УДК 519.2:303.732.4

01.00.00 Физико-математические науки

ЗАДАЧА ИССЛЕДОВАНИЯ ИТОГОВОГО РАНЖИРОВАНИЯ МНЕНИЙ ГРУППЫ ЭКСПЕРТОВ С ПОМОЩЬЮ МЕДИАНЫ КЕМЕНИ

Жуков Михаил Станиславович
аналитик
РИНЦ SPIN-код: 8029-2401
Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5, mikhail.zhukov@yahoo.ca

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н., профессор
РИНЦ SPIN-код: 4342-4994
Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru

В различных прикладных областях возникает необходимость анализа нескольких экспертных упорядочений, т.е. кластеризованных ранжировок объектов экспертизы. К таким областям относятся технические исследования, экология, менеджмент, экономика, социология, прогнозирование и т.д. В качестве объектов могут выступать образцы продукции, технологии, математические модели, проекты, кандидаты на должность и др. При построении итогового мнения комиссии экспертов необходимо найти кластеризованную ранжировку, усредняющую ответы экспертов. В статье описан ряд методов усреднения совокупности кластеризованных ранжировок, среди которых выделяется метод расчета медианы Кемени, основанный на использовании расстояния Кемени. Настоящая статья посвящена вычислительной стороне задачи исследования итогового ранжирования мнений группы экспертов с помощью медианы Кемени. В настоящее время неизвестно ни одного точного алгоритма поиска множества всех медиан Кемени для заданного множества перестановок (ранжировок без связей), кроме полного перебора. *Однако, существуют различные подходы поиска части или всего множества медиан, которые проанализированы в этой работе.* Эвристические алгоритмы Жихарева служат хорошим инструментом для исследования множества всех медиан Кемени: выявления каких-либо закономерностей при изучении взаимного расположения медиан по отношению к экспертной совокупности или экспертному подмножеству

UDC 519.2:303.732.4

Physics and mathematical sciences

THE PROBLEM OF RESEARCH OF FINAL RANKING FOR GROUP OF EXPERTS BY MEANS OF KEMENY MEDIAN

Zhukov Mikhail Stanislavovich
analyst

Bauman Moscow State Technical University, Moscow, Russia

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci., professor
Bauman Moscow State Technical University, Moscow, Russia

In various applications, it is necessary to analyze several expert orderings, i.e. clustered rankings objects of examination. These areas include technical studies, ecology, management, economics, sociology, forecasting, etc. The objects can be some samples of products, technologies, mathematical models, projects, job applicants and others. In the construction of the final opinion of the commission of experts, it is important to find clustered ranking that averages responses of experts. This article describes a number of methods for clustered rankings averaging, among which there is the method of Kemeny median calculation, based on the use of Kemeny distance. This article focuses on the computing side of the final ranking among the expert opinions problem by means of median Kemeny calculation. There are currently no exact algorithms for finding the set of all Kemeny medians for a given number of permutations (rankings without connections), only exhaustive search. However, there are various approaches to search for a part or all medians, which are analyzed in this study. Zhikharev's heuristic algorithms serve as a good tool to study the set of all Kemeny medians: identifying any connections in mutual locations of the medians in relation to the aggregated expert opinions set (a variety of expert answers permutations). Litvak offers one precise and one heuristic approaches to calculate the median among all possible sets of solutions. This article introduces the necessary concepts, analyzes the advantages of median Kemeny among other множества перестановок экспертных ответов. Литвак предлагает один точный и один эвристический подход к вычислению одной медианы среди всего возможного множества решений задачи. В настоящей статье введены необходимые понятия, проанализированы преимущества медианы Кемени среди других возможных поисков экспертного упорядочивания. Выявлены сильные и сравнительно слабые стороны рассматриваемых способов вычисления

possible searches of expert orderings. It identifies the comparative strengths and weaknesses of examined computational ways

Ключевые слова: МАТЕМАТИКА, ЭКОНОМИКА, УПРАВЛЕНИЕ, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, ПРИКЛАДНАЯ СТАТИСТИКА, СТАТИСТИКА НЕЧИСЛОВЫХ ДАННЫХ, ТЕОРИЯ ПРИНЯТИЯ РЕШЕНИЙ, ЭКСПЕРТНЫЕ ОЦЕНКИ, КЛАСТЕРИЗОВАННЫЕ РАНЖИРОВКИ, СРЕДНИЕ ВЕЛИЧИНЫ, МЕДИАНА КЕМЕНИ

Keywords: MATHEMATICS, ECONOMICS, MANAGEMENT, MATHEMATICAL STATISTICS, APPLIED STATISTICS, STATISTICS OF NON-NUMERICAL DATA, DECISION THEORY, EXPERT ESTIMATORS, CLUSTERED RANKINGS, AVERAGE, KEMENY MEDIAN

1. Введение. Постановка задачи

Экспертные процедуры применяются во многих областях деятельности [1 - 3]. К таким областям относятся прежде всего менеджмент (особенно производственный менеджмент), экономика, экология, социология, прогнозирование, технические исследования, и т.д., особенно те их разделы, что связаны с экспертными оценками. В качестве объектов могут выступать образцы продукции, технологии, математические модели, проекты, кандидаты на должность и др.

Во многих экспертных процедурах ответы экспертов - кластеризованные ранжировки объектов [4]. Кластеризованные ранжировки могут быть получены как с помощью экспертов, так и объективным путем, например, при сопоставлении математических моделей с экспериментальными данными и упорядочения этих моделей по точности с помощью того или иного критерия качества [5].

При построении итогового мнения комиссии экспертов необходимо найти кластеризованную ранжировку, усредняющую ответы экспертов. Разработан ряд методов усреднения совокупности кластеризованных ранжировок, среди которых выделяется метод расчета медианы Кемени, основанный на использовании расстояния Кемени [6]. Хотя этот метод имеет большое теоретическое значение, вычислительные вопросы нахождения медианы Кемени требуют дальнейшего изучения.

Настоящая статья посвящена вычислительной стороне задачи исследования итогового ранжирования мнений группы экспертов с помощью медианы Кемени.

2. Понятие кластеризованной ранжировки

Пусть имеется конечное число объектов, которые будем обозначать натуральными числами 1, 2, 3, ..., k и называть «носителем». Под кластеризованной ранжировкой, определенной на заданном носителе, понимаем следующую математическую конструкцию (специальный класс бинарных отношений [7]).

Пусть объекты разбиты на группы, которые будем называть кластерами. В кластере может быть и один элемент. Входящие в один кластер объекты будем заключать в фигурные скобки. Например, объекты 1, 2, 3, ..., 10 могут быть разбиты на 7 кластеров: {1}, {2, 3}, {4}, {5, 6, 7}, {8}, {9}, {10}. В этом разбиении один кластер {5, 6, 7} содержит три элемента, другой - {2, 3} - два, остальные пять - по одному элементу. Кластеры не имеют общих элементов, а объединение их (как множеств) есть все рассматриваемое множество объектов.

Вторая составляющая кластеризованной ранжировки - это строгий линейный порядок между кластерами. Задано, какой из них первый, какой второй, и т.д. Будем изображать строгую упорядоченность с помощью знака <. При этом кластеры, состоящие из одного элемента, будем для простоты изображать без фигурных скобок. Тогда кластеризованную ранжировку на основе введенных выше кластеров можно изобразить так:

$$A = [\ 1 < \{2, 3\} < 4 < \{5, 6, 7\} < 8 < 9 < 10\ ] .$$

Введенная описанным образом кластеризованная ранжировка является бинарным отношением на множестве {1, 2, 3, ..., 10}.

Каждую кластеризованную ранжировку, как и любое бинарное отношение, можно задать матрицей $\| x(a, b) \|$ из 0 и 1 порядка k х k. При этом $x(a, b) = 1$ тогда и только тогда, когда $a < b$ либо $a = b$. В первом случае $x(b, a) = 0$, а во втором $x(b, a) = 1$. При этом хотя бы одно из чисел $x(a, b)$ и $x(b, a)$ равно 1, причем $x(a, b) = x(b, a) = 1$ тогда и только тогда, когда $a = b$.

Введенный математический объект иногда называют по-другому. Он известен в математической статистике как "ранжировка со связями" (см., например, [8, с.48 и др.]). Дж. Кемени и Дж. Снелл используют термин "упорядочение" [6, с.20], Б.Г. Миркин - "квазисерия" [9, с.37], Ю.А. Шрейдер - "совершенный квазипорядок" [10, с.127, 130]. Учитывая разнобой в терминологии, мы сочли полезным применять термин "кластеризованная ранжировка", поскольку в нем явным образом названы основные элементы изучаемого математического объекта - кластеры, рассматриваемые на этапе согласования ранжировок как классы эквивалентности, и ранжировка - строгий совершенный порядок между ними (в терминологии [10, гл. IV]).

3. Итоговое мнение комиссии экспертов

Разработаны разные способы поиска итогового ранжирования мнений группы экспертов, например **Метод средних арифметических рангов,** в котором каждому объекту присваивается числовой ранг каждым экспертом и рассчитывается сумма рангов по каждому проекту. Затем эта сумма делится на число экспертов, в результате получается средний арифметический ранг (именно эта операция дала название методу) По средним рангам строится итоговая ранжировка. Другим примером является **Метод медиан рангов,** где упорядочивание строится по медиане присвоенных рангов, а не по среднему арифметическому. По эти двум ранжировкам с помощью специально разработанного алгоритма строят согласующую ранжировку [2, 4].

Обсудим условный пример. Пусть рассматриваются 20 альтернатив $a_1, ..., a_{20}$. Их ранжирования (упорядочения), соответствующие мнению 10 экспертов P_1, P_2, ..., P_{10}, например, таковы (табл.1).

Таблица 1. Ранжирования экспертов

Ранг альтернативы	Эксперты						
	P_1	P_2	P_3	P_4	P_5	...	P_{10}
1	a_2	a_2	a_1	a_3	a_4	...	a_9
2	a_1	a_1	a_3	a_1	a_1	...	a_1
3	a_3	a_3	a_4	a_4	a_3	...	a_3
4	a_4	a_4	a_5	a_5	a_5	...	a_4
...	...	...	...	...	...	...	...
19	a_{19}	a_{19}	a_{20}	a_{20}	a_{20}	...	a_{20}
20	a_{20}	a_{20}	a_2	a_2	a_2	...	a_2

Рассмотрим задачу определения наилучшей альтернативы, исходя из коллективного мнения экспертов. Для решения этой задачи есть несколько подходов.

По правилу большинства подсчитывается число экспертов, отдавших предпочтений каждой из альтернатив. Наилучшей объявляется альтернатива, которую назвали лучшей большинство экспертов. В данном случае, это альтернатива a_2 . Однако, вряд ли такое ранжирование можно назвать справедливым.

Кондорсе предложил следующий способ определения наилучшей альтернативы (см., например, [11]). Каждый эксперт ранжирует альтернативы по предпочтениям. На основании полученных ранжирований для каждой пары альтернатив a_i, a_j подсчитывается s_{ij} - число экспертов, считающих a_i более предпочтительней, чем a_j . Если $s_{ij} > s_{ji}$, то a_i признается более предпочтительной, чем a_j . Альтернатива a_i объявляется наилучшей альтернативой (альтернативой Кондорсе), если эта альтернатива признана более предпочтительной, чем любая из остальных. В данном примере (табл.1) такой альтернативой Кондорсе будет являться a_1. Однако при использовании принципа Кондорсе может возникать указанный им же парадокс, являющийся следствием нетранзитивности коллективных предпочтений. Пусть три эксперта проранжировали альтернативы следующим образом: $a_1 > a_2 > a_3$, $a_3 > a_1 > a_2$, $a_2 > a_3 > a_1$. Тогда $s_{12} > s_{21}$, $s_{23} > s_{32}$, но $s_{13} < s_{31}$. Альтернативы Кондорсе в этом случае не существует.

Вычисление средних величин для тех или иных совокупностей данных –основная статистическая процедура. В центре теории вероятностей и математической статистики находятся Законы больших чисел: эмпирические средние сходятся к теоретическим. В классическом варианте: выборочное среднее арифметическое при определенных условиях сходится по вероятности при росте числа слагаемых к математическому ожиданию.

Развитие математического инструментария решения прикладных задач, прежде всего в экспертных технологиях и социологии, привело к необходимости использования средних значений в пространствах нечисловой природы [12, 13]. Сначала в качестве средних значений бинарных отношений применяли медианы Кемени. Затем оптимизационный подход к построению средних величин стал стержнем нечисловой статистики - новой области прикладной математической статистики [14, 15].

В книге Дж. Кемени и Дж. Снелла [6] (на основе более ранней статьи Кемени [16]) в качестве итогового мнения комиссии экспертов предложено применять «медиану Кемени», т.е. результат минимизации суммы расстояний Кемени от мнений экспертов до произвольного бинарного отношения X.

Расстоянием Кемени между бинарными отношениями A и B, описываемыми матрицами $\| a_{ij} \|$ и $\| b_{ij} \|$ соответственно, называется число

$$D(A, B) = \sum \left| a_{ij} - b_{ij} \right|,$$

где суммирование производится по всем i, j от 1 до k. Расстояние Кемени - это число несовпадающих элементов в матрицах $\| a_{ij} \|$ и $\| b_{ij} \|$.

С помощью расстояния Кемени находят итоговое мнение комиссии экспертов. Пусть n ответы экспертов представлены как упорядочения (кластеризованные ранжировки) $A(1)$, $A(2)$, …, $A(n)$, или как бинарные отношения других типов (отношения эквивалентности, толерантности и др.). Для их усреднения используют т.н. **медиану Кемени**

$$\mathrm{Argmin} \sum_{s=1}^{n} D(A_s, A).$$

Медиана Кемени используется для описания усреднения ответов экспертов, представленных в виде бинарных отношений из рассматриваемого класса бинарных отношений.

Согласно [17] медиана Кемени является кондорсетовым ранжированием, так как если существует наилучшая альтернатива Кондорсе, то она всегда будет ранжироваться наиболее высоко. Медиана Кемени также удовлетворяет большинству критериев Эрроу [17]:

1) универсальность множества допустимых отношений – для любой тройки альтернатив должны найтись отношения такие, что первое связывает все три альтернативы попарно, второе и третье только первые две альтернативы и требование транзитивности результирующего отношения.

2) условие монотонности – если какой-то эксперт изменил своё мнение в пользу результирующего отношения, то оно от этого не изменится.

3) ненавязанность – для любой пары альтернатив существуют множества отношений, такие, что для первого множества пара альтернатив принадлежит оптимальному решению, а для второго нет.

4) отсутствие диктатора – нет эксперта, мнение которого определяет решение независимо от остальных экспертов.

Из сказанного выше можно заключить, что медиану Кемени можно считать одним из наиболее корректных результирующих ранжирований.

4. Вычисление медианы Кемени

Если ответы экспертов - кластеризованные ранжировки, то вычисление медианы Кемени является задачей целочисленного программирования. Для ее нахождения используют различные алгоритмы дискретной математики, в частности, основанные на методе ветвей и границ, на разных подходах кластерного анализа. В настоящее время неизвестно ни одного точного алгоритма поиска множества всех медиан Кемени для заданного множества перестановок (ранжировок без связей), кроме полного перебора.

Однако существуют алгоритмы, разработанные российскими учеными Б.Г. Литваком [17], В.Н. Жихаревым [18].

Рассмотрим подход В.Н. Жихарева. Для практического использования вместо полного перебора и селекции множества всех медиан в множестве всех кластеризованных ранжировок (как он пишет - перестановок) В.Н.

Жихарев предлагает эвристические алгоритмы [18], в которых исследует меньшие по количеству элементов множества - производные от совокупности экспертных ответов - и выделяет в них т.н. псевдомедианы в предположении, что псевдомедианы могут быть медианами. Псевдомедианы ищутся так же как медианы Кемени, только на выделенном множестве перестановок.

Псевдомедианой совокупности E экспертных перестановок в множестве V называется перестановка, $pm(E, V)$, для которой сумма расстояний $D(x, E)$ до элементов совокупности E принимает минимальное значение среди всех перестановок x множества V.

Жихарев использует также следующие понятия в пространстве G кластеризованных ранжировок (перестановок), которое рассматривается как граф (ребрами соединены перестановки, отличающиеся одной инверсией):

Сферой $S(x, R)$ радиуса R с центром в точке (перестановке) x называется множество всех таких вершин (перестановок) y графа G, для которых выполняется условие $d(x, y) = R$.

Шаром $B(x, R)$ радиуса R с центром в точке (перестановке) x называется множество всех таких вершин (перестановок) y графа G , для которых расстояние $d(x, y)$ не превосходит R.

***R* - окрестностью** $B(V, R)$ множества перестановок V радиуса R называется объединение шаров радиуса R с центрами в точках y множества V. В частности, $B(V, 0) = V$.

Границей (*R* - фронтом) $S(V, R)$ R - окрестности $B(V, R)$ множества перестановок V называется теоретико–множественная разность

$$B(V, R) - B(V, R - 1),$$

причем $S(V, 0) = V$.

Центром *C*(*V*) множества *V* (центром минимальной окрестности) называется первое непустое по последовательности R = 0, 1, 2, …

пересечение всех шаров $B(x, R)$ радиуса R с центрами во всех точках x множества V, а соответствующее минимальное значение R называется радиусом минимальной окрестности $U(C(V), R)$, включающей в себя все элементы множества V.

Взвешенным (или смещенным) центром $Cw(E)$ совокупности E называется первое непустое по последовательности $R = 0, 1, 2, \ldots$ пересечение всех шаров $B(x, [R/k(x)])$ с центрами во всех точках x множества $N(E)$ совокупности E, где $k(x)$ – кратность элемента x в совокупности E, $[R/k(x)]$ – целая часть рационального числа $R/k(x)$.

Окрестность $U(Cw(E), R)$ включает в себя все элементы множества $N(E)$ совокупности E.

Используемые производные множества перестановок Жихарев строит путем нахождения центров множества ответов экспертов, далее алгоритмически специальным образом задаются сферы, шары, окрестности на данном множестве, тем самым ограничивая полный перебор на множестве всех перестановок экспертных ранжировок.

Дадим краткое описание алгоритмов В.Н. Жихарева.

Алгоритм 1.

1. Построить центр $C(N(E))$ множества $N(E)$ совокупности E.
2. Найти множество всех псевдомедиан $PM(E, C(N(E)))$.

Алгоритм 2.

1. Построить центр $C(N(E))$ множества $N(E)$ совокупности E.
2. Построить окрестность $U(C(N(E)), R)$, центра $C(N(E))$ минимального радиуса R, содержащую все элементы множества $N(E)$.

3. Найти множество всех псевдомедиан $PM(E, U(C(N(E)), R))$.

Алгоритм 3.

1. Построить центр $C(N(E))$ множества $N(E)$ совокупности E.
2. Построить шаровую окрестность $B(e, R)$, произвольной точки e центра $C(N(E))$ минимального радиуса R, содержащую все элементы множества $N(E)$.
3. Найти множество всех псевдомедиан $PM(E, B(e, R))$.

Алгоритм 4.

1. Построить смещенный центр $Cw(E)$ совокупности E.
2. Найти множество всех псевдомедиан $PM(E, Cw(E))$.

Алгоритм 5.

1. Построить смещенный центр $Cw(E)$ совокупности E.
2. Построить окрестность $U(Cw(E), R)$ центра $Cw(E)$ минимального радиуса R, содержащую все элементы множества $N(E)$.
3. Найти множество всех псевдомедиан $PM(E, U(Cw(E), R))$.

Алгоритм 6.

1. Построить смещенный центр $Cw(E)$ совокупности E экспертных ответов.
2. Построить шаровую окрестность $B(ew, R)$ произвольной точки ew смещенного центра $Cw(E)$ минимального радиуса R, содержащую все элементы множества $N(E)$.
3. Найти множество всех псевдомедиан $PM(E, B(ew, R))$.

Алгоритм 7.

1. *Старт*: $N = N1 = N(E)$, $R = 3$, $N(E)$- совокупность экспертных ответов.

2. *Операция “захвата”*:

Построить окрестность $B(N, R)$, где R - радиус окрестности, построенной в каждой из точек N.

3. *Операция селекции*:

Найти множество всех псевдомедиан $N1 = PM(E, B(N, R))$.

4. *Условие остановки*. Сравнить N и $N1$. Если N = N1, то остановка, иначе переход к п.2.

В работе В.Н. Жихарева использовалось несколько моделей ответов экспертов. В наиболее общей модели 1: совокупность экспертных ответов формируется методом случайной неупорядоченной выборки с повторениями, при этом любая перестановка множества X с одинаковой вероятностью может быть выбрана в качестве экспертного ответа. В модели 2: локализованная совокупность экспертных ответов, т.е. все экспертные ответы лежат внутри шара с фиксированным радиусом и равновероятны. |В модели 3: центрированная и локализованная совокупность экспертных ответов, т.е. существует одна "истинная" и наиболее вероятная перестановка - центр шара. Вероятности выбора других перестановок совокупности зависят от расстояния до центра шара.

В результате численных экспериментов В.Н. Жихарев проводит сравнение множества псевдомедиан, найденных с помощью предложенных эвристических алгоритмов, с множеством медиан Кемени, найденным им путем полного перебора (табл. 2). Жирным шрифтом в таблице указано описание серии экспериментов методом Монте-Карло, которое состоит из набора характеристических параметров, а именно:

номер модели совокупности экспертных ответов,

количество объектов,

количество экспертов,

количество испытаний (тестов).

(Например: **1-5-5-1000**).

В первых двух колонках табл. 2 указан порядковый номер и “условный” символ алгоритма, разработанного В.Н. Жихаревым. Используя поочередно каждый из предложенных алгоритмов, автор сравнивает количество найденных псевдомедиан с числом медиан Кемени, вычисленным методом полного перебора на множестве всех перестановок *X*. В третьей, четвертой и пятой колонках указаны проценты испытаний (тестов) для каждого типа алгоритмов, в которых удалось найти все медианы Кемени, некоторые или ни одной. В последней колонке указаны средние отношения времени работы каждого тестируемого эвристического алгоритма ко времени работы алгоритма полного перебора.

В.Н. Жихарев выделяет один из предложенных им эвристических алгоритмов (номер 7 в таблице) в качестве наиболее эффективного по количеству найденных медиан в условиях наиболее общей модели ответов экспертов (модель 1). Однако, как показывают данные последней колонки табл.2, алгоритм полного перебора в большинстве экспериментов работает более быстро.

Таблица 2. Сводная таблица результатов испытаний поиска медиан Кемени по алгоритмам В.Н.Жихарева

Номер алгоритма	Символ алгоритма	Найдены все медианы (% тестов)	Найдены не все медианы (% тестов)	Не найдено ни одной медианы (% тестов)	Относительное быстродействие
1-5-5-1000					
1	**C**	23.7	13.7	62.6	3.9
2	**U(C)**	99.5	0.2	0.3	0.00087
3	**U(Ce)**	98.	1.2	0.8	0.001
4	**Cw**	24.7	14.	61.3	3.78
5	**U(Cw)**	99.5	0.2	0.3	1.23
6	**U(Cwe)**	98.	1.2	0.8	1.46
7	**Move**	100.	0	0	1.11
1-5-10-1000					
1	**C**	7.4	38.1	54.5	1.7
2	**U(C)**	99.3	0.7	0	0.001
3	**U(Ce)**	94.2	5.7	0.1	0.0012
4	**Cw**	7.7	38.3	54	1.6
5	**U(Cw)**	99.3	0.7	0	0.69
6	**U(Cwe)**	94.8	5.1	0.1	0.76
7	**Move**	100.	0	0	0.81
1-7-5-61					
1	**C**	14.7541	8.19672	77.0492	4.02
2	**U(C)**	100	0	0	0.0205
3	**U(Ce)**	98.3607	1.63934	0	0.0257

4	**Cw**	14.7541	8.19672	77.0492	4.04
5	**U(Cw)**	100	0	0	1.43
6	**U(Cwe)**	98.3607	1.63934	0	1.8
7	**Move**	95.082	4.91803	0	11.1
1-6-5-500					
5	**U(Cw)**	100	0	0.3	1.3
7	**Move**	99.2	0.8	0	2.8
1-6-10-500					
5	**U(Cw)**	99	1	0	2.63
7	**Move**	100	0	0	3.26
2-6-10-500					
5	**U(Cw)**	100	0	0	4.4
7	**Move**	100	0	0	2.52
3-6-10-500					
5	**U(Cw)**	100	0	0	2.3
7	**Move**	100	0	0	2.4

В.Н. Жихарев считает, что метод полного перебора на современных компьютерах может быть достаточно эффективно применен для поиска всех медиан Кемени при относительно небольшом количестве экспертов (для числа объектов в перестановке вплоть до восьми и даже девяти). Описанные выше эвристические алгоритмы созданы для решения более сложных задач: выявления каких-либо закономерностей при изучении взаимного расположения медиан по отношению к экспертной совокупности или экспертному множеству.

К недостаткам предложенных методов вычисления медианы Кемени В.Н. Жихарев относит тот факт, что множество, построенное на первом этапе этих алгоритмов, может оказаться слишком большим – сравнимым по размерам (по числу \лементов) со множеством всех перестановок, и тогда на втором этапе фактически осуществляется полный перебор перестановок. Это обстоятельство следует иметь в виду прежде всего в том случае, когда имеется сильный разброс в ответах экспертов и они приблизительно могут быть описаны моделью №1. В этом случае может оказаться целесообразным не вычислять медианы, а провести новую экспертизу с другим подбором объектов [18].

В работе Б.Г. Литвака [17] предложены два алгоритма вычисления медианы Кемени: точный и эвристический. Расстояние от произвольного ранжирования P до всех ранжирований P_1, ..., P_m определяется как:

$$\sum_{v=1}^{m} d(P,P_v) = \sum_{v=1}^{m}\sum_{i<j} | p_{ij}^{(v)} - p_{ij} | = \sum_{v=1}^{m}\sum_{i<j} d_{ij}(P,P_v) \ ,$$

где

$$d_{ij}(P,P_v) = | p_{ij}^{(v)} - p_{ij} | .$$

Используется матрица потерь с элементами:

$$r_{ij} = \sum_{v=1}^{m} d_{ij}(P,P_v) .$$

Поиск медианы Кемени эквивалентен отысканию упорядочивания строк и столбцов матрицы потерь, обладающей минимальной суммой наддиагональных элементов. Точный алгоритм Литвака строится на теореме о необходимом и достаточном условии транзитивности матрицы потерь, обладающей свойством Кондорсе. Ранжирование P обладает свойством Кондорсе, если любое подмножество альтернатив $a_1,...,a_n$ обладает альтернативой Кондорсе. Пусть матрица потерь построена для единственного ранжирования *P*. Если в упорядочивании *P* альтернатива a_i предпочтительнее альтернативы a_j, то $r_{ij} < r_{ji}$. Если в матрице потерь $r_{il} \le r_{li}$, как только $r_{ij} \le r_{ji}$ и $r_{jl} \le r_{lj}$, такая матрица называется транзитивной. Матрица потерь может быть нетранзитивна в случае непоследовательности ответов экспертов при парных сравнениях. Пусть ранжирования P_1, ..., P_m обладают альтернативой Кондорсе - это означает, что матрица потерь содержит строку i_1 такую, что $r_{i_1 j} \le r_{j i_1}$, $j \in \{1,...,n\}$. Благодаря свойству транзитивности (последовательности экспертов в своих выборах) после удаления некой альтернативы a_{i_1}, альтернативой Кондорсе становится a_{i_2}.

Точный алгоритм Литвака использует метод ветвей и границ, в нем предполагается последовательное фиксирование расположения части альтернатив (наилучших альтернатив Кондорсе и наименее предпочтительных альтернатив), определение верхней и нижней границ значений целевой функции на уменьшенном таким образом множестве ранжирований и отбрасывание заведомо бесперспективных при поиске медианы Кемени вариантов. Точный алгоритм Б.Г. Литвака, в частности, описан Д.А. Сумкиным (в коллективной монографии [19]).

Если ранжирования P_1, ..., P_m не обладают свойством Кондорсе, то Литвак предлагает использовать разработанный им эвристический алгоритм поиска медианы Кемени. В этом алгоритме на первом этапе последовательно находятся минимумы из сумм:

$$S_1 = \sum_{j}^{n} r_{1j}\ ,\ S_n = \sum_{j}^{n} r_{nj}\ ,$$

после чего соответствующую минимальной сумме

$$S_{\min} = \sum_{j}^{n} r_{\min j}$$

альтернативу $a_{(\min)}$ ставят на первое место, вычеркивая ее из матрицы потерь. Далее аналогичные действия проводятся с уменьшенной матрицей потерь до тех пор, пока не получится ранжировка: $a_{i_1}...a_{i_k}$. На втором этапе для полученной ранжировки последовательно проверяются соотношения: $r_{i_k i_{k+1}} \le r_{i_{k+1} i_k}, k = n-1,...,1$. Как только для некоторого k оно нарушено, альтернативы a_{i_k} и $a_{i_{k+1}}$ меняются местами в новом ранжировании. Данный алгоритм описан, например, в работе польских авторов [20].

Недостатком подходов Б.Г. Литвака является то, что предложенные методы поиска могут найти лишь одну медиану из большого множества возможных, поэтому в общем случае они должны быть модифицированы.

В простейшем случае, представляет интерес подсчет итогового мнения комиссии экспертов только среди указанных экспертами ответов (“модифицированная” медиана Кемени, предложенная в [21]).

5.Изучение свойств медианы Кемени

В работах [22, 14, 15] приведены результаты изучения свойств медианы Кемени, полученные с помощью расчетов по алгоритмам В.Н. Жихарева [18], описанным выше. В частности, приводятся следующие наблюдения.

1. Средняя величина количества медиан сравнительно невелика, однако больше единицы.

2. В отдельных случаях количество медиан может значительно превосходить количество экспертов.

3. Диаметр множества медиан во всех экспериментах меньше диаметра множества экспертных ответов. Это позволяет с большой долей уверенности утверждать, что оператор вычисления множества всех медиан является

сжимающим оператором на множестве всех подмножеств множества всех перестановок.

4. Множество медиан в значительном количестве испытаний (тестов) пересекается с множеством экспертных ответов. Частота этого события в серии экспериментов недостаточно велика для того, чтобы на этой основе строить эвристический алгоритм поиска медиан. Исключение составляет модель № 3, в которой частота равна 92. 6%.

5. Имеются тесты, в которых медианы достаточно далеко удалены от множества экспертов.

Одной из важных задач является исследование зависимости итогового решения комиссии от назначенной некоторыми способами “авторитетности” того или иного эксперта. Например, авторитетность каждого эксперта v может задаваться весовым коэффициентом - мультипликатором k_v таким образом, что расстояние от произвольного ранжирования P до всех ранжирований P_1, ..., P_m вычисляется следующим образом:

$$\sum_{v=1}^{m} d(P, P_v) = \sum_{v=1}^{m} \sum_{i<j} \mathrm{k}_v * | p_{ij}^{(v)} - p_{ij} | = \sum_{v=1}^{m} \sum_{i<j} d_{ij}(P, P_v) \ ,$$

где

$$d_{ij}(P, P_v) = k_v * | p_{ij}^{(v)} - p_{ij} | .$$

Интуитивно это означает, что чем больше значимость эксперта (больше коэффициент k_v), тем сильнее будет учитываться вклад разногласия с ним в матрице потерь, измеряемое расстоянием в матрице ранжировок.

Также представляет интерес изучение расстояния между медианой Кемени и модифицированной медианой Кемени, найденной среди ответов экспертов.

Медиана Кемени заслуживает дальнейшего изучения. Речь идет как об алгоритмических аспектах, так и о свойствах самой медианы и ее оценок с помощью различных методов при конкретных объемах выборки (числе экспертов). Отметим, что и такой инструмент исследования, как метод Монте-Карло (статистических испытаний), требует дальнейшего развития [23, 24].

Целесообразно широко применять медиану Кемени при решении различных прикладных задач, анализировать опыт такого применения.

Литература

1. Орлов А.И. Экспертные оценки // Заводская лаборатория. Диагностика материалов. 1996. Т.62. № 1. С.54-60.
2. Орлов А.И. Организационно-экономическое моделирование: учеб. Ч.2. Экспертные оценки. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. - 486 с.
3. Орлов А.И. Теория экспертных оценок в нашей стране // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 1-11.
4. Орлов А.И. Анализ экспертных упорядочений // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 112. С. 21–51.
5. Орлов А.И. Эконометрика. - М.: Издательство "Экзамен", 2002. - 576 с.
6. Кемени Дж., Снелл Дж. Кибернетическое моделирование. Некоторые приложения. - М.: Советское радио, 1972. - 192 с.
7. Кострикин А.И. Введение в алгебру. Основы алгебры.— М.: Физматлит, 1994. — 320 с.
8. Холлендер М., Вулф Д. Непараметрические методы статистики. - М.: Финансы и статистика, 1983. - 518 с.
9. Миркин Б.Г. Проблема группового выбора. - М.: Наука, 1974. - 256 с.
10. Шрейдер Ю.А. Равенство, сходство, порядок. - М.: Наука, 1971. - 256 с.
11. Литвак Б.Г. Экспертные оценки и принятие решений. - М.: Патент, 1996. - 272 с.
12. Орлов А.И. Организационно-экономическое моделирование: : учебник : в 3 ч. Ч.1: Нечисловая статистика. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2009. — 542 с.
13. Орлов А.И. О развитии статистики объектов нечисловой природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 41-50.
14. Орлов А. И. О средних величинах // Управление большими системами. Выпуск 46. М.: ИПУ РАН, 2013. С.88-117.
15. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 89. С. 175-200.
16. Kemeny J. Generalized random variables // Pacific Journal of Mathematics. – 1959. – Vol. 9. – P. 1179 –1189.

17. Литвак Б.Г. Экспертная информация. Методы получения и анализа. - М. «Радио и связь», 1982. - 184 с.

18. Жихарев В.Н. Медиана Кемени. [Электронный ресурс] URL: http://www.bmstu.ru/ps/%7Eorlov/ (дата обращения 05.09.2016).

19. Космонавтика XXI века. Попытка прогноза развития до 2101 года / Батурин Ю.М., Черток Б.Е., Шуров А.И. Кричевский С.В., Сумкин Д.А. - М.: РТСофт, 2011. - 864 с.

20. Hanna Bury, Dariusz Wagner. Kemey’s Median Algorithm. Appliction for determining group judgement.July 15-17, 2002, IIASA, Laxenburg, Austria.

21. Орлов А.И. Роль медиан Кемени в экспертных оценках и статистическом анализе данных // Теория активных систем: Труды международной научно-практической конференции (14-16 ноября 2011 г., Москва, Россия). Том I. Общая редакция – В.Н. Бурков, Д.А. Новиков. – М.: ИПУ РАН, 2011. – С.172-176.

22. Жихарев В. Н., Орлов А. И. Законы больших чисел и состоятельность статистических оценок в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. – Пермь: Изд-во Пермского госуд. ун-та, 1998. – С. 65-84.

23. Орлов А.И. Взаимосвязь предельных теорем и метода Монте-Карло // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 114. С. 27–41.

24. Орлов А.И. Предельные теоремы и метод Монте-Карло // Заводская лаборатория. Диагностика материалов. 2016. Т.82. №7. С. 67-72.

References

1. Orlov A.I. Jekspertnye ocenki // Zavodskaja laboratorija. Diagnostika materialov. 1996. T.62. № 1. S.54-60.

2. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie: ucheb. Ch.2. Jekspertnye ocenki. - M.: Izd-vo MGTU im. N.Je. Baumana, 2011. - 486 s.

3. Orlov A.I. Teorija jekspertnyh ocenok v nashej strane // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 93. S. 1-11.

4. Orlov A.I. Analiz jekspertnyh uporjadochenij // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 112. S. 21–51.

5. Orlov A.I. Jekonometrika. - M.: Izdatel'stvo "Jekzamen", 2002. - 576 s.

6. Kemeni Dzh., Snell Dzh. Kiberneticheskoe modelirovanie. Nekotorye prilozhenija. - M.: Sovetskoe radio, 1972. - 192 s.

7. Kostrikin A.I. Vvedenie v algebru. Osnovy algebry.— M.: Fizmatlit, 1994. — 320 s.

8. Hollender M., Vulf D. Neparametricheskie metody statistiki. - M.: Finansy i statistika, 1983. - 518 s.

9. Mirkin B.G. Problema gruppovogo vybora. - M.: Nauka, 1974. - 256 s.

10. Shrejder Ju.A. Ravenstvo, shodstvo, porjadok. - M.: Nauka, 1971. - 256 s.

11. Litvak B.G. Jekspertnye ocenki i prinjatie reshenij. - M.: Patent, 1996. - 272 s.

12. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie: : uchebnik : v 3 ch. Ch.1: Nechislovaja statistika. — M.: Izd-vo MGTU im. N. Je. Baumana, 2009. — 542 s.

13. Orlov A.I. O razvitii statistiki ob#ektov nechislovoj prirody // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 93. S. 41-50.

14. Orlov A. I. O srednih velichinah // Upravlenie bol'shimi sistemami. Vypusk 46. M.: IPU RAN, 2013. S.88-117.

15. Orlov A.I. Srednie velichiny i zakony bol'shih chisel v prostranstvah proizvol'noj prirody // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 89. S. 175-200.

16. Kemeny J. Generalized random variables // Pacific Journal of Mathematics. – 1959. – Vol. 9. – P. 1179 –1189.

17. Litvak B.G. Jekspertnaja informacija. Metody poluchenija i analiza. - M. «Radio i svjaz'», 1982. - 184 s.

18. Zhiharev V.N. Mediana Kemeni. [Jelektronnyj resurs] URL: http://www.bmstu.ru/ps/%7Eorlov/ (data obrashhenija 05.09.2016).

19. Kosmonavtika XXI veka. Popytka prognoza razvitija do 2101 goda / Baturin Ju.M., Chertok B.E., Shurov A.I. Krichevskij S.V., Sumkin D.A. - M.: RTSoft, 2011. - 864 s.

20. Hanna Bury, Dariusz Wagner. Kemey’s Median Algorithm. Appliction for determining group judgement.July 15-17, 2002, IIASA, Laxenburg, Austria.

21. Orlov A.I. Rol' median Kemeni v jekspertnyh ocenkah i statisticheskom analize dannyh // Teorija aktivnyh sistem: Trudy mezhdunarodnoj nauchno-prakticheskoj konferencii (14-16 nojabrja 2011 g., Moskva, Rossija). Tom I. Obshhaja redakcija – V.N. Burkov, D.A. Novikov. – M.: IPU RAN, 2011. – S.172-176.

22. Zhiharev V. N., Orlov A. I. Zakony bol'shih chisel i sostojatel'nost' statisticheskih ocenok v prostranstvah proizvol'noj prirody // Statisticheskie metody ocenivanija i proverki gipotez: mezhvuz. sb. nauch. tr. – Perm': Izd-vo Permskogo gosud. un-ta, 1998. – S. 65-84.

23. Orlov A.I. Vzaimosvjaz' predel'nyh teorem i metoda Monte-Karlo // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 114. S. 27–41.

24. Orlov A.I. Predel'nye teoremy i metod Monte-Karlo // Zavodskaja laboratorija. Diagnostika materialov. 2016. T.82. №7. S. 67-72.