

Retraction

Retracted: Prediction of Chinese Semantic Word-Building Patterns Based on Complex Network Features

Wireless Communications and Mobile Computing

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] M. Wang, "Prediction of Chinese Semantic Word-Building Patterns Based on Complex Network Features," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 4162998, 10 pages, 2022.

Research Article

Prediction of Chinese Semantic Word-Building Patterns Based on Complex Network Features

Minfeng Wang 

Zhengzhou Shengda University of Economics, Business & Management, Zhengzhou, Henan 451191, China

Correspondence should be addressed to Minfeng Wang; 100711@shengda.edu.cn

Received 2 January 2022; Revised 17 January 2022; Accepted 18 January 2022; Published 14 February 2022

Academic Editor: Xin Ning

Copyright © 2022 Minfeng Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The research on prediction of Chinese semantic word-formation patterns based on complex network features has certain practical and theoretical significance in the field of natural language understanding. In this paper, complex networks are introduced into the prediction of Chinese semantic word-formation patterns, and a new prediction method of Chinese semantic word-formation patterns based on complex networks is proposed. And a solution that combines the semantic word-building rules of Chinese language with pattern recognition algorithm is put forward. Aiming at this scheme, a variety of pattern recognition algorithms are compared and analyzed, and the most suitable binary logistic regression model and naive Bayes model are found to predict Chinese semantic word-building patterns. The semantic loss is reduced, and the text classification model and corresponding classification algorithm are constructed, by introducing the maximum common subgraph theory to calculate text similarity under the complex network representation. The results of the experiments show that using complex networks to predict Chinese semantic word-formation patterns is both effective and feasible. The computer can judge the semantic word-formation pattern more accurately using the semantic word-formation pattern prediction model based on this theory.

1. Introduction

With the explosive growth of Internet information, automatic text classification technology, which can facilitate users to locate the required information quickly and accurately, becomes more and more important [1]. In the text representation method based on vector space model widely used in traditional text classification, it is easy to cause the lack of text semantic information [2] because it assumes that the feature words are independent of each other, that is, it ignores the semantic relationship between words. With the rapid development of Chinese information processing technology, the problems of polysemy and ambiguity at the text and dialogue level and unregistered words are becoming more and more prominent because computers cannot fully “understand” the semantics of words. In order for a computer to accurately “understand” the semantics of Chinese words, it must first “understand” the rules for the transformation and integration of various Chinese characters into words from a semantics perspective [3]. The word can be

regarded as a watershed moment in Chinese studies. On the one hand, it is about using words to form words and sentences, and on the other, it is about combining sound, form, and meaning. It completes the core step of all language structure, combining meaning and form on a sound tangible carrier, through “word”. A complex network is a new method for studying complex systems in a global context. It uses two basic elements of nodes and edges to study complex network systems, regardless of how complex the structure or scale of the network is [4]. Complex networks can also be found in human language, according to research. The human language network is neither completely random nor completely regular, but rather a complex network with “small-world” characteristics. Language networks show high global connectivity and local aggregation.

In a semantic network, nodes are notional words and the relationship between nodes is semantic relationship [5]. Because the semantic network is an intermediate layer between syntactic and conceptual network, the dynamic semantic network based on real text is helpful to the research

of the following three problems: the organization of human semantic or conceptual knowledge, human semantic processing mechanism, and semantic retrieval process [6]. Syntax and semantics are closely related. Therefore, observing the similarities and differences between them from the perspective of complex network is also helpful to the study of syntactic semantic interface [7]. The study of Chinese word formation starts from the study of semantics. The focus of semantic word formation is compound words. The focus of solving compound word formation is to straighten out the complex semantic relationship between word meaning and morpheme meaning. Word-forming morphemes have less explicit things and more implicit things. In this process, the meaning of some words is just a simple addition of morpheme meaning [8]. On the contrary, the relationship between morpheme meaning and word meaning of some words is indirect, which needs a certain “semantic bridge” for transition, so as to better clarify the process of implicit fusion. This has also become the focus and difficulty of linguistics, computer linguistics, and dictionary interpretation, including the study of Chinese as a foreign language. In order to solve the lack of text structure and semantic information in the above vector space model, in this paper, the complex network theory is introduced into the prediction of Chinese semantic word-formation mode, and the semantic theory is effectively integrated with text representation.

The study of Chinese vocabulary benefits greatly from research into the semantic word formation of disyllabic compound words. The traditional word-formation method of “pattern analysis word formation by applying syntax” simplifies the complex internal relationships of compound words to some extent, so Chinese word-formation research will almost certainly be combined with lexical semantics [9]. The first step in “word formation” is the creation of compound words with two or more syllables. In Chinese grammar, the “law” embodied in this structure is fundamental, and larger language units are established step by step as a “paradigm” [10]. Chinese is an ideographic language, meaning that the semantics of word connotation take precedence over the form of words. Furthermore, there are many instances of word semantic correlation and context dependence in Chinese. As a result, the semantic information contained in Chinese text is richer than that contained in other languages that focus on form and structure, making it more difficult for a text representation model based on vector space to fully describe and predict the semantic information contained in Chinese text [11]. Therefore, this paper proposes a prediction model of Chinese semantic word formation based on complex network features. Based on the co-occurrence relationship between words, this paper constructs a single text weighted complex network to represent Chinese text. This text representation method can not only contain the information of feature words but also reflect the semantic correlation and context structure information between feature words. At the same time, in order to reduce the computational complexity, the text feature selection is carried out through the small-world characteristics of the complex network, and the keywords

reflecting the text theme are extracted as the text feature words by using the comprehensive characteristics of nodes, so as to optimize the text network structure and reduce the complexity of the text network. Simulation results show that this method can significantly improve the efficiency and accuracy of unlisted word interpretation, ambiguity elimination, dictionary compilation, and machine translation.

2. Related Work

On the basis of constructing a word co-occurrence text complex network, Akbar and Hossein [12] used the impact of node deletion on the average shortest path length in the network to extract Chinese text keywords. Koronovskii et al. [13] comprehensively considered node degree and aggregation coefficient for keyword extraction but did not consider the impact of nodes on the global network. Fu et al. [14] comprehensively considered the weighted aggregation coefficient and node intermediate number to extract text keywords but ignored the importance of node degree. Wang et al. [15] used the community discovery algorithm of complex networks to complete feature selection. After understanding the related problems in the field of natural language processing, Shen et al. [16] found that the research on word formation based on semantic perspective is of great significance to solve the ambiguity and polysemy of unregistered words and texts. Yin et al. [17] combined conceptual integration theory with physical structure, which complements each other, and provided a good reference method for us to analyze the specific ways of semantic word formation. Chen et al. [18] considered that noun use belongs to logical metonymy, which can be explained by event coercion in generative thesaurus theory, and put forward that physical structure is equivalent to noun and argument structure is equivalent to verb. Li et al. [19] used the theory of complex networks to create a “modern Chinese complex sentence relational word collocation network” based on 560 complex sentence relational words. It is used to determine the collocation ability and intensity between Chinese complex sentence relational words using network statistics such as average path length, aggregation coefficient, and degree distribution. These findings may aid in determining the hierarchical relationship and logical semantics of complex sentences automatically. Tian and Wang [20] proposed that complex network theory be introduced into the field of Chinese text classification, that it be used to effectively integrate the theories of semantics and syntax with text representation, and that the text semantic information be retained to the greatest extent possible. A Chinese document keyword extraction algorithm based on complex network features was proposed in [21]. The algorithm extracts keywords from a document language network using the complex network eigenvalues of vocabulary nodes. The average accuracy of keyword extraction is high, according to experimental results. A Chinese semantic role network was built, and its statistical characteristics were studied in [22]. Although the semantic network is a small-world and scale-free network, the results show that it differs significantly from the syntactic network in terms of hierarchical structure and node degree correlation. He et al. [23] focused on the grammatical form of words, word meaning, and dictionary

interpretation, describing the relationship between word meaning and morpheme meaning. It enriches and develops lexicology to pay attention to and explain the semantic structure of implicit and metonymic compounds. Li and Xiao [24] claim that language, at all levels, including pronunciation, morphology, syntax, and semantics, embodies the nature of complex networks. Guo et al. [25] investigated the subject of predicting Chinese semantic word-formation patterns. We can understand the research status and accomplishments of semantics by consulting a large amount of literature and materials. Hou et al. [26] used Chinese semantic word-formation knowledge such as part of speech, word-formation structure, and morpheme meaning to calculate the semantic similarity of Chinese words based on “morpheme concept” based on the idea of Chinese character standard. The word knowledge is expressed in a simple, intuitive, and easy-to-expand way. The calculation model is straightforward, with as few features and parameters as possible.

This paper constructs a prediction model of Chinese semantic word-formation pattern based on complex network features. This paper introduces the research status, research objects, and applied theories of foreign language semantic word formation in China and explains the source of corpus and the construction of corpus. On the basis of previous studies, more detailed information annotation is carried out to meet the research needs, and finally three subcorpora are formed respectively. In order to verify the effect of the Chinese semantic word-formation pattern prediction algorithm based on complex network proposed in this paper, in the research on the prediction of Chinese semantic word-formation pattern, the Chinese semantic word-formation pattern prediction model based on binary logistic regression and naive Bayes is constructed and simulated. The confusion matrix of the prediction results of the two models is calculated, and the prediction accuracy of the two models is compared.

3. Methodology

3.1. Language Complex Network. A complex network is a complex system that combines the perspective and basic principles of network. Essentially, it is a graph structure composed of a large number of nodes and their interactions. Among them, the basic elements or phenomena in the system are abstracted as nodes, while the relationship between research objects is defined as the connection edge of nodes. The research of complex networks has strong interdisciplinary characteristics. At present, complex networks have been widely used in all levels and fields. In the research of complex network, the language network, as a new research direction, has quietly emerged. A language complex network is the language structure studied with complex network view. Research shows that human language is also a complex network in the human complex system. Although the construction principles and methods of different language networks are different, all language networks have similar statistical characteristics.

The complex network is an interdisciplinary research field full of vitality. On the one hand, in the research of

complex network theory, new theoretical models and new analysis methods are constantly put forward. On the other hand, new structures and phenomena in the real network are constantly discovered, and a large number of important application problems emerge. The language complex network usually defines morphemes in language as nodes and the relationship between morphemes as edges. Common conjunctions include co-occurrence relation, conceptual synonymy, syntactic relation, etc. Relational words are words used in complex sentences, sentence groups, or discourse to connect sentences to show logical relations. It is one of the important symbols of the logical and semantic relationship between sentences, and it has both formal and semantic functions in syntactic structure analysis.

Complex networks are everywhere. Almost all complex systems have multiple agents, each with their own set of relationships and constraints, resulting in a complex network system. The formation and evolution of a real network with a large number of interacting nodes is not as random as originally thought but exhibits a number of peculiar properties. The “small-world” feature is one of the most striking features. The small-world phenomenon states that, despite the fact that many networks are quite large in scale, there is a relatively short “fast distance” between any two nodes if the number of edges of the shortest path connecting nodes is defined as the distance between nodes. That is, the network’s topology shows relatively high aggregation characteristics across the board, but the path between any two nodes is relatively short. Despite their differing construction principles, studies on various language networks show that these networks are small-world and scale-free. To put it another way, the overall statistical characteristics of language networks appear to have nothing to do with the structure or type of language.

Each word or phrase is regarded as a node in language, and the words or phrases that appear together in the same sentence form the link between words or phrases. If only the connections between two adjacent words or phrases are collected in the edge collection, some long-range connections may be lost, while the importance of some useless words in the network is increased. As a result, the relevance span of words or phrases in sentences must be determined. Many important correlations cannot be recorded if the time span is too short. If the interval is too long, a lot of redundant data may be produced. There is no such thing as a completely random or completely regular language. Discrete and limited words can form sentences, paragraphs, and texts with rich semantics because words in natural language have semantic correlation with each other and form sentences in a nonrandom manner. This language’s construction engineering is extremely robust. As a result, using the graph that depicts the relationship between words, we can investigate the organizational characteristics of language.

3.2. Prediction of Chinese Semantic Word-Formation Patterns. Chinese information processing is facing the research difficulties of sentence processing and text processing. In the aspect of sentence processing, it is mainly divided into single

sentence processing and complex sentence processing. At present, there are many achievements in the research of Chinese simple sentence information processing. However, complex sentences are the bridge between simple sentences and texts and are important entity units of Chinese grammar, which express rich and complex semantic information. Therefore, it has important research value in the field of information processing. In the semantic classification system of the semantic classification information database of Chinese words and phrases, the semantic categories, middle categories, and small categories are relatively consistent in the absolute number of Chinese words and phrases, that is to say, the distribution number of words in the semantic category with the largest number of words is also the largest in the “Chinese words and phrases semantic information database.” On the contrary, if the number of words in this semantic category is the least, then the number of words contained in this semantic category is also the least in the “Chinese Words Semantic Category Information Solution.”

The collocation of Chinese complex sentence relational words is the syntactic co-occurrence of two or more complex sentence relational words in Chinese discourse. It is the marking component used to connect clauses, mark the semantic relations between clauses and form complex sentence patterns, and is the formal symbol of syntactic relations and semantic relations between clauses. There is a certain relationship between words and phrases in semantic composition, and there is a certain correspondence in the semantic category distribution system. The semantics of most words may be integrated on the basis of each word meaning that constitutes the vocabulary. That is, two words form a vocabulary with certain semantic meaning, not just a combination of two words. The collocation of Chinese complex sentence relational words not only affects the semantics of clauses but also affects the division of hierarchical relations of complex sentences. Chinese text classification based on complex network is shown in Figure 1.

When constructing relevant models for the prediction of Chinese semantic word-formation patterns studied in this paper, a relatively large corpus is required as the material for machine learning. The vocabulary system is open and infinite, whereas the Chinese characters that make up the vocabulary are limited, and these limited Chinese characters are combined to form many words. We can use single words as resources, label information, and build a corpus in Chinese information processing to investigate the semantic word-building rules of Chinese disyllabic compound words. To ensure that the semantic word-building rules learned by the prediction model used in this paper in the corpus machine learning [27, 28] process are authentic, the vocabulary quantity, scale, representativeness of the corpus, and semantic category labeling accuracy can be used as a training sample data set source for the Chinese word-building pattern prediction algorithm for machine learning to some extent.

It is necessary to determine the prediction model to be used when using a computer to predict semantic word-formation patterns. The prediction model can judge the semantic word-formation patterns of new words using the known

characteristic attribute factor values of new words and the learned rules after learning the relationship between the characteristic attribute factor values that affect the semantic word-formation patterns. Because words in a text are not combined at random, discrete words and words are combined to form a text using semantic correlation. Rich semantic relations of words, especially in Chinese, form sentences, paragraphs, texts, and other structures with rich semantics, which then convey information through texts. The majority of traditional feature selection methods rely on statistical data such as word and document frequency, ignoring the semantic correlation between important words in the text.

Analyze the language network, look for words that play an important and central role in the whole language network, and extract these words as keywords. In the language network, the degree of a word reflects the relationship between the word and other words. The greater the degree of the word, the greater the importance of the word. The aggregation coefficient of the vocabulary reflects the interconnection density of nodes in the local scope of the vocabulary and the aggregation of the vocabulary in the local scope. After further study of Chinese word formation, it is found that the part of speech of the two word positions that make up the vocabulary has an influence on the semantic word-formation mode of the vocabulary to a certain extent, so the part of speech of the two word positions is subdivided into big categories and small categories as the other four characteristic-influencing factors. Based on the assumption that these characteristic attribute variables are independent of each other, the model with simple algorithm logic and easy realization but good algorithm prediction effect is selected. The experimental flow of the Chinese semantic word-formation pattern prediction model is shown in Figure 2.

On the basis of argument structure and metonymy theory, this paper makes a detailed explanation of the escape ways of verbs in the research corpus and finds that verbs are mostly metaphorical or metonymic based on the similarity or correlation between word-formation morpheme events and word meaning representative events. Among them, for compound verbs with verb-noun, noun-noun, and noun-noun, the interpretation of the physical role of noun morpheme plays an important supporting role in the interpretation of verb meaning. In order to better study the rules of Chinese semantic word formation, it is necessary to build a large-scale database with words as the basic unit and mark the semantic categories of words and Chinese characters constituting each word in this database with reference to the “Chinese Characters and Word Meaning Classification Information Database” built above and make statistical induction and analysis on this database, so as to sum up the rules between lexical semantics and Chinese characters constituting this word.

3.3. Prediction Model of Semantic Word-Formation Pattern Based on Complex Network. Before constructing a complex network, I must first preprocess the text, which includes word segmentation and the removal of stop words. The original feature set is created after the text has been preprocessed. This paper represents the text as a weighted complex network

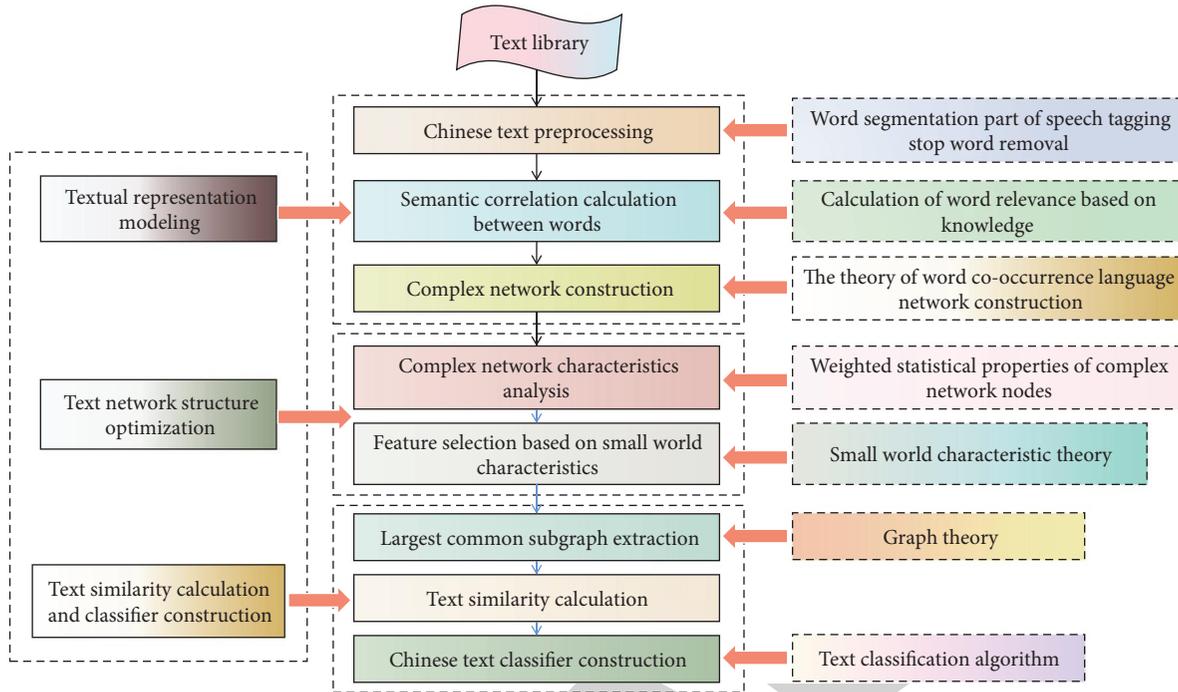


FIGURE 1: Chinese text classification based on complex network.

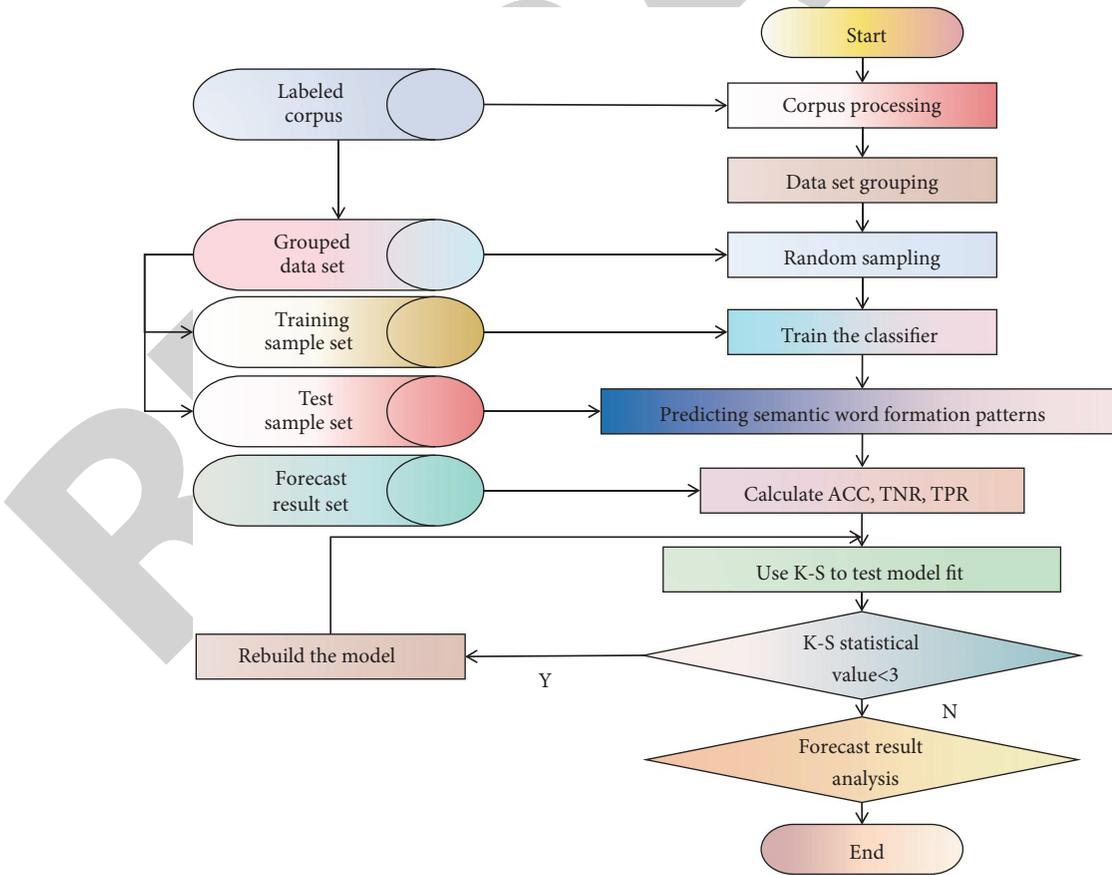


FIGURE 2: Prediction model of Chinese semantic word-formation pattern.

structure to retain more text information and reflect the structure and semantic features of the text. This paper explains the escape ways of nouns in corpus using transitivity structure and metaphorical metonymy theory and finds that most nouns are metaphorically or metonymically based on some kind of transitivity role in which word-building morphemes partially or entirely act as word meanings.

Not only the collocation network of modern Chinese compound sentence relational words is a typical complex network, but these statistical characteristics also reflect compound sentence relational words' collocation ability and the strong-weak relationship between collocation objects. They provide a solid foundation for further research into the automatic detection and processing of compound sentence relational words, compound sentence hierarchical relationships, and compound sentence logic semantics in modern Chinese.

If a network has a smaller $\langle d \rangle$ ($\langle d \rangle \sim \langle d_{\text{rand}} \rangle$) and a larger aggregation coefficient C ($C \gg C_{\text{rand}}$), this kind of network is a small-world network. Accordingly, although the C of the semantic network is slightly smaller than the C of the two syntactic networks, the semantic network can still be regarded as a small-world network. If the degree distribution of a network obeys the power law:

$$P(k) \sim k^{-\gamma}, \quad (1)$$

and the value of the power exponent γ is between 2 and 3. Such a network is also called a scale-free network. Figure 3 shows the cumulative distribution of the semantic network. It can be seen from Figure 3 that the semantic network is a scale-free network.

If a certain characteristic attribute value of a new sample does not appear or does not appear in a certain category when predicting the category to which it belongs, the quality of the prediction model will be greatly reduced. Once a certain feature attribute value in the test sample does not appear in the training sample, it will lead to the inability to calculate the conditional probability of the feature attribute value, thus seriously affecting the classification of the test sample. In order to avoid wrong results when predicting test samples, calibration is needed to solve this problem.

In this article, $f(d(a_i, b_j), C, \alpha)$ is calculated using the function shown in formula (2).

$$\begin{aligned} D(A, B) &= f(d(a_i, b_j), C, \alpha), \\ &= \alpha * \sum_{i=1,2} d(a_i, b_j) * C_{ij}. \end{aligned} \quad (2)$$

Calculate the semantic similarity $\text{Sim}(A, B)$ of the word pair A and B . Considering the distribution characteristics of $D(A, B)$, it is agreed as formula (3).

$$\begin{aligned} \text{Sim}(A, B) &= f(d(A, B)), \\ &= \begin{cases} 1, & d(A, B) = 0, \\ \frac{1}{1 + e^{(a*d(A,B)+c)}}, & d(A, B) > 0, \end{cases} \end{aligned} \quad (3)$$

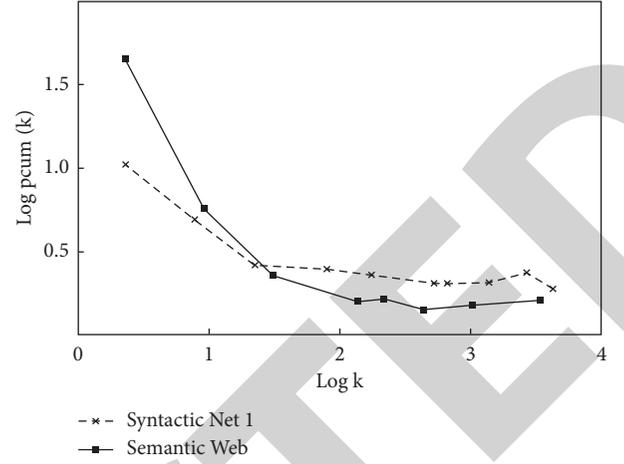


FIGURE 3: Cumulative degree distribution of the semantic network.

where a is used to adjust the overall trend of the function, and c is used to adjust the center of symmetry of the function. In this article, take $a = 0.8$, $c = -20$.

Function words are not included in the semantic analysis of sentences. This also means that the semantic network obtained from semantic analysis does not contain function words. On the contrary, function words play a very important role in the syntactic network. Therefore, the statistical characteristics of semantic network and syntactic network should be different. Because of the similarity between semantic analysis and concept map, the research of semantic network is also helpful to discover some features of conceptual network. As for adjectives, their independence is poor, and most of them come from metaphor or metonymy through nouns and verbs or descriptions of things and events.

I use the theory of physical structure under the generated lexicon theory to describe the semantics of nouns in detail, while for verbs, we form a semantic framework on the premise of describing their events or argument structures when analyzing the "communication mode" between the morpheme meaning and the word meaning of undirected words. The semantic word-formation pattern prediction model predicts two types of semantic word-formation patterns in the test sample data set based on the probability of the word-formation pattern of the language to which the word belongs. As a result, when calculating the probability of the semantic word-formation pattern to which the word belongs, the semantic word-formation pattern prediction model should make p of the semantic word-formation pattern to which the word to be predicted actually belongs as large as possible, and $1 - p$ as small as possible.

Basic indicators need to be combined to comprehensively measure the performance of text classification. Use the F_β measurement value method to comprehensively consider recall and precision. The F_β calculation formula is as follows:

$$F_\beta = \frac{(\beta^2 + 1) \cdot r \cdot p}{\beta^2 \cdot p + r}. \quad (4)$$

Among them, β is a parameter that adjusts the proportion of recall and precision in the F_β measurement value. Among them, when $\beta > 1$, recall is more important. Conversely, when $\beta < 1$, precision is more important. When $\beta = 1$, recall and precision have equal importance. The F_1 value is the balance point, and its calculation formula is as follows:

$$F_1 = \frac{2 \cdot r \cdot p}{r + p}. \quad (5)$$

From the definition of conditional probability and the formula of total probability:

$$P(A_j | B) = \frac{P(A_j B)}{P(B)} = \frac{P(A_j)P(B | A_j)}{P(B)}, \quad 1 \leq j \leq n, \quad (6)$$

$$P(B) = \sum_{i=1}^n P(A_i)P(B | A_i).$$

Inferred:

$$P(A_j | B) = \frac{P(A_j)P(B | A_j)}{\sum_{i=1}^n P(A_i)P(B | A_i)}, \quad 1 \leq j \leq n. \quad (7)$$

The complex network built in this paper has a large number of nodes and edges, with nodes representing feature words and edges representing semantic correlation between feature words. The co-occurrence and adjacent position relationship of feature words in the text reflects this, and the edge weight represents the degree of semantic correlation of feature words. The closer the semantic correlation between feature words, the higher the weight. Create a language network for documents. We set up a connection edge for the node pairs with a span of 1 or 2 in each sentence, connect the network composed of each sentence, and merge the same nodes and connection edges to form the document's language network, using the words in the preprocessed document as nodes.

4. Result Analysis and Discussion

The number of training sample data sets and the quality and representativeness of data directly affect the quality of the prediction model, while the test sample data sets can help us screen better prediction models through the prediction effect, so it is equally important. In this study, because there are many models built, the training samples and test samples applicable to each model are different, so this paper adopts the method of random sampling to generate training and test sample data sets. The semantic word-formation pattern prediction model based on complex network proposed in this paper is verified by experiments, that is, the prediction method based on complex network and vector space model are used to predict the text. The recall test results are shown in Figure 4.

As can be seen from Figure 4, compared with the traditional method based on vector space, the prediction method proposed in this paper improves the recall rate. From the average recall rate, the model based on complex

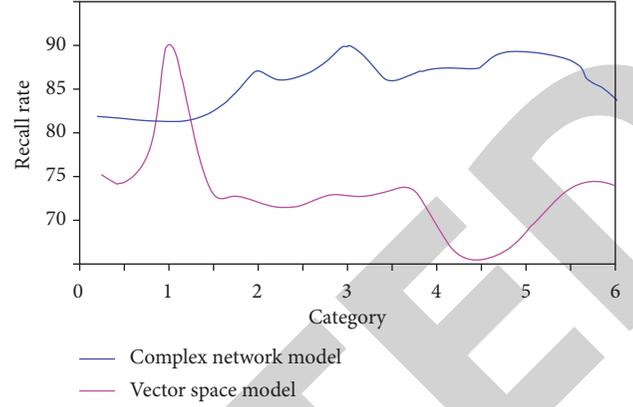


FIGURE 4: Comparison of recall test results.

network is dominant. This shows that it is feasible and effective to introduce the complex network model into the semantic word-formation pattern prediction.

In order to measure the complexity of a network, the most commonly used complex network parameters are average path length, aggregation coefficient, and degree distribution. Semantic analysis is a kind of language structure analysis based on dependency, which aims at clarifying the deep semantic structure of sentences. In syntactic analysis, every word in a sentence will appear in the final syntactic diagram, but the semantic analysis of a sentence generally only needs to pay attention to the relationship between notional words in the sentence. In order to further verify the effectiveness of keyword extraction in the semantic word-formation prediction model based on complex network, the TFIDF method and this method are used to extract keywords, and the extracted results are compared with the keywords marked in corpus. The experimental results are shown in Figure 5.

It can be seen from Figure 5 that the average accuracy rate obtained by the method based on complex network features proposed in this paper is higher than that obtained by the TIFDF method. The word-formation structure of Chinese words reflects the different contributions of morphemes to the whole word meaning under different word-formation structures. For example, in the joint structure, each morpheme contributes almost the same to the whole word meaning, while in the centering structure, the central element contributes more to the whole word meaning. The semantic average path length of the network represents the average shortest path length between any two nodes in the network, which is expressed by $\langle d \rangle$. The largest and shortest path in the semantic network is D . Figure 6 shows that the shortest path distribution of semantic network has a long tail. The $\langle d \rangle$ and d of the semantic network are larger than that of the syntactic network.

Taking Wikipedia knowledge base as the data source, the concept distance and category distance are calculated by using its link structure and category system, and then the correlation between concepts is calculated by linear combination of these two values. Firstly, the feature words are transformed into topic concepts, that is, word-concept matching is carried out. Secondly, the semantic correlation

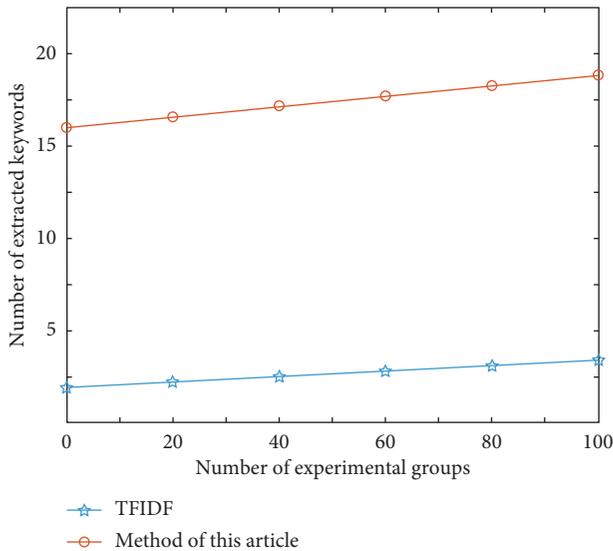


FIGURE 5: Comparison of different methods of extracting keywords.

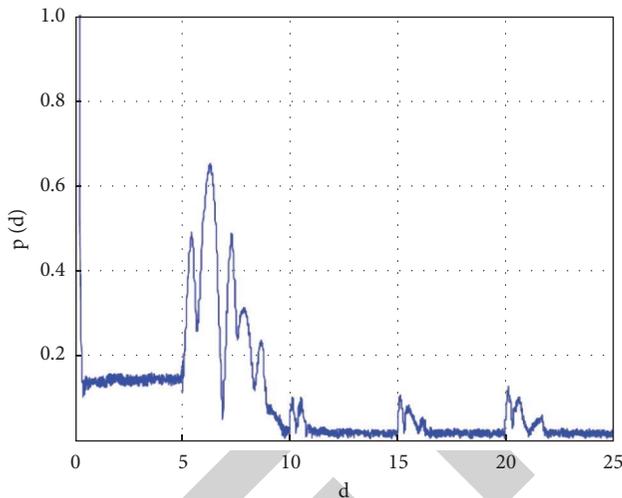


FIGURE 6: The shortest path distribution of the semantic network.

between concepts is quantified to complete the calculation of word correlation.

The similarity or correlation between the events constructed by word-formation morphemes and the events represented by word meaning itself creates the indirect relationship between word meaning and morpheme meaning. Among them, analyzing the physical role of noun morphemes in word-building morphemes is very helpful for us to further clarify the detailed ways of verb meaning transfer. For adjectives, metaphor or metonymy mainly occurs in two ways based on the characteristics of things or events, and the characteristics expressed by the word meaning of adjectives are embodied by word-building morphemes. Figure 7 shows the prediction accuracy of word-formation patterns based on the logistic regression model and naive Bayes model to word-formation patterns in the test sample data set.

The prediction accuracy of the eight semantic word-formation pattern prediction models based on logistic

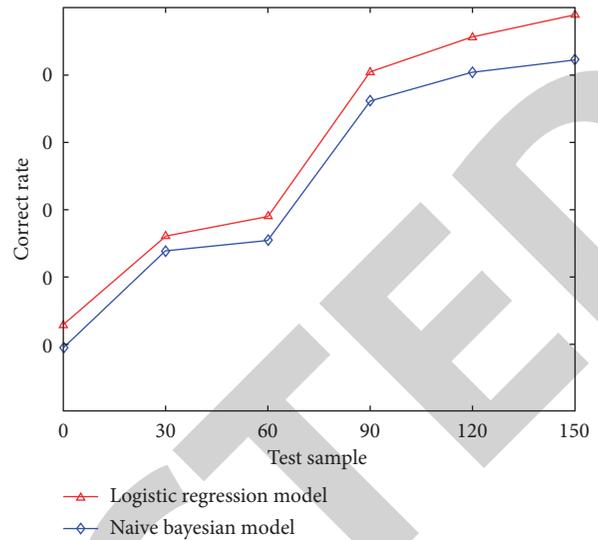


FIGURE 7: The prediction accuracy of the two models.

regression is relatively high, as can be seen in Figure 7, indicating that the prediction effect is better. In a vector space model, feature selection aims to keep as much text information as possible while improving text category discrimination and removing noise. However, most commonly used feature selection methods are based on statistical data such as word frequency or the relationship between words and categories, while ignoring the semantic correlation between feature items, making the selected feature words unable to fully express the text's theme and category features, and thus affecting the feature selection effect to some extent.

The two lexical features that make up words are marked in this paper by searching the electronic dictionary by computer for narrow lexical features. Parts of speech are divided into large and small categories, which are then entered into the semantic word-building database separately. Finally, the structure of feature attribute factors and semantic word-building patterns in the processed labeled corpus sample data set is clear, and the compiler can easily learn the prediction model. Based on the text representation of a complex network, this paper's feature selection method aims to reduce the complexity of the complex network of text and optimize its network structure, in order to extract keywords that can reflect the text's theme as feature words of the text, in which node characteristics are calculated by considering node degree, weighted aggregation coefficient, and node betweenness, in order to comprehensively measure the segregation. In this experiment, the K nearest neighbor (KNN) classification algorithm is used for classification, and included angle cosine method is used for text similarity calculation. After repeated experiments, in order to get the best experimental results, K in the KNN classification algorithm is 20. The experimental results are shown in Figure 8.

It can be seen from Figure 8 that the classification performance of the Chinese semantic word-formation pattern prediction model based on complex network proposed in this paper is better than the traditional information gain (IG), document frequency (DF), and chi-square test

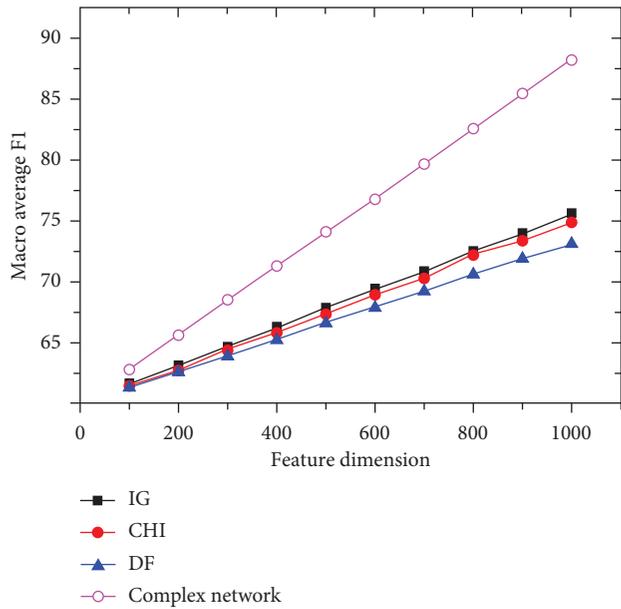


FIGURE 8: Predictive text classification effect under different feature selection methods.

(CHI). It shows that considering the semantic relationship between words and text structure information in Chinese text feature selection can effectively improve the text classification effect. The overall effect of the complex network is better, which shows that using the word relevance method of knowledge base can further improve the effect of text representation and feature selection.

This section mainly studies the simulation of this model. By comparing the predicted value with the actual value, the accuracy rate, the coverage rate of positive cases, and the coverage rate of negative cases of several models are counted, and it is found that the prediction effect of this model is better than that of other models. It shows the superiority and practicability of this model.

5. Conclusions

With the rapid development of natural language processing, more and more experts and scholars are devoted to semantic research. Before computers can understand the semantics of sentences, they should first understand the semantics of words. Polysemy and ambiguity of vocabulary semantics are the bottlenecks that restrict the further development of natural language information processing, and the identification of unknown words is also an extremely difficult problem in Chinese information processing nowadays. In this paper, a prediction model of Chinese semantic word-formation pattern based on complex network is proposed. The theory of complex network is introduced into the prediction field of Chinese semantic word-formation pattern, and the theories of semantics and syntax are effectively integrated with text representation. Making use of the small-world characteristics of the language network and comprehensively considering the characteristics of nodes for feature selection can keep the text information and reduce

the complexity of the text network. At the same time, the text similarity calculation method based on the maximum common subgraph is used to improve the accuracy of calculation, thus improving the prediction effect of Chinese semantic word-building patterns. The results show that combining the small-world characteristics of complex networks and making use of the comprehensive characteristics of nodes to select text features can accurately and comprehensively measure the characteristics of nodes and their influence on the text theme content, thus improving the accuracy of text feature selection. Combining the research results of Chinese language and literature, the rules of semantic word formation of Chinese words with computer technology can not only lay a solid foundation for effectively solving the prediction problem of Chinese semantic word-formation mode but also further apply the ingenious results to practical application.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author does not have any possible conflicts of interest.

Acknowledgments

This study was supported by the 2022 Research on the Emergency Language Service for Public of Humanities and Social Sciences in Universities in Henan Province (2022-ZZJH-534) and the 2021 14th Five-Year Plan Project of Education Science in Henan Province (2021YB0371).

References

- [1] C. K. Morley and C. W. Nixon, "Topological characteristics of simple and complex normal fault networks," *Journal of Structural Geology*, vol. 84, no. 3, pp. 68–84, 2016.
- [2] J. Kong, C. Yang, J. Wang et al., "Deep-stacking network approach by multisource data mining for hazardous risk identification in IoT-based intelligent food management systems," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 1194565, 2021.
- [3] T. Wang, J. Fei, and X. Cheng, "Estimating the correlated observation-error characteristics of the Chinese Feng Yun microwave temperature sounder and microwave humidity sounder," *Advances in Atmospheric Sciences*, vol. 35, no. 11, pp. 94–107, 2018.
- [4] Y. Chuanqi, G. Ping, and Z. Caineng, "Formation conditions and distribution patterns of N-1 tight oil in Zhahaquan Area, Qaidam Basin, China," *Energy Exploration & Exploitation*, vol. 34, no. 3, pp. 339–359, 2016.
- [5] I. Moreno, E. Boldrini, P. Moreda, and M. T. Romá-Ferri, "DrugSemantics: a corpus for named entity recognition in Spanish summaries of product characteristics," *Journal of Biomedical Informatics*, vol. 72, pp. 8–22, 2017.
- [6] Z. Pan, Q. Ma, J. Ding, and L. Wang, "Research on the stock correlation networks and network entropies in the Chinese green financial market," *The European Physical Journal B*, vol. 94, no. 2, p. 56, 2021.

- [7] M. Guerrero, F. G. Montoya, R. Baños, A. Alcayde, and C. Gil, "Adaptive community detection in complex networks using genetic algorithms," *Neurocomputing*, vol. 266, no. 11, pp. 101–113, 2017.
- [8] M. Wiedermann, J. F. Donges, J. Kurths, and R. V. Donner, "Spatial network surrogates for disentangling complex system structure from spatial embedding of nodes," *Physical Review A*, vol. 93, no. 4, Article ID 42308, 2016.
- [9] X. Liu, Z. Yang, Z. Zhou et al., "The impact of protein interaction networks' characteristics on computational complex detection methods," *Journal of Theoretical Biology*, vol. 439, pp. 141–151, 2018.
- [10] Y. Hwang, W. B. William, and R. T. David, "Measuring ecological characteristics of environmental building performance: suggestion of an information-network model and indices to quantify complexity, power, and sustainability of energetic organization," *Ecological Indicators*, vol. 83, pp. 201–217, 2017.
- [11] G. Li, I. Svogor, and G. Beltrame, "Long-term pattern formation and maintenance for battery-powered robots," *Swarm Intelligence*, vol. 13, no. 1, pp. 21–57, 2019.
- [12] G. Akbar and S. Hossein, "The solar flare complex network," *The Astrophysical Journal*, vol. 847, no. 2, p. 115, 2017.
- [13] A. A. Koronovskii, O. I. Moskalenko, A. A. Pivovarov, and A. E. Hramov, "A method of distinguishing between the characteristic phases of behavior in complex networks in the intermittent generalized synchronization regime," *Technical Physics Letters*, vol. 43, no. 4, pp. 328–330, 2017.
- [14] X. Fu, Y. Yang, W. Dong, C. Wang, and Y. Liu, "Spatial structure, inequality and trading community of renewable energy networks: a comparative study of solar and hydro energy product trades," *Energy Policy*, vol. 106, no. 7, pp. 22–31, 2017.
- [15] Q. Wang, Z. Zhang, M. Wang, J. Li, J. Fang, and Y. Lai, "Self-assembled three-dimensional carbon networks with accessorial Lewis base sites and variational electron characteristics as efficient oxygen reduction reaction catalysts for alkaline metal-air batteries," *Chinese Journal of Catalysis*, vol. 39, no. 7, pp. 1210–1218, 2018.
- [16] W. Shen, Q. Qu, and X. Li, "Semantic information mediates visual attention during spoken word recognition in Chinese: evidence from the printed-word version of the visual-world paradigm," *Attention, Perception, & Psychophysics*, vol. 78, no. 5, pp. 1–18, 2016.
- [17] C. Yin, X. Xiao, V. Balaban et al., "Network science characteristics of brain-derived neuronal cultures deciphered from quantitative phase imaging data," *Scientific Reports*, vol. 10, no. 1, Article ID 15078, 2020.
- [18] B. Chen, X. B. Wang, Y. L. Li, Q. Yang, and J. S. Li, "Energy-induced mercury emissions in global supply chain networks: structural characteristics and policy implications," *The Science of the Total Environment*, vol. 670, no. 6, pp. 87–97, 2019.
- [19] L. Li, X. Meng, C. Lin, D. Shao, Y. Xiong, and M. Wu, "The somatic and germline mutation pattern of 21 hereditary ovarian cancer genes in 62 Chinese ovarian cancers," *Gynecologic Oncology*, vol. 154, no. 1, Article ID 272, 2019.
- [20] M. Tian and S. Wang, "Dynamic transition and pattern formation for chemotactic systems," *Discrete and Continuous Dynamical Systems - Series B*, vol. 19, no. 9, pp. 2809–2835, 2017.
- [21] R. Kurita, "Control of pattern formation during phase separation initiated by a propagated trigger," *Scientific Reports*, vol. 7, no. 1, Article ID 6912, 2017.
- [22] S. Kumar, B. S. Panda, and D. Aggarwal, "Community detection in complex networks using network embedding and gravitational search algorithm," *Journal of Intelligent Information Systems*, no. 1, pp. 1–22, 2020.
- [23] L. He, L. Zhu, and Z. Zhang, "Turing instability induced by complex networks in a reaction-diffusion information propagation model," *Information Sciences*, vol. 578, pp. 762–794, 2021.
- [24] X. Li and R. Xiao, "Analyzing network topological characteristics of eco-industrial parks from the perspective of resilience: a case study," *Ecological Indicators*, vol. 74, no. 3, pp. 403–413, 2017.
- [25] J. Guo, S. Wang, D. Wang, and T. Liu, "Spatial structural pattern and vulnerability of China-Japan-Korea shipping network," *Chinese Geographical Science*, vol. 27, no. 5, pp. 697–708, 2017.
- [26] R. Hou, S. Yuan, C. Feng, Y. Xu, K. Rao, and Z. Wang, "Toxicokinetic patterns, metabolites formation and distribution in various tissues of the Chinese rare minnow (*Gobiocypris rarus*) exposed to tri(2-butoxyethyl) phosphate (TBOEP) and tri-n-butyl phosphate (TNBP)," *The Science of the Total Environment*, vol. 668, no. 6, pp. 806–814, 2019.
- [27] X. B. Jin, W. Z. Zheng, J. L. Kong et al., "Deep-learning temporal predictor via bidirectional self-attentive encoder-decoder framework for iot-based environmental sensing in intelligent greenhouse," *Agriculture*, vol. 11, no. 8, Article ID 802, 2021.
- [28] M. Zhao, A. Jha, Q. Liu et al., "Faster Mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, Article ID 102048, 2021.