

С. Д. Штовба, д. т. н., доц.; Е. В. Штовба, к. э. н.

SH-ИНДЕКС – НОВАЯ ДРОБНАЯ МОДИФИКАЦИЯ ИНДЕКСА ХИРША

Предложен новый наукометрический показатель – Sh-индекс. В Sh-индексе целая часть является обычным индексом Хирша, а дробная показывает, насколько ученый приблизился к следующему значению индекса Хирша. Дробная часть рассчитывается как доля опубликованных статей, необходимых для достижения следующего значения индекса Хирша.

Ключевые слова: наукометрия, индекс цитирования, индекс Хирша, дробная модификация, Sh-индекс.

Введение

Сегодня для оценивания продуктивности научной деятельности совместно с экспертными заключениями все чаще используются и наукометрические показатели. Интерес к наукометрическим показателям катализировали работы советского ученого В. В. Налимова, самой известной из которых является монография [1]. Сначала основным наукометрическим показателем было количество печатных работ ученого – общее или по отдельным типам: монографии, статьи, тезисы, публикации в специальных изданиях, вошедших в список ВАК Украины, внесенных в электронные базы Web of Science или Scopus, проиндексированных Google Scholar и т. п. Для искусственного улучшения этого показателя используют такие типовые приемы, как: дробление результатов для опубликования в разных изданиях или даже дублирование описания одних тех же материалов в статьях с разными названиями. Таким образом, погоня за количеством публикаций часто снижает качество научных работ.

Вторым популярным наукометрическим показателем стал индекс цитирования – количество публикаций со ссылками на автора, работы которого анализируются. Индекс цитирования отражает реакцию научного сообщества на публикации автора, поэтому его целесообразно использовать для оценки результатов фундаментальных исследований. Как правило, слабые работы не цитируют, кроме случаев особых взаимоотношений между авторами. Разные модификации индекса цитирования очищают его от «шумов», например, исключают самоцитирования и цитирования соавторами или требуют наличия в одной базе данных и статьи, и работ с ссылками на эту статью. Чтобы попасть в эту базу данных статьи должны соответствовать определенным требованиям. Однако и этот показатель не защищает от дробления результатов исследования по мелким статьям. Возможно, большой индекс цитирования обеспечивает только одна статья, которую ученый написал много лет назад и после этого ничего существенного не публиковал. Поэтому, только по одному индексу цитирования трудно достоверно определить уровень продуктивности ученого, который важен для приёма на работу, финансирования исследований и т. п.

Чтобы выявить ученых, пишущих много и качественно, в 2005 г. физик Х. Хирш предложил новый показатель – индекс Хирша [2]. Индекс Хирша или h-индекс – это максимальное целое число h , указывающее, что автор опубликовал h статей, каждая из которых процитирована хотя бы h раз. Эти h статей составляют ядро Хирша или h-ядро. Чтобы попасть в ядро Хирша, статью должны процитировать хотя бы h раз, поэтому не следует дробить результаты по нескольким публикациям. Простота расчетов и нечувствительность к типовым приемам искусственного улучшения вышерассмотренных показателей мгновенно сделали индекс Хирша популярным наукометрическим индикатором.

Рассмотрим двух авторов, получивших при минимально допустимых требованиях соседние значения индекса Хирша h и $h+1$. Количество цитирований этих авторов отличается существенно, а именно – на величину $(h+1)^2 - h^2 = 2h+1$. Поэтому по индексу Хирша можно просто кластеризовать авторов, например, на группы с одинаковыми его значениями. Однако целочисленность этого показателя не позволяет ранжировать авторов в пределах группы с одинаковым значением индекса Хирша. **Целью статьи** является разработка нового наукометрического показателя, который, наследуя все преимущества индекса Хирша и дополнительно позволит осуществить такое ранжирование. Новый показатель представляет собой дробную модификацию индекса Хирша. В нем целая часть является обычным индексом Хирша, а дробная – показывает, насколько близко автор приблизился к следующему значению индекса Хирша. В соответствии с наукометрическими традициями новый показатель назовем Sh-индекс.

Формализованная постановка задачи

Обозначим количество публикаций некоторого автора через N . Количество цитирований его публикаций отсортируем по убыванию и представим таким вектором:

$$C = (c_1, c_2, \dots, c_N), \quad (1)$$

где c_i – количество цитирований i -ой публикации, причем $c_1 \geq c_2 \geq \dots \geq c_N$, $i = \overline{1, N}$.

Задача исследования состоит в нахождении функционального отображения

$$C \rightarrow Sh_index \in [h(C), h(C)+1], \quad (2)$$

где Sh_index – дробная модификация индекса Хирша (Sh-индекс);

$$h(C) = \max_{\substack{i=1, N \\ i \geq c_i}}(i) \text{ – индекс Хирша.}$$

На отображение (2) наложим следующие ограничения:

(а) если индекс Хирша $h(C)$ набран при минимально возможном количестве цитирований, т. е. в случае $(c_1 = h, c_2 = h, \dots, c_h = h)$, тогда $Sh_index(C) = h(C)$;

(б) чем больше дробная часть Sh_index , тем ближе автор к увеличению своего индекса Хирша.

Новый Sh-индекс

Новый наукометрический Sh-индекс для вектора цитирований (1) определим таким образом:

$$Sh_index(C) = h(C) + \Delta, \quad (3)$$

где $\Delta = \frac{1}{h+1} \cdot \max_{\substack{j=1, h \\ c_j > h}}(j)$ – дробная часть Sh-индекса, показывающая долю наполненности

текущего h -ядра статьями, которые по числу цитирований входят и в следующее ядро.

Для удобства расчетов дробную составляющую (3) перепишем в таком виде:

$$\Delta = \frac{\sum_{j=1, h} r_j}{h+1}, \quad (4)$$

где $r_j = \begin{cases} 1, & \text{якщо } c_j > h \\ 0, & \text{інакше} \end{cases}$ – индикатор принадлежности статьи следующему ядру Хирша.

Пример. Векторы цитирований трёх авторов приведены в табл. 1. В ней все авторы имеют одинаковый индекс Хирша, равный $h = 4$. Чтобы увеличить его на 1, необходимо 5

статей, каждая из которых должна быть процитирована хотя бы 5 раз. У автора А – 4 таких статьи, у автор Б – 3 таких статьи, а у автора С – 0. Следовательно, оценки авторов по предложенному Sh-индексу составляют:

$$sh_index(\text{Автор } A) = 4 + \frac{4}{5} = 4.8;$$

$$sh_index(\text{Автор } B) = 4 + \frac{3}{5} = 4.6;$$

$$sh_index(\text{Автор } C) = 4 + \frac{0}{5} = 4.$$

Таблица 1

Векторы цитирования

Авторы	Номер статьи							
	1	2	3	4	5	6	7	8
Автор А	20	9	7	5	4	2	1	0
Автор Б	6	5	5	4	1	0	0	0
Автор В	4	4	4	4	0	0	0	0

Сравнение с конкурентными подходами

Чаще всего для ранжирования ученых с одинаковыми индексами Хирша используют дополнительные показатели, наиболее популярные среди которых сведены в табл. 2. Преимущество предложенного показателя состоит в том, что он обобщает индекс Хирша, тогда как показатели из табл. 2 используются с ним совместно. Преимущества и недостатки этих дополнительных наукометрических показателей проанализированы во многих статьях, например в [9, 10], поэтому мы на них не останавливаемся.

Таблица 2

Дополнительные наукометрические показатели

Показатель	Описание	Отношения	Источник
g-индекс	Максимальное число g самых популярных статей, получивших совместно не меньше g^2 цитирований. Учитывает превышение суммарного цитирования ядра Хирша минимального теоретического значения.	$h \leq g$	[3]
hg-индекс	Среднее геометрическое h-индекса и g-индекса: $hg = \sqrt{h \cdot g}$.	$h \leq hg \leq g$	[4]
e-индекс	Квадратный корень избыточного цитирования ядра Хирша h^2 : $e = \sqrt{\sum_{j=1, h} c_j - h^2}$.	–	[5]
A-индекс	Среднее число цитирований ядра Хирша: $A = \frac{1}{h} \sum_{j=1, h} c_j$.	$R = \sqrt{h \cdot A}$	[6]
R-индекс	Квадратный корень суммарного цитирования ядра Хирша: $R = \sqrt{\sum_{j=1, h} c_j}$.	$R = \sqrt{h \cdot A}$	[7]
m-индекс	Медиана количества цитирования ядра Хирша.	$h \leq m$	[8]

Ближайшим аналогом предложенного Sh-индекса является h_{rat} -индекс [11]:

$$h_{rat}(C) = h(C) + 1 - \frac{n}{2h(C) + 1}, \quad (5)$$

где $n = (h(C) + 1 - c_{j+1}) + \sum_{j=1, h} (1 - r_j)$ – минимальное количество цитирований для достижения следующего значения индекса Хирша;

$2h(C) + 1$ – теоретически минимальное количество цитирований для достижения следующего значения индекса Хирша ($h + 1$) в случае наихудших начальных условий, когда $(c_1 = h, c_2 = h, \dots, c_h = h, c_{h+1} = 0)$.

Сравнивая (3) и (5), видим, что целые части обоих показателей тождественны и соответствуют индексу Хирша. Дробные части различаются из-за разных принципов интерполяции отрезка $[h, h + 1]$. В (5) этот единичный отрезок разбивается на $2h(C) + 1$ равные части в соответствии с минимально необходимым количеством цитирований. Затем к h – количеству публикаций из текущего ядра Хирша – прибавляется доля выполнения плана цитирования по формированию следующего ядра Хирша. Следовательно, в h_{rat} существует противоречие между составляющими, поскольку одно слагаемое соответствует количеству публикаций, а второе формируется по количеству цитирований. В предложенном подходе этот недостаток отсутствует, т. к. и целая, и дробная части Sh-индекса (3) составлены по количеству статей. В (3) целая часть равна количеству статей из текущего ядра Хирша, а дробная – доле заполнения статьями следующего ядра. Другими словами, дробную часть Sh-индекса можно интерпретировать как долю выполнения плана по статьям для получения следующего значения индекса Хирша.

Выводы

Предложен новый наукометрический показатель – Sh-индекс. Этот показатель является дробной модификацией одного из самых популярных наукометрических показателей – индекса Хирша. В Sh-индексе целая часть представляет собой обычный индекс Хирша, а дробная показывает, насколько приблизился ученый к следующему значению индекса Хирша. Дробная часть рассчитывается как доля уже опубликованных статей, которые необходимы для получения следующего значения индекса Хирша. Наследуя все преимущества индекса Хирша, новый показатель дополнительно позволяет ранжировать авторов с одинаковыми индексами Хирша. Кроме того, по новому показателю можно отследить прогресс ученого в улучшении своего индекса Хирша за определённый промежуток времени.

СПИСОК ЛИТЕРАТУРЫ

1. Налимов В. В. Наукометрия. Изучения развития науки как информационного процесса / В. В. Налимов, З. М. Мульченко. – М.: Наука. – 1969. – 192 с.
2. Hirsch J. E. An index to quantify an individual's scientific research output / J. E. Hirsch // Proceedings of the National Academy of Sciences of the USA. – 2005. – Vol. 102, № 46. – P. 16569 – 16572.
3. Egghe L. Theory and practice of the g-index / L. Egghe // Scientometrics. – 2006. – Vol. 69, № 1. – P. 131 – 152.
4. Alonso S. Hg-index: a new index to characterize the scientific output of researchers based on the hand g-indices / S. Alonso, F. Cabrerizo, E. Herrera-Viedma, F. Herrera // Scientometrics. – 2010. – Vol. 82, № 2. – P. 391 – 400.
5. The e-index, complementing the h-index for excess citations [Электронный ресурс] / Zhang C.-T. // PLoS ONE. – Vol. 4, №5. – 2009. Режим доступа: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0005429>.
6. New developments related to the Hirsch index [Электронный ресурс] / Rousseau R. // Science Focus. – Vol. 1, № 4. – 2006. – P. 23 – 25. Режим доступа: <http://eprints.rclis.org/6376/>.
7. Liang B. J. L. The R- and AR-indices: complementing the h-index / B. J. L. Liang, R. Rousseau, L. Egghe // Наукові праці ВНТУ, 2011, № 3

Chinese Science Bulletin. – 2007. – Vol. 52, № 6. – P. 855 – 863.

8. Bornmann L. Are there better indices for evaluation purposes than the h-index? A comparison of nine different variants of the h-index using data from biomedicine / L. Bornmann, R. Mutz, H. Daniel // Journal of the American Society for Information Science and Technology. – 2008. – Vol. 59, № 5. – P. 830 – 837.

9. Alonso S. H-index: a review focused in its variants, computation and standardization for different scientific fields / S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, F. Herrera // Journal of Informetrics. – 2009. – Vol. 3, № 4. – P. 273 – 289.

10. Egghe L. The Hirsch-index and related impact measures / L. Egghe // Annual Review of Information Science and Technology. – 2010. – Vol. 44. – P. 65 – 114.

11. Guns R. Real and rational variants of the h-index and the g-index / R. Guns, R. Rousseau // Journal of Informetrics. – 2009. – Vol 3, № 1 – P. 64 – 71.

Штовба Сергей Дмитриевич – д. т. н., профессор кафедры компьютерных систем управления.

Штовба Елена Валериевна – к. т. н., доцент кафедры менеджмента и моделирования в экономике.

Винницкий национальный технический университет.