

Метод оценки показателей публикационной активности

А.С. Козицын, С.А. Афонин, Д.А. Шачнев

НИИ механики МГУ им. М.В. Ломоносова

Аннотация. В современном научном мире одним из способов оценки успешности научной деятельности ученого является вычисление различных показателей, основанных на количестве его публикаций и их цитируемости. При этом каждый соавтор публикации получает за нее одинаковое количество баллов. Подобный способ оценки приводит к искусственному увеличению количества соавторов, что, в свою очередь, влечет за собой искажение рейтинговых оценок научной деятельности в организации, а также значительно снижает качество тематического поиска экспертов, конференций и журналов по библиографическим данным. Представленный в работе метод позволяет оценить степень влияния указанного фактора на показатели, основанные на учете количества и цитируемости научных публикаций. Апробация проводилась на данных наукометрической системы ИАС ИСТИНА.

Ключевые слова: ранжирование, наукометрия, наукометрические системы, соавторство, системы цитирования, научный рейтинг.

Method for refining publication activity indicators

A.S. Kozitsyn, S.A. Afonin, D.A. Shachnev

Institute of Mechanics Lomonosov Moscow State University

Abstract. One way to evaluate the scientific activity of a scientist is to calculate various indicators based on the number of his publications and their citation. In this case, each co-author of the publication receives the same number of points for it. A similar ranking method leads to an artificial increase in the number of co-authors in our publications. This leads to a distortion of ratings, and to a significant decrease in the quality of thematic search. The method presented in the paper allows us to evaluate the contribution of the author to his published works. Approbation of the method was done on the data of the scientometric system IAS ISTINA.

Keywords: ranking, scientometrics, scientometric systems, co-authorship, citation systems, scientific rating, ranking, scientometrics, scientometric systems, co-authorship, citation systems, scientific rating

1. Введение

В современном научном мире одним из способов оценки успешности научной деятельности ученого является подсчет различных показателей, основанных на количестве его публикаций и их цитируемости. Наиболее распространенным показателем оценки деятельности ученого является h-индекс (индекс Хирша). Несмотря на его популярность, необходимо учитывать имеющиеся у него недостатки. Во-первых, индекс Хирша существенно зависит от области охвата анализируемых публикаций. В зависимости от выбранной системы цитирования (WoS, Scopus, Google Scholar, РИНЦ) для ученого будут получены разные результаты оценки. Во-вторых, этот индекс не учитывает тематическую направленность работ, которая оказывает существенное влияние на абсолютное значение индекса. Средний индекс цитирования в области физики и медицины в несколько раз превышает средний индекс цитирования в области математики и информационных технологий. В-третьих, индекс не учитывает самоцитирования и цитирования соавторов. В-четвертых, не учитывается авторитетность цитирующей статьи или издания, что особенно важно для систем с очень большим охватом разнородных источников, например, Goggle Scholar. На основе индекса Хирша было построено множество модификаций [1], создатели которых пытались в той или иной мере устранить перечисленные выше недостатки.

Следует отметить еще один существенный недостаток таких индексов. При их вычислении для всех соавторов статья учитывается одинаково. Это создает существенное искажение данных для коллабораций, когда в список соавторов включаются сотни и тысячи ученых (Atlas collaboration, Cms collaboration, Alice collaboration, Ecoteam4 и другие). Кроме того, при публикации статьи в список соавторов могут по разным причинам включаться авторы, которые в реальности не принимали существенного участия в написании статьи и получении описанных в ней научных результатов. Все эти факторы приводят к искусственному увеличению количества соавторов. По данным системы ИСТИНА среднее количество соавторов в статьях без коллабораций в период с 2012-2022 года увеличилось на 35% (Таб.1).

Таблица 1.

Распределение среднего числа соавторов статей за период 2010-2022

Год	Среднее число соавторов
2010	2.74
2011	2.88
2012	3.02
2013	3.05
2014	3.03
2015	3.13

2016	3.21
2017	3.18
2018	3.24
2019	3.34
2020	3.53
2021	3.65
2022	3.7

Для частичной коррекции ситуации возможно использование нормировки баллов на количество соавторов. Например, такой способ расчета предлагается методике расчета комплексного бала публикационной результативности (КБПР) Минобрнауки. Однако, КБПР применяется для расчета результативности работы организации в целом, и количество соавторов из одной организации никак не влияет на этот показатель. Реальная оценка персональной деятельности производится по другим критериям. В профиле автора РИНЦ указывается цитируемость и индекс Хирша без нормировки по соавторам. В формулах рейтинга МГУ нормировка на количество соавторов используется только в 50% случаев.

Поскольку каждый соавтор публикации получает за статью одинаковое количество баллов, представленные выше способы оценки цитируемости автора приводят к искажению рейтинговых оценок. Кроме того, ухудшается качество работы алгоритмов тематического поиска авторов.

2. Критерий оценки авторов

Одним из методов оценки вклада авторов в опубликованные им работы может являться оценка статистического распределения его позиции в списке соавторов статей. В работе [2] проводилась оценка достоверности гипотез о распределении авторов в библиографическом описании статьи. Рассматривалось два варианта: «Первый соавтор располагается по алфавиту» и «Первый соавтор располагается по вкладу в работу». Для оценки использовались данные наукометрической системы ИАС «ИСТИНА», в которой на настоящий момент подключено более 30 организаций, в том числе МГУ им. М.В. Ломоносова.

В работе [2] показано, что для статей, имеющих четырех и более соавторов, процент библиографических описаний с указанием первого автора на основе алфавита составляет менее 9%.

Этот факт необходимо учитывать как при разработке алгоритмов тематического анализа наукометрических данных [3], в том числе, алгоритмов поиска экспертов по заданным предметным областям [4,5] и тематического поиска с использованием графов соавторства [6], так и при оценке научной деятельности ученого.

Одним из наиболее простых критериев такой оценки являются количество и процент статей, в которых автор является первым соавтором. Однако, подобные критерии не чувствительны к количеству соавторов статьи. Для построения нового индекса оценки предъявлялись следующие требования:

для авторов, которые все статьи опубликовали без соавторов индекс должен быть равен 1;

для авторов, которые никогда не являлись первыми соавторами в библиографическом списке индекс должен быть равен 0;

для «первых» соавторов статей индекс должен увеличиваться с увеличением количества соавторов.

Рассматривалось два варианта индексов.

$$P(a) = \frac{\sum_{d \in D_a} (C(d)) \cdot |\{d \in D_a \mid ord(a,d)=1\}|}{|D_a|^2}$$

и

$$K(a) = \frac{\sum_{d \in \{d \in D_a \mid ord(a,d)=1\}} C(d)}{|D_a|}$$

где $C(d)$ - количество соавторов в статье d , D_a - множество статей автора a , $ord(a,d)$ - порядковый номер автора a в статье d .

Апробация предложенных индексов производилась на данных наукометрической системы ИСТИНА[7]. Для проверки зависимости индексов от первой буквы фамилии автора были рассчитаны индексы за 5 лет для всех сотрудников зарегистрированных в системе организаций, вся группа сотрудников разбита на подгруппы по первой букве фамилии и посчитан средний индекс в каждой подгруппе (таб. 2)

Таблица 2. Распределение индексов P и K по первой букве фамилии.

	Р	К		Р	К
А	1.54	1.35	П	1.2	0.99
Б	1.44	1.2	Р	1.21	1.02
В	1.3	1.12	С	1.16	1
Г	1.45	1.17	Т	1.15	0.98
Д	1.3	1.1	У	1.24	1.01
Е	1.28	1.12	Ф	1.22	1.05
Ж	1.22	0.99	Х	1.21	1.02
З	1.31	1.08	Ц	1.02	0.86
И	1.21	1.02	Ч	1.22	1.01
К	1.22	1.03	Ш	1.14	0.93
Л	1.16	1	Щ	1.08	0.93
М	1.22	1.02	Э	1.2	1
Н	1.23	1.04	Ю	1.09	0.95

Из приведенной таблицы можно видеть, что зависимость от первой буквы фамилии существует, но разница между суммами первых трех букв алфавита и последних трех букв алфавит составляет менее 30%. Можно сделать вывод, что индивидуальная оценка по такому индексу не является корректной, однако, ее можно использовать для агрегированных оценок по научным коллективам в целом.

3. Программная реализация

Программная реализация аналитических отчетов для получения агрегированных оценок различных срезов выполнена с использованием модуля SQLReport[8]. Срезы могут выполняться по подразделениям, должностям, научным коллективам и другим достаточно репрезентативным признакам.

В таб. 3 приведено распределение процента сотрудников по должностям в зависимости от значения индекса K . Для каждой должности в столбце «<0.25» указана доля занимающих эту должность сотрудников, для которых значения индекса K не превышает 0.25, в столбце «0.25-0.5» доля сотрудников, для которых значение индекса K находится в диапазоне от 0.25 до 0.5, и так далее.

Таблица 3. Гистограмма распределения для должностей.

Должность	<0.25	0.25-0.5	0.5-0.75	0.75-1	1-1.5	1.5-2	2-3	>3
Аспирант	20	4	8	7	35	8	11	8
Ассистент	17	7	12	13	33	6	7	5
Ведущий научный	22	11	11	12	23	10	7	4
Ведущий специалист	41	5	6	7	15	8	9	8
Доцент	14	5	9	14	42	8	5	2
Заведующий кафедрой	14	6	6	12	39	14	5	2
Заведующий лабораторией	24	11	11	17	18	7	6	6
Инженер 1-ой категории	30	6	6	10	17	11	11	9
Лаборант	28	4	4	6	25	6	17	10
Магистрант	17	2	5	8	46	6	7	9
Младший научный	19	6	8	9	22	11	14	11
Научный сотрудник	23	9	11	10	20	9	12	6
Преподаватель	7	2	9	15	58	4	4	1
Профессор	17	7	8	11	39	10	6	2
Специалист	30	7	11	6	16	6	11	12
Старший научный	22	10	13	10	22	11	8	4
Старший преподаватель	16	5	10	13	48	5	3	1

На основе анализа данных в таб. 3 важно отметить, что для всех должностей выделяются два значимых столбца. Столбец $K \in [1, 1.5)$ — сотрудники, фамилии которых оказываются в библиографическом списке на первой позиции немного чаще, чем при случайном распределении. $K < 0.25$ - сотрудники, которые имеют опубликованные статьи, но никогда или почти никогда не оказываются на первом месте в библиографическом списке статей. Для должностей из категории научно-вспомогательного персонала («Ведущий специалист», «Инженер 1-ой категории», «Лаборант», «Специалист») высокое значение показателя в этом столбце является вполне закономерным, поскольку эти сотрудники, как правило, обеспечивают техническое сопровождение научных исследований, а не проводят их самостоятельно. Однако, для остальных должностей большие значения в столбце $K < 0.25$ могут означать наличие определенных проблем с учетом публикаций и, как следствие, с распределением финансирования в соответствии с публикационной активностью сотрудников.

Следует отметить, что 1013 из 9 тысяч сотрудников научного и профессорского состава имеют более 2 статей (в среднем по 18 статей) за последние 5 лет, но при этом ни разу не являлись первым соавтором или единственным автором какой-либо публикации. Эти данные совместно с данными таблицы 1 о росте соавторов необходимо учитывать при построении наукометрических оценок для анализа результатов деятельности научной организации.

4. Заключение

Предложенные в работе индексы позволяют оценить степень участия авторов в написании статей. В силу наличия определенного влияния фамилии автора на значение индексов не рекомендуется применять подобные оценки для построения персональных рейтингов и проведения конкурсов. Однако, эти индексы могут применяться для построения аналитических срезов с агрегированием данных по различным условиям, а также для уточнения результатов тематического поиска, в том числе, с использованием графов соавторства, и могут являться дополнением к любым стандартным индексам цитирования.

Литература

1. URL:<https://science.bsu.by/index.php/info/indexes/h-index> (10.04.2023)
2. Козицын А. С., Афонин С. А., Шачнев Д. А. Метод оценки тематической близости научных журналов // Программная инженерия. - 2020. - № 6. - С. 335-341.

3. Козицын А. С. Алгоритмы тематического поиска данных в наукометрических системах // Программная инженерия. - 2022. - Т. 13, № 6. - С. 291-300.
4. Shachnev D. A. Searching for activity results and experts in a given subject area, taking results significance into account // Programmnaia inzheneriia. 2021. Tom. 12, No. 5. P. 260-266.
5. Козицын А.С., Афонин С.А., Шачнев Д.А. Алгоритм поиска по ключевым словам специалистов в заданной предметной области // Современные информационные технологии и ИТ-образование. 2021. Т. 17. № 1. С. 124-133.
6. Козицын А. С., Афонин С. А., Шачнев Д. А. Метод оценки тематической близости научных журналов // Программная инженерия. - 2020. - № 6. - С. 335-341.
7. Васенин В.А., Занчурин М.А., Козицын А.С., Кривчиков М.А., Шачнев Д.А., Архитектурно-технологические аспекты разработки и сопровождения больших информационно-аналитических систем в сфере науки и образования// Программная инженерия. 2017. Том 8, № 10. С. 448-455.
8. Afonin S., Kozitsyn A., Astapov I. Sqlreports: Yet another relational database reporting system // Proceedings of the 9th International Conference on Software Engineering and Applications. 2014. P. 529-534.

References

1. URL:<https://science.bsu.by/index.php/info/indexes/h-index> (10.04.2023)
2. Kozitsyn A. S., Afonin S. A., Shachnev D. A. Metod otsenki tematicheskoi blizosti nauchnykh zhurnalov // Programmnaia inzheneriia. - 2020. - № 6. - P. 335-341.
3. Kozitsyn A. S. Algoritmy tematicheskogo poiska dannykh v naukometricheskikh sistemakh // Programmnaia inzheneriia. - 2022. - V. 13, № 6. - P. 291-300.
4. Shachnev D. A. Searching for activity results and experts in a given subject area, taking results significance into account // Programmnaia inzheneriia. 2021. V. 12, No. 5. P. 260-266.
5. Kozitsyn A.S., Afonin S.A., Shachnev D.A. Algoritm poiska po kliuchevym slovam spetsialistov v zadannoi predmetnoi oblasti // Sovremennye informatsionnye tekhnologii i IT-obrazovanie. 2021. V. 17. № 1. P. 124-133.
6. Kozitsyn A. S., Afonin S. A., Shachnev D. A. Metod otsenki tematicheskoi blizosti nauchnykh zhurnalov // Programmnaia inzheneriia. - 2020. - № 6. - P. 335-341.
7. Vasenin V.A., Zanchurin M.A., Kozitsyn A.S., Krivchikov M.A., Shachnev D.A., Arkhitekturno-tekhnologicheskie aspekty razrabotki i

soprovozhdeniia bolshikh informatsionno-analiticheskikh sistem v sfere nauki i obrazovaniia// Programmnaia inzheneriia. 2017. V. 8, № 10. P. 448-455.

8. Afonin S., Kozitsyn A., Astapov I. Sqlreports: Yet another relational database reporting system // Proceedings of the 9th International Conference on Software Engineering and Applications. 2014. P. 529-534.