Accumulated Citation Count as Fertileness of Scientific Article

Sachio Hirokawa Research Institute for Information Technology, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka, 812-8581, Japan Email: hirokawa@cc.kyushu-u.ac.jp

Hayato Nakanishi Graduate School of Integrated Frontier Sciences, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka, 812-8581, Japan Email: hayato.nakanishi.460@s.kyushu-u.ac.jp

Abstract—The literature survey by scientific bibliographic data base is indispensable in the research activities. We can find related articles with appropriate keywords. However, the threads of related research are not easy to grasp from the search result. It is necessary to repeat a search, judge a citation relation and figure out the thread. The present paper proposes the index "accumulated citation count" of a scientific article to measure the thread of citations that starts from the article.

I. INTRODUCTION

It is important to find related appropriate articles for a scientific activity. Data bases of scientific articles have been expanding their contents and functions to respond the requirement. Search with appropriate keywords returns related articles. The articles in the search results are sorted by the relevance to the query or by the citation count of each article. A researcher has to check articles in turn to gasp the situation of related work.

Citation count of an article is the total number of articles that cite the article. It is one of the widely used objective bibliometric index. Citation count evaluates the direct linkage between articles. The present paper expands the citation count for threads of citation linkage. Imagine the process how a new idea and a new scientific notion spread from the original article to other articles. The idea would be developed in the first generation and then further developed and modified or might generate another new ideas in the descendant. It may be the case that some descendant do not have direct citation to the original article. The simple citation count does not capture the size of the whole tree of citation linkage.

The present paper proposes an index to evaluate the influence of an scientific article which generated the thread of citations. The idea is to evaluate the fertileness. We would like to call an article as fertile research when the article is not only highly evaluated but also there are many highly evaluated articles that cite the article. Articles which cause the ripple effect like avalanche should be evaluated. An article with a small citation count should be highly evaluate if there are Tetsuya Nakatoh Research Institute for Information Technology, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka, 812-8581, Japan Email: nakatoh@cc.kyushu-u.ac.jp

many good articles that cite the article or if there are many good articles in the descendant. We would like to call such an article as the origin of the flourished research field. Conversely, event hough an article had a large number of citation count, the influence of the article is limited in one generation if there were no descendant.

The present paper proposes "accumulated citation count" as a bibliometric index to evaluate scientific articles. The accumulated citation count sums up the citation counts among all the descendant articles in the thread of citation linkage that begins from the article. Comparisons of the accumulated citation count and the conventional citation count are shown.

II. RELATED WORK

Citation count is a useful measure for search of scientific articles. The analysis of citation is one of the major fields of bibliometrices [16] and has a large number of research. Martin [11] reported that the cation count is recognized as useful measure by many researchers. Scopus ¹ is one of the most used databases that provides the sorted listing by citation count. On the other hand, Kostoff [9] raised the following problems of citation count.

- (a) The citing article and the cited article do not always have reasonable relation.
- (b) Incorrect work may be cited too much.
- (c) Methodological articles tend to be highly cited.
- (d) Self-citation may cause inflation of citation count.

Nakatoh et. al [13] solved (a) by restricting the article to the target field when considering the citation count. Another problem of the citation count is that newly published articles have no citation count. So, the citation count cannot be applied as the evaluation of the latest research. It will take a few years to settle the evaluation of an article as the increase of citation count. Nakanishi et.al [15] analyses the cause of steep increase of citation count to predict the new trends of research.

¹http://www.scopus.com/



Evaluation of the journal is reasonable measure to select appropriate scientific articles. Journal Impact Factor [4], [10], [5] is the most popular measure for journals. To get to know the ranking of journals may be one of the first steps to be a professional of a particular field. However, Pudovkin and Garfield [17] warns that the comparison of the journal impact factors among different genres is not appropriate. There are normalization approaches of journal impact factor to overcome this problem [10], [17]. Bergstrom [1] proposed EigenFactor by adjusting the weight of citation linkage to get rid of the problem. Nakatoh et.al [12] proposed the focused impact score by restricting the citing articles in counting the cation. Evaluation of researcher is a related research subject. Hindex [6] is a measure of researcher which evaluates the impact of a researcher based on the number of publication and the citation count.

Seach engine is useful to search for related articles of particular research field. However, interpreting and grasping the search result are not easy task yet. Garfield et.al [3] constructed the network diagrams of citation and analysed the history of science. Igami and Saka [8], [18] proposed Science Map to visualize co-citation and analysed the trends of scientific research. Nakatoh et.al [14] constructed a visualization system of citation graph and applied the sytem to view the citation diagrams of articles related to bibliometrics. The purpose of the present paper is quantification of citation diagram and the evaluation of articles based on citation relation.

III. ACCUMULATED CITATION COUNT

The present paper gives a formulation of fertileness of a scientific article. Citation count is a good start point as evaluation index of articles. A good article would be cited by many articles and hence have a high citation count. A fertile article should be good article with high citation count. However, a good article is fertile only when the articles that cite the article are good as well. We would like to formulate the evaluation of an article not just by the citation of the article but with the evaluation of related, i.e., citing articles.

We would say that a paper is fertile not because that it is cited a lot but because that the research field is developed after another papers which cite the paper and after the papers which cite those papers. When the influence of the paper is transfered and expanded to the next generation of citation linkage, the origin of the thread is worthwhile being called fertile.

On the other hand, if the linkage terminates in one step, then the article is not recognized as fertile even if the citation count is very large. Such an article has an impact at the time. But, the influence did not last and failed to establish a new research field. The citation graph of such paper would look like an umbrella. The citation graph of fertile paper would look like a big tree with thick trunks and many branches.

Accumulated citation count is a formulation of this idea. The **Accumulated Citation Count** $ACC(a_i)$ of an article a_i is the sum of the citation count $CC(b_j)$ of articles $b_j(j = 1, \dots, n)$ that cite the article a_i .

IV. CITATION DATA BASE FOR EVALUATION

To conduct empirical evaluation of the accumulated citation count, we constructed a citation database for articles related to "bibliometrics". We collected 10,186 articles that contain the keyword "bibliometrics" published from 1976 to 2015. The data of articles were retrieved from Scopus ². There are 3,024 articles that are not cited by other article. There is an article that is cited by 2,977 article. The total citation count is 116,743.

These data are written in JSON format which consists of the following fields: "Content Type", "Search identifier", "Complete author list", "Resource identifiers", "Abstract Text", "First author", "Page range", "SCOPUS Cited-by URI", "Result URL", "Document identifier", "Publication date", "Source title", "Article title", "Cited by count", " ISSN", "Issue number" and "Volume."

Table I shows the top 10 articles with respec to citation count of the article. We see three articles by Garfield which are famous in bibliometrics.

V. CITATION COUNT VS ACCUMULATED CITATION COUNT

We use the focused citation count (FCC) introduced by Nakatoh et.al [13] as the citation count $cc(a_i)$ of an article. The focused citation count of an article a_i is the number of articles which cite a_i and contain the keyword, in this experiment, "bibliometrics". FCC is a modified version of citation count. By restricting the articles to contain the keyword, highly related articles in the field gain high score of citation. Nakatoh et.al [13] reports that FCC is effective to select more appropriate papers.



Fig. 1. Citation Count vs Accumulated Citation Counts

Fig. 1 plots the citation count (CC) and the accumulated citation counts (ACC) of the top 500 articles sorted by citation count. We see the positive correlation. The larger the citation

²http://www.scopus.com/

count, the larger the accumulated citation count. However, we can see there are many articles of high ACC whose CC are low. We see a few articles of high CC with low ACC. Articles with high ACC compared with CC may be considered as influential fertile articles. In other words, the ratio of ACC/CC may be a good measure to evaluated the ripple effect.

Table II shows the top 10 articles with respect to accumulated citation count. The first column displays the rank of each article with respect to citation count. We can see that there are 4 articles Rousseual1997, Cronin2001 Moed1996 and Moed1995 ranked up in top 10 from out of the top with respect to CC. We would not find those articles when we were checking according to the ranking in CC.

Table III shows the top 10 articles with respect to ACC/CC. If we pay attention to the ranks of those articles, we see that they are out of top 100 with respect to CC. In fact, their CC's are less than 40. However, they have 20 or 30 times ACC compared with their CC.

VI. CONCLUSION

Search of appropriate articles is crucial to scientific activities. Quantitative evaluation of articles is one of the key issues in bibliometrics. The present paper proposed a novel measure of "accumulated citation count" of an article. The measure captures the quantity of citation graph given an article. Comparison of the proposed accumulated citation with the focused citation count were carried out on 10,186 articles related to bibliometric. It turned out that there are 4 articles ranked within top 10 with respect to ACC but ranked out of top 10 with respect to CC. Thus, the proposed measure is effective to find fertile articles that would be missed in conventional ranking.

The ranking by ratio ACC/CC yields quite different list. Further analysis of citation graph of each article is necessary to confirm if they are really valuable paper or if there were any reason that caused the increase of ACC.

ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI Grant Number 24500176 and 15K00426.

REFERENCES

- C. Bergstrom, "Measuring the value and prestige of scholarly journals," in *College & Research Libraries News*, 2007, vol. 68, No. 5, pp.314– 316.
- [2] M. Callaham, R. L. Wears, E. Weber, "Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals," in *Journal of the American Medical Association*, 2002, Vol. 287, No. 21, pp.2847–2850.
- [3] E. Garfield, I. H. Sher, R. J. Torpie, "The Use of Citation Data in Writing the History of Science," in *Institute for Scientific Information*, in *Philadelphia*, 1964.
- [4] E. Garfield, "Citation indexes for science," in *Science*, 1955, Vol. 122, No. 3159, pp.108–111.
- [5] E. Garfield, "The history and meaning of the journal impact factor," in *Journal of the American Medical Association*, 2006, Vol. 295, No. 1, pp.99–93.
- [6] J. E. Hirsch, "An index to quantify an individual's scientific research output," in Proc. of *the National Academy of Sciences of the United States of America*, Vol.102(46), pp.16569–16572, 2005.

- [7] J. E. Hirsch, "Does the h index have predictive power?" in Proc. of the National Academy of Sciences of the United States of America, 104(49), pp.19193–19198, 2007.
- [8] M. Igami, A. Saka, "Observation of the evolution of science via the Science Map : Methodology and application," in Journal of *Information Processing and Management*, (2009, Vol. 52, No. 5, pp.255–266.
- [9] R. N. Kostoff, "Performance measures for government-sponsored research: Overview and background," *Scientometrics*, 1996, Vol.36, No.3, pp.281–292.
- [10] Marshakova-Shaikevich Irina, "The standard impact factor as an evaluation tool of science fields and scientific journals." *Scientometrics*, 1996, Vol.35, No.2, pp.283–290.
- [11] B. R. Martin, "The use of multiple indicators in the assessment of basic research," *Scientometrics*, 1996, Vol.36, No.3, pp.343–362.
- [12] T. Nakatoh, H. Nakanishi, and S. Hirokawa, "Journal Impact Factor Revised with Focused View," in Proc. of 7th KES International Conference on Intelligent Decision Technologies (KES-IDT 2015), 2015, pp.471– 481.
- [13] T. Nakatoh, H. Nakanishi, and S. Hirokawa, "Focused Citation Count: A Combined Measure of Relevancy and Quality," Proc. of *IIAI 4th International Congress on Advanced Applied Informatics (IIAI AAI* 2015), 2015, pp.12–16.
- [14] T. Nakatoh, H. Nakanishi, and S. Hirokawa, "Threads and History of Bibliometrics," Proc. of *ICADL2015*, 2015. (to appear)
- [15] H. Nakanishi, T. Nakatoh, and S. Hirokawa, "Cause Analysis for Steep Increase of Citation," Proc. of *The 10th International Conference on Knowledge, Information and Creativity Support Systems (KICSS2015)*, 2015. (to appear)
- [16] A. Pritchard, "Statistical Bibliography or Bibliometrics?" in Journal of Documentation, 1969, Vol. 25, No. 4, pp.348–349.
- [17] A. I. Pudovkin, and E. Garfield, "Rank-Normalized Impact Factor: A Way to Compare Journal Performance Across Subject Categories," Proc. of *the ASIST Annual Meeting*, 2004, Vol. 41, No. 1, pp.507–515.
- [18] A. Saka, M. Igami, "Mapping modern science using co-citation analysis," in *Information Visualization. IV'07, 11th International Conference*, 2007, pp.453–458.
- [19] G. Salton, "Automatic Text Processing: The Transformation Analysis and Retrieval of Information by Computer," in Addison-Wesley series in computer science. 1988.
- [20] R. Yan, C. Huang, J. Tang, Y. Zhang, X. Li, "To better stand on the shoulder of giants," in Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, 2012, Vol.51, No. 60.

	TABLE I	
TOP 10 ARTICLES	WITH RESPECT TO	CITATION COUNTS

rank	ACC/CC	CC	ACC	short name	title
1	6.48	399	2584	Seglen1997	Why the impact factor of journals should not be used for evaluating research
2	4.27	260	1110	Garfield2006	The history and meaning of the journal impact factor
3	7.53	241	1815	Borgman2002	Scholarly communication and bibliometrics
4	1.22	202	246	Daim2006	Forecasting emerging technologies: Use of bibliometrics and patent analysis
5	7.09	195	1382	White1998	Visualizing a discipline: An author co-citation analysis of information science, 1972-1995
6	8.83	158	1395	Garfield1996	How can impact factors be improved?
7	7.35	137	1007	Garfield1999	Journal impact factor: A brief review
9	5.63	134	755	Meho2007	Impact of data sources on citation counts and rankings of LIS faculty:
					Web of science versus scopus and google scholar
8	11.72	134	1571	Almind1997	Informetric analyses on the world wide web: Methodological approaches to 'webometrics'
11	10.75	130	1397	Cronin2001	Bibliometrics and beyond: some thoughts on web-based citation analysis

 TABLE II

 TOP 10 ARTICLES WITH RESPECT TO ACCUMULATED CITATION COUNTS

rank	ACC/CC	CC	ACC	short name	title
1	6.48	399	2584	Seglen1997	Why the impact factor of journals should not be used for evaluating research
3	7.53	241	1815	Borgman2002	Scholarly communication and bibliometrics
23	18.15	91	1652	Rousseau1997	Situations: An exploratory study
8	11.72	134	1571	Almind1997	Informetric analyses on the world wide web: Methodological approaches to 'webometrics'
11	10.75	130	1397	Cronin2001	Bibliometrics and beyond: some thoughts on web-based citation analysis
6	8.83	158	1395	Garfield1996	How can impact factors be improved?
5	7.09	195	1382	White1998	Visualizing a discipline: An author co-citation analysis of information science, 1972-1995
53	22.94	54	1239	Moed1996	Impact factors can mislead.
20	12.23	92	1125	Moed1995	New bibliometric tools for the assessment of national research performance:
					Database description, overview of indicators and first applications
2	4.27	260	1110	Garfield2006	The history and meaning of the journal impact factor

TABLE III TOP 10 ARTICLES WITH RESPECT TO ACC/CC

rank	ACC/CC	CC	ACC	short name	title
326	38.83	18	699	Cronin2000	Semiotics and evaluative bibliometrics
477	34.86	14	488	Snyder1995	What's the use of citation? Citation analysis as a literature topic
					in selected disciplines of the social sciences
480	31.64	14	443	Yitzhaki1998	The 'language preference' in sociology: Measures of 'language self-citation',
					'relative own-language preference indicator', and 'mutual use of languages'
305	30.16	19	573	Ding1999	Mapping the intellectual structure of information retrieval studies:
					An author co-citation analysis 1987-1997
191	28.48	27	769	Taubes1993	Measure for measure in science
120	26.92	37	996	Snyder1999	Can search engines be used as tools for web-link analysis? A critical view
249	26.09	23	600	White1997	A qualitative study of citing behavior: Contributions, criteria,
					and metalevel documentation concer