Postprint: Intelligent Evaluation of Content Innovativeness in Academic Papers Based on Knowledge Elements

Authors: Li He, Du Xingye

Date: 2023-04-01T16:15:47+00:00

Abstract

[Purpose/Significance] Innovation represents the most fundamental requirement for academic paper quality, constituting the soul of academic papers and the core of academic paper evaluation. Knowledge element serves as the basic composition unit of academic papers. Based on knowledge element theory and machine learning related theories and algorithms, this study investigates from the content perspective of academic papers how computers can intelligently perform innovation evaluation and its implementation process and methods. [Method/Process] First, we construct four knowledge element ontologies for academic papers—research question, theory, method, and conclusion—and subsequently propose an academic paper innovation evaluation model based on knowledge elements. Second, according to the research characteristics of academic papers, we construct a machine classification model for theories and methods as well as extraction rules and methods for knowledge elements, establishing a rule base and knowledge corpus. Finally, based on semantic similarity calculation methods, we score the innovation of academic papers across four dimensions according to evaluation rules and relevant weights. [Results/Conclusion] Empirical results from the academic paper innovation scoring system based on knowledge element extraction demonstrate that this intelligent evaluation method possesses certain feasibility and can provide methodological reference for the ultimate realization of intelligent evaluation systems for academic paper content innovation.

Full Text

Research on Intelligent Evaluation of Content Innovation in Academic Papers Based on Knowledge Elements

Li He[1], Du Xingye[1,2,3]

[1]School of Management, Jilin University, Changchun 130002
[2]National Science Library, Chinese Academy of Sciences, Beijing 100190
[3]Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract: [Purpose/Significance] Innovation is the most fundamental requirement for academic paper quality, representing the soul of academic papers and the core of academic evaluation. Knowledge elements constitute the basic units of academic papers. Based on knowledge element theory and machine learning algorithms, this study investigates how computers can intelligently evaluate innovation from the content perspective of academic papers, including the implementation process and methodology. [Method/Process] First, we constructed four knowledge element ontologies for academic papers: research problem, theory, method, and conclusion, and proposed a knowledge element-based innovation evaluation model. Second, according to the characteristics of academic research, we built machine classification models for theories and methods, established extraction rules and methods for knowledge elements, and created rule bases and knowledge corpora. Finally, based on semantic similarity calculation methods, we scored the innovation of academic papers across four dimensions according to evaluation rules and relevant weights. [Result/Conclusion] Empirical results from the knowledge element extraction-based academic paper innovation scoring system demonstrate the feasibility of this intelligent evaluation method, providing methodological references for the ultimate realization of intelligent evaluation systems for academic paper content innovation.

Keywords: Academic papers; Knowledge element; Content innovation; Intelligent evaluation

---

Innovation serves as the crucial foundation for national development. From the perspective of national knowledge innovation strategy, knowledge innovation reflects a country's research output capacity, knowledge dissemination ability, and overall scientific and technological strength. Scientific research achievements, primarily represented by academic papers, constitute an essential component of knowledge innovation. The quality and quantity of academic papers serve as important indicators for measuring a nation's innovation capacity and vitality. Academic paper evaluation forms the basis and essential content of knowledge innovation capacity assessment, representing one of the key metrics in the national comprehensive innovation capability measurement system. The core of academic paper evaluation lies in assessing paper quality, academic value, and academic impact. Currently, pre-publication evaluation of academic papers mainly relies on expert anonymous review, which is limited by factors such as experts' academic levels and disciplinary fields. This approach has certain limitations that may cause some excellent achievements to be overlooked or published with delay, while some inferior results appear in influential journals, thereby negatively impacting national comprehensive innovation capability evaluation.

Developments in knowledge management, big data, and artificial intelligence provide new possibilities for overcoming the drawbacks of academic paper review. Moreover, since knowledge elements in academic papers can be used not only to express, store, retrieve, and utilize knowledge but also to describe the evolution of knowledge and facilitate knowledge discovery, this paper attempts to investigate the theories and methods for intelligent evaluation of academic paper innovation based on knowledge element theory, leveraging big data and artificial intelligence technologies.

1. Overview of Academic Paper Innovation Evaluation

Innovation represents the most fundamental requirement for academic paper quality, the soul of academic papers, and the core of academic evaluation. Innovation evaluation of academic papers encompasses multiple dimensions. From a content perspective, it includes: viewpoint innovation (proposing new viewpoints or research questions that others have not raised in a particular field); academic theory innovation (discovering new phenomena or revealing new patterns, or proposing new theories); structural or methodological innovation (proposing new perspectives or research methods based on existing research, improving or perfecting existing methods, or applying existing methods to solve new problems in application fields); and results/conclusion innovation (accompanying theoretical and methodological innovation, referring to obtaining different results and conclusions from previous achievements based on the first three types of innovation).

Regarding innovation levels, Chen Jianqing [1] categorized innovation into three hierarchical levels: pioneering, original, and improvement. Pioneering research achievements refer to those with far-reaching impact, global significance, forward-looking vision, strategic importance, breakthrough nature, and disruptive innovation in a professional discipline or field at home and abroad, representing the most innovative achievements. Original research achievements refer to new topics proposed in existing professional fields with original innovation or independent intellectual property rights. Improvement-based research achievements refer to supplementary, improved, or perfected research work based on existing research topics, objects, and achievements. Innovation is hierarchical [2], with pioneering innovation at the highest level, followed by stage-based innovation, and then applied innovation. From the perspective of definitions by authoritative international academic journals, *Nature* believes that innovative scientific research should possess novelty and attract attention while having broad significance beyond the field. *Science* considers innovation as proposing new insights into nature or theory rather than re-demonstrating existing research conclusions.

Regarding innovation evaluation methods, peer review is currently the most recognized approach in the international academic community [3], with citation analysis also being widely used. For instance, scholars in the field of bibliometrics [4] believe that papers with high innovation potential are more likely to become highly cited papers. However, both peer review and citation-based bibliometric analysis have limitations when evaluating academic paper innovation. Peer review limitations mainly include: (1) heavy reliance on reviewers' subjective judgments with inconsistent evaluation standards; (2) narrow scope for selecting qualified peers, making it difficult to determine the credibility of expert evaluations; and (3) difficulty ensuring the fairness of anonymous review. Additionally, as a qualitative evaluation method, peer review suffers from subjective arbitrariness, inefficiency, concealed evaluation processes, and results that are difficult to replicate and supervise [5]. Quantitative evaluation, on the other hand, has limited universal applicability across different disciplines, and applying metrology to evaluate academic achievements may mislead scholars [6].

Currently, the inherent defects of pre-publication evaluation based on peer review and post-publication evaluation based on bibliometrics (including altmetrics) have not been fundamentally resolved. Research aimed at improving peer review defects, such as open peer review [7], expert selection for peer review [8], and quantitative processing of peer review forms [9], has addressed certain issues like fairness requirements to some extent but cannot solve the problem of subjective expert judgment in peer review. Although bibliometrics-based academic paper impact evaluation has undergone multiple improvements, such as altmetrics supplementing citation-based evaluation, it still cannot fundamentally address impact issues from the perspective of academic content.

Researchers have therefore shifted their focus to analyzing the content of academic papers themselves, identifying innovation points through content analysis and mining to construct innovation measurement indicators for academic papers. Several relevant studies have emerged. For example, Shen Yang [10] evaluated paper innovation from the perspective of keywords, extracting existing keywords based on statistical analysis of keyword frequency across different periods, arguing that higher frequency, longer duration, and lower user evaluation indicate lower innovation. He Wanying [11] evaluated paper innovation from multiple dimensions of innovation absorption and diffusion, achieving certain results but still paying insufficient attention to content factors within academic papers themselves. Suo Chuanjun et al. [12] measured the aging and innovation of individual academic papers using the quantity of knowledge element transfer. Yang Jing et al. [13] evaluated paper innovation based on research topics, arguing that if a paper's research topic aligns with current scientific research frontiers and is published in a high-impact journal, the paper possesses high innovation. Ruan Guangce [14] used the Doc2Vec method for vector calculation and similarity computation of text content to generate hot topic paper collections, then applied topic models and clustering algorithms for topic identification and mining, achieving better results in semantic feature identification, which can serve as an important foundation for identifying and judging the novelty and innovation of paper topics—the core of this study's content innovation intelligent evaluation.

2. Concepts and Process of Intelligent Evaluation of Academic Papers

Evaluation refers to the process of comparing evaluation objects under certain standards to help users better understand them and guide decision-making [15]. Intelligent evaluation applies artificial intelligence theories, methods, and technologies to the evaluation process to understand evaluation objects. Intelligent evaluation of academic papers refers to the evaluation process where key indicators for judging paper quality can be assessed by artificial intelligence technologies or automatically completed by computer programs to provide evaluation results. Through intelligent evaluation, the drawbacks of subjective judgment inherent in traditional peer review can be partially or fully resolved.

Theories and methods for intelligent evaluation of academic papers initially adapted to academic impact after publication (e.g., judging academic paper impact from 计量指标 based on massive data). With the development of big data technologies, especially knowledge representation, knowledge reasoning, text recognition and analysis, knowledge discovery, and machine learning, it has become possible for computers to intelligently judge and evaluate the quality of individual academic papers (manuscripts or submissions) that have not yet entered the dissemination domain. The application of intelligent technologies to pre-publication evaluation of academic papers is an evolutionary process that gradually penetrates various processes [16] or main content stages [17] of academic paper evaluation as technology and methods advance.

Evaluation factors include evaluation purpose, subject, object, indicator system, criteria, model, and results. Based on the degree of automation achieved by intelligent technology in evaluation or participation in traditional review processes, intelligent evaluation of academic papers can be divided into three stages: (1) early-stage computer-aided evaluation focusing on external indicators; (2) mid-stage evaluation primarily relying on computers for content-level identification and evaluation, focusing on content innovation; and (3) mature-stage fully automated evaluation where computers automatically generate main review comments. Among these, technologies for evaluating external factors have matured, while technologies for evaluating content itself represent current research hotspots and constitute the main focus of this paper. Fully automated evaluation that generates review comments represents the future direction.

The intelligent evaluation process involves using intelligent systems to understand evaluation objects. Through self-organization, self-learning, self-adaptation, self-recognition, and self-coordination functions, models become intelligent comprehensive evaluation models that better serve users' decision-making needs. The intelligent evaluation process for academic paper content includes three aspects: (1) intelligent content identification; (2) intelligent content extraction; and (3) intelligent content comparison. Intelligent identification recognizes content features based on intelligent methods, such as identifying research topics. Intelligent extraction extracts content based on description rules following identification. Intelligent comparison is a crucial step in evaluation, involving semantic similarity calculation and comparison of identified and extracted content, followed by automatic feature classification using machine learning to determine novelty and innovation. The intelligent evaluation process for academic papers is shown in [Figure 1: see original paper].

3. Knowledge Element Theory

3.1 Concept of Knowledge Elements

Knowledge elements are the basic units for representing, controlling, managing, and manipulating knowledge. They emerged to address the limitation that knowledge organization based on documents as units contains too little knowledge content to meet users' growing knowledge demands [18]. In the late 1970s, American information scientist Vladimir Slamecka pointed out that delving into knowledge elements within documents as the unit of knowledge control would generate significant knowledge value-added, thereby improving knowledge utilization and creation efficiency [19]. British information scientist B.C. Brookes subsequently proposed using the concept of "cognitive viewpoint" maps to connect and represent knowledge content and knowledge creation [20], while evolving document networks into conceptual networks of knowledge element associations, transforming knowledge systems from external macrostructures to internal microstructures [21].

Knowledge elements can be used not only to express, store, retrieve, and utilize knowledge but also to describe knowledge evolution trajectories, facilitate knowledge discovery, and predict future development directions. Knowledge elements manifest differently across disciplines and periods. In education, they refer to "knowledge points" in knowledge systems; in artificial intelligence, they refer to "semantic webs"; and in library and information science, they represent basic concepts in documents [22]. Wen Youkui considers knowledge elements as the basic units constituting knowledge structures, representing the smallest units into which knowledge can be decomposed for independent use. They can express complete knowledge content or concepts, forming a set of information units containing certain knowledge components [23]. Based on the definition of knowledge elements as basic units, they can effectively address: (1) free segmentation and access of knowledge; (2) free organization and retrieval of knowledge; (3) free combination and retrieval of knowledge; and (4) accurate measurement and evaluation of knowledge.

3.2 Description and Extraction of Knowledge Elements

Knowledge element description includes both description models and description rules. Knowledge element description models [18] are abstract representations that reveal the semantic content and structure of knowledge elements, serving as methods for knowledge element representation with the goal of facilitating knowledge element management and utilization. Description rules are formulated for knowledge element identification and extraction, representing the sum of knowledge element representations developed based on description models and feature analysis.

Knowledge element description models generally include three aspects: attributes, content, and relationships. Suo Chuanjun et al. [18] used semantic triples to describe innovative knowledge elements, arguing that each innovative knowledge element can be decomposed into at least one subject-predicate-object form. These semantic triples possess certain logical relationships because they describe knowledge content under the same theme. Yuan Mingyi et al. [25] proposed an ontology-based knowledge element representation method. Ontology provides standardized specifications of terms and their relationships in a particular domain, offering common understanding and description for knowledge sharing, communication, and reuse, composed of precisely defined concepts and their relationships. The knowledge element ontology includes five categories: Creator, KnowledgeElement, KnowledgeElementAbstract, KnowledgeElement-Description, and History. Creator describes the creator, KnowledgeElement describes different knowledge units, KnowledgeElementAbstract represents the knowledge element abstract, KnowledgeElementDescription represents the knowledge element description body, and History records the evolution process of knowledge elements.

Extracting knowledge elements from digital resources forms the foundation for knowledge element application. Current methods proposed by scholars can be broadly categorized into text structure-based extraction methods and rule-based extraction methods. Text structure-based methods include Jiang Yongchang's [26] approach based on physical and logical structures; Zhou Ning et al.'s method extracting text fragments according to predefined structural constraints; and Fang Long et al.'s [27] identification based on functional structures of academic texts. Rule-based methods include Wang Zhongyi et al.'s [28] approach, which first established description rules for conceptual, factual, numerical, methodological, and relational knowledge elements, detailed feature words for each knowledge element type, and then identified and extracted knowledge elements based on these rules.

Building on relevant knowledge element research, this paper constructs an academic paper knowledge element ontology, establishes extraction rules for knowledge elements based on the ontology, extracts knowledge elements from academic papers, and evaluates their innovation through semantic similarity calculation between knowledge elements and academic paper semantic information.

4. Innovation Evaluation Process for Academic Papers Based on Knowledge Elements

As discussed above, academic paper content innovation primarily concerns new arguments and new evidence. New arguments include new problems, new theories, and new conclusions, while new evidence includes new methods and new data. Academic paper achievements should demonstrate innovation in micro-level aspects such as research problems, theories, methods, and conclusions. Therefore, this paper divides academic paper content innovation evaluation into four dimensions: research problem innovation, theoretical innovation, methodological innovation, and conclusion innovation.

Research problem innovation (also called research topic or research selection innovation) refers to researchers proposing a new research problem, topic, viewpoint, or perspective, from which research value and innovation can be preliminarily judged. Theoretical innovation refers to researchers making new rational analyses and solutions to emerging problems in social practice activities, revealing and predicting the essence, laws, and development trends of cognitive or practical objects, and sublimating historical and practical experiences with new rationality. It represents new breakthroughs in original theoretical systems or frameworks, new corrections and developments to existing theories, and new explorations in unknown fields. Methodological innovation refers to proposing new methods for existing research objects, improving existing methods, or applying existing methods to solve problems in application fields. Conclusion innovation accompanies research problem, theoretical, and methodological innovation, referring to obtaining different results or conclusions from previous achievements based on the above innovations. The specific dimensions are shown in [Figure 2: see original paper].

To judge the innovation of the above content, we must first identify and extract relevant knowledge from target academic papers and conduct comparative analysis or similarity calculation with existing academic paper knowledge bases to determine innovation. The specific steps are:

1. Establish knowledge element ontologies, description rule bases, and terminology bases: Describe extraction rules for knowledge elements reflecting the four innovation dimensions to establish a description rule base; describe standardized terminology to establish a terminology base.
2. Based on knowledge element description rules, extract knowledge elements from published academic paper knowledge bases within a certain time window to establish knowledge element ontology bases, including research problem, theory, method, and conclusion knowledge element ontology bases.
3. Establish a knowledge element graph base (i.e., knowledge linking network). Identify knowledge elements in academic papers within a certain time window and construct knowledge element graphs for each paper, marking timestamps to form a knowledge element graph base.
4. Extract knowledge elements from target academic papers and construct knowledge element graphs for target papers.
5. Calculate the innovation of target academic paper knowledge elements to obtain the target paper's innovation index. Match and calculate similarity between target papers and the knowledge element ontology base and

graph base to obtain innovation indices at theoretical, methodological, and application levels.

The overall process for academic paper content innovation evaluation is shown in [Figure 3: see original paper]. First, describe academic paper knowledge elements, establish knowledge element extraction rules to form an extraction rule base, and use the rule base to extract knowledge elements from academic papers based on the knowledge element ontology. Construct the academic paper knowledge element ontology, perform similarity calculation with knowledge elements extracted from target papers, and obtain innovation evaluation results.

4.1 Academic Paper Knowledge Element Ontology

Academic papers have universal metadata. Basic metadata features include title, author, affiliation, abstract, keywords, classification number, DOI, journal name, publication date, supporting fund, topic, keywords, cited references, etc. Among these, title, abstract, keywords, topic, classification number, and cited references are related to paper content. Academic paper content structures are also similar. Taking papers from *Library and Information Service* as an example, they typically include introduction, research status or related research, theory, method research or model construction, examples or experiments, conclusion, references, and other basic structures. These structures constitute a hierarchical system and respective functions for academic papers. Focusing on innovation evaluation, this paper primarily extracts core content knowledge such as research problems, theoretical knowledge points, methodological knowledge points, and result knowledge points. Research problems belong to the topic domain and can be obtained from titles, keywords, abstracts, and introductions; theoretical knowledge points originate from titles, keywords, abstracts, related theories, theoretical model construction, and conclusions; methodological knowledge points originate from titles, keywords, abstracts, related theories, method research, examples or experiments; result knowledge points originate from abstracts, examples or experiments, and conclusions. Based on this analysis, we construct knowledge element description rules and ontologies for academic paper innovation evaluation.

4.1.1 Overall Structure of Academic Paper Knowledge Element Ontology Based on analysis of knowledge elements and ontology models, this paper constructs a logical structure description model for knowledge element ontologies. Current knowledge element description models include triple, quadruple, quintuple, and sextuple models, lacking unified standards and frameworks. Considering the use of RDF format for storing knowledge element ontologies, this paper selects the triple model as the logical description model. The RDF model, proposed by W3C, describes web resources using subject-predicate-object semantic triples, including resource metadata. The subject refers to the described resource, the predicate is the attribute, and the object is the attribute value. Based on this analysis, this paper constructs academic paper ontology, research problem ontology, theory ontology, method ontology, and conclusion ontology, with relationships shown in [Figure 4: see original paper].

4.1.2 Academic Paper Knowledge Element Ontology Academic paper attributes primarily consist of metadata, including title, author, abstract, keywords, classification number, DOI, journal name, publication date, supporting fund, topic, and cited references. Academic literature types generally include books, reports, and conference articles. This paper focuses on journal articles, classifying them as review papers or research papers rather than by literature type, treating these as subclasses of academic paper entities. To improve ontology sharing and reuse, the ontology constructed in this paper inherits concepts from doco, fabio, and deo ontologies. The specific conceptual hierarchy is shown in [Figure 5: see original paper].

4.1.3 Academic Paper Research Problem Knowledge Element Ontology Research problems are the cornerstone of academic paper research. Without research problems, academic research loses its meaning. Research problems in academic papers are reflected through research objects, backgrounds, purposes, and significance. The hierarchical model of the academic paper research problem knowledge element ontology is shown in [Figure 6: see original paper].

4.1.4 Academic Paper Theory Knowledge Element Ontology Theoretical innovation is an important part of research innovation, including innovation in ideas and doctrines. Theoretical innovation content in academic papers is mainly reflected through title feature words, subject terms, theoretical viewpoints, hypothetical models, framework models, and conclusions contained in the paper. Based on different paper structures, academic theory entities are divided into theoretical viewpoints, hypothetical models, and framework models. Accordingly, this paper constructs a hierarchical model of the academic paper theory knowledge element ontology, as shown in [Figure 7: see original paper], providing a foundation for theoretical innovation evaluation data preparation.

4.1.5 Academic Paper Method Knowledge Element Ontology Method usage in academic papers is relatively complex, generally divided into scientific research methods and problem-solving methods. Scientific research methods mainly include questionnaire surveys, expert interviews, case studies, observation, literature research, and experiments. Problem-solving methods include algorithms, technical methods, evaluation models, and mathematical models. Since different methods have different attributes, in the method entity, questionnaire surveys, expert interviews, case studies, observation, literature research, experiments, algorithms, technical methods, evaluation models, and mathematical models are all treated as method subclasses, constructing the academic method knowledge element ontology hierarchy shown in [Figure 8: see original paper].

4.1.6 Academic Paper Conclusion Knowledge Element Ontology Paper conclusions are important components, including main conclusionary, viewpoint, and innovative knowledge. Core conclusion elements mainly include countermeasures, suggestions, implications, research value, advantages, and innovation points. Based on this, the hierarchical structure of the academic paper conclusion knowledge element ontology is constructed, as shown in [Figure 9: see original paper].

4.2 Academic Paper Knowledge Element Extraction

4.2.1 Knowledge Element Extraction Rules By analyzing academic knowledge element extraction requirements and examining academic paper characteristics, this paper designs a knowledge element extraction scheme. Academic paper titles, keywords, and content structures typically consist of introduction, related theories, research methods/content, experiments/cases, and conclusions, allowing for knowledge element extraction based on content. To facilitate extraction, several constraints are established: (1) English case is not distinguished during extraction; (2) long strings are prioritized in regular expression design; and (3) greedy matching mode is selected to obtain as much information as possible. Before text analysis, preprocessing is performed to filter out formatting symbols, line breaks, and other unnecessary characters. Academic paper text content has strong regularity, especially in abstracts. During rule-based knowledge element identification, standardized terminology base data is combined to finally obtain paper knowledge elements.

(1) **Academic Paper Research Problem Knowledge Element Extraction.** Target text identification and extraction includes: Title and keywords: perform topic identification on titles, match with keywords to determine research objects and content; Research purpose, significance, and practical implications in abstracts; Research content mentioned in introductions. Partial extraction rules for academic paper research problem knowledge elements are as follows:

- Keyword extraction
- Research/analysis of (.*?) problems in (.*?) field
- For/toward (.*?) field, research (.*?) problems
- Based on (.*?), research (.*?), solve (.*?) problems
- Propose/elaborate (.*?) solutions/problems/methods
- Discuss/analyze (.*?) influencing factors/elements

(2) **Academic Paper Theory Knowledge Element Extraction.** Target text identification and extraction includes: Theoretical explanations mentioned in abstracts; Discussions on theoretical deficiencies in introductions and literature reviews; Related theoretical statements in theoretical foundation sections; Theoretical contributions mentioned in conclusions. Partial extraction rules for academic paper theory knowledge elements are as follows:

- Propose (.*?) theory/hypothesis/framework/model
- Improve/enhance/perfect (.*?) theory
- Have (.*?) theoretical significance
- Review/analyze/evaluate (.*?) theory
- Apply/borrow (.*?) theory

(3) **Academic Paper Method Knowledge Element Extraction.** Target text identification and extraction includes: Method explanations mentioned in abstracts; Discussions on methodological deficiencies in introductions and literature reviews; Methods in experimental sections; Methods mentioned in conclusions. Partial extraction rules for academic paper method knowledge elements are as follows:

- Propose (.*?) method/process/algorithm/procedure
- Improve/develop/perfect (.*?) method
- Improve/develop/perfect (.*?) algorithm
- Improve/develop/perfect (.*?) process
- Review/analyze/evaluate (.*?) method
- Apply/borrow (.*?) method
- Such as survey questionnaire quantity (.*?)
- Questionnaire reliability/validity (.*?)
- Precision/recall (.*?)

(4) **Academic Paper Conclusion Knowledge Element Extraction.** Target text identification and extraction includes: Results and conclusions mentioned in abstracts; Research purposes in introductions and literature reviews; Results in experimental sections; Conclusions. Partial extraction rules for academic paper conclusion knowledge elements are as follows:

- Improve (.*?)
- Draw (.?) conclusions: (1) (.?), (2) (.?), (N) (.?)
- Verify/improve/enhance (.*?)
- Verify/prove (.?) feasibility/validity/is feasible/accuracy reaches (.?)
- Research finds (.*?)
- Results show (.*?)

4.2.2 Theory and Method Classification Model Based on Machine Learning

(1) **Construction of Theory and Method Classification Models.** Innovation judgment is the process of evaluating knowledge element innovation. Machine learning-based theory and method classification enables computers to automatically discover and fully understand basic rules and semantics from training sets (published paper knowledge elements), representing them in computer-recognizable forms as the basis for unknown text judgment—i.e., automatic computer classification. In recent years, numerous studies have applied machine learning methods to text classifi-

cation, particularly in sentiment classification. For instance, Du Hui et al. [29] used word vectors containing contextual semantic information to construct text feature representations for sentiment classification using machine learning. Li Huifu et al. [30] used principal component analysis, latent semantic analysis, Word2Vec, and TF-IDF feature extraction methods as multi-type classifier fusion features, achieving excellent performance across various corpora.

After examining multiple machine learning algorithms, this paper ultimately selected the Naive Bayes model. The Naive Bayes classifier is a commonly used machine learning method, a supervised learning algorithm with good robustness and speed, particularly suitable for big data processing, and frequently used in text classification in recent years.

The theory and method classification model based on Bayesian theory is shown in [Figure 10: see original paper].

(2) **Quantification of Theories and Methods.** Authors' descriptions of theories and methods in academic papers follow certain rules, generally expressed as descriptions of these two aspects. Analysis of numerous academic papers reveals that this description structure can be summarized as "verb + feature word + adverb" or "verb + feature word." For example, in methodological discussions, phrases like "Based on ***, propose a*** method" or "Improve the *** method" are common. Therefore, evaluation indicator quantification uses *** in the description as feature words, locating intervals [−u, u] before and after feature words, and assigning values to verbs or adverbs within the interval based on corpora. Partial assignments (5-point scale) are shown in .

4.2.3 Rule Base Construction Rule-based knowledge element extraction matches rules with text content to extract required information. This deterministic information extraction method is typically implemented through regular expressions, offering high accuracy but lacking flexibility. This paper constructs extraction rule bases by analyzing paper content. The specific construction process is shown in [Figure 11: see original paper].

The process mainly includes: (1) Based on academic paper knowledge element ontology entities, classifications, and attributes, 梳理 and analyze paper content, filter complete sentences containing required information to form a primary selection set; (2) Use SVM models to classify sentences in the primary selection set, forming a rule corpus set; (3) Perform word segmentation and part-of-speech tagging on sentences in the rule corpus, analyze sentence structures (e.g., "…construct a…model," "…propose a…method," "…build a…framework"), determine knowledge element ontology types, summarize these sentence structures to form a candidate rule set; (4) Build rules based on the candidate rule set, express rules through regular expressions, optimize and refine rules through continuous information extraction experiments, construct stable rule templates, and ultimately obtain the rule base.

4.3 Intelligent Innovation Evaluation Process for Academic Papers

Based on the constructed academic paper knowledge element base, innovation evaluation is achieved by extracting paper knowledge elements (after training) and comparing them with existing academic paper knowledge elements to obtain basic data for innovation evaluation. The intelligent evaluation process for academic paper innovation is shown in [Figure 12: see original paper].

The specific steps are:

Step 1: Academic Paper Knowledge Element Extraction. First, preprocess text to filter unnecessary characters; then combine standardized terminology base data in rule-based knowledge element identification to finally obtain academic paper knowledge elements.

Step 2: Numerical Comparison. Academic paper knowledge elements include both numerical and textual elements. Numerical knowledge elements mainly include methodological knowledge elements, such as survey questionnaire quantities, questionnaire reliability and validity, primarily involving paper scientificity. Conclusion knowledge elements, such as precision and recall, involve conclusion innovation evaluation. This paper focuses on paper innovation evaluation, therefore selecting conclusion numerical values and comparing them according to their specific definitions to determine conclusion innovation.

Step 3: Text Similarity Calculation. When knowledge element types are textual, text similarity must be judged. This paper primarily uses word vector methods. Word2Vec is a model that generates word vectors, enabling each word to obtain a corresponding word vector. Word similarity is obtained by calculating cosine values between word vectors. After introducing word vectors, words with different forms but related or similar meanings can be identified, compensating for shortcomings of traditional text similarity algorithms. This paper uses Chinese Wikipedia corpora to train word vectors.

Step 4: Target Academic Paper Innovation Evaluation. Based on text similarity and numerical comparison results, this paper calculates paper innovation results. First, calculate research problem innovation evaluation results; second, evaluate theoretical innovation; third, evaluate methodological innovation; and finally, evaluate conclusion innovation results.

4.4 Empirical Testing of Intelligent Innovation Evaluation for Academic Papers

Based on the above research process, this paper uses Python, combines Python's NLP toolkit with the Flask framework, and conducts empirical testing of intelligent evaluation across four innovation dimensions.

4.4.1 Dataset Acquisition The experimental dataset primarily consists of submissions to *Library and Information Service* from 2015-2018 and published academic papers from core library and information science journals from 2015-2017 (approximately 6,000 entries). Considering experimental convenience, this experiment only collected paper titles, Chinese abstracts, and keywords. Some abstracts were directly entered as structured abstracts (e.g., dividing abstracts into [Purpose/Significance], [Process/Method], [Result/Conclusion]) to enhance computer recognition effectiveness.

4.4.2 Experimental Module Composition and Functions The experimental modules include dataset management, extraction rule management, innovation evaluation, dataset training, and paper comprehensive scoring modules.

Dataset Management Module: Maintains datasets in the database, primarily referring to published academic paper knowledge bases.

Extraction Rule Management Module: Sets scoring rules for theoretical and methodological dimensions, calculating scores for these two dimensions based on established rules.

Innovation Evaluation Module: Based on system datasets and combined with extraction rules, derives innovation comprehensive scores from four dimensions. Current module functions include: (1) Basic dataset management (add, modify, delete); (2) Custom scoring rules, including rule descriptions, scores, and priorities; (3) Based on custom scoring rules and regular expression matching, scores theoretical and methodological dimensions and displays matched scoring rules; (4) Calculates research problem and conclusion dimension innovation scores based on semantic similarity; (5) Research problem and conclusion innovation scores can be calculated for different annual datasets, displaying corresponding semantic similarity values and experimental data analysis/results; (6) Combines four dimensions with coefficients to calculate comprehensive paper scores; (7) Generates innovation distributions based on paper innovation scores and provides adoption/rejection recommendations according to set thresholds.

Dataset Training Module: Uses machine learning methods combined with innovation evaluation indicators to construct training models for dataset scoring.

Paper Comprehensive Scoring Module: Provides comprehensive innovation scores for published and newly submitted papers based on trained models, then uses machine learning to evaluate whether papers should be published based on these comprehensive scores.

4.4.3 Partial Experimental Results and Analysis Randomly selecting papers published in *Library and Information Service* in 2016, 2017, and 2018 (the database currently contains papers from 2015-2018) and papers published in other journals for innovation scoring calculation, partial results are shown in .

From the partial random results in , most papers show decreasing innovation scores as calculation years increase, consistent with the general law of innovation diffusion. As data volume increases and learning/training iterations grow, innovation dimension judgments will become more precise.

4.5 Conclusion Innovation is an important criterion for academic paper acceptance. Identifying important innovation points or contributions in arguments (i.e., research problems), theories, evidence/data, methods, conclusions, and value forms the basis for judging whether academic paper content possesses innovation. Based on knowledge element research and academic paper content analysis, this paper constructed academic paper knowledge element ontologies reflecting four innovation dimensions, determined knowledge element extraction rules for these dimensions, used Word2Vec and Naive Bayes methods to classify theoretical and methodological innovation, and employed SVM models to construct knowledge element extraction rule bases. Based on the constructed academic paper knowledge element base, this paper proposed basic methods for intelligent evaluation of research problem, theoretical, methodological, and conclusion innovation, constructing an intelligent innovation evaluation process.

Finally, using papers published in *Library and Information Service* from 2015-2017 as the experimental database, this paper extracted knowledge elements according to extraction rules, performed machine learning classification on theoretical and methodological knowledge elements, making them self-weighted knowledge element categories. Further word vector training was conducted on extracted four-dimension knowledge elements to establish corpora. Using 2018, 2017, and 2016 papers as test data, innovation identification and judgment were performed, with final scoring results demonstrating feasibility and basically reflecting the innovation diffusion process (i.e., decreasing innovation).

Further analysis of scoring results revealed some issues in the scoring system, such as overly strict rule settings for theoretical and methodological dimensions resulting in zero scores for some papers' methodological innovation, requiring further methodological adjustments. The research conclusion innovation calculation also requires further adjustment, and conclusion metadata extraction rules need improvement for better results.

References

[1] Chen Jianqing. Reflections on the evaluation of academic paper innovation in China[J]. Youth Journalist, 2013(18): 33-35.

[2] Wang Wenyan. On the hierarchy of innovation[J]. Journal of Henan Normal University (Philosophy and Social Sciences Edition), 2006(1): 218-219.

[3] Lu Wanhui, Tan Zongying. Research on methods for measuring the novelty of academic achievement topics—Based on Doc2Vec and HMM algorithms[J]. Data Analysis and Knowledge Discovery, 2018(3): 22-30.

[4] Uzzi B, Mukherjee S, Stringer M, et al. Atypical combinations and scientific impact[J]. Science, 2013, 342(6157): 468-472.

[5] Li Chong, Su Yongjian. Academic evaluation: Reflections and transcendence of quantitative models[J]. Studies in Dialectics of Nature, 2017, 33(2): 59-63.

[6] Selvarajoo K. Measuring merit: Take the risk[J]. Science, 2015, 347: 139-140.

[7] Peng Lin, Du Xingye. Investigation on the implementation of open peer review in academic journals[J]. Chinese Journal of Scientific and Technical Periodicals, 2018, 29(11): 1114-1121.

[8] He Ying. Research on Peer Review Expert Selection[M]. Beijing: China Social Sciences Press, 2016.

[9] Vieira ES, Gomes JANF. The peer-review process: The most valued dimensions according to the researcher's scientific career[J]. Research Evaluation, 2018, 27(3): 246-261.

[10] Shen Yang. An innovation evaluation method based on keywords[J]. Information Studies: Theory & Application, 2007(1): 125-127.

[11] He Wanying. Research on Academic Paper Innovation Evaluation Based on Machine Learning[D]. Nanjing: Nanjing University, 2019.

[12] Suo Chuanjun. Research on academic paper aging and innovation from the perspective of knowledge transfer[J]. Library and Information Service, 2014, 58(5): 5-12.

[13] Yang Jing, Wang Fang, Bai Rujiang. A method for evaluating the innovation of individual academic papers based on research topic comparison[J]. Library and Information Service, 2018, 62(17): 75-83.

[14] Ruan Guangce, Xia Lei. Identification of hot topics in journal papers based on Doc2Vec[J]. Information Studies: Theory & Application, 2019, 42(4): 107-111, 106.

[15] Qiu Junping. Evaluation Science: Theory, Method, and Practice[M]. Beijing: Science Press, 2010.

[16] Qiu Junping, Zhao Yanjie, Luo Li. Research on paper classification methods in scientific evaluation[J]. Journal of the China Society for Scientific and Technical Information, 2011(5): 554-560.

[17] Xu Jianqiang, Cui Huijie, Li Xiaoping, et al. A new intelligent evaluation system for dissertations[J]. Computer Engineering, 2013, 39(7): 224-227, 232.

[18] Suo Chuanjun, Gai Shuangshuang. Research on the connotation, structure, and description model of knowledge elements[J]. Journal of Library Science in China, 2018, 44(4): 54-72.

[19] Wen Youkui, Xu Guohua, Lai Bonian, et al. Knowledge Element Mining[M]. Xi'an: Xidian University Press, 2005.

[20] Brookes BC. The developing cognitive viewpoint in information science[C]//International workshop on the cognitive viewpoint. Ghent: Ghent University, 1977: 195-203.

[21] Brookes BC. The foundations of information science part I. Philosophical aspects[J]. Journal of information science, 1980, 2(3/4): 125-133.

[22] Wen Youkui, Jiao Yuying. Knowledge Discovery Based on Knowledge Elements[M]. Xi'an: Xidian University Press, 2011.

[23] Wen Youkui, Xu Guohua. Knowledge element linking theory[J]. Journal of the China Society for Scientific and Technical Information, 2003, 22(6): 665-670.

[24] Wen Tingxiao, Hou Jingchuan, Gong Jiaoteng, et al. Construction of Chinese text knowledge elements and its practical significance[J]. Journal of Library Science in China, 2007, 33(6): 91-95.

[25] Yuan Mingyi, Xie Shenquan. Unified representation of knowledge based on knowledge element ontology[J]. Modern Computer (Professional Edition), 2008(5): 46-48, 57.

[26] Jiang Yongchang, Yang Hongyan, Zhang Libo. Research on knowledge organization based on knowledge elements and its system service functions[J]. Information Studies: Theory & Application, 2007(1): 37-40.

[27] Fang Long, Li Xin, Huang Yong, et al. Structural function identification of academic texts—Application in automatic keyword extraction[J]. Journal of the China Society for Scientific and Technical Information, 2017, 36(6): 599-605.

[28] Wang Zhongyi, Shen Xueying, Huang Jing. Chinese text hierarchical segmentation based on knowledge elements[J]. Library and Information Service, 2019, 63(7): 105-115.

[29] Du Hui, Xu Xueke, Wu Dayong, et al. Micro-blog sentiment classification based on sentiment word vectors[J]. Journal of Chinese Information Processing, 2017, 31(3): 170-176.

[30] Li Huifu, Lu Guang. Research on text classification methods based on multi-type classifier fusion[J]. Application Research of Computers, 2019, 36(3): 752-755.

Author Contributions

Li He: Responsible for topic selection, framework, revision, and finalization.
Du Xingye: Responsible for initial draft writing and revision.

Research on Intelligent Evaluation for the Content Innovation of Academic Papers

Li He[1], Du Xingye[1],[2],[3]

[1]School of Management, Jilin University, Changchun 130002
[2]National Science Library, Chinese Academy of Sciences, Beijing 100190
[3]Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract: [Purpose/Significance] Innovation is the key factor of academic paper evaluation. Based on knowledge element theory and machine learning theory and algorithm, this paper studies how to intelligently evaluate the innovation of academic papers from the content perspective. [Method/Process] Firstly, we constructed 4 knowledge element ontologies of academic papers including 'research problem ontology', 'theory ontology', 'method ontology' and 'conclusion ontology', and proposed the model of innovation evaluation. Secondly, we put forward the methods of machine classification models for theories and methods of academic papers and the extraction rules and methods of knowledge elements, and established rule bases and knowledge corpora. Finally, based on semantic similarity calculation methods, we scored the innovation of academic papers across four dimensions according to judgment rules and relevant weights. [Result/Conclusion] The feasibility of the intelligent evaluation method is verified by the experiment and could provide the references for the realization of intelligently evaluation of academic paper.

Keywords: Academic paper evaluation; Knowledge element; Content innovation; Intelligent evaluation