгоритму Литвака в системе Microsoft Excel. В результате получено итоговое консенсусное ранжирование альтернатив:

$$\text{ТК1} \succ \text{ТК2} \succ \text{ТК3} \succ \text{ТК28} \succ \text{ТК25} \succ \text{ТК22} \succ \text{ТК23} \succ \text{ТК26} \succ \text{ТК27} \succ \text{ТК11}.$$

Правильность расчетов была проверена с использованием комплекса Matlab-программ, основанных на решении задачи поиска медианы Кемени как задачи целочисленного программирования с использованием метода ветвей и границ [19]. Результаты расчета полностью совпадают, при этом при расчете с помощью Matlab-програм затраты процессорного времени примерно в 1,8-2 раза больше, чем при расчетах в Microsoft Excel. Для сравнения тот же набор индивидуальных ранжирований был обработан по методу Борда [6]. Затраты процессорного времени при этом сопоставимы с расчетами в Microsoft Excel. Полученное агрегированное консенсусное ранжирование выглядит так:

$$\text{ТК1} \succ \text{ТК2} \succ \text{ТК3} \succ \text{ТК28} \succ \text{ТК22} \succ \text{ТК25} \succ \text{ТК23} \succ \text{ТК26} \succ \text{ТК27} \succ \text{ТК11}.$$

Полученный результат показывает, что медиана Кемени, рассчитанная по эвристическому приближенному алгоритму Литвака, позволяет получить более точное консенсусное агрегирование – видно, что счет Борда не позволяет различить превосходство альтернативы 6 над альтернативой 5.

## Выводы

В работе исследованы медианные методы построения консенсусных предпочтений в ранговых шкалах. Систематизированы данные по медианам Кука-Сейфорда, Литвака и Кемени. Поскольку все медианные методы являются NP-полными задачами, проведен анализ приближенных методов оценивания консенсусных ранжирований в задачах большой размерности. Установлено, что приближенный алгоритм расчета медианы Кемени, предложенный Литваком, позволяет получить достаточно точное консенсусное ранжирование с минимальными вычислительными затратами. Это позволяет рекомендовать алгоритм Литвака для расчета рангового консенсуса в многоагентных задачах большой размерности.

## Список литературы


1. Elkind, E. Distance rationalization of voting rules / E. Elkind, P. Faliszewski, A. Slinko. // Social Choice and Welfare. — 2015. — Vol. 45, No. 2. — Pp. 345–377.
2. D'Ambrosio, A. Kemeny's axiomatic approach to end consensus ranking in tourist satisfaction / A. D'Ambrosio, V.A. Tutore. // Statistica Applicata. — 2008. — Vol. 20, No. 1. — Pp. 21–32.
3. Bury, H. Group Judgement With Ties. Distance-Based Methods / H. Bury, D. Vagner. // In: H. Aschemann (ed.). New Approaches in Automation and Robotics. — Vienna: I-Tech I-Tech Education and Publishing. — 2008. — Pp. 153–172.
4. Borda, J. C. M´emoire sur les ´elections au scrutin [Electronic resource] // Histoire de l'Acad´emie Royale des Sciences. — Paris, 1781. Mode of access: http://gerardgreco.free.fr/IMG/pdf/MA_c_moire-Borda-1781.pdf
5. Condorcet, J. A. N. Essai sur L´Application de L´Analyse a la Probabilite des Decisions Rendues a la Pluralite des Voix [Electronic resource] // A Paris: de l'Imprimerie royale, 1785. — Paris, 1781. Mode of access: http://www.e-rara.ch/zut/content/titleinfo/1175327
6. Петровский, А. Б. Теория принятия решений. / А. Б. Петровский. — М.: Академия, 2009. — 400 с.
7. Cook, W.D. Distance-based and adhoc consensus models in ordinal preference ranking. / W.D. Cook. // European Journal of Operational Research. — 2006. — No. 172. — Pp. 369–385.
8. Писарук, Н. Н. Исследование операций / Н. Н. Писарук. — Минск: БГУ, 2015. — 304 с.
9. Литвак, Б.Г. Экспертная информация: Методы получения и анализа / Б. Г. Литвак. — М.: Радио и связь, 1982. — 184 с.
10. Bury, H. Zastosowanie mediany Litvaka do wyznaczania oceny grupowej w przypadku występowania obiektów równoważnych / H. Bury, D. Wagner. // Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą. — 2007. — Vol. 10. — Pp. 19–34.
11. List, C. Judgement aggregation: a survey / C. List, C. Puppe. // In: Oxford handbook of rational and social choice. — Oxford: Oxford University Press. — 2009. — Pp. 457–482.
12. Hudry, O. Complexity of computing median linear orders and variants / O. Hungry. // Electronic Notes in Discrete Mathematics. — 2013. — Vol. 42. — Pp. 57–64.
13. Hudry, O. NP-hardness results on the aggregation of linear orders into median orders / O. Hungry. // Annals of Operations Research. — 2008. — Vol. 163, No. 1. — Pp. 63–88.
14. Самохвалов, Ю.Я. Экспертное оценивание. Методический аспект / Ю. Я. Самохвалов, Е.М. Науменко. — Киев: ДУИКТ, 2007. — 262 с.
15. Amodio, S. Accurate algorithms for identifying the median ranking when dealing with weak and partial rankings under the Kemeny axiomatic approach // S. Amodio, A. D'Ambrosio, R. Siciliano. // European Journal of Operational Research. — 2015. — No. 249. — Pp. 667–676.
16. Alnur, A. Experiments with Kemeny Ranking: What Works When? / A. Alnur, M. Meila. // Mathematical Social Sciences. — 2012. — Vol. 64, No. 1. — Pp. 28-40.
17. Куваева, В.И. Применение методов экспертного оценивания при построении систем поддержки принятия решений / В.И. Куваева, А.В. Позняк, В.А. Болтенков. // Системи та засоби штучного інтелекту: тези доповідей Міжнародної наукової молодіжної школи. — Київ: ІПШІ «Наука і освіта». — 2017. — С. 104-108.
18. Писарева, О.М. Методы прогнозироваиия развития социально-экономических систем / О.М. Писарева. — М: Высшая школа, 2007. — 591 с.
19. Compute the median ranking according to the Kemeny axiomatic approach // [Electronic resource]: Mode of access: https://www.mathworks.com/matlabcentral/fileexchange/52235-compute-the-median-ranking-according-to-the-kemenyaxiomaticapproach?focused =3889946&tab =function&requestedDomain=true.

## АНАЛІЗ МЕТОДІВ КОНСЕНСУСНОГО АГРЕГУВАННЯ РАНГОВИХ ПЕРЕВАГ

В.О. Больтьонков, В.І. Куваєва, О.В. Позняк

Одеський національний політехнічний університет,
просп. Шевченка, 1, Одеса, 65044, Україна; e-mail: vaboltenkov@gmail.com

У роботі досліджені медіанні методи консенсусного агрегування багатоагентних індивідуальних переваг, виконаних в рангових шкалах. Розглянуто три медіанні ранжирування: медіана Кука-Сейфорда в просторі позицій альтернатив, медіана Литвака в просторі векторів переваг, медіана Кемені в просторі парних порівнянь. Оскільки всі медіанні ранжирування є NP-повними задачами, проведена оцінка обчислювальної складності різних варіантів наближених алгоритмів медіанного агрегування. Введено поняття задач агрегування великої розмірності з числом альтернатив більше п'яти. Встановлено, що для задач рангового агрегування великої розмірності точне рішення, при мінімальних обчислювальних витратах, забезпечує евристичний алгоритм побудови медіани Кемені, запропонований Б.Г. Литваком. Евристичний алгоритм Литвака реалізований в електронних таблицях Microsoft Excel для практичної задачі побудови системи підтримки прийняття рішень при виборі абітурієнтом IT-спеціальності для навчання в університеті. Отримані результати дозволяють рекомендувати евристичний алгоритм Литвака для обчислення медіанного консенсусу за Кемені для практичного застосування в задачах агрегування рангів великої розмірності.

**Ключові слова:** агрегування рангових переваг, рангові шкали, консенсусні ранжирування, медіанні алгоритми, системи підтримки прийняття рішень

## ANALYSIS OF MEDIAN METHODS FOR CONSENSUS RANK PREFERENCES AGGREGATION

V.A. Boltenkov, V.I. Kuvayeva, A.V. Pozniak

Odesa National Polytechnic University,
1, Shevchenko Ave., Odesa, 65044, Ukraine; e-mail: vaboltenkov@gmail.com

The median methods of consensus aggregation of multi-agent individual preferences performed in rank scales has been investigated. Three median rankings were considered: the Cook-Seyford’s median in the space of alternatives positions, the Litvak’s median in the space of preference vectors, the of Kemeny’s median in the space of paired comparisons. Since all median rankings are NP-complete problems, the computational complexity of various variants of approximate median aggregation algorithms has been studied. The notion of large-dimensional aggregation problems with the number of alternatives more than five has been introduced. It is established that for the problems of rank-based aggregation of large dimension, the exact solution with minimal computational costs is provided by the heuristic algorithm for constructing the Kemeny’s median proposed by B.G. Litvak. The heuristic algorithm of Litvak is implemented in Microsoft Excel spreadsheets for the practical task of constructing a decision support system for applicants selecting an IT specialty for studying at a university. The obtained results make it possible to recommend the heuristic algorithm of Litvak for calculating the median consensus according Kemeny for practical application in problems of large dimension ranks aggregation.

**Keywords:** aggregation of rank preferences, rank scales, consensus ranking, median algorithms, decision support systems