Complex & Intelligent Systems
https://doi.org/10.1007/s40747-022-00757-y

ORIGINAL ARTICLE

An associative knowledge network model for interpretable semantic representation of noun context

Yulin Li[1,2] · Zhenping Xie[1,2] · Fanyu Wang[1,2]

Received: 28 August 2021 / Accepted: 15 April 2022

Abstract
Uninterpretability has become the biggest obstacle to the wider application of deep neural network, especially in most human–machine interaction scenes. Inspired by the powerful associative computing ability of human brain neural system, a novel interpretable semantic representation model of noun context, associative knowledge network model, is proposed. The proposed network structure is composed of only pure associative relationships without relation label and is dynamically generated by analysing neighbour relationships between noun words in text, in which incremental updating and reduction reconstruction strategies can be naturally introduced. Furthermore, a novel interpretable method is designed for the practical problem of checking the semantic coherence of noun context. In proposed method, the associative knowledge network learned from the text corpus is first regarded as a background knowledge network, and then the multilevel contextual associative coupling degree features of noun words in given detection document are computed. Finally, contextual coherence detection and the location of those inconsistent noun words can be realized by using an interpretable classification method such as decision tree. Our sufficient experimental results show that above proposed method can obtain excellent performance and completely reach or even partially exceed the performance obtained by the latest deep neural network methods especially in F1 score metric. In addition, the natural interpretability and incremental learning ability of our proposed method should be extremely valuable than deep neural network methods. So, this study provides a very enlightening idea for developing interpretable machine learning methods, especially for the tasks of text semantic representation and writing error detection.

Keywords Text semantic modelling · Interpretable computing · Associative knowledge network · Semantic coherence of noun context · Incremental learning

Introduction

Text semantic representation is one of the core contents of natural language processing and plays an indispensable role in different applications, such as text classification [1], sentiment analysis [2] and information extraction [3]. Existing text semantic representation models can be divided into vector space models [4] and neural net-based methods. The former includes the latent semantic analysis model (LSA) [5] and latent Dirichlet allocation (LDA) [6], and the latter has word2vec [7] and doc2vec [8] methods. However, vector space model and traditional topic model method cannot model the contextual semantic information of text with high-precision. Although neural network methods have achieved relatively better precision, but their interpretability is very poor, which seriously limits its application scope.

In recent years, knowledge graph technology has been widely introduced in the field of text analysis. A knowledge graph is a tool of describing knowledge and modelling the relationships between things based on a graph structure [9], which has shown strong practical value in intelligent question answering [10, 11], natural language understanding [12], big data analysis [13–15], interpretability enhancement of machine learning [16], semantic search [17, 18], etc. When using knowledge graph to model text semantic information, entities in the text are represented as nodes in network, and edges represent the relationships between entities. In traditional knowledge graph construction methods, no matter for

✉ Zhenping Xie
xiezp@jiangnan.edu.cn

1 School of Artificial Intelligence and Computer Science, Jiangnan University, No. 1800 Lihu Avenue, Wuxi 214122, Jiangsu, People's Republic of China

2 Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, No. 1800 Lihu Avenue, Wuxi 214122, Jiangsu, People's Republic of China

Published online: 13 May 2022

handcrafting rule methods [19] or deep learning methods [20], knowledge relationships among knowledge concepts are all considered to own extra semantic labels (like 'belong to', 'is a', 'located at', etc.). However, the above assumption is quite different from the basic computing process of our brain neural network, that is, there should be no label difference for different neural connections of human brain neurons. So, existing knowledge graph model will be not general for some text semantic analysis applications such as error detection in text writing.

For error correction, there are commonly two types of errors: word spelling errors, and grammatical or syntactic errors. Word spelling error correction methods are mainly based on word dictionaries, generally without considering whether the contextual semantic relations of words are reasonable. In contrast, grammatical, or syntactic error correction is relatively complex, and it is also a difficult issue in high-quality text writing even for people. Wherein four types of problems can be concluded including redundant words, missing words, word selection errors, and word ordering errors [21]. Such errors can only be found by means of semantic analysis on text context. However, current Chinese text error correction methods mainly focus on spelling error corrections, and few semantic error corrections.

Based on the above motivations, inspired by the associative computing mechanism of human brain, we propose a new model, associative knowledge network model, which uses the neighbour relationships between noun entities in the text to model the semantic relationships. Different from existing knowledge graph construction methods, semantic labels among knowledge relationships are no longer specifically considered in our proposed model, but only one type of relationship between knowledge concepts is considered, that is, a unified associative relationship with strength. And then, the performances of the new model are studied by solving the problem of checking the semantic coherence of noun context. That is a new text error detection method in word granularity and its main function is to detect and locate those noun entities with inconsistent contextual semantic in texts.

The main contributions of this study can be summarized as follows:

(1) An interpretable text semantic representation model of noun context, named as associative knowledge network, is proposed, in which an improved associative strength computing equation is newly designed.
(2) An interpretable method for checking the semantic coherence of noun context is designed by taking the learned associative knowledge network as background knowledge network. New method can realize the Chinese text word error detection using multilevel contextual word semantic relations.
(3) The experimental results indicate that, the proposed method has not only good interpretability but also excellent detection performance compared to latest state-of-art neural network methods.

Related work

Text semantic representation modelling

Recently, great progress has been made in research work related to the modelling of text semantic representation. In the technical aspect based on knowledge graph, Etaiwi et al. [22] proposed a graph-based semantic representation model for Arabic text. The core idea is to use predefined rules to identify the semantic relationship between words and build the final semantic graph. Wei et al. [23] proposed a multilevel text representation model within background knowledge, which captures the semantic content of the text at three levels, machine surface code, machine text base and machine situational model. Furthermore, external background knowledge is introduced to enrich the text representation so that the machine can better understand the semantic content in the text. Geeganage et al. [24] proposed a semantic-based topic representation using frequent semantic patterns, and in new method the text semantic can be captured by matching the words in each topic with concepts in the Probase ontology.

In recent years, in addition to using knowledge graph to model text semantic representations, an increasing number of researchers have devoted themselves to the study of text semantic representations combined with deep neural networks to extract deeper text semantic features. Chen et al. [25] proposed a neural knowledge graph evaluator to effectively predict the reliability of answers in an automatic question answering system, in which the prediction performance is mainly improved by jointly encoding structural and semantic features in a knowledge graph. Wang et al. [26] proposed a novel text-enhanced knowledge graph representation model. They introduced a mutual attention mechanism between the knowledge graph and text to mutually reinforce the relationship between knowledge graph representation and textual relation representation. Wang et al. [27] proposed a graph-based neural network model for early fake news detection based on enhanced text representations. They modelled the global pair-wise semantic relations between sentences as a complete graph, and learned the global sentence representations via a graph convolutional network with self-attention mechanism. Although deep neural networks have shown good advantages in the study of text semantic representation learning, one of their well-known problems is that their learning representation is difficult to interpret. Accordingly,

Xie et al. [28] proposed a novel neural sparse topic model called semantic reinforcement neural variational sparse topic model for explainable and sparse latent text representation learning. Ennajari et al. [29] proposed a Bayesian embedded spherical topic model that combines both knowledge graph and word embeddings in a non-Euclidean curved space, the hypersphere, for better topic interpretability and discriminative text representation. These developments all enhance the interpretability of neural networks by adding interpretable semantic module to neural networks, but the whole models are still not completely interpretable.

When using a knowledge graph to model the semantic representation of text, the measurement of semantic relationships between knowledge is also an essential task. To compensate for the incomplete measurement ability of the co-occurrence frequency and mutual information method in quantifying the relevance relation between words, Zhong et al. [30] proposed a quantitative computing method for the relationship between words that integrates co-occurrence frequency and mutual information. Wang et al. [31] proposed a new semantic relationship measurement method according to the number of times and intensity of knowledge co-occurrence in the text. Li et al. [32] proposed a lightweight algorithm for learning word single-meaning embeddings to enhance the accuracy of semantic relatedness measurement by developing WordNet synsets and Doc2vec document embeddings.

Chinese text error correction

Chinese text error correction is an important technology for realizing automatic checking and error correction of Chinese writing. Its importance in the fields of automatic question answering systems [33], machine translation [34] and summary generation [35] is self-evident. To solve the problems of mismatching words and unsmooth context sentences in text paragraphs, many text error correction techniques have been developed.

Cui et al. [36] proposed a new pre-trained model called MacBERT that mitigates the gap between the pre-training and fine-tuning stage by masking the word with its similar word, which has proven to be effective on downstream tasks. Liu et al. [37] proposed a pre-trained masked language model with misspelled knowledge (PLOME) for Chinese spelling correction, which jointly learns how to understand text semantic and correct spelling errors. Zhang et al. [38] proposed a new neural network model based on BERT for Chinese spelling error correction, which consists of two networks respectively for error detection and error correction. This model can detect the correctness of every position of Chinese sentences, which is an effective application extension of the original BERT model.

The proposed method

Overview

This study proposes a new interpretable semantic representation model of text corpus, associative knowledge network model. And, the performance of the proposed model is studied by developing new method for checking the semantic coherence of noun context. The whole framework is divided into two parts. The one part is the modelling process of associative knowledge network. And another part is the process of checking the semantic coherence of noun context. The whole framework of this study can be concluded as the following Fig. 1.

As shown in Fig. 1, the left part is the modelling realization of associative knowledge network. Wherein, the text corpus is first preprocessed to generate noun nodes in "Noun entity node creation". Next, the associative relationships between knowledge nodes are created in "Associative relationship creation", whose associative strength is computed in "Associative strength computation". Then the relationships with strength are incrementally updated to the network in "Incremental updating of associative relationship". Finally, the whole network is reduced and reconstructed to form constructed associative knowledge network. In addition, extra cycles can be performed to learn more texts. And in the right part, a novel interpretable method for the practical problem of checking semantic coherence of noun context is introduced. Here, an associative knowledge network constructed on given text corpus is firstly considered as a background knowledge network. Next, for a given document required to be checked, all noun words are all extracted in "Current document preprocessing", and their multilevel contextual relationships are extracted in "Multilevel contextual relationships acquiring". Then, a group of interpretable semantic features are computed according to the coupling degree from the prior knowledge network to the multilevel contextual relationships of current document in "Associative coupling degree computing". Finally, a classification method is employed to realize the non-coherence error detection of noun context.

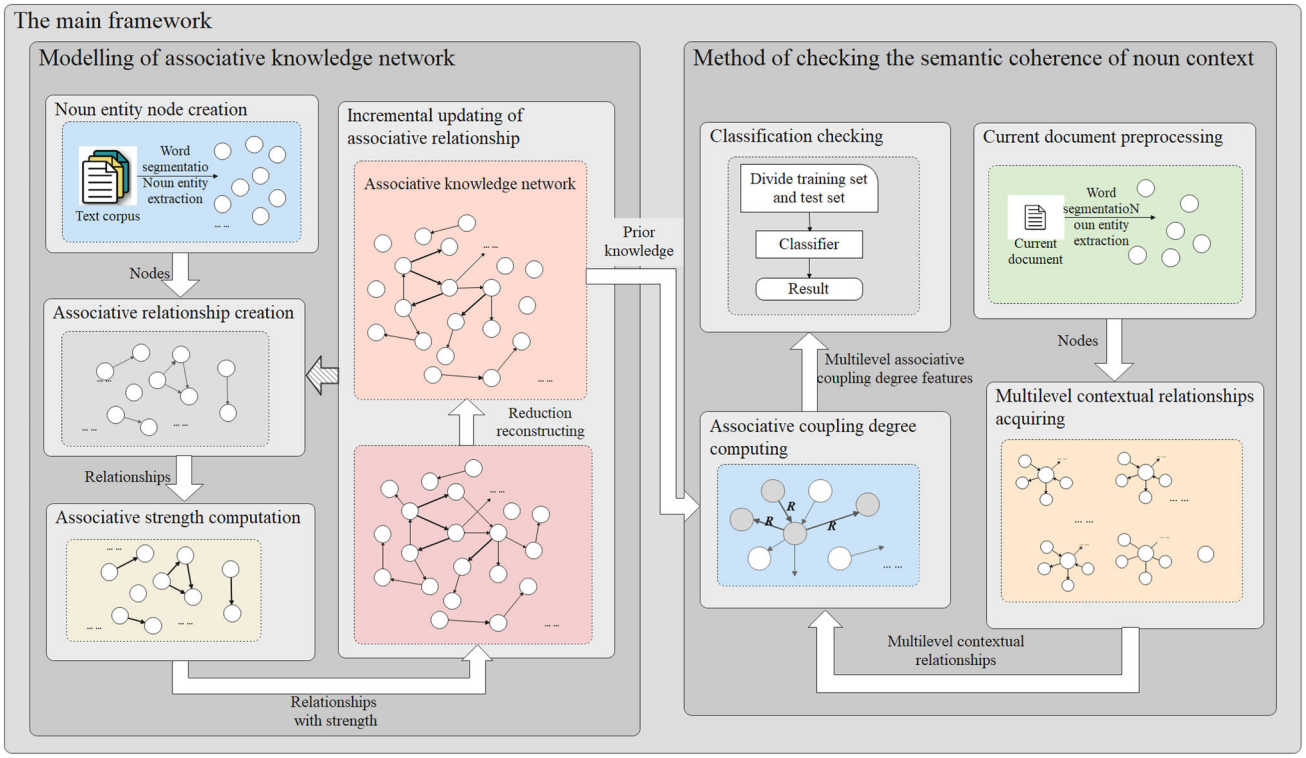


Fig. 1 The framework of this study

Associative knowledge network modelling

Next, from the perspective of associative memory of human brain, the construction process of associative knowledge network will be described in details. Associative memory is a basic way of human brain thinking, which is a process of forming, deleting, and changing the relationship between information neurons. Accordingly, we consider that the main process of associative knowledge network modelling includes the creation of noun entity nodes and associative relationships, the computing of associative strength, and the incremental updating of associative relationship.

Noun entity node creation

The main function of this part is to extract noun entities from given text and to create their mapping as network nodes in associative knowledge network. First, noun words with certain conditions are extracted from the text corpus and then extracted noun entities are directly added as nodes in the associative knowledge network. Before extracting noun entities, the text corpus is preprocessed by word segmentation tools, including sentence extraction, Chinese word segmentation and part-of-speech tagging. Then, only noun entities are extracted from the results of word segmentation and part-of-speech tagging.

Associative relationship creation

Similarly, the main function of this part is to extract the associative relationships from given text and to create their mapping as network relationships in associative knowledge network. Concretely, according to extracted noun entities, the direct pointing relationships are directly created according to the front and back positions of noun entities in sentences. If entity a_1 precedes entity a_2 in a sentence, a direct associative relationship from a_1 to a_2 is created in the associative knowledge network. Here, we think that the entity appearing later in the same sentence is produced associatively by the previous entity, and only when there is a direct pointing relationship between two entities can there be a direct associative relationship. $\langle a_1, a_2 \rangle$ is used to represent a directed direct associative relationship pair, which means that there is a directed edge at which a_1 points to a_2 in the associative knowledge network.

Associative strength computation

The main function of this part is to compute the associative strength when a new associative relationship is extracted. We consider that the associative strength between two adjacent knowledge nodes in an associative knowledge network is related to their co-occurrence times and co-occurrence positions in the text. In addition, a more reasonable computing

method for direct associative strength between knowledge nodes is further designed by developing the quantitative computing method of semantic relevance relation given by Zhong et al. [30] and the definition of associative weight given by Wang et al. [31]. Concretely, we propose the following Definition 1.

Definition 1 In the text corpus with a given statistical window size, direct associative strength R_{ab} between any two noun entities a and b is defined as:

$$R_{ab} = \log \frac{p(a, b)}{p_a * p_b} / \log \frac{2}{p_a + p_b} \tag{1}$$

$p(a, b)$ in the above formula represents the neighbour probability of noun entities a and b in the statistical window; p_a and p_b represent the probabilities of noun entities a and b appearing in the statistical window. Furthermore, it is defined as follows:

$$p(a, b) = \frac{u_{ab}}{u_{all}} \tag{2}$$

$$u_{ab} = \sum_{1 \le k \le q} \frac{1}{I_b\left(\left\langle a^k, b^k \right\rangle\right) - I_a\left(\left\langle a^k, b^k \right\rangle\right)} \tag{3}$$

$$u_{all} = \sum_{xy \in M} u_{xy} \tag{4}$$

$\left\langle a^k, b^k \right\rangle$ indicates a direct associative relationship pair in the statistical window;q represents the sum of the co-occurrence times of knowledge items a and b in the statistical window; I_a and I_b represent the relative position index values of two noun entities, respectively. Obviously, in the same statistical window, the minimum difference value between them is 1. M is the set of all associative relationship pairs in the statistical window. When building the general associative knowledge network model, the statistical window is naturally considered as a natural sentence.

Different from Zhong's semantic relevance relation measurement method, we not only consider the frequency of their co-occurrence but also the relative proximity of two co-occurrence entities in the window when calculating the neighbour probability of two noun entities in the statistical window. That is, when the distance between two entities is closer, their relationship is closer and their strength is greater, and conversely, their strength is smaller. In the experimental part of this paper, a comparative study on above two computing strategies is also executed.

Incremental updating of associative relationship

Furthermore, the main function of this part is to incrementally update the associative strength when a new associative relationship is extracted. Like the process of human brain knowledge updating, associative knowledge network should have dynamic updating ability; that is, knowledge network can be updated incrementally with the increase of learning corpus. However, from the perspective of human brain memory, this incremental updating has not only the addition and enhancement of associative relationships but also the process of weakening, deleting, or forgetting associative relationships. Nevertheless, there is no clear overall consideration in the existing knowledge graph construction strategies, which is also presented in our previous research on knowledge network modelling [31]. Therefore, this study further considers the incremental updating mechanism of associative relationships, that is, with the increase of material texts in the corpus, knowledge nodes can be inserted incrementally, and the strength of new and existing associative relationships can be updated effectively.

Regarding the principle of associative relationship updating between neurons in human brain, Donald, a famous Canadian physiologist, proposed the Hebb learning rule [39]. He believes that the learning process of human brain neural network occurs at synapses between neurons, the strength of synaptic connections changes with the neuronal activity before and after synapses, and the amount of change is proportional to the total activities of two neurons. That is, in a certain period, the connection between activated neurons is strengthened, while the connection between neurons is weakened when two neurons are not activated for a long time. Combining above ideas, if knowledge nodes in associative knowledge network are related to neurons and associative relationships between knowledge nodes are related to synapses connected between neurons, we can give the following strategies for updating knowledge and associative relationships:

(a) Considering the characteristics that neurons in the human brain are "stimulated" and "activated" by the brain, the connection between neurons will be strengthened. When nodes in associative knowledge network increase or are triggered, the associative strength on the corresponding node edges are also enhanced. The corresponding strategy schematic is given in Fig. 2, in which shadow nodes are newly inserted knowledge

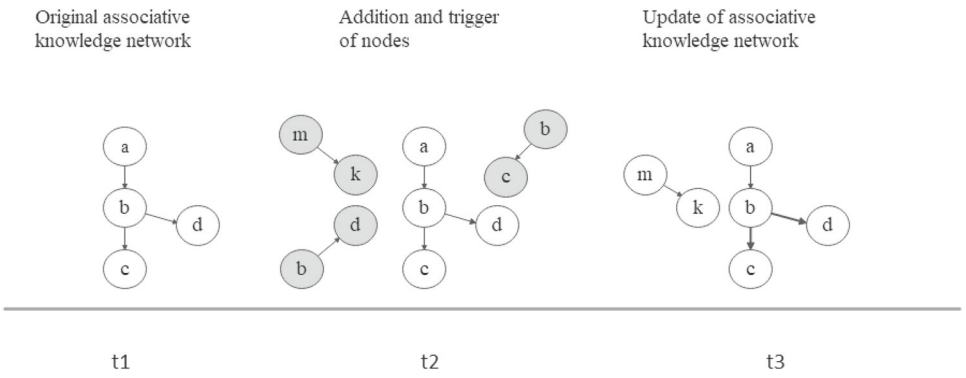


Fig. 2 Node addition or trigger enhancement process in an associative knowledge network

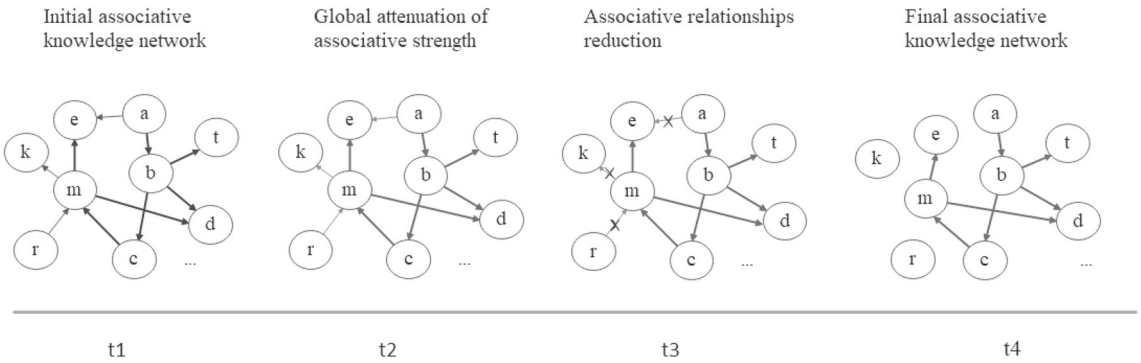


Fig. 3 Global attenuation of associative strength and associative relationship reduction in an associative knowledge network

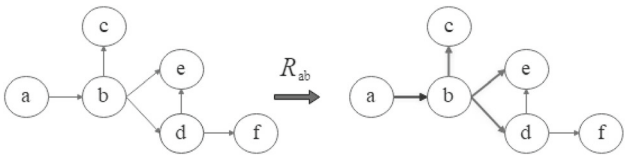


Fig. 4 Chain enhancement process of associative strength between nodes in an associative knowledge network

nodes, and the edge thickness indicates the size of associative strength.

(b) If neurons in the human brain are not "stimulated" for a long time, the "connection" between neurons will be weakened or even "forgotten". We introduce that, in every learning period, global attenuation of associative strength and reduction in associative relationships are carried out one time. Global attenuation simulates the process of "memory weakening" in the human brain, while associative relationship reduction simulates the process of "memory forgetting". The corresponding strategy schematic is given in Fig. 3. The thickness of the edges in the figure indicates the size of associative strength. After global attenuation of associative strength, the strength value of network edges will decrease, while associative relationship reduction will delete those edges with less strength in the network.

(c) In addition, according to the neuron chain reaction characteristics and from the perspective of information dissemination in complex networks [40], it is further considered that in the dynamic "learning" process of associative knowledge network, when nodes are "activated", neighbouring nodes are also "activated", and the corresponding associative strength is also enhanced. The corresponding process schematic is given in Fig. 4, and the thickness of edges in the figure indicates the size of associative strength. When nodes a and b are activated, not only the associative strength between a and b is enhanced but the associative strength of b's direct associative relationships are also enhanced. Wherein, R_{ab} is the strength of edges generated after nodes a and b are activated.

To summarize above discussions, we can give the following associative knowledge network construction algorithm.

Algorithm 1: Construction algorithm of associative knowledge network

Input: Given text corpus, an empty network structure $G=(V,E,U)$ as a combination of a knowledge nodes set, an associative edge set, and an edge strength set.

Output: Associative knowledge network $G'=(V',E',U')$

1. Extract noun entities to get a set *NounSet*
2. For $list_{batch}$ in *NounSet* //Incremental learning
3. For w_i, w_j in $list_{batch}$ //Get neighbour entity pair
4. Create nodes v_i and v_j corresponding to entities w_i and w_j
5. Create a direct associative relationship from node v_i to v_j, and get edge e_{ij}.
6. Calculate the associative strength R_{ij} of v_i and v_j corresponding to entities w_i and w_j by Formula (1)
7. According to v_i, v_j, R_{ij} and e_{ij}, call Algorithm 2 to update associative knowledge network
8. End For
9. //Global attenuation of associative strength
10. For e_{ij} in E
11. Let $R_{ij}^{t+1} = R_{ij}^{t} * \gamma$ (5)
12. End For
13. //Associative relationship reduction
14. If $Count(E) > T$
15. delete $(T - Count(E))$ edges in sort(E)
16. End If
17. End For

Algorithm 2: Dynamic updating of associative relationship

Input: Associative knowledge network $G=(V,E,U)$, nodes v_i and v_j, associative strength R_{ij}, edge e_{ij}

Output: New associative knowledge network $G'=(V',E',U')$

1. If $e_{ij} \in E$
2. Let $R_{ij}^{t+1} = R_{ij} * y_i + R_{ij}^{t}$ (6)
3. Else
4. Insert edge e_{ij} into network G, and let $R_{ij}^{t+1} = R_{ij}$ (7)
5. End If
6. Acquire all direct associative knowledge node sets Set of nodes v_j in the network G
7. For x in Set do
8. Let $R_{jx}^{t+1} = R_{jx} + R_{ij} * x_i * \frac{R_{jx}}{\sum_{s \in Set} R_{js}}$ (8)
9. End for

In Algorithm 1, the global attenuation of associative strength between nodes is executed after learning a batch of material texts for every increment. Concretely, the strengths of all associative edges in the network are multiplied by an attenuation value to simulate the process of memory decline caused by long-term no stimulation of neurons in the human brain, where γ is the attenuation rate and 0.95 is taken as the default according to our empirical analysis.

In a large-scale knowledge graph, there will be many associative edges with weak associative strength (close to 0), so the existence of these edges will lead to unnecessary costs in the knowledge querying process. Therefore, in Algorithm 1, we consider setting the scale of network edges as a constraint capacity value T to simulate the "forgetting" process of the connection between neurons of human brain. The specific rule is that, after learning a batch of material texts, if the total number of edges exceeds the pre-set constraint capacity, the part of edges with smaller associative strength will be deleted directly, so that the total number of associative edges will be less than the constraint capacity value. This process is related to the step 14 to step 16 in the Algorithm 1.

In Algorithm 2, when a new associative edge e_{ij} is generated, if the edge already exists in the network, the corresponding associative strength is updated according to formula (6). If there is no edge e_{ij} in the network, the algorithm adds the edge e_{ij} to the network, and directly updates the associative strength to R_{ij} according to formula (7).

In addition, when edge e_{ij} in the network is updated, because nodes v_i and v_j are activated, the direct associative knowledge of node v_j is also considered to be activated and enhanced. The corresponding update computing is shown in formula (8), where x_i and y_i are learning rates, and $y_i = 0.95$ and $x_i = 0.85$ are taken as default values according to our empirical analysis on complex network.

Method of checking the semantic coherence of noun context

In this section, we take the associative knowledge network as the background knowledge, and judge whether the noun entity is semantic coherence by analysing the differences of their context information between background knowledge and current document.

In text writing, improper use of words in sentences is a common problem, and their semantic coherence checking will be an effective aid for such problems. Table 1 gives some simulated representative sentence examples, in which correctly used noun entities are marked with shadow background and wrong noun entities are marked with a double underline. The examples given in Table 1 include ① redundant words, ② word selection errors, and ③ word ordering errors. To effectively check and find these errors, in this study, we carry out context analysis on each noun entity. That is,

Table 1 Examples of sentences with incoherence context semantic of noun entities

Label	Text sentence
Correct:	杜仲是一种很滋补的药材, 对我们很多的疾病都有很好治疗效果。 Eucommia ulmoides is a very nourishing medicinal material, which has a good therapeutic effect on many of our diseases
Incorrect: ① ②	杜仲是一种很滋补的食物, 对我们很多的疾病都有很好治疗效果作用。 Eucommia ulmoides is a very nourishing food, which has a good therapeutic effect and function on many of our diseases
Correct:	食欲不振的人, 吃龙眼可以得到很好的改善。 People with poor appetite can be well improved by eating longan
Incorrect: ① ③	吃龙眼食欲不振的人, 可以得到很好的作用改善。 People with poor appetite after eating longan can get a good effect improvement
Correct:	对于贫血的人、体质虚弱的人吃龙眼是很有益处的。 Longan is very beneficial for anaemic people and people with weak constitution
Incorrect: ① ②	对于贫血的人、健康虚弱的人吃龙眼水果是很有益处的。 Longan fruit is very beneficial for anaemic people and people with weak healthy
Correct:	大力发展社会主义先进文化。 Vigorously developing advanced socialist culture
Incorrect: ②	大力发展社会主义先进艺术。 Vigorously developing advanced socialist art
Correct:	历史和现实都告诉我们, 法治兴则国兴, 法治强则国强。 Both history and reality tell us that, the prosperous rules of law make the country prosperous, and the strong rules of law make the country strong
Incorrect: ② ③	历史和今天都告诉我们, 国兴则法治兴, 法律强则国强。 Both history and today tell us that, the prosperity of the country makes the rule of law flourish, while the strong of the law makes the country strong
Correct:	中国共产党是中国工人阶级的先锋队, 同时是中国人民和中华民族的先锋队。 The Communist Party of China (CPC) is the vanguard of the Chinese working class, and also the vanguard of the Chinese people and the Chinese nation
Incorrect: ②	中国共产党是中国人民的先锋队, 同时是中国人民和中华民族的未来。 The Communist Party of China (CPC) is the vanguard of the Chinese people, and also the future of the Chinese people and the Chinese nation

we take our associative knowledge network as a background knowledge network to provide empirical knowledge and then take a word as an observation perspective to analyse whether the context words of this word in the current document can effectively support it semantically or interpret it associatively.

Combined with the previous discussions, the method of checking the semantic coherence of the noun context is given below. Specifically, the criterion is whether the contextual relationships of noun entities in the current document have good associative characteristics in the background knowledge network. That is, if the contextual relationships of some noun entities in the current document do not have good associative characteristics in the background knowledge network, it can be considered that the contextual semantic of these noun entities are inconsistent or mismatched. According to the above algorithm principle, our coherence checking method can accurately locate the semantic consistency of a single noun entity in the context instead of giving a rough score of the coherence of the whole sentence or paragraph. Combined with the overall technical framework in this study given in Fig. 1, the following process for checking the contextual semantic coherence of nouns based on an associative knowledge network can be given.

Algorithm 3: Checking the contextual semantic coherence of nouns based on associative knowledge network

Input:	Given the current document
Output:	Coherence checking and positioning results of noun entities
1.	Preprocess a given document to extract noun entities.
2.	Acquire multilevel contextual relationships of noun entities.
2.1	Extract contextual relationships in a sentence.
2.2	Extract contextual relationships from the front and rear sentences.
2.3	Extract contextual relationships among sentences in a paragraph.
3.	Carry out the associative computing on background knowledge network, and obtain multilevel associative coupling degree features of noun entities.
4.	Coherence checking based on a classification method.
5.	Get the check result of the whole given document.

Current document preprocessing

This part is to preprocess the detection document. First, sentence extraction, Chinese word segmentation and part-of-speech tagging are performed on the current document, and then noun entities are extracted. Different from natural sentence extraction in the building module of an associative knowledge network, in this module, a short sentence extraction method is proposed to avoid inaccurate contextual entity relationships caused by sentences that are too long. That is, comma extraction is added to traditional sentence extraction.

Acquisition of multilevel contextual relationships of noun entities

This part is to extract multilevel contextual relationships of noun entities in the given detection document. In general, the semantic coherence of noun entities in document is related to the location of entities and other entities in context. To evaluate the semantic coherence of a noun entity in a document, it is necessary to obtain the contextual relationships of this noun entity at first, that is, to determine the context-related entities of this noun entity from different perspectives and to form multi-perspective correlation pairs related to this noun entity. When obtaining context-related entities, we consider the following three perspectives:

(1) Intra-sentence relevance. The associative contextual relational network is constructed inside the current sentence to obtain the context-related nouns of noun entities. Considering a short sentence sequence $S = a, b, c, d, e$ in detection document, the intra-sentence associative contextual relational network constructed by S is shown in Fig. 5a, in which the contextual relationships obtained by noun entity c are shown in Fig. 5b, and the corresponding correlation pair is $Pair_c = \{\langle a, c\rangle, \langle c, e\rangle, \langle b, c\rangle, \langle c, d\rangle\}$.

(2) Inter-sentence relevance. At first, two short sentences before and after the current sentence are taken to construct an associative contextual relational network. For a short sentence sequence $S_q = a, b, c, d, e$ in the current detection document, Fig. 6a shows two short sentences before and after the short sentence S_q, Fig. 6b shows the associative contextual relational network based on inter-sentences, and Fig. 6c shows contextual relationships obtained by noun entity e. Then, the correlation pair of e is $Pair_e = \{\langle e, k\rangle, \langle e, r\rangle, \langle e, f\rangle, \langle t, e\rangle, \langle a, e\rangle, \langle b, e\rangle, \langle c, e\rangle, \langle d, e\rangle\}$.

(3) Intra-paragraph relevance. First, the paragraph containing the target noun entity is located, then other nouns in the paragraph are taken as the context of the target noun, and several sets of contextual relationships of the target noun in the paragraph are extracted. Let $Pks = \{k_1, k_2, ..., k_n, k_m\}$ represents the set of noun entities in a paragraph. For noun entity $k_n \in Pks$, we can obtain multiple groups of correlation pairs $Mu(Pair_{k_n}) = \left\{\{(k_n, k_i)\}_{1\le i\le m, i\ne n}\right\}$ based on

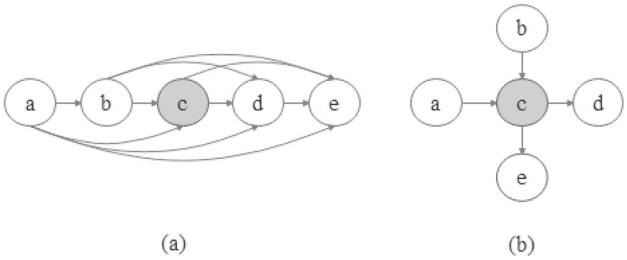


Fig. 5 Schematic diagram of extracting correlation pair of intra-sentences

an intra-paragraph, and the contextual relationships in the paragraph do not consider the directionality of edges.

Associative coupling degree computing

To quantitatively evaluate the semantic coherence of noun entities in a document with given background knowledge network, this part introduces the associative coupling degree computing strategy. In the previous section, how to extract the correlation pairs of target noun entities has been described. Next, how to further compute the multilevel associative coupling degree features of target noun entities in the background knowledge network G is discussed. Concretely, the correlation pairs of the target noun entity are mapped to the background knowledge network, and whether these relation pairs have direct associative relationships in the background knowledge network are queried. Obviously, if there is a direct associative relationship, it can show that this correlation pair has a good associative experience in the background knowledge network; that is, the target noun is more coherent in context. Figure 7 shows the associative computing process of a noun entity in the background network, in which the correlation pair of noun entity f is $Pair_f = \{\langle f, k\rangle, \langle f, h\rangle, \langle f, v\rangle, \langle e, f\rangle, \langle b, f\rangle\}$, the bold edges e_{ef}, e_{fv} and e_{fh} in Fig. 7 indicate that, the correlation pair of entity f has direct associative relationships in background network G. Ra_{ef}, Ra_{fv} and Ra_{fh} are the associative strength values

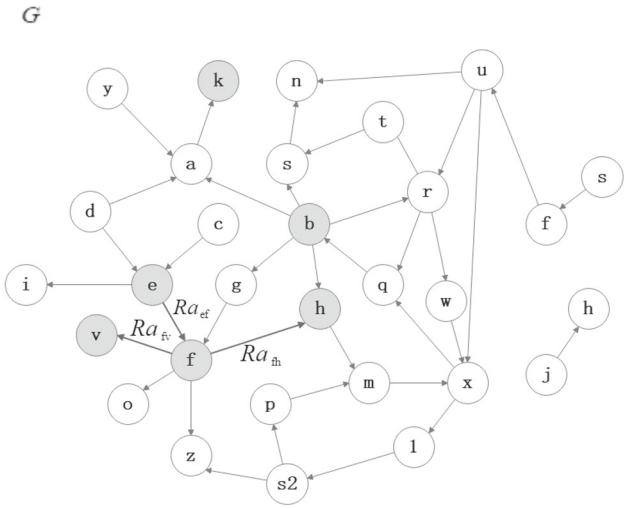


Fig. 7 Associative computing process of the correlation pair of noun entity f in the background knowledge network

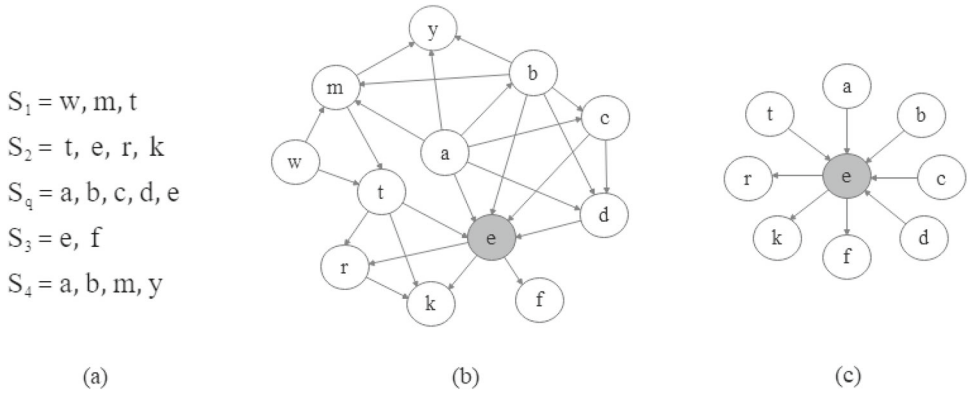
on the corresponding edges. In above computing process, the edge directionality is not considered for the intra-paragraph correlation pairs.

Furthermore, to quantitatively evaluate the semantic coherence of the target noun in context, a computing method of the multilevel associative coupling degree features is further designed as follows.

Definition 2 Let a correlation pair of a noun entity k_i be $Pair_{k_i} = \{\langle k_1, k_i\rangle, \langle k_2, k_i\rangle, ..., \langle k_m, k_i\rangle\}$, $i \neq m$. Then, its associative coupling degree feature in the background knowledge network G is computed as follows:

$$Vacd(k_i) = \frac{\sum_{\langle k_n, k_i\rangle \in (Pair_{k_i} \cap G)} Ra_{k_n k_i}}{\sum_{\langle k_n, k_i\rangle \in (Pair_{k_i} \cap G)} 1} * \log_2\left(1 + \sum_{\langle k_n, k_i\rangle \in (Pair_{k_i} \cap G)} 1\right) \tag{9}$$

Fig. 6 Schematic diagram of extracting correlation pair of inter-sentences



In the formula, $\langle k_n, k_i \rangle \in \left(Pair_{k_i} \cap G\right)$ indicates that entity k_i has a direct associative relationship in background network G, and $Ra_{k_n k_i}$ represents the associative strength value of edge $e_{k_n k_i}$ in the background knowledge network.

Multilevel associative coupling degree features

This part further expands the multilevel associative coupling degree computing methods. By acquiring the contextual relationships of noun entities at multiple levels, we can obtain multiple groups of correlation pairs of noun entities including inside a sentence, between sentences, and in a same paragraph. Further, by mapping each group of correlation pairs to the background network for associative computing, we can obtain multiple groups of associative coupling degree features corresponding to the noun entity. For the assumption that there are multiple groups of correlation pairs of noun entity k, the associative coupling degree feature $Vacd$(k) inside the sentence is called $Vacd_{inside}$, and the associative coupling degree feature $Vacd(k)$ between sentences is called $Vacd_{between}$. Moreover, we call $Vacd_{inside}$ and $Vacd_{between}$ basic features. Additionally, the associative coupling degree features $\{Vacd_1, Vacd_2, Vacd_3, ..., Vacd_n, ..., Vacd_m\}$ can be obtained based on multiple groups of intra-paragraph correlation pairs, in which the value sequence is sorted from largest to smallest and m is related to the number of noun entities in a paragraph. We take the top n $Vacd$ values as paragraph features. In summary, the features $\{Vacd_{inside}, Vacd_{between}, Vacd_1, Vacd_2, Vacd_3, ..., Vacd_n\}$ of the multilevel associative coupling degree features of noun entity k can be used. In the experiment, we will study the influence of the number n to the method performance.

Coherence checking using interpretable classification

For coherence checking using interpretable classification decision, we simply use an interpretable classification method decision tree [41] to judge the coherence based on the multilevel associative coupling features. In the experimental part, more details will be discussed.

Experimental methods

Method parameters

The method parameters mainly include the constraint capacity value T of associative knowledge network and the number n of paragraph features. In the following experimental analysis, we will discuss the influence of these parameters on the method performance.

Table 2 Dataset size used in the experiments

	Corpus I	Corpus II
Quantity of text for constructing a background knowledge network	10,697	7149
Quantity of text for coherence checking	100	100

Evaluation metrics

In this study, precision (P), recall (R) and F1-score (F1) are considered as performance evaluation metrics, and the corresponding definitions are as follows:

$$P = \frac{|M \cap B|}{|M|} * 100 \tag{10}$$

$$R = \frac{|M \cap B|}{|B|} * 100 \tag{11}$$

$$F1 = \frac{P \times R}{(P + R)/2} * 100 \tag{12}$$

wherein, M is the output result of the classification method, and B is the result of the test sample. P can measure the precision of the model's error detection, R can measure the information coverage of the model's error detection, and F1 can balance the influence of P and R. Moreover, time complexity and space complexity are also used as metrics to measure the model performance in subsequent analysis.

Experimental datasets

In this study, we introduced two experimental corpus datasets. The first dataset is 10,797 texts related to the diet on the topics "healthy knowledge", "dietary nutrition" and "dietary errors", which are crawled from "Meishi-Baike" [42] and "Foodbk" [43] and recorded as Corpus I. The second dataset is 7249 texts provided by Yozosoft, which comes from the party and government corpus of various provinces in the "National Learning Platform Exhibition and Broadcast" module crawled from the official website of "Xuexi.cn" [44] and is recorded as Corpus II.

As the research method includes constructing a background knowledge network and the coherence checking application, the experimental data quantity of these two parts is shown in Table 2.

After constructing associative knowledge network, the network on Corpus I owns 102,942 nodes and 5,024,139 edges. And for Corpus II, the network owns 43,576 nodes and 4,136,888 edges.

In addition, in the coherence checking experiment, we also need to build incorrect samples. We consider randomly

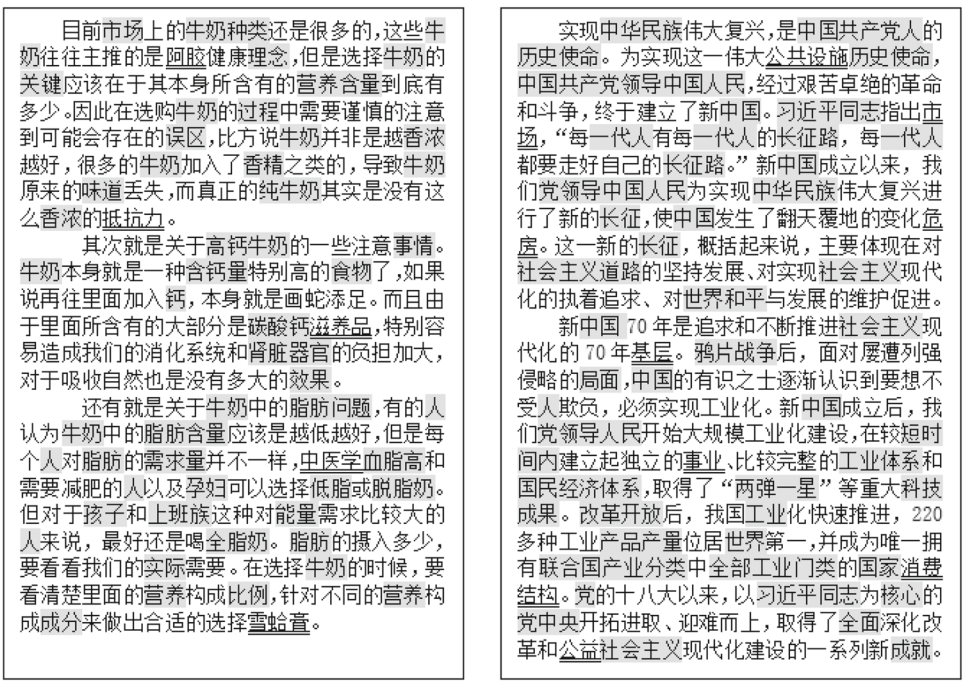
inserting 1800 noun entities into 100 documents of two corpora as context semantic inconsistency nouns in text, namely, negative sample data, in which randomly inserted nouns are uniformly taken from the noun set in the background knowledge network. To ensure that the semantic information of nouns in the original text is not changed in the process of randomly inserting, a noun entity is inserted every 2–3 sentences. The corpora after insertion are called Dataset I and Dataset II. Figure 8 shows a typical result of randomly inserting noun entities into text paragraphs from the corpus [43, 44], in which the shaded part is the existing noun entities in text, and the double-underline denotes the noun entities we randomly inserted. The left sample text in the figure is taken from Dataset I, and the right sample text is taken from Dataset II.

Numerical results and discussions

In this part, a group of numerical results will be reported. In the simulation experiments, based on the Dataset I and Dataset II constructed above and the multilevel coupling degree features extraction method, the decision tree model is introduced to judge the semantic coherence of the noun context. In addition to using most of materials in the corpus to construct a background knowledge network, for the training of classification model, the positive sample is from those noun entities already existing in original texts, and the negative sample data are constructed by randomly inserting noun entities in original texts. Because the number of original positive samples in Dataset I and Dataset II is far greater than the number of negative samples, the random under-sampling method is adopted to randomly sample the positive sample data to maintain the balance between the positive and negative sample data.

In the experimental analysis of checking the semantic coherence of noun context, all comparative experiments are performed by five-fold cross-validation for performance analysis. The output performance of the model takes the mean and mean square deviation of five experimental performance metrics.

Tables 3 and 4 show some results of the multilevel associative coupling degree features of noun entities in the example text of Fig. 8. The double underlined parts in the table are error entities with inconsistent context semantic. By analysing the data in Tables 3 and 4, it can be found that the associative coupling degree features of noun entities with correct semantic in document is usually greater than 0, while the associative coupling degree features of noun entities with wrong context and incoherent semantic usually has more 0 values. This result is very in line with our intuitive cognition. That is, for a noun entity with correct semantic meaning in text, its contextual words can effectively support the semantic meaning of this noun, and it must also have good associative interpretation ability in the background network. However, for the wrong entity with incoherent context semantic, the semantic support ability of its contextual words to itself is weak, and it is usually impossible to obtain better associative characteristics in the background knowledge network.


目前市场上的牛奶种类还是很多的，这些牛奶往往主推的是阿胶健康理念，但是选择牛奶的关键应该在于其本身所含有的营养含量到底有多少。因此在选购牛奶的过程中需要谨慎的注意到可能会存在的误区，比方说牛奶并非是越香浓越好，很多的牛奶加入了香精之类的，导致牛奶原来的味道丢失，而真正的纯牛奶其实是没有这么香浓的抵抗力。

其次就是关于高钙牛奶的一些注意事情。牛奶本身就是一种含钙量特别高的食物了，如果说再往里面加入钙，本身就是画蛇添足。而且由于里面所含有的大部分是碳酸钙滋养品，特别容易造成我们的消化系统和肾脏器官的负担加大，对于吸收自然也是没有多大的效果。

还有就是关于牛奶中的脂肪问题，有的人认为牛奶中的脂肪含量应该是越低越好，但是每个人对脂肪的需求量并不一样，中医学血脂高和需要减肥的人以及孕妇可以选择低脂或脱脂奶。但对于孩子和上班族这种对能量需求比较大的人来说，最好还是喝全脂奶。脂肪的摄入多少，要看看我们的实际需要。在选择牛奶的时候，要看清楚里面的营养构成比例，针对不同的营养构成成分来做出合适的选择雪蛤膏。

实现中华民族伟大复兴，是中国共产党人的历史使命。为实现这一伟大公共设施历史使命，中国共产党领导中国人民，经过艰苦卓绝的革命和斗争，终于建立了新中国。习近平同志指出市场，“每一代人有每一代人的长征路，每一代人都要走好自己的长征路。”新中国成立以来，我们党领导中国人民为实现中华民族伟大复兴进行了新的长征，使中国发生了翻天覆地的变化危房。这一新的长征，概括起来说，主要体现在对社会主义道路的坚持发展、对实现社会主义现代化的执着追求、对世界和平与发展的维护促进。

新中国 70 年是追求和不断推进社会主义现代化的 70 年基层。鸦片战争后，面对屡遭列强侵略的局面，中国的有识之士逐渐认识到要想不受人欺负，必须实现工业化。新中国成立后，我们党领导人民开始大规模工业化建设，在较短时间内建立起独立的事业、比较完整的工业体系和国民经济体系，取得了“两弹一星”等重大科技成果。改革开放后，我国工业化快速推进，220 多种工业产品产量位居世界第一，并成为唯一拥有联合国产业分类中全部工业门类的国家消费结构。党的十八大以来，以习近平同志为核心的党中央开拓进取、迎难而上，取得了全面深化改革和公益社会主义现代化建设的一系列新成就。

Fig. 8 Two examples obtained by randomly inserting noun words in text paragraphs

Table 3 Multilevel associative coupling degree features of some noun entities in the left example text in Fig. 8

Name		$Vacd_{inside}$	$Vacd_{between}$	$Vacd_1$	$Vacd_2$	$Vacd_3$	$Vacd_4$
市场	Market	19.273	19.273	12.628	11.691	11.076	6.4567
牛奶	Milk	97.443	12.715	206.32	112.00	28.551	20.597
种类	Type	11.973	11.973	27.971	14.623	11.691	10.680
牛奶	Milk	97.443	1.3333	206.32	112.00	28.551	20.597
阿胶	Donkey-hide gelatine	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
理念	Idea	1.3333	1.3333	5.6206	2.4995	1.5000	1.3333
牛奶	Milk	97.443	134.87	206.32	112.00	28.551	20.597
关键	Key	9.1841	9.1841	12.968	9.9160	8.5892	3.7447
营养	Nutrition	154.41	154.41	112.00	109.71	66.647	32.737
含量	Content	132.22	132.22	206.32	109.71	27.971	19.570
…		…	…	…	…	…	…
抵抗力	Resistant	0.0000	0.0000	9.3108	4.8435	1.0000	0.0000
高钙	High calcium	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
牛奶	Milk	42.992	0.0000	79.439	31.765	22.357	13.311
事情	Thing	0.0000	0.0000	5.9122	2.3333	1.0000	0.0000
牛奶	Milk	28.479	28.479	79.439	31.765	22.357	13.311
含钙量	Calcium content	5.3162	5.3162	13.311	6.7186	2.5000	1.0000
食物	Food	26.531	26.531	58.442	37.833	31.765	5.9122
钙	Calcium	0.0000	1.0000	79.439	37.833	8.9932	6.7186
碳酸钙	Calcium carbonate	0.0000	0.0000	2.1067	1.0000	0.0000	0.0000
滋养品	Nourishment	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
…		…	…	…	…	…	…
中医学	Traditional Chinese medicine	0.0000	0.0000	1.0000	1.0000	1.0000	0.0000
…		…	…	…	…	…	…
雪蛤膏	Snow clam paste	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000

Based on above experimental results, we will further quantitatively analyse the performance impact of different parameters and carry out performance comparison to existing methods. And for our proposed method of checking the semantic coherence of noun context for given text, a method name AssoCheck is used for the convenience of description.

Performance analysis on different paragraph feature numbers

In this section, we will analyse the performance influence of different paragraph feature number n. The F1-score difference value is used to evaluate whether the comprehensive performance of the model is improved after adding paragraph features. The F1-score difference value is the F1-score of the model after adding paragraph features minus the F1-score of the model with only basic features. The experimental results are shown in Fig. 9, and the detailed analysis is as follows.

Figure 9 reflects the change in the F1-score difference value before and after adding paragraph features, and the abscissa shows the number of paragraph features. By analysing Fig. 9, compared with only basic features, the error detection performance of the model is improved to varying degrees after adding paragraph features. It can be seen in the figure that the comprehensive performance is best when the number of paragraph features in both datasets is 4. When there are too many paragraph features, the performance of the model will decline instead, which may be due to random interference caused by too many feature quantities. Therefore, we suggest that the number of paragraph features n is set to 4 by default.

Furthermore, by observing Table 5, it can be found that the performance of the model improved after adding paragraph features in both Dataset I and Dataset II. In Dataset I, after adding paragraph features, the F1-score increased by 0.82 performance points, and in Dataset II, the F1-score increased by 0.65 performance points. So, paragraph features are valuable in the coherence checking method, and when n is set to 4, the comprehensive performance is the highest.

Table 4 Multilevel associative coupling degree features of some noun entities in the right example text in Fig. 8

Name		$Vacd_4$	$Vacd_{inside}$	$Vacd_{between}$	$Vacd_1$	$Vacd_2$	$Vacd_3$
中华民族	Chinese nation	6.6407	6.6407	11.851	9.0341	6.1143	2.6826
中国共产党	Chinese Communist	11.343	11.343	33.167	18.969	15.120	9.6730
人	People	5.8105	5.8105	7.4738	4.1817	3.5780	3.4692
历史使命	Historical mission	80.034	80.034	86.296	52.163	43.588	33.167
公共设施	Public facilities	0.0000	0.0000	1.9668	1.0000	0.0000	0.0000
历史使命	Historical mission	15.565	15.565	14.709	11.529	1.5750	1.4210
中国共产党	China Engineering party	24.830	0.0000	4.9653	2.4979	1.0000	0.0000
领导	Leader	2.9734	2.9734	2.2328	1.5893	1.0000	0.8201
…		…	…	…	…	…	…
市场	Market	3.6933	1.0000	15.275	5.3381	3.5209	3.1440
一代人	A generation	42.909	42.909	71.694	55.993	16.068	12.536
一代人	A generation	0.7358	0.7358	1.0000	0.7358	0.6734	0.5329
长征路	Long march road	57.313	4.2077	222.94	37.288	34.500	18.595
…		…	…	…	…	…	…
危房	Dilapidated houses	0.0000	0.0000	1.0000	0.7597	0.5794	0.2893
基层	Grass roots	11.828	11.828	11.828	1.0000	0.0000	0.0000
…		…	…	…	…	…	…
事业	Cause	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
…		…	…	…	…	…	…
消费结构	Consumption structure	0.0000	0.0000	1.2438	1.0000	0.8161	0.0000
…		…	…	…	…	…	…
公益	Public welfare	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000

From the above analysis, we can conclude that the proper introduction of paragraph features enables noun entities to acquire richer contextual semantic information, thus improving the error detection performance of the method.

Performance influence under different capacity scales of background knowledge network

In the next experiment, we will consider the influence of the constraint capacity ratio r, which represents the current network edge capacity T divided by the original scale of the background knowledge network. Wherein, the original scale of the background knowledge network refers to the network formed without any connection edge deletion. The experiment is also carried out on Dataset I and Dataset II, and the F1-score difference value is still used to evaluate the performance influence. The experimental results are shown in Fig. 10.

Figure 10 shows the performance changes by setting different constraint capacity ratio r for Dataset I and Dataset II. By analysing the change curve of Dataset I, it can be seen that as the constraint capacity ratio r gradually decreases, the F1-score difference value gradually increases in the beginning. That is, the model has better comprehensive performance. However, when the constraint capacity ratio r is further reduced, the F1-score difference value of the model begins to decrease in reverse until a negative effect is appeared. We can think that some meaningless connections in the network are removed to a certain extent by properly restricting the scale of background knowledge network edges, thus improving the comprehensive performance of the model. While,

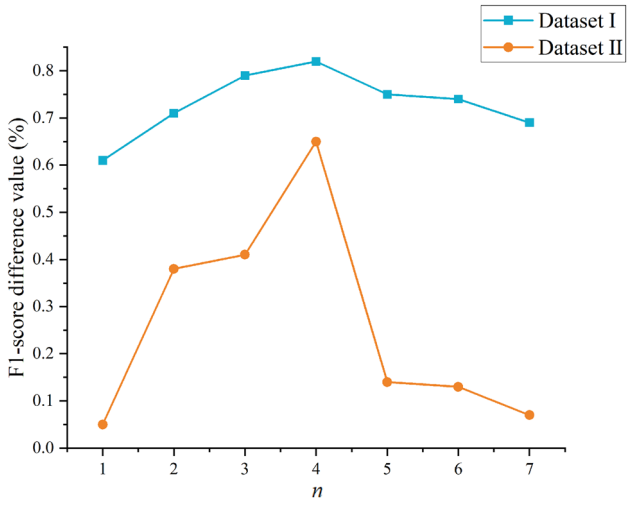


Fig. 9 Influence of the number of paragraph features n on model performance

Table 5 F1-score values obtained by two compared models with no and added paragraph features

	Only basic features	Add 4 paragraph features
Dataset I	91.92 ± 0.53	**92.74 ± 1.12**
Dataset II	91.70 ± 0.65	**92.35 ± 0.97**

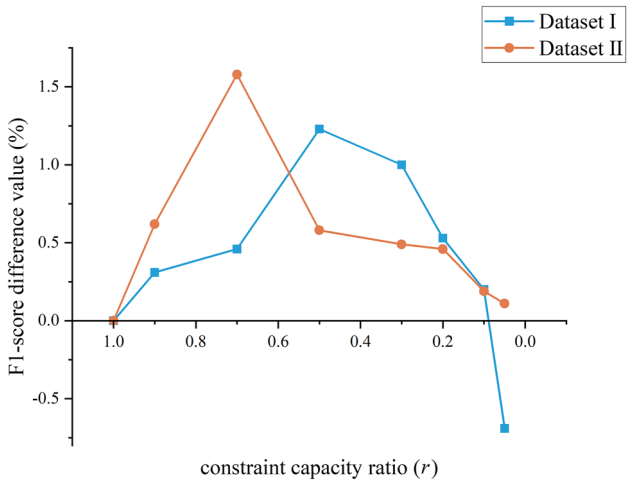


Fig. 10 Performance comparison with different constraint capacity ratios for background knowledge network

if the degree of scale limitation is too large, some necessary connection edges will be discarded, and some necessary semantic connections will be ignored, which will reduce the semantic representation ability of the model. As shown in Fig. 10, the model has the best modelling performance for Corpus I when the constraint capacity ratio r is 0.5. However, for Dataset II, when the constraint capacity ratio is 0.7, the model has the best modelling performance.

Table 6 Performance comparison using different relationship measurements on Dataset I

	AssoCheck	Zhong [30]	Wang [31]
P (%)	92.27 ± 0.67	92.03 ± 0.67	93.32 ± 0.67
R (%)	95.73 ± 1.13	91.51 ± 0.67	91.36 ± 0.67
$F1$ (%)	**93.97 ± 0.79**	91.75 ± 0.67	92.33 ± 0.67

Table 7 Performance comparison using different relationship measurements on Dataset II

	AssoCheck	Zhong [30]	Wang [31]
P (%)	93.49 ± 0.62	92.20 ± 0.87	92.66 ± 0.82
R (%)	94.38 ± 0.83	91.46 ± 1.01	93.05 ± 1.24
$F1$ (%)	**93.93 ± 0.58**	91.82 ± 0.60	92.85 ± 0.9

Performance analysis using different relationship measurements

Here, the performance impact of different knowledge relationship measurement strategies is further analysed. As comparison, two related measurements Zhong [30] and Wang [31] are considered, wherein their relationship strength computing equations are used to compute the associative relationship strength of our model. Corresponding experimental results are reported in Tables 6 and 7 related to Dataset I and Dataset II respectively.

Above results clearly indicate that our proposed measurement strategy can gain slightly better performance than two compared strategies. Accordingly, we can think that our measurement strategy can capture the semantic relationship between noun entities more effectively.

Performance analysis of comparable methods

In this part, performance analysis of the method AssoCheck will be examined compared to two following neural network methods. In the experiment, fivefold cross-validation is also used.

1. **ERNIE** [45]. In 2019, Baidu put forward the ERNIE 1.0 pretraining model inspired by the masking strategy of BERT, in which BERT's random masking strategy is replaced by entity-level or phrase-level masking strategy. In our experiments, original text sentences are extracted from Dataset I and Dataset II as positive samples, and then negative sentences are constructed by inserting entities with inconsistent context semantic in positive sample sentences. Based on

Table 8 Comparison of error detection performance of different methods on Dataset I

	AssoCheck	ERNIE [45]	SoftMB [38]
P (%)	92.27 ± 0.67	91.39 ± 0.72	76.68 ± 0.15
R (%)	95.73 ± 1.13	96.47 ± 4.99	74.44 ± 0.63
*F*1 (%)	**93.97 ± 0.79**	93.84 ± 2.74	75.54 ± 0.39

Table 9 Comparison of error detection performance of different methods on Dataset II

	AssoCheck	ERNIE [45]	SoftMB [38]
P (%)	93.49 ± 0.62	86.02 ± 1.34	82.97 ± 3.23
R (%)	94.38 ± 0.83	92.56 ± 0.27	72.21 ± 2.05
*F*1 (%)	**93.93 ± 0.58**	89.17 ± 0.59	77.21 ± 2.57

above samples and ERNIE 1.0 training model, semantic inconsistency of every sentence can be recognized. In the experiment, we set the batch size as 4, the learning rate as 5e-5 and the epoch as 1 respectively.

2. **SoftMB** [38]. Researchers from ByteDance and Fudan University proposed a new model framework for Chinese spelling error correction in 2020, soft-masked BERT. We quote the first part of the framework, the detection network, as comparison method. The detection network is a bidirectional GRU model. The input is a sequence of sentences, and the output is a classification label. It encodes each sentence sequence bidirectionally to obtain bidirectional hidden states. Then, the hidden states in two directions are spliced and sent to the fully connected layer to obtain a probability value between 0 and 1. In the experiment, we consider that the probability value with greater than 0.5 is related to the wrong word. And in the experiment, the batch size is set to 16, the embedding size is set to 256, and the number of layers is 2. Based on the above setting, Tables 8 and 9 give comparative experimental results on Dataset I and Dataset II, respectively.

According to results in Tables 8 and 9, our method shows the best comprehensive performance. In addition, our method is an interpretable semantic modelling method compared to the advanced neural network semantic modelling method, and it has no disadvantage in performance too.

Besides, although the comprehensive performance of ERNIE is relatively good, its sentence checking and judging can only give a judging result for the whole sentence but cannot locate which words are inconsistent in contextual semantic. For SoftMB, although semantic consistency can be judged for every position in the sentence, its overall performance is significantly lower than other methods. In contrast, the method AssoCheck can not only check the semantic coherence of words in each specific position but also accurately locate misused words, and the overall performance results are also very high.

Table 10 Computational complexity comparison of two methods by using Dataset I

	Training stage (3000 texts) time/memory usage	Detecting stage (10 texts) time/memory usage
AssoCheck	5.0 h/170.700 MB	108.5 s/85.504 MB
SoftMB [38]	2.9 h/336.289 MB	26.40 s/22.427 MB

Complexity analysis

In this section, we will further study the computational complexity of our method AssoCheck by compared to the neural network-based method SoftMB, in which Dataset I is used. In experiment, the computing requirements of dynamically updating 3000 texts with 340,770 words in training stage, and 10 texts in detecting stage. Detailed results are shown in Table 10.

According to the results in Table 10, we can find that the computational complexity should be higher than the neural network method. However, we can think that, deep neural network has been a very compact computational structure and been effectively optimized by using GPU. In contrast, our proposed method has no further computational optimization, and even so, their computational complexity is at the same level.

Extended experimental analysis

In previous experiments, the decision tree is considered to realize the judgment of noun coherence checking. Here, we further consider the influence of different classification algorithms on the error detection performance of our proposed detection method. The compared classification algorithms include SVM [46], KNN [47], Random Forest (RF) [48], Multilayer Perceptron (MLP) [49], and decision tree (DT) method. Based on the same datasets and experimental methods of the previous experiments, the results are shown in Table 11.

In experimental settings, for decision tree, entropy is used as the attribute selection measure. For SVM, linear kernel function is adopted. For KNN, the number of nearest neighbours is set to 3. For Random Forest, entropy is used as the attribute selection measure too, and the number of trees in the forest is 100. For MLP, the batch size is set to 16, the

Table 11 Error detection performance $F1$ (%) within different classification algorithms

	DT [41]	SVM [46]	KNN [47]	RF [48]	MLP [49]
Dataset I	**93.97 ± 0.79**	93.49 ± 0.52	93.76 ± 0.69	93.88 ± 0.52	94.39 ± 0.45
Dataset II	**93.93 ± 0.58**	91.91 ± 0.78	93.86 ± 0.47	93.89 ± 0.42	94.50 ± 1.02

learning rate to 0.001 and the numbers of hidden neurons to 128, 64 and 32 respectively.

From the Table 11, the classification methods based on decision tree and MLP can obtain better classification performance on both datasets. However, from the perspective of interpretable semantic modelling, we think that the classification method based on decision tree is more in line with our proposed task. Besides, according to the results shown in Table 11, all classification algorithms can achieve good performance. This results again demonstrate the effectiveness of out proposed semantic coherence checking method.

Discussions on meta-heuristic algorithm to enhance associative knowledge network modelling performance

The nodes and edges in the associative knowledge network are crucial to the acquisition and dissemination of semantic information. At present, meta-heuristic algorithms [50] are widely used in practical problems. For associative knowledge network modelling, we think that meta-heuristic algorithms will be effective to optimize the representation and dissemination of semantic information of network nodes and edges. It will be further expanded in our future work.

Conclusion and future work

Inspired by what the human brain has a strong pure associative computing ability, an associative knowledge network is proposed for the semantic representation of noun context. Moreover, a completely novel and highly interpretable method is proposed for checking the contextual semantic coherence of noun words in a document, which is very valuable for error detection in text writing. By introducing existing comparable related methods, the rich experimental analysis results show that, the proposed method has better performance in F1-score metric than the methods based on deep neural network. In addition, the proposed model has incomparable advantages in natural interpretability and incremental learning ability.

Even so, in the construction of associative knowledge network, there is no strong theoretical basis for the computing of edge strength, which is mainly supported by research experience. In the future, we will explore stricter logical base in computing associative relationship strength. In addition, the proposed coherence checking method can only detect and locate those noun words with non-coherence context, but can't make them correct modification. This also be further studied.

Author contributions Proposed an interpretable semantic representation model of noun context, and demonstrated its effectiveness by developing a novel method of checking semantic coherence of noun context for a detection document and known text corpus.

Funding This work was supported in part by the grants from NSFC of China (Grant no. 61872166), Science and Technology Planning Project of Wuxi (Grant no. G20201004) and Six Talent Peaks Project of Jiangsu Province of China (Grant no. 2019 XYDXX-161).

Availability of data and materials The data used to support the findings of this study will be available from the corresponding authors.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix 1

The corresponding English translations of two text paragraphs in Fig. 8.

At present, there are many types of milk in the market. These milk often promote the health concept of donkey-hide gelatin, but the key of choosing milk should be how much nutrition it contains. Therefore, in the process of buying milk, you need to be cautious about possible misunderstandings. For example, the more fragrant the milk, the better the quality may not be. A lot of milk is added with flavors, which causes the original taste of milk to be lost, but the real pure milk is actually not so fragrant and resistant.

Secondly, there are some things to note about high-calcium milk. Milk itself is a food with a particularly high calcium content. If you add calcium to it, it is superfluous in itself. And because most of it contains calcium carbonate nourishment, it is particularly easy to increase the burden on our digestive system and kidney organs, and it has little effect on absorption.

There is also the problem of fat in milk. Some people think that the fat content in milk should be as low as possible, but everyone's demand for fat is different. Traditional Chinese Medicine people with high blood fat and those who need to lose weight and pregnant women can choose low-fat or skim milk. But for children and office workers who have a greater need for energy, it is best to drink whole milk. The amount of fat intake depends on our actual needs. When choosing milk, it is necessary to see the nutritional composition ratio inside, and make the appropriate choice of snow clam paste for different nutritional composition.

Realizing the great rejuvenation of the Chinese nation is the historical mission of the Chinese Communists. In order to realize this great historical mission of public facilities, the Chinese Communist Party led the Chinese people to finally establish New China after arduous revolutions and struggles. Comrade Xi Jinping pointed out that in the market, "every generation has a long march for every generation, and every generation must walk its own long march." Since the founding of New China, our party has led the Chinese people in a new long march to realize the great rejuvenation of the Chinese nation, which has caused earth-shaking changes in China's dilapidated houses. In a nutshell, this new long march is mainly embodied in the persistence and development of the socialist road, the persistent pursuit of realizing socialist modernization, and the maintenance and promotion of world peace and development.

The 70 years of New China have been the 70-year grassroots that pursued and continuously promoted socialist modernization. After the Opium War, in the face of repeated aggressions by foreign powers, people of insight in China gradually realized that in order not to be bullied by others, they must realize industrialization. After the founding of New China, our party led the people to begin large-scale industrialization, established an independent enterprise, a relatively complete industrial system and a national economic system in a relatively short period of time, and achieved major scientific and technological achievements such as "two bombs and one star". After the reform and opening up, our country's industrialization progressed rapidly, and the output of more than 220 industrial products ranked first in the world. And it has become the only country's consumption structure that has all the industrial categories in the United Nations Industrial Classification. Since the 18th National Congress of the Communist Party of China, the Party Central Committee with Comrade Xi Jinping at its core has forged ahead, faced difficulties, and achieved a series of new achievements in comprehensively deepening reforms and socialist modernization of public welfare.

References

1. Moreo A, Esuli A, Sebastiani F (2020) Learning to weight for text classification. IEEE Trans Knowl Data Eng 32:302–316. https://doi.org/10.1109/TKDE.2018.2883446
2. Shobana J, Murali M (2021) An efficient sentiment analysis methodology based on long short-term memory networks. Complex Intell Syst 7:2485–2501. https://doi.org/10.1007/s40747-021-00436-4
3. Huang Z, Xie Z (2021) A patent keywords extraction method using TextRank model with prior public knowledge. Complex Intell Syst. https://doi.org/10.1007/s40747-021-00343-8
4. Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. Commun ACM 18:613–620. https://doi.org/10.1145/361219.361220
5. Dumais ST, Furnas GW, Landauer TK, Deerwester S, Harshman R (1988) Using latent semantic analysis to improve information retrieval. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, NY, USA, pp 281–285. https://doi.org/10.1145/57167.57214
6. David MB, Andrew YN, Michael IJ (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022. https://doi.org/10.5555/944919.944937
7. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. In: Proceedings of workshop at ICLR. https://arxiv.org/abs/1301.3781
8. Quoc L, Mikolov T (2014) Distributed representations of sentences and documents. In: Proceedings of the 31st international conference on machine learning, vol 32, pp 1188–1196. https://doi.org/10.5555/3044805.3045025
9. Singhal A (2012) Official Google Blog: introducing the knowledge graph: things, not strings. Retrieved from http://googleblog.blogspot.pt/2012/05/introducing-knowledge-graph-things-not.html
10. Cui W, Xiao Y, Wang H, Song Y, Hwang S, Wang W (2017) KBQA: learning question answering over QA corpora and knowledge bases. In: Proceedings of the VLDB endowment, vol 10, pp 565–576. https://doi.org/10.14778/3055540.3055549
11. Turing AM (1950) Computing machinery and intelligence. Mind 59:433–460. https://doi.org/10.1093/mind/LIX.236.433
12. Yu YH, Simmons RF (1990) Truly parallel understanding of text. In: Proceedings of the eighth national conference on artificial intelligence, vol 2, pp 996–1001. https://doi.org/10.5555/1865609.1865649
13. Kaminski M, Grau BC, Kostylev EV, Motik B, Horrocks I (2017) Foundations of declarative data analysis using limit datalog programs. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence main track, pp 1123–1130. https://doi.org/10.24963/ijcai.2017/156
14. Chen J, Lécué F, Pan J Z, Chen H (2017) Learning from ontology streams with semantic concept drift. In: Proceedings of the 26th international joint conference on artificial intelligence. AAAI Press, pp 957–963. https://doi.org/10.24963/ijcai.2017/133
15. Bellomarini L, Gottlob G, Pieris A, Sallinger E (2018) Swift logic for big data and knowledge graphs. In: Tjoa A, Bellatreche L, Biffl S, Leeuwen J, Wiedermann J (eds) SOFSEM 2018: theory and practice of computer science. SOFSEM 2018. Lecture notes in computer science, vol 10706, pp 3–16. https://doi.org/10.1007/978-3-319-73117-9_1
16. Chen J, Lécué F, Pan JZ, Horrocks L, Chen H (2018) Knowledge-based transfer learning explanation. In: Proceedings of the sixteenth international conference on principles of knowledge representation and reasoning. AAAI Press, pp 349–358. https://aaai.org/ocs/index.php/KR/KR18/paper/view/18054
17. Dong X, Gabrilovich E, Heitz G, Horn W, Lao Li, Murphy K, Strohmann T, Sun S, Zhang W (2014) Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, NY, USA, pp 601–610. https://doi.org/10.1145/2623330.2623623
18. Guha R, McCool R, Miller E (2003) Semantic search. In: Proceedings of the 12th international conference on World Wide Web. ACM, New York, NY, USA, pp 700–709. https://doi.org/10.1145/775152.775250
19. Rau LF (1991) Extracting company names from text Proceedings. In: Proceedings of the seventh IEEE conference on artificial intelligence application. IEEE Computer Society, pp 29–32. https://doi.org/10.1109/CAIA.1991.120841
20. Socher R, Chen D, Manning C, Andrew YN (2013) Reasoning with neural tensor networks for knowledge base completion. In: Proceedings of the 26th international conference on neural information processing systems, vol 1, pp 926–934. https://doi.org/10.5555/2999611.2999715
21. Liu Y, Han Y, Zhuo L, Zan H (2016) Automatic grammatical error detection for Chinese based on conditional random field. In: Proceedings of the third workshop on natural language processing techniques for educational applications, pp 57–62. [Online]. Available via DIALOG. https://aclanthology.org/W16-4908 of subordinate document
22. Etaiwi W, Awajan A (2020) Graph-based Arabic text semantic representation. Inf Process Manage 57:102183. https://doi.org/10.1016/j.ipm.2019.102183
23. Wei X, Zhang J, Zeng D, Li Q (2016) A multi-level text representation model within background knowledge based on human cognitive process for big data analysis. Clust Comput 19:1475–1487. https://doi.org/10.1007/s10586-016-0616-3
24. Geeganage D, Xu Y, Li Y (2021) Semantic-based topic representation using frequent semantic patterns. Knowl-Based Syst 216:106808. https://doi.org/10.1016/j.knosys.2021.106808
25. Chen Q, Xiao H (2020) A neural knowledge graph evaluator: combining structural and semantic evidence of knowledge graphs for predicting supportive knowledge in scientific QA. Inf Process Manage 57:102309. https://doi.org/10.1016/j.ipm.2020.102309
26. Wang Y, Zhang H, Shi G, Liu Z, Zhou Q (2020) A model of text-enhanced knowledge graph representation learning with mutual attention. IEEE Access 8:52895–52905. https://doi.org/10.1109/ACCESS.2020.2981212
27. Wang Y, Wang L, Yang Y, Lian T (2021) SemSeq4FD: integrating global semantic relationship and local sequential order to enhance text representation for fake news detection. Expert Syst Appl 166:114090. https://doi.org/10.1016/j.eswa.2020.114090
28. Xie Q, Tiwari P, Gupta D, Huang J, Peng M (2021) Neural variational sparse topic model for sparse explainable text representation. Inf Process Manage 58:102614. https://doi.org/10.1016/j.ipm.2021.102614
29. Ennajari H, Bouguila N, Bentahar J (2021) Combining knowledge graph and word embeddings for spherical topic modelling. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2021.3112045
30. Zhong M, Liu H, Liu L (2008) Method of semantic relevance relation measurement between words. J Chinese Inf Process 23:37–47
31. Wang K, Xie Z, Liu Y (2020) On learning associative relationship memory among knowledge concepts. Int J Netw Distribut Comput 8:124–130. https://doi.org/10.2991/ijndc.k.200515.005
32. Li X, You S, Chen W (2021) Enhancing accuracy of semantic relatedness measurement by word single-meaning embeddings. IEEE Access 9:117424–117433. https://doi.org/10.1109/ACCESS.2021.3107445
33. Yao X, Durme BV (2014) Information extraction over structured data: question answering with freebase. In: Proceedings of the 52nd

annual meeting of the association for computational linguistics, vol 1, pp 956–966. https://doi.org/10.3115/v1/P14-1090
34. Simmons RF (1986) Technologies for machine translation. Futur Gener Comput Syst 2:83–94. https://doi.org/10.1016/0167-739X(86)90002-6
35. Kupiec J, Pedersen J, Chen F (1995) A trainable document summarizer. In: Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval. ACM, New York, NY, USA, pp 68–73. https://doi.org/10.1145/215206.215333
36. Cui Y, Che W, Liu T, Qin B, Wang S, Hu G (2020) Revisiting pre-trained models for Chinese natural language processing. Association for computational linguistics, pp 657–668. https://doi.org/10.18653/v1/2020.findings-emnlp.58
37. Liu S, Yang T, Yue T, Zhang F, Wang D (2021) PLOME: pre-training with misspelled knowledge for Chinese spelling correction. Association for computational linguistics, pp 2991–3000. https://doi.org/10.18653/v1/2021.acl-long.233
38. Zhang S, Huang H, Liu J, Li H (2020) Spelling error correction with soft-masked BERT. In: Proceedings of the 58th annual meeting of the association for computational linguistics online, pp 882–890. https://doi.org/10.18653/v1/2020.acl-main.82
39. Brown RE (2020) Donald O. Hebb and the organization of behavior: 17 years in the writing. Mol Brain 13:1–28. https://doi.org/10.1186/s13041-020-00567-8
40. Tulu MM, Hou R, Younas T (2018) Identifying influential nodes based on community structure to speed up the dissemination of information in complex network. IEEE Access 6:7390–7401. https://doi.org/10.1109/ACCESS.2018.2794324
41. Quinlan JR (1986) Induction of decision trees. Mach Learn 1:81–106. https://doi.org/10.1007/BF00116251
42. Meishi-baike. [EB/OL]. [Online]. Available via DIALOG. https://meishibaike.lofter.com/ of subordinate document
43. Foodbk. [EB/OL]. [Online]. Available via DIALOG. http://www.foodbk.com/ of subordinate document
44. Xuexi.cn. [EB/OL]. [Online]. Available via DIALOG. https://www.xuexi.cn/ of subordinate document
45. Sun Y, Wang S, Li Y, Feng S, Chen X, Zhang H, Tian X, Zhu D, Tian H, Wu H (2019) ERNIE: enhanced representation through knowledge integration. CoRR abs/1904.09223. http://arxiv.org/abs/1904.09223
46. Vapnik V, Lerner A (1963) Recognition of patterns with help of generalized portraits. Avtomat Telemekh 24:774–780
47. Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13:21–27. https://doi.org/10.1109/TIT.1967.1053964
48. Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324
49. Pal S, Mitra S (1992) Multilayer perceptron, fuzzy sets, and classification. IEEE Trans Neural Networks 3:683–697. https://doi.org/10.1109/72.159058
50. Aghaee Z, Ghasemi M, Beni H, Bouyer A, Fatemi A (2021) A survey on meta-heuristic algorithms for the influence maximization problem in the social networks. Computing 103:2437–2477. https://doi.org/10.1007/s00607-021-00945-7