https://doi.org/10.51408/issi2025_104

Enhancing Scientometrics Prediction under Uncertainty: A DIKW-Based Framework and Methodological Synthesis

Shuya Chen[1], Guo Chen[2]

[1]996512152@qq.com,[2]delphi1987@qq.com

Nanjing University of Science and Technology, No. 200 Xiao Ling Wei, Nanjing, Jiangsu (China)

Abstract

This paper analyzes the uncertainties present in predictive-oriented scientometric research and, through a literature review, organizes and categorizes information analysis tasks related to prediction under uncertain conditions. Furthermore, to better adapt to these tasks, we approach the issue from the perspective of the DIKW model and summarize various methods for handling uncertainty. Finally, we propose a research framework for conducting predictive-oriented scientometric studies in uncertain environments, using scenario analysis and signal analysis to dealing with uncertainty.

Introduction

The advancement of information technology and the continuous progress of globalization have led to an exponential growth of open-source information. Its multi-source, complex, abundant, and uncertain nature has become the norm in modern information environment. Such an information environment has prompted profound changes in information analysis tasks, gradually shifting from targeted services with clear objectives to innovative and foresight-oriented information services in an environment filled with uncertainty (Zhao & Zeng, 2022). Alongside the increase in open-source information, uncertainties in scientometric research have also become more pronounced (Zhao, 2022). Although the widespread application of machine learning and artificial intelligence has driven the transformation of scientometrics toward a model-based analytical paradigm, such quantitative analysis methods ignore the uncertainties that are intrinsic to prediction problems. However, the incomplete and complex nature of information inevitably leads to uncertainties in predicting future trends. Existing scientometric analysis methods are unable to predict and evaluate the direction of future changes and their ramifications while ignoring the issue of uncertainty (Sun & Ke, 2007). Although the uncertainties inherent in prediction tasks, arising from various factors, cannot be entirely eliminated, scientific methods can be employed to significantly reduce uncertainties in scientometric analysis to the greatest extent possible (Wu et al., 2022).

Methods

This study employs literature review and content analysis methods, focusing on informetric tasks related to prediction under uncertain environments and approaches for handling uncertainty. The scope of the review primarily centers on the field of library and information science, encompassing various literature resources such as academic journal articles, professional books and government reports. During the retrieval process, multiple authoritative databases were utilized. Chinese literature was mainly sourced from CNKI (China National Knowledge Infrastructure), while English literature was obtained from databases such as Web of Science, ScienceDirect, and SpringerLink. Initially, a simple search query was constructed using core concepts and related terms, such as ("uncertainty environment" OR "uncertainty" OR "uncertainty handling") AND "prediction". Following preliminary research, the search terms were expanded to include synonyms. For example, "uncertainty environment" was expanded to include "complex environment", and "prediction" was extended to "foresight" and "early warning". Additionally, sentence-level searches were conducted to ensure that relevant literature lacking specific keywords was not omitted. Table 1 lists the keywords and filters used in the two rounds of retrieval. We systematically reviewed the research content, methods, and conclusions of the selected literature, analyzing their contributions to handling uncertainty. The screened literature was then categorized and summarized to provide a structured overview of the issue.

Table 1. Search Items and Filters.

Search Items		Filters
Initial Search Items	**Expanded Search Items**	
Uncertainty	Fuzzy, Rough, Deep uncertainty	• Library and Information Science •Journal Articles OR book OR Report OR Conference Paper • Citations > 0
Uncertainty environment	Complex Environment	
Prediction	Early Warning, Forecast	
Uncertainty Handling	Uncertainty Representation, Uncertainty Measurement	

Results

Uncertainty in Predictive-Oriented Scientometrics

Uncertainty is an inherent challenge in scientometrics. Throughout the entire analytical process, from data input and analysis to the generation of results, various forms of uncertainty are always present. The goal of informetric research is to leverage available information to reduce uncertainty in understanding and predicting phenomena, making prediction an essential information service. (Li & Sun, 2024). In predictive-oriented scientometric research, uncertainty manifests in three dimensions: information, process, and outcomes, as shown in Figure 1.

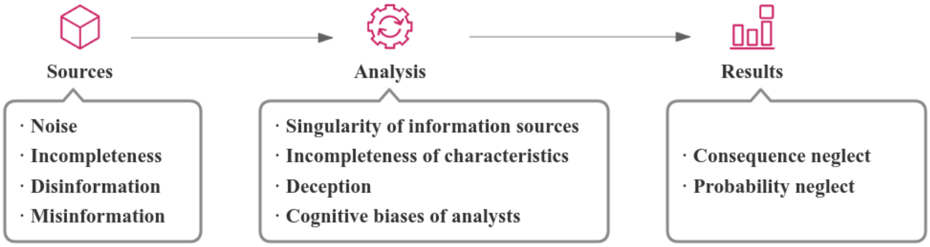


Figure 1. Uncertainty of Predictive-Oriented Scientometrics.

The Signal-Noise Theory suggests that an imbalanced signal-to-noise ratio is a significant source of uncertainty in intelligence analysis. Chinese scholar Wang Yanfei proposed the concept of "information fog", highlighting the falsehood and incompleteness of information. During the information analysis process, factors such as single-sided information sources, incomplete features, the deception, and cognitive biases of analysts can all contribute to uncertainty (Chen et al., 2022).The uncertainty issues in predictive intelligence outcomes include Consequence Neglect and Probability Neglect (Friedman & Zeckhauser, 2012). Consequence Neglect refers to overemphasizing the probabilities of various outcomes while neglecting their potential consequences. Probability Neglect refers to presenting possible outcomes without paying attention to the probabilities associated with each outcome.

Trend extrapolation is commonly used for prediction in informetric studies. To further illustrate why predictive-oriented scientometrics methods need to consider uncertainty, we use trend extrapolation as an example. Traditional data-driven scientometrics methods rely on quantitative analysis, aiming to identify the most likely patterns, thereby enhancing the certainty of a particular outcome and replacing all possibilities with the highest probability. For instance, as shown in Figure 2, curve l2 provides the best fit at time T1. However, at time T2, curve l3 becomes the best fit. During the transition from T1 to T2, changes in the phenomenon may be influenced by new factors or may no longer be affected by previously relevant factors.

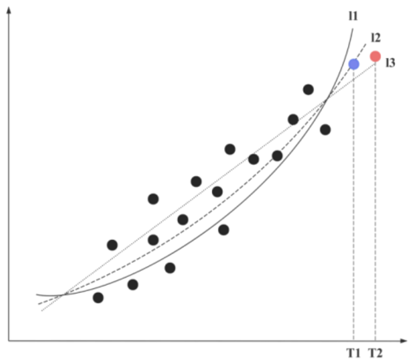


Figure 2. Future Trend Prediction Based on Historical Data.

Woodrow J. Kuhns (2003) argued that prediction methods should not simply produce results but also describe trends and the factors or variables influencing the development of a situation. Proposing future scenarios can lay a solid foundation for decision-making. Traditional trend extrapolation methods are limited by their overemphasis on identifying the best methods and outcomes. The issue lies in the fact that almost all scientometrics research is based on incomplete information, and the quantity and quality of information can both contribute to uncertainty in predictive-oriented scientometrics research (Mandel, 2020). Our goal is not to find a method that can completely eradicate uncertainty, but rather to develop one that can further minimize uncertainty in the analysis process and go beyond static and simplistic conclusions.

Predictive-Oriented Tasks under Uncertain Environments

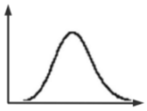
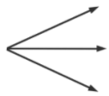
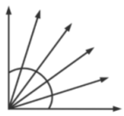
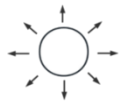
In the context of big data, the explosive growth of open-source information has profoundly influenced scientometric research. While deterministic information has become increasingly accessible, this very accessibility underscores the importance of analyzing uncertain information. Concurrently, the intricate and layered nature of information uncertainty has not only amplified the demand for predictive capabilities but also imposed more rigorous standards on predictive tasks. Based on a comprehensive review and synthesis of the literature, we have delineated seven key predictive-oriented scientometric tasks in complex environments. These tasks are systematically classified based on their role in the information chain within the DIKW (Data-Information-Knowledge-Wisdom) model, the degree of uncertainty they involve, and their strategic significance, as illustrated in Figure 3.

The DIKW model, rooted in Ackoff's classification of human cognitive content, distinguishes four levels: Data, Information, Knowledge, and Wisdom (Bosancic, 2016). Data consist of symbols representing the attributes of things or objects, which are inherently devoid of meaning. To convert data into information, it is essential to collect, organize, and process data relevant to the target problem, extract meaningful components, and contextualize them. Transitioning from fragmented information to systematic and theoretical knowledge requires extensive induction, analysis, and synthesis. Knowledge encompasses theories and patterns derived by individuals, while wisdom involves applying knowledge to solve problems and make decisions. In uncertain environments, rather than striving for optimal decisions, the emphasis shifts to flexible and nuanced decision-making that can adapt to multiple future scenarios.

The second dimension is strategic significance. From the perspective of information analysis, "strategy" focuses on the "information activities of the subject". Without a subject, there is no drive or uniqueness in competition and confrontation, because information activities of the subject are closely connected to its economic, social, and cultural background (Yang, 2022). The closer the analysis is to the tactical level, the finer the granularity of the problems, such as resource replenishment and information fusion. Conversely, the higher the strategic significance, the more macroscopic the problems to be considered, requiring the mobilization and coordination of resources across various domains, as seen in tasks like early warning.

The third dimension is the degree of information uncertainty. Marchau (2019) categorized uncertainty into four levels based on the nature of the problems, as shown in Table 2. Across all levels of uncertainty, resource replenishment remains a necessary task. Information fusion, however, is relevant except in scenarios with fully determined objectives and information. When multiple foreseeable futures exist but their likelihoods are uncertain, tasks such as technology foresight and intelligence assessment become essential. Level 4 uncertainty, the deepest level, can be divided into two scenarios: one where the future is constrained by many plausible possibilities (4a) and another where we only know that we do not know (4b). Tasks such as clue discovery, situational awareness, counterintelligence, and early warning aim at addressing this profound level of uncertainty, requiring not only the exploration of possibilities based on existing knowledge but also forward-looking expertise and insights.

Table 2. Levels of Information Uncertainty.

Degree of Uncertainty	*Level 0*	*Level 1*	*Level 2*	*Level 3*	*Level 4*	
Objectives	Complete Determinism	A clear enough future	Alternate futures with probabilities	A few plausible futures	Many plausible futures	Unknown future
						

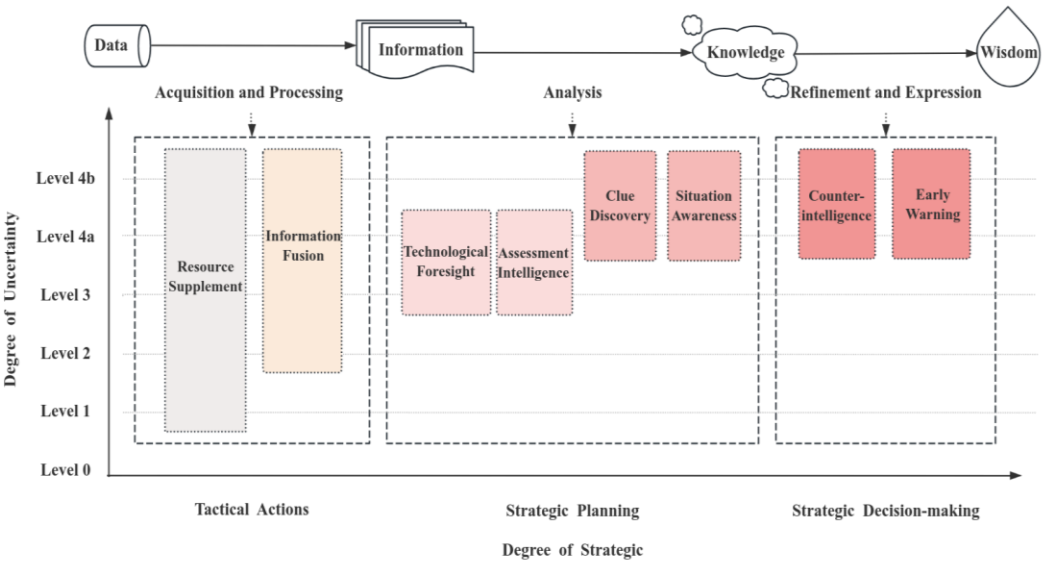


Figure 3. Predictive-Oriented Tasks and Their Classification in Uncertain Environments.

Scientometrics Methods for Addressing Uncertainty

Information uncertainty has become increasingly prominent in complex information environments, and the focus of information analysis has gradually shifted from descriptive intelligence to predictive, evaluative, and early-warning intelligence. These tasks in uncertain environments, as illustrated in Figure 3, often rely on traditional data-driven methods, which are ill-suited for tasks aimed at reducing future uncertainty. Predictive work in uncertain environments emphasizes the extensive collection of information, the handling of knowledge uncertainty, and the acknowledgment of multiple future possibilities to support decision-making.

In past scientometric research, numerous methods have been developed to address various uncertainties, including uncertainties in information sources, uncertainty relationship mining, and uncertainty representation. Building on the DIKW model, we categorize the collected methods into three types based on the uncertainties present at each stage of cognitive content transformation. As shown in Figure 4, uncertainties such as information representation exist during the transition from Data to Information. To better quantify uncertainty and represent uncertain information, we need to use foundational uncertainty representation methods to store and express information and knowledge in a more scientific form, making them applicable to complex uncertain scenarios. During the transition from Information to Knowledge, the core interest is identifying which "signals" or knowledge can help us predict future events or capture potentially unnoticed information in uncertain environments. Ultimately, during the transition from Knowledge to Wisdom, we need to synthesize, analyze, and organize knowledge at a macro level to generate insightful content to support decision-making across diverse issues. This method collection, constructed based on the characteristics of each stage of the DIKW model, covers various uncertainty issues that may arise throughout the entire knowledge chain, from raw data collection to the generation of high-value intelligence.

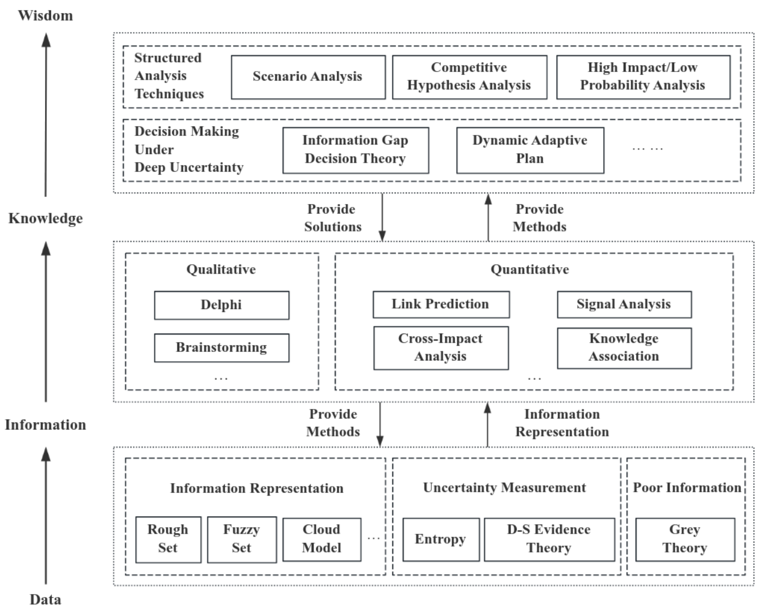


Figure 4. Method Framework for Scientometrics in Uncertain Scenarios.

Toward New Scientometric Approaches: A Case Study of Signal Analysis

In previous research, we have revealed the widespread presence of uncertainty in predictive tasks from a scientometric perspective and explored a collection of methods to address these uncertainties. Future research will further focus on optimizing the combination of methods at the practical level, proposing to integrate scenario analysis with weak signal analysis to construct a novel analytical framework for technology trend foresight, as shown in Figure 5.

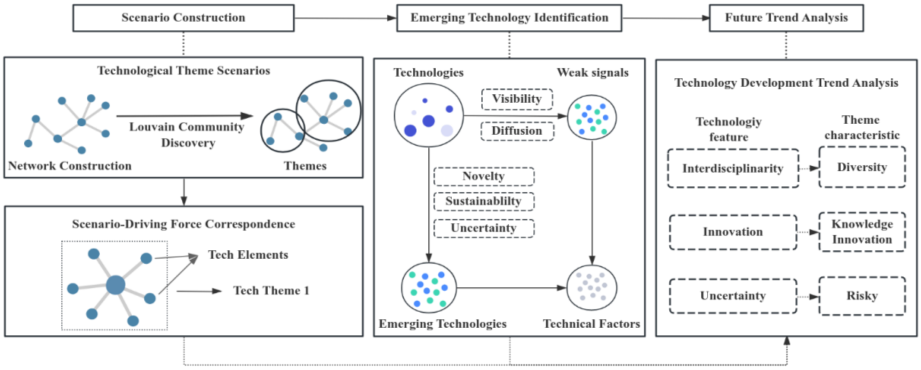


Figure 5. Practical Framework for Technology Foresight in Uncertain Environments.

Specifically, due to the inherently high uncertainty of technological evolution, especially in long-term trend foresight, where potential influencing factors are complex and intertwined, traditional single-metric methods struggle to comprehensively capture their dynamic characteristics. Thus, this study proposes to incorporate weak signal analysis by identifying and filtering technologies with potential influence as driving forces, combined with scientometric methods (such as network analysis and topic modeling) to quantitatively analyze the evolutionary paths of technological themes. To concretely reveal the trends of future scenarios, we set three core predictive objectives: diversity, innovation, and risk, each scenario characteristic, corresponding scientometric indicators will be established. Through driving force analysis and scientometrics analysis, scenario-based predictions will be made for trend foresight of each technological theme.

Acknowledgement

This work is supported by Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. KYCX24_0791).

References

Bosancic B. (2016). Information in the knowledge acquisition process. *Journal of Documentation,* 72(5), 930-60.

Chen Y., Ma X. J., et al. (2022). Studies on Analytical Method System of Warning Intelligence: From a Perspective of Probability Analysis. *Journal of Intelligence*, 41(02), 20-28+11.

Friedman J. A., (2012). Zeckhauser R. Assessing Uncertainty in Intelligence. *Intelligence and National Security*, 27(6), 824-847.

Kuhns W. J. (2003). *Paradoxes of Strategic Intelligence*. London: Routledge.

Li Y. & Sun J. J. (2024). Predictive Intelligence Research Based on Data Intelligence: A New Paradigm for Predictive Intelligence Research in a New Environment. *Journal of the China Society for Scientific and Technical Information*, 43(6), 633-643.

Mandel D.R. (2020). Assessment and Communication of Uncertainty in Intelligence to Support Decision Making: Final Report of Research Task Group SAS-114.

Marchau V. A. W. J., Walker W. E., et al. (Ed.). (2019). Decision Making under Deep Uncertainty: From Theory to Practice. Cham: Springer.

Sun J. J. & Ke Q. (2007). A Method for Information Analyses under Incomplete Information Environment: Scenarios Analysis and Its Application in Information Research. *Library and Information Service*, 51(02), 63-66+120.

Wu Y., Ma J., et al. (2022). Uncertainty in Intelligence Analysis: A Study from the Perspective of National Security Intelligence. *Information studies: Theory & Application*, 45(05), 66-74.

Yang G. L. (2022). Development of Intelligence Studies under the National Strategies. *Journal of the China Society for Scientific and Technical Information*, 41(07), 762-773.

Zhao Z. Y. & Zeng W. (2022). Constructing a Scientific and Technological Intelligence Theoretical System in a Complex Information Environment. *Journal of the China Society for Scientific and Technical Information*, 41(06), 549-557

Zhao Z. Y. (2022). Outstanding and Intelligent Empowerment of Scientific and Technological Intelligence in Complex Information Environments. Journal of the China Society for Scientific and Technical Information, 41(12), 1229-1237.