DSI DATA SCIENCE AND INFORMETRICS
VOL.1, NO.4, NOV. 2021

Application of Article–Level Metrics and comparison with Journal–Level Metrics in differentiated document recommendation: An empirical study in artificial intelligence field

Feifei Wang, Wanzhao Lu, Yusi Hou

College of Economics and Management, Beijing University of Technology, Beijing, China

ABSTRACT

The measurement indexes of literature include citation frequency, H index, etc. The evaluation of core journals mainly relies on the indexes such as impact factor or comprehensive evaluation method. With the in-depth development of research, these indicators are not comprehensive and accurate for literature and journals. Therefore, according to various literature needs for different researchers, "core literature" in this study was divided into three types: classical, popular and frontier; and measurement system of document value was constructed with comprehensive use of entropy weight method and principal component analysis from the perspective of Article-Level Metrics. In the case study of artificial intelligence (AI), three types of document sets were acquired with the threshold value of specific indicators, and then measured by a combination of multi-index, achieving identification and recommendation of core documents for different research needs. At the same time, this paper further calculates the total score of the journals according to the literature score, and finds that the journal distribution of different types of core literature is quite different. The difference between the ranking of journal scores and the ranking of impact factor after literature classification is relatively large, but the ranking of journal normalized Eigenfactor and the ranking of impact factor are similar. These research directions, loading journals, selected indicators, and temporal effects in three types of core documents were revealed in the study, which can provide a certain reference for promoting scientific research in AI and launching scientific research management services.

KEYWORDS

Article-Level Metrics; Journal-Level Metrics; Core documents; Entropy weight method; Principal component analysis; Artificial intelligence

1 Introduction

Professor Bradford, a famous British bibliographer, once felt from his long-term scientific research and literature work that the distribution of relevant papers in periodicals for a certain topic, specialty and subject area was extremely uneven. Papers that are useful to a scholar are not only concentrated in professional journals of their own discipline, but also may be scattered from time to time in journals of other disciplines (Brookes, 1969). In fact, in the actual distribution of journal papers, this distribution phenomenon is universal. For a particular discipline and specialty, a few journals contain a large amount of relevant information, while most journals have a small amount of relevant information. That is, a large number of papers in a given subject are highly concentrated in "core journals" (Chen & Leimkuhler, 1986; Drott et al., 1979; Goffman & Morris, 1970).

At present, the method of bibliometrics is usually adopted for core journals. First, indexes such as journal citation amount and impact factor provided by SCI and its subsidiary product JCR are used. The second is to measure the core journals according to certain steps of metrology. Meanwhile, traditional bibliometrics indexes such as the citation frequency and the journal impact factor are also challenged. For example, the lack of evaluation justice caused by a defect in time lag which is inherent for citation analysis (Adams, 2014; Yang, 2011), selection obstacle caused by the redundancy and complexity of indexes (van Ypersele, 2013), single assessment results that are not scientific and comprehensive enough (Lane, 2010), value orientation bias caused by index dissimilation and excessive focus (Werner, 2015), etc. In the traditional academic evaluation system, the practice of evaluating articles through publishing and individual or institution through influence factors has drawn more and more disputes gradually. Under this background, more and more scholars start to pay attention to the selective and comprehensive evaluation with multi-index (Chen & Guan, 2011; Garfield, 2009). Meanwhile, it restores the essence of scientific research evaluation-the research of building the dynamic measurement of single document value and hierarchic evaluation system has gained more attention, too.

2 Related study

At present, in the view of Article-Level Metrics, the research of the value measurement and assessment of a single paper mainly concentrates on many aspects such as the presentation and promotion of the evaluating indicator of a single paper (Priem et al., 2012), data source and tool selection (Qiu & Yu, 2015), index change rule (Wang et al., 2014), applicability discussion (Adie & Roe, 2013) and interdependency comparative analysis (de Winter, 2015). Relevant research on the application of the comprehensive traditional document measuring index and substitute measuring index in the identification and value measuring of core documents is relatively obvious. In foreign countries, Handel (2014) proposed the idea of combining the traditional citation index and Altmetrics index to cover the evaluation of long-term and short-term influence. In China, Wang also put forward the idea of constructing a continuous, dynamic and compound single paper evaluation system, and selected cited frequency, HTML views, PDF downloads and 7 indicators of Altmetrics series which is the data of Facebook, Twitter, Mendeley and CiteULike, through the use of AHP in different periods of different ways to assign different weight index, carried out empirical research (Wang et al., 2014; Wang et al., 2015). However, it is undeniable that the correlation between the Altmetrics series index and the cited frequency is questionable (Haustein et al., 2014), and the use of these indicators in the evaluation of the academic influence of the literature is also controversial (Bornmann, 2014). Especially for some classic literature identification and recommendation on the effectiveness of the use of social media indicators have yet to be discussed.

Based on the summary of present situation and problems, with the applicable aim of literature recommendation and tracking, this paper puts forward the research ideas of value measure of classic- popular- frontier literature, the identification and value measuring of core document, aiming to meet the specific needs of different groups, by recommending the core literature , and at the same time ordering the core document set according to objective evaluation measure; Take artificial intelligence field, which is a representative field for distinctive interdisciplinary and high overlapping utilization of new and old literature as an example to carry out evidential research, and conduct related analysis of results obtained from Article-Level Metrics (ALM) and data of Journal-Level Metrics (JLM), so that to further show the difference of two measurement methods in the application of realizing core literature recommendation when facing different scientific research requirements, thus giving objective true response to the question of "Judge Paper by journal" on a certain level.

3 Data and indexes

The general research thought of this paper is: get three basic literature collections of the traditional literature, the classic literature getting from the network database and the popular literature and the frontier literature through limiting the threshold of the total citation frequency, the annual average citation frequency and the index of network concern and influence to realize the initial recognition of the three types of core literature. Then establish corresponding index systems for the three kinds of documents, conduct the value measurement of literature by using entropy weight method and major-component-analysis method respectively, and finally realize the ranking and recommendation of core documents meeting the needs of different scientific researches. The correlation analysis of the indicators obtained from the two measurement perspectives of ALM and JLM is carried out, and then the difference in the core document identification and recommendation of ALM and JLM is found. The research process design is shown in Figure 1.

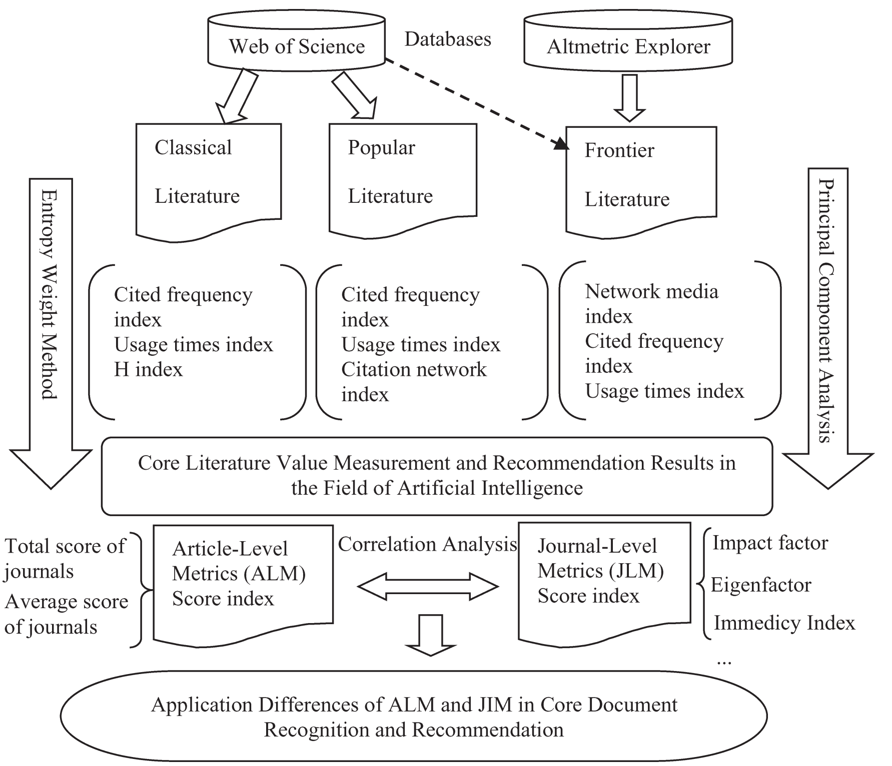

Figure 1 Research approach of the paper

In the course of the study, the bibliometrical method will be adopted to realize the identification and establishment of initial literature collection, and it obtained the indexes such as cited frequency and H index; Synthetically use the methods of citation analysis and social network analysis to structure the network of co-citation, coupling and cross-citation of literature composed by popular literature collection, and get the corresponding index of network measurement. The above process is mainly realized through VBA programming and Ucinet software. Given the difference in the characteristics of each index of different literature sets, this research uses the entropy method and the principal component analysis method to conduct the comprehensive estimation of the value of classic, the popular literature and the frontier literature. Both methods are objective weight methods. Among them, the entropy method is to use the information entropy to calculate the entropy weight of each index according to the variation of each index and revises the weight of each index through the entropy weight to get a more objective index weight, which is suitable for the objective evaluation of the distribution index of discrete data; Principle component analysis is a method of simplifying data structures by reducing dimensions, which reflects the most of the information of the original multitudinous variables by using the linear combination of a few aggregate variables, and suitable for multitudinous index evaluation on the strong correlations in the indexes of the components and difference and weak correlation among those of the components (Chen, 2013). This method has already had applied cases in the evaluation on the influence of academic theses based on Altmetrics (Zhao et al., 2016). The process of evaluation measurement and index correlation analysis can be achieved through Excel, VBA programming and SPSS software.

3.1 Data acquisition and processing

In this study, we select the literature in the field of artificial intelligence (Computer Science, Artificial Intelligence) before 2016 as the research object, and divide the core literature into the following three categories:

1) Classical Literature

Classical literature is usually accompanied by high citation frequency. In this paper, the threshold of citation frequency is set to 1000 times. The literature collection of subcategories and h index (Schubert, 2008) were obtained from Web of Science. Literature types are limited to Article or Proceedings Paper or Review. After searching, 160 pieces of classical literature were obtained.

2) Popular literature

According to the theory proposed by Price, 5 years can be used as the standard to classify the utilization degree of literature information (Qiu, 2007). Therefore, we set the retrieval time of popular literature as 2011 to 2016, and the retrieval date as March 2016. A total of 164,317 records were obtained from Web of Science. Further screening the core data sets of popular literature, we calculated the annual citation frequency of these literature.

According to the statistics, 1127 pieces of literature with an average annual cited number greater than 9 accounted for more than 20% of the total annual cited number of all literature. According to the Pareto law (Newman, 2005), the secondary core of popular literatures may mainly gather in the first 20% of literature, so 1127 pieces of literatures were selected as the data set of popular literature.

3) Frontier Literature

Frontier literature refers to literature that has not been widely cited because it was published recently. One of the important characteristics of frontier literature is that it attracts extensive attention from other scholars after publication, so we choose Altmetric Explorer database to collect frontier literature data.

Using ISSN number of the 160 journals under Computer Science and Artificial Intelligence of JCR database to retrieve in batch in Altmetric Explorer base. Due to the time lag of citation (Wang et al., 2015), the citation frequency of literature may gradually increase one year after its publication, so select 1-year from the option of "Mentioned in the past" in the database for data acquisition and download, and 2440 articles in the document set can be obtained. This data set was further compared with the records in the Web of Science, and a total of 1072 frontiers of documents were obtained after comparison and deduplication.

3.2 Selection of article–level metrics

According to the characteristics of the three types of literature, we designed different article-level Metrics sets:

1) Classical Literature

For classical literature, we select the total citations (TC), h index, annual average total citations and the clicking times or saving times of full document of the literature (U2). U2 is a newly-added index of Web of Science, and this counting is considered to be able to reflect the times a certain document meets the information needs of users (Thomson Reuters, 2016), so this research also includes it in the measurement scope. Finally, the entropy weight method was used to integrate the four indexes, and the standardized index values were used to calculate, and the comprehensive score of each literature was obtained.

2) Popular literature

For popular literature, we also choose TC, U2 and annual average total citations to describe. At the same time, we construct the direct citation network, co-citation network and coupling network. Then the degree centrality of the nodes of the three networks is calculated. The following four indexes are obtained: direct citation network-indegree, direct citation network-outdegree, co-citation network-degree centrality, coupling network-degree centrality. The entropy weight method was used to calculate the score of the literature as the classical literature.

3) Frontier Literature

Because Altmetric Explorer collects all online attention about a paper according to a series of information sources, it has more than a dozen measurement indexes, and some indicators have too many 0 values and are not related to long-term academic influence, so this study supplemented and selected the indexes, the process is as follows:

a. After data cleaning, the final frontier literature selected by the 8 evaluation indexes include Bloggers, Tweeters, Google+ authors, News outlets, Facebook walls, Wikipedia pages, Mendeley readers, CiteULike readers from Altmetric Explorer.

b. To wholly investigate the relationship between the social media index and citation frequency, which is a traditional academic impact index, obtain all the data mentioned by "any time" in Altmetric Explorer; compare that in the database of Web of Science in accordance with DOI, to get the indexes TC and U2.

c. Correlation analysis with SPSS (as shown in Table 1), we found that Wikipedia pages, Mendeley readers and CiteULike readers were significantly correlated with TC, and these three indexes were selected; Weibo users and TC, the annual average cited, U2 The correlation is 0 and is discarded.

d. Though the six indexes of Bloggers, Tweeters, Google+authors, News outlets and Facebook walls have a relatively low or negative correlation with citation frequency, they have a more obvious correlation with U2. In addition, many other studies also show that these indexes are significantly associated to some extent with the citation frequency (Eysenbach, 2011; Wang et al., 2015; You et al., 2014). But since the studies in the artificial intelligence field usually cross with each other and people concerned about the studies are scattered, so it may make the data not accurate. As a whole, however, Altmetric score is still significantly correlated with citation frequency, and the score is mainly calculated with the above indexes. Perhaps some of the studies in this field have attracted the attention of mass netizens indeed, but they have not gained the attention of scholars in the industry and are not cited by them. To some extent, it reflects that the web impact is different from the realistic academic impact. But from the long-term perspective, articles with relatively strong web impact can also produce certain academic impacts. Therefore, in the frontier literature value measurement part, we still use these indexes.

e. Principal component analysis (PCA) was used to carry out value measurement. Based on the comprehensive consideration of the results of total variance interpretation and the situation of gravel plots, the number of principal components was limited to 3 in this study. According to the component score coefficients of the 10 indexes in the Frontiers Collection and the contribution ratios of the explanatory variance of the three components respectively, the final score of each literature was calculated by compound weighting.

Table 1 Pearson correlation analysis results of each index.

	Score	Blog-gers	Tweet-ers	Google+ authors	News outlets	Face-book walls	Weibo users	Wikipedi a pages	Mende-ley readers	CiteU-Like readers
Score	1	.572**	.752**	.614**	.682**	.304**	.180**	.167**	.037**	.082**
TC	.029*	.013	−.011	.009	−.010	−.007	.000	.284**	.550**	.316**
Average Annual Citation	.049**	.017	.018	.030*	−.007	.007	.000	.287**	.603**	.334**
U2	.058**	.001	.080**	.062**	.004	.024	.000	.077**	.209**	.146**

Note: * Indicates Significant Correlation at the 0.05 Level (Bilateral), ** Indicates Significant Correlation at the 0.01 Level (Bilateral).

4) Journal–Level Metrics

According to the journal to which the literature belongs, the sum of the scores of the literature can be obtained. In this paper, the total score and the average score are used to describe the Journal-Level Metrics (JLM) of the journal. The relatively well-known index in JLM is Impact Factor (Mering, 2017). The impact factor was used as the control index in this paper.

4 Results and discussion

4.1 Recommendation Results in the Field of Artificial Intelligence

After calculating, the recommended results of the three types of literature are obtained, as shown in table 3. In the classical literature, the most recommended is Lowe's "Distinctive image features from scale-invariant key points" (Lowe, 2004), which is one of the most classic papers in the field of image recognition scale-invariant feature conversion algorithm (SIFT). Followed by Breiman's "Random forests" (Breiman, 2001), which first proposed the random forest method is the foundation of machine learning. The third one is Kennedy and Eberhart's "Particle swarm optimization" (Kennedy & Eberhart, 1995), which first proposed the particle swarm optimization (PSO) and is an important milestone in the field of artificial intelligence. After discussion with the relevant experts in this field, we can find that the above results are in line with the classical literature of artificial intelligence.

In the popular literature, four of the top ten papers are about 1/Hesitant fuzzy information aggregation in decision making (Xia & Xu, 2011), 9/Hesitant fuzzy linguistic term sets for decision making (Rodriguez et al., 2012), 10/On distance and correlation measures of hesitant fuzzy information (Xu & Xia, 2011) and 2/Hesitant fuzzy prioritized operators and their application to multiple attribute decision making (Wei, 2012). It is also a hot research direction in the field of artificial intelligence, which is widely used in pattern recognition, data mining and decision analysis. In addition, 3/4/7 are all about the feedback tracking and control system, and the extreme learning machine regression fitting and classification, differential evolution algorithm is also the research hotspot in this field (Tong et al., 2011; Zhou et al., 2012; Zhou et al., 2011).

For the frontier literature, what requires special attention is that the document "Deep learning in neural networks: An overview" (Schmidhuber, 2015) gets the highest score and has a big gap with following literature, which reflects that deep learning on neural networks is the focal question noted by researchers and network users and also the leading topic in this field. In addition, frontier researches on Brainprint, ImageNet, Turing tests, Human-Computer Interaction Intelligence, Smart Home Internet and so on have aroused considerable concern, which is also the main direction for innovative researchers to explore.

Table 3 Recommended results for the three types of core literature

Ranking	Classical	Popular	Frontier
1	Distinctive image features from scale –invariant key points. International Journal of Computer vision,2004	Hesitant fuzzy information aggregation in decision making. International Journal of Approximate Reasoning, 2011	Deep learning in neural networks: An overview. Neural Networks,2015
2	Random forests. Machine Learning,2001	Hesitant fuzzy prioritized operators and their application to multiple attribute decision making. Knowledge–based Systems,2012	Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics. Neurocomputing,2015
3	Particle swarm optimization. 1995 IEEE International Conference On Neural Networks Proceedings,1995	Observer –based adaptive fuzzy backstepping control for a class of stochastic nonlinear strict –feedback systems. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics,2011	Soft robotic glove for combined assistance and at–home rehabilitation. Robotics and Autonomous Systems,2015

Ranking	Classical	Popular	Frontier
4	A fast and elitist multiobjective genetic algorithm: NSGA –II. IEEE Transactions on Evolutions Computation,2002	Adaptive output –feedback fuzzy tracking control for a class of nonlinear systems. IEEE Transactions on Fuzzy Systems,2011	Towards computational models of animal cognition, an introduction for computer scientists. Cognitive Systems Research, 2015
5	Support –vector networks. Machine Learning,1995	Differential evolution: A Survey of the State –of –the –Art. IEEE Transactions on Evolutionary Computation,2011	ImageNet large –scale visual recognition challenge. International Journal of Computer vision,2015
6	Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine, 1984	Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics,2012	Using the collective intelligence for inventive problem solving: A contribution for open computer–aided innovation. Expert Systems with Applications,2015
7	A theory for multiresolution signal decomposition – the wavelet representation. IEEE Transactions on Pattern Analysis and Machine, 1989	Neural–network–based decentralized adaptive output –feedback control for large –scale stochastic nonlinear systems. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics,2012	A Panorama of artificial and computational intelligence in games. IEEE Transactions on Computational Intelligence & AI in Games,2015
8	A computational approach to edge –detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986	Global contrast–based salient region detection. 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , 2011	Towards computational models of animal communications, an introduction for computer scientists. Cognitive Systems Research,2015
9	A tutorial on Support Vector Machines for pattern recognition. Data Mining and Knowledge Discovery,1998	Hesitant fuzzy linguistic term sets for decision–making. IEEE Transactions on Fuzzy Systems,2012	Human misidentification in Turing tests. Journal of Experimental & Theoretical Artificial Intelligence,2015
10	Bagging predictors. Machine Learning,1996	On distance and correlation measures of hesitant fuzzy information. International Journal of Intelligent Systems,2011	Visual simultaneous localization and mapping: A survey. Artificial Intelligence Review,2015

4.2 Article–Level Metrics (ALM) vs Journal–Level Metrics (JLM)

In the Article-Level Metrics (ALM) and measurement results, it has been revealed that there is a large difference in the distribution of journals of different types of core literature, such as the top of the classic literature mostly distributed in the "Machine Learning", "IEEE Transactions on Pattern Analysis and Machine" and other journals. The top popular literature is distributed in IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics, IEEE Transactions on Fuzzy Systems and other journals. The top literature is distributed, but it seems to be more advanced in the research of neural network/computing and cognitive systems.

The paper carries on the comprehensive statistics to a number of articles published in the three kinds of literature journals collection. It is found that the journals with the most papers published are IEEE T PATTERN ANAL, INT J COMPUT VISION, MACH LEARN; IEEE T PATTERN ANAL, EXPERT SYST APPL, IEEE T IMAGE PROCESS; MED IMAGE ANAL, EXPERT SYST APPL, AUTON ROBOT.

The corresponding journal's published paper scores are summarized, that is, the total score and the average score of each periodical in different literature collections are obtained, which is the summary of the paper's scores, as shown in Table 4 (JTA-Journal Title Abbr. , PA-Paper Amount, SS-SUM Score, AS-Average Score). Most of the top 10 journals in the three categories are not listed in the top 10 of the influencing factor rankings (see the red-labeled journals in the table), regardless of the number of papers published from the journal, or the paper score. In particular, the top ten sources in the literature are only the eighth most influential factor of MED IMAGE ANAL, and the others are 15 or even 20.

The results also showed that the recommended literature generated by the ALM were not published in the journals with high impact, but the consistency with the Normalized Eigenfactor was more obvious. For example, IEEE T IMAGE PROCESS impact factor ranked 11th, and in the classic and popular literature of ALM total scores in the fourth and fifth, Normalized Eigenfactor ranked third; EXPERT SYST APPL ranks second in popular and frontier literature of ALM total scores and Normalized Eigenfactor rankings, but its impact factor ranking is 19. Some of the leading ALM-SS journals are not even in the Q1 of JCR ranking, such as NEUROCOMPUTING in the popular and frontier literature of ALM total scores in the top ten, but in the 2014 JCR partition has been down out of Q1 the new version of 2015 was a slight advantage into Q1. ARTIF INTELL MED, AUTON ROBOT belongs to Q2 journals in JCR ranking. COGN SYST RES belongs to Q3 journals in JCR ranking. Both of them are in the top ten in the ALM scores of the frontier literature collection. It is still significantly different between the results of "to review the article with publication" and "to review the article with pieces".

Table 4 Total score and average score of three types of literature journals (Top 10)

Num.	Classic				Popular				Frontier			
	JTA	PA	SS	AS	JTA	PA	SS	AS	JTA	PA	SS	AS
1	IEEE T PATTERN ANAL	49	5.776	0.118	IEEE T FUZZY SYST	82	10.917	0.133	NEURAL NETWORKS	52	1.137	0.022
2	INT J COMPUT VISION	15	1.945	0.13	EXPERT SYST APPL	107	8.504	0.079	EXPERT SYST APPL	76	0.851	0.011
3	MACH LEARN	11	1.884	0.171	IEEE T PATTERN ANAL	116	7.969	0.069	MED IMAGE ANAL	84	0.611	0.007
4	IEEE T IMAGE PROCESS	9	0.992	0.11	IEEE T NEUR NET LEAR	68	7.1	0.104	NEUROCOMPUTING	41	0.47	0.011
5	IEEE T EVOLUT COMPUT	6	0.95	0.158	IEEE T IMAGE PROCESS	105	7.002	0.067	AUTON ROBOT	64	0.397	0.006

Num.	Classic				Popular				Frontier			
	JTA	PA	SS	AS	JTA	PA	SS	AS	JTA	PA	SS	AS
6	NEURAL COMPUT	8	0.806	0.101	KNOWL –BASED SYST	48	4.717	0.098	COGN SYST RES	15	0.343	0.023
7	NEURAL NETWORKS	6	0.562	0.094	IEEE TRANS CYBERN	41	4.669	0.114	ROBOT AUTON SYST	22	0.336	0.015
8	J MACH LEARN RES	4	0.466	0.117	APPL SOFT COMPUT	74	4.379	0.059	APPL SOFT COMPUT	27	0.231	0.009
9	CHEMOMETR INTELL LAB	3	0.465	0.155	NEURAL NETWORKS	38	3.691	0.097	DECIS SUP-PORT SYST	30	0.224	0.007
10	ARTIF INTELL	4	0.35	0.088	NEURO-COMPUTING	45	2.997	0.067	ARTIF INTELL MED	29	0.212	0.007

In order to obtain a more clear relationship and the difference between results of measurement at article level and measurement at journal level, the author analyses the total score, the average score, the impact factors, the characteristic factors, the immediacy indexes and so on of the periodical after summary. And the results are shown in Table 5. In general, ALM-SUM Score in the popular literature is significantly associated with other indexes. Regarding the value measurement and recommendation of popular literature, measurement results at article level are relatively consistent with those of academic journals. Besides Immediacy Index, ALM-SUM Score for classical literature has a relatively high correlation with other traditional journal measuring indexes. While ALM-Average Score doesn't have an obvious correlation with all the other indexes, and it even has a negative correlation with Immediacy Index. It is clear that for the whole classical literature, the two measuring methods have consistency at a certain level, but the consistency is higher for popular literature; As for the accurate recommendation of single classical literature, differences between the two results are very obvious. Although there is a significant correlation above the level of 0.05 between the score index of the frontier literature and most indexes, the correlation coefficient is not high, which is no more than 0.5. So, it can be seen that the gap of the results worked out by two kinds of measurement methods is still evident; but it is worth noting that among the three indexes of ALM-Average Score, Immediacy Index only has a prominent correlation with this index of the latest information, and so does Article Influence Score. It reflects that to find and recognize the latest information in the field, besides building an index system on ALM, we can also refer to the immediacy index or journal articles with high Article Influence Score on a certain level. In addition, the data still represents that the relevance between Normalized Eigenfactor and third-kind document ALM-SUM Score is the highest, which is particularly evident higher than the Impact Factor which is more emphasized by people. This is closely related to the connotation and essence of Normalized Eigenfactor. Normalized Eigenfactor is the normalization of Eigenfactor, both of them have the same effects in expressing journal features. Eigenfactor investigates not only the quantity of citation, but it also takes the influence of the journals into consideration, which means the more a journal is cited by influential journals, the higher the influence it will have. Eigenfactor evaluates the importance of each paper (or each web page) based on the whole social network structure. Therefore, this explains the significant correlation between the indexes of Normalized Eigenfactor and

ALM-SUM Score.

Table 5 Correlation analysis results of ALM-JLM indexes

	Classic		Popular		Frontier	
	ALM–SUM Score	ALM–Average Score	ALM–SUM Score	ALM–Average Score	ALM–SUM Score	ALM–Average Score
ALM–SUM Score	1	0.153	1	.298**	1	.269**
ALM–Average Score	0.153	1	.298**	1	.269**	1
Impact Factor	.466*	.077	.617**	.194	.272**	.234*
Immediacy Index	.187	–.188	.546**	.170	.482**	.208*
5–Year Impact Factor	.590**	0.093	.570**	.260*	.222*	.236*
Article Influence Score	.668**	0.1	.408**	0.151	0.169	.351**
Normalized Eigenfactor	.677**	0.013	.770**	0.119	.482**	0.145

Note: * Indicates Significant Correlation at the 0.05 Level (Bilateral), ** Indicates Significant Correlation at the 0.01 Level (Bilateral).

5 Conclusion

In view of the bias and limitations in simply using cites or journal impact factors to evaluate the literature value in the past, and at the same time taking into account the different needs of different types of researchers for the access, identification and recommendation of core documents in the field, the paper proposes the line of thinking to divide the core document collection into three categories of document objects on different levels, including the classic literature, the popular literature and the frontier literature; And then the paper Article-Level Metrics thesis level evaluation index system: the classic literature index mainly citation index, and the index number and paper H index; The measurement indexes of popular literature consist mainly of cited frequency index, usage times index and cited network index; the measurement indexes of leading documents consist mainly of network media index, cited frequency index and the index of usage times. According to the characteristics of indicators and data, we used entropy weight method and principal component analysis to measure the value of classics, popular and frontier literature respectively, and then successfully realize the best core document identification and recommendation for the needs of scientific research. However, the main value measurement method of this paper is to adopt the pure objective weighting method. Although on a certain level, it avoids human intervention, it will inevitably increase contingency and uncertainty if simply from data. Therefore, future research will explore the subjective and objective comprehensive weighting method to achieve a more rigorous and credible evaluation of comprehensive literature.

In the process of the research, the traditional literature database and the network media database of resources realized docking, thus it made the index of Article-Level Metrics more abundant and complete. Especially for the measurement of the frontier literature, found in different periods, the correlation between the index number of published literature and classification and scale to measure academic value (mainly the cumulative variance explained in common factor variance) and factor accumulated explanation variance altmetrics system in a different period in the future and realized the value measurement of literature more effective.

In addition, through comparison between measurement results and indicators on Article-Level Metrics and traditional journal measurement level, differences and connections were found. Judging from the journal distribution of the core literature, although "to review the article with publication" can recommend excellent papers to a certain degree, it is still significantly different from the result of "to review the article with pieces". The gap between ALM-SS ordering of journals and traditional influence factors ordering is relatively large, while it is much closer to Normalized Eigenfactor ranking. The classic and popular literature which ranks top ten in journals is almost in zone one of JCR, but even the frontier literature has involved in zone two and zone three of JCR, which proposes an objective question about the method on using JCR partition to evaluate the quality of the article on a certain level.

Judging from the overall results of the core document identification and recommendation, there's the most significant correlation between the recognition result at the level of the popular literature thesis and that of the traditional measuring method; secondly, it is the classical literature; thirdly, the difference of the frontier literature is the biggest. This indicates that if you pay attention to the popular papers in recent 5 years, you can adopt the traditional method of "evaluating papers according to the journals" to receive some paper objects in a certain degree and range; However, the classic literature more focuses on the academic value of itself, while the frontier papers more focus on the novelty and potential influence of the thesis. The feasibility of using journals to indirectly locate and accurately get these core literature has been decreasing. The effectiveness is also indeed not enough to use traditional journal evaluation index to test the literature value indirectly. The users of results under both ALM and JLM modes should choose appropriate methods according to specific needs and situations. In the existing evaluation system, it could be betterto use the more easily implemented "review articles by journal" to achieve more accurate recommendations for core literature. This study suggests that the normalized characteristic factor may be a better choice to replace the impact factor index. We also advocate the use of article-level measurement to meet the different needs of literature value measurement and rapid recommendation from customers, which is the direction we will adhere to in the future.

Acknowledgements

This research is funded by Beijing Social Science Fund (grant number 19GLC057) and Project of High-level Teachers in Beijing Municipal Universities -Youth Top-notch Talent Support Program (grant number PXM2020_014204_000019).

Reference

Adams, J. (2014). Bibliometrics: The citation game. *Nature, 510* (7506), 470–471.

Adie, E., & Roe, W. (2013). Altmetric: Enriching scholarly content with article–level discussion and metrics. *Learned Publishing, 26* (1), 11–17.

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics, 8* (4), 895–903.

Breiman, L. (2001). Random forests. *Machine Learning, 45* (1), 5–32.

Brookes, B. C. (1969). Bradford´s law and the bibliography of science. *Nature, 224* (5223), 953–956.

Chen, K., & Guan, J. (2011). A bibliometric investigation of research performance in emerging nanobiopharmaceuticals. *Journal of Informetrics, 5* (2), 233–247.

Chen, Y., & Leimkuhler, F. (1986). A relationship between lotka´s law, bradford´s law, and zipf´s law. *Journal of* *the American Society for Information Science, 37*, 307–314.

Chen, Z. (2013). Technology and application of comprehensive evaluation (in chinese). *Press of Southwestern University of Finance and Economics.*

de Winter, J. C. F. (2015). The relationship between tweets, citations, and article views for plos one articles. *Scientometrics, 102* (2), 1773–1779.

Drott, M., Mancall, J., & Griffith, B. (1979). Bradford´s law and libraries: Present applications–potential promise. *Aslib Proceedings – ASLIB PROC, 31*, 296–304.

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research, 13* (4), e123.

Garfield, E. (2009). From the science of science to scientometrics visualizing the history of science with histcite software. *Journal of Informetrics, 3* (3), 173–179.

Goffman, W., & Morris, T. G. (1970). Bradford´s law and library acquisitions. *Nature, 226* (5249), 922–923.

Handel, M. J. (2014). Article–level metrics–it´s not just about citations. *Journal of Experimental Biology, 217* (24), 4271–4272.

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the American Society for Information Science & Technology, 65* (4), 656–669. (in Chinese).

Kennedy, J., & Eberhart, R. (1995, 27 Nov.–1 Dec. 1995). Particle swarm optimization. *Proceedings of ICNN´95 – International Conference on Neural Networks.* IEEE.

Lane, J. (2010). Let´s make science metrics more scientific. *Nature, 464* (7288), 488–489.

Lowe, D. G. (2004). Distinctive image features from scale–invariant keypoints. *International Journal of Computer Vision, 60* (2), 91–110.

Mering, M. (2017). Bibliometrics: Understanding author–, article– and journal–level metrics. *Serials Review, 43* (1), 41–45.

Newman, M. E. J. (2005). Power laws, pareto distributions and zipf´s law. *Contemporary Physics, 46* (5), 323–351.

Priem, J., Groth, P., & Taraborelli, D. (2012). The altmetrics collection. *PLOS ONE, 7* (11), e48753.

Qiu, J. (2007). *Information measurement*. Wuhan University Press.

Qiu, J., & Yu, H. (2015). Some basic problems in advancing development of altmetrics. *Journal of Library Science in China, 2015* (1), 4–15.

Rodriguez, R. M., Martinez, L., & Herrera, F. (2012). Hesitant fuzzy linguistic term sets for decision making. *IEEE Transactions on Fuzzy Systems, 20* (1), 109–119.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117.

Schubert, A. (2008). Using the h–index for assessing single publications. *Scientometrics, 78* (3), 559.

Thomson Reuters. (2016). *Usage times*. Retrieved June 20 from http://images.webofknowledge.com/WOKRS522_1R3/help/zh_CN/WOS/hp_usage_score.html

Tong, S., Li, Y., Li, Y., & Liu, Y. (2011). Observer–based adaptive fuzzy backstepping control for a class of stochastic nonlinear strict–feedback systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 41* (6), 1693–1704.

van Ypersele, J. P., et al. (2013). The maze of impact metrics. *Nature, 502* (7471), 271–271.

Wang, X., Fang, Z., & Wang, H. (2015). Study on the continuous, dynamic and comprehensive article–level e–valuation system. *Science of Science and Management of S.& T., 36* (8), 37–48. (in Chinese)

Wang, X., Mao, W., Xu, S., & Zhang, C. (2014). Usage history of scientific literature: Nature metrics and metrics of nature publications. *Scientometrics, 98* (3), 1923–1933.

Wang, Y., Guo, S., & Zhang, J. (2015). Study on correlation of researcher influence indicators. *Library and Information Service, 59* (5), 106–127. (in Chinese)

Wang, Y., Ruifeng, H., & Die, H. (2015, 15–17 Aug. 2015). Citing time lag: A new determinant for paper quality? A case study on management journal in China. *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD).* IEEE.

Wei, G. (2012). Hesitant fuzzy prioritized operators and their application to multiple attribute decision making.

Knowledge–Based Systems, 31, 176–182.

Werner, R. (2015). The focus on bibliometrics makes papers less useful. *Nature, 517* (7534), 245–245.

Xia, M., & Xu, Z. (2011). Hesitant fuzzy information aggregation in decision making. *International Journal of Approximate Reasoning, 52* (3), 395–407.

Xu, Z., & Xia, M. (2011). On distance and correlation measures of hesitant fuzzy information. *International Journal of Intelligent Systems, 26* (5), 410–425.

Yang, S. (2011). The problems of citation analysis and their causes. *Journal of Library Science in China, 37* (3), 108–117.

You, Q., Wei, B., & Tang, S. (2014). Study on correlation of different article–level metrics. *China Academic Journal Electronic Publishing House, 58* (8), 79–84.

Zhao, R., Guo, F., & Tan, J. (2016). Evaluation of academic papers impact based on altmetrics, a case study of Chinese language and literature. *Journal of Library Science in China, 2016* (1), 96–108.

Zhou, Q., Shi, P., Liu, H., & Xu, S. (2012). Neural–network–based decentralized adaptive output–feedback control for large–scale stochastic nonlinear systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42* (6), 1608–1619.

Zhou, Q., Shi, P., Lu, J., & Xu, S. (2011). Adaptive output–feedback fuzzy tracking control for a class of nonlinear systems. *IEEE Transactions on Fuzzy Systems, 19* (5), 972–982.