

УДК 005.935.3 + 025.2

ББК 65.050.2-73

ОБ ОДНОМ ПОДХОДЕ К ИСПОЛЬЗОВАНИЮ ЭКСПЕРТНЫХ МЕТОДОВ ДЛЯ ОЦЕНКИ НАУЧНЫХ ЖУРНАЛОВ

Абаев Л. Ч.¹

*(Российский институт
стратегических исследований, Москва)*

Проводится анализ статьи О. В. Федорца «Коллективная экспертиза научных журналов: методика агрегирования экспертных оценок и построения рейтинга». Обсуждаются вопросы уточнения постановки задачи и проведения коллективной экспертизы. Рассматриваются различные методы агрегирования ранговых экспертных оценок.

Ключевые слова: организация экспертного опроса, рейтинг научных журналов, ранговые оценки, агрегирование экспертной информации.

1. Актуальность проблемы

Статья О. В. Федорца «Коллективная экспертиза научных журналов: методика агрегирования экспертных оценок и построения рейтинга» [7] является актуальной как в теоретическом плане, так и в практическом.

Проблема оценки и рейтингования российских научных журналов представляется весьма важной. В действующих в настоящее время индексах цитирования научных статей отечественные работы встречаются достаточно редко, во всяком случае – русскоязычные. Отечественный индекс цитирования,

¹ Лев Черменович Абаев, доктор технических наук, ведущий научный сотрудник (abaev_lev@mail.ru).

который мог бы в определенной степени характеризовать и научные журналы, находится в настоящее время на начальном этапе своего развития. Поэтому рассмотрение проблемы оценки качества российских научных журналов представляется вполне обоснованным. К тому же, как справедливо указывается в статье О. В. Федорца, решение данной проблемы позволит производить эффективный отбор научных журналов для их реферирования в ВИНТИ.

В теоретическом плане данная работа также является актуальной, поскольку способствует исследованию весьма важной задачи экспертного ранжирования объектов при неполных исходных данных.

Не будучи специалистом по науковедению, позволю себе, тем не менее, высказать ряд соображений и замечаний, касающихся постановки задачи, организации экспертного опроса и методики обработки экспертной информации.

2. О постановке задачи и некоторых аспектах организации экспертного опроса при построении рейтингов научных журналов

Что касается постановки задачи, то в работе О. В. Федорца вызывает настороженность определенная ее «раздвоенность». С одной стороны, автор предлагает экспертам проводить ранжирование научных журналов от «лучшего» к «худшему», а с другой – утверждает, что подобная процедура необходима для отбраковки «плохих» журналов (действительно, формирование множества реферируемых в ВИНТИ журналов требует не определения «лучших», а отсеивания «худших»).

В случае получения от эксперта полной информации, т.е. ранжирования всех объектов противоречия не возникает, так как экспертное упорядочение определяет и «лучшие» и «худшие» объекты. Однако если эксперты упорядочивают лишь относительно небольшую долю объектов (а именно такой случай и рассматривает автор), то решить задачу отсеивания «пло-

хих» объектов, на мой взгляд, становится затруднительным. Дело в том, что при упорядочении от «лучшего» к «худшему» объекту в случае неполного ранжирования эксперт будет оценивать группу журналов-«лидеров», а не «аутсайдеров».

Представляется, что более обоснованной в данной ситуации была бы процедура упорядочения журналов от «худшего» к «лучшему». В этом случае эксперт должен был бы определять наиболее, на его взгляд, «некачественные» журналы, что и позволило бы путем агрегирования экспертной информации определить подмножество отбраковываемых изданий.

Такой подход, кстати, в определенной степени позволит снизить риск манипулирования, вольного или невольного, со стороны экспертов. А именно, при ранжировании от «лучшего» к «худшему» объекту эксперт, заинтересованный в «продвижении» того или иного журнала будет ставить его на высокое место, что может привести к существенному увеличению множества «хороших» журналов (особенно если для агрегирования экспертной информации используется методика, предлагаемая автором). В то же время, при ранжировании от «худшего» к «лучшему» объекту «продвижение» вверх «своего» издания окажется, во всяком случае, гораздо более затруднительным. Поэтому представляется целесообразным продумать исходную постановку задачи и, при необходимости, скорректировать ее.

Второй момент, вызывающий сомнения, заключается в «общем» ранжировании всех журналов – и специализированных, причем различной тематики, и политематических (во всяком случае, именно это следует из рассмотрения представленного в статье списка «ТОР–20 научных журналов»).

На мой взгляд, более надежным представляется ранжирование научных журналов «внутри» своей специализации и, соответственно – отдельно политематических. Конечно, здесь могут возникнуть трудности с отнесением некоторых журналов к тому или иному классу, к тому же возникнет проблема приоритетности тех или иных «классов» журналов. Тем не менее, упорядочение «однотипных» журналов может быть проведено с

гораздо большей обоснованностью, нежели упорядочение «разнотипных» журналов. Я вообще не очень представляю, как можно достаточно адекватно сравнить, например, «физический» журнал с «машиностроительным», а ведь ранжирование предполагает получение именно сравнительных оценок. Если же в результате неполного ранжирования фактически сравнимыми окажутся лишь «однотипные» издания, то и практически любое итоговое упорядочение всех журналов окажется весьма ненадежным.

Кстати, ранжирование журналов «внутри» своей «специализации» может помочь в решении и той важной проблемы, на которую совершенно справедливо указал автор, а именно – на сложность получения высоких итоговых оценок для журналов довольно узкой специализации, интересующих лишь небольшой круг специалистов. Вследствие этого, они получают небольшое число экспертных оценок и по критерию суммы оценок, предложенному автором, оказываются «неконкурентоспособными» по сравнению с более «популярными» журналами (автор, правда, решает эту проблему, используя критерий максимальной оценки, однако данный критерий, как будет показано ниже, имеет ряд недостатков). В случае, когда узкоспециализированные издания будут рассматриваться отдельно, часть из них обязательно попадет в список реферируемых изданий.

Проблема отнесения журнала к тому или иному классу может быть решена с использованием методов классификации на базе, например, существующего классификатора научных направлений. Что касается процедуры формирования общего множества реферируемых журналов, то в первом приближении, например, можно потребовать, чтобы доли журналов различных направлений (включая политематические) в итоговом списке совпадали с долями этих журналов в исходном множестве оцениваемого массива. С учетом данного требования и может быть выделено «необходимое» количество журналов каждой тематики, подлежащих реферированию в ВИНТИ.

3. О методике агрегирования экспертных оценок при построении рейтинга научных журналов

Предложенная автором методика агрегирования экспертных оценок при построении рейтинга научных журналов имеет, на мой взгляд, ряд серьезных недостатков.

Предлагаемый автором в качестве основной критерий максимальной оценки не позволяет провести эффективную ранжировку объектов. Всего одна максимальная оценка дает возможность тому или иному журналу оказаться в лидирующей группе (причем такая оценка может оказаться и результатом манипулирования со стороны какого-либо эксперта). Сам автор отмечает, что по данному критерию из исходного множества в «лидирующий» кластер было включено 45,5% журналов, имеющих хотя бы одну максимальную оценку (почти половина). По моему, это очень большая доля¹.

Автор, понимая это, вводит второй критерий – сумму оценок, полученную рассматриваемым изданием.

Здесь, по моему мнению, автор совершает существенную методологическую ошибку, так как предлагаемый им подход оказывается логически противоречивым.

Действительно, автор утверждает, что экспертные ранжирования представляют собой измерения в порядковой шкале. Но тогда вербальная шкала и соответствующие ей «количественные» оценки могут использоваться лишь для удобства. При этом в случае использования порядковых шкал такая операция с ранговыми оценками, как их суммирование (применяемая, например, в процедуре Борда), не представляется вполне обоснованной даже в том случае, если все эксперты проранжировали все объекты (по крайней мере, в этом случае требуется наличие

¹ Кстати, интерес представляет информация о том, какая доля от всего объема оцениваемых изданий оказалась по данному критерию в «лидирующем» кластере, а также – какую долю в этом кластере составляют журналы с одной максимальной оценкой.

высокой согласованности групповых экспертных ранжирований).

В гораздо большей степени такой подход оказывается необоснованным в случае неполного экспертного ранжирования. Рассмотрим пример ранжирования восьми объектов шестью экспертами, представленный автором в разделе 6 своей работы.

Поскольку каждый эксперт, оценив некоторые объекты, сформировал на них отношение порядка, становится возможным для k -го эксперта ($k = 1, \dots, 6$) построить матрицу неполных парных сравнений A_k , где a_{ij}^k принимает значение 1, если для k -го эксперта i -й объект предпочтительнее j -го, $\frac{1}{2}$ – если данные объекты для эксперта равноценны и 0 – если j -й объект предпочтительнее i -го.

Рассмотрим матрицу A с ячейками

$$(1) \quad a_{ij} = \sum_{k=1}^6 a_{ij}^k,$$

где $i, j = 1, \dots, 8$, причем в тех случаях, когда некоторые a_{ij}^k не определены, будем считать их нулевыми. Результаты расчета по формуле (1) представлены в таблице 1 (отметим, что, хотя элементы (6, 8) и (8, 6) матрицы A имеют значение н/о – нет оценок, в дальнейших расчетах их значения будут считаться нулевыми).

Из построенной матрицы видно, что объект F , по мнению экспертов, предпочтительнее (с перевесом в один голос) лишь объекта G . Объекты A и D эквивалентны F , а объекты H и F – не сравнимы. Что касается объектов B , C и E , то они, по мнению экспертной группы, оказались предпочтительнее F (с перевесом в один голос каждый).

Возникает вопрос, можно ли считать на основании данных экспертных оценок объект F лучшим? Очевидно, что нельзя. Однако методика автора рассматриваемой статьи именно объект F выделяет в качестве лучшего. Что определило такой результат? Именно необоснованное использование критерия суммы, предлагаемого автором.

Таблица 1. Матрица неполных парных сравнений

	Объекты								*	**
	A	B	C	D	E	F	G	H		
A	–	0,5	0	1	1,5	1,5	2	0	6,5	50
B	1,5	–	0	0,5	1	2	2	0	7	53,8
C	2	1	–	1	1	1	1	1	8	72,7
D	0	1,5	1	–	0,5	1	1	0	5	50
E	0,5	1	1	0,5	–	1	2	0	6	50
F	1,5	1	0	1	0	–	1,5	н/о	5	41,7
G	0	0	0	0	0	0,5	–	0	0,5	4,5
H	1	1	1	1	2	н/о	1	–	7	87,5
	6,5	6	3	5	6	7	10,5	1		
	Сумма потерянных объектами очков									

(в таблице 1 * – сумма набранных объектами очков; ** – процент набранных объектами очков)

Возвращаясь к матрице A , отметим, что построение такого рода матриц может оказаться весьма полезным, даже если не использовать для ее обработки те или иные формальные процедуры.

Действительно, она позволяет понять, какие объекты имеют достаточное количество сравнительных оценок, а какие объекты необходимо дополнительно сравнивать, существуют ли противоречия в экспертных оценках, определяемые нарушением свойства транзитивности (например, в рассматриваемом случае данное свойство оказывается нарушенным для объектов A , B и D). В ряде случаев матрицы такого рода позволяют, не прибегая к относительно сложной процедуре итогового упорядочения объектов (ввиду неполных парных сравнений, а также нарушений свойства транзитивности), тем не менее, решить задачу выделения подмножества «лучших» объектов (например, выделяя множество недоминируемых вариантов). Кстати, эта процедура может, на мой взгляд, оказаться эффективной при отбраковке «плохих» журналов в случае, если экспертные ранжирования проводятся от «худшего» объекта к «лучшему».

Естественно, что рассмотренная матрица групповых парных сравнений может использоваться и для итогового упорядочения объектов. В качестве примера рассмотрим подход, основанный на введении уровней «надежности» l . В нашем случае l принимает значения 0,5; 1; 1,5 и 2 и содержательно интерпретируется как «объем» используемой экспертной информации для упорядочения объектов. Определим матрицы A_l следующим образом:

$$(2) \quad a_{ij}^l = \begin{cases} 1, & \text{если } a_{ij} \geq l, \\ 0, & \text{в противном случае,} \end{cases} \quad i, j = 1, \dots, 8.$$

Тогда получим семейство обычных матриц парных сравнений, по которым и определим итоговое упорядочение объектов с точностью до классов эквивалентности (будем считать, что объекты, образующие циклические предпочтения, входят в один класс).

В итоге придем к следующим результатам для различных l .

$$1. l = 0,5 - \{C, H\} \succ \{A, B, E\} \succ \{D, F\} \succ G.$$

$$2. l = 1 - \{C, H\} \succ \{A, B, D\} \succ E \succ F \succ G,$$

при этом объекты A , B и D образуют цикл предпочтений.

$$3. l = 1,5 - \{C, D, H\} \succ B \succ \{A, F\} \succ E \succ G.$$

$$4. l = 2 - \{B, C, D, H\} \succ \{A, E, F\} \succ G.$$

Отметим, что рассмотренная процедура идентична известной процедуре представления нечетких бинарных отношений предпочтения в виде семейства α -срезов и решения на их основе задачи выбора.

Конечно, представленный выше подход носит иллюстративный характер, но и он показывает необоснованность выбора в качестве оптимального объекта F , полученного с использованием процедуры суммирования экспертных оценок, предложенного автором.

В то же время необходимо отметить, что критерий суммы на самом деле может в ряде случаев использоваться для анализа

неполных матриц групповых парных сравнений. Однако здесь необходимо понимание сложности этой задачи.

Например, если использовать критерий суммы набранных очков, то результат будет следующим (см. таблицу 1):

$$C \succ \{B, H\} \succ A \succ E \succ \{D, F\} \succ G.$$

А если использовать критерий суммы потерянных очков, то результат изменится:

$$H \succ C \succ D \succ \{B, E\} \succ A \succ F \succ G.$$

Очевидно, что в связи с различным числом парных сравнений для обоснованного упорядочения необходимо учитывать и число набранных, и число потерянных очков (для матриц полных парных сравнений такой проблемы не возникает).

Например, в квалификационных шахматных соревнованиях, в случае, когда турнир оказывается незаконченным, для оценки игроков используется показатель «процент набранных очков», но соответствующий разряд присваивается лишь в том случае, если игрок сыграл не меньше определенного числа партий. В случае использования вышеуказанного критерия результат будет следующий:

$$H \succ C \succ B \succ \{A, D, E\} \succ F \succ G.$$

Представленные результаты показывают, к каким неоднозначным результатам может привести непродуманное использование критерия суммы оценок.

В связи с этим рекомендую автору обсуждаемой статьи обязательно ознакомиться с работами П. Ю. Чеботарева [8, 9 и др.], в которых получены важные результаты, касающиеся как метода строчных сумм (для полных и неполных парных сравнений), так и в целом процедур агрегирования неполных экспертных предпочтений.

Наконец, для решения задачи, поставленной автором в рассматриваемой статье, могут быть использованы и вероятностные (статистические) модели обработки экспертной информации. К таковым относятся модели Тэрстоуна [5], Брэдли-Терри [2], стохастическая модель Ушакова [6] и др.

Возможно, данные модели окажутся более устойчивыми к неполноте экспертной информации, однако это требует дополнительной проверки.

К указанным моделям примыкают и такие методы, как метод взвешенных парных сравнений [3] и равномерного сглаживания [4]. Групповые экспертные ранжирования могут быть преобразованы тем или иным способом к матрицам степенной калибровки, после чего эти методы тоже могут использоваться (естественно, также после проверки их устойчивости к неполноте экспертной информации) для решения рассматриваемой задачи. Некоторые результаты, касающиеся этих моделей, представлены в [1].

В заключение следует отметить, что автором рассмотренной статьи, О. В. Федорцом, проделана достаточно серьезная работа по сбору, систематизации и анализу информации, касающейся рейтингования научных журналов. Используя весьма объемную информационную базу ВИНТИ, автору следует, на мой взгляд, продолжить свои исследования, уточняя и улучшая разработанную им методику. Не вызывает сомнений, что ее совершенствование до высокого научного уровня позволит решить целый ряд задач как теоретического, так и прикладного характера.

Литература

1. АБАЕВ Л. Ч. *Проблемы преобразования «качественной» экспертной информации в «количественную»* // Вторая международная конференция по проблемам управления. – Т. 1. – Москва: ИПУ РАН, 2003. – С. 229-236.
2. БЕЛКИН А. Р., ЛЕВИН М. Ш. *Принятие решений: комбинаторные модели аппроксимации информации* – М.: Наука. Гл. ред. физ.-мат. лит., 1990. – 160 с.
3. БРУК Б. Н., БУРКОВ В. Н. *Методы экспертных оценок в задачах упорядочения объектов* // Изв. АН СССР. Техн. кибернетика. – 1972. – №3. – С.29-39.

4. КИСЕЛЕВ Ю. В. *Оценка важности программ методом парных сравнений* // Изв. АН СССР. Техн. Кибернетика. – 1971. – №3. – С. 41-47.
5. МИРКИН Б. Г. *Проблема группового выбора*. – М.: Наука, 1974. – 256 с.
6. УШАКОВ И. А. *Задача о выборе предпочтительного объекта* // Изв. АН СССР. Техн. кибернетика. – 1971. – №4. – С. 3-8.
7. ФЕДОРЕЦ О. В. *Коллективная экспертиза научных журналов: методика агрегирования экспертных оценок и построения рейтинга* // Управление большими системами. – 2009. – Выпуск 27. – С. 12 – 35.
8. ЧЕБОТАРЕВ П. Ю. *Агрегирование неполных предпочтений* // III Всесоюзная конференция «Методы социологических исследований». – Т. 5. – Москва: ИС АН СССР, 1989. – С. 59-62.
9. ЧЕБОТАРЕВ П. Ю. *Метод строчных сумм и приводящие к нему модели* // Сб. тр. ВНИИ системных исследований. – 1989. – №3. – С. 94-110.

AN APPROACH TO APPLY EXPERT-BASED METHODS TO ESTIMATE SCIENTIFIC JOURNALS

Lev Ch. Abaev, Russia's Institute for Strategic Studies, Moscow, Dr.Sc., leading researcher (abaev_lev@mail.ru).

Abstract: The analysis of the article «Collective expert examination of scientific journals: Procedure of expert judgments aggregation and rating construction» by O. V. Fedorets is carried out. Setting of scientific journals rating construction problem and aspects of carrying out the collective expertise are discussed. Various aggregation methods of rank experts' estimations are considered.

Keywords: the expertise organization, rating of scientific journals, rank judgments, aggregation of expert information.