Topics and Methodological Trends in Data Fusion Research (Postprint)

Authors: Li Jie, Yu Qianqian, Wang Yuju

Date: 2023-11-07T10:52:02+00:00

Abstract

Data fusion constitutes an important pathway for realizing the value of multi-source data, and a comprehensive analysis of the overall thematic landscape of global data fusion research holds significant scientific and technological intelligence value for the current understanding and study of data fusion. [Methods/Process] Using term frequency and co-word analysis methods, we analyzed the hotspot topics and research methods of 16,053 data fusion research papers in the Web of Science Core Collection. [Results/Conclusion] Data fusion research has demonstrated a significant overall growth trend, and after more than 30 years of development, has established core research hotspots and data fusion methodologies. In the research, sensor data fusion (including wireless sensor data fusion) represents a hotspot direction in this field. Fault diagnosis, remote sensing, security, and smart grids constitute hotspot application scenarios for data fusion. Kalman filtering, neural networks, Dempster-Shafer evidence theory, and machine learning (including deep learning, support vector machines, etc.) are hotspot methods for data fusion, and a multi-method co-occurrence collaborative network has already formed in data fusion research.

Full Text

Trends of Topics and Methods in Data Fusion Research

Li Jie[1,2], Yu Qianqian[1], Wang Yuju[1]

[1]National Science Library, Chinese Academy of Sciences, Beijing 100190, China
[2]Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China

Abstract

[Purpose/Significance] Data fusion represents a critical pathway for realizing the value of multi-source data. A comprehensive analysis of the global landscape of data fusion research topics holds significant scientific and technological intelligence value for current understanding and investigation of data fusion. [**Method/Process**] This study employs word frequency and co-word analysis to examine hot topics and research methods across 16,053 data fusion papers from the Web of Science Core Collection. [**Result/Conclusion**] Data fusion research has demonstrated a marked growth trend overall, and over more than three decades of development has established core research hotspots and methodological frameworks. Sensor data fusion (including wireless sensor networks) constitutes the primary research direction in this field, with fault diagnosis, remote sensing, security, and smart grids emerging as key application scenarios. Kalman filtering, neural networks, Dempster-Shafer evidence theory, and machine learning (including deep learning and support vector machines) represent the dominant methods in data fusion research, with a co-occurring synergistic network of multiple methods already taking shape.

Keywords: Data fusion; Information fusion; Knowledge fusion; Multi-source data integration; Co-word analysis; VOSviewer

---

Introduction

The adage "listening to both sides makes one wise, while heeding only one side makes one benighted" reflects the long-standing limitations of relying on single data sources to reveal intelligence characteristics and objective patterns. Recognizing the critical application value of data fusion in integrating object intelligence, the U.S. military began focusing on and researching data fusion issues in the 1990s. In the 1980s, the U.S. Department of Defense Joint Directors of Laboratories (JDL) established the Data Fusion Working Group to systematically conduct research on data fusion. Based on military requirements, JDL defined data fusion as the process of associating, correlating, and combining data from multiple sensors and information sources to obtain accurate position and identity estimation, thereby enabling comprehensive and detailed assessment of situations, threats, and importance. This JDL concept and model of data fusion has played a significant role in subsequent research and applications.

Although discussions and research on data fusion have spanned over 30 years, the connotation and extension of data have evolved with changing times, making it necessary to re-examine data fusion issues in the new digital era. Currently, as digitalization, datafication, and intelligentization deepen further, integrating and fusing multi-source data for problem-solving has become an emerging hotspot in data-driven problem resolution within complex information environments. The data-information-knowledge-wisdom (DIKW) model in knowledge management demonstrates the data empowerment process from raw data to wisdom, wherein the fusion of data, information, and knowledge forms the foundation for achieving wisdom. Through literature investigation, we find that the boundaries among data fusion, information fusion, and knowledge fusion remain unclear in current academic research, with scholars often using these terms randomly based on personal or disciplinary preferences. Some scholars have compared and discussed these concepts; for instance, Professors Zhu Zhenyuan and Li Guangjian' s comparative analysis of data fusion, information fusion, and knowledge fusion revealed that research content in data fusion and information fusion is increasingly converging, with research boundaries not clearly demarcated. Knowledge fusion primarily focuses on the integration of literature and knowledge, specifically manifested in knowledge base construction, knowledge mapping, and ontology development, emphasizing semantic organization and expression of relationships between resources. However, from the perspective of knowledge fusion research, the object of fusion has merely changed (i.e., treating literature data and knowledge-related data as knowledge for fusion), with no substantial difference from data fusion and information fusion. We argue that, according to the DIKW model, data fusion represents the most fundamental level of fusion, serving as the basis for information fusion, knowledge fusion, and intelligence fusion. Theoretically, information fusion, knowledge fusion, and intelligence fusion constitute high-dimensional forms of data fusion; therefore, data fusion should encompass information fusion, knowledge fusion, and intelligence fusion.

Data fusion technologies and methods have permeated all aspects of scientific research. In astronomy, scientists use multi-source data and information fusion technologies to explore and map the universe (e.g., synthesizing the first black hole image). In automation, fusing multi-source sensor data (e.g., temperature, images, sound, vibration) enables intelligent control systems (e.g., autonomous driving). In scientific and technological intelligence, the fusion empowerment of various literature and intelligence sources has become essential for achieving a "holistic intelligence perspective" and "precise intelligence perspective." The comprehensive integration philosophy and decision-making approaches that incorporate expert knowledge from various fields have become crucial methods in complex problem decision-making processes. For a long time, scholars from different specialized fields have conducted relevant research and analysis on data fusion theories, technologies, and methods from the perspective of domain requirements. However, due to varying degrees of urgency, practical applications of data fusion differ across fields. In the scientific and technological intelligence domain, data fusion development has been relatively slow, primarily focusing on scientific papers integrated with patents, policies, social media, and other data types to study science-technology (S-T), science-policy (S-P), and science-media interactions from different perspectives. In medical technology information research, fusing various diagnostic information from medical records can effectively identify potential health risks, providing decision-making support for "precision medicine" and "preventive treatment." Beyond these mentioned scenarios, data fusion continuously influences human production, life, and survival, making comprehensive understanding of data fusion research profoundly significant.

Scholars both domestically and internationally have conducted surveys and review analyses on multi-source data fusion theories, models, methods, and applications. For example, Han Zengqi et al. provided a comprehensive review of information fusion technology, examining domestic and international development status and offering detailed coverage of definitions, hierarchical structures, classical methods, and applications, establishing a solid foundation for understanding research trends in information fusion. Yu Jiahui et al. analyzed domestic multi-source and multi-dimensional data fusion research trends from theoretical, methodological, and application perspectives using CNKI data. Castanedo and Alofi et al. successively provided comprehensive reviews and summaries of data fusion technologies, offering panoramic materials for understanding data fusion techniques and processes. Additionally, relevant scholars have systematically reviewed and discussed data fusion models, methods, and applications. Analysis of previous data fusion reviews reveals that sensor data constitutes the core domain of data fusion, where a relatively complete set of data fusion processes, models, and technical methods has been developed. Building upon previous research, this study utilizes the international Web of Science database to comprehensively collect data fusion literature, employing keyword frequency and co-word analysis to investigate global research trends and core data fusion methods, aiming to provide scientific and technological intelligence support for data-driven decision-making in China' s complex information environment.

1.1 Data Sources

This study' s data originates from the Web of Science Core Collection, selecting SCI/SSCI journal articles and CPCI conference papers. Based on preliminary investigation results and given the lack of clear demarcation among data fusion, information fusion, and knowledge fusion, this paper uses "data fusion" as an umbrella term for all three concepts. In data retrieval, "fusion,""integration,"and "aggregation" were used as English equivalents for "融合." The search timeframe was set to 1900–2021 to capture all data fusion literature up to 2021. The final search query was: (TI=("data fusion" *OR " data integrat*" OR "data aggregat" *OR " information fusion*" OR "information integrat" *OR " information aggregat*" OR "knowledge fusion" *OR " knowledge integrat*" OR "knowledge aggregat* ") AND FPY=1900-2021) AND LA=(English). This search yielded 16,053 data fusion research papers published between 1966 and 2021, with the output trend shown in Figure 1 [Figure 1: see original paper].

From the perspective of stage characteristics in paper output, data fusion development can be divided into four periods: embryonic period, slow development period, rapid development period (I), and rapid development period (II). From 1966 to 1989, research attention on data fusion was very low, related to limited demand for data fusion and low levels of digital development. During this stage, overall data processing scale was small, and analysis of single data sources remained in the era of small data, with minimal demand for multi-source data fusion. From 1990 to 1998, paper output on data fusion became more active than in the previous stage, but overall output remained relatively low, even experiencing a slight decline in 1999. From 1999 to 2009, data fusion paper output entered rapid development period (I), showing significant growth trends and indicating increased demand and attention for data fusion. After a slight decline in 2010, output began rapid growth again, entering rapid development period (II). From 2010 to 2021, technologies for data collection, processing, and analysis such as big data, artificial intelligence, and the Internet of Things developed rapidly, leading to unprecedented increases in data fusion technologies and practical application demands, and ushering in a new growth period for data fusion research output.

1.2 Research Methods

To explore the thematic characteristics of data fusion research, this study analyzes data fusion topics from two dimensions: word frequency and co-word analysis. Word frequency statistical analysis involves extracting and counting each keyword' s occurrences from the keyword fields (DE) of the collected data fusion papers, using frequency to measure keyword popularity. In scientific literature analysis, high-frequency keywords typically characterize research hotspots within a dataset. However, keyword frequency analysis only presents research content from a single dimension, lacking revelation of semantic relationships between keywords. Therefore, based on word frequency analysis, co-word analysis further examines the data fusion research topic network.

Co-word analysis can be traced back to the 1980s, represented by the publication of *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World* by Callon et al. from the French National Center for Scientific Research. With increasing widespread use, its basic analytical processes and patterns have been established. First, meaningful terms must be identified through text mining or extraction techniques. Then, under user-defined "association or co-occurrence" rules, a relationship matrix between terms is constructed. Finally, statistical or mathematical processing of the generated co-word matrix interprets thematic characteristics of the research field. Early co-word matrices were often processed using SPSS to complete two-dimensional mapping and hierarchical clustering of word pairs. In recent years, rapid development of database technologies and knowledge graph tools has made co-word analysis more convenient. For example, emerging software such as CiteSpace, VOSviewer, and SCIMAT, particularly VOSviewer, which integrates term mapping and clustering technologies, has greatly improved co-word analysis efficiency and effectiveness. Therefore, this study selects keywords from data fusion papers as analysis objects and uses the VOSviewer scientific knowledge mapping tool to analyze keyword frequency and co-word networks.

2.1 Overall Landscape of Hot Topics in Data Fusion Research

Keywords directly represent paper research content and constitute important elements for analyzing scientific research hotspots through literature data. This study extracted 418 keywords with frequency $ $10 from 16,053 papers to construct a data fusion topic network, shown in Figure 2 [Figure 2: see original paper] (displaying TOP 500 connections). Keywords underwent cleaning and disambiguation, deleting meaningless keywords and merging synonymous variants. To understand usage preferences for different forms of data fusion, synonymous keywords representing data fusion in the search query were not merged during analysis.

Word frequency analysis of data fusion keywords reveals that, beyond search terms, hot keywords with frequency $ $100 include: wireless sensor networks (872), Kalman filter (228), multi-sensor data fusion (213), sensor fusion (209), neural network (205), D-S evidence theory (201), internet of things (195), ontology (195), machine learning (180), fault diagnosis (174), sensor networks (164), deep learning (152), support vector machine (148), remote sensing (144), clustering (135), sensors (135), multi-sensor data fusion (127), multi-sensor (124), classification (116), big data (114), security (112), and smart grid (102). These high-frequency keywords indicate that sensor data fusion represents the current research hotspot (involving wireless sensor networks, multi-sensor data fusion, sensor fusion, sensor networks, sensors, and multi-sensor). This trend inevitably follows developments in automation, artificial intelligence, and the Internet of Things. Wireless sensor networks (WSN) integrate sensors, microelectromechanical systems, and network technologies, representing a novel information acquisition and processing technology with important applications in military, environmental science, medical detection, and space exploration. WSN features significant data redundancy, ensuring robust operation, convenient and rapid information transmission and sharing, and enhanced emergency response capabilities and intelligence timeliness. Hotspot keywords for data fusion application scenarios primarily include fault diagnosis, remote sensing, big data, security, and smart grids. The temporal distribution of high-frequency keywords is shown in Figure 2(b), where warmer node colors indicate more recent keyword activity. Analysis results demonstrate that both application scenarios and employed technologies and methods in data fusion research evolve with the times. Currently, the ten most active keywords in the data fusion keyword cluster are covid-19, task analysis, computational modeling, knowledge graph, transfer learning, cameras, blockchain, convolutional neural network, sentinel-2, and deep learning.

Cluster analysis of topics from the co-word network perspective is presented in Table 1 . The analysis reveals that although Chinese translates "fusion," "integration," and "aggregation" uniformly as "融合," clustering results show significant usage preferences across different scenarios. Current keyword clustering primarily uses search terms as representative words, dividing research into #1 data fusion, #2 data integration, and #3 data aggregation. Clusters #1 and #3 both focus on sensor data fusion, with sensor data fusion research predominantly using "data fusion" and "data aggregation" to represent data fusion. However, differences exist between #1 and #3: cluster #3 data aggregation keywords concentrate on WSN data fusion research directions, while #1 data fusion focuses on multi-source sensor data fusion. Cluster #2 data integration keywords differ significantly from #1 and #3, representing network or literature data fusion.

2.2 Hot Methods and Trends in Data Fusion

Analysis and summarization of data fusion methods provide important guidance for practical data fusion research. In previous studies, Jitendra systematically summarized mathematical methods for sensor data fusion as: probabilistic data fusion methods; fuzzy logic and possibility theory-based data fusion; filtering, target tracking, and kinematic data fusion; decentralized data fusion systems; component analysis and data fusion; and image algebra data fusion. In China, review and scientometric literature on data fusion and information fusion has also summarized data fusion analysis methods at different levels. Building upon previous research and the overall keyword network, this section further extracts the TOP 30 data fusion methods, shown in Figure 3 [Figure 3: see original paper]. The ten most applied methods in data fusion research are Kalman filter (228), neural network (205), D-S evidence theory (201), machine learning (180), deep learning (152), support vector machine (148), clustering (135, upper-level method), classification (116, upper-level method), big data (114, upper-level method), data mining (90, upper-level method), and fuzzy logic (90). Examining average publication years for method applications reveals that fuzzy logic, wavelet analysis, neural networks, and genetic algorithms were among the earliest data fusion analysis methods. From early applications, Kalman filter, evidence theory, and support vector machine have high usage frequencies, establishing them as hotspot methods in data fusion. Recently, against the backdrop of big data and artificial intelligence, emerging data fusion methods include convolutional neural networks, deep learning, fog computing, and machine learning. Through co-occurrence network analysis of methods, relationships involving hierarchical and synergistic patterns were extracted. Figure 4 [Figure 4: see original paper] presents the data fusion method synergy network, where connections between two methods indicate simultaneous usage, with wider connections representing higher co-occurrence frequencies. Primary method relationship pairs include: deep learning-convolutional neural network, neural network-fuzzy logic, neural network-D-S evidence theory, deep learning-machine learning, deep learning-feature extraction, neural network-genetic algorithm, support vector machine-classification, machine learning-neural network, and support vector machine-feature extraction.

Based on the above analysis, several widely applied and clearly defined data fusion methods are introduced:

(1) Kalman Filter

In 1960, Kalman published a paper using recursive methods to solve discrete data linear filtering problems, later known as the Kalman filter. The fundamental idea employs minimum mean square error as the optimal estimation criterion, utilizing signal and noise state space models to update state variable estimates using previous estimates and current observations to obtain current estimates. The algorithm provides minimum mean square error estimates for signals based on established system and observation equations. Kalman filter characteristics make it highly suitable for complex multi-sensor estimation and data fusion problems, leading to extended Kalman filter and unscented Kalman filter methods for data fusion. Currently, Kalman filter has been widely applied in data fusion research and practice. For example, Sun et al. proposed a new matrix-weighted multi-sensor optimal information fusion criterion under linear minimum variance, presenting a general multi-sensor optimal information fusion decentralized Kalman filter with a two-layer fusion structure. Smyth et al. proposed a multi-rate Kalman filtering method to address data fusion problems in dynamic system measurement of displacement and acceleration responses. Engineering data fusion applications span aerospace, autonomous driving, and unmanned systems.

(2) D-S Evidence Theory

In 1967, Harvard mathematician Dempster developed the theory' s prototype using upper and lower probabilities to solve multi-valued mapping problems. Building upon this foundation, Shafer further refined and developed the theory in 1976 by introducing belief functions and establishing a mathematical method for handling uncertainty based on "evidence" and "combination," publishing the seminal work *A Mathematical Theory of Evidence*. This theory for handling uncertainty has found widespread applications in information fusion, expert systems, and pattern recognition. The essential process of D-S evidence theory involves mathematical fusion of evidence, representing an important method and direction in current data fusion research. Practical applications include robot data fusion, diagnostic information fusion, reliability data fusion, and fire detection. Additionally, research has addressed method corrections and multi-method fusion.

(3) Neural Network, Machine Learning, Deep Learning, and Support Vector Machine

In 1943, psychologist McCulloch and mathematician Pitts proposed neural network concepts and models, pioneering neural network research. The fundamental idea simulates human brain information processing, dividing information processing into input, hidden, and output layers. Neural networks have achieved significant development since their inception, encompassing BP neural networks, RBF neural networks, and recently developed convolutional neural networks. Neural network methods have a long history in data fusion research and analysis, with the earliest application in this dataset tracing back to 1989 when Whittington et al. used neural networks to analyze tactical and sensor data fusion problems in modern naval environments. Neural network-based analysis methods have been widely applied in data fusion for crack detection, multi-sensor data fusion, target tracking, fault diagnosis, and remote sensing.

Recently, machine learning and deep learning methods have been increasingly applied in data fusion research, representing new growth points for data fusion methods and technologies. Machine learning enables computers to learn decision-making methods from prior datasets, thereby gaining predictive capabilities for similar data, constituting an important research area in artificial intelligence. Machine learning can be categorized into supervised, unsupervised, and semi-supervised learning. Deep learning represents a branch of machine learning development aimed at enabling machines to possess analytical capabilities like humans. Machine learning enables automated data clustering, classification, or prediction mechanisms, making it highly suitable for feature analysis and extraction from multi-source data and driving data fusion under learning rules. Currently, machine learning-based data fusion methods have been extensively applied across various domains, including corrosion testing, smart city security, and urban big data fusion. Within the machine learning family, support vector machine, proposed by Cortes and Vapnik in 1995, is a supervised learning method for binary classification that offers simple structure and strong generalization capabilities, providing an effective method for small-sample, nonlinear, high-dimensional data fusion. Support vector machine has been applied in traffic information fusion, target tracking, and remote sensing data fusion.

(4) Fuzzy Logic

In 1965, Zadeh published the paper "Fuzzy Sets," marking the birth of fuzzy mathematics. In data fusion, expressions include fuzzy control, fuzzy inference, fuzzy sets, and fuzzy theory. Fuzzy logic powerfully promotes existing data fusion methods, significantly improving data fusion feasibility. For example, fuzzy logic can generate data fusion rule weights to facilitate data fusion operations. Currently, fuzzy logic has become a core method in data fusion analysis, extensively applied in scenario-specific data fusion tasks. As early as 1994, Abdulghafour et al. utilized fuzzy logic to develop fusion formulas based on fuzzy measures, applying fuzzy theory to data fusion research. Fuzzy logic has also been applied in data fusion for target recognition, vehicular networks, and mobile robot navigation.

(5) Wavelet Analysis

Wavelet analysis, also known as wavelet transform, was first proposed in 1974 by French engineer J. Morlet working in petroleum signal processing. Wavelet transform extends Fourier transform methods, providing a time-frequency domain representation. As a branch of mathematics, this method has been extensively applied in signal analysis, image processing, and numerical analysis. Particularly against the backdrop of growing image data demands, wavelet analysis has played important roles in image processing and image fusion. For example, research has used wavelet transform coefficients to perform image fusion analysis and processing on blurred aircraft images, converting them into clear photographs. Wavelet transform methods have also been applied in remote sensing data fusion, multi-sensor data fusion, and Internet of Things data fusion.

(6) Bayesian Networks

Bayesian networks, also known as belief networks or directed acyclic graphical models, are probabilistic graphical models. As probability-based data fusion methods, early research and applications focused primarily on reliability combination, representing trust/belief degrees through prior credibility, conditional PDFs (probability density functions), and posterior probabilities, then performing fusion based on Bayesian rules to infer overall reliability. In data fusion practice based on Bayesian theory, Z. Chair published a paper in 1988 discussing distributed Bayesian and distributed data fusion issues. Subsequently, this method has been widely applied in data fusion for sensor data, pipeline leak detection, and EEG/MEG and fMRI data fusion.

(7) Semantic Web

The Semantic Web, also known as Web 3.0, was proposed by World Wide Web founder Tim Berners-Lee in 1998. The core concept involves adding computer-understandable semantics to web documents, transforming the entire internet into a universal information exchange medium. RDF (Resource Description Framework) and OWL (Web Ontology Language) constitute core technologies in the Semantic Web. The Semantic Web essentially enables interconnection and integrated use of internet data through unified standards, representing the core theory of internet data fusion. Currently, the Semantic Web is widely applied in knowledge fusion, including enterprise knowledge fusion, drilling risk management knowledge fusion, and traditional Chinese medicine-biomedicine knowledge fusion. From a technical characteristics perspective, the Semantic Web represents the core technology for scientific literature data integration and fusion using networks as carriers, serving as the key to multi-source literature data convergence and fusion empowerment.

3 Summary and Discussion

Data fusion represents an important pathway for realizing multi-source data value in the current big data era and constitutes a key focus in scientific and technological literature data research within complex information environments. Data fusion adopts a holistic thinking approach to achieve systematic revelation of specific tasks and scenarios through multi-source data integration. This study's analysis of global data fusion research topics and methods may offer valuable reference for understanding current data fusion research.

(1) Academic research on data fusion has a long history, but due to influences from technological development stages and practical demands, data fusion has not received sufficient attention and widespread application. The years 1999 and 2010 represent two turning points in data fusion research. From 1999 to 2009, global data fusion research demonstrated unprecedented research enthusiasm, with paper output showing significant growth trends. After a brief decline in 2010, overall output resumed significant growth. Post-2010 data fusion research has re-emerged against the backdrop of big data, artificial intelligence, and Internet of Things technologies. Meanwhile, data fusion has become one of the core tasks in data-driven problem-solving.

(2) From the overall thematic landscape of data fusion research, distinct thematic communities have formed according to search terms. In current scientific research, sensor data fusion represents the primary research direction, with wireless sensor network data fusion independently forming a substantial thematic cluster. In terms of research maturity, sensor data fusion has the longest history and is most mature in data fusion. Military demands for sensor data fusion promoted data fusion development and accumulated theoretical and technical foundations. After civilian adoption, multi-source sensor data fusion technology greatly advanced industrial and social production and governance. Under new era conditions, rapid network technology development enables fast multi-source data transmission and fusion processing. Currently, data fusion technologies and methods are widely applied in fault diagnosis, security, and smart grids. Against the backdrop of rapid intelligentization development, data fusion will play increasingly important roles in the digital intelligence process.

(3) Data fusion integrates multi-source and multi-type data, exhibiting significant task complexity. In practical data fusion work, a relatively mature data fusion methodology system has been developed for different data fusion requirements and data characteristics. Among all data fusion methods, Kalman filtering is most representative. Additionally, D-S evidence theory, clustering/classification, and support vector machines are widely applied across different data fusion scenarios. In recent data fusion research, methods based on deep learning, machine learning, and convolutional neural networks have emerged. Through continuous development, data fusion research methods have progressed toward multi-method fusion, forming a "method network" for data fusion research.

(4) Although this study did not distinguish among data fusion, information fusion, and knowledge fusion in data retrieval, analysis results show usage preferences across application scenarios. "Fusion" and "aggregation" are most frequently used in sensor data fusion, with "fusion" leaning toward multi-sensor data fusion and "aggregation" leaning toward wireless sensor network data fusion. "Integration" scenarios lean toward other data fusion types beyond sensor data, such as network data and textual knowledge data. The linguistic meanings of these three terms may fundamentally explain their application scenarios, though this paper' s primary task precludes further elaboration.

Furthermore, regarding specific data fusion scenarios, sensor data fusion research is relatively mature and has formed a large research scale. Therefore, current scientific literature multi-source data fusion research can draw upon its fusion theories, methods, and technologies to some extent to develop a unique theoretical and technical system for scientific literature data fusion. Against the backdrop of data as a key production factor, data must and will inevitably move toward fusion to maximize value.

References

[1] Hall D L, Llinas J. An introduction to multisensor data fusion[J]. Proceedings of the IEEE, 1997, 85(1): 6-23.
[2] Jitendra R. Raol. Data Fusion Mathematics: Theory and Practice[M]. Translated by Wang Gang, He Zhenghong, Wang Rui, et al. Beijing: National Defense Industry Press, 2021.
[3] Zhu Zhenyuan, Li Guangjian. A preliminary exploration of knowledge fusion from a holistic "data-information-knowledge" perspective: The correlation and comparison of data fusion, information fusion, and knowledge fusion[J]. Information Studies: Theory & Application, 2017, 40(2): 12-18.
[4] Yu Jingyuan, Zhou Xiaoji. From comprehensive integration thought to comprehensive integration practice: Method, theory, technology, and engineering[J]. Chinese Journal of Management, 2005(1): 4-10.
[5] Han Zengqi, Yu Junjie, Li Ningxia, et al. A review of information fusion technology[J]. Journal of Intelligence, 2010, 29(S1): 110-114.
[6] Yu Jiahui, Liu Jiajing, Zheng Jianming. Research trends in multi-source and multi-dimensional data fusion: Theory, method, and application[J]. Journal of Intelligence, 2022, 41(5): 133-138, 207.
[7] Castanedo F. A review of data fusion techniques[J]. The Scientific World Journal, 2013, 2013: 704504.
[8] Alofi A, Alghamdi A A, Alahmadi R F, et al. A review of data fusion techniques[J]. International Journal of Computer Applications, 2017, 167: 37-41.
[9] Esteban J, Starr A, Willetts R, et al. A review of data fusion models and architectures: Towards engineering guidelines[J]. Neural Computing & Applications, 2005, 14(4): 273-281.
[10] Lau B P L, Marakkalage S H, Zhou Y, et al. A survey of data fusion in smart city applications[J]. Information Fusion, 2019, 52: 357-374.
[11] Callon M, Rip A, Law J. Mapping the dynamics of science and technology: Sociology of science in the real world[M]. Springer, 1986.
[12] Li Jie. Principles and Applications of Scientific Knowledge Mapping[M]. Beijing: Higher Education Press, 2018.
[13] Van Eck N J, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping[J]. Scientometrics, 2010, 84(2): 523-538.
[14] Ren Fengyuan, Huang Haining, Lin Chuang. Wireless sensor networks[J]. Journal of Software, 2003(7): 1282-1291.
[15] Khaleghi B, Khamis A, Karray F O, et al. Multisensor data fusion: A review of the state-of-the-art[J]. Information Fusion, 2013, 14(1): 28-44.
[16] Kalman R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960, 82(1): 35-45.
[17] Peng Dingcong. Basic principles and applications of Kalman filtering[J]. Software Guide, 2009, 8(11): 32-34.
[18] Roumeliotis S I, Bekey G A. An Extended Kalman Filter for frequent local and infrequent global sensor data fusion[C]// Proceedings of the Conference on Sensor Fusion and Decentralized Control in Autonomous Robotic Systems, Oct

14-15, 1997, Pittsburgh, Pa. Bellingham: SPIE - Int Soc Optical Engineering, 1997: 11-22.
[19] Cao J J, Fang J C. Fuzzy adaptive unscented Kalman filter for MIMU/MMC/GPS data fusion[C]// Proceedings of the 7th International Conference on Electronic Measurement and Instruments, Aug 16-18, 2005, Beijing, P R China. Hong Kong: International Academic Publishers Ltd, 2005: 3.380-3.386.
[20] Sun S L, Deng Z L. Multi-sensor optimal information fusion Kalman filter[J]. Automatica, 2004, 40(6): 1017-1023.
[21] Smyth A, Wu M. Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring[J]. Mechanical Systems and Signal Processing, 2007, 21(2): 706-723.
[22] Dempster A P. Upper and lower probabilities induced by a multivalued mapping[J]. The Annals of Mathematical Statistics, 1967, 38(2): 325-339.
[23] Shafer G. A mathematical theory of evidence[M]. Princeton University Press, 1976.
[24] Sun Y S, Zheng C L, Ma P. D-S evidence theory and its application in robot information fusion[C]// Proceedings of the International Conference on Information Science, Automation and Material System, May 21-22, 2011, Zhengzhou, P R China. Stafa-Zurich: Trans Tech Publications Ltd, 2011.
[25] Wu S Q, Jiang W L. Research on data fusion fault diagnosis method based on D-S evidence theory[C]// Proceedings of the International Conference on Measuring Technology and Mechatronics Automation, Apr 11-12, 2009, Zhangjiajie, P R China. Los Alamitos: IEEE Computer Soc, 2009.
[26] Liang L Q, Cai Q, Shen Y J, et al. A reliability data fusion method based on improved D-S evidence theory[C]// Proceedings of the 11th International Conference on Reliability, Maintainability and Safety (ICRMS) - Integrating Big Data, Improving Reliability & Serving Personalization, Oct 26-28, 2016, Hangzhou, P R China. New York: IEEE, 2016.
[27] Cai Z S, Chen M S. Application of data fusion technology based on D-S evidence theory in fire detection[C]// Proceedings of the 6th International Conference on Electronics and Information Engineering (ICEIE), Sep 26-27, 2015, Dalian, P R China. Bellingham: Spie-Int Soc Optical Engineering, 2015.
[28] Pan G, Wu L L. Information fusion based on improved D-S evidence theory[C]// Proceedings of the 2nd International Conference on Information Technology and Management Innovation (ICITMI 2013), Jul 23-24, 2013, Zhuhai, P R China. Stafa-Zurich: Trans Tech Publications Ltd, 2013.
[29] Sun R, Huang H Z, Miao Q. Improved information fusion approach based on D-S evidence theory[J]. Journal of Mechanical Science and Technology, 2008, 22(12): 2417-2425.
[30] Zhou Y M, Xu H J, Sun J F, et al. Multisensor data fusion based on modified D-S evidence theory[C]// Proceedings of the International Conference on Computer Modeling, Simulation and Algorithm (CMSA), Apr 22-23, 2018, Beijing, P R China. Paris: Atlantis Press, 2018.
[31] Chen B, Wang J F, Chen S B. Prediction of pulsed GTAW penetration status based on BP neural network and D-S evidence theory information

fusion[J]. International Journal of Advanced Manufacturing Technology, 2010, 48(1-4): 83-94.
[32] Ding H, Hou R C, Ding X Q. A data fusion equipment monitoring method based on fuzzy set and improved D-S evidence theory[C]// Proceedings of the 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Jul 29-31, 2017, Guilin, P R China. New York: IEEE, 2017.
[33] Mcculloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The Bulletin of Mathematical Biophysics, 1943, 5(4): 115-133.
[34] Whittington G, Spracklen C T. The application of neural networks to tactical and sensor data fusion problems[C]// Proceedings of the 1989 First IEEE International Conference on Artificial Neural Networks, (Conf Publ No 313), 16-18 Oct. 1989.
[35] Chen F, Jahanshahi M R. NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion[J]. IEEE Transactions on Industrial Electronics, 2018, 65(5): 4392-4400.
[36] Kolanowski K, Świetlicka A, Kapela R, et al. Multisensor data fusion using Elman neural networks[J]. Applied Mathematics and Computation, 2018, 319: 236-244.
[37] Chin L. Application of neural networks in target tracking data fusion[J]. IEEE Transactions on Aerospace and Electronic Systems, 1994, 30(1): 281-287.
[38] Jing L, Wang T, Zhao M, et al. An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox[J]. Sensors, 2017, 17(2): 414.
[39] Cheng L, Wang L, Feng R, et al. Remote sensing and social sensing data fusion for fine-resolution population mapping with a multi-model neural network[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 5973-5987.
[40] Birgersson M, Hansson G, Franke U. Data integration using machine learning[C]// Proceedings of the 20th IEEE International Enterprise Distributed Object Computing Conference (EDOC), Sep 05-09, 2016, Univ Vienna, Fac Comp Sci, Vienna, Austria. New York: IEEE, 2016.
[41] Dong X L, Rekatsinas T. Data integration and machine learning: A natural synergy[J]. Proceedings of the VLDB Endowment, 2018, 11(12): 2094-2097.
[42] Waltz E. Machine learning (ML) support to information fusion[C]// Proceedings of the Conference on Signal Processing, Sensor/Information Fusion, and Target Recognition XXVIII, Apr 15-17, 2019, Baltimore, MD. Bellingham: Spie-Int Soc Optical Engineering, 2019.
[43] Voelker C, Kruschwitz S, Ebell G. A machine learning-based data fusion approach for improved corrosion testing[J]. Surveys in Geophysics, 2020, 41(3): 531-548.
[44] Abid A, Abbas A, Khelifi A, et al. An architectural framework for information integration using machine learning approaches for smart city security profiling[J]. International Journal of Distributed Sensor Networks, 2020, 16(10): 16.

[45] Liu J, Li T R, Xie P, et al. Urban big data fusion based on deep learning: An overview[J]. Information Fusion, 2020, 53: 123-133.
[46] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
[47] Luo Y, Wang Y Z, Sun M. A data fusion approach based on parallel support vector machine[C]// Proceedings of the 1st IITA International Joint Conference on Artificial Intelligence, Apr 25-May 26, 2009, Hainan Isl, P R China. Los Alamitos: IEEE Computer Soc, 2009.
[48] Liu H H, Wang X Y, Tan D R, et al. Study on traffic information fusion algorithm based on support vector machines[C]// Proceedings of the 6th International Conference on Intelligent Systems Design and Applications (ISDA 2006), Oct 16-18, 2006, Jinan Univ, Jinan, P R China. Los Alamitos: IEEE Computer Soc, 2006.
[49] Vasuhi S, Vaidehi V, Midhunkrishna P R, et al. Multiple target tracking using support vector machine and data fusion[C]// Proceedings of the 3rd International Conference on Advanced Computing (ICoAC), Dec 14-16, 2011, Anna Univ, Dept Comp Technol, Chennai, India. New York: IEEE, 2011.
[50] Zhao S H, IEEE. Remote sensing data fusion using support vector machine[C]// Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Sep 20-24, 2004, Anchorage, AK. New York: IEEE, 2004.
[51] Zadeh L A. Fuzzy sets[J]. Information and Control, 1965, 8(3): 338-353.
[52] Abdulghafour M, Fellah A, Abidi M A, et al. Fuzzy logic-based data integration: Theory and applications[C]// Proceedings of the 1994 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Oct 02-05, 1994, Las Vegas, Nv. New York: IEEE, 1994.
[53] Chen B H, Lin C F, Thomopoulos S C A. Fuzzy logic information fusion for object recognition[C]// Proceedings of the Conference on Applications of Fuzzy Logic Technology II, Apr 19-21, 1995, Orlando, Fl. Bellingham: Spie - Int Soc Optical Engineering, 1995.
[54] Tal I, Muntean G M. Using fuzzy logic for data aggregation in vehicular networks[C]// Proceedings of the 16th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications, Oct 25-27, 2012, Dublin, Ireland. New York: IEEE, 2012.
[55] Parasuraman S, Shirinzadeh B. Fuzzy logic based sensors data fusion for mobile robot navigation[C]// Proceedings of the International Conference on Materials, Mechatronics and Automation (ICMMA 2011), Jan 15-16, 2011, Melbourne, Australia. Durnten-Zurich: Trans Tech Publications Ltd, 2011.
[56] Hariharan G. Wavelet analysis—An overview[M]. Wavelet Solutions for Reaction–Diffusion Problems in Science and Engineering. Singapore: Springer Singapore, 2019: 15-31.
[57] Ranchin T. Wavelets for modeling and data fusion in remote sensing[C]// Proceedings of the Conference of the NATO-Advanced-Study-Institute on Multisensor Data Fusion, Jun 25-Jul 07, 2000, Pitochry, Scotland. Dordrecht: Springer, 2000.
[58] Xu L J, Zhang J Q, Yan Y, et al. A wavelet-based multi-sensor data fusion algorithm[C]// Proceedings of the 20th IEEE Instrumentation and

Measurement Technology Conference, May 20-22, 2003, Vail, Co. New York: IEEE, 2003.
[59] Guidi B, De Salve A, Ricci L. A data aggregation strategy based on Wavelet for the Internet of Things[C]// Proceedings of the 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Sep 21-24, 2017, Timisoara, Romania. New York: IEEE, 2017.
[60] Vechet S, Krejsa J, Ondrousek V, et al. Sensors data fusion via bayesian filter[C]// Proceedings of the 14th International Power Electronics and Motion Control Conference (EPE-PEMC), Sep 06-08, 2010, Ohrid, Macedonia. New York: IEEE, 2010.
[61] Guerriero M, Wheeler F, Koste G, et al. Bayesian data fusion for pipeline leak detection[C]// Proceedings of the 19th International Conference on Information Fusion (FUSION), Jul 05-08, 2016, Heidelberg, Germany. New York: IEEE, 2016.
[62] Trujillo-Barreto N J, Martinez-Montes E, Valdes-Sosa P A, et al. Bayesian model for EEG/MEG and fMRI data fusion[J]. NeuroImage, 2001, 13(6): S270.
[63] Gu W D, Xia G P, You W J. Enterprise knowledge integration by semantic web[C]// Proceedings of the International Conference on Research and Practical Issues of Enterprise Information Systems, Apr 24-26, 2006, Vienna, Austria. New York: Springer, 2006.
[64] Xu Y Z. Research of knowledge integration based on semantic web for drilling risk management[C]// Proceedings of the 3rd International Conference on Information Computing and Applications (ICICA 2012), Sep 14-16, 2012, Chengde, P R China. Berlin: Springer-Verlag Berlin, 2012.
[65] Yu T, Liu J, Yang S, et al. Semantic web for knowledge integration between traditional Chinese medicine and biomedicine[C]// Proceedings of the 7th International Conference on Information Technology in Medicine and Education (ITME), Nov 13-15, 2015, Anhui, P R China. New York: IEEE, 2015.

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