Research on Predatory Journal Identification Incorporating Altmetrics Indicators

Authors: Wang Linzi, Boxin Zhang, Chen Ming, Chen Ming

Date: 2024-01-26T00:00:00+00:00

Abstract

Objective To investigate the effectiveness of applying Altmetrics indicators in identifying predatory journals and to provide a new perspective. Methods Based on the Logit regression model, three predatory journal discriminant models were constructed: one containing only average citation count, one containing only Altmetrics presence rate indicator, and one integrating both indicators simultaneously. The fitting effects of the three models were compared using ROC curves. Finally, journal data from invitation-to-submit emails were used to validate the model effectiveness. Results The predatory journal discriminant model integrating both average citation count and Altmetrics presence rate achieved the optimal performance, with both indicators showing significant negative correlation with journal predatory nature. Validation using data from 14 journals solicited through invitation emails revealed that over 85% were identified as predatory journals, demonstrating the rationality of the model. Conclusion Altmetrics indicators demonstrate good effectiveness in identifying predatory journals, providing a new and beneficial supplement for predatory journal identification.

Full Text

A Study on Predatory Journal Identification Using Altmetrics Indicators

Wang Linzi[1], **Zhang Boxin**[1], **Chen Ming**[1]
[1] School of Information Management, Nanjing University, No. 163 Xianlin Avenue, Qixia District, Nanjing, Jiangsu Province, 210046

Abstract:
[Purpose] To explore the effectiveness of applying Altmetrics indicators in identifying predatory journals and provide a new perspective for predatory journal identification. [Methods] Based on the Logit regression model, three predatory

journal identification models were constructed using only the average citation count, only the Altmetrics presence rate, and a combination of both indicators. The ROC curve was used to compare the fitting effects of the three models. Finally, journal data from invitation emails were used to validate the model effects. [Findings] The model integrating both average citations and Altmetrics presence rate performed best, with both indicators showing significant negative correlation with journal predatoriness. Validation using data from 14 journals that sent solicitation emails found that over 85% were identified as predatory journals, demonstrating the model' s rationality. [Conclusions] The application of Altmetrics indicators demonstrates good effectiveness in identifying predatory journals and provides a valuable new supplement for predatory journal identification.

Keywords: Predatory journals; Altmetrics; Citation frequency; Logit model; Identification studies

Open Access (OA) is an academic journal publishing model proposed to promote the sharing of academic information resources and facilitate academic communication [1]. As the main product of open access, OA journals provide free access to published articles for the public, promoting more efficient and convenient dissemination of academic achievements. OA journal publishers need to charge authors article processing fees to maintain operations and profitability. However, some unscrupulous publishers exploit this model by publishing paid literature in large volumes without proper review to maximize profit. Jeffrey Beall, a librarian at the University of Colorado, termed journals that exploit the open access model through falsification and false promises, irresponsibly charging authors article processing fees for profit, as "predatory journals" [2]. Predatory journals have since come to scholars' attention and attracted widespread concern and discussion.

Predatory journals exploit the characteristics of open access to conceal their deceptive nature and seek private gain, causing certain interference and damage to the open access movement [3]. In recent years, predatory journals have shown a rapid growth trend, seriously harming the academic ecosystem. Data published by Nature in March 2022 indicated that the number of predatory journals has exceeded 15,500 and is growing rapidly [4]. However, current academic research on identifying predatory journals remains relatively weak. Existing methods for identifying predatory journals mainly rely on list-based approaches, such as Beall' s list of "potential, possible, or probable predatory scholarly open-access publishers" and the Kscien' s list established by the Kscien Predatory List Committee (PLC) [5]. However, such list-based approaches have obvious limitations: first, they are difficult to generate and update, making it hard to include newly emerged predatory journals in a timely manner; second, they cannot cover all predatory journals, resulting in a certain false-negative rate. Therefore, as the number of articles published in predatory journals increases year by year, how to identify and judge predatory journals has become an issue requiring urgent attention. In recent years, Altmetrics, as a new indicator for

measuring academic impact, has attracted widespread attention and provided a new perspective for journal evaluation. This study intends to use Altmetrics indicators to judge journal predatoriness, aiming to establish a more effective identification method.

1.1 Characteristics and Identification Methods of Predatory Journals

Given the significant impact and harm caused by predatory journals globally, scholars have studied the characteristics that differentiate predatory journals from other journals across various dimensions. The 2022 report "Combatting Predatory Academic Journals and Conferences" published by the InterAcademy Partnership (IAP) [6] used mapping methods to specifically differentiate characteristics among various types of journals, with typical features of predatory journals including non-existent or incorrect peer review, imitation of other journals or websites, absence of editorial boards or fake editorial boards, and alternative or false impact factors. Shamseer et al. [7] pointed out that English-language predatory scientific journals have 13 characteristics, including spelling and grammatical errors on websites, promises of rapid publication, and absence of retraction policies. Frandsen [8] found that authors who publish more frequently in predatory journals have higher average publication counts (and medians) across Scopus, showing a certain positive correlation. Based on multi-dimensional research, predatory journals have been found to have certain loopholes and problems in various aspects such as fees, promotion, review processes, and policies, which have become relatively obvious characteristics of predatory journals.

As characteristics of predatory journals continue to be discovered and summarized, researchers have proposed many perspectives on identification indicators and methods for predatory journals in recent years. Regarding the high publication fees of predatory journals, Xia [9] studied the charging practices of predatory journals and found that article processing charges (APC) cannot be used as the sole criterion for distinguishing whether a journal is predatory. Wang Lingfeng et al. [10] proposed the Journal Price Index (JPI) to evaluate the reasonable level of academic journal publication fees, providing a simple and effective objective method for precisely defining predatory journals from the perspective of page charges. Ruiter-Lopez L et al. [11] examined the editorial boards of predatory journals using quantitative methods and found that most members were high-level scholars with a median publication count of 43, citation count of 664, and h-index of 14, indicating that examining editorial board membership is difficult for identifying journal predatoriness. Regarding journal websites and terminology, Chen L-X et al. [12] used machine learning methods to extract website text content, keywords, and other features from mainstream predatory journal websites and regular journal websites, proposing a predatory journal classification system based on a new model. Chen et al. [13] also found that measuring the differences in specific word frequencies between journals through differential scoring can improve the classification efficiency of

bag-of-words models and TF-IDF algorithms, helping to identify characteristic words of predatory journals. However, journal websites and vocabulary usage can also be adjusted and modified, and the boundary between predatory journals and low-quality journals remains relatively ambiguous. Yeo-The et al. [14] believe that the most important criterion for distinguishing predatory journals from legitimate academic journals is the rigor of peer review, and the motivation or intent of submitting authors is also crucial. Kong Yehan et al. [15] conducted a comparative analysis from six dimensions including academic publishing legitimacy, commercial fraud, and academic misconduct, establishing 18 secondary indicators to classify identified "warning journals." Analyzing these motivations and related behavioral factors yields deeper conclusions, but considering that some subjective factors are difficult to collect and judge, this may affect the accuracy of identification results. Therefore, many scholars also study and analyze predatory journals from the perspective of citation patterns. Frandsen [16] tracked the citation patterns of 124 predatory journals in Scopus from 2013-2016, finding that these journals were cited 1,295 times, approximately 10.5 times per journal, and concluded that citations to predatory journals in non-predatory literature are limited. Bo-Christer Björk et al. [17] randomly selected 250 articles published in predatory journals from Google Scholar, studied their citation data over five years, and found that each article was cited an average of 2.6 times, with 56% of articles receiving no citations at all.

1.2 The Value and Application of Altmetrics

Although traditional citation indicators are often used to measure the impact of academic achievements, they have problems such as time lag, negative citations, and self-citations, and cannot comprehensively reflect the impact of academic achievements [18]. With the continuous development of social online media platforms, scientific communication is becoming increasingly networked, and academic activities are gradually becoming more open. More and more researchers are using online academic network platforms and social media platforms to acquire, disseminate academic resources, and conduct academic exchanges. In 2010, J. Priem first proposed the concept of Altmetrics on Twitter [19] to evaluate the impact of academic papers on social networks, measuring their social impact by tracking the dissemination and communication of academic achievements such as scholarly papers on the internet. As an emerging research impact indicator, Altmetrics has attracted the attention and research of numerous scholars since its proposal. Existing literature shows that applied research on Altmetrics mostly focuses on the comprehensive evaluation of paper or journal impact. For example, regarding papers, Wang Yanbo et al. [20] proposed combining Altmetrics indicators, which measure social impact, with traditional citation indicators, which measure academic impact, to construct a more comprehensive and holistic indicator system for evaluating academic paper impact. At the journal level, Wang Kaili et al. [21] integrated citation analysis and Altmetrics methods to construct a journal impact evaluation system and applied it to the impact analysis of international library and information

science journals. Yu Zhenglu et al. [22] conducted statistical analysis on the social impact of Chinese English-language scientific journals based on Altmetrics mention counts. The above studies indicate that the effectiveness of Altmetrics indicators in the comprehensive evaluation of paper or journal impact has been preliminarily verified.

1.3 Using Altmetrics for Predatory Journal Identification

Although research using Altmetrics indicators to identify predatory journals is limited, preliminary studies suggest its potential application. The authors previously analyzed the performance of predatory and non-predatory journals in library and information science in terms of Altmetrics, finding that predatory journals had much lower Altmetrics presence rates than non-predatory journals in this field [23]. However, due to limitations in predatory journal data volume, a clear conclusion that Altmetrics can effectively identify predatory journals could not be drawn, and its effectiveness remains to be verified through systematic large-sample studies. Based on the above, this study builds upon existing research by selecting the biomedical field, where predatory journals are widely distributed, introducing Altmetrics indicators on the basis of traditional citation-based identification of predatory journals, and constructing a discriminant model for predatory journals using the Logit regression model to provide new indicators and methods for predatory journal identification.

2.1 Research Hypotheses

Based on the previous literature review and analysis of journal predatoriness and Altmetrics and citation indicators, this study constructs the following two hypotheses and intends to use the Logit regression model, with journal citation indicators and Altmetrics indicator values as independent variables and journal predatoriness as the dependent variable, to verify these hypotheses.

Hypothesis 1: There is a negative correlation between citation indicators and journal predatoriness; that is, when the average citation indicator is higher, the probability of a journal being identified as predatory is smaller.

Hypothesis 2: There is also a negative correlation between Altmetrics indicators and journal predatoriness; that is, when Altmetrics indicators are higher, the probability of a journal being identified as predatory is smaller.

2.2 Research Methods

The Logit model, also known as the "rating model" or "classification rating model," is a type of discrete choice model used to predict the probability of event occurrence, mainly divided into two categories: binary Logit regression model and multinomial Logit regression model. The dependent variable in binary Logit regression models is dichotomous, typically defining event occurrence as "1" and non-occurrence as "0". Multinomial Logit regression models are suitable for

cases where the dependent variable has multiple categories. The Logit model is widely applied in fields such as sociology, biostatistics, and econometrics. In the library and information science field, Ren Haizhi et al. [24] used a multinomial Logit regression model to conduct empirical analysis on influencing factors of communication effectiveness of official WeChat accounts of book publishing enterprises. This study mainly explores the identification of predatory journals, and the binary Logit model can well characterize the two types of "predatory journals" and "non-predatory journals." Since this paper hypothesizes that citation indicators and Altmetrics indicators have negative correlations, the binary Logit model is selected for journal type identification analysis.

where y' s range is (-∞, +∞), while the probability of event occurrence ranges from [0,1]. Therefore, it is necessary to perform a Logit transformation on the linear regression model. First, introduce the Odds ratio, where Odds represents the ratio of the probability of event occurrence to the probability of event non-occurrence. The calculation method for Odds is shown in Equation 1, where Odds' range is [0, +∞).

Taking the natural logarithm of the Odds ratio yields Logit, which maps P from [0,1] to (-∞, +∞), enabling multiple linear regression modeling. The process from probability P to Odds to Logit is called the Logit transformation. The Logit model expression is:

$$y = \text{logit}(P) = \ln(\text{Odds}) = \ln\left(\frac{p}{1-p}\right) = \alpha_0 + \alpha_{jx} j + \epsilon$$

where p represents the probability of a certain value of the dependent variable, x_j are independent variables, α_j are coefficients of independent variables, α_0 is the intercept term, and ϵ represents the error term.

2.3.1 Dependent Variable Definition and Data Sources

This study uses a dichotomous discrete variable of journal type as the dependent variable, defining predatory journals = 1 and non-predatory journals = 0. Shen and Bjork [25] found in their research that the number of articles in predatory journals has increased rapidly year by year, rising from over 53,000 articles in 2010 to over 420,000 articles in 2014, with approximately 8,000 active predatory journals, and the vast majority of these predatory journal articles come from fields such as biomedicine. Therefore, this study randomly selected 100 predatory journals and 100 non-predatory journals from the biomedical field as research samples. Predatory journal data were sourced from Kscien' s list, while non-predatory journal data were sourced from the DOAJ directory. DOAJ is a directory website of peer-reviewed open access (OA) journals created and maintained by Lund University Library in Sweden, with strict journal inclusion standards and evaluation processes [26]. Since both Kscien' s list and DOAJ consist of open access journals, the two types of journal samples selected in this paper are comparable.

2.3.2 Independent Variable Definition and Data Sources

Citation counts reflect the academic impact of articles or journals to a certain extent, serving as an important indicator for evaluating paper quality and value, and representing the core indicator for calculating journal impact factors. To eliminate the influence of differences in journal article numbers on citation counts, this study uses the average number of citations per journal to measure citation performance, represented by variable x_1. The number of articles published by journals from 2012-2022 was collected by crawling journal websites, and journal citation counts during this period were queried using Web of Science. The specific calculation formula is shown in Equation 3:

$$\text{Average citations per journal} = \frac{\text{Total citations of journal articles}}{\text{Number of articles published by journal}}$$

Altmetrics, as a new indicator for measuring the social impact of academic achievements, expands traditional citation-based impact assessment and can more comprehensively reflect the impact of academic achievements in the digital network. Since the concept of Altmetrics was proposed, various measurement tools have emerged, such as Altmetric.com, PlumX, and Crossref Event Data (CED). Among them, Altmetric.com has extensive information sources, covering over 5,000 mainstream media outlets and 15,000 academic and non-academic blogs worldwide [27], and provides free access to researchers. Therefore, most Altmetrics-related studies in recent years have adopted the Altmetric.com tool [28]. This study uses the Altmetric Explorer tool from Altmetric.com, with journal ISSN as the search criterion, to obtain Altmetrics data for journals from 2012-2022. To eliminate the influence of journal size differences, this study defines the Altmetrics presence rate indicator, represented by independent variable x_2, with the specific calculation formula shown in Equation 4:

$$\text{Altmetrics presence rate} = \frac{\text{Number of articles with Altmetrics scores}}{\text{Number of articles published by journal}}$$

3 Empirical Analysis

3.1 Descriptive Statistics

From the descriptive statistics of Altmetrics presence rate and average citation count for the collected 200 predatory and non-predatory journals (Table 1), it can be observed that the average Altmetrics presence rate of these predatory journals is only 0.0088. Among these 100 journals, 81 have an Altmetrics presence rate of 0, meaning 81% of predatory journal articles have no Altmetrics score, indicating that predatory journal articles receive little attention and discussion on social media. The mean average citation count for these predatory journals is 0.7210, meaning each article is cited approximately 0.7210 times on

average. Compared with predatory journals, non-predatory journals have an average Altmetrics presence rate of 0.338, far greater than the 0.0088 of predatory journals. This indicates that non-predatory journals have far greater influence on social networks than predatory journals. For non-predatory journals, only 23 journals have an Altmetrics presence rate of 0, and the highest Altmetrics presence rate among journals can reach 0.9916, indicating that non-predatory journals have greater attention and influence on social networks compared to predatory journals. In terms of journal citations, the mean average citation count for non-predatory journals is 5.1763, higher than that of predatory journals. This indirectly reflects that the quality of articles in predatory journals is difficult to guarantee, may not provide constructive viewpoints, and are less frequently cited, while articles in non-predatory journals may be of higher quality and provide more substantial help to researchers' work, thus being cited more frequently.

Table 1 Descriptive statistical analysis of predatory and non-predatory journal data: Altmetrics presence rate and average citation count for predatory journals and non-predatory journals

3.2 Logit Model Results

From the above analysis, there are significant differences in average citation count and Altmetrics presence rate between predatory and non-predatory journals. Can we use them as indicators to determine journal predatoriness? The authors used Stata 14.0 software to establish and compare the following three Logit regression models. First, a predatory journal identification model based on citation indicators was constructed (Model 1). Then the effectiveness of a discriminant model built using Altmetrics presence rate was explored (Model 2). Finally, a predatory journal identification model combining average citation count and Altmetrics presence rate was constructed (Model 3), and a comparative analysis of these three models and their effectiveness was conducted. The specific models and effectiveness analysis are as follows.

3.2.1 Model 1: Citation-based Model Using Stata 14.0 software to conduct Logit model regression analysis on journals' average citation counts, the results are shown in Table 2 . The expression for the Logit model fitting results is:

$$y = \ln\left(\frac{p}{1-p}\right) = 0.864 - 0.438 \times \text{average\_citation\_count}$$

The coefficient of the average citation count variable is -0.438 $<$ 0 (sig $<$ 0.05), which indicates that average citation count is significantly negatively correlated with whether a journal is predatory, confirming that Hypothesis 1 holds. That is, when the average citation count is lower, the probability of a journal being predatory is higher.

It can be seen that,

Table 2 Logit model regression results for citation indicators: Average citation count

After completing the calculation of model parameters, it is necessary to evaluate whether the expected probabilities calculated by the model and the actual probabilities can effectively fit. If actual observations have high consistency with model predictions, the model is considered to fit the data; otherwise, the model cannot be accepted and variables need to be reset. In this study, the authors use the Hosmer-Lemeshow test to examine the goodness-of-fit of the binary Logit regression model. The Hosmer-Lemeshow test indicates the degree of agreement between fitted values and observed values. If Sig < 0.05, it indicates significant differences between model predictions and observations, suggesting poor model performance; conversely, if Sig > 0.05, it is considered that the model estimation fits the data at an acceptable level, indicating good model performance. Conducting the Hosmer-Lemeshow test on the citation indicator discriminant model yields a Sig value of $0.000 < 0.05$. This result indicates that the Logit discriminant model based solely on average citation count has poor fit. Therefore, we attempt to introduce Altmetrics indicators for discrimination.

3.2.2 Model 2: Altmetrics-based Model First, a Logit discriminant model was established using only the Altmetrics presence rate of journals, obtaining the Logit model fitting table. The modeling results are shown in Table 3 . It can be seen that the coefficient of Altmetrics presence rate is -14.871 < 0 (sig < 0.05), indicating that the Altmetrics presence rate of journals is significantly negatively correlated with whether a journal is predatory, confirming that Hypothesis 2 holds. That is, the higher the Altmetrics presence rate of a journal, the smaller the probability of the journal being predatory. The Altmetrics presence rate can reflect the social impact of a journal' s academic achievements, so this result is also reasonable. This part continues to use the Hosmer-Lemeshow test to examine model fit, obtaining a Sig value of $0.997 > 0.05$ for the discriminant model built with Altmetrics presence rate, indicating that the model has good fit.

The Logit model expression is:

$$y = \ln\left(\frac{p}{1-p}\right) = 1.036 - 14.871 \times \text{Altmetrics\_presence\_rate}$$

Table 3 Logit model regression results for Altmetrics indicators: Altmetrics presence rate

3.2.3 Model 3: Combined Model In addition to constructing a Logit discriminant model using Altmetrics presence rate alone, the authors also established a Logit regression model simultaneously incorporating both average

citation count and Altmetrics presence rate variables. The expression for the Logit model fitting results is:

$$y = \ln\left(\frac{p}{1-p}\right) = 1.220 - 0.204 \times \text{average\_citation\_count} - 12.015 \times \text{Altmetrics\_presence\_rate}$$

with regression results shown in Table 4 . Among them, the significance levels of the average citation count and Altmetrics presence rate indicators are 0.033 and 0.002 respectively, both less than 0.05. This indicates that the fitting effect is significant, and both average citation count and Altmetrics presence rate have negative effects on whether a journal is predatory, confirming Hypotheses 1 and 2. That is, when both average citation count and Altmetrics presence rate are lower, the probability of a journal being predatory is greater. This is consistent with the results obtained from Model 1 and Model 2.

Table 4 Logit model regression results after introducing Altmetrics: Average citation count and Altmetrics presence rate

Using the Hosmer-Lemeshow test to examine the goodness-of-fit of the model, the result is Sig = 0.3568 > 0.05. In the Hosmer-Lemeshow test, when the Sig value is greater than 0.05, the fit is considered good, and when the Sig value is greater than 0.1, the fit is even better, indicating that the discriminant model after introducing Altmetrics presence rate has good fit.

3.2.4 Comparison of Three Models The area under the ROC curve is used to test model prediction accuracy, and it is believed that when the area under the curve is greater than 0.75, the model has sufficient discriminatory power. Based on the curve' s position, the ROC curve divides the entire graph into two parts. The area of the portion under the curve is called AUC (Area Under Curve), used to represent prediction accuracy. The higher the AUC value, that is, the larger the area under the curve, the higher the prediction accuracy. The closer the curve is to the upper left corner (the smaller the X, the larger the Y), the higher the prediction accuracy. This paper uses ROC analysis to conduct comparative analysis of the three models.

Table 5 Model ROC comparison: AUC area and asymptotic 95% confidence interval

From the model ROC results (Table 5), it can be seen that Model 3' s prediction performance is superior to Model 1 and Model 2, meaning Model 3 has higher prediction accuracy. This means that in constructing discriminant models, the model simultaneously integrating average citation count and Altmetrics presence rate performs better than models using only average citation count or Altmetrics presence rate alone. The discriminant model constructed using only Altmetrics presence rate performs better than the model using only

average citation count, thereby proving the rationality and correctness of introducing Altmetrics indicators for predatory journal discrimination proposed in this paper.

4 Model Validation

To validate the model' s effectiveness, the authors collected journal data from invitation emails received within the past year. Sureda-Negre et al. [29] analyzed journals from emails inviting three professors in the education field at a Spanish university to publish articles within three months, finding that most journals (69.7%) were on predatory journal lists, and concluded that journals sending submission invitations to scholars via email are mostly of low quality. Therefore, we believe it is reasonable to select invitation journals from email to validate the discriminant model' s effectiveness. After excluding journals that appeared in Kscien' s list, the obtained journals are shown in Table 6 .

Table 6 Journal information from invitation emails

Review of Contemporary Business Research
American Journal of Information Science and Technology (AJIST)
Journal of Intercultural Communication (JICC)
Journal of Management Information System and E-commerce
International Journal of Business & Economic Development
International Journal of Business and Applied Social Science
International Journal of Library and Information Studies
Journal of Economics & Management Research
International Journal of Sustainability Management and Information Technologies
Health Informatics Journal
International Journal of Business and Social Science Research
Journal of Business and Social Science Review (JBSSR)
International Journal of Business & Management Studies (IJBMS)

Using the same data collection method, citation data and Altmetrics data for these journals were collected and input into Model 3 for validation. The final results are shown in Table 7 . It can be seen that among these 14 validation journals, only the Health Informatics Journal had a predicted probability of being predatory of $0.003 < 0.5$. Upon investigation, the authors found that Health Informatics Journal is an SCI-indexed journal in JCR Q3 category, which matches our prediction of a low probability of being predatory. The remaining 13 journals had predicted probabilities greater than 0.5, with 12 having probabilities greater than 0.7. This indicates that although these journals have not yet appeared in Kscien' s predatory journal list, they have a high probability of being predatory, which aligns with Sureda-Negre et al.' s [29] finding that journals sending submission invitations to scholars via email are mostly of low quality. This also demonstrates that the predatory journal discriminant model integrating average citation count and Altmetrics presence rate is reasonable.

Table 7 Model validation results: Altmetrics presence rate, average citation count, and predicted probability

5 Discussion and Conclusion

Based on the Logit regression model, this paper constructed three predatory journal discriminant models: one containing only the journal' s average citation count, one containing only the Altmetrics presence rate indicator, and one simultaneously containing both the journal' s average citation count and Altmetrics presence rate indicators. Through comparative model analysis, it was found that the predatory journal discriminant models containing only the Altmetrics presence rate indicator and simultaneously introducing both average citation count and Altmetrics presence rate have better performance. Although citation indicators are traditionally considered suitable for measuring journal academic impact, Oviedo-García' s analysis of the predatory journal publisher MDPI shows that some predatory journals have high self-citation rates [30], resulting in some predatory journals having high citation counts. This suggests that high citation counts in these journals may be manipulated, and judging whether a journal is predatory based purely on citation indicators is not accurate. In the Web 2.0 environment, Altmetrics indicators fully utilize academic social networks for bibliometric analysis. Because Altmetrics data are updated more timely, they can avoid the lag in evaluating academic achievement impact and can supplementarily reflect the social impact of academic achievements. The combined application of Altmetrics indicators and citation indicators can more comprehensively evaluate the impact of academic achievements. In this study, the discriminant model constructed by simultaneously introducing average citation count and Altmetrics indicators integrates the advantages of both types of indicators, thus having better effectiveness in judging journal predatoriness.

Based on this, the authors collected journal data from submission invitation emails in their mailbox and input them into the predatory journal discriminant model that simultaneously incorporates average citation count and Altmetrics presence rate to validate the model. The results indicate that journals inviting submissions via email have a high probability of being predatory. This demonstrates that Altmetrics indicators have good effectiveness in identifying predatory journals, can better judge journal quality, and provide new indicators and methods for predatory journal identification.

Undeniably, this study still has some limitations. First, due to difficulties in data acquisition, we can only preliminarily verify the effectiveness of using journal citation indicators and Altmetrics indicators to identify predatory journals. Second, the methods applied in the article still need optimization. In future research, we will attempt improvements from two aspects: first, expanding indicator types by adding more indicators that can effectively identify predatory journals to enrich model output; second, increasing sample size and attempting to adopt more advanced methods, such as random forest models in machine learning, to enhance identification effectiveness and model precision. Overall,

this study offers some inspirational significance for predatory journal identification, but further expansion in data and methods is still needed.

References

[1] Li Wu, Liu Ziheng. Analysis of a brand-new academic publishing model: Open access publishing model [J]. Journal of Library Science in China, 2004(6): 67-70.
[2] Beall J. Predatory journals and the breakdown of research cultures[J]. Information Development, 2015, 31(5): 473-476.
[3] Dai Qi, Yuan Xilin. Research on academic reputation risk and early warning of open access journals[J]. Chinese Journal of Scientific and Technical Periodicals, 2018, 29(11): 1063-1071.
[4] Nature: Global predatory journals have exceeded 15,500 types [EB/OL]. [2022-08-16]. https://mp.weixin.qq.com/s/WJ_{4URZ2MqUVhjuzl2Jm}-g.
[5] Kakamad F H, Mohammed S H, Najar K A, et al. Kscien' s list; a new strategy to hoist predatory journals and publishers[J]. International Journal of Surgery Open, 2019, 17: 5-7.
[6] The InterAcademy Partnership. Combatting predatory academic journals and conferences[EB/OL]. [2023-08-21]. https://www.interacademies.org/project/predatory-publishing.
[7] Shamseer L, Moher D, Maduekwe O, et al. Potential predatory and legitimate biomedical journals: Can you tell the difference? A cross-sectional comparison[J]. BMC Medicine, 2017, 15(1): 1-14.
[8] Frandsen T F. Authors publishing repeatedly in predatory journals: An analysis of Scopus articles[J]. Learned Publishing, 2022, 35(4): 598-604.
[9] Xia J. Predatory journals and their article publishing charges[J]. Learned Publishing, 2015, 28(1): 69-74.
[10] Wang Lingfeng, Du Yanling, Tang Biqun. Quantitative definition of domestic "predatory journals" : Journal Price Index[J]. Journal of Shanxi Library, 2018(03): 5-9.
[11] Ruiter-Lopez L, Lopez-Leon S, Forero D A. Predatory journals: Do not judge journals by their Editorial Board Members[J]. Medical Teacher, 2019, 41(6): 691-696.
[12] Chen L X, Wong K S, Liao C H, et al. Predatory Journal Classification Using Machine Learning[C]. 2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII). IEEE, 2020: 193-196.
[13] Chen L X, Su S W, Liao C H, et al. An open automation system for predatory journal detection[J]. Scientific Reports, 2023, 13: 2976.
[14] Yeo-Teh N S L, Tang B L. Wilfully submitting to and publishing in predatory journals - a covert form of research misconduct?[J]. Biochemia Medica, 2021, 31(3): 0-0.
[15] Kong Yehan, Peng Lin, Yuan Junpeng, Yu Yihao. Construction of early warning monitoring system for scientific journals[J]. Acta Editologica, 2022, 34(04): 369-377.
[16] Frandsen T F. Are predatory journals undermining the credibility of

science? A bibliometric analysis of citers[J]. Scientometrics, 2017, 113(3): 1513-1528.
[17] Björk B C, Kanto-Karvonen S, Harviainen J T. How frequently are articles in predatory open access journals cited[J]. Publications, 2020, 8(2): 17.
[18] Qiu Junping, Yu Houqiang. The proposal process and research progress of altmetrics[J]. Library and Information Service, 2013, 57(19): 5-12.
[19] Priem J. I like the term #articlelevelmetrics, but it fails to imply *diversity* of measures. Lately, liking #altmetrics [EB/OL]. [2023-08-20]. https://twitter.com/jasonpriem/status/25844968813.
[20] Wang Yanbo, Wang Lei, Wu Xiaolan. Research on impact evaluation of domestic academic papers based on Altmetrics and traditional bibliometric indicators[J]. Information Studies: Theory & Application, 2020, 43(06): 76-81.
[21] Wang Kaili, Huang Xiao, Wu Jiang. Comprehensive evaluation of journal impact integrating Altmetrics and citation analysis[J]. Information Science, 2021, 39(02): 169-177.
[22] Yu Zhenglu, Ma Zheng, Tian Ruiqiang, et al. International social impact performance of Chinese English scientific journals: Based on Altmetrics mention count indicator[J]. Acta Editologica, 2021, 33(03): 349-354.
[23] Chen M, Wang L. An Altmetrics and citation analysis of selected predatory journals in library and information science field[J]. The Journal of Academic Librarianship, 2022, 48(6): 102618.
[24] Ren Haizhi, Pan Yingjie. Research on influencing factors of communication effectiveness of official WeChat accounts of book publishing enterprises based on Logit model[J]. Information Research, 2022(06): 62-70.
[25] Shen C, Björk B C. 'Predatory' open access: a longitudinal study of article volumes and market characteristics[J]. BMC Medicine, 2015, 13(1): 1-15.
[26] Wang Linhui, Ni Ming. Analysis on strategies for Chinese open access journals applying to join DOAJ database[J]. Publishing & Printing, 2022(05): 74-83.
[27] Altmetric. Attention sources tracked by Altmetric [EB/OL]. [2022-11-24]. https://help.altmetric.com/support/solutions/articles/6000235983-attention-sources-tracked-by-altmetric.
[28] Ortega J L. Altmetrics data providers: A meta-analysis review of the coverage of metrics and publication[J]. Profesional de la Información, 2020, 29(1): e290103.
[29] Sureda-Negre J, Calvo-Sastre A, Comas-Forgas R. Predatory journals and publishers: Characteristics and impact of academic spam to researchers in educational sciences[J]. Learned Publishing, 2022, 35(4): 441-447.
[30] Oviedo-García M A. Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI)[J]. Research Evaluation, 2021, 30(3): 405-419.

This work is supported by the Major Project of Philosophy and Social Science Research in Jiangsu Universities "Research on Early Warning of Reputation Risks of Academic Journals in Open Environment" (Project No.: 2022SJZD093).

Author Contributions: Wang Linzi (ORCID: 0009-0009-9092-5915), Master' s student, linzi.wang@smail.nju.edu.cn: Research design, data collection and organization, paper writing and revision; Zhang Boxin (ORCID: 0009-0006-6749-3358), Master' s student: Data collection and organization, paper writing and revision; Corresponding author: Chen Ming (ORCID: 0000-0001-5061-6821), Ph.D., Associate Professor, Master' s supervisor, chenming@nju.edu.cn: Research design, paper writing and revision.

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