Identifying scientific problems and solutions: Semantic network analytics and deep learning

Lu Huang[1], Xiaoli Cao[2,3,*], Hang Ren[1], Chunze Zhang[1,4] and Zhenxin Wu[2,3]

[1] *School of Management and Economics, Beijing Institute of Technology, Beijing, China, 100081*
[2] *National Science Library, Chinese Academy of Sciences, Beijing, China, 100190*
[3] *Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China, 100190*
[4] *Zhejiang Sineva Intelligent Technology Co., Ltd, Zhejiang, China, 314499*

Abstract

As critical building blocks of scientific research, scientific problems and solutions are put forward to reveal the existing issues and primary methods in scientific and technological practice. In this paper, we proposed a novel method for identifying scientific problems and solutions using semantic network analytics and deep learning. Firstly, the BERT-CRF model constructed is combined with BIO tagging to identify four entity types: research object, problem, solution, and fundamental principle. Then, the Levenshtein algorithm is applied to align entities, and a knowledge network is constructed integrating semantic information and co-occurrence associations, comprehensively and accurately depicting the relations between entities. Finally, the correlations between the four entity types are thoroughly explored using semantic network analytics and topological structure analytics. A case study on artificial intelligence domain demonstrates the reliability of the proposed methodology, and the results provide intelligent support for raising and solving scientific problems in the field.

Keywords

Scientific problems and solutions, Semantic network analytics, BERT-CRF, Knowledge network, Entity identification

1. Introduction

The rapid increase in scientific articles lays a strong foundation for identifying problems and solutions in a field [1]. The intelligent mining of scientific problems and solutions aims to identify the real-world issues existing in the scientific and technological practices of a field, find corresponding solutions, and explore the underlying theoretical foundations. It facilitates a deep exploration of the intrinsic logical relationships among research objects, problems, solutions, and fundamental principles. Identifying scientific problems and solutions can help scholars map the scientific field, enhance the speed of information retrieval and processing, and offer reference solutions for real-world issues in industrial practices [2,3].

Some scholars have mentioned that problems and corresponding solutions constitute the "key insights" within scientific articles [4]. Many significant studies focus on extracting key viewpoints (e.g., research problems, and solutions) from scientific papers using entity extraction techniques [5,6]. However, these methods usually involve supervised learning on pre-annotated datasets, a process that requires significant resources for domain-specific

[*] Corresponding Author

Joint Workshop of the 5th Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2024) and the 4th AI + Informetrics (AII2024), April 22-26, 2024, Changchun, China and Online

EMAIL: huanglu628@163.com (Lu Huang); cxl163990307@163.com (Xiaoli Cao); renhang0988@163.com (Hang Ren); zhangchunze@sineva.com.cn (Chunze Zhang); wuzx@mail.las.ac.cn (Zhenxin Wu)

CEUR Workshop Proceedings (CEUR-WS.org)

annotation [7]. Deep learning is an efficient and accurate technology for extracting information from complex unstructured data (e.g. graphics, text) and converting data into vector representations [8,9]. Combining deep learning with bibliometrics is often used to address problems in science, technology, and innovation (ST&I) management [10,11]. As an important area of deep learning, text representation learning effectively extracts information from text data and has been widely applied in data mining [12].

Additionally, some scholars employ methods such as keyword network analysis and citation analysis to construct academic knowledge graphs, deeply exploring the relationships among knowledge entities in papers [13,14]. For example, Zhang et al. [15] integrate multiple relationships such as co-occurrence, citation, and co-authorship to explore the processes of knowledge creation, knowledge transfer, and other knowledge evolution dynamics. However, these researches ignore the specific semantic functions of keywords in different contexts, leading to a lack of accuracy in the representation of knowledge structures [16]. Semantic network analysis incorporates the rich semantic information of keywords into network analysis, providing a more intensive and accurate analysis for the mining of scientific problems and solutions [17,18].

To address these concerns, we propose a novel framework for identifying problems and solutions using semantic network analytics and deep learning. The proposed method advances the fields of entity extraction and knowledge graph analysis by delineating three specific functions: 1) the BERT-CRF model is constructed to generate textual representations with enhancing semantics, and it is combined with BIO tagging to identify four entity types: research object, problem, solution, and fundamental principle, improving the accuracy of identifying entities; 2) the Levenshtein algorithm is applied to align entities, and the semantic relations and co-occurrence associations are integrated to construct knowledge network, comprehensively and accurately revealing the relationships between entities; 3) the combination of semantic network analytics and topological structure analytics are applied to thoroughly explore the correlations between the four entity types. We use a case study on artificial intelligence domain to demonstrate the reliability of our proposed method.

2. Method

The framework of identifying scientific problems and solutions is shown in Figure 1.

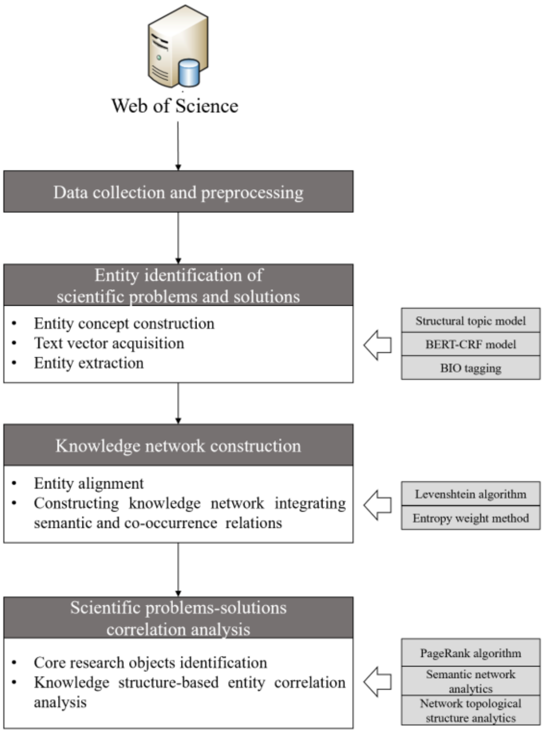
Web of Science
Data collection and preprocessing
Entity identification of scientific problems and solutions
• Entity concept construction
• Text vector acquisition
• Entity extraction
Structural topic model
BERT-CRF model
BIO tagging
Knowledge network construction
• Entity alignment
• Constructing knowledge network integrating semantic and co-occurrence relations
Levenshtein algorithm
Entropy weight method
Scientific problems-solutions correlation analysis
• Core research objects identification
• Knowledge structure-based entity correlation analysis
PageRank algorithm
Semantic network analytics
Network topological structure analytics

Figure 1: Framework of identifying scientific problems and solutions

2.1. Entity identification of scientific problems and solutions

2.1.1. Entity concept construction based on Structural Topic Model

The paper data gathered is acquired from the Web of Science (WoS) and pre-processed via VantagePoint (VP) [19].

Then, four abstract entity concepts are constructed based on the concept of Structural Topic Model (STM), which includes "research object", "problem", "solution" and "fundamental principle". These entities serve as a structured representation of knowledge, characterizing scientific problems and solutions. The STM, an advancement over the Latent Dirichlet Allocation (LDA) topic model, extracts topics from document-level metadata and establishes latent connections between these topics and the document data [20]. This approach facilitates the discovery of hidden knowledge structures within texts and the accurate delineation of implicit relationships among them [21]. Therefore, this study employs the STM to construct entity concepts.

Within this knowledge structure, "research object" refers to the research subfields, serving as the starting point of the research; "problem" focuses on the scientific issues to be resolved and the goals to be achieved, jointly defining the problems space with the research object; "solution" describes the overall solution to the problem, representing the key steps towards achieving the goals; "fundamental principle" refers to the theoretical foundation underlying the solution methods. The four-entity concept constructed provides a comprehensive analytical framework reflecting the essence of research literature. By capturing and mining these four key entities, this study can excavate the scientific problems and solutions within the paper data, unveiling research hotspots in the domain.

Finally, the four types of entities in a small number of literatures are manually identified and a pre-trained dataset is generated based on the BIO tagging [22].

2.1.2. Text vector acquisition based on BERT-CRF model

This section aims to construct an enhanced semantic BERT-CRF model, transforming scientific texts into feature vector matrices. The Bidirectional Encoder Representation from Transformers (BERT), a deep learning technology based on the bidirectional transformer architecture [23], can capture contextual semantic information and latent relationships from large-scale corpora, achieving more precise textual semantic representations [24]. BERT can process vast amounts of textual corpus data with a need for minimal training datasets. Combined with the Conditional Random Field (CRF) model, it can effectively improve the efficiency and quality of text sequence labelling [25].

First, the BERT model is trained on each of the four types of entities by using a pre-training dataset and setting the model parameters. To acquire enhanced vector representations that include contextual positional information, this paper integrates the CRF model [26] with BERT based on the embedding during the training process, further training and optimizing the configuration of feature function. Moreover, the multi-head self-attention mechanism of BERT is applied to better capture contextual semantic information [27], obtaining enhanced textual semantic representation vectors. When the model converges, the well-trained BERT-CRF model is generated.

Then, the well-trained model is used to transform the dataset into vector representations with enhanced contextual positional information and multiple semantic information.

2.1.3. Entity extraction

The purpose of this section is to extract scientific problem and solution entities by transforming textual semantic vectors into probabilistic representations of text sequence labeling using BERT-CRF model and BIO tagging.

First, the well-trained BERT-CRF model is applied to process the text vectors using the SoftMax function [28], generating predicted labels corresponding to the text sequence. Assuming the sentence length is n, and the input text sequence is represented as $X = (x_1, x_2, \cdots, x_n)$, the corresponding predicted label sequence is represented as $Y = (y_1, y_2, \cdots, y_n)$. The method for calculating the final prediction score $score(X, Y)$ for the text sequence X is:

$$score(X,Y) = \sum_{i=1}^{n} P_{x_i,y_i} + \sum_{i=2}^{n} T_{y_{i-1},y_i} \quad (1)$$

where P_{x_i,y_i} represents the probability of the text sequence element x_i being predicted as y_i, and T_{y_{i-1},y_i} is the score for the transition from label y_{i-1} to label y_i.

Thus, the probability distribution matrix corresponding to the text sequences is obtained.

Then, this paper integrates BIO tagging [22] into the CRF layer to generate four sequence labeling matrices of "research object", "problem", "solution" and "fundamental principle" based on the probability distribution matrix. Finally, the entity categories corresponding to the text sequences are identified based on the sequence annotation results.

2.2. Knowledge network construction

After identifying the four types of entities corresponding to scientific problems and solutions, a knowledge network containing multiple semantic and structural information between entities is constructed. This part includes two sections: 1) Entity alignment based on Levenshtein algorithm and 2) Constructing knowledge network integrating multiple relations.

2.2.1. Entity alignment based on Levenshtein algorithm

Considering that entities extracted from different literature may have multiple names for the same entity, we apply the Levenshtein method for measuring the difference between two sequences, to disambiguate. Levenshtein algorithm can consider both the contextual information and semantic similarity, enhancing the accuracy of entity alignment [29].

Furthermore, this paper constructs an entity dictionary based on expert knowledge, which is used for further checking and proofreading of entity alignment results.

2.2.2. Constructing knowledge network integrating semantic and co-occurrence relations

The purpose of this section is to construct a heterogeneous knowledge network including four types of entities, integrating semantic and co-occurrence information among entities, and improving the accuracy of relationship identification between entities.

First, the cosine distance between entity vectors is used to measure the semantic similarity between entities. The calculation method of the semantic similarity $sim(a,b)$ between entity a and b is:

$$sim(a,b) = \frac{U(a)^T U(b)}{||U(a)||_2 \cdot ||U(b)||_2} \quad (2)$$

where $U(a)$ and $U(b)$ denote the textual vectors of entities a and b respectively.

Then, the co-occurrence relation between entities is obtained based on the literature data. The co-occurrence association between entities a and b is denoted as $co\_entity(a,b)$, represented by the number of co-occurrences between a and b.

Finally, the entropy weight method [30] is introduced to integrate the semantic information and co-occurrence information between entities and get the relation between entities in the knowledge network. Link weight $weight(a,b)$ between entities a and b can be calculated as:

$$weight(a,b) = \alpha * sim(a,b) + \beta * co\_entity(a,b) \quad (3)$$

where α and β are coefficients of semantic similarity and co-occurrence correlation obtained by entropy weight method respectively, and α+β=1.

In this section, we generate a knowledge network $G = (V, E, W)$ containing rich semantic and structural information among entities, where V, E, and W denote the entities, edges, and edge weights in G respectively.

2.3. Scientific problems-solutions correlation analysis

This part aims to identify the primary research problems corresponding to the core research objects and find the main solutions and theoretical basis based on the topological structure analysis of knowledge networks.

2.3.1. Core research objects identification based on PageRank algorithm

In this section PageRank algorithm is used to measure the importance score of research objects and thus identify core research objects in the knowledge network. This algorithm fully considers multiple factors including the local topological structure and semantic information of the target node and the importance of the nodes connected with it [31], which has been widely applied to identify core nodes in various complex knowledge networks [32]. Therefore, we use PageRank algorithm to rank the importance of research objects in knowledge network. The calculation method of the importance score $PR(a)$ of research object a can be calculated as:

$$PR(a) = d \times \sum_{j=1}^{n} \frac{PR(T_j)}{C(T_j)} + (1 - d) \qquad (4)$$

where d is the damping factor ($0 \leq d \leq 1$), generally 0.85, T_j denotes the entity linked to the research object a, $C(T_j)$ is the number of entities linked with T_j, and n is the number of entities linked with research object a.

Finally, we sort the research objects in the knowledge network based on the importance score, and select the top-K research objects as the core research objects. The core research objects set O is represented as:

$$O = \{O_1, \cdots, O_i, \cdots, O_K\} \qquad (5)$$

where O_i denotes the *i*-th core research object in the knowledge network, and *K* is the number of core research objects that has been identified.

2.3.2. Knowledge structure-based entity correlation analysis

After identifying the core research objects within the knowledge network, this section will deeply analyze the correlation between entities in the domain based on the topological structure analysis of the knowledge network.

First, based on the link weights between the core research object O_i and the research questions, the primary problems corresponding to O_i are identified. The primary problems set *P* is represented as:

$$P = \{P_1, \cdots, P_i, \cdots, P_m\} \qquad (6)$$

where P_i denotes the *i*-th primary problem corresponding of O_i, and m is the number of primary problems.

Similarly, this paper identifies the main solutions corresponding to the primary problem P_i, and the main fundamental principle for the corresponding solutions.

Finally, we generate a series of complete chains of scientific problems and solutions, which can be represented as "research object - problem - solution - fundamental principle".

3. Case study

Artificial Intelligence (AI) is a multidisciplinary domain, composed of a diverse and heterogeneous network of innovations. It has emerged as a significant force driving technological innovation [33]. This field encompasses many emerging research questions and research methods, offering extensive data support for empirical analysis. Therefore, this paper analyzed the scientific problems and solutions in-depth in the AI domain to verify the effectiveness of the proposed method.

3.1. Entity identification based on BERT-CRF model

Following the study of Liu et al. [34], a total of 375608 papers published between 2021 to 2023 were retrieved from the Web of Science (WoS). Then, VantagePoint (VP) was used to process titles and abstracts of papers. Finally, a total of 310456 papers were retained as the textual corpus, and a total of 3000 papers were randomly selected in proportion to the publication year as the pre-training dataset. Based on the BIO tagging, the titles and abstracts of the pre-training dataset were annotated with "research object", "problem", "solution", and "fundamental principle".

Following the design in Section 2.1.2, this study constructed an enhanced semantic BERT-CRF model based on the textual dataset to transform text data into feature vectors. During the experimental process, the performance of the model was assessed based on evaluation metrics (Precision, Recall, and F1-score) [35], with model parameters being continuously adjusted. The optimal model was determined when the evaluation metrics reached their maximum values. Finally, when the Precision of the model reaches 89.2%, the Recall rate reaches 87.4%, and the F1-score reaches 88.3%, the optimal model was generated. The results show that the trained BERT-CRF model exhibits better performance.

Finally, the trained BERT-CRF model and BIO tagging method were used to identify four types of entities. We identified 24,254 "research object" entities, 23,839 "problem" entities, 20,670 "solution" entities, and 17,550 "fundamental principle" entities from the dataset.

3.2. Constructing knowledge network in AI

After identifying the four types of entities corresponding to scientific problems and solutions from the text dataset, this paper comprehensively considered the semantic similarity between entities and expert knowledge to achieve entity synonym alignment. The Levenshtein algorithm is employed for aligning entities within the same category and across different categories. Finally, a total of 887 "research object" entities, 4,136 "problem" entities, 13,858 "solution" entities, and 5,518 "fundamental principle" entities were obtained.

Then, we integrated semantic similarity and structural similarity between entities to build a knowledge network in AI domain. The statistical information on the edges between each type of entity in the knowledge network is shown in Table 1.

Table 1

Descriptive statistics of edges in knowledge network

	Research object	Problem	Solution	Fundamental principle
Research object	/	12683	9105	8468
Problem	12683	/	17691	10565
Solution	9105	17691	/	13783
Fundamental principle	8468	10565	13783	/

3.3. Entity correlation analysis

The next stage was to analyze the correlation between entities. Following Section 2.3.1, the PageRank algorithm was applied to calculate the important scores of research objects and thus identify the core research objects in the knowledge network. The hot research topics in the artificial intelligence domain can be explored according to the core research objects.

According to the results of the PageRank algorithm, the hot research objects in artificial intelligence domain mainly include deep learning, neural network, medical image, facial image, robot, and electric system. Following the design in Section 2.3.2, the top-2 problems were identified corresponding to the core research objects within the knowledge network.

Finally, we explored the correlations between entities and generated a series of complete chains including four types of entities. The partial entity correlation results of Top-6 core research objects are shown in Figure 2.

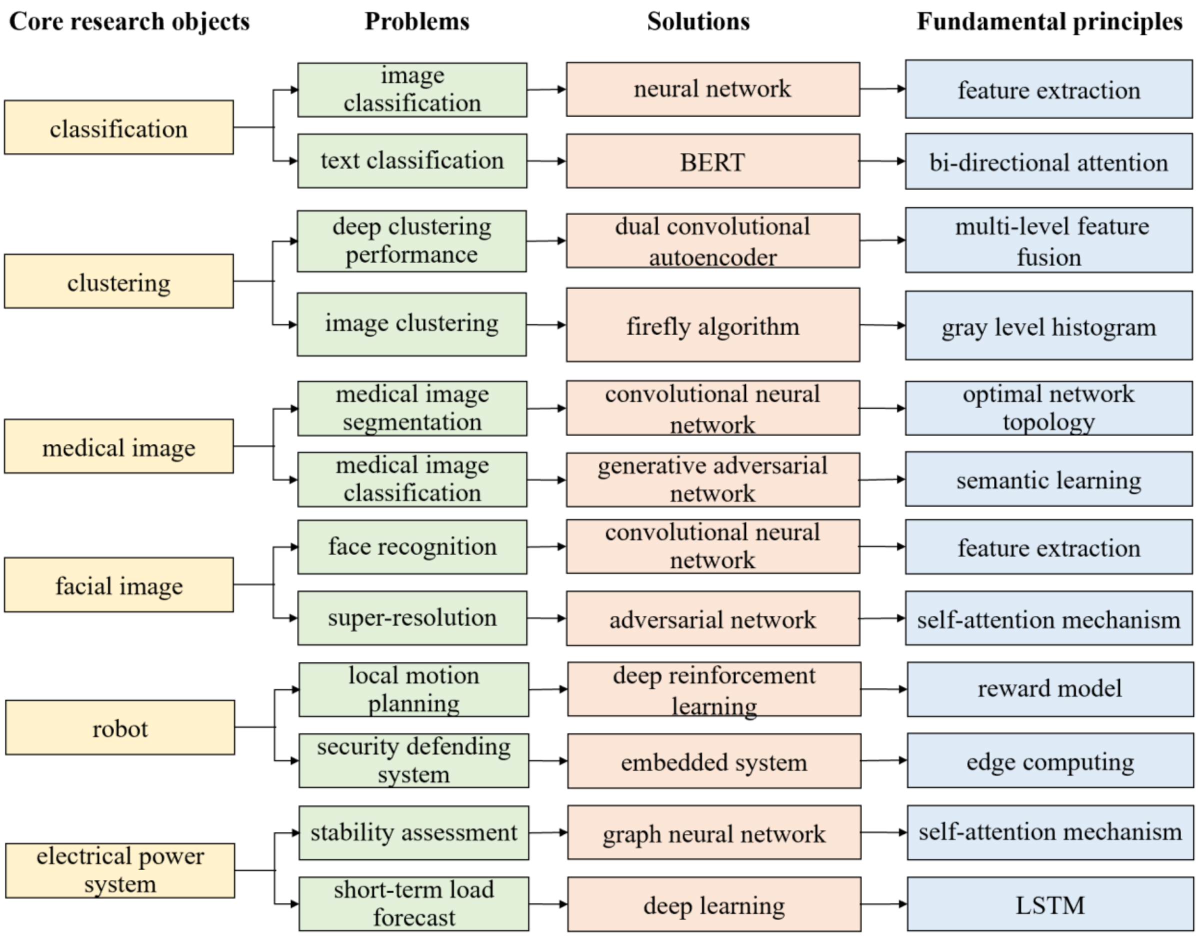

Figure 2: Entity correlation results

Several observations can be acquired based on the above results. The research object represents a subfield, where the problems refer to the issues contained within that subfield, the solution refers to the methods or technologies required to solve the problems, and the fundamental principles refer to the inherent principles involved in the implementation process of the methods and technologies. The "research object" and "problem" together constitute the complete scientific problem, and the "solution" and "fundamental principle" together constitute the complete solution. For example, for the identified "classification - image classification - neural network – feature extraction", it refers that the neural network can be used to solve image classification problems through feature extraction [36].

This paper identifies a complete chain of "research object - problem - solution - fundamental principle". On the one hand, it can identify the core research objects and corresponding primary problems in the field of artificial intelligence. On the other hand, it is able to detect the corresponding solutions to the real problems in the scientific and technological practice and explore the theoretical basis behind them, and thus realize the in-depth excavation of the intrinsic logical connection among scientific problems, solutions and fundamental principles.

3.4. Validation

In this part, we conducted the quantitative and qualitative methods to verify the reliability of our proposed method and entity identification results.

3.4.1. Verification of the trained model

To quantitatively verify the advantages of the combination of BERT-CRF model trained in this paper and the BIO tagging method, we select three advanced models, ALBERT [37], SciBERT [38], and XLNet [39], for comparison experiments. Referencing the model parameters of BERT-CRF in this paper, the three models were fine-tuned respectively to achieve the optimal effect. The specific parameter settings of the models are shown in Table 2.

Table 2
Parameter configurations of models

	Our method	ALBERT	SciBERT	XLNet
Maximum input length	64	64	64	64
Training epoch	30	40	30	35
Batch size	4	16	4	8
Number of layers	12	12	12	12
Learning rate	1e-5	5e-6	1e-5	1e-6
CRF learning rate multiplier	100	/	50	/

Then, the performance of our method was validated based on Recall and Precision by comparing with three state-of-the-art methods. The comparison results are given in Table 3.

Table 3
The comparison of prediction performance

Methods	Research objects		Problems		Solutions		Fundamental principles	
	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
Our method	**0.980**	**0.965**	**0.964**	**0.942**	**0.856**	**0.877**	**0.724**	**0.759**
ALBERT	0.934	0.936	0.924	0.896	0.848	0.815	0.638	0.702
SciBERT	0.928	0.955	0.906	0.883	0.834	0.827	0.704	0.740
XLNet	0.962	0.919	0.944	0.907	0.838	0.859	0.680	0.741

It can be seen that our method outperforms baseline methods in two evaluation indicators. Concretely, in the entity recognition of the research objects, the Recall value of our method increases by 4.6%, 5.2%, and 1.8%, respectively, and the Precision value increases by 2.9%, 1.0%, and 4.6%, respectively. In the problems identification, the Recall value of our method increases by 4.0%, 5.8%, and 2.0%, respectively, and the Precision value increases by 4.6%, 5.9%, and 3.5%, respectively. In the solutions identification, the Recall value of our method increases by 0.8%, 2.2%, and 1.8%, respectively, and the Precision value increases by 6.2%, 5.0%, and 1.8%, respectively. In the entity identification of fundamental principles, the Recall value of our method increases by 8.6%, 2.0%, and 4.4%, respectively, and the Precision value increases by 5.7%, 1.9%, and 1.8%, respectively. These results demonstrate the combination of BERT-CRF model and BIO tagging used in this paper has achieved good performance on our dataset.

3.4.2. Verification of entity identification

In this section, the qualitative method was applied to verify the reliability of the entity identification results by searching relevant articles published in 2021 and beyond. Table 4 shows the detailed empirical evidence of partial entity identification results.

Table 4

Relevant documentary proof of partial entity identification results

No	Research object - problem - solution - fundamental principle	Relevant documentary proof
1	classification - image classification - neural network – feature extraction	In 2021, Nadendla et al. proposed a neural network-based classifier by feature extraction and classification to solve the image classification problem [36].
2	clustering - deep clustering performance - dual convolutional autoencoder - multi-level feature fusion	In 2022, Hou et al. used a dual convolutional autoencoder to extract features of multi-levels and fuse them to improve the performance of deep clustering [40].
3	medical image - medical image segmentation - convolutional neural network - optimal network topology	In 2024, Tamilmani et al. used the convolutional neural network with the optimal network topology to solve the problem of medical image segmentation [41].
4	facial image - face recognition - convolutional neural network - feature extraction	In 2022, Wei-Jie et al. applied the convolutional neural network by extracting the masked face features to solve the problem of masked facial recognition [42].
5	robot - local motion planning - deep reinforcement learning - reward model	In 2023, Garrote et al. proposed a deep reinforcement learning strategy based on a reward model to solve the local motion planning problem of robots [43].
6	electrical power system - stability assessment - graph neural network - self-attention mechanism	In 2022, Gu et al. used the graph neural network based on the self-attention mechanism to evaluate the stability of the power system [44].

Table 4 demonstrates the alignment between our entity identification results and the literature. Therefore, the four types of entities identified and the relations between entities in this paper are reliable, and the effectiveness of the proposed method has been further verified.

4. Conclusion

In this paper, we proposed a novel methodology to identify scientific problems and solutions using semantic network analytics and deep learning. First, the deep learning method is applied to extract textual semantic information and identify entities, capturing the hidden semantic association in different textual contexts effectively, and improving the accuracy of the entity recognition. Then, the machine learning method was used to construct the knowledge network, fully considering the knowledge structure and semantic structure between entities, and thus containing more abundant information. Finally, the PageRank algorithm and semantic network analytics were introduced to deeply explore the linkages among research objects, problems, solutions, and fundamental principles.

Semantic network analytics and deep learning were combined to identify scientific problems and solutions from scientific text, which provides technical intelligence for field scientific innovation and industrial technology upgrading. In addition, this method can not only explore the association between entities, reveal the primary research problems and corresponding solutions in the field of artificial intelligence, but also discover the knowledge structure in this field and promote the development of scientific knowledge network analysis methods.

Several limitations of our proposed method require further improvement: 1) The scientific and technological output of a certain field includes not only papers, but also patents and product data. Further research should be conducted based on more data sources; 2) The methodology of entity alignment can be further optimized. More advanced methods and professional expert knowledge could be introduced in the future to improve the efficiency and quality of entity alignment; 3) The evolution mechanism of "research object - problem - solution - fundamental principle" needs to be further explored.

5. Acknowledgements

This work was supported by the National Nature Science Foundation of China Funds (Grant No. 72274013), and Fundamental Research Funds for the Central Universities.

6. References

[1] Y. Zhang, M. Wang, M. Saberi, et al, From big scholarly data to solution-oriented knowledge repository. Frontiers in Big Data, 2: 38, 2019.

[2] P. Li, W. Lu, Q. Cheng, Generating a related work section for scientific papers: an optimized approach with adopting problem and method information. Scientometrics, 127(8): 4397-4417, 2022.

[3] Z. Luo, W. Lu, J. He, et al, Combination of research questions and methods: A new measurement of scientific novelty. Journal of Informetrics, 16(2): 101282, 2022.

[4] G. Chen, J. Peng, T. Xu, et al, Extracting entity relations for “problem-solving” knowledge graph of scientific domains using word analogy. Aslib Journal of Information Management, 75(3): 481-499, 2023.

[5] V. Giordano, G. Puccetti, F. Chiarello, et al, Unveiling the inventive process from patents by extracting problems, solutions and advantages with natural language processing. Expert Systems with Applications, 229: 120499, 2023.

[6] R. B. Mishra, H. Jiang, Classification of problem and solution strings in scientific texts: evaluation of the effectiveness of machine learning classifiers and deep neural networks. Applied Sciences, 11(21): 9997, 2021.

[7] H. Liu, T. Brailsford, J. Goulding, et al, Towards idea mining: problem-solution phrase extraction from text. International Conference on Advanced Data Mining and Applications (pp. 3-14). 2022.

[8] Y. Zhang, J. Lu, F. Liu, et al, Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. Journal of Informetrics, 12(4): 1099–1117, 2018.

[9] R. Xiang, E. Chersoni, Q. Lu, et al, Lexical data augmentation for sentiment analysis. Journal of the Association for Information Science and Technology, 72(11): 1432-1447, 2021.

[10] X. Chen, P. Ye, L. Huang, et al, Exploring science-technology linkages: A deep learning-empowered solution. Information Processing & Management, 60(2): 103255, 2023.

[11] J. Chen, Y. Chen, Y. He, et al, A classified feature representation three-way decision model for sentiment analysis. Applied Intelligence, 52(7): 7995–8007, 2022.

[12] X. Xi, F. Ren, L. Yu, et al, Detecting the technology's evolutionary pathway using HiDS-trait-driven tech mining strategy. Technological Forecasting and Social Change, 195: 122777, 2023.

[13] J. Wang, Q. Cheng, W. Lu, et al, A term function–aware keyword citation network method for science mapping analysis. Information Processing & Management, 60(4): 103405, 2023.

[14] G. Garechana, R. Río-Belver, E. Zarrabeitia, et al, TeknoAssistant: a domain specific tech mining approach for technical problem-solving support. Scientometrics, 127(9): 5459-5473, 2022.

[15] X. Zhang, Q. Xie, C. Song, et al, Mining the evolutionary process of knowledge through multiple relationships between keywords. Scientometrics, 127(4): 2023-2053, 2022.

[16] X. Cao, X. Chen, L. Huang, et al, Detecting technological recombination using semantic analysis and dynamic network analysis. Scientometrics, Doi: 10.1007/s11192-023-04812-4, 2023.

[17] J. Liu, Z. Zhou, M. Gao, et al, Aspect sentiment mining of short bullet screen comments from online TV series. Journal of the Association for Information Science and Technology, 74(8): 1026-1045, 2023.

[18] J. Won, D. Lee, J. Lee, Understanding experiences of food-delivery-platform workers under algorithmic management using topic modeling. Technological Forecasting and Social Change, 190: 122369, 2023.

[19] L. Huang, X. Chen, Y. Zhang, et al, Identification of topic evolution: network analytics with piecewise linear

representation and word embedding. Scientometrics, 127(9): 5353-5383, 2022.

[20] S. Bai, D. Yu, C. Han, et al, Enablers or inhibitors? Unpacking the emotional power behind in-vehicle AI anthropomorphic interaction: A dual-factor approach by text mining. IEEE Transactions on Engineering Management, Doi: 10.1109/TEM.2023.3327500, 2023.

[21] Z. Zhang, H. Mu, S. Huang, Playing to save sisters: how female gaming communities foster social support within different cultural contexts. Journal of Broadcasting & Electronic Media, 67(5): 693-713, 2023.

[22] J. Wei, T. Hu, J. Dai, et al, Research on named entity recognition of adverse drug reactions based on NLP and deep learning. Frontiers in Pharmacology, 14: 1121796, 2023.

[23] Z. Xue, G. He, J. Liu, et al, Re-examining lexical and semantic attention: Dual-view graph convolutions enhanced BERT for academic paper rating. Information Processing & Management, 60(2): 103216, 2023.

[24] X. Zhu, Z. Kuang, L. Zhang, A prompt model with combined semantic refinement for aspect sentiment analysis. Information Processing & Management, 60(5): 103462, 2023.

[25] C. Zhang, Improved word segmentation system for Chinese criminal judgment documents. Applied Artificial Intelligence, 38(1): 2297524, 2024.

[26] K. Gupta, A. Ahmad, T. Ghosal, et al, A BERT-based sequential deep neural architecture to identify contribution statements and extract phrases for triplets from scientific publications. International Journal on Digital Libraries, Doi: 10.1007/s00799-023-00393-y, 2024.

[27] Z. Wang, X. Xu, X. Song, et al, Multigranularity pruning model for subject recognition task under knowledge base question answering when general models fail. International Journal of Intelligent Systems, 2023: 1202315, 2023.

[28] N. Xu, Y. Liang, C. Guo, et al, Entity recognition in the field of coal mine construction safety based on a pre-training language model. Engineering, Construction and Architectural Management, Doi: 10.1108/ECAM-05-2023-0512, 2023.

[29] M. Mansurova, V. Barakhnin, A. Ospan, et al, Ontology-driven semantic analysis of tabular data: an iterative approach with advanced entity recognition. Applied Sciences, 2023, 13(19): 10918, 2023.

[30] B. Jiang, W. Tang, M. Li, et al, Assessing land resource carrying capacity in China's main grain-producing areas: Spatial–temporal evolution, coupling coordination, and obstacle factors. Sustainability, 15(24): 16699, 2023.

[31] P. Marjai, A. Kiss, Influential Performance of Nodes Identified by Relative Entropy in Dynamic Networks. Vietnam Journal of Computer Science, 8(1): 93-112, 2021.

[32] T. Liang, C. Li, H. Li, Top-k Learning Resource Matching Recommendation Based on Content Filtering PageRank. Computer Engineering, 43(2): 220-226, 2017.

[33] K. Song, A selection method for industry-university cooperation from the perspective of patentometrics. Library Tribune, 41(11): 19-27, 2021.

[34] N. Liu, P. Shapira, X. Yue, Tracking developments in artificial intelligence research: constructing and applying a new search strategy. Scientometrics, 126(4): 3153-3192, 2021.

[35] L. Lü, T. Zhou, Link prediction in complex networks: A survey. Physica A: statistical mechanics and its applications, 390(6): 1150-1170, 2011.

[36] H. R. Nadendla, A. Srikrishna, K. G. Rao, Rider and Sunflower optimization-driven neural network for image classification. Web Intelligence. IOS Press, 19(1-2): 41-61, 2021.

[37] J. Li, Q. Huang, S. Ren, et al, A novel medical text classification model with Kalman filter for clinical decision making. Biomedical Signal Processing and Control, 82: 104503, 2023.

[38] S. Shen, J. Liu, L. Lin, et al, SciBERT: A pre-trained language model for social science texts. Scientometrics, 128(2): 1241-1263, 2023.

[39] J. Sirrianni, E. Sezgin, D. Claman, et al, Medical text prediction and suggestion using generative pretrained transformer models with dental medical notes. Methods of Information in Medicine, 61(05/06): 195-200, 2022.

[40] H. Hou, S. Ding, X. Xu. A deep clustering by multi-level feature fusion. International Journal of Machine Learning and Cybernetics, 13(10): 2813-2823, 2022.

[41] G. Tamilmani, C. H. Phaneendra Varma, V. Brindha Devi, et al, Medical image segmentation using grey wolf based U-Net with bi-directional convolutional LSTM. International Journal of Pattern Recognition and Artificial Intelligence, Doi: 10.1142/S0218001423540253, 2023.

[42] L. C. Wei-Jie, S. C, Chong, T. S. Ong, Masked face recognition with principal random forest convolutional neural network (PRFCNN). Journal of Intelligent & Fuzzy Systems, 43(6): 8371-8383, 2022.

[43] L. Garrote, J. Perdiz, U. J. Nunes, Costmap-based local motion planning using deep reinforcement learning. In 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE (pp. 1089-1095). 2023.

[44] S. Gu, J. Qiao, Z. Zhao, et al, Power system transient stability assessment based on graph neural network with interpretable attribution analysis. In 2022 4th International Conference on Smart Power & Internet Energy Systems (SPIES). IEEE (pp. 1374-1379). 2022.