

Analysis of Temporal Logarithmic Perspective Phenomenon Based on Changing Density of Information

Yonghe Lu

School of Information Management
Sun Yat-sen University
Guangzhou, China
luyonghe@mail.sysu.edu.cn

Weiting Zhang

School of Information Management
Sun Yat-sen University
Guangzhou, China
zhangwt7@mail2.sysu.edu.cn

Abstract—Considering the influence of exponential growth of network information and literature on the density of information as time goes by, this paper proposes logarithmic transformation expression based on changing density of information. To evaluate the applicability of the expression, time distributed data of references from SCI literatures is extracted to conduct curve fitting using the original and improved expression respectively. Coefficient of determination (R square) is used as the evaluation indicator to verify the validity of the improved expression through the comparison with the original expression. The results show that the improved expression has good ability to explain the rule of information utilization as time goes by. At the same time, the improved expression is better than original expression, which has a fixed density of information, in explanation ability for a long time.

Keywords—*logarithmic perspective; information density; curve fitting; literatures growth;*

I. INTRODUCTION

Principle of logarithmic perspective was proposed by Brookes in 1980. He gained his enlightenment from Weber-Fechner law and inferred that if our sensory mechanisms behave according to a log law, it seems possible that all our neural mechanisms, including those of our mental neural system[1]. According to the principle, we can consider our sensory and neural mechanisms as a logarithmic scale. Every entity in the physical space is reflected in our sensory systems in terms of its logarithmic scale[2].

Principle of logarithmic perspective is of great value for research as a quantitative method of information science. First, it can (1)reveal the differences in quantities and characteristics between information in physical space and knowledge in understanding space as well as between information carriers and contents[3]; second, it (2)well explains the transforming rules of information and knowledge as time, space and discipline varies[4]; finally, the existing logarithmic perspective limits a lot on application conditions, which implies (3)the principle has new development space especially in 21st Century, when Internet greatly developed and network information grows explosively.

Finding new expressions of logarithmic perspective is helpful to discover quantitative rule of information acquisition in special situations such as network environment.

Some fundamental assumptions are provided by the principle:(1) The density of information space is uniform from the macro perspective. Meanwhile the quantity of information and knowledge of each paper is the same from the micro perspective;(2)The receptivity of different people who receive the information and knowledge is the same;(3)No auxiliary means and technologies are used to support the acquisition of information and knowledge;(4)Knowledge can be inherited well[5].To some extent, these assumptions limit the application of the logarithmic perspective theory. For instance, the theory requires the distribution of information and knowledge is uniform on the time line of the physical information space when exploring the transforming rules of information and knowledge as time goes by. However, This is impossible as it is known that the most outstanding achievements and progress are “law of exponential growth” and “law of logical growth” in the research of increasing law of knowledge and literature of science[6], which implies the increasing of literature follows the exponential law or even the more complicated logical law as time goes by. Furthermore, Statistics of network information[7,8] also illustrated the exponential growth of network information. Thus the existing application of logarithmic perspective is inadequate since it ignored the influence of information growth. The two studies[7,8] analyzing principle of logarithmic perspective in the network environment both demonstrated the time distribution of information and knowledge is uneven. But neither of them noticed that the existing expression of logarithmic perspective theory is unfit for the condition of information growth.

In this paper, the influence of information growth as time goes by is considered to calculate the time distribution of information. After that, expression of temporal logarithmic perspective based on changing density of information is proposed. Furthermore, to evaluate the applicability of the improved expression, the literature citation dataset from SCI, on which the curve fitting method is performed, is extracted to simulate information acquisition in network environment. Finally, the comparison of coefficient of determination(R square) between the original

expression and the improved one demonstrates the validity of improvement.

II. RELATED WORKS

The existing research of logarithmic perspective theory can be divided mainly into three categories: (1) Statement, analysis and commentary of logarithmic perspective theory; (2)evaluating the transforming rules of information and knowledge as time, space and discipline varies; (3)application of logarithmic perspective theory to explore the pattern of information and knowledge growth in special situations such as network environment.

A. Statement, analysis and commentary of logarithmic perspective theory

In addition to Brookes' paper in 1980[1], which first propose the theory, and its Chinese version which was translated by information scientists in China[9], Ma[2,3] analyzed and evaluated the generation and application of logarithmic perspective theory as well as introducing the basic principle and method of information science. Belkin [10] expounded the cognitive view of information science which includes academic system of information science established by Brooks. He considered that one of the most interesting aspects of Brookes' work is the explicit relationship that Brookes draws between individual cognition, especially perceptual and judgemental abilities, and the statistics necessary for considering the interactions of groups of people with objective knowledge.

B. Evaluating the transforming rules of information and knowledge

The presentation of logarithmic perspective theory will have a significant impact on the quantification of information science. But the applicability of logarithmic perspective theory in measuring the quantity of cognitive information in the information space when receiving message still needs further evaluations through experiments[5]. Jiang et al. evaluated the logarithmic rule of quantity of references in data mining area from the perspectives of time and discipline and further discussed the utilization rate of literatures according to the citation of literatures in data mining area[4].Although the paper performed curve fitting method towards the relationship between time and quantity of references, it ignored the problem that the density of information is uneven. Xiao et al. discussed the logarithmic perspective of knowledge growth in network environment, knowledge acquisition with time and discipline as well as retrieval efficiency. The paper implied the logarithmic perspective theory should develop new forms of expression and evolutionary trend in network environment though without detailed discussion and analysis[7]. Deng analyzed the link data of websites of Chinese commercial banks and demonstrated the applicability of logarithmic perspective theory in website links[11].

C. Applications on the pattern of information and knowledge growth in network environment

Zhang et al. used logarithmic perspective theory to explore the quantity of knowledge according to the quantity of information in network environment. The growth patterns of network information and knowledge are analyzed respectively[8]. Yuan et al. discussed the perspective transformation in the communications of tacit knowledge, which can be further divided into four aspects including spatial, temporal, expectation and knowledge correlation. Effective strategies, characteristics and effects of each aspect of the perspective transformation were analyzed[12].

This paper is an empirical research on the basis of paper[4,5]. First of all, an improvement of expression of logarithmic perspective based on changing density of information is proposed. Then the improved expression is evaluated using references data from retrieval platform of literatures. Finally, the limitations of the improved expression are presented .

III. RESEARCH METHODS

A. Expression of Logarithmic Perspective Based on Changing Density of Information

Brookes' expression of logarithmic perspective[1,9] which reflects the quantity of perceived information is formulated as (1).

$$aPI = \int_a^{a+n} \rho \times \frac{1}{x} dx = \rho[\ln(a+n) - \ln a] \quad (1)$$

Where the ρ is the density of potential information, $a+n$ and a are scale values of linear scale in the physical space of information as well as the instances of x .

However, the distribution of potential information in the physical space is uneven, which means the density of potential information in the linear scale of physical space is also diverse.

Taking time as an example to explore the changing regulation of density of potential information. The density of potential information is proportional to the quantity of potential information as time goes by. According to Price's law of literature growth, the quantity of scientific literatures increases exponentially as time goes by[6] as shown in (2).

$$F(t) = ae^{bt} \quad (a > 0, b > 0) \quad (2)$$

Where the $F(t)$ is the accumulation quantity of literatures; t is time calculated in years; a is a conditional constant which means the quantity of literatures at the initial moment; b is a time constant which represents the constant growth rate of literatures. According to (2) the increment of literatures as time goes by can be derived as (3).

$$(F(t))' = abe^{bt} \quad (a > 0, b > 0) \quad (3)$$

As shown in (3), the increment of literatures with time also accords with exponential growth. According to the principle of logarithmic perspective, the increment of information that can be perceived is the logarithm of the increment of literatures as shown in (4). Information that can be perceived doesn't refer to the information that has been perceived by individual, but refer to the actual

quantity of information. Considering that the information received includes a lot of repetition, the actual quantity of information should be less than the quantity of information received.

$$PI = K \ln(F(t))' = K \ln(ae^{bt}) = Kbt + K \ln a \quad (a > 0, b > 0) \quad (4)$$

As shown in (4), the increment of information that can be perceived is a linear growth with time. So this raises a question: the application of logarithmic perspective theory in paper[7] showed the quantity of network knowledge increases linearly as time goes by. In other words, the increment of knowledge with time is a constant. According to (2), the accumulation quantity of information that can be perceived can be derived as (5).

$$aPI = K \ln(F(t)) = K \ln(ae^{bt}) = Kbt + K \ln a \quad (a > 0, b > 0) \quad (5)$$

As shown in (5), the increment of information that can be perceived as time goes by can be derived as (6).

$$(aPI)' = Kb \quad (a > 0, b > 0) \quad (6)$$

It seems that (4) and (6) are contradictory. But it is essential to clarify that the meanings of the increment of information that can be perceived in two equations are different. In (6), the increment of information represents the increment of new information. The equation implies the increment of new information that can be perceived with time is a constant. However, the increment of information includes not only new information, but also the information that has been produced in the past. So (2) represents the total increment of new information and old information that can be perceived with time.

When scholars retrieve literatures from retrieval platforms, they don't care whether the information acquired is new or existed. So the increment of information that can be perceived with time is in accordance with (4). Furthermore, the density of information that can be perceived is proportional to the increment of information that can be perceived with time. According to (4), the density of information that can be perceived with time can be calculated as (7).

$$\rho = Kt \quad (K > 0) \quad (7)$$

Where t represents time calculated in years. Replacing the ρ of (1) with (7). One important thing to note is the t in (7) is calculated from the earliest year, which is contrary to the meaning of x in (1). Thus the accumulation quantity of perceived information is derived as (8).

$$aPI = \int_a^{a+n} \rho \times \frac{1}{x} dx = \int_a^{a+n} K(N-x) \times \frac{1}{x} dx = (KN \ln x - Kx) \Big|_a^{a+n} = KN \ln \frac{a+n}{a} - Kn \quad (K > 0) \quad (8)$$

Where the N is the maximum of x. That is to say, $1 \leq x \leq N$.

B. Evaluation of the Improved Expression of Logarithmic Perspective

1) Source of data:

Searching all the related literatures from Web of Science with user query, of which the topic is "machine

learning" and the time interval is "1999-2016". The total number of retrieved literatures is 32592. As shown in Fig.1, the growth of literatures as time goes by conforms to the law of exponential growth, which illustrates that the basis(as shown in (2)) of improved expression (as shown in (7)) is tenable.

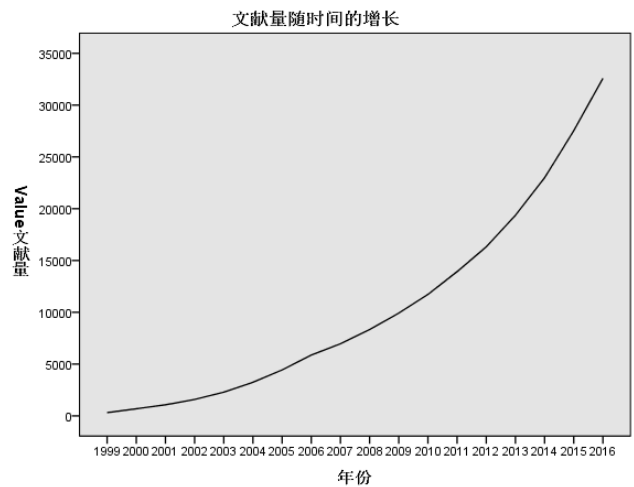


Fig. 1 Growth of Literatures as Time Goes by

To research the temporal logarithmic perspective phenomenon based on the changing density of information, all references cited by the literatures published in 2016 are selected as experimental data. The initial number of references is 226921. After removing the duplicate and informal references, the number of remaining references is 165638. On the basis of logarithmic perspective theory, when obtaining information people prefer to select the information produced recently. If it's true, the same rule should be shown in statistic of references cited by scholars' papers, which means most references were published close to the citing year. Table I shows the time distribution of references cited by literatures published in 2016 with topic "machine learning". The cited references published in 2016 and 2017 are left out of consideration, since the published literatures are incomplete and scholars cannot cite the literatures that has not been published.

TABLE I (Part of)Time Distribution of References Cited by Literatures with Topic "Machine Learning"

Published time of references	Time interval between citing and cited	Number of references	Accumulative number of references
2015	1	13157	13157
2014	2	16248	29405
2013	3	15832	45237
2012	4	14304	59541
2011	5	12676	72217

2010	6	11102	83319
2009	7	9846	93165
2008	8	8589	101754
2007	9	7648	109402
2006	10	6821	116223
2005	11	6070	122293
2004	12	5297	127590
2003	13	4616	132206
2002	14	3996	136202
2001	15	3360	139562
2000	16	3078	142640
1999	17	2558	145198
1998	18	2286	147484
1997	19	1898	149382
1996	20	1789	151171
1995	21	1536	152707
1994	22	1295	154002
1993	23	1077	155079
1992	24	968	156047
1991	25	824	156871
1990	26	770	157641
1989	27	710	158351
1988	28	551	158902
1987	29	468	159370
1986	30	498	159868

From Table I it can be known that the time distribution of references is in accordance with the mainly idea of logarithmic perspective theory. That is to say, the smaller the time interval between citing and cited is, the bigger the quantity of references will be. But whether the distribution is in accordance with the expression of logarithmic perspective theory in statistics still needs further evaluation.

2) Curve fitting and comparative analysis of the expressions before and after improvement:

In this section, the curve fitting method is performed to evaluate the validity and explanatory ability of improved expression of logarithmic perspective. Expression before and after improvement are adopted respectively to fit the experimental data. The process of curve fitting is completed with the help of function `nlinfit()`, a nonlinear fitting function in MATLAB R2014b. When using `nlinfit()`, regarding the time interval as independent variable and the accumulative number of references as dependent variable.

The time interval between citing and cited is likely to reach a maximum of 116 years in experimental data. The references published earliest are very few, which is attributed not only to people's preference of acquiring information from literatures published recently, but also to the lack of topic related literatures in the past. Table II shows the R-Squares of expression before and after improvement with experimental data when the maximum of time interval varies. Maximum of time interval between citing and cited in Table II represents the value of N in (8).

TABLE II R-Squares of Expression Before and After Improvement with Experimental Data

Maximum of time interval between citing and cited	R-Square before improvement	R-Square after improvement
15	0.4847	0.5345
30	0.9828	0.9463
45	0.9497	0.9779
60	0.8926	0.9807
75	0.8214	0.9658
90	0.7418	0.9394
105	0.6576	0.9054
116	0.5940	0.8771

According to Table II, R-Squares of expression after improvement are always better than those before improvement except for when the maximum of time interval is 30 years. Meanwhile, the R-Square values of expressions before and after improvement are both increasing first and decreasing afterwards. The optimal curve fitting figures of expressions before and after improvement are drawn respectively with drawing function `plot()` in MATLAB. Fig.2 shows the optimal curve fitting figure of expression before improvement when the maximum of time interval is 30 years. Fig.3 shows the optimal curve fitting figure of expression after improvement when the maximum of time interval is 60 years. As shown in Fig.2 and Fig.3, the

development trend of fitted curve of expression after improvement is more accordance with the observations.

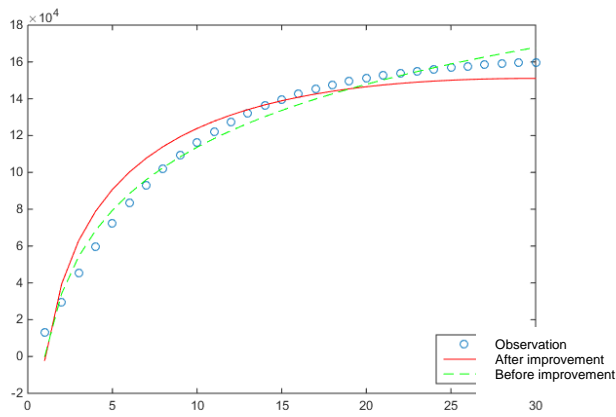


Fig. 2 Optimal Curve Fitting Figure of Expression Before Improvement

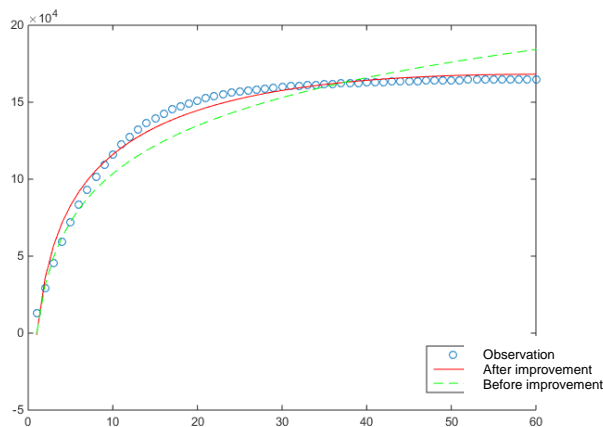


Fig. 3 Optimal Curve Fitting Figure of Expression After Improvement

3) Discussions:

As shown in Table II, the highest value of R-Squares of expression after improvement reaches 0.9807, which indicates the expression of logarithmic perspective based on changing density of information proposed in this paper can explain the perceived information with time to a certain degree.

In addition, it can be seen that the R-Squares of expressions before and after improvement reach their best values at different maximums of time interval respectively. The expression before improvement has better fitting effect when the maximum of time interval is smaller. In contrast, The expression after improvement has better fitting effect when the maximum of time interval is bigger, which indicates the improved expression has stronger ability in long-term explanation.

Although the expression before improvement has better fitting effect than the expression after improvement when the maximum of time interval is smaller (such as 30 years), we cannot ignore that the quantity of references that was published more than 30 years before the citing literatures is big, which implies the sample data is not comprehensive when the

maximum of time interval is small. So as shown in Fig.2, the fitted curve of expression before improvement is deviated from the development trend of observations of experimental data, which is not comprehensive enough to demonstrate the complete pattern of references quantity.

IV. CONCLUSIONS

In this paper the expression of logarithmic perspective is improved. Considering the exponential growth of network information and literatures with time, the expression of logarithmic perspective based on changing density of information is proposed. At the same time, using Web of Science to retrieve literatures related to “machine learning”, and extracting the references of these literatures as the experimental data to evaluate the applicability of improved expression through curve fitting method. The fitting results show the improved expression can explain the perceived information with time to a certain degree. Moreover, the improved expression has stronger ability in long-term explanation than the original expression, whose density of information is a constant.

In spite of the better fitting effect of improved expression, there are still limitations of the improvement in theory and application. In theory, the improved expression is on the basis of the assumption that information is increasing exponentially. The assumption may be applicable for network information, but limitation appears when it comes to literature growth. That is because the growth of literature is in stages, which means the quantity of literatures is more in accordance with logical growth than exponential growth. So, the improved expression proposed in this paper may not be suitable for all kinds of circumstances of information growth. In application, the improved expression is especially applicable to analyze the temporal logarithmic perspective phenomenon of perceived information. The applicability of the expression in other aspects such as discipline needs to be developed.

As one of the basic principles of information science, principle of logarithmic perspective is universal in the process of information receiving and utilizing. The improvement and application of logarithmic perspective theory will contribute to the discovery and prediction of user pattern of knowledge acquisition. In this paper, the function of density of information with time is introduced to improve the existing expression of logarithmic perspective. The improved expression makes the explanation and prediction of user pattern of knowledge acquisition more precious.

ACKNOWLEDGMENT

This research was supported by National Natural Science Foundation of China (Grant No. 71373291). This work also was supported by Science and Technology Planning Project of Guangdong Province, China (Grant No. 2016B030303003).

REFERENCES

[1] B. C. Brookes. The Foundations of Information Science. Part III: Quantitative Aspects: Objective Maps and Subjective Landscapes. Journal of Information Science, vol. 2, no. 6, pp.269-275, 1980.

[2] F. C. Ma. Quantitative Methods of Information Science Theory of Brooks. Information Science, vol. 4, no. 4, pp.1-9, 1983.

[3] F. C. Ma. Basic Principle and System Construction of Information Science. Journal of the China Society for Scientific and Technical Information, vol. 26, no. 1, pp.3-13, 2007.

[4] X. Q. Jiang, Z. P. Zhang, L. N. Li. Research on the Logarithmic Perspective of Knowledge Acquiring Based on a Case of Data Mining. Journal of Intelligence, vol. 33, no. 7, pp.156-160, 2014.

[5] J. P. Jing, F. C. Ma, X. X. Zhang. Theory of information science. Science Press, 2009.

[6] J. P. Qiu. Informetrics(II) Chapter II: Rule and Application of Literature Information Growth. Information Studies: Theory & Application, vol. 23, no. 2, pp.153-157, 2000.

[7] N. Xiao N, Q. E. Ren, F. Hu. A Study on the Principle of Logarithmic Perspective Under Network Environment. Document Information & Knowledge, vol. 23, no. 3, pp.60-64, 2007.

[8] J. Z. Zhang, G. X. Li. Research on the Logarithmic Perspective of Network Knowledge Growth—A Case of Competitive Intelligence. Library & Information Service, vol. 55, no. 2, pp.78-82, 2011.

[9] B. C. Brookes, C. D. Wang, Y. J. Deng, J. G. Liu. The Foundations of Information Science. Part III. Information Science, vol. 4, no. 6, pp. 84-91, 1983.

[10] N. J. Belkin. The Cognitive Viewpoint in Information Science. Journal of Information Science, vol. 16, no. 1, pp.11-15, 1990.

[11] H. M. Deng. A Study of the Principle of Logarithmic Perspective based on Links Analysis. Journal of the Graduates Sun Yat-sen University(Social Sciences), vol. 34, no. 3, pp.51-61, 2013.

[12] Q. Y. Yuan, G. Cheng. Perspective Transformation in the Communications of Tacit Knowledge. Journal of Library Science in China, vol. 33, no. 5, pp.95-98, 2007.