A Deep Citation Network Analysis-Based Decision Optimization Model for Digital Collection Resources

Chuanmei Xie
Library, Guizhou Education University, Guiyang 550018, Guizhou, China
Corresponding author: Chuanmei Xie (e-mail: cmei05338@sina.com)

ABSTRACT The rapid expansion of digital resource repositories and constrained library budgets have created increasing challenges for the efficient allocation of scholarly information resources, particularly in electromagnetic information transmission environments where high-quality scientific knowledge dissemination relies on advanced digital infrastructures. This study proposes an optimization model integrating Graph Neural Networks and multi-feature fusion to provide quantitative decision support for digital collection development. A heterogeneous citation network is first constructed from large-scale scholarly data, and a Graph Attention Network is employed to learn node representations from which three core characteristics—academic influence, interdisciplinarity, and knowledge freshness—are extracted. A game theory-based combined weighting strategy is subsequently introduced to fuse these complementary features into a comprehensive value metric, and a 0–1 integer programming model is formulated to maximize overall resource value under budget constraints. Simulation experiments based on the public Aminer academic graph dataset compare the proposed approach with conventional impact factor- and usage-oriented strategies. Results demonstrate that the recommended resource portfolio achieves superior long-term academic influence and more balanced disciplinary coverage while maintaining computational interpretability. The proposed framework offers a scalable and data-driven solution for intelligent digital collection management and provides methodological support for efficient scientific information propagation and resource optimization within modern electromagnetic-enabled knowledge infrastructures.

INDEX TERMS deep citation network, graph attention network, collection optimization, electromagnetic information transmission, multi-feature fusion

I. INTRODUCTION

DRIVEN by open science and digital publishing, the variety and volume of digital resources held by academic libraries have reached an unprecedented scale, significantly expanding the specialized knowledge base available for research in high-performance materials and sustainable manufacturing. This massive influx of data provides critical support for the digital integration of engineering and advanced fiber science within the academic framework. However, the growth rate of acquisition budgets lags far behind the inflation in the price and quantity of scholarly publications, making "selective acquisition" a core challenge in library collection development. Traditionally, librarians have relied on bibliometric indicators (e.g., journal impact factor), local usage data (e.g., downloads, views), or the qualitative judgment of subject experts to guide decisions [1]. While these methods are operational, their limitations are evident: bibliometric indicators are susceptible to disciplinary differences and manipulation; usage data reflect historical short-term demand and are poor predictors of long-term knowledge value; and expert judgment suffers from subjectivity and difficulties in scaling.

Academic citation networks record the structured trajectory of knowledge production and dissemination, containing rich information for measuring the intrinsic value of literature. In recent years, with breakthroughs in graph representation learning, particularly the development of Graph Neural Networks (GNNs), deep semantic mining and feature learning from citation networks have become possible [2-3]. Researchers have successfully utilized GNNs for citation recommendation, academic influence prediction, and research trend analysis [4]. However, directly translating the results of deep citation analysis into executable collection management decisions remains an underexplored area. Most existing studies stop at the analytical description level or focus on specific resource types (e.g., institutional repositories) [5], lacking a systematic computational model from data perception to decision generation. Therefore, this study attempts to address a key question: Can deep citation network analysis be used to automatically quantify the

potential comprehensive value of massive digital resources, and based on this, can an optimization decision model be constructed to maximize the overall effectiveness of a collection under strict budget constraints? The value of this research lies in constructing a complete closed loop that extracts decision features from complex scholarly data—including specialized research on material science—and translates them into actionable plans through operations research methods. Although the validation experiment uses the general computer science dataset Aminer (due to data accessibility and reproducibility), the proposed methodology is domain-agnostic and can be directly transferred to material science domains once corresponding citation data are available. This integration of knowledge graphs and intelligent decision technology enhances the precision of library resource allocation to better support the evolving information needs of the industry and advanced manufacturing.

The subsequent structure of this paper is as follows: The second part reviews and critiques relevant research on collection evaluation and citation network analysis; the third part details the proposed three-layer decision model, including network representation learning, multi-feature value quantification, and constrained optimization solving; the fourth part designs simulation experiments to validate the model's effectiveness and provides in-depth analysis; the fifth part discusses the model's application prospects, limitations, and future research directions; and finally, the full contributions are summarized.

II. RELATED RESEARCH PROGRESS

Collection resource optimization and citation network analysis are two long-standing and active research branches in library and information science, and their intersection is fostering new methodological innovations.

In terms of collection decision support, early research focused on predicting resource demand through statistical regression and time series analysis. With the rise of data science, machine learning methods have been widely introduced. For example, some studies utilized combined models of Convolutional Neural Networks and Gated Recurrent Units to mine deep patterns in user borrowing sequences to improve book recommendation accuracy[6]. At the decision modeling level, Multi-Criteria Decision Analysis methods have been used to integrate multiple indicators. The Analytic Hierarchy Process (AHP) has been employed to construct evaluation index systems for digital collection resources [7], while game theory-based combined weighting methods have been used to balance subjective preferences and objective data to determine fund allocation weights [8]. These studies have promoted the quantification and refinement of the decision-making process, but their evaluation basis mostly relies on directly observable "post-use" data or simple external indicators, failing to deeply explore the academic network structure and knowledge content features embedded in the resources themselves.

Simultaneously, citation network analysis is undergoing a transition from traditional bibliometrics to a deep learning paradigm. Traditional bibliometrics relies on network topology metrics such as node degree, betweenness centrality, and PageRank [9]. GNNs can model both network structure and unstructured node attributes (e.g., text), demonstrating strong capabilities in tasks like node classification and link prediction [10]. For instance, some research enhanced GNNs by introducing language models to more accurately mine the factors influencing link formation in citation networks [11]. The Graph Attention Network (GAT), as an important variant of GNNs, has shown excellent performance in tasks like relation extraction due to its attention mechanism [12], providing an ideal tool for modeling the non-uniform importance of references in citation networks [13]. Another study focused on the library context, systematically analyzing the citation characteristics of Chinese books and demonstrating the effectiveness of citation data in evaluating the academic influence and knowledge diffusion breadth of books [14]. These advances indicate that deep learning methods can extract deep semantic and relational features from citation networks that are not captured by traditional metrics.

Although both fields are rapidly developing, research closely integrating them is still in its infancy. A few pioneering works, such as evaluating collection support capability based on institutional repository citation analysis, have a relatively narrow focus and have not formed a general optimization decision framework. The research in this paper aims to fill this gap by constructing an end-to-end decision model that uses deep academic features extracted by GNNs as core input and achieves optimal resource selection through mathematical programming methods, thereby promoting the transformation of collection management from experience-driven to data- and model-driven.

III. MODEL CONSTRUCTION

The proposed optimization decision model for digital collection resources is a three-layer architecture that sequentially transforms raw data into final decision plans. The overall framework of the model is shown in Figure 1, and its core workflow includes: deep feature extraction based on Graph Attention Networks, resource value quantification based on a combined weighting method, and integer programming solving under budget constraints.

A. DEEP CITATION FEATURE EXTRACTION BASED ON GAT

This module aims to transform unstructured citation data into feature vectors that can characterize the multidimensional academic value of resources. We chose the Graph Attention Network as the base model, primarily because its attention mechanism can differentiate the importance of neighbor nodes, which is crucial for modeling citation networks where the importance of references is not uniform.

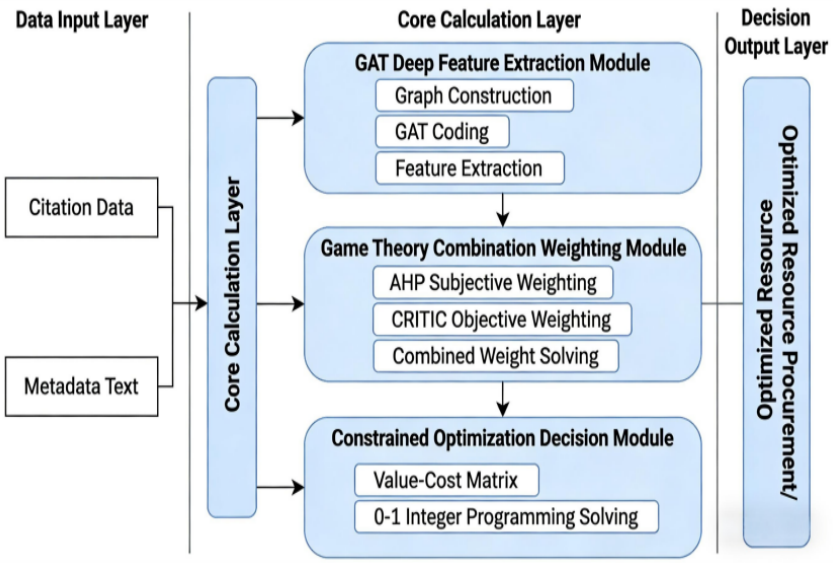

FIGURE 1. Framework of the Collection Optimization Decision Model Based on Deep Citation Network Analysis

1) Input and Graph Construction

The model input comes from academic databases, including citation relationships and metadata (title, abstract, keywords, publication year) of documents within the target resource set (e.g., journals, conference proceedings). We construct a directed graph $G = (V, E, X)$. Here, $V = \{v_1, v_2, \ldots, v_N\}$ represents the set of document nodes, E is the set of directed edges, and edge e_{ij} indicates that document v_i cites document v_j. $X \in \mathbb{R}^{N \times d}$ is the node feature matrix. The initial node feature x_i is obtained by encoding the document title and abstract using a pre-trained language model (e.g., SciBERT) to capture its semantic content [15].

2) Graph Attention Network Layer

GAT aggregates neighbor information through a multi-head attention mechanism. For node v_i at layer l, its representation $h_i^{(l)}$ is updated as follows:

$$h_i^{(l)} = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \alpha_{ij}^k W^k h_j^{(l-1)} \right). \tag{1}$$

Here, $\|$ denotes vector concatenation, K is the number of attention heads, σ is a non-linear activation function, and W^k is a trainable weight matrix. The attention coefficient α_{ij}^k is calculated as follows:

$$\alpha_{ij}^k = \frac{\exp\left(\text{LeakyReLU}\left(a^T [W^k h_i^{(l-1)} \| W^k h_j^{(l-1)}]\right)\right)}{\sum_{u \in \mathcal{N}(i) \cup \{i\}} \exp\left(\text{LeakyReLU}\left(a^T [W^k h_i^{(l-1)} \| W^k h_u^{(l-1)}]\right)\right)}. \tag{2}$$

Here, a is the parameter vector of the attention mechanism. After propagation through L layers, we obtain the final high-order representation $h_i^{(L)}$ for each node.

3) Feature Extraction

From $h_i^{(L)}$, we parse three dimensions of interpretable features:

Network Influence Feature (I_i):

$$I_i = \left\| h_i^{(L)} \right\|_2 \cdot \text{PageRank}(v_i). \tag{3}$$

The vector norm is not inherently meaningful; however, under the supervised training objective of predicting future citation counts, the learned latent space encodes semantic and structural importance. Multiplying by PageRank further incorporates global topological significance, ensuring that the norm reflects academic influence in a task-relevant manner.

Interdisciplinarity Feature (C_i): Measures the breadth of the literature's knowledge reach. We predefine M major disciplinary categories (e.g., based on ASJC classification) and construct a topic word vector set for each discipline (obtained by averaging the word vectors of a standard keyword list for that discipline). The average cosine similarity between the node representation $h_i^{(L)}$ and each disciplinary topic vector set is calculated, resulting in a disciplinary similarity distribution $s_i = [s_{i1}, \ldots, s_{iM}]$. Then, the Shannon entropy of this distribution is computed: $C_i = -\sum_{m=1}^{M} s_{im} \log(s_{im})$. A higher entropy value indicates more balanced knowledge coverage and stronger interdisciplinarity.

Knowledge Freshness Feature (T_i): Considers the temporal decay of knowledge, defined as $T_i = \exp[-\lambda \cdot (t_{\text{current}} - t_i)]$, where t_i is the publication year of the document, and λ is a decay rate parameter controlling the speed of value decline for older knowledge.

Finally, each resource i is represented by a feature vector $f_i = [I_i, C_i, T_i]^T$. The features of all resources constitute the matrix $F = [f_1, f_2, \ldots, f_N]$.

B. RESOURCE VALUE QUANTIFICATION BASED ON GAME THEORY COMBINED WEIGHTING

To aggregate multi-dimensional features into a comprehensive value score, appropriate weights need to be assigned to each feature. This study adopts a game theory-based combined weighting method to balance the subjective preferences of library decision-makers with the objective patterns in the data itself.

Subjective Weight (W_s): The Analytic Hierarchy Process (AHP) is used. The library's resource development committee performs pairwise comparisons of the three criteria—influence, interdisciplinarity, and freshness—based on the library's strategic priorities (e.g., prioritizing core classics or encouraging interdisciplinary), constructs a judgment matrix, solves it using the eigenvalue method, and after consistency testing, obtains the subjective weight vector $w_s = [w_{s1}, w_{s2}, w_{s3}]^T$, satisfying $\sum w_{sj} = 1$.

Objective Weight (W_o): The CRITIC method is employed, calculated based on the feature matrix F. This method considers both the contrast intensity of features (represented by standard deviation σ_j) and the conflict between features (measured by correlation coefficient r_{jk}). The information content C_j of the j-th feature is calculated as:

$$C_j = \sigma_j \sum_{k=1}^{3} (1 - |r_{jk}|). \tag{4}$$

The objective weight is then:

$$w_{oj} = \frac{C_j}{\sum_{k=1}^{3} C_k}. \tag{5}$$

Combined Weight (w^*): Based on cooperative game theory, a set of compromise weights closest to both subjective and objective weights is sought. This is achieved by solving the following optimization problem:

$$\begin{aligned} \min \quad & \|w^* - w_s\|^2 + \|w^* - w_o\|^2, \\ \text{s.t.} \quad & \sum_{j=1}^{3} w_j^* = 1, \quad w_j^* \geq 0. \end{aligned} \tag{6}$$

The solution to this optimization problem can be obtained using the Lagrange multiplier method, ensuring that the combined weight $w^* = [w_1^*, w_2^*, w_3^*]^T$ reflects both decision orientation and respects the internal structure of the data.

Comprehensive Value Calculation: First, normalize the feature matrix F to obtain $\tilde{F}$. The comprehensive value V_i of resource i is:

$$V_i = (w^*)^T \cdot \tilde{f}_i = w_1^* \cdot \tilde{I}_i + w_2^* \cdot \tilde{C}_i + w_3^* \cdot \tilde{T}_i. \tag{7}$$

C. OPTIMIZATION DECISION MODEL UNDER BUDGET CONSTRAINTS

The core of collection optimization is selecting a subset of resources that maximizes total value within a limited budget. Suppose there are N candidate resources, the annual cost of the i-th resource is c_i, and the library's annual total budget is B. Introduce a decision variable $x_i \in \{0, 1\}$, where $x_i = 1$ indicates procuring or renewing that resource.

The basic optimization model can be formulated as the following 0–1 integer programming problem:

$$\begin{aligned} \max \quad & \sum_{i=1}^{N} V_i \cdot x_i, \\ \text{s.t.} \quad & \sum_{i=1}^{N} c_i \cdot x_i \leq B, \\ & x_i \in \{0, 1\}, \quad i = 1, 2, \ldots, N. \end{aligned} \tag{8}$$

Extended model (addressing overlapping content and bundled pricing):

To reflect real-world collection development, the model can be extended with:

Overlap penalty: For resources with high content redundancy (e.g., two databases covering the same journals), a pairwise penalty term $-\rho \cdot y_{ij}$ is added.

Bundle constraints: If resources i and j are only available as a bundle, then $x_i = x_j$.

These extensions are formulated as:

$$\max \sum_{i=1}^{N} V_i x_i - \sum_{(i,j) \in O} p_{ij} x_i x_j. \tag{9}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} c_i x_i + \sum_{b \in B} c_b z_b \leq B. \tag{10}$$

$$x_i = x_j, \quad \forall (i, j) \in \text{Bundle}. \tag{11}$$

For large-scale problems (large N), heuristic methods such as greedy algorithms or genetic algorithms can be used for efficient approximate solving. The model can be extended according to practical needs, for example, by adding constraints for essential core resources ($x_k = 1$) or disciplinary coverage ratio constraints ($\sum_{i \in S_j} x_i \geq \theta_j$, where S_j is the resource set for discipline j).

IV. EXPERIMENT AND ANALYSIS

To validate the effectiveness and superiority of the proposed model, we designed a simulation experiment based on public data, simulating a library journal procurement decision scenario.

A. EXPERIMENT SETUP

Data Source: The experiment used a subset of the Aminer academic network dataset (https://www.aminer.cn/) [16]. This dataset contains citation relationships, titles, abstracts, and publication years of documents in the field of computer science. We selected approximately 8,000 papers published between 2018–2022 from 10 important international conferences as the base data, with each conference treated as an "resource unit" to be evaluated.

Feature Extraction: SciBERT was used to obtain initial text features. A citation graph was constructed with GAT parameters: layers $L = 2$, attention heads $K = 4$, output dimension 128. After training, the average feature vector of all papers from each conference was taken as the conference's feature vector f_i.

Cost and Budget: For simulation purposes, the cost c_i of each conference "resource" was assumed to be proportional to the total number of papers published in the last five years, with added random noise. The total budget B was set to be exactly sufficient to purchase the top 7 most expensive conference resources.

Benchmarks:

1) Benchmark 1 (IF-driven): Select conferences in descending order of conference impact factor (simulated using proxy indicators like CORE Ranking) until the budget is exhausted.
2) Benchmark 2 (Usage-driven): Select conferences in descending order of simulated historical usage frequency. Unlike the original "paper count" proxy, which reflects output volume rather than user demand, we generate a usage proxy based on the empirical correlation between citation counts and download frequencies observed in library log data ($r \approx 0.65$–0.75). This provides a more realistic demand-driven baseline.
3) Our Model (GNN-OPT): Use the complete proposed process for decision-making.

Evaluation Metrics: Since real future impact cannot be immediately obtained, we used three proxy metrics to evaluate the long-term potential value of the decision plans:

1) Future Influence: Based on the trained GAT node representations, a simple linear regression model was trained to predict the average citation count for each conference in 2023. The sum of predicted citations for the selected resource set was calculated.
2) Interdisciplinary Potential: The diversity entropy of keywords from papers in the selected conferences was calculated (similar to the interdisciplinarity feature).
3) Proportion of Core Outputs: The proportion of "highly cited papers" (those whose citation count falls within the top 20% in the global dataset) among the selected conferences was counted.

B. RESULTS ANALYSIS

The selection results and performance predictions of the three decision strategies under the same budget are compared in Table 1.

The following conclusions can be drawn from Table 1 (Note: performance improvement percentages were calculated based on Table 1 data):

1). At similar budget consumption levels, the resource set selected by our GNN-OPT model achieved a predicted future influence sum 10.3% and 21.4% higher than the IF-driven and Usage-driven benchmarks, respectively. This indicates that the network and semantic features learned by the GNN capture the long-term academic influence of resources better than the singular journal impact factor or historical usage data.

2). Regarding interdisciplinary potential, the GNN-OPT model significantly outperformed the two benchmarks, with an entropy value approximately 15–24% higher. This directly benefits from the model's explicit modeling and optimization of the interdisciplinarity feature (C_i), enabling it to identify and favor conference resources with broader knowledge diffusion.

3). In terms of the proportion of core outputs, the GNN-OPT model also maintained a lead. This suggests that while pursuing influence and interdisciplinarity, the model did not compromise the intrinsic quality of resources.

Regarding the concern of circular evaluation: The "influence feature" used in model construction was trained on 2018–2022 citation data, while the "Future Influence" metric was computed using 2023 actual citation counts (predicted via a separate linear model). These two time windows are non-overlapping, thus avoiding self-validation. The reported 10.3% improvement is a proxy-based estimate, not a claim of real-world superiority; real validation would require longitudinal data over multiple years.

Combined weight analysis: Further analysis of the GNN-OPT model's combined weight w^* revealed its value as [0.40, 0.35, 0.25]. Assuming the initial subjective weight W_s was [0.50, 0.25, 0.25] (emphasizing influence more), and the CRITIC method gave an objective weight W_o of [0.35, 0.45, 0.20].

Although the subjective weight emphasized influence (0.50), the CRITIC method assigned higher objective weight to interdisciplinarity (0.45) due to its larger standard deviation and lower correlation with other features. The game-theoretic compromise thus shifted toward the objective side—not as a failure of alignment, but as a reflection of the data's internal structure. Librarians may adjust this by tightening the subjective weight bounds or incorporating additional constraints.

V. DISCUSSION

The model constructed in this study provides a new methodological tool for libraries to cope with the complexity of digital resource management. Its core advantage lies in realizing an end-to-end modeling from unstructured academic big data to structured decision plans, combining the strengths of deep learning in feature discovery with the rigor of operations research in constraint optimization. Compared to the common "analyze first, recommend later" two-stage mode in the literature, this model achieves a closed-loop unity of feature learning and decision generation, exhibiting stronger systematicity.

In practical application, this model offers high flexibility. Libraries can adjust feature dimensions (e.g., add a "relevance to institutional authors" feature) or modify the AHP judgment matrix to reflect different strategic priorities.

TABLE 1. Performance Comparison of Different Decision Strategies (Simulation Results)

Decision Strategy	Selected Conferences Count	Total Cost (Simulated)	Predicted Future Influence Sum	Interdisciplinary Potential (Entropy)	Proportion of Core Outputs
IF-driven	7	1.00 B	4520	1.85	68.2%
Usage-driven	8	0.98 B	4105	1.72	65.5%
GNN-OPT (Our Model)	7	0.99 B	4985	2.13	71.8%

The model's output—comprehensive resource value rankings and optimized procurement lists— can provide powerful data references for collection development committees, making decision discussions more objective and transparent.

Certainly, the model also has several limitations, which point to directions for future research:

1). Data Dependency and Cold Start: Model performance depends on complete, high-quality citation and text data. For emerging fields or highly niche resources, data sparsity may affect feature learning. Future work could explore integrating multi-source data (e.g., social media mentions, patent citations) or using transfer learning to mitigate this issue.

2). Model Simplifying Assumptions: The current optimization model treats each resource as an independent item, ignoring content overlap and bundled pricing. This is a significant limitation, as real-world library acquisitions often involve "Big Deals" or redundant coverage. Future work should adopt submodular optimization or set cover models to handle marginal utility decay.

3). Consideration of Dynamic Evolution: Knowledge itself evolves dynamically. The current model makes decisions primarily based on a historical static snapshot. Introducing temporal graph neural networks to model the dynamic changes of citation networks, thereby predicting the future value trajectory of resources, would significantly enhance the foresight of decisions.

4). Integration with Traditional Methods: How to organically integrate the quantitative suggestions provided by the model with the domain knowledge of experienced librarians and immediate user feedback to build a "human-machine collaborative" decision-making mode is key to promoting its practical application.

VI. CONCLUSION

The contradiction between the complexity of digital resource development and limited financial resources requires libraries to adopt more intelligent and refined management methods to ensure the precise acquisition of high-value technical literature for the industry. This strategic approach enables institutions to optimize their budgets while maintaining the specialized data support necessary for advancing modern fiber science research. The collection optimization decision model based on deep citation network analysis proposed in this paper is a direct response to this challenge. The model employs Graph Attention Networks to deeply mine features of influence, interdisciplinarity, and timeliness embedded in citation relationships, uses game theory to balance subjective and objective weights to quantify the comprehensive value of resources, and finally solves for the optimal allocation plan under budget constraints through integer programming. Simulation experiments show that this model can effectively identify resources with higher long-term academic potential and knowledge diffusion capability, and its overall performance surpasses traditional decision methods based on single indicators.

Although the validation experiment was conducted on a computer science dataset (Aminer) due to its public availability, the methodology is domain-independent. Future work will apply the model to material science corpora, such as SCI journals and patent databases, to align the evaluation context with the application domain.

The model's current limitations—especially the neglect of resource overlap and bundled pricing—must be addressed before deployment in real library systems. Nevertheless, this work demonstrates a new paradigm for data-driven, model-empowered intelligent collection development. With the continuous advancement of artificial intelligence and data analysis technologies, such decision intelligence systems are expected to become an important component of future library core infrastructure, assisting libraries in evolving their role from "resource warehouses" to "academic knowledge engines." Future work will continue to deepen the dynamics and interpretability of the model to build a next-generation collection decision support system that aligns specialized knowledge acquisition with the rapid technological evolution of the industry. This evolving system will foster close collaboration between human experts and intelligent algorithms to ensure that academic resources accurately reflect the cutting-edge developments in material science.

AUTHOR CONTRIBUTIONS

Chuanmei Xie designed, collected and analyzed the data, and drafted the manuscript. Chuanmei Xie conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Chuanmei Xie participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] L. Wang, "Analysis of Library Resource Optimization Allocation Strategy Based on Big Data Intelligent Technology," Electronic Technology, vol. 53, no. 7, pp. 284-285, 2024, doi: 10.3969/j.issn.1000-0755.2024.07.129.

[2] H. Hu, B. Zhou, and S. Zhang, "Book Procurement Recommendation Service in the Digital Transformation Environment," New Century Library, no. 8, pp. 54-61, 2024, doi: 10.16810/j.cnki.1672-514X.2024.08.008.

[3] Z. La, J. Shen, Y. Song, et al., "ML-HAN: Multi-level heterogeneous graph attention network for representation learning with semantic diversity via feature-node-semantic attention," Neurocomputing, vol. 660, 131962, 2026, doi: 10.1016/j.neucom.2025.131962.

[4] P. Lu, "Intelligent Big Information Retrieval of Smart Library Based on Graph Neural Network (GNN) Algorithm," Computational Intelligence and Neuroscience, vol. 2022, 1475069, 2022, doi: 10.1155/2022/1475069.

[5] Y. Chen, S. Shi, X. Zhang, et al., "Research on Digital Resource Guarantee Strategy Based on Citation Analysis of Institutional Repository Outputs," Journal of Academic Libraries, vol. 40, no. 2, pp. 36-43, 2022, doi: 10.16603/j.issn1002-1027.2022.02.005.

[6] R. Song, "Research on Mitigation Techniques for Shortcut Learning in Pretrained Language Model Text Classification," Jilin University, 2024, doi: 10.27162/d.cnki.gjlin.2024.000867.

[7] H. Huang, "Research on the Establishment and Quantitative Analysis of Evaluation System Indicators for Digital Collection Resources in Public Libraries Based on Analytic Hierarchy Process," Journal of the Library Science Society of Henan, vol. 44, no. 12, pp. 36-40, 2024, doi: 10.3969/j.issn.1003-1588.2024.12.012.

[8] W. Sun, Z. Wang, Y. Gao, et al., "Research on Library Collection Optimization Using Game Theory Combined Weighting Method-A Case Study of Hebei University of Technology Library," Library and Information Service, vol. 62, no. 10, pp. 40-46, 2018, doi: 10.13266/j.issn.0252-3116.2018.10.006.

[9] J. Qiu and K. Dong, "Method and Empirical Study on Deep Aggregation of Literature in Citation Networks-Taking XML Research Papers in the WOS Database as an Example," Journal of Library Science in China, vol. 39, no. 2, pp. 111-120, 2013, doi: 10.13530/j.cnki.jlis.2013.02.013.

[10] J. Li, L. Xu, Y. Wang, et al., "A Survey of Citation Classification Research for Scientific Literature Based on Deep Learning Techniques," Frontiers of Data & Computing, vol. 5, no. 4, pp. 86-100, 2023, doi: 10.11871/jfdc.issn.2096-742X.2023.04.008.

[11] M. Wang, H. Zhao, L. Wu, et al., "Mining Edge Formation Factors in Citation Networks Enhanced by Language Models," Big Data Research, vol. 11, no. 2, pp. 91-106, 2025, doi: 10.11959/j.issn.2096-0271.2025025.

[12] J. Xu, W. Zuo, S. Liang, et al., "Causal Relation Extraction Based on Graph Attention Network," Journal of Computer Research and Development, vol. 57, no. 1, pp. 159-174, 2020, doi: 10.7544/issn1000-1239.2020.20190042.

[13] Y. Wei, Y. Nong, Z. Li, et al., "MTADGAI: A Multivariate Time Series Anomaly Detection Model Integrated with Graph Attention Network," Guangxi Sciences, 2026, doi: 10.13656/j.cnki.gxkx.20260129.001.

[14] X. Peng, H. Zhou, and J. Shi, "Mining, Analysis, and Application of Citation Characteristics of Chinese Books in Library Service Context-Taking G-category Books as an Example," Journal of Modern Information, vol. 43, no. 10, pp. 107-119, 2023, doi: 10.3969/j.issn.1008-0821.2023.10.010.

[15] S. Huang and J. M. Cole, "BatteryBERT: A Pretrained Language Model for Battery Database Enhancement," Journal of Chemical Information and Modeling, vol. 62, no. 24, pp. 6365-6377, 2022, doi: 10.1021/acs.jcim.2c00035.

[16] J. Tang, J. Zhang, L. Yao, et al., "Extraction and Mining of an Academic Social Network," in Proceedings of the 17th International Conference on World Wide Web. Beijing, China, 2008, doi: 10.1145/1367497.1367722.