Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Application Of Scientometric Laws On Deep Learning Research Output Using R Programming

M.Thamaraiselvi [1] * , Dr. S. Lakshmi[2]

[1*] Part-Time Research Scholar (LIS), Srimad Andavan Arts and Science College (A), Affiliated to Bharathidasan University, Thiruvanaikoil, Tiruchirappalli – 620005, Tamil Nadu, India.

[2] Librarian & Head, DLIS, Srimad Andavan Arts and Science College (A), Affiliated to Bharathidasan University, Thiruvanaikoil, Tiruchirappalli- 620005, Tamil Nadu, India.

Abstract

The present study investigates the validity of Lotka's law, Bradford's law and Zipf's law on deep learning research publications using R programming. Data was collected from SCOPUS database for the period of four years from 2017 to 2020. A total of 28536 records were retrieved from the database during the study period. The findings of the study shows that single author dominated in the field of deep learning research with 77.40% of total publications. Lotka's law of scientific productivity of authors reveals that 14.32% of publications was contributed by single author which is less than 60% of author productivity, Bradford's law of journals are scattered as 76:766:5019 in three zones and Zipf's law indicates that deep learning is the word secured top rank in frequency of occurrence. The result also concluded that the application of Lotka's law and Zipf's law are did not fit to the dataset whereas Bradford's law is fit to the deep learning research output.

Keywords: Deep learning, Lotka's law, Bradford's law, Zipf's law, R programming, Scientometrics, SCOPUS.

INTRODUCTION

The publication productivity of the topics in various subjects is measured by applying the scientometric laws. It examines the research performance of the subject specialist in various levels of organizations. It helps to the research specialist, librarian, and information science researchers to apply the appropriate tools and technique to evaluate the detailed information about the topics to the faculty and researchers. It enables the strength, weakness and trends of a topic which takes place in the author's productivity, scattering of journals and the occurrence of keywords etc. Scientometrics are applied to evaluate the performance of the science related subject research output with the parameters and laws.

Deep learning is comes under the category of computer science applications and programming language. It is a subset of machine learning and artificial intelligence. Deep learning helps human to gain specific types of knowledge. It acts as a important element for data science, statistics and predictive modeling fields. It is very useful to the data scientists for collecting, analyzing and interpreting huge amount of data in faster and easier than other technology.

Deep learning is an easy way to analyze the predictive future in the society. In tradition, machine learning algorithms are mostly used to process data while deep learning algorithms are in the form of hierarchy, short and increase the complex and abstraction of the data. It recognizes and processes the images and speech quicker and reply immediately than machine learning algorithms. Deep learning is more popular among data scientists for the accuracy of result which analyzed from the huge amount of data set.

In this study, R programming is applied to verify the application of scientometric laws on deep learning research output from 2017 to 2020.

REVIEW OF LITERATURE

Chaturbhuj and Batcha (2020) verified Lotka's law of author productivity in the field of thermodynamics dataset from 2015 to 2019. Web of Science database is used to collect data for this study. They concluded that Lotk'as law does not fit to the inverse square root method of the present dataset.

Sudhier (2020) studied the scholarly distribution of research publications scattered on doctoral theses citations at Indian Institute of Science from 2004 to 2008. He examined the scattering of citations through Bradford's law. He found that 79 doctoral theses contain 690 periodicals with 11319 references. Physics Review – B is the most cited journal with 9.53% citations and the citation distribution pattern of journals does not fit to the Bradford's law of scattering during the study period.

Thamaraiselvi et al. (2020) applied scientometrics laws in scientometric journal publications. Data was collected from Web of Science database from 2010 to 2019. They found that there is an increasing trend in the publications of scientometric journal for the next 10 years and the applicability of scientometrics laws such as Lotka's law, Price Square Root law and Pareto's Principle shows that there are not fit to the dataset during the study period.

Thirumagal et al. (2020) analyses the application of Lotka's law, Price's square root law, Pareto principle and collaborative measures on the research productivity of Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu. Web of Science is the database which is used to download the data for the study (1992-2020). The findings of the study shows that the scientometric laws do not fit to the dataset of Manonmaniam Sundaranar University research publications.

Lalrempuii, Ngurtinkhuma and Mishra (2019) investigated the validity of bibliometric laws, especially Lotka's law and Zipf's law in their study. They applied bibliometric laws in doctoral theses of library and information science who have completed in the University of North and East India. They concluded that a total of 12707 citations were cited in the doctoral theses for the period from 2006 to 2015. They also revealed that the validity of Lotka's law and Zipf's law is does not fit to the dataset.

OBJECTIVES OF THE STUDY

The aims of the present study are denoted as follows

1. To find out the authorship pattern and examine the validity of Lotka's law of scientific productivity on deep learning research publications.
2. To find out the core journals contributed in deep learning research productivity and verify the Bradford's law.
3. To known the frequency of word occurrences and test the data using Zipf's law of word occurrences.

HYPOTHESES

1. Multi authors dominated in deep learning research publications.
2. Lotka's law fit to the observed frequency of author productivity which distributed on deep learning research publications.
3. The scattering of deep learning research publications in journals confirms the implication of Bradford's law.
4. The frequency of word occurrences in the deep learning research publications obeys the Zipf's law.

SCOPE OF THE STUDY

The scope of the study is confined to the application of scientometric laws on deep learning over a period of 4 years (2017-2020). The data on deep learning is collected from SCOPUS database.

METHODOLOGY

The present study contains a total of 32718 records which was available on the database, by using the search string as deep learning for the time span of 2017 to 2020. Data was collected from SCOPUS databases. The collected data was analyzed, tabulated and interpreted using R programming software.

DATA ANALYSIS AND FINDINGS

AUTHORSHIP PATTERN ANALYSIS

The study of authorship pattern is one of the important aspects of scientometric study. It helps to know the contribution and collaboration of authors on deep learning research publications.

Table 1: Authorship Pattern of Deep Learning Research Publications

Authorship Pattern	No. of Authors	%	Cumulative No. of Authors	Cumulative %
Single	22087	77.40	22087	77.40
Two	1485	5.20	23572	82.60

Three	1637	5.74	25209	88.34
Four	907	3.18	26116	91.52
Five	559	1.96	26675	93.48
Above Five	1861	6.52	**28536**	**100.00**
Total	**28536**	**100.00**		

From the above table, it clearly indicates the contribution of authors on deep learning research publications. A total of 28536 authors contributed in the deep learning research publications during the study period. Out of which single author contributed 22087 (77.40%) publications followed by 1485 (5.20%) of two authors publications, 1637 (5.74%) of three authors, 907 (3.18%) of four authors, 559 (6.52%) of five authors publications and 1861 (6.52%) more than five authors contributed on deep learning research. Hence it concluded that single authors dominated in the field of deep learning research with 77.40% of publications.

Table 1 shows that the single author contributed (77.40%) of publications whereas multi authors contributed (22.6%) publications from total contribution. Hence, it stated that **hypothesis 1 “Multi authors dominated in deep learning research publications” has not statistically proved.**

APPLICATION OF LOTKA’S LAW OF SCIENTIFIC PRODUCTIVITY

The productivity of author can be measured using Lotka’s law. It states that the major contribution should contributed by small number of authors. Alfred Lotka’s stated that the scientific productivity is based on the frequency of author’s contribution in a given field or dataset. It define that the number of authors making n contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that make a single contribution is about 60 percentage. The application of Lotka’s law for the present study has been verified from the below table.

Table 2: Lotka’s Law on Deep Learning Research Publications

Authorship Pattern	No. of Contributions	Total No. of Contributions	%	Cumulative No. of Total Contributions	Cumulative %
1	**22087**	**22087**	**14.32**	**22087**	**14.32**
2	1485	2970	1.93	25057	16.25
3	1637	4911	3.18	29968	19.43
4	907	3628	2.35	33596	21.79
5	559	2795	1.81	36391	23.60
6	609	3654	2.37	40045	25.97
7	90	630	0.41	40675	26.38
8	96	768	0.50	41443	26.87
9	146	1314	0.85	42757	27.73
10	47	470	0.30	43227	28.03
11	77	847	0.55	44074	28.58
12	77	924	0.60	44998	29.18

Authorship Pattern	No. of Contributions	Total No. of Contributions	%	Cumulative No. of Total Contributions	Cumulative %
13	69	897	0.58	45895	29.76
15	50	750	0.49	46645	30.25
16	31	496	0.32	47141	30.57
17	32	544	0.35	47685	30.92
18	72	1296	0.84	48981	31.76
21	55	1155	0.75	50136	32.51
23	30	690	0.45	50826	32.96
25	24	600	0.39	51426	33.35
26	26	676	0.44	52102	33.79
27	23	621	0.40	52723	34.19
28	29	812	0.53	53535	34.72
33	25	825	0.53	54360	35.25
39	22	858	0.56	55218	35.81
41	20	820	0.53	56038	36.34
42	21	882	0.57	56920	36.91
43	19	817	0.53	57737	37.44
52	18	936	0.61	58673	38.05
57	15	855	0.55	59528	38.60
62	33	2046	1.33	61574	39.93
78	14	1092	0.71	62666	40.64
80	13	1040	0.67	63706	41.31
122	12	1464	0.95	65170	42.26
152	11	1672	1.08	66842	43.35
192	10	1920	1.25	68762	44.59
197	9	1773	1.15	70535	45.74
266	8	2128	1.38	72663	47.12
343	7	2401	1.56	75064	48.68
578	6	3468	2.25	78532	50.93
958	5	4790	3.11	83322	54.03
1822	4	7288	4.73	90610	58.76
3413	3	10239	6.64	100849	65.40
9879	2	19758	12.81	120607	78.21
33421	1	33421	21.67	**154028**	**100.00**
Total	**28536**	**154028**	**100.00**		

The highlighted value in the table 2 indicates that single authorship pattern contributed i.e. 22087 (14.32%) articles followed by the two author's contribution i.e. 1485 (1.93%), three authors who have contributed 1637 (3.18%), four authors who have contributed 907 (2.35%), five authors contributed 559 (1.81%) publications from the total publications and so on.

Table 2 also represents the author productivity data for Lotka's law. Of the 28536 total contributions 22087 (14.32%) publications was contributed by single authors. Based on the Lotka's law 60% of the contribution must contributed by single authors. So the dataset did not match the generalized form of Lotka's law. Hence, it concluded that **hypothesis 2 "Lotka's law fit to the observed frequency of author productivity which distributed on deep learning research publications" has not statistically proved.**

IMPLICATION OF BRADFORD'S LAW OF SCATTERING

The aim of Bradford's law is to define that a group of journals act as a core journal for the dataset. It is arranged in an order of decreasing productivity to identify which journals yield most productive articles. According to this law, journals are grouped in to three zones. First zone considered as nucleus zone with a set of productive articles and other two zones must have the similar number of articles. However, the number of journals in each zone will be in the form of increasing mode. The relationships between the zones are calculated as 1: n: n^2. The total number of articles 20884 was scattered in 5861 journals.

Table 3: Bradford's Law of Scattering on Deep Learning Research Publications

Rank	No. of Journals	Cumulative No. of Journals	Frequency	Cumulative Frequency	Zone
1	1	1	1393	1393	Zone 1
2	1	2	959	2352	Zone 1
3	1	3	540	2892	Zone 1
4	1	4	409	3301	Zone 1
5	1	5	361	3662	Zone 1
6	1	6	309	3971	Zone 1
7	1	7	283	4254	Zone 1
8	1	8	238	4492	Zone 1
9	1	9	208	4700	Zone 1
10	1	10	203	4903	Zone 1
11	1	11	197	5100	Zone 1
12	1	12	192	5292	Zone 1
13	1	13	184	5476	Zone 1
14	1	14	182	5658	Zone 1
15	1	15	162	5820	Zone 1
16	2	17	155	5975	Zone 1
17	1	18	146	6121	Zone 1
18	1	19	143	6264	Zone 1
19	1	20	132	6396	Zone 1
20	1	21	130	6526	Zone 1
21	1	22	120	6646	Zone 1
22	2	24	118	6764	Zone 1
23	1	25	115	6879	Zone 1

Rank	No. of Journals	Cumulative No. of Journals	Frequency	Cumulative Frequency	Zone
24	1	26	114	6993	Zone 1
25	1	27	110	7103	Zone 1
26	2	29	107	7210	Zone 1
27	1	30	101	7311	Zone 1
28	1	31	100	7411	Zone 1
29	1	32	96	7507	Zone 1
30	1	33	94	7601	Zone 1
31	1	34	92	7693	Zone 1
32	1	35	91	7784	Zone 1
33	1	36	88	7872	Zone 1
34	2	38	87	7959	Zone 1
35	1	39	86	8045	Zone 1
36	1	40	82	8127	Zone 1
37	1	41	79	8206	Zone 1
38	1	42	77	8283	Zone 1
39	1	43	76	8359	Zone 1
40	1	44	75	8434	Zone 1
41	2	46	74	8508	Zone 1
42	2	48	71	8579	Zone 1
43	1	49	70	8649	Zone 1
44	3	51	69	8718	Zone 1
45	1	52	68	8786	Zone 1
46	1	53	66	8852	Zone 1
47	2	55	64	8916	Zone 1
48	1	56	62	8978	Zone 1
49	2	58	61	9039	Zone 1
50	3	61	57	9096	Zone 1
51	4	65	56	9152	Zone 1
52	1	66	55	9207	Zone 1
53	3	69	54	9261	Zone 1
54	2	71	52	9313	Zone 1
55	3	74	51	9364	Zone 1
56	**1**	**75**	**50**	**9414**	**Zone 1**
57	2	77	49	9463	Zone 2
58	1	78	48	9511	Zone 2
59	4	82	47	9558	Zone 2
60	5	87	46	9604	Zone 2
61	3	90	45	9649	Zone 2
62	3	93	44	9693	Zone 2
63	1	94	43	9736	Zone 2

Rank	No. of Journals	Cumulative No. of Journals	Frequency	Cumulative Frequency	Zone
64	2	96	42	9778	Zone 2
65	3	99	41	9819	Zone 2
66	2	101	40	9859	Zone 2
67	3	103	39	9898	Zone 2
68	2	105	38	9936	Zone 2
69	2	107	37	9973	Zone 2
70	7	114	36	10009	Zone 2
71	3	117	35	10044	Zone 2
72	8	125	34	10078	Zone 2
73	5	130	33	10111	Zone 2
74	7	137	32	10143	Zone 2
75	7	144	31	10174	Zone 2
76	3	147	30	10204	Zone 2
77	8	155	29	10233	Zone 2
78	15	170	28	10261	Zone 2
79	8	178	27	10288	Zone 2
80	10	188	26	10314	Zone 2
81	10	198	25	10339	Zone 2
82	8	206	24	10363	Zone 2
83	9	215	23	10386	Zone 2
84	16	231	22	10408	Zone 2
85	12	243	21	10429	Zone 2
86	15	258	20	10449	Zone 2
87	14	272	19	10468	Zone 2
88	20	292	18	10486	Zone 2
89	15	307	17	10503	Zone 2
90	19	326	16	10519	Zone 2
91	28	354	15	10534	Zone 2
92	25	379	14	10548	Zone 2
93	29	408	13	10561	Zone 2
94	51	459	12	10573	Zone 2
95	46	505	11	10584	Zone 2
96	52	557	10	10594	Zone 2
97	65	612	9	10603	Zone 2
98	91	703	8	10611	Zone 2
99	**127**	**830**	**7**	**10618**	**Zone 2**
100	177	1007	6	10624	Zone 3
101	227	1234	5	10629	Zone 3
102	420	1654	4	10633	Zone 3
103	565	2219	3	10636	Zone 3

Rank	No. of Journals	Cumulative No. of Journals	Frequency	Cumulative Frequency	Zone
104	1064	3283	2	10638	Zone 3
105	**2566**	**5849**	**1**	**10639**	**Zone 3**

2566 different journals were published only one publication which is related to deep learning, 1064 journals were produced each two publications, 565 journals were published each three publications, 420 journals were produced each four publications, 227 journals were published each five publications and so on. The bold letters mentioned the values of three zones of Bradford's law of scattering on deep learning dataset.

Table 3.1: Bradford's Distribution of Journals on Deep Learning Research Publications

Zones	No. of Journals	No. of Publications	Bradford Multiplier
z1	76	9414	
z2	766	10172	10.08
z3	5019	10266	6.55
Total	**5861**	**29852**	**8.31**

The above table 3.1 shows the observation of group journals produced research publications on deep learning. The researcher has grouped them as three zones. They are z1 – zone1, z2 – zone2 and z3 – zone3. The first zone indicated as nuclear zone or core zone which has small groups of journals. From the analysis, first zone representing 76 journals which produced 9414 research publications. So the 76 journals were identified as core journals on the deep learning research publications. The second zone has 766 journals with 10172 publications and the third zone of 5019 journals yields the remaining 10266 publications. The Bradford multiplier between zone 1 and zone 2 is 10.08 while it is 6.55 between zone 2 and zone 3. The average value of multiplier is 8.31.

According to the Bradford's law, the scattering of journals are in the form of 76:766:5019. Here k is 8.31. Therefore the value is 76:76x8.31:76x8.312 is 76:631.56:5248.264.

Percentage of error = (5955.824-5861)/5861*100 = 1.62

on the basis of above calculation, it is found that the percentage of error (1.62%) is very low, and the data of deep learning research output fit to the Bradford expression. It is found from the above analysis that the **hypothesis -3 "The scattering of deep learning research publications in journals confirms the implication of Bradford's law" is accepted.**

APPLICATION OF ZIPF'S LAW OF WORD OCCURRENCE

The law define that the list of words occurring in the dataset is arranged in decreasing frequency. The multiplication of rank and frequency of word must be constant. The equation of the Zipf's law is rxf=k, where k is constant, r is the rank of the word, and f is the frequency of word occurred. While taking in to consideration, top 25 most occurred words are listed in the table.

Table 4: Zipf's Law of Word Occurrence

S. No.	Words	No. of Occurrences	Rank	rxf = k (constant)
1	deep learning	28876	1	28876
2	neural networks	6292	2	12584
3	deep neural networks	6180	3	18540
4	learning systems	5882	4	23528
5	convolutional neural network	4977	5	24885
6	machine learning	3992	6	23952
7	human	3695	7	25865
8	learning algorithms	3558	8	28464
9	convolution	3394	9	30546
10	reinforcement learning	3335	10	33350
11	article	3227	11	35497
12	humans	2612	12	31344
13	forecasting	2402	13	31226
14	image processing	2379	14	33306
15	classification (of information)	2370	15	35550
16	female	2334	16	37344
17	convolutional neural networks	2227	17	37859
18	male	2164	18	38952
19	image segmentation	1958	19	37202
20	artificial intelligence	1761	20	35220
21	long short-term memory	1751	21	36771
22	procedures	1748	22	38456
23	algorithm	1658	23	38134
24	network architecture	1601	24	38424
25	adult	1552	25	38800

Table 4 shows the words occurrence of Zipf's law. The word deep learning secured as 1st rank with 28876 frequency of occurrence followed by neutral networks (6292) as second rank, deep neural networks (6180) as third, leaning systems (5882) as fourth and convolutional neural network (4977) as fifth rank and so on. The relationships of the word occurrences in the dataset are not always related and multiplication of them is not equal to the constant. Therefore, Zipf's law did not match with the dataset. From the above table it was found that the **hypothesis 4 – "The frequency of word occurrences in the deep learning research publications obeys the Zipf's law" has not statistically proved.**

CONCLUSION

This study was widely used to evaluate the research output of deep learning by applying scientometric laws using R programming. Data was downloaded from SCOPUS database for the study period of 2017 to 2020. The collected data was analysed, tabulated and interpreted using R programming language and MS Excel softwares. It concluded that single author dominated in this research with 77.40% of publications whereas multi authorship contribution is 22.6%. According to the application of scientometric laws, Lotka's law of author productivity of single author contribution is less than 60%, Bradford's law, the scattering of journals are in the form of 76:766:5019 distribution and Deep learning is the word secured as 1st rank with 28876 frequency of occurrence followed by neutral networks (6292) as second rank, deep neural networks (6180) as third. It also tested the hypotheses that "Multi authors dominated in deep learning research publications" has not statistically proved. "Lotka's law fit to the observed frequency of author productivity which distributed on deep learning research publications" has not statistically proved. "The scattering of deep learning research publications in journals confirms the implication of Bradford's law" is accepted and "The frequency of word occurrences in the deep learning research publications obeys the Zipf's law" has not statistically proved.

REFERENCES

1. Chaturbhuj, S. B., & Batcha, M. S. (2020). Application of Lotka's Law to the research productivity in the field of Thermodynamics during 2015-2019. Library Philosophy and Practice (e-journal) .4523, 1-12.

2. Lalrempuii, S., Ngurtinkhuma, R. K., & Mishra, R. N. (2019). Application of bibliometric laws in doctoral theses of library and information science submitted to the Universities of NorthEast India. JOURNAL OF INDIAN LIBRARY ASSOCIATION , 55 (4), 47-58.

3. Sudhier, K. G. (2020). Application of Bradford's law of Scattering to the Physics literature: A study of Doctoral theses citations at the Indian Institute of Science. DESIDOC Journal of Library & Information Technology , 30 (2), 3-14.

4. Thamaraiselvi, M., Lakshmi, S., Manthiramoorthi, M., & Thirumagal, A. (2020). Application of bibliometrics laws in scientometric journal publications during 2010 - 2019: a scientometric analysis. Library Philosophy and Practice (e-journal). 4802, 1-14.

5. Thirumagal, A., Manthiramoorthi, M., Thamaraiselvi, M., & Mani, M. (2020). Application of Lotka's law Price's square root and Pareto principle on research publications of

Manonmaniam Sundaranar University - a scientometric analysis. Library Philosophy and Practice (e-journal). 4771 , 1-15.

6. Lotka, A. J. (1926). The frequency distribution of scientific productivity. Journal of Washington Academy of Sciences , 16 (12), 317-323.
7. Bradford, S. C. (1934). Sources of information on specific subjects. Engineering , 137, 85-86.
8. https://en.wikipedia.org/wiki/Zipf%27s_law. (n.d.). Retrieved 10 19, 2021, from https://en.wikipedia.org.
9. definition/deep-learning-deep-neural-network. (n.d.). Retrieved 10 19, 2021, from Searchenterpriseai.techtarget.com: https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network