Research Progress on Academic Paper Citation Prediction Postprint

Authors: Xia Wanjun, Xiaohong Chen, Jiang Yanping

Date: 2023-04-01T00:00:00+00:00

Abstract

[Objective/Significance] To systematically review the influencing factors and prediction methods for academic paper citation prediction, analyze existing challenges, and propose future research directions.

[Method/Process] This study employs a literature survey methodology to review domestic and international research advances, summarizing the relevant content and characteristics of prediction influencing factors and methods.

[Results/Conclusions] Current research exhibits numerous influencing factor indicators without unified standards, weak theoretical foundations for prediction methods, insufficient investigation into the dynamic nature of citation prediction, and limited generalizability of prediction models. Future research should strengthen theoretical investigations in citation prediction, enhance the integration of traditional bibliometrics with altmetrics, deepen the application of natural language processing, establish unified baseline standards, and develop more accurate prediction models.

Full Text

Research Progress on Academic Paper Citation Prediction

Xia Wanjun[1],[2], **Chen Xiaohong**[1], **Jiang Yanping**[1] [1]Library of Southwest Jiaotong University, Chengdu 611756 [2]School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756

Abstract: [Purpose/Significance] This paper systematically reviews the influencing factors and prediction methods for academic paper citation prediction, analyzes existing problems, and proposes future development directions. [Method/Process] Using literature research methodology, we review domestic and international research progress, summarizing the relevant content and characteristics of prediction influencing factors and methods. [Result/Conclusion]

Current research faces several challenges: numerous influencing factor indicators without unified standards, weak theoretical foundations for prediction methods, insufficient research on the dynamic nature of citation prediction, and limited generalizability of prediction models. Future work should strengthen theoretical research on citation prediction, integrate traditional bibliometrics with altmetrics, deepen the application of natural language processing, establish unified baseline standards, and construct more accurate prediction models.

Keywords: citation prediction; influencing factors; prediction methods

Academic papers serve as crucial media for disseminating scientific knowledge, with most new technologies and discoveries made public through this channel. New achievements typically build upon previous work and cite others' literature, reflecting the inheritance and development of scientific research. Due to the rapid development of scientific research, massive numbers of academic papers are produced annually, growing exponentially. This makes it increasingly challenging for researchers to quickly identify influential papers from vast literature resources, especially those recently published with few citations but representing cutting-edge research. Currently, the simplest, most effective, and objective metric for measuring paper impact is citation frequency, which is widely considered to reflect contributions to scientific progress. Consequently, scientific evaluation often relies on citation counts. Predicting citation frequency not only helps researchers identify valuable papers but also assists administrators in resource allocation, making it a task of significant practical value. Therefore, this study focuses on predicting the citation status of individual papers—that is, citations *to* the paper from other works, rather than citations *by* the paper. While numerous studies exist on this topic, including review articles such as Bao Yufang et al.'s summary of common citation prediction methods, these approaches have limitations, primarily relying on regression analysis without comprehensive systematic examination. Given this context, this paper employs literature research methodology to systematically analyze paper citation prediction research, with particular emphasis on recent advances, to provide reference for future work.

2. Data Sources and Analysis

To understand the current state of academic paper citation prediction research domestically and internationally, we conducted preliminary searches using the keywords "citation" and "prediction/predicting/predictor." We formulated Chinese search queries as "主题 =(论文 or 文献) and (引用 or 被引 or 引文 or 影响力) and 预测" and English queries as "Title=citation* and predict*," searching in CNKI, Web of Science Core Collection, SDOS, and IEEE Xplore. After deduplication, manual screening, and reference-based expansion, we selected 25 Chinese and 112 English papers based on maximum relevance. English literature on this topic dates back to the 1980s, attracting attention from library and information science, computer science, life sciences, economics, and other fields. In contrast, Chinese literature is relatively limited and concentrated in recent five years, indi-

cating that research in this area is still in its infancy in China. Further analysis reveals that existing citation prediction methods primarily achieve prediction by selecting relevant influencing factors for model construction. Therefore, this paper examines citation prediction from two perspectives: influencing factors and specific prediction methods.

3. Research Status of Citation Prediction Influencing Factors

3.1 Diverse and Open Influencing Factors

Researchers often investigate factors affecting paper citations to increase their own work's visibility. The citation process is complex, influenced not only by scientific content but also by other factors including publishing journals, author reputation, and social impact. Early studies on citation prediction influencing factors primarily considered indicator accessibility. As research progressed, investigators began incorporating additional information, such as temporal dynamics of citation patterns and social media data, to predict future citations, providing new perspectives for factor analysis. Current research on the relationship between future citations and influencing factors can be categorized into paper-related factors, author-related factors, journal-related factors, temporal factors, and altmetric factors, with specific indicators and effects shown in Table 1 .

Table 1. Paper Citation Prediction Influencing Factors and Their Effects

Category	Specific Indicators	Effects
Paper Factors	Short-term citation history	More short-term citations correlate with higher future citation probability [7]

Category	Specific Indicators	Effects
	References (quantity, age, impact, diversity)	Moderate correlation between reference count and citations [10]; lower average reference age increases future citation likelihood [11-12]; papers citing high-impact references receive more citations [13]; diverse reference fields increase attention and citation probability [7]
	Title characteristics	Entertaining titles, compound titles, or question titles attract more citations [14]
	Paper length	Longer articles, often with more detailed methods and results, increase scientific impact and dissemination, making future citations more likely [8]

Category	Specific Indicators	Effects
	Abstract and keywords	Keyword frequency in abstracts and journal-level keyword frequency show significant positive correlation with future citations [15]
	Review articles	Review articles typically receive more citations than research papers [9]
	Interdisciplinarity	Interdisciplinary papers tend to gain more future citations [7]
Author Factors	h-index and derivatives	Papers by high h-index authors are more likely to be cited [16]
	Previous citation frequency	Authors with high previous citations tend to gain more citations [17]
	Publication quantity	More published papers correlate with higher future citations [7]
	Field diversity	Authors publishing across diverse fields gain more citations [7]

Category	Specific Indicators	Effects
	Number of authors	Positive correlation between author count and future citations [13]
	International collaboration	Significant positive correlation between international collaboration and citation rates [13]; interdisciplinary collaboration promotes citation increase [8]
Journal Factors	Impact (impact factor, citation frequency)	Positive correlation between journal influence and future paper citations [18]
Temporal Factors	Publication time	Citation probability decays exponentially over time [19]
	Time to first citation	Papers receiving first citations faster tend to accumulate more citations [10]
	Citation accumulation rate	Faster citation accumulation predicts higher future citation counts [20]

Category	Specific Indicators	Effects
Altmetric Factors	Usage (clicks, downloads, saves, views)	Moderate or significant positive correlation between usage metrics and future citations [21-25]
	Reference management tool users	
	Social media mentions	

3.2 Characteristics of Influencing Factor Research

3.2.1 Multi-dimensionality. Early research examined relationships between future citations and individual factors through correlation analysis. However, single-category factors contain limited information, so researchers have attempted to integrate multi-dimensional factors and analyze indicator importance. T. Chakraborty used 16 indicators encompassing paper, author, and journal factors to build prediction models, finding that removing each feature individually reduced overall accuracy to varying degrees [7], with author factors being the most effective. Geng Qian et al. utilized 23 features across the same three categories for prediction, discovering that using any single category or any two-category combination yielded worse results than using all factors. Among them, paper download counts, reference numbers, and author citation counts consistently ranked high in importance across different prediction periods [1]. R. Yan examined different feature groups, finding that the combination of paper, author, and journal factors achieved the best prediction effect ($R^2 = 0.927$), while using single categories yielded maximum R^2 of only 0.659, with author influence and journal impact being the most important indicators [16].

3.2.2 Cross-domain. Most current citation prediction research lacks generalizability because data is limited to specific fields, and identified factors often apply only to single disciplines. To explore universal patterns, researchers have begun investigating multi-disciplinary applications. D. Wang's *Science* paper derived a citation dynamics model for individual papers, finding that papers across different disciplines and journals follow similar temporal patterns, suggesting common temporal factors enable cross-domain long-term citation prediction [26]. N. Onodera selected six different disciplines, identifying common influencing indicators such as recent five-year reference ratio and reference count [27]. F. Didegah studied citation factors across three disciplines, finding that while the same indicators had varying effects across fields, common factors increasing citations included journal impact, reference influence, and reference count [13]. These common indicators can be applied across disciplines for cita-

tion prediction.

3.2.3 Real-time capability. Networked scientific communication has dramatically improved dissemination efficiency, giving rise to altmetrics. Altmetric data collected through public APIs is open and accumulates rapidly [28], compensating for the time lag of traditional bibliometric indicators. As altmetric indicators have matured in academic data applications, they have introduced new influencing factors for citation prediction, enriching the existing indicator system. Xiong Zequan et al. found that early high-download and low-download papers are more predictive [21]. H. Shema demonstrated that articles cited in science blogs tend to receive more subsequent citations [22]. B. K. Peoples discovered strong positive correlation between Twitter mention counts and citations, outperforming journal impact factors in predicting citation rates [23]. D. Zoller found moderate correlation between BibSonomy additions, views, exports, and queries with future citations [24]. M. Thelwall studied multiple Altmetric.com indicators, finding Mendeley reader count to be a consistent predictor of future citation impact [25].

4. Research Status of Prediction Methods

4.1 Diverse and In-depth Prediction Methods

With advancements in artificial intelligence, machine learning algorithms have demonstrated excellent performance in numerous prediction tasks. In academic paper citation prediction, researchers have applied machine learning to large-scale academic data, yielding three main approaches:

4.1.1 Statistical Methods. Statistical methods were the most widely used approach in early citation prediction research. These methods analyze relevant feature indicators to obtain statistical data for predicting future citation counts. Current statistical methods fall into two categories:

(1) **Regression Analysis.** To determine causal relationships between influencing factors and future citations, most scholars employ regression analysis, including stepwise regression, negative binomial regression, linear regression, and quantile regression [10, 27, 30-31]. T. Yu used stepwise regression to predict five-year impact in library and information science; C. Stegehuis employed quantile regression to predict probability distributions of citations five and fifteen years after publication [31].

(2) **Custom Models.** M. E. J. Newman utilized Z-scores (calculating the mean and standard deviation of citations for papers published in a time window, then computing the standard deviation of a paper's citations from the mean) to predict highly-cited papers in physics [32-33].

4.1.2 Machine Learning Methods. Machine learning algorithms have shown superior performance in many prediction tasks. In citation prediction, three main approaches exist:

(1) **Classification.** Many researchers treat citation prediction as a classification problem due to better generalization ability. Various classification standards have been defined: M. Wang [34] defined three classes (high, medium, or low citations after three years); L. D. Fu classified whether citation counts exceed threshold t (t = 20, 50, 100, 500) after ten years [9]; H. S. Bhat used citation distribution percentiles (0, 33%, 66%) for three-class classification [35]. Common algorithms include SVM, Naive Bayes, decision trees, random forest, AdaBoost, and XGBoost, with SVM, random forest, and XGBoost achieving approximately 90% accuracy.

(2) **Clustering.** X. Cao et al. used Gaussian Mixture Models (GMM) to cluster papers with similar citation patterns, revealing multiple future citation trends and their probabilities [36]. This method is simple, effective, and robust.

(3) **Regression.** A. Abrishami treated citation prediction as a regression learning problem, using Recurrent Neural Networks (RNN) as a powerful model for predicting future citation counts [37]. This method achieved excellent results (R^2 up to 0.9) using only citation counts as a feature, representing successful deep learning application in citation prediction.

4.1.3 Graph Model Methods. With the popularity of PageRank and HITS algorithms, graph-based methods have been widely applied to network entity ranking. In academic networks, many studies iteratively rank papers and researchers through citation and co-authorship relationships [38]. For citation prediction, graph models calculate papers' "future scores" for impact ranking, validated against actual citation rankings to indirectly achieve prediction. Two main approaches exist:

(1) **Simple Graph Networks.** N. Pobiedina treated citation count prediction as a link prediction problem in citation networks, introducing GER-score based on frequent graph pattern mining for citation prediction [39]; Chen Chaomei proposed a predictive bibliometric method from a scientific mapping perspective using structural variation models [40-41]. These methods mine citation prediction-related information from network structures but have relatively simple structures that may ignore important factors.

(2) **Complex Heterogeneous Graph Networks.** Academic networks contain multiple entity and relationship types with mutual influences [42], creating highly complex and heterogeneous structures that can reveal more hidden information. FutureRank [43] was among the first to use heterogeneous networks for future citation ranking, constructing paper citation and author-paper networks through random walk iterations, achieving 75% accuracy. Subsequent research has optimized this approach: Liu Dayou et al. improved performance significantly by calculating author authority values without computing paper PageRank values [44]; MRCoRank constructed time-aware weighted networks incorporating textual information

and burst word detection to predict groundbreaking papers [48]; NERank embedded paper, author, and journal nodes into the same low-dimensional vector space while considering global and local network structure information, improving prediction accuracy by 6% over MRCoRank [47].

4.2 Comparison of Different Prediction Methods

A comparative analysis of the three mainstream prediction methods is shown in Table 2 .

Table 2. Comparative Analysis of Different Prediction Methods

Method	Application Scenario	Common Metrics	Advantages	Disadvantages
Statistical	Primarily for exploring relationships between influencing factors and citations, suitable for small datasets	Correlation coefficient (R), coefficient of determination (R^2), root mean square error (RMSE), mean square residual (MSR)	Strong mathematical theory support, rigorous mathematical explanations and inference formulas, good for discovering correlations when data is limited	Prediction accuracy typically lower than machine learning, prone to overfitting, limited explanatory power
Machine Learning	Suitable for large-scale datasets with high-dimensional features	AUC, F1 score, ROC, etc.	High prediction accuracy (up to 90%), can fully utilize high-dimensional features in big data for precise prediction	Lacks interpretability, prone to overfitting

Method	Application Scenario	Common Metrics	Advantages	Disadvantages
Graph Model	Primarily for citation network research	TOP-N ranking accuracy, etc.	Can mine network structure information, assign different weights to citations to distinguish quality	Prediction accuracy typically lower than machine learning, high computational complexity

While both statistical and machine learning methods include regression, they differ fundamentally: statistical regression emphasizes unbiased fitting of historical data, whereas machine learning regression reduces variance to avoid overfitting and achieve higher prediction accuracy. Statistical and machine learning methods typically treat all citations as "equal," while graph models can leverage citation networks, author networks, and other structural information to assign different weights to citations, distinguishing high-quality from low-quality citations or those from influential scholars, thereby more clearly revealing citation trends.

4.3 Characteristics of Prediction Method Research

4.3.1 Big Data. Early citation prediction relied on small-sample statistical analysis with only hundreds of data points. However, big data technology has transformed research paradigms. With increasing data scale, researchers can rapidly extract valuable information from massive datasets. Citation prediction datasets have grown from hundreds to hundreds of thousands or even millions of records. For example, H. S. Bhat applied machine learning to a large dataset of nearly 8 million papers by over 3 million authors, with JSON files reaching 220 GB [35]. Larger sample sizes facilitate better machine learning model construction, as these models often perform better with more data.

4.3.2 Intelligence. Machine learning enables intelligent data processing by leveraging knowledge and value embedded in data [46]. Compared with manually designed models, machine learning avoids subjective interference and focuses on the data itself, automatically learning rules from historical data to predict new data. For instance, L. D. Fu used decision trees to automatically extract feature patterns from bibliometric characteristics, reducing manual intervention [9]. Scholars have also begun integrating prediction algorithms into intelligent retrieval systems, such as R. Yan's prototype design for a personal-

ized paper recommendation system combining citation prediction [16] and Shen Lei's mobile-accessible paper impact prediction system [47]. While these intelligent systems are not yet fully implemented, they provide practical guidance for deploying citation prediction algorithms.

4.3.3 Structuring. Citation prediction no longer relies on single influencing factors but comprehensively considers interactions among author authority, journal impact, and other factors within structured networks. In these networks, each paper node connects to others through citation links, providing more information about authors and papers. For example, S. Wang constructed multiple sub-networks for papers, authors, journals, and textual features, using mutual reinforcement relationships among sub-networks for citation prediction ranking [48]; Zeng Wei built prediction models using paper-paper and author-paper relationship matrices, achieving high efficiency [49]. Structured networks incorporate known information into academic networks, creating "interactions" between nodes and enhancing understanding of citation behavior through network structure and topology.

5. Existing Problems and Future Prospects

5.1 Existing Problems

Overall, academic paper citation prediction has attracted substantial research attention, with new methods enriching and expanding the field. However, numerous unresolved issues remain:

(1) **Numerous influencing indicators without unified selection criteria.** Research shows that feature selection is crucial for effective prediction [35], but current indicators are numerous and determining primary factors remains complex. Despite extensive research, no consensus exists, with some contradictory conclusions. This partly stems from studies focusing on single factors (or treating multiple factors as independent) without adequately considering interactions between factors. Additionally, different fields have varying feature selection standards and effects, and cross-domain studies have yielded inconsistent conclusions.

(2) **Weak theoretical foundations for prediction methods.** Most existing methods are parametric, requiring accurate parameter estimation for correct predictions. However, citation dynamics' complex patterns are difficult to describe with simple parametric models. Parameter settings often require manual adjustment to determine optimal values, introducing subjective factors and lacking deep theoretical research. Identified multi-dimensional indicators lack scientific justification, having only correlational rather than causal relationships with future citations, making it impossible to explain confounding factors or prediction errors.

(3) **Insufficient research on citation prediction dynamics.** Citation prediction is a dynamic process where factor effects may change over time.

Whether models from previous years remain applicable is unclear, and whether different prediction windows require different factors needs further investigation. Most current graph-based methods are static, while networks evolve dynamically with continuously updating nodes and links. Research on temporal topology information remains inadequate, affecting prediction accuracy.

(4) **Limited generalizability of prediction models.** Existing studies cannot apply to all citation patterns. Although D. Wang proposed a model following common temporal patterns, it cannot handle papers with secondary citation peaks years later [26]. Current studies select prediction timeframes arbitrarily without considering disciplinary literature half-life effects [3]. While some argue that five-year post-publication impact reflects paper quality and ten-year prediction is unnecessary, fields like mathematics and economics may require longer prediction periods. Model effectiveness across different time cycles remains unknown.

5.2 Future Prospects

(1) **Strengthen theoretical research on citation prediction.** Future work should focus on theoretical exploration of citation drivers to better explain prediction results. Citation behavior is a dynamic process with complex temporal heterogeneity. Model construction can draw on complex networks, system dynamics, and evolutionary theory to deepen understanding of citation network topology and evolution, grasp dynamic trends and characteristics, and explore citation dynamic patterns, thereby solidifying theoretical foundations.

(2) **Integrate traditional bibliometrics with altmetrics.** Networked academic communication has created new venues for information dissemination and discussion. Altmetrics overcome traditional bibliometrics' limitations with real-time capabilities but face issues with data sources and indicator reliability. Future research should fully combine traditional and altmetric indicators, leveraging altmetrics' complementary role to explore their interaction effects on future citations and conduct deeper correlation analysis to construct more effective feature spaces.

(3) **Deepen natural language processing applications.** Combining scientometric theory with natural language processing technology can leverage NLP's powerful semantic association and mining capabilities to explore deeper indicators, avoiding "manipulable" factors and enabling content-based analysis. Mining more content features from full-text academic literature and incorporating citation content analysis can reveal citation motivations and objectively assess literature value to improve prediction accuracy.

(4) **Establish unified baseline standards.** Unified baseline standards facilitate continued scientific discovery. Current diverse datasets and in-

dicators limit feature robustness and interpretability. Without unified standards, designed features lack generalizability. Building a unified academic data benchmark requires researchers to share data and indicators, representing an important issue that citation prediction research urgently needs to address.

(5) **Construct more accurate prediction models.** Future research should strengthen investigation of different citation patterns, such as "Sleeping Beauty" papers [50], to explore unified prediction frameworks. Fine-grained studies of citation behavior—such as self-citations, negative citations, or citations from scholars with different influence levels—can further examine author social relationships and various citation distribution effects. Incorporating more precise temporal information to build dynamic sequence prediction models can fully describe observed phenomena, reveal underlying mechanisms, and establish accurate causal relationships.

In summary, in the intelligent, digital, and networked environment, academic paper citation prediction research continues to evolve, generating new influencing factors and prediction methods. This paper has systematically reviewed recent advances, but many open questions remain. Future work should deepen theoretical research, promote theoretical innovation, strengthen rational application of new indicators and methods, facilitate dataset sharing and open data, and integrate prediction methods into intelligent academic search platforms to meet users' diversified and personalized needs, enabling scientific application of citation prediction methods.

References

[1] Geng Qian, Jing Ran, Jin Jian, et al. Analysis of academic paper citation prediction and influencing factors [J]. Library and Information Service, 2018, 62(14): 29-40.

[2] Yang L, Zhang Z, Cai X, et al. Citation recommendation as edge prediction in heterogeneous bibliographic network: a network representation approach [J]. IEEE Access, 2019, 7: 23232-23239.

[3] Bao Yufang, Ma Jianxia. Current status analysis and research on scientific paper citation frequency prediction [J]. Journal of Intelligence, 2015, 34(5): 66-71.

[4] Stewart JA. Achievement and asccriptive processes in the recognition of scientific articles [J]. Social forces, 1983, 62(1): 166-189.

[5] Willis DL, Bahler CD, Neuberger MM, et al. Predictors of citations in the urological literature [J]. BJU international, 2011, 107(12): 1876-1880.

[6] Kosteas VD. Predicting long-run citation counts for articles in top economics journals [J]. Scientometrics, 2018, 115(3): 1395-1412.

[7] Chakraborty T, Kumar S, Goyal P, et al. Towards a stratified learning approach to predict future citation counts [C]//2014 IEEE/ACM joint conference on digital libraries (JCDL). London: IEEE computer society, 2014: 351-360.

[8] Antoniou GA, Antoniou SA, Georgakarakos EI, et al. Bibliometric analysis of factors predicting increased citations in the vascular and endovascular literature [J]. Annals of vascular surgery, 2015, 29(2): 286-292.

[9] Fu LD, Aliferis CF. Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature [J]. Scientometrics, 2010, 85(1): 257-270.

[10] Yu T, Yu G, Li PY, et al. Citation impact prediction for scientific papers using stepwise regression analysis [J]. Scientometrics, 2014, 101(2): 1233-1252.

[11] Haslam N, Ban L, Kaufmann L, et al. What makes an article influential? Predicting impact in social and personality psychology [J]. Scientometrics, 2008, 76(1): 169-185.

[12] Roth C, Wu J, Lozano S. Assessing impact and quality from local dynamics of citation networks [J]. Journal of informetrics, 2013, 6(1): 111-120.

[13] Didegah F, Thelwall M. Which factors help authors produce the highest impact research? Collaboration, journal and document properties [J]. Journal of informetrics, 2013, 7(4): 861-873.

[14] Subotic S, Mukherjee B. Short and amusing: the relationship between title characteristics, downloads, and citations in psychology articles [J]. Journal of information science, 2014, 40(1): 115-124.

[15] Sohrabi B, Iraj H. The effect of keyword repetition in abstract and keyword frequency per journal in predicting citation counts [J]. Scientometrics, 2017, 110(1): 243-251.

[16] Yan R, Tang J, Liu X, et al. Citation count prediction: learning to estimate future citations for literature [C]//ACM international conference on information & knowledge management. Glasgow: ACM, 2011: 1247-1252.

[17] Tahamtan I, Afshar AS, Ahmadzadeh K. Factors affecting number of citations: a comprehensive review of the literature [J]. Scientometrics, 2016, 107(3): 1195-1225.

[18] Bornmann L, Leydesdorff L, Wang J. How to improve the prediction based on citation impact percentiles for years shortly after the publication date? [J]. Journal of informetrics, 2014, 8(1): 175-180.

[19] Zhang Meiping, Shang Mingsheng. Important paper prediction based on continuous attention decay [J]. Complex Systems and Complexity Science, 2015, 12(3): 77-84.

[20] Bornmann L, Daniel HD. Citation speed as a measure to predict the attention an article receives: An investigation of the validity of editorial decisions

at Angewandte Chemie International Edition [J]. Journal of informetrics, 2010, 4(1): 83-88.

[21] Xiong Zequan, Duan Yufeng. Can early download counts predict later citation counts?—Taking library and information science journals as an example [J]. Library and Information Knowledge, 2018(4): 32-40.

[22] Shema H, Bar-Ilan J, Thelwall M. Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics [J]. Journal of the Association for Information Science and Technology, 2014, 63(3): 431-449.

[23] Peoples BK, Midway SR, Sackett D, et al. Twitter predicts citation rates of ecological research [J]. PloS One, 2016, 11(11): e0166570.

[24] Zoller D, Doerfel S, Jäschke R, et al. Posted, visited, exported: Altmetrics in the social tagging system BibSonomy [J]. Journal of informetrics, 2016, 10(3): 732-749.

[25] Thelwall M, Nevill T. Could scientists use Altmetric.com scores to predict longer term citation counts? [J]. Journal of informetrics, 2018, 12(1): 237-248.

[26] Wang D, Song C, Barabasi AL. Quantifying long-term scientific impact [J]. Science, 2013, 342(6154): 127-132.

[27] Onodera N, Yoshikane F. Factors affecting citation rates of research articles [J]. Journal of the Association for Information Science and Technology, 2015, 66(4): 739-764.

[28] Yu Houqiang, Qiu Junping. On the application of altmetrics in library literature services [J]. Journal of Intelligence, 2014(9): 163-166.

[30] Abramo G, D'Angelo, Felici G. Predicting publication long-term impact through a combination of early citations and journal impact factor [J]. Journal of informetrics, 2019, 13(1): 32-49.

[31] Stegehuis C, Litvak N, Waltman L. Predicting the long-term citation impact of recent publications [J]. Journal of informetrics, 2015, 9(3): 642-657.

[32] Newman MEJ. The first-mover advantage in scientific publication [J]. Europhysics letters, 2009, 86(6): 68001.

[33] Newman MEJ. Prediction of highly cited papers [J]. Europhysics letters, 2014, 105(2): 28002.

[34] Wang M, Wang Z, Chen G. Which can better predict the future success of articles? Bibliometric indices or altmetrics [J]. Scientometrics, 2019, 119(3): 1575-1595.

[35] Bhat HS, Huang LH, Rodriguez S, et al. Citation prediction using diverse features [C]//IEEE international conference on data mining workshop, USA: IEEE, 2015: 589-596.

[36] Cao X, Chen Y, Ray Liu KJ. A data analytic approach to quantifying scientific impact [J]. Journal of informetrics, 2016, 10(2): 471-484.

[37] Abrishami A, Aliakbary S. Predicting citation counts based on deep neural network learning techniques [J]. Journal of informetrics, 2019, 13(2): 485-499.

[38] Wu Zhiyong. Research on academic paper ranking prediction algorithms [D]. Inner Mongolia: Inner Mongolia University, 2015.

[39] Pobiedina N, Ichise R. Citation count prediction as a link prediction problem [J]. Applied intelligence, 2016, 44(2): 252-268.

[40] Chen C. Predictive effects of structural variation on citation counts [J]. Journal of the Association for Information Science and Technology, 2012, 63(3): 431-449.

[41] Yu Zhitao, Mou Xiaoqing. Review of Chen's predictive indicators and their applications in bibliometrics [J]. Library Tribune, 2013, 33(4): 32-41.

[42] Bai Xiaomei. Academic influence evaluation and prediction based on social network analysis [D]. Dalian: Dalian University of Technology, 2017.

[43] Sayyadi H, Getoor L. FutureRank: ranking scientific articles by predicting their future PageRank [C]//Proceedings of the SIAM international conference on data mining, USA: Society for industrial and applied mathematics, 2009: 533-544.

[44] Liu Dayou, Xue Ruiqing, Qi Hong. Paper value prediction algorithm based on author authority [J]. Acta Automatica Sinica, 2012, 38(10): 1654-1662.

[45] Fan Wei, Han Jianing, Zhang Yuxiang. Paper influence prediction algorithm based on network representation learning [J/OL]. Computer Engineering. [2019-06-15]. https://doi.org/10.19678/j.issn.1000-3428.0053395.

[46] Chinese Association for Artificial Intelligence. Machine Learning White Paper. [EB/OL]. [2019-05-20]. http://www.caai.cn/index.php?s=/home/article/detail/id/49.html.

[47] Shen Lei. New paper impact prediction based on academic networks [D]. Jinan: Shandong University, 2018.

[48] Wang S, Xie S, Zhang X, et al. Coranking the future influence of multiobjects in bibliographic network through mutual reinforcement [J]. ACM transactions on intelligent systems and technology, 2016, 7(4): 1-28.

[49] Zeng Wei. Research on literature ranking prediction algorithms and author influence evaluation algorithms [D]. Chongqing: Southwest University, 2014.

[50] Du Jian, Wu Yishan. Important characteristics, predictive clues, and policy implications of "Sleeping Beauty" literature [J]. Studies in Science of Science, 2018, 36(11): 1938-1945.

Author Contributions

Xia Wanjun: Conceptualized the framework, collected materials, and wrote the paper; Chen Xiaohong: Guided the framework and revised the paper; Jiang Yanping: Revised the paper.

Note: Figure translations are in progress. See original paper for figures.