ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202509.00124

Fine-grained Credibility Assessment of Social Media Health Information via Semantic Enhancement

Authors: Liu Chengshan, Tang Xuan, Qin Chunxiu, Tang Xuan

Date: 2025-09-20T18:42:27+00:00

Abstract

[Objective/Significance] To address the problems of weak model generalization capability caused by scarce annotated data for social media health information, sparse semantics in short texts, and coarse classification granularity. This study aims to investigate methods for enhancing semantic features of health information, refine credibility classification, and improve the interpretability and accuracy of health information credibility evaluation.

[Method/Process] First, a BERT-CNN-BiLSTM semantic enhancement architecture was constructed, introducing dependency syntactic analysis to obtain candidate relation tuples and fusing external medical knowledge bases to generate standard relation tuples. Second, a credibility evaluation index system was built by integrating dimensions such as comment semantics and sentiment features. Finally, health information was classified into five categories based on credibility levels, achieving fine-grained evaluation of health information.

[Results/Conclusion] Comparative validation results with expert-annotated datasets and baseline models demonstrate that the proposed model achieves accuracy and F1-score of 90.8% and 89.0%, respectively, on the five-class classification task, verifying that the fine-grained credibility evaluation model incorporating semantic enhancement can effectively improve the evaluation performance of social media health information.

Full Text

Fine-Grained Credibility Assessment of Social Media Health Information with Semantic Enhancement

Liu Chengshan, Tang Xuan*, Qin Chunxiu
School of Economics and Management, Xidian University, Xi'an 710126, China

Email: 23061212962@stu.xidian.edu.cn

[**Objective/Significance**] This study aims to address the challenges of weak model generalization, semantic sparsity in short texts, and coarse classification granularity caused by scarce annotated health information data on social media. It explores methods to enhance semantic features of health information, refine credibility classification, and improve the interpretability and accuracy of health information credibility assessment. [**Methods/Process**] First, we constructed a BERT-CNN-BiLSTM semantic enhancement architecture. Dependency parsing was employed to extract candidate relation tuples, which were then fused with external medical knowledge bases to generate standardized relation tuples. Second, we integrated dimensions such as comment semantics and sentiment features to establish a credibility assessment metric system. Finally, health information was categorized into five levels based on credibility, enabling fine-grained evaluation. [**Results/Conclusions**] Comparison with expert-annotated datasets and baseline models demonstrates that the proposed model achieves 90.8% accuracy and 89.0% F1 score in the five-classification task. This validates that the proposed fusion semantic enhancement-based fine-grained credibility assessment model effectively improves the evaluation of health information on social media.

Keywords: Semantic enhancement; Information credibility; Fine-grained assessment; Social media
Classification Number: G206

Funding: This work was supported by the National Social Science Fund Key Project "Research on Fine-Grained Organization and Precise Service Models of Literature Resources in China's Key Core Fields Driven by Scenarios" (Project No. 22ATQ002).

Authors: Liu Chengshan, Associate Researcher, Master's Supervisor; Tang Xuan, Master's Student, Email: 23061212962@stu.xidian.edu.cn; Qin Chunxiu, Professor, Ph.D., Master's Supervisor.

---

1 Research Status

To address the issues of colloquialism, semantic ambiguity, and insufficient semantics in social media health information, this paper introduces semantic enhancement methods to explore approaches for augmenting textual semantic features, thereby enabling fine-grained credibility assessment based on relevant health information credibility research.

1.1 Related Research on Semantic Enhancement

Semantic enhancement refers to the use of semantic analysis and annotation methods to improve the capture and processing of semantic information, addressing deficiencies in semantic revelation and encoding representation in current

electronic documents [3]. This section analyzes the topic from two perspectives: research domains and methodological approaches. At the domain level, significant progress has been made in natural language processing. Yang et al. [4] employed a Semantic Enhancement Mechanism (SEM) in relation extraction to extract multi-channel salient features of sentences, obtaining enhanced sentence representations through entity-aware word embeddings and Salient Feature Perception (SFP), thereby improving feature learning capability and relation extraction performance. For short symptom texts with insufficient semantics, Word2Vec and TF-IDF algorithms were used to extract keywords based on term importance and relevance, which were then added to the text for semantic enhancement, significantly improving classification performance with machine learning algorithms [5]. Credibility assessment can be transformed into a text classification problem to some extent, and semantic enhancement can address insufficient semantic information and enrich textual feature representation by fusing multimodal information [6], leading to improved accuracy in identifying false information and rumors. Wang et al. [7] incorporated sentiment features extracted from a BERT-BiLSTM sentiment analysis model on top of single textual features, improving the accuracy of false health information detection. Sun et al. [8] input text into BERT to obtain vector matrices, then fed them into CNN/RNN to obtain rumor probabilities, which were combined with other discrete features and input into XGBoost (eXtreme Gradient Boosting) to obtain rumor classification results, significantly improving detection accuracy.

These studies demonstrate that semantic enhancement can mine deep semantic features of text, enrich information dimensions, solve the problems of sparse features and insufficient semantics in short texts, and effectively improve information identification accuracy. The combination of models such as BERT and BiLSTM plays a significant role in semantic enhancement and text classification tasks. For texts containing complex medical relationships, using a single model may result in limited representation capability and insufficient generalization performance, whereas combined models can collaboratively enhance semantic representation and improve semantic understanding and classification accuracy. Currently, there is relatively little research on fusing multiple model methods in semantic feature enhancement studies. Given the application achievements of model combination architectures in related fields such as rumor and false health information detection, this paper considers introducing a BERT-CNN-BiLSTM architecture to achieve deeper semantic understanding and feature extraction of health information, thereby improving the discrimination and accuracy of subsequent credibility classification.

1.2 Related Research on Health Information Credibility

Information credibility influences users' attitudes toward online health information [9] and represents a key concern for information users. Users' credibility judgments of health information affect their health decision-making. Current research on health information credibility has gradually deepened, focusing pri-

marily on theoretical studies, exploration of influencing factors, and investigation of credibility assessment methods.

First, at the theoretical level, Song et al. [10] analyzed credibility research from three dimensions—conceptual connotation, theoretical foundation, and research methodology—and summarized an evolutionary framework of credibility themes from the iField perspective. At the methodological level, most studies are based on user experiments and questionnaire data. Some studies designed user experiments to investigate how information sources and consumers' e-health literacy affect users' evaluation of online health information credibility [11]. Others used questionnaire surveys to explore the influence of source credibility and content reliability on users' judgments of health information credibility [12]. In recent years, with extensive research on deep learning, deep learning has been increasingly introduced to solve credibility assessment problems. For example, P. Swarup et al. [13] proposed an Attention-based Recurrent Multi-channel Convolutional Neural Network (ARMCNN) model to predict the credibility of online health information, yielding better performance than other benchmark techniques. S. Wafa et al. [14] employed a hybrid LDA-LSTM approach to evaluate latent topics and concluded that credibility assessment based on hybrid machine learning methods is more accurate than existing recommendation models.

Second, at the level of exploring credibility influencing factors, Song et al. [15] conducted a literature review and summarized factors influencing consumers' credibility judgments of distorted health information from the perspectives of information characteristics and individual characteristics, finding that both the framing of distorted information and individual factors such as age and gender significantly affect credibility judgments. Other studies have used information credibility as a mediating variable to investigate the impact of information framing on HPV vaccination in online health communities [16].

Third, regarding the application of credibility assessment in rumor detection, Qian et al. [17] used information credibility assessment methods to quantify the credibility of public opinion information, ultimately achieving credibility grading of rumor detection results. Liu [18] divided rumor sources into high-credibility and low-credibility categories and used Kaplan-Meier survival estimation and Cox proportional hazard models for analysis, finding that source credibility had no significant effect on rumor survival time.

In summary, although existing research on health information credibility has made certain progress, studies have mainly focused on the Twitter platform [19-20], with most attention given to public health emergencies [21] and user credibility assessment [22], while lacking systematic evaluation of daily massive basic health information. In terms of assessment methods, most studies adopt binary classification approaches, and coarse-grained information classification results cannot reflect differences in credibility levels, making it difficult to effectively identify "partially credible" health information types such as "incomplete content," "misleading information," or "lack of evidence," and failing to provide users with more accurate and discriminative credibility assessment basis.

Furthermore, existing credibility assessment models have limited research on extracting features unique to health information, such as professional expressions and medical relationships, necessitating further investigation into health information semantic extraction methods to enhance model applicability and accuracy.

Therefore, this paper focuses on health information under specific topics on social media platforms, fuses semantic enhancement methods, proposes a BERT-CNN-BiLSTM deep semantic extraction architecture, and introduces dependency syntax and external medical knowledge bases to strengthen the identification of medical entities and relationships. By integrating comment text and sentiment analysis, we construct a social media health information credibility assessment metric system across four dimensions, ultimately inputting into a DNN model to achieve fine-grained assessment of health information credibility under five categories, aiming to provide more precise and interpretable health information quality assessment results.

---

2 A Semantic Enhancement Method

This section expands the dataset through keyword-based data augmentation methods, fuses external medical knowledge bases and dependency syntax analysis to annotate, align, and fuse medical terms and relationships in text, and constructs a BERT-CNN-BiLSTM deep semantic extraction architecture.

2.1 Keyword-Based EDA Method

Currently, there are no open-source datasets for health information credibility assessment. Relying solely on expert annotation as an important basis for information credibility assessment results incurs high costs. To reduce this cost, we invited medical experts in the field to annotate a small amount of text. Based on the labeled dataset, to prevent semantic changes and loss of key information caused by replacing or deleting core medical terms, we propose a keyword-based EDA (Easy Data Augmentation) data augmentation method. Considering both term frequency and contextual semantic relationships, we use TF-IDF to extract high-frequency terms and TextRank to extract semantically core words, and designate other medically related terms as keywords to serve as objects for synonym replacement, insertion, and exchange, while excluding them from random deletion operations to ensure no loss of key information. An example of the keyword-based method is shown in Figure 1 [Figure 1: see original paper].

Figure 1 Examples of keyword-based EDA

The synonym lexicon was built by crawling medical terms and their synonyms from disease encyclopedias, based on the “Harbin Institute of Technology Synonym Forest Extended Edition.”

2.2 Entity-Relationship Based Semantic Enhancement

Information on social media platforms mostly consists of unstructured text with complex semantics and professional medical knowledge, where colloquial expressions lead to truncated or metaphorical professional terms, posing difficulties for models to extract textual semantics. Therefore, we combine dependency syntax analysis with entity and relationship alignment from external medical knowledge bases to achieve accurate deep semantic extraction.

In the entity alignment phase, we use the spaCy library in Python to identify and extract medical entities from text, including disease names, symptom manifestations, medical drugs, and other types, obtaining the initial entity set for each data point. By fusing external knowledge bases such as ICD-11, "Disease Encyclopedia," and "Baidu Baike," we align medical terms in health information with the knowledge bases. In the entity relationship extraction phase, we introduce dependency syntax analysis to parse sentence structures, identify core verbs and their grammatical dependency relationships, and obtain real associations between entities based on dependency path constraints to generate candidate relations such as {disease, symptom}, {disease, treatment method} binary tuples or {disease, symptom, treatment method} ternary tuples. These are then semantically matched and structurally validated against standard relationships from external knowledge bases to resolve relationship expression ambiguity and complete missing association paths between entities, providing structured semantic input for subsequent deep neural networks. For example, a health information text states: "Helicobacter pylori infection generally does not heal itself without medication. If not treated with medication in time, it may induce gastric diseases such as gastric ulcers and chronic gastritis. If the infection persists, it will induce gastric cancer." An example of dependency syntax is shown in Table 1 .

Table 1 Dependent Syntax Example

Using spaCy's Chinese model to identify entities in the text, we obtain the entity set T={Helicobacter pylori, gastric ulcer, chronic gastritis, gastric cancer}, all of which can be standardly mapped to medical knowledge bases. We identify the key core verb "induce" in the sentence. For texts containing multiple entities, we obtain multiple groups of candidate relations according to proximity principles and dependency path constraint rules, and determine the final relation tuples by combining them with medical knowledge bases.

Relying on BERT's mature pre-training capability to provide deep global contextual semantic representations for text, we use BERT to output the data-augmented text as 768-dimensional feature vectors. The Convolutional Neural Network (CNN) extracts local key features on this basis, where convolution kernels can effectively capture short-range dependencies between entities, and multi-scale characteristics can capture semantic relationships of different lengths between medical entities, as shown in Formula (1).

$$\langle\langle MATH_1 \rangle\rangle$$

where the convolution kernel sizes fs=[3,5,7], Wk represents the weight of the k-th convolution kernel, and X(h,d) represents the input features. We introduce an adaptive convolution kernel size selection mechanism to respectively adapt to local relationship features of short symptom phrases such as "persistent stomach pain," medium-length treatment descriptions such as "daily oral antibiotics," and complex medical expressions such as "gastroscopy shows Helicobacter pylori infection." Through max pooling, we extract salient features at each scale and concatenate them, introducing an Attention mechanism to dynamically adjust the weights of multi-scale features, enabling the model to autonomously enhance the representation contribution of medical entities. Finally, the weighted features are input into a fully connected layer to generate low-dimensional relation vectors.

2.3 Text Semantic Feature Fusion

The BiLSTM model captures forward and backward semantic dependency relationships in text sequences simultaneously through its bidirectional recurrent neural network structure. Through progressive processing at global-local-sequence levels, it compensates for the limitations of single models and achieves semantic enhancement of health information from multiple dimensions. When processing health information, the model performs bidirectional encoding on the hidden states of each word, dynamically regulating information flow through gating mechanisms where the input gate prioritizes writing professional terms and the forget gate suppresses historical memory retention of vague expressions. Professional medical terms such as "Helicobacter pylori" and "gastric ulcer" receive higher gating weights, while vague symptom descriptions such as "stomach discomfort" have their semantic contribution adjusted through contextual context.

In the feature fusion phase, the high-dimensional vectors generated by BERT carry global semantics, and the local features extracted by CNN with adaptive convolution kernels are concatenated with BiLSTM's hidden states, as shown in Formula (2).

$$\langle\langle MATH_2 \rangle\rangle$$

where hsemantic represents the semantic feature output by BiLSTM as the concatenated vector of bidirectional hidden states, and h1 and hT represent the forward and backward hidden states, respectively, with || denoting vector concatenation operation. The fused feature vector output from BERT-CNN is used as the input sequence for BiLSTM, and after bidirectional encoding, we obtain semantic representations containing dynamic weights. Finally, the model maps high-level semantics to the classification space through a fully connected layer,

using the sigmoid function to map health information semantic credibility features to [0,1], outputting semantic credibility probability, as shown in Formula (3).

$$\langle\langle MATH_3 \rangle\rangle$$

where Ptrust represents the semantic probability value, σ represents the sigmoid function, Wc represents the weight matrix of the fully connected layer, and bc represents the bias of the fully connected layer. This probability value serves as the core semantic feature and is fused with other features as the final input for health information classification.

---

3 Construction of Fine-Grained Credibility Assessment Model

This section fuses the semantic enhancement methods described above to construct a credibility assessment model comprising a feature extraction layer, feature fusion layer, and fine-grained assessment layer. The credibility assessment model is shown in Figure 2 [Figure 2: see original paper].

Figure 2 A fine-grained evaluation model for the credibility of social media health information

3.1 Feature Extraction Layer

Considering the user interaction characteristic of social media platforms, this paper adds comment semantics and sentiment feature analysis to the feature extraction layer based on existing metric systems, constructing a metric system from four dimensions: source credibility, content credibility, sentiment features, and propagation features, yielding 12 specific indicators as shown in Table 2 , with measurement methods provided for each indicator.

Table 2 Multi-dimensional characteristic index system of credibility

Source Credibility Feature Extraction
Hovland's source credibility theory posits that information acceptance depends on information credibility, which in turn depends on the source's expertise and reliability [23]. This paper selects four indicators for measurement: user authority, publisher influence, activity level, and publisher information completeness [8]. Verified users have higher credibility than unverified users. We categorize verification information into five types and assign values from high to low: central government assigned 5 points, followed by official organizations, online media, medical experts, and non-medical individual users, with unverified users assigned 0. Users with higher influence are more likely to attract attention from other users and thus have higher credibility, which can be calculated through the

ratio of followers to (followers + following) [8]. The historical activity level of information publishers can reflect user credibility to some extent, with the most direct measurement indicator being the number of posts per unit time [24]. We count all users' posting volumes and perform equal-interval division, assigning values from 1 to 5 from low to high. The completeness of personal information of information publishers affects users' credibility judgments; more complete information indicates higher authority and credibility, including personal profile, birthday, location, education information, and occupation information. We construct a vector V=(v1, v2, … vi) to represent user information completeness, where vi represents the i-th personal information item, marked as 0 if information is missing and 1 if it contains valid information, as shown in Formula (4).

$$\langle\langle MATH_4 \rangle\rangle$$

where m represents user information completeness and n represents vector dimension.

Content Credibility Feature Extraction

Content quality is a crucial factor determining content credibility, which is a key factor in evaluating information credibility. Content quality indicators include consistency, relevance, and professionalism. This paper mainly measures from two aspects: professionalism and objectivity [25]. To increase information attention, publishers use exaggerated terms such as "rejuvenation" and "cure-all," which can easily mislead users into forming incorrect health perceptions. Therefore, samples containing exaggerated vocabulary are labeled as 0, and those without are labeled as 1. Compared with other information, users have higher quality requirements for health information. If literature is cited and information sources are indicated, information accuracy and user-perceived credibility can be improved. Samples containing source basis are labeled as 1, otherwise 0. Additionally, commercial promotion phenomena exist where information is mixed with commercial advertisements and health product sales, which cannot guarantee information quality [25]. Therefore, samples containing commercial promotion components are labeled as 0, otherwise 1.

Sentiment Feature Extraction

Sentiment features include emotions and sentiment tendencies contained in the main text and comments, which significantly affect users' attitudes and cognition toward information [26]. Sentiment feature analysis can assist credibility assessment. The more authentic the information, the more it tends toward positive or neutral sentiment; more negative sentiment indicates lower credibility. The sentiment polarity of hot comments affects users' trust in information, so we select the top ten hot comments under each piece of information and use the SnowNLP tool in Python to classify the sentiment tendencies of the main text and comments into positive and negative emotions. The more positive the tendency, the closer to 1; the more negative, the closer to 0. We accumulate

the sentiment value of each comment and divide it by the number of crawled comments to obtain comment sentiment degree.

Propagation Feature Extraction

The propagation characteristics of social media can reflect the popularity and influence of content. User comments contain users' real reactions and judgments, which affect users' evaluation of health information credibility. If comments contain questioning vocabulary such as "fake" or "rumor," the content may contain false information [17]. We construct a questioning vocabulary list and use regular expressions to match whether comments contain questioning terms, assigning 0 if present and 1 if absent. Finally, we calculate comment questioning degree through weighted computation of each comment's assignment. Propagation typically considers repost volume, like volume, and comment count [8]. The propagation value calculation is shown in Formula (5).

$$\langle\langle MATH_5 \rangle\rangle$$

where repost volume, comment count, and like volume are represented by repd, comd, and attd, respectively. Combined with weight parameters γ, , and , the propagation value is normalized and mapped to [0,1].

3.2 Feature Fusion Layer

The feature fusion layer fuses the four dimensions and 12 feature indicators extracted in the previous section. First, the Analytic Hierarchy Process (AHP) is used, where two medical experts judge the relative importance among the four dimensions and 12 feature indicators according to the 1-9 scale method, discuss to determine the final weight values in the judgment matrix, and test the consistency ratio of the initial matrix. The first-level indicator CR=0.0888, and second-level indicator CR values are 0.07926 and 0.08296, respectively, all less than 0.1, passing the consistency test. The indicator weights are shown in Table 3 .

Table 3 Credibility evaluates the weight of the indicator

According to Table 3 results, experts believe that content features are the most important factor affecting information credibility, with a weight close to 0.6, among which semantic features are the most important second-level indicator, with a weight of 0.37 among the 12 indicators. Next are source features, propagation features, and sentiment features. Second, after normalizing the feature indicators in each dimension, we obtain four-dimensional feature vectors f1, f2, f3, and f4 through direct concatenation, which are then concatenated into a multi-feature vector Tf. The feature concatenation process is shown in Formula (6).

$$\langle\langle MATH_6 \rangle\rangle$$

3.3 Fine-Grained Assessment Layer

Current research on health information credibility assessment divides information into credible and non-credible categories. This one-size-fits-all evaluation method cannot accurately identify information with mixed truthfulness, and the uneven quality of information affects users' health cognition and may mislead users into making incorrect health decisions. Therefore, based on existing literature, the fine-grained assessment layer clarifies the basis for credibility degree division and categorizes information into five classes according to credibility degree. After revision by medical experts, the credibility classification basis is shown in Table 4 .

Table 4 Credibility classification basis

This study focuses on health information credibility assessment in social media scenarios, where texts vary in length, contain complex semantics including both popular expressions and professional medical terms, and the extracted indicator features include source, sentiment, and medical relationship features. It requires coordinating heterogeneous representations of textual features and knowledge graph features and precise classification under fuzzy text classification boundaries. The DNN model can fuse heterogeneous features through fully connected layers, making it suitable for multi-source heterogeneous feature fusion scenarios. Its deep nonlinear structure can depict more complex decision boundaries, making it applicable to multi-classification tasks. Therefore, this paper selects the DNN model as the classification model for credibility assessment tasks. The fused feature indicators are input into the DNN. The model input is normalized feature vectors. Features are mapped to hidden space through fully connected layers, and high-order nonlinear features are extracted level by level through dual hidden layers (512$→$256 neurons) with ReLU activation functions, capturing complex patterns such as medical entity conflict detection and multi-source feature interaction. The output uses the sigmoid function to generate binary classification probability distribution, the Softmax function to generate five-classification probability distribution, and cross-entropy loss function to optimize classification boundaries. Finally, health information is divided into five categories according to credibility degree, achieving fine-grained assessment of health information credibility.

---

4 Experiments

4.1 Data Source and Preprocessing

This experiment crawled Weibo text on the health topic of gastric cancer from the web version of Sina Weibo, with a time range of 2024.1.01-2024.10.10, including each Weibo's content information (main text content, like count, comment count, repost count, and top ten comment contents) and Weibo publisher information (user personal information, follower count, following count, and posting

volume within a certain period).

After removing irrelevant text, we performed jieba word segmentation and stop word removal on Weibo text, removed irrelevant characters such as emojis, @, and hashtags from Weibo text, and obtained 1052 Weibo texts as the initial dataset.

4.2 Data Annotation and Augmentation

Two medical experts in the field of gastric cancer separately annotated the original dataset and performed consistency testing, with a KAPPA coefficient of 0.884, indicating good consistency. Data with consistent annotations from both experts were retained. Based on the five categories of data with consistent expert annotations, the dataset was expanded through the keyword-based data augmentation operation proposed in this paper. The dataset distribution is shown in Table 5 .

Table 5 Data distribution before and after data enhancement

4.3 Experimental Design

(1) Experimental Parameters
The experimental environment of this paper is Python 3.8, using Pytorch to construct deep learning models and conduct model training. The main experimental parameters are shown in Table 6 .

Table 6 The main parameters of the experiment

(2) Credibility Assessment
We use spaCy for named entity recognition on each text to obtain a set of medical entities, map and align the identified entities with external medical knowledge bases such as ICD-11 to complete entity standardization, obtain relation tuples based on dependency syntax and external knowledge bases, and use the vectors obtained from the BERT-CNN layer as BiLSTM input to obtain semantic feature values for each data point to fuse with other features. We use Python's train_{{{test}}_{{split}}} function to divide the training set and test set at a ratio of 8:2, input them into the DNN classification model, and finally obtain the interval distribution results of the health information dataset in five-level classification, as shown in Table 7 .

Table 7 Distribution of Health Information Credibility

According to the data in the table, the model demonstrates high identification accuracy on the "very credible" and "non-credible" information dimensions, indicating that these types of information possess more discriminative features. Horizontal analysis of the data reveals that "very credible" information contains clear factual basis, authoritative sources, and professional expressions, while "non-credible" information contains significant errors, low-credibility sources, or commercial content, making them easily identifiable and accurately judged

by the model. However, for intermediate types of information, identification results show deviations, with their feature representations being more dispersed in semantic space, suffering from insufficient features and fuzzy division intervals, causing certain difficulties in the identification process.

(3) Comparative Experiments
This paper sets up three groups of comparative experiments to verify the performance of each model component.

Data Augmentation Effect Verification
We compare no data augmentation, EDA method, and keyword-based EDA method to verify the effectiveness of the keyword-based data augmentation method proposed in this paper. The experimental results are shown in Table 8 .

Table 8 Data augmentation of experimental results

Compared with no data augmentation and general EDA methods, the keyword-based EDA data augmentation method improves accuracy and F1-score in binary classification tasks, but the improvement is not obvious in five-classification tasks, indicating that the keyword-based EDA method proposed in this paper can improve credibility assessment effectiveness to a certain extent.

Semantic Enhancement Effect Verification
We set up two groups of comparative experiments: (a) compare evaluation results before and after extracting semantic features while keeping other features unchanged to verify the importance of semantic features in credibility assessment; (b) compare the semantic extraction model proposed in this paper with other commonly used semantic extraction models:

a. BiLSTM [27]: Uses forward and backward propagation of BiLSTM to extract deep semantic features.

b. TextCNN [28]: Uses TextCNN to extract high-order semantic features of text.

c. TextRNN [29]: Uses multiple TextRNN groups as semantic feature expert networks to extract semantic features.

d. BiLSTM+CNN [30]: BiLSTM extracts text contextual semantic features, while CNN captures local semantic features.

e. BiGRU+Att [31]: Uses bidirectional GRU neural networks and attention mechanisms to learn and extract semantic features.

All baseline model experimental results on the dataset are shown in Table 9 .

Table 9 Results of semantic enhancement experiments

The experimental results show that the method without semantic feature

extraction is lower than other methods on all indicators, especially performing poorly in five-classification tasks, indicating that semantic features play an important role in credibility assessment. The semantic extraction model BERT+CNN+BiLSTM proposed in this paper performs best among all methods, achieving 90.8% accuracy and 89.0% F1-score in five-classification tasks. This indicates that the model can better capture deep semantic features and improve classification performance.

Credibility Assessment Model Performance Verification
The health information credibility assessment studied in this paper can be regarded as a text classification problem. Therefore, we select commonly used baseline classification models to verify the effectiveness of the DNN classification model. The experimental results are shown in Table 10 .

Table 10 Confidence evaluates the results of the experiment

The experimental results show that traditional machine learning models perform weakly in five-classification tasks, with Random Forest performing best with an F1-score of 78.2%, but still significantly lower than deep learning models. CNN and RNN achieve F1-scores of 82.5% and 80.7% respectively in five-classification tasks, indicating that deep learning models can better capture text features. The classification model proposed in this paper achieves 90.8% accuracy and 89.0% F1-score in five-classification tasks, both superior to all baseline models, demonstrating that the DNN classification model possesses recognition capability and generalization ability for multi-classification tasks.

(4) Result Analysis
We select one sample from each of the five categories for analysis, as shown in Table 11 .

Table 11 Sample Examples

Sample 1 has accurate and complete content, supported by authoritative medical evidence (WHO/IARC classification, epidemiological data), and contains clear and specific action recommendations. **Sample 2** has correct content but lacks elaboration on "quantity" and "conditions," making the content incomplete. **Sample 3** contains correct core facts (Hp is a Class I carcinogen and a risk factor for gastric cancer) but also contains exaggerated components ("directly related") and imprecise expressions, implying that symptoms are universal manifestations of infection. **Sample 4** uses absolute and exaggerated terms like "coming signals" for common non-specific gastrointestinal symptoms such as abdominal bloating, belching, irregular abdominal pain, and post-meal hiccups, causing panic rather than scientific popularization. **Sample 5** is explicitly identified by China's official Internet rumor-refuting platform as having no effect on gastric cancer treatment, with comments also containing questioning words like "fake" and "boasting."

Comprehensive analysis of the above comparative experimental results shows that the model proposed in this paper performs excellently in both semantic ex-

traction and fine-grained credibility assessment. At the semantic enhancement level, keyword-based EDA data augmentation is an effective method to solve the problem of insufficient labeled data, and the semantic extraction model BERT-CNN-BiLSTM outperforms other models, indicating that the fusion of dependency syntax and medical knowledge bases can accurately capture deep semantic associations. At the fine-grained assessment level, the final credibility classification task performance is superior to baseline models. Fine-grained assessment can effectively divide health information into five levels—very credible, relatively credible, generally credible, less credible, and non-credible—optimizing classification results and providing methodological references for information filtering and quality improvement on social media platforms.

5 Conclusion and Outlook

The health information credibility assessment model with fused semantic enhancement proposed in this paper improves assessment granularity and accuracy. First, to address the problem of scarce annotated data, we propose a keyword-based EDA method to effectively expand the dataset and improve model generalization capability. We construct a deep semantic extraction architecture BERT-CNN-BiLSTM to solve problems of semantic ambiguity and complex medical relationships, enhancing semantic features. We use BERT and CNN to capture global and local semantic information of text, respectively, and introduce dependency syntax and external medical knowledge bases to improve the model's understanding and accuracy of medical entities and relationships. Second, based on existing research, we add comment semantics and sentiment features to construct a credibility assessment system across four dimensions. Finally, we achieve fine-grained assessment of health information credibility through a DNN classification model. Experimental results show that the model proposed in this paper achieves 89.0% assessment accuracy under five-classification, outperforming baseline models in binary classification tasks. On the basis of improving the accuracy of "credible information" and "non-credible information," it effectively distinguishes "partially credible" health information.

Future work will further expand the model's capability to process multimodal information such as images and videos, fusing multimodal information to provide more comprehensive technical support for health information credibility assessment on social media platforms and promote content governance and health communication on social media platforms.

References

[1] General office of the state council of the People's Republic of China. China to promote national health during 14th five-year plan[R]. General office of the state council of the People's Republic of China, 2022.

[2] DENG S L, GU Y F. A review of online health misinformation: recognition,

action and governance[J]. Library journal, 2022, 41(05): 14-22.

[3] SONG N Y, PEI L, WAGN C Y. The survey and tendency of semantic enrichment for scientific papers[J]. Library and information service, 2021, 65(01): 82-90.

[4] YANG D W, XU X H, SONG W. Relation extraction method combining semantic enhancement and perception attention[J/OL]. Journal of computer applications, 2025.

[5] ZHANG M Y, DING J D. Research on semantic enhancement for short text classification[J]. Library and information service, 2023, 67(09): 4-11+3.

[6] LI Y, LU Z X, LIU S Y, et al. Research on micro-video multi-label classification based on deep multimodal association learning[J]. Data analysis and knowledge discovery, 2024, 8(07): 77-88.

[7] Wang H, Gong L J, Zhou Z Y, et al. Research on a method for detecting false information in social media by integrating semantic enhancement [J]. Data analysis and knowledge discovery, 2023, 7(02): 48-60.

[8] Sun R, An L. Research on rumor identification during public health emergencies [J]. Information and documentation work, 2021, 42(05): 42-49.

[9] YAO Y B, CAO J D, FENG F Z, et al. A meta-analysis of factors influencing digital health literacy among residents at home and abroad[J]. Medicine and society, 2024, 37(11): 122-129.

[10] SONG S J, ZHAO Y X, ZHU Q H. Information credibility research in the iField: conceptual development, topic evolution, and future direction[J]. Journal of library science in China, 2022, 48(01): 107-126.

[11] YUNG S C, YAN Z, JACKE G. The effects of information source and eHealth literacy on consumer health information credibility evaluation behavior[J]. Computers human behavior, 2021, 115: 106629.

[12] S. B. EGALA, D. LIANG, D. BOATENG. Social media health-related information credibility and reliability: an integrated user perceived quality assessment[J]. IEEE transactions on engineering management, 2024, 71: 5018-5029.

[13] SWARUP P, SANTOSH S R. An attention-based deep learning model for credibility assessment of online health information[J]. Computational intelligence, 2023, 39(5): 832-859.

[14] SHAFQAT W, BYUN Y C, PARK N. Effectiveness of machine learning approaches towards credibility assessment crowdfunding projects reliable recommendations[J]. Applied sciences, 2020, 10(24): 9062.

[15] SONG S J, ZHAO Y X, SONG X K, et al. Investigating the influential factors of consumer's credibility judgment on health misinformation[J]. Journal of library science in China, 2019, 45(04): 72-85.

[16] XU X T, ZHANG T T, ZHU Q H. Research on the impact of message framing on HPV vaccination in online health communities: a mediation of message credibility[J]. Library & Information, 2020, (05): 39-47.

[17] QIAN D M, ZHENG J M, WANG W J, et al. Rumor identification of public health emergency based on information credibility assessment[J]. Information science, 2024, 42(02): 35-42.

[18] LIU X, ZHANG L, SUN L. et al. Survival analysis of the duration of rumors during the COVID-19 pandemic[J]. BMC public health, 2024, 24(1).

[19] KAUFHOLD M A, BAYER M, HARTUNG D. et al. Design and evaluation of deep learning models for real-time credibility assessment on twitter[J]. Lecture notes artificial intelligence, 2021, 12895: 396-408.

[20] AHMAD F, RIZVI S. Identification of credibility content measures for twitter and sina-weibo social networks[C]//Proceedings icetit 2019: Emerging trends in information technology, 2020, 605: 372-384.

[21] LI Y M, XU X K, WANG P C. Research on credibility evaluation of network public opinion and government microblog guidance for emergencies[J]. Journal of intelligence, 2021, 40(11): 87-92+120.

[22] WANG Z N, ZHANG G F. Microblog users' credibility evaluation algorithm considering multi-interaction relationships and emotional tendencies[J]. Application research of computers, 2024, 41(10): 3000-3007.

[23] HOVLAND C.I., WEISS W. The Influence of source credibility on communication[J]. Public opinion quarterly, 1951(15): 635-650.

[24] ZHU M J, JIANG H X, XU W. Weibo moods and propagation factors based stock prices prediction[J]. Journal of shandong university (Natural science), 2016, 51(11): 13-25.

[25] Lu Q, Yue XQ, Liu T, et al. Research on the credibility assessment of online health information based on multidimensional features [J]. Journal of information resource management, 2021, 11(03): 121-131.

[26] PAUL V S, KAREN R, CHRISTOPHER W, et al. Risk as affect: the affect heuristic in cybersecurity [J]. Computers & Security, 2020(90): 101651-101651.

[27] CAO B W, CAO J X, GUI J, et al. Character relation extraction from Chinese literature[J]. Journal of Chinese information processing, 2023, 37(05): 88-100.

[28] YUAN P Y, QIU L. Web attacks detection method based on BERT with multi-model fusion[J]. Computer engineering, 2024, 50(11): 197-206.

[29] WANG Z Q, CHEN T, ZHANG B Y, et al. Multi-domain fake news detection based on cross-feature perception fusion[J]. Computer systems & applications, 2024, 33(03): 264-272.

[30] FENG L P, DONG C C, XU X K. Research on weibo rumor identification in public health emergencies based on hybrid neural network[J]. Journal of intelligence, 2022, 41(12): 81-88.

[31] HUANG X J, WANG G S, LUO Y S, et al. Weibo rumors real-time detection model based on fusion of multi-user features and content features[J]. Journal of Chinese computer systems, 2022, 43(12): 2518-2527.

Corresponding Author: Tang Xuan, E-mail: 23061212962@stu.xidian.edu.cn

Author Contributions:
Liu Chengshan: Proposed research ideas, reviewed and revised the paper;
Tang Xuan: Collected and analyzed data, wrote the paper;
Qin Chunxiu: Discussed research ideas, revised the final version of the paper.

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

Source: ChinaXiv — Machine translation. Verify with original.