Evaluation of Authors' Academic Influence Based on Weighted h-Index

Ganbin Shen
College of Science, Wuhan University of Science and Technology, Wuhan China
1820386429@qq.com

Juan Chen
College of Science, Wuhan University of Science and Technology, Wuhan China
772145477@qq.com

Junan Lu
School of Mathematics and Statistics, Wuhan University, Wuhan China
jalu@whu.edu.cn

ABSTRACT

The h-index is a commonly used mixed quantitative index used to evaluate the academic level of the authors of the paper. But in practice the h index is insensitive to high-quality papers in all papers published by scientists. By giving weight to high-quality papers, this paper puts forward a new index: h_w index. It makes up for this deficiency. The h_w index is improved on the basis of the traditional h-index which gives weight to different levels of papers published by scientists and evaluates the academic influence of scientists with a weighted h index. Obtaining JCR partition data and paper citations from the WOS database of scientists in different fields such as mathematics, computer science, medicine, etc.And Calculation and comparison the difference between h_w index and h-index. Empirical analysis shows that the h_w index can evaluate the academic level of scientists steadily, has a better differentiation from the h index, and the h_w index can better highlight the index level of scientists with high paper quality and can appropriately reduce the index level of scientists with lower paper quality.This paper proposes a new h_w index.While retaining the characteristics of simplicity, ease of use and easy calculation, it is effective, reasonable and can be popularized in practical application.

CCS CONCEPTS

• **Human-centered computing**; • **Collaborative and social computing**; • **Collaborative and social computing devices**;

KEYWORDS

Weighted h-index, Journal influence factor, Academic influence evaluation, Number of papers cited

ACM Reference Format:
Ganbin Shen, Juan Chen, and Junan Lu. 2021. Evaluation of Authors' Academic Influence Based on Weighted h-Index. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3487075.3487148

CSAE 2021, October 19–21, 2021, Sanya, China

ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487148

1 INTRODUCTION

As a link in scientific research activities, academic level evaluation is also a very important and indispensable link. Due to the differences between the research levels of different scientists and the popularity of research fields, it is very necessary to use a reasonable, fair, simple and efficient index to evaluate the academic level of different scientists [1]. This is very helpful for the rational allocation of scientific research resources and the promotion of scientists' research level. H-index is a commonly used evaluation index at present. Professor Hirsch J.E. of the Department of medicine of University of California San Diego first proposed the concept of h-index in his paper published in 2005. A scientist's h-index is that h papers in his N papers are cited at least h times [2]. The definition of h index can be illustrated in Figure 1. It estimates the impact and importance of scientists' cumulative research contributions.After h-index was put forward, it has been widely recognized and used. This is largely because h-index is simple in concept and easy to calculate, but at the same time, it can provide a strong assessment of the influence of scientists' long-term accumulated research results. However, with the deepening of practical application, some shortcomings of h-index are gradually exposed, For example, the h-index is not sensitive enough to high-quality papers. It depends on the academic career of scientists. The h-index will only grow with time, and has a poor response to the recent academic influence of scientists. Then scholars from all over the world put forward other varieties of H index. Egghe [3] put forward g-index in 2006. G-index means that the number of citations of at least g papers is not less than G^2 , that is, the number of citations corresponding to the $G + 1$ paper will be less than $(G + 1)^2$. According to Egghe, the advantage of h-index is that it is not sensitive to papers with low citations, but it should be more sensitive to all papers with high citations. G-index can well reflect papers with high citations and the impact of the increase of citations on the academic influence of authors [4, 5]. M-quotient was put forward by Hirsch J E [3]. M-quotient is defined as the quotient of H divided by the academic career time. M-quotient is proposed to solve the problem that h-index is not sensitive to time, that is, h-index will only grow or remain unchanged with the passage of time. By dividing the academic career of scholars, the influence of time length on h-index can be eliminated, so as to better highlight the potential young scholars. M. Kosmulski [6] put forward the H (2) index in 2006, which is defined as the maximum natural number that makes each H (2) article with the highest citation frequency not less than $[H (2)]^2$ times. The purpose of H (2) index is to enhance the sensitivity of h index to highly cited articles, that is, to highlight its value by giving more weight to highly cited articles. Hirsch J E [2] put forward in

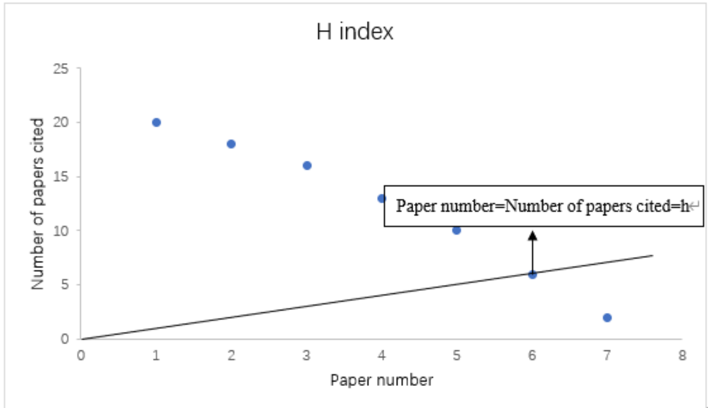

Figure 1: Definition of h-Index.

2010 that the definition of hbar index is to include other articles into the hbar core on the basis of H index. The composition of the hbar nucleus is: independent author; H index is higher than that of the co authors; The h-index is lower than that of the collaborators, but the citation frequency of the papers is higher than that of the collaborators. The proposed index of hbar mainly considers the contribution of collaborators [7], but it is obvious that its calculation is more complex and needs more information.The indexes proposed in these literatures either focus on the impact of optimizing academic career [8-13] or the impact of highly cited papers [14-19], but none of them focus on the impact of paper quality on the index.In this paper, we propose an improved h-index with weight, which is called h-index, based on the weight theory [20, 21], aiming at the problem that h-index is not sensitive to high-quality papers_ W index. It is very necessary to improve the sensitivity of H index to high-quality papers. According to the statistics of nature, in 2005, 25% of the papers in nature got 89% of the total citations, and the other 75% only got 11% of the citations [22]. That is to say, most of the papers published in journals such as nature have been cited for a limited number of times, and these papers are likely to be ignored in the calculation of h-index, which is unfair to the evaluation of the author's academic influence. By giving reasonable weight to better highlight the overall high-level scientists, it can also ease the situation of "winning by quantity".

2 h_w INDEX

2.1 Design Weight

In order to study the influence relationship of journals in different regions, the journal influence factor IF is used to compare. Generally speaking, the greater the impact factor, the greater its academic influence. Due to the different development levels of different disciplines, the average impact factors of journals in different disciplines are quite different. Therefore, IF cannot directly compare the influence of journals of different disciplines. JCR partition is based on the ranking of impact factors of all journals under a certain discipline classification. The top 25% journals are Q1 area, 25% - 50% journals are Q2 area, 50% - 75% journals are Q3 area, and 75% journals are Q4 area. For certain journals that belong to multiple

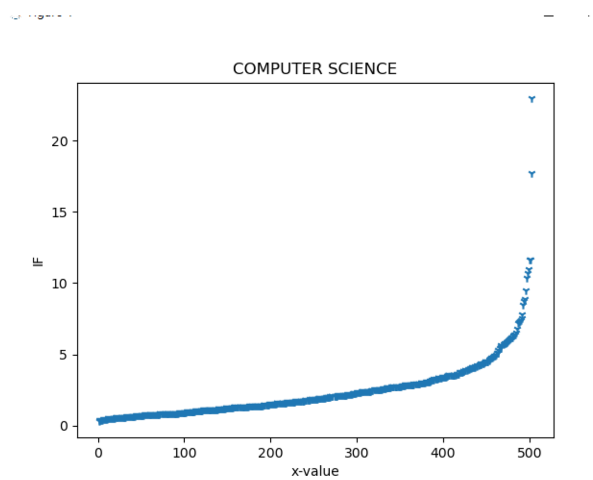

Figure 2: The Distribution of Influencing Factors in Computer Science Journals.

fields, we select the research field in which the author publishes the paper to determine the journal field that is located in multiple fields at the same time.According to the impact factor IF of 12558 journals obtained from Web of Science (WOS), calculate the integral ratio of Q1, Q2, Q3 and Q4 to determine the weight ratio. Taking the field of computer science as an example, the influencing factors of journals in the field of computer science are collected. The influencing factors of journals are arranged from large to small as shown in Figure 2

IF is a discrete distribution, so the trapezoidal method is used to calculate the integral of different regions.

$$S = \int_a^b f(x)\,dx \approx \sum_{i=1}^{n} S_i \tag{1}$$

In Eq (1),x_i is the journal whose influence factors rank i, and $f(x_i)$ is the influence factor of journal x_i. S_i is the area of the nth partition (Q1, Q2, Q3, Q4). Therefore, Eq (1) can be used to calculate the area of each partition in the IF curve as shown in Figure 3

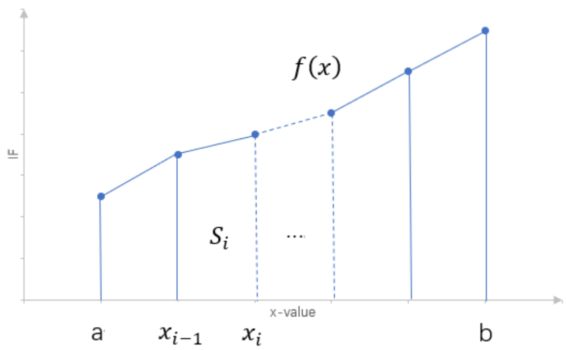

Figure 3: Trapezoidal Method.

The weight ratio can be determined by determining the integral ratio of different zones:

$$W_1 W_2 W_3 W_4 = S_1 : S_2 : S_3 : S_4 \tag{2}$$

H-index can also be regarded as the equal weight case of h_w index.

$$W(q_i) = \begin{cases} 1 & q_i = Q4 \\ 1 & q_i = Q3 \\ 1 & q_i = Q2 \\ 1 & q_i = Q1 \end{cases} \tag{3}$$

By calculating the area ratio, the journal partition weight ratio of computer science field is 96:160:308:640.After determining the weight ratio, we need to determine the benchmark. Obviously, using Q1 and Q4 as the benchmark cannot achieve our goal. If Q2 is used as the benchmark, the weight ratio is: 0.3:0.5:1:2. This means that only when authors publish articles that belong to the Q1 area will they have a beneficial effect on the h_w index, and will reduce the h_w index of almost the vast majority of authors, making the h_w index unable to achieve the desired effect.So,in order to keep the level of h_w index and h index relatively stable, the weight coefficient is reduced to the interval of [0,4], and the weight of Q3 area is set to 1, that is, taking Q3 area as the benchmark, h_w index improves the weight of articles in Q1 and Q2, and reduces the weight of articles in Q4 area. In order to calculate the weight coefficient accurately to 0.1

$$W(q_i) = \begin{cases} 0.6 & q_i = 4 \\ 1.0 & q_i = 3 \\ 1.9 & q_i = 2 \\ 4.0 & q_i = 1 \end{cases} \tag{4}$$

2.2 Definition of h_w Index

Suppose an author publishes N articles. When calculating its h_w index, a new standard is set for the number of citations C_i of the i-th paper. In order to make the new h-index reflect the quality of published articles, the weight W_i is introduced in the calculation of C_i. The weighted C_i is denoted as C_i.

$$C_i (C_i, W_i) = C_i \times W_i \tag{5}$$

In Eq (5), W_i is the weight of the partition where the article is located. Considering that the h index is mostly annual statistics, and the JCR division of the journal may change in different years, and some journal divisions may change, but in general these changes are relatively subtle, and we select the Q3 area as the benchmark. The h_w index can also be kept stable, so in order to facilitate data collection and calculation, the partition data obtained by the h_w index is set as the partition data of the statistical year. h_w index is the weighted h index. The h_w index of a scientist is the maximum of m papers in N papers published by him, if the number of citations C with weight in each paper is not less than m times, then m is the value of h_w index.

$$h_w(m) = max\{m : C_i \geq m \,; i \leq m\} \tag{6}$$

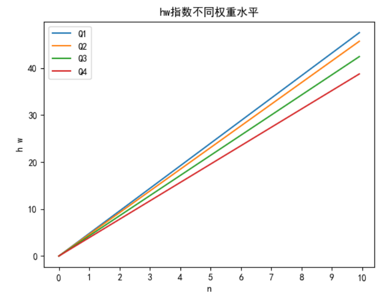

Figure 4: The Level of Growth under Different Weights of the h_w Index.

2.3 The Contribution of Weight to h_w Index

Assuming that an author publishes p articles steadily every year, and each published paper receives c new citations every year, the total number of citations obtained by the author after n years S_c is

$$S_c = \sum_{j=1}^{n} pcj = \frac{n(n+1)pc}{2} \tag{7}$$

Assuming that all papers up to year y contribute to h_w index, there is

$$wc(n-y) = h_w \tag{8}$$

$$py = h_w \tag{9}$$

Where Eq (7) is the number of citations of the latest papers contributing to the h_w index, and Eq (8) is the total number of papers contributing to the h_w index. It can be obtained from equation Eq (7) ,Eq (8).

$$h_w = \frac{wc}{1 + wc/p} * n \tag{10}$$

The h_w index is linearly related to time n under different weight levels, and it grows with the growth of academic career. This point should be quite common for those scientists who publish stable quality papers at a steady rate during their academic careers. And very reasonable. The relationship between h_w index and n can be expressed as

$$h_w \sim a * n \tag{11}$$

The growth rate of the h_w index will accelerate as the quality of the papers published by authors in their academic careers increases. As shown in Figure 4, for the scientists with the top distribution of papers, the value of his h_w index and the annual growth rate are better than those with the low distribution of the papers when the publication speed of the papers is relatively stable.

Table 1: Some Authors in the Field of Mathematics h Index with h_w Index

Name	h	$\mathbf{h_w}$	variety
Bai, ZD	16	17	+1
Guo,SJ	16	28	+12
Chen, FD	30	54	+24
Xu, DY	25	38	+13
Wang, WD	25	37	+12
Wang, JB	33	57	+24
He, JH	32	42	+10
Sun, JT	30	52	+22
Xu,JM	16	14	-2
Li, YK	25	38	+13

Table 2: Some Authors in the Field of Computer Science h Index with h_w Index

Name	h	$\mathbf{h_w}$	variety
Freitas, Alex A	25	26	+1
Andre Schiper	24	21	-3
Ben Taskar	24	24	+0
Christian Bizer	24	26	+2
Pierre Vandergheynst	33	43	+10
Rahul Sukthankar	27	29	+2
Maurice Heemels	18	23	+5
Carlo Zaniolo	11	10	-1
Mario Piattini	28	26	-2
Peter Meer	43	44	+1

3 EMPIRICAL ANALYSIS

This paper selects scholars in the fields of mathematics, computer science and medicine as the empirical research objects, and uses WOS (web of Science) database as the data source. In order to verify the stability of h_w index and the difference between h_w index and h index, the authors' published papers were obtained from the database, and their h index and h_w index were calculated.

3.1 h-Index and h_w Index in Computational Mathematics

In Table 1, the h_w index of Chen,FD and Wang, JB increased from 30 to 54 and 33 to 57, respectively. By comparing the distribution of articles in Figure 5, it is found that most of their papers are in the first quarter and the second quarter, and their articles are of high quality. In the h_w index, the desensitivity of the h index to its product quality has been improved. At the same time, the h_w index of jmxu was lowered and its distribution was observed. Most articles are distributed in Q3 and Q4. Author Bai, ZD's articles are mostly distributed in the Q3 area, so the index remains relatively stable.

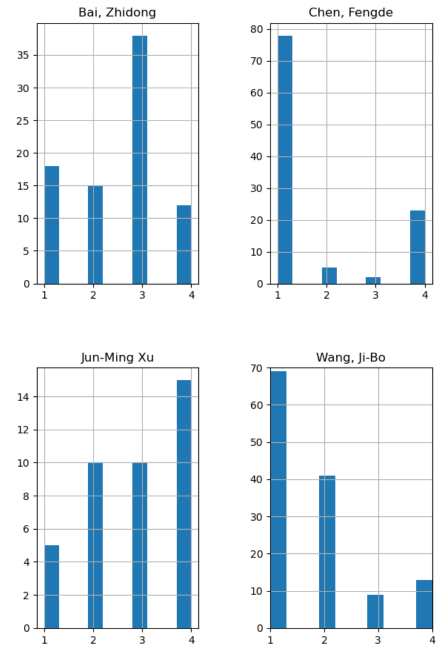

Figure 5: Article Distribution.

3.2 h Index and h_w Index in Computer Science

In Table 2, comparing the partition status of Andre schiper, Ben taskar and Christian bizer authors with the same h-index of 24, as shown in Figure 6. We can find that h_w index can better respond to the high-quality papers of authors, promote the authors with higher partition distribution, and appropriately reduce the ones with lower partition distribution. Among them, Andre schiper's papers were mainly distributed in Q3 and Q4, with the h_w index adjusted from 24 to 21. Ben taskar's papers were mainly distributed in Q3, with the h_w index stable at 24. Christian bizer's papers were in Q1 and Q2, with the h_w index adjusted from 24 to 26. At the same time, the index change of Pierre vandergheynst was adjusted from 33 to 43. By observing the regional distribution of his papers, we can find that Pierre vandergheynst is mainly concentrated in Q1 and Q2, that is to say, Pierre vandergheynst's papers have a high level. Ben taskar's papers mainly focus on Q3, and the $W(q_i)$ value of Q3 is 1, so its index level has not changed.

3.3 h Index and h_w Index in Medical

As shown in Figure 7, the distribution of author Emsley, P's papers is concentrated in the Q1 area, and the h_w index increased from 22 to 26 in Table 3, which better reflects the quality of his articles; the papers of authors Hilgen and Gerrit are relatively evenly distributed, so The h_w index has a slight change; the author Rajakulendran, Sanjeev's articles in Q1 also increased the level of his h_w index; the author Yousaf, Z, the article distribution is mostly concentrated

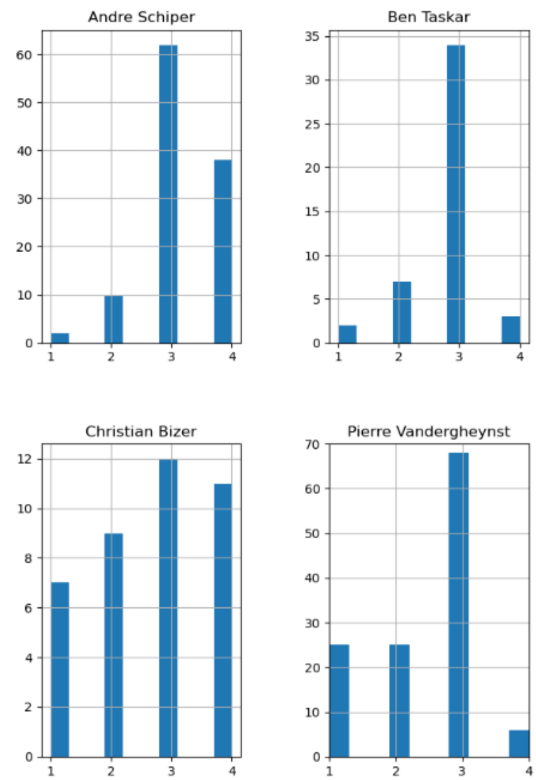

Figure 6: Article Distribution.

Table 3: Some Authors in the Field of Medical h Index with h_w Index

Name	h	$\mathbf{h_w}$	variety
Hilgen, Gerrit	11	12	+1
Yousaf, Z	33	29	-4
Emsley, P	22	26	+4
Rajakulendran, Sanjeev	11	14	+3
Tumasyan, A	101	112	+11
Khachatryan, V	99	100	+1
Chatrchyan,S	79	73	-6
Schael,S	108	110	+2
Addy, T. N.	75	72	-3
Aat, E	10	11	+1

in the Q3 area, the Q4 area, the h_w index is also adjusted down appropriately, from 33 To 29.

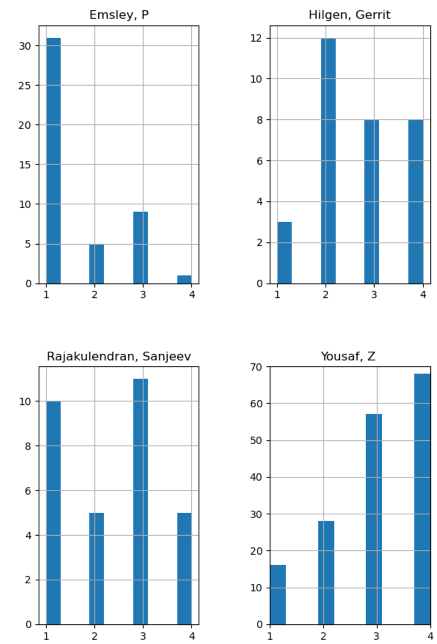

Figure 7: Article Distribution.

4 CONCLUSION

The h index is a relatively complete academic evaluation index, but there are also problems such as low sensitivity to high-quality papers. It is obviously unfair to use the h index to evaluate the academic influence of authors, because the h index may ignore some high-quality papers. Aiming at this defect of the h index, this paper proposes the h_w index on the basis of maintaining the simplicity and ease of use of the h index, and applies it to mathematics, computer science and medicine for empirical analysis. The results show that the h_w index can better respond to scientists with higher paper quality and give a certain improvement, and make certain adjustments to lower quality papers. At the same time, it also shows that the h_w index is stable in different fields and can stably reflect the quality of the author's article. The results of the empirical analysis are in line with expectations and relatively satisfactory. To sum up, the h_w index has the characteristics of simplicity, ease of use and easy calculation of the h index, and it pays more attention to high-quality papers while reducing the contribution of lower-quality papers to the h_w index. The h_w index It is effective, reasonable and can be promoted in actual application. However, because the h_w index and h index depend on the citation of the paper as the measurement standard, the citation of a paper is a relatively complicated process, which includes the influence of space, time and other factors. There are many factors that can be considered for the improvement of the h index. And will continue to develop and progress in the future.

ACKNOWLEDGMENTS

Thanks to Professor Chen and Professor Lu for their instructive opinions and recommendations on the research direction of my thesis.

REFERENCES

[1] Hirsch J E (2005). An index to quantify an individual's scientific research output[J]. Proceedings of the National Academy of Sciences of the United States of America, 102(46): 16569-16572.

[2] Hirsch J E (2010). An index to quantify an individual's scientific research output that takes into account the effect of multiple coau- thorship[J]. Seientometrics, 85(3): 741-754.

[3] Egghe L (2006). Theory and practise of the -index[J]. Scientometrics, 69(1): 131-152.

[4] Braun T, Glanzel W, Schubert A (2006). A Hirsch-type index for journals[J]. Scientometrics, 69(1):169-173.

[5] Bihui J, Liming L, Rousseau R, *et al* (2007). The R- and AR-indices: Complementing the h-index[J]. Chinese Science Bulletin, 52(6): 855-863.

[6] Kosmulski M (2006). A new Hirsch-type index saves time and works equally well as the original h-index [J].ISSI Newsletter, 2(3): 4-6.

[7] Egghe L, Rousseau R (2006). An informetric model for the Hirsch-index[J]. Scientometrics, 69(1):121-129.

[8] Vieira E S, Gomes J (2011). The journal relative impact:an indicator for journal assessment[J].Scientometrics, 89(2): 631–651.

[9] Seglen P O (1997). Why the impact factor of journals should not be used for evaluating research [J]. BMJ, 314(2): 498–502.

[10] Prathap G (2010). Is there a place for a mock h - index[J]. Scientometrics, 84(1): 153–165.

[11] Jan R, Ahmad R (2020). H-Index and Its Variants: Which Variant Fairly Assess Author's Achievements[J]. Journal of Information Technology Research, 13(1):68-76.

[12] Karadag E (2020). Academic (dis)qualifications of Turkish rectors: their career paths, H-index, and the number of articles and citations[J]. Higher Education, (3):1-23.

[13] Ain Q U , Riaz H , Afzal M T (2019). Evaluation of h-index and its citation intensity based variants in the field of mathematics[J]. Scientometrics, 119(1):187-211.

[14] Prathap G (2011). Qualifying Scholarly Impact using an iCX (impact-Citations-Exergy) Analysis [J]. Desidoc Journal of Library & Information Technology, 31 (5):382-386.

[15] Prathap G (2012). Evaluating journal performance metrics[J]. Scientometrics, 92(2): 403-408.

[16] Prathap G (2010). The 100 most prolific economists usingthe p-index[J]. Scientometrics, 84(1): 167-172.

[17] Xinjun Xu (2020). Application of Z-index in Periodical Network Transmissibility [J]. Journal of Academic Libraries, 38(6): 111-117.

[18] Liping Yu, Yaru Han, Zaijie Zhang (2020). Research on h-index Optimization Considering the Amount of Papers–hq2-index[J]. Modern Information, (2): 114-121.

[19] Chao Wang, Chaoqun Peng (2016). Efficiency analysis of P index for the evaluation of Chinese SCI-covered journals[J]. Evaluation & Management, 27(8): 893–899.

[20] Pu Zhang, Xinnian Wu, Xuefei Chen (2012). The Development and Research Route About the h-type Indexes[J]. Document,Information & Knowledge, (1):93-101,108.

[21] Zheqian Chen, Xiaotong Yu, Najing Zheng ,et al (2020). Construction and Application of Evaluation Index for Public Cultural Cloud Service[J].Document, Informaiton & Knowledge, (6):54-66.

[22] C Zhang (2010). Assessing academic assessment[J].Bulletin of National Natural Science Foundation of China, 24(06): 328–332.