Research Article

Research on the Natural Language Recognition Method Based on Cluster Analysis Using Neural Network

Guang Li [iD],[1] **Fangfang Liu** [iD],[2] **Ashutosh Sharma** [iD],[3] **Osamah Ibrahim Khalaf** [iD],[4] **Youseef Alotaibi** [iD],[5] **Abdulmajeed Alsufyani** [iD],[6] **and Saleh Alghamdi** [iD][7]

[1]*Institute of Data Science, City University of Macau, Macau 999078, China*
[2]*High School Attached to Capital Normal University, Beijing 100048, China*
[3]*Institute of Computer Technology and Information Security, Southern Federal University, Rostov-on-Don, Russia*
[4]*Al-Nahrain University, Al-Nahrain Nano-renewable Energy Research Center, Baghdad, Iraq*
[5]*Department of Computer Science, College of Computers and Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia*
[6]*Department of Computer Science, College of Computers and Information Technology, Taif University, P.O.Box 11099, Taif 21944, Saudi Arabia*
[7]*Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia*

Correspondence should be addressed to Guang Li; liguang015@gmail.com and Ashutosh Sharma; sharmaashutosh1326@gmail.com

Withthe technological advent, the clustering phenomenon is recently being used in various domains and in natural language recognition. This article contributes to the clustering phenomenon of natural language and fulfills the requirements for the dynamic update of the knowledge system. This article proposes a method of dynamic knowledge extraction based on sentence clustering recognition using a neural network-based framework. The conversion process from natural language papers to object-oriented knowledge system is studied considering the related problems of sentence vectorization. This article studies the attributes of sentence vectorization using various basic definitions, judgment theorem, and postprocessing elements. The sentence clustering recognition method of the network uses the concept of prereliability as a measure of the credibility of sentence recognition results. An ART2 neural network simulation program is written using MATLAB, and the effect of the neural network on sentence recognition is utilized for the corresponding analysis. A postreliability evaluation indexing is done for the credibility of the model construction, and the implementation steps for the conjunctive rule sentence pattern are specifically introduced. A new method of structural modeling is utilized to generate the structured derivation relationship, thus completing the natural language knowledge extraction process of the object-oriented knowledge system. An application example with mechanical CAD is used in this work to demonstrate the specific implementation of the example, which confirms the effectiveness of the proposed method.

1. Introduction

Clustering is a fundamental approach that explores data mining, image analysis, and various other pattern recognition methods for grouping the data into clusters. Natural language processing also deals with the introduction of clustering phenomenon utilizing the various aspects of deep learning [1]. The previous studies demonstrated the utilization of artificial intelligence [2, 3] methods in order to refine it to form structured knowledge using the primitive knowledge of human natural language. This link is called knowledge extraction. At present, in terms of natural language processing theory and technology, there are mainly formal language processing methods, semantic-based natural language understanding methods, and research on grammar inference. In terms of formal language processing, people have achieved many mature results, especially the processing technology of unrelated grammar, and the

compilation principal technology based on the derivation tree and automata theory has penetrated into various related fields of computer research. However, the formal language understanding mechanism requires a high degree of consistency in language expression and lacks flexibility, so it is lacking in the adaptability of natural language processing [4].

So far, there is still a lack of effective methodology for the determination of engineering characteristics in intelligent systems and realizing the function of knowledge extraction. In fact, the understanding of natural language is a formal matching process, and this understanding can be reduced to a cluster recognition process [5–7]. The basic block diagram of the natural language processing method using the cluster recognition approach is depicted in Figure 1.

For intelligent systems for engineering applications, the natural language concepts mainly involve knowledge concepts in a limited field and limited sentence expressions, showing obvious clustering characteristics in sentence structure. Therefore, the natural language vectorization method is adopted [8]. First, the natural language is vectorized, and then the adaptive neural network is used to cluster the vectors in the hyperspace of the obtained sentence vector. After the recognition processing, each group in the same cluster set is recognized. Sentences will have similar or the same way of understanding, which can alleviate the contradiction between the variability of natural language and the strictness of formal language recognition [9]. In addition, it is necessary to establish a dynamic update mechanism of the knowledge system, which can not only automatically create a structured knowledge system but also can dynamically structure and maintain the knowledge system in the intelligent system. According to the inclusive relationship between the various knowledge objects involved in natural language, structural modeling methods can be used to dynamically generate knowledge objects, their structure, and the relationship between them (such as the derived relationship between objects). Based on this knowledge structure generation mechanism, as long as a new knowledge description (such as natural sentences) appears, the system can describe the augmented knowledge. The collection is automatically reconstructed, and then an updated knowledge system is obtained.

This article proposes a dynamic knowledge extraction method based on sentence clustering recognition. The work provides a novel research framework of the dynamic knowledge extraction method while considering the conversion process from natural language processing to object-oriented knowledge system. In this work, sentence vectorization is studied, some basic definitions and a judgment theorem are given, and the postprocessing of sentence element attribute vectors is discussed. The credibility of sentence recognition results is achieved using the concept of prereliability while utilizing the ART2 network and MATLAB analysis. The article proposes a new method of structural modeling for the generation of structured derivation relationship, thus completing the natural language knowledge extraction process of the object-oriented knowledge system. The application example of mechanical CAD is considered in this work which uses the background throughout the text, demonstrates the specific implementation of the example, and confirms the effectiveness of the proposed method.

The rest of the article is arranged as follows: Section 2 discusses the review of the existing state-of-the-art work in the literature. Section 3 presents the research foundation of the knowledge dynamic extraction method followed by the statement clustering recognition based on neural network in Section 4. Section 5 provides the intermediate code generation and Section 6 analyzes the results of dynamic generation of object-oriented knowledge structure followed by the concluding remarks in Section 7.

2. Literature Review

Wang et al. considered the data of the P2P network loan platform as a reference and selected 24 typical P2P network loan platforms as the sample of empirical research in order to construct the comprehensive evaluation index system. Secondly, using factor analysis and cluster analysis, the comprehensive score and grade division of 24 P2P network loan platforms are given. Finally, according to the research results, some reasonable suggestions are put forward to provide a reference for investors to choose platform and platform business model optimization [10]. Sun et al. took the carrying capacity of resources and environment, the existing development density, and the development potential as the leading factors, combined with the development advantage and development orientation of Xuzhou, and established the evaluation index system of each region of Xuzhou. Using SPSS software, the whole Xuzhou city is divided into the development core area, the key development area, the optimization development area, and the restricted development area. Finally, the author put forward a way to optimize the allocation of land resources in Xuzhou so as to promote the industrial structure adjustment and economic development of Xuzhou [11]. Ran et al. compared the clustering results and algorithm efficiency of different clustering algorithms for a small number of load datasets and then compared the clustering results and clustering efficiency of these algorithms for a large number of load datasets. The results show that the partition-based analysis algorithm has a good clustering effect on load datasets. The user behavior patterns are divided into eight categories: bimodal, wind-avoiding, stationary, single-peak stable, single-peak, back-peak, fluctuating, and forward; finally, according to the results of cluster analysis, the single-peak type and single-peak stability are selected and double peak type and large power consumption customers are adjustable users. Different fault peak control schemes are formulated for this type of users, and the economic benefits brought by different control schemes are compared [12].

A variety of language processing methods have been developed for the search and classification of texts and to acquire dynamic connection knowledge within them [13]. Pimm et al. [14] presented an approach for automated analysis of linguistic features highlighting various aspects of challenges in this field. Tanguy et al. [15] proposed an
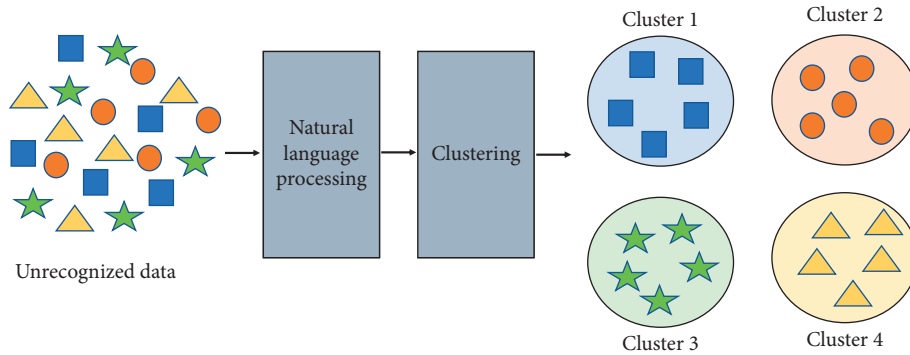
FIGURE 1: Block diagram of natural language processing method using cluster recognition.

application of natural language processing in the classification of aviation safety data. This method provides a way to analyze the text for the quantification of key elements in aviation incidents. Subramanian and Rao [16] developed a method for the categorization of different flight safety event categories for the extraction of textual narratives for time-series forecast of various incidents related to it. The clustering analysis has significantly been used for exploring the context of aviation for the prediction of accidental risk [17, 18]. Natural language processing has also found its applications in various domains of predictive operation in the airline industry for classification and pattern recognition [19, 20]. The literature has presented the usage of natural language processing in multiple domains, therefore encouraging its usage in dynamic knowledge extraction and clustering recognition. The utilization of natural language processing to object-oriented knowledge has also shown various manifold possibilities for investigation.

The literature suggests that the previous studies utilizing the neural network platforms largely focus on the typical clustering perspective of the networks. However, this article utilizes dynamic knowledge for addressing the challenges of natural language processing in the object-oriented knowledge system. A credible sentence clustering recognition method is presented in this work that uses the concept of prereliability while utilizing the ART2 network and MAT-LAB analysis.

3. Research Foundation of the Knowledge Dynamic Extraction Method

3.1. Research Framework. The basic steps and functions of knowledge extraction based on sentence clustering recognition are as follows: directly input the Chinese natural language in a human-computer interaction environment. Then, use the background thesaurus and Chinese word segmentation algorithm to segment the initial sentence into keywords and punctuation. The sequence of sentence elements is composed of symbols and the vector expression of sentences is obtained through the nonmonotonic sentence vectorization method based on sentence elements. In view of the randomness of natural language, the pattern recognition function of the neural network is used to analyze sentence

clustering [21–23]. In turn, various static and dynamic knowledge structures are obtained, and finally, the structural modeling method is used to generate an object-oriented knowledge system based on the mutually inclusive relationship between static knowledge structures. The dynamic knowledge that has been obtained is filled into the object body so as to realize the knowledge dynamic extraction of sentence clustering recognition. The overall framework of this research is shown in Figure 2.

3.2. Related Definitions and Theorems

Definition 1. Sentence element (the smallest logical unit of a sentence): a sentence through word segmentation consists of a series of words and punctuation marks.

To complete, use Li to represent the words or punctuation in the sequence and define Li as a sentence element. Such a sentence can be expressed as $L_1 \longrightarrow L_2 \longrightarrow \ldots \longrightarrow L_n$. The sentence element is the smallest unit of the logical meaning of the sentence, which is the grammatical analysis and semantic analysis. The smallest segmentation is the primitive of "meaningful sentences."

Definition 2. The attributes of the sentence element: the attribute x of the sentence element L reflects the possible functional categories of the sentence element in the sentence, and it is the functional description of the sentence element [24–26].

Definition 3. The attribute set of the sentence element: the attribute set of the definition sentence element L is $X = \{x_1, x_2, \ldots, x_m\}$. The sentence element attribute set reflects all the sentence elements in the sentence possible functional categories; it is a description of all the functions of the sentence element.

Definition 4. The certainty of the attributes of a sentence element: the certainty of the attribute of a sentence element CF is a fuzzy measure of the effect of each attribute of a sentence element in a sentence, which satisfies the following expression:
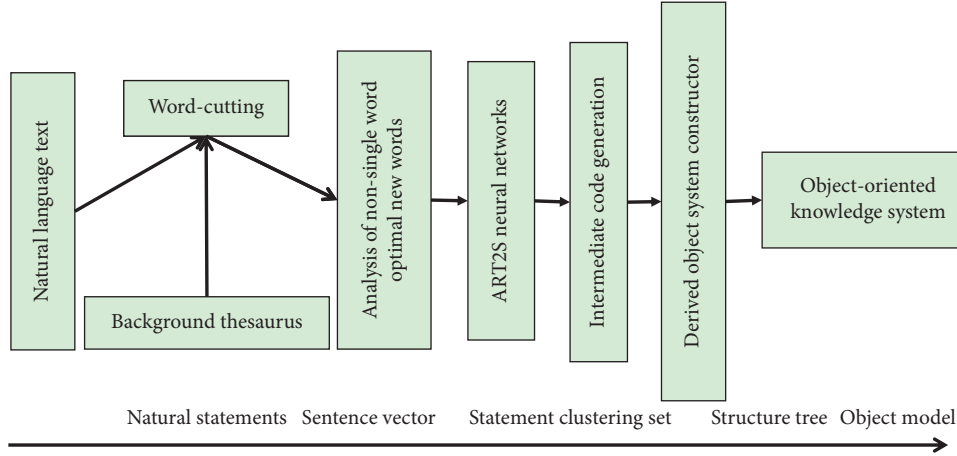
FIGURE 2: Research framework of knowledge extraction based on cluster recognition.

$$0 \leq CF \leq 1,$$
$$\sum_{i=1}^{m} CF_i = 1. \tag{1}$$

Definition 5. The attribute support of sentence element $\tilde{x}$ is given by the following expression:

$$\tilde{X} = \{x | \forall x \in X, x' \in X, x \neq x', \exists CF_x > CF_{x'}\}. \tag{2}$$

Definition 6. The certainty of sentence element attribute support CF is given by the following:

$$CF = CF_x. \tag{3}$$

Among them, $x \in \tilde{x}$.

Theorem 1. *The sufficient conditions for the judgment of sentence ambiguity are as follows:*

(1) In the sentence element sequence corresponding to the sentence, the number of elements with the attribute support of a sentence element is greater than 1.

(2) There is a sentence element attribute support set. The degree of certainty CF is less than the fuzzy level value K.

3.3. Postprocessing of Sentence Element Attribute Vector. The method of sentence vectorization has been introduced in detail in the literature. The basic step is to divide a sentence into a sequence of several sentence elements through a word segmentation algorithm and then use a nonmonotonic optimal part-of-speech analysis model to determine each sentence subjected to the functional attributes of the element. Obviously, the dimension of the sentence element attribute vector obtained in this way is equal to the length of the sentence element sequence generated by the word segmentation, but some elements in this vector can also be combined to a certain extent [27–29]. Sentences are a kind of division. Layer structure, under

different granularities, in the same sentence can present different content. Generally speaking, the more granular the level of analysis of the sentence, the richer the content of the sentence and the more accurate the understanding of the sentence. Sentence element postprocessing determines at which granular level the sentence is to be recognized and analyzed, that is, to extract the main components of the sentence and remove the secondary components. At the same time, the sentence element postprocessing can also remove the natural language. At a certain level of granularity, the function is equivalent, resulting in repeated ingredients.

The following are a few typical postprocessing rules:

(1) Combine the "{adverb} + adjective" component in the vector corresponding to the parallel modification to make it an abstract "{adverb} + adjective" modifier component, that is, a subvector in the sentence vector.

(2) Put multiple nonrestrictive attributive clauses after the commas in the vector are merged into an abstract nonrestrictive attributive subvector. (3) Multiple nouns are combined as attributive components and abstracted into a modifier subvector. The following is an application example with mechanical CAD as the background. This example will run through the full text to illustrate the specific implementation of the method in this article. Suppose there are the following two sentences a and *b*, which are regarded as standard sample sentences [30–32].

Statement a: if attribute B of object A is C, then attribute *E* of object *D* is F. (a-0)
Statement b: attribute H of object *G* is I, if and only if attribute K of object *J* is *L*. (b-0)

According to the word segmentation algorithm, their respective sentence element sequences can be obtained as follows:

if" $\longrightarrow$ object A $\longrightarrow$ attribute B $\longrightarrow$ C $\longrightarrow$ "then" $\longrightarrow$ object D $\longrightarrow$ attribute E $\longrightarrow$ F $\longrightarrow$ (a-1)

Object $G \longrightarrow$ attribute H $\longrightarrow$I $\longrightarrow$ only if" $\longrightarrow$ object J $\longrightarrow$attribute K $\longrightarrow$L $\longrightarrow$ (B-1)

Among them, "if," "then," and "only if" are keywords retrieved from the background thesaurus.

Set the quantized value of the abstract relational operator to 1, the quantized value of the logical operator is 2, the quantized value of the keyword "if" is 3, the quantized value of the keyword "then" is 4, and the quantized value of the comma is 5. The quantized value of the period is 6, the quantized value of the comma is 7, the quantized value of the general noun is 8, the quantized value of the keyword "only if" is 9, the quantized value of the keyword " = " is 10, rhe quantized value of the keyword ">" is 11, the quantized value of the keyword "<"is 12, etc. Based on this, the vectorized result of the sample sentence a, b is as follows:

Statement a: 3 2 8 2 8 2 10 2 8 2 4 2 8 2 8 2 10 2 8 2 6 (a-2).

Statement b: 8 2 8 2 10 2 8 2 9 2 8 2 8 2 10 2 8 2 6 (b-2).

Here is a sentence to be processed:

Statement c: if the model of the milling machine is small, its feed type is desktop feed, and its table width is greater than 80.0, then the motor type of the continuously variable motor is "Z2 2 61," and its motor _ j is equal to 0.56, its motor power is 5.5, its maximum motor speed is equal to 1000, and its minimum motor speed is equal to 10. (c-0).

By cutting words, you can get its sentence element vector as "if" $\longrightarrow$ milling machine $\longrightarrow$ milling machine model $\longrightarrow$ "=" $\longrightarrow$ small $\longrightarrow$ " and " $\longrightarrow$ it $\longrightarrow$ feed type $\longrightarrow$ "=" $\longrightarrow$ table feed $\longrightarrow$ "and" $\longrightarrow$ it $\longrightarrow$Workbench width $\longrightarrow$ ">" $\longrightarrow$ 800. 0 $\longrightarrow$ " then" $\longrightarrow$ Infinitely variable motor $\longrightarrow$ Motor type $\longrightarrow$ "=" $\longrightarrow$Z2261 $\longrightarrow$ "and" $\longrightarrow$ its $\longrightarrow$ mo2tor _j $\longrightarrow$ "=" $\longrightarrow$ 0.56 $\longrightarrow$ "and" $\longrightarrow$ it $\longrightarrow$ motor power $\longrightarrow$ "=" $\longrightarrow$ 5. 5 $\longrightarrow$ "and" $\longrightarrow$ it $\longrightarrow$ maximum motor speed $\longrightarrow$ "=" $\longrightarrow$ 1000$\longrightarrow$ "and" $\longrightarrow$ its $\longrightarrow$ minimum motor speed $\longrightarrow$ "=" $\longrightarrow$ 10 $\longrightarrow$. (c-1).

After part-of-speech analysis and postprocessing, the parallel modifier in sentence c has been reduced to an abstract modifier, and the vectorized result is 3 2 8 2 8 2 1 2 8 2 4 2 8 2 8 2 1 2 8 2 6 (c-2).

4. Statement Clustering Recognition Based on Neural Network

4.1. Selection and Design of Neural Network.
The sentence element attribute vector obtained in the previous section $Y = (\widetilde{X_1}, \widetilde{X_2}, \widetilde{X_3}, \ldots, \widetilde{X_n})$ represents a point in n-dimensional space, for a set of M sentence element attribute vector s$J = \{Yi | 1 \le i \le M\}$. In other words, it is the mapping result of the initial natural sentence to this n-dimensional vector space. Regarding these points as pattern samples, the patterns with the same category or some similar characteristics are also relatively close in the n-dimensional space. These points are in the space and the clusters are distributed in clusters. To determine the location and distribution of each cluster, you first need to determine the typical samples of the cluster and treat them as the center of each cluster. These

typical samples may be one of the sample sets, or it may be a vector calculated from each cluster sample and located at the center of each sample group [33, 34].

This article uses Carpenter and Grossberg's adaptive resonance network ART2 and Kohonen learning method to measure the Euclidean distance of each point in n-dimensional space. As long as any mode is located in a hypersphere with a certain radius near the center of a cluster, it is considered to have the mode characteristics of this cluster; that is, each vector point in the same cluster has the same sentence function structure.

4.2. Front Reliability.
Set the critical cluster radius of a sentence cluster set to R, and the Euclidean distance from the sentence vector to the cluster and cluster center to dst.Then, the prereliability of sentence recognition can be defined [35–44].

Definition 7. The prereliability PT is a function of dst, which satisfies the following conditions: (1) PT (dst > R) < 0; (2) PT (dst = R) = 0; (3) PT (0<dst < R) ∈ (0, 1); (4) PT (dst = 0) = 1.

The smaller the distance from the sentence vector to the center of the cluster, the higher its prereliability. When the sentence vector is outside the critical cluster radius, its prereliability is negative; that is, it is not credible. The form can be linear or nonlinear, depending on the actual situation. Obviously, prereliability is a measure of sentence recognition results.

4.3. The Effect of ART2 Network on Sentence Recognition.
Continue to use the example in the previous section. Among them, three sentence vectors (c-2), (a-2), and (b-2) are not equal, but in this simpler example, it can be seen that (c-2) and (a-2) have greater similarities. Accordingly, MATLAB is used to write an ART2 network simulation program, which clusters the input sentence vector and graphics function to convert the sentence into dimensional space. The problem of clustering recognition in dimensional space is represented by a plane diagram (see Figures 3–6, corresponding to the number of adaptive learning times 400, 600, 800, and 1000, respectively), which is the projection result of this high-dimensional problem on a hyperplane. In the figures, the two asterisks on the left represent sentences (a-2) and (c-2), the asterisk on the right represents sentences (b-2), and the two circles represent the adaptive output pass vector of the ART2 network. They also indicate that the center of each cluster under the current adaptive learning times can be known from the above figures. Through cluster recognition, it is found that (c-2) and (a-2) belong to the same cluster set, so you can follow the semantic model of sentence a used to construct the semantic model of sentence c, which is also consistent with the result of direct observation.

In addition, we also used the ART2 network to recognize sentences with vector dimensions. Figure 7 shows the recognition effect with the vector dimension and the number of neural network output nodes as independent variables. In actual processing, the sentences with different vector dimensions are expanded to 20 dimensions and recognized by the ART2 network. Therefore, the vector dimension in
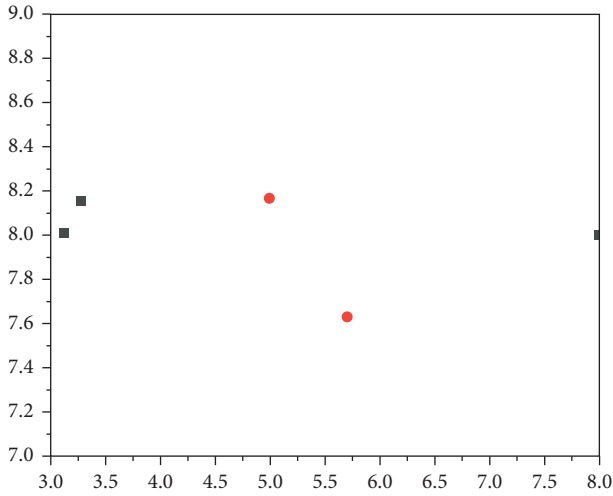
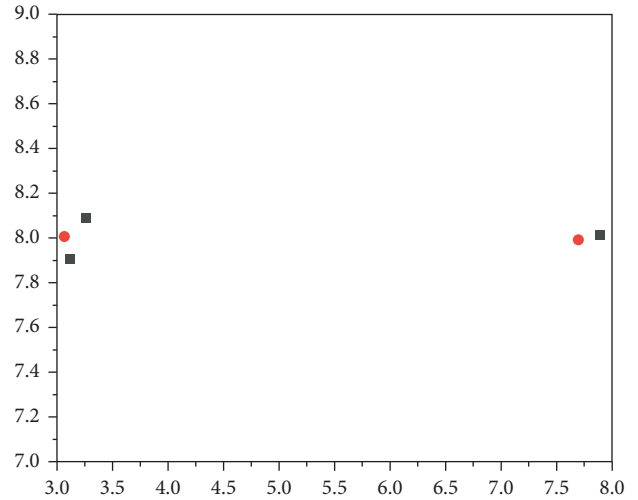FIGURE 3: The number of adaptive learning times 400.
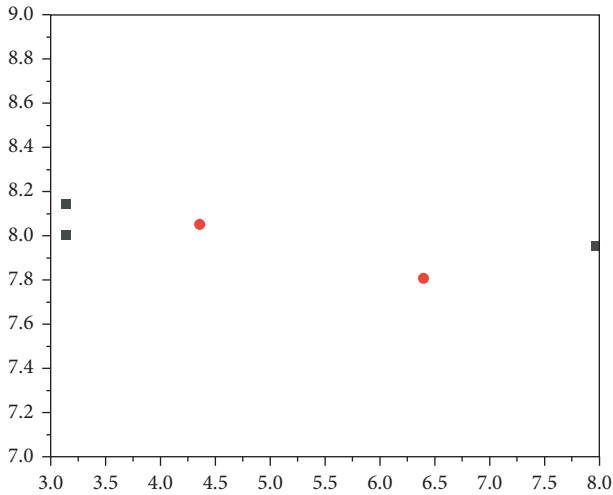

FIGURE 4: The number of adaptive learning times 600.

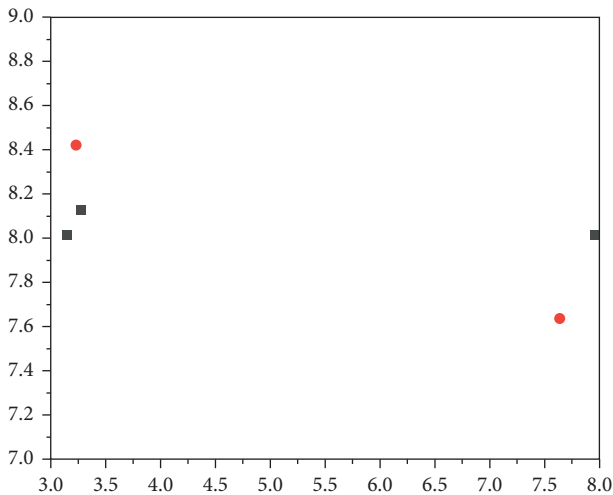FIGURE 5: The number of adaptive learning times 800.

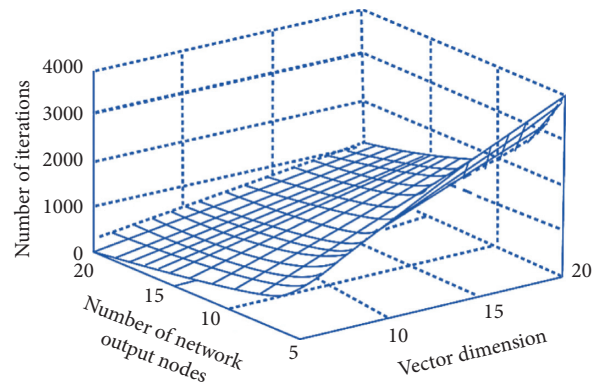FIGURE 6: The number of adaptive learning times 1000.

FIGURE 7: The clustering recognition effect of ART2 network.

Figure 7 actually refers to the effective vector dimension before the expansion transformation.

From the surface characteristics shown in Figure 7, the following can be seen:

(1) In the case of a certain vector dimension, the fewer the number of output nodes of the neural network, the slower the recognition speed of the network; conversely, the more the number of output nodes of the neural network, the faster the recognition speed of the network.

(2) When the number of network output nodes is constant, the number of iterations identified by the neural network is basically proportional to the vector dimension.

(3) When the number of spatial clusters is much larger than the number of network output nodes, the recognition effect of the ART2 network gradually deteriorates. Based on this, the following conclusions can be drawn: in order to ensure the recognition effect of the ART2 network on sentences (including recognition accuracy and recognition speed), the

network must have a sufficient number of output nodes, generally not less than the number of standard sentence samples.

5. Intermediate Code Generation

Intermediate code generation is based on the result of sentence recognition by the ART2 network, converting each sentence after clustering into knowledge form and generating a knowledge expression form that can be recognized by subsequent processing. Here is a "conjunctive" form of rule knowledge expression; the details give the intermediate code generation method. Regarding the intermediate code generation method of other forms of statements, similar promotion can be made according to the specific situation.

5.1. Width-First Method of Intermediate Code Generation. The sentence is regarded as a tree structure, the sentence elements belonging to the main sentence component are located close to the root node of the tree structure, and the sentence elements belonging to the secondary sentence component are located far away from the root node of the tree structure. This can be based on intelligence and the breadth-first method in the search generates the corresponding intermediate code structure while traversing the sentence structure tree. The specific steps are as follows:

Step 1. Construct the structure tree of the sentence corresponding to the "sample vector" in the cluster, and set N as the depth of the structure tree.

Step 2. Construct the intermediate code framework corresponding to the sentence structure tree corresponding to the "sample vector" in the cluster.

Step 3. Set the level depth pointer $n = 0$.

Step 4. According to the sentence elements of the nth layer in the sample sentence structure tree, according to certain rules, extract the corresponding sentence elements of the currently processed sentence (nonsample sentence), and put them into the corresponding unit of the sample intermediate code framework obtained in Step 2.

Step 5. n plus 1. If $n < N$, then return to Step 4; otherwise, go to

Step 6. End.

5.2. Width-First Code Generation Method Based on Conjunctive Rules. In an intelligent system, the "conjunctive" form of the rule structure is shown in Figure 8. The intermediate code generation method for conjunctive rules is given as follows:

Step 1. Construct a structure tree of sample sentences expressing conjunctive rules.

Step 2. Construct the intermediate code framework corresponding to the conjunctive rule sample sentence structure tree.

Step 3. Read the uppermost sentence elements "if" and "then" in the processed nonsample sentences, and put them into the uppermost unit of the intermediate code framework.

Step 4. Read the sentence elements of each expression object in the processed sentence and put them into the sublevel unit of the intermediate code framework.

Step 5. Read the sentence elements of each attribute in the processed sentence and put them into the lower unit of the corresponding object sentence element of the intermediate code frame.

Step 6. End.

Continue to use the previous example, first construct the structure tree of sample sentences a and b, as shown in Figures 9 and 10.

According to the result of the ART2 pattern recognition, the recognized sentence c and the sample sentence a belong to the same cluster set. Therefore, the semantic structure of sentence c can be constructed by the breadth-first method according to the structure tree of sentence a, and the structure tree is shown in Figure 11.

5.3. Postreliability. For a sentence clustering set, if the sample vector is set to be the cluster center, then there are two situations that may occur through the width-first method generated by the above code:

(1) For nonsample sentences with previous reliability less than 1, the structure tree is different from the structure tree of sample sentences; that is, there may be nodes that are not filled.

(2) Of course, there may also be cases where the content of nonsample sentences is richer than that of sample sentences. In this case, the width-first method will cause part of the semantics to be lost. For the former case, the definition of posterior reliability is given as follows.

Definition 8. Postreliability $AT = \sum q(i)$.

Here, $q(i)$ is the weight of the node i actually filled in the structure tree obtained by the width-first method for the nonsample sentence. Let the weight of the node i of the structure tree of the sample sentence be $q(i)$, then the sum of the weights of each node of the tree is obviously 1, that is, $\sum q(i) = 1$. For nonsample sentences, the sum of the weights of each node actually filled by the structure tree obtained according to the width-first method should satisfy $\sum q(i) \le 1$. It can be seen that postreliability is a final evaluation index for the credibility of sentence recognition and semantic model construction. For nonsample sentences with prereliability of 1, the final postreliability must also be 1. For sentences with high prereliability, the corresponding postreliability will also be higher in a statistical sense. Therefore, the prereliability and postreliability are both beliefs about sentence recognition and semantic model construction incomplete matching methods and degree measurement.
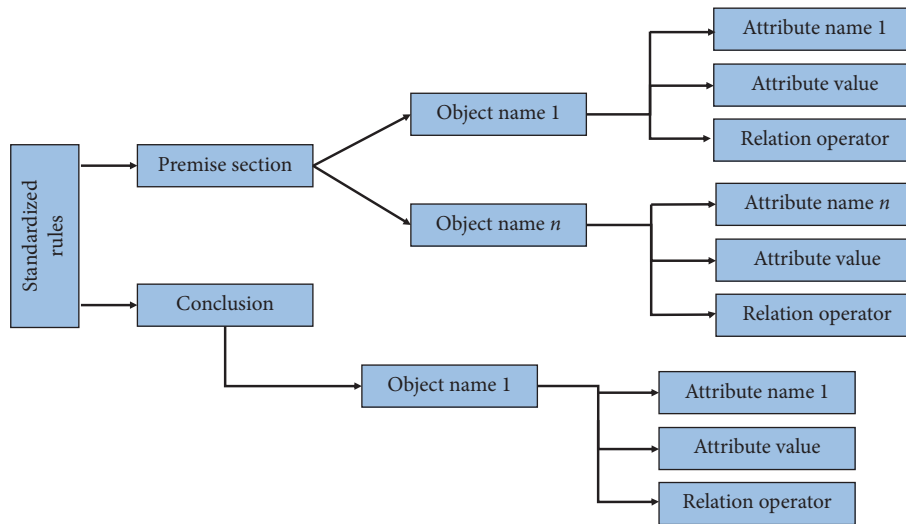
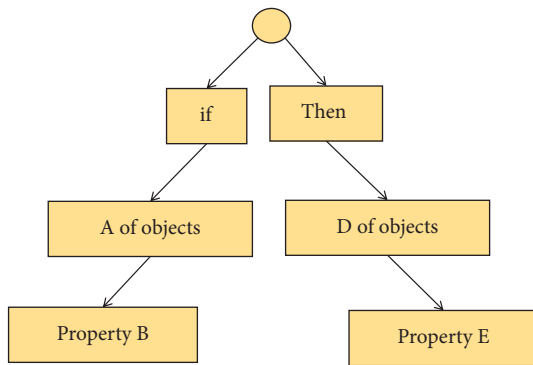FIGURE 8: The structure of the "conjunctive" formal rule.
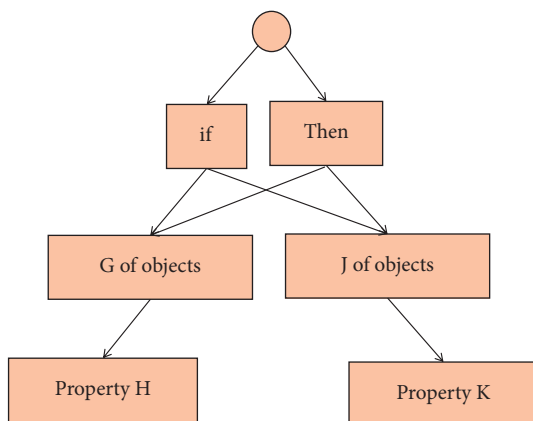

FIGURE 9: Structure tree of statement a.

FIGURE 10: Structure tree of sentence b.

6. Dynamic Generation of Object-Oriented Knowledge Structure

6.1. Conventions of the Object Model. The object is used as the basic constituent unit of the knowledge system, and the object is composed of attributes, rules, and methods.

(1) Attribute: an attribute is a combination of attribute name, attribute type, attribute value, and operation type, which reflects the static knowledge structure of the object. The attribute type can be numeric, Boolean, string, or multimedia information; of course, it can also be an object.

Definition 9. Attributes are equal.

Attribute A is equal to attribute B if and only if their names, types, values, and operation types are equal respectively, denoted as Attr A = Attr B.

(2) Rules: they reflect the dynamic knowledge of different attributes in the contact object or the same object. In actual use, the main purpose of the rules is to provide a basis for filling in some unknown object attributes according to known conditions, thus reflecting links between static knowledge structures.

(3) Method: it embodies the steps taken by humans to solve specific problems. It finally achieves specific functions by alternately calling rule inference engines and various numerical calculation tools.

Definition 10. Object reachability relationship.

Suppose two objects OBJ1 and OBJ2; for each attribute Attr2j ∈ OBJ2 in OBJ2, there exists an Attr 1i ∈ OBJ1, which satisfies Attr2 j = A ttr 1i. It is said that OBJ1 can reach OBJ2 (that is, the attributes of OB J1 include the attributes of OBJ2).

6.2. Steps to Structure Object-Oriented Knowledge. For using our new structural modeling method to further structure object-oriented knowledge, the steps are as follows:

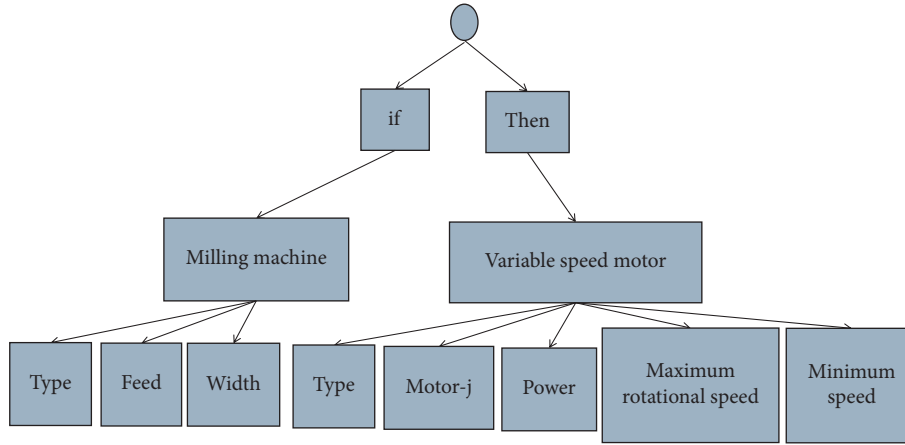*Step 1.* Read each object structure generated by the intermediate code.

FIGURE 11: Structure tree of sentence c to be recognized.

Step 2. According to the static knowledge containment relationship between the objects, use the structural modeling method to find the inheritance relationship between the objects and then obtain the derived object architecture. The decomposition steps are as follows:

2.1. First, according to the static structure of the object, construct the system reachability matrix according to Definition 10.
2.2. Find the strongly connected subset of the system. Each object in the same strongly connected subset has the same static knowledge structure, but their dynamic knowledge is generally different.
2.3. Pick any element from each strongly connected subset to characterize the subset it is in, which constitutes a reduced system of the original system.
2.4. Perform area division, if and only if there is an inheritance relationship between objects in the same area.
2.5. Find the skeleton matrix of the reduced system.
2.6. From the element a $ij = 1$, which is 1 in the skeleton matrix, determine that element j is the parent of element i, thereby generating the inheritance system of the object.

Now, it is still explained by the example in mechanical CAD; suppose the milling machine is represented by $x1$; the continuously variable motor is represented by $x2$; the model is "small," and the milling machine whose feed type is "table feed" is a milling machine 1. It is expressed by $x3$; the model is "medium," the feed type is "table feed," and the milling machine whose worktable width is less than 2500. 0 is milling machine 2. It is expressed by $x4$; the model is "small," the feed type is "table feed," and the milling machine with a worktable width greater than 800. 0 is milling machine three, which is represented by $x5$; because of the different parameters of the continuously variable motor, there are motor one, motor two, and motor three, represented by $x6$, $x7$, and $x8$, respectively. They form system $S = \{x1, x2, x3, x4, x5, x6, x7, x8\}$; for convenience, use subscripts instead of elements; that is, $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$. In terms of statistics and the mutually inclusive relationship between its static knowledge structures, the reachable matrix of S is obtained as follows:

$$
Re = \begin{matrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{matrix}, \tag{4}
$$

Find from Re

$$
Re \cap R_e^T = \begin{matrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{matrix}. \tag{5}
$$

Among the row vectors of composition $Re \cap R_e^T$, there is no row vector that is not equal to the other and has more than one component of 1. According to the literature inference 1, this result indicates that there is no strong connected subset in the system.

Among the row vectors of composition $t (R_e \cup R_e^T)$, the ones that are not equal to each other are $(1, 0, 1, 1, 1, 0, 0, 0)$ and $(0, 1, 0, 0, 0, 1, 1, 1)$. According to inference 2, this result shows that the system consists of two nonmolecular sets $\{1, 3, 4, 5\}$ and $\{2, 6, 7, 8\}$, which are two regions of the system. $R_e^{'}$ can be found from $R_e^{'} - I_m$. Then, inferred from literature [3] and example $R_e = R_e^{'}$, the following is available:
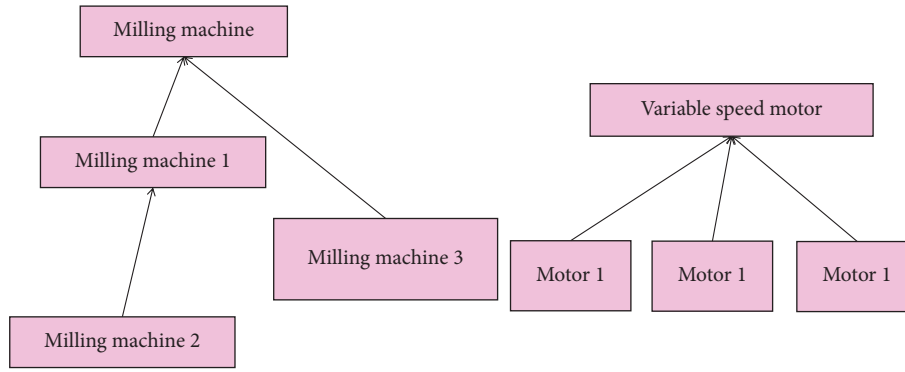
FIGURE 12: Object inheritance relationship.

TABLE 1: Comparative analysis of presented object-oriented knowledge-based system.

Method	Clustering algorithm	Outcomes
Yang et al. [45]	Centroid update k-means	Well known neural network-based clustering method but a limited amount of data is utilized
Kilinc and Uysal [46]	Soft clustering	Utilized the nonparametric clustering approach, which reduces model complication
Hsu and Lin [47]	k-means	Robust outcomes are obtained using the combination of k-means and feature attribute engineering
Method presented in this article	Carpenter and Grossberg's adaptive resonance neural network ART2 and Kohonen learning method	Established a knowledge dynamic extraction framework for conversion of natural language to the object-oriented knowledge system

$$S_k' = \left(R_e' - I_m\right) - \left(R_e - I_m\right)^2 = \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}. \tag{6}$$

Because system S does not have a strongly connected subset, the skeleton matrix Sk is the same as the reduced skeleton matrix $S'k$. According to the above calculation results, the object inheritance relationship shown in Figure 12 can be obtained, where the arrow points to the parent node, so the final object-oriented structured knowledge system is obtained [45–55].

6.3. Validation of the Object-Oriented Knowledge-Based System against Other Methods. For the validation of the object-oriented knowledge-based system presented in this article, various other methods have been reviewed in the literature. Some of the methods are compared in Table 1 to the technique presented in this article for the verification of the proposed approach.

This tabular comparison reveals the effectiveness of the presented method evaluating its credibility for sentence recognition and semantic model construction. This method utilizes structural modeling in order to create a link between natural language processing and the object-oriented knowledge system [56–60].

7. Conclusion

This article proposes a method of dynamic knowledge extraction based on sentence clustering recognition and neural networking. The proposed system is based on the clustering phenomenon of natural language in the engineering field and accomplishes the requirements for the dynamic update of the knowledge system. The credibility of the sentence clustering recognition method is demonstrated using the concept of prerealiability while utilizing the Carpenter and Grossberg adaptive resonance neural network ART2 and Kohonen learning method. The work presented in this article establishes a research framework of knowledge dynamic extraction, which demonstrates the conversion process from natural language documents to object-oriented knowledge system. The conversion of the knowledge form of each sentence after clustering is done utilizing the breadth-first method of intermediate code generation. It was analyzed from the compilation of the ART2 neural network program that in order to ensure the recognition effect of ART2 network on sentence (including recognition accuracy and recognition speed), the network must have sufficient output. The number of nodes should generally be no less than the

number of standard sentence samples. This work proves the effectiveness of the proposed method through an application example. The concepts of prereliability and postreliability are defined in this work to measure and evaluate the credibility of sentence recognition and semantic model construction. The structural modeling method is used to generate structured derivation relationships, thereby completing the knowledge extraction process from natural language processing to object-oriented knowledge systems.

Data Availability

All data are available within the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] M. R. Karim, O. Beyan, A. Zappa et al., "Deep learning-based clustering approaches for bioinformatics," *Briefings in Bioinformatics*, vol. 22, no. 1, pp. 393–415, 2021.

[2] J. Dogra, S. Jain, A. Sharma, R. Kumar, and M. Sood, "Brain tumor detection from MR images employing fuzzy graph cut technique," *Recent Advances in Computer Science and Communications*, vol. 13, no. 3, pp. 362–369, 2020.

[3] G. Dhiman, D. Oliva, A. Kaur et al., "BEPO: a novel binary emperor penguin optimizer for automatic feature selection," *Knowledge-Based Systems*, vol. 211, p. 106560, 2021.

[4] W. Wan, Q. Zeng, and Z. Wen, "Research on software runtime credibility based on cluster analysis," *IOP Conference Series Materials Ence and Engineering*, vol. 768, Article ID 072115, 2020.

[5] G. Dhiman, K. K. Singh, M. Soni et al., "MOSOA: a new multi-objective seagull optimization algorithm," *Expert Systems with Applications*, vol. 167, Article ID 114150, 2020.

[6] G. Rathee, A. Sharma, H. Saini, R. Kumar, and R. Iqbal, "A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology," *Multimedia Tools and Applications*, vol. 79, pp. 9711–9733, 2020.

[7] G. Rathee, A. Sharma, R. Kumar, F. Ahmad, and R. Iqbal, "A trust management scheme to secure mobile information centric networks," *Computer Communications*, vol. 151, pp. 66–75, 2020.

[8] Q. Tian, Y. Mei, Y. Du, and X. Zhou, "Research on large capacity coupling design task planning based on cluster analysis," *Zhongguo Jixie Gongcheng/China Mechanical Engineering*, vol. 29, no. 5, pp. 544–551, 2018.

[9] W. Chang, Z. Xu, S. Zhou, and W. Cao, "Research on detection methods based on doc2vec abnormal comments," *Future Generation Computer Systems*, vol. 86, no. SEP, pp. 656–662, 2018.

[10] K. Wang, J. Zhu, and J. Lu, *Research on P2P Lending Platform Based on Factor Analysis and Cluster Analysis*, vol. 34, pp. 33–39, Harbin Normal University, Harbin, China, 2018.

[11] Q. Sun, W. Zhu, q. Hu, and C. Xu, "Research on r main functional zoning of Xuzhou City based on system clustering analysis research on xuzhou's main function zoning based on system cluster analysis," *Land and Natural Resources Research*, vol. 000, no. 004, pp. 33–37, 2018.

[12] C. S. Ran, Y. Liu, and L. Zhao, "Based on cluster analysis, power pattern discrimination research of electricity consumption pattern recognition based on cluster analysis," *Guizhou Electric Power Technology*, vol. 022, no. 004, pp. 43–49, 2019.

[13] K. R. Chowdhary, "Natural language processing," in *Fundamentals of Artificial Intelligence*, pp. 603–649, Springer, New Delhi, India, 2020.

[14] C. Pimm, C. Raynal, N. Tulechki, E. Hermann, G. Caudy, and L. Tanguy, "Natural Language Processing (NLP) tools for the analysis of incident and accident reports," in *Proceedings of the International Conference on Human-Computer Interaction in Aerospace (HCI-Aero)*, Brussels, Belgium, September 2012.

[15] L. Tanguy, N. Tulechki, A. Urieli, E. Hermann, and C. Raynal, "Natural language processing for aviation safety reports: from classification to interactive analysis," *Computers in Industry*, vol. 78, pp. 80–95, 2016.

[16] S. V. Subramanian and A. H. Rao, "Deep-learning based time series forecasting of go-around incidents in the national airspace system," in *Proceedings of the 2018 AIAA Modeling and Simulation Technologies Conference*, p. 0424, Kissimmee, FL, USA, January 2018.

[17] A. O. Alkhamisi and R. Mehmood, "An ensemble machine and deep learning model for risk prediction in aviation systems," in *Proceedings of the 2020 6th Conference on Data Science and Machine Learning Applications (CDMA)*, pp. 54–59, IEEE, Riyadh, Saudi Arabia, March 2020.

[18] X. Zhang and S. Mahadevan, "Ensemble machine learning models for aviation incident risk prediction," *Decision Support Systems*, vol. 116, pp. 48–63, 2019.

[19] P. Korvesis, S. Besseau, and M. Vazirgiannis, "Predictive maintenance in aviation: failure prediction from post-flight reports," in *Proceedings of the 2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pp. 1414–1422, IEEE, Paris, France, April 2018.

[20] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 140–150, 2019.

[21] X. Huang, J. Jiang, D. Zhao, Y. Feng, and Y. Hong, "A novel community detection method based on cluster density peaks," *Natural Language Processing and Chinese Computing Volume 10619 (Lecture Notes in Computer Science)*, pp. 515–525, 2018.

[22] F. Mao, "Research on the second-hand housing recommendation based on cluster analysisr—taking Beijing as an example," *Science and Technology Entrepreneurship Journal*, vol. 31, no. 5, pp. 149–153, 2018.

[23] C. Cheryl, "26. Novel approaches to psychosis risk: movement, stress modulation, reward and language," *Schizophrenia Bulletin*, vol. 44, no. S1, p. S42.

[24] M. Bibi, W. Aziz, M. Almaraashi, I. H. Khan, M. S. A. Nadeem, and N. Habib, "A cooperative binary-clustering framework based on majority voting for twitter sentiment analysis," *IEEE Access*, vol. 8, pp. 68580–68592, 2020.

[25] N. Ouerhani, A. Maalel, and H. Ben Ghézela, "Spececa: a smart pervasive chatbot for emergency case assistance based on cloud computing," *Cluster Computing*, pp. 1–12, 2019.

[26] J. Chen, U. Liji, H. Wang, and Z. Yan, "Community mining in signed networks based on dynamic mechanism," *IEEE Systems Journal*, vol. 13, no. 1, pp. 447–455, 2019.

[27] W. Hu, G. Tian, Y. Kang, C. Yuan, and S. Maybank, "Dual sticky hierarchical dirichlet process hidden markov model and its application to natural language description of motions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 10, pp. 2355–2373, 2018.

[28] G. D. Chavan, R. Chaudhuri, and W. J. Johnston, "Industrial-buying research 1965-2015: review and analysis," *Journal of Business & Industrial Marketing*, vol. 34, no. 1, pp. 205–229, 2019.

[29] M. Alhawarat and M. Hegazi, "Revisiting k-means and topic modeling, a comparison study to cluster arabic documents," *IEEE Access*, p. 1, 2018.

[30] C. R. Green, "On the link between onset clusters and codas in mbat (jarawan Bantu)," *Natural Language & Linguistic Theory*, pp. 1–26, 2020.

[31] P. Yang, Y. Yao, and H. Zhou, "Leveraging global and local topic popularities for lda-based document clustering," *IEEE Access*, vol. 8, pp. 24734–24745, 2020.

[32] P. Das and A. K. Das, "Graph-based clustering of extracted paraphrases for labelling crime reports," *Knowledge-Based Systems*, vol. 179, pp. 55–76, 2019.

[33] M. M. Kubek, T. Böhme, and H. Unger, "Empiric experiments with text-representing centroids," *Theory and Application of Text-Representing Centroids*, vol. 863, pp. 39–54, 2019.

[34] D. Robert, O. Mira, and D. Lisa, "On the relation between speech perception and loanword adaptation: cross-linguistic perception of Korean-illicit word-medial clusters," *Natural Language & Linguistic Theory*, vol. 37, pp. 825–868, 2018.

[35] M. Krichen, S. Mechti, R. Alroobaea et al., "A formal testing model for operating room control system using internet of things," *Computers, Materials & Continua*, vol. 66, no. 3, pp. 2997–3011, 2021.

[36] O. I. Khalaf, K. A. Ogudo, and M. Singh, "A fuzzy-based optimization technique for the energy and spectrum efficiencies trade-off in cognitive radio-enabled 5G network," *Symmetry*, vol. 13, no. 1, p. 47, 2021.

[37] O. I. Khalaf, F. Ajesh, A. A. Hamad, G. N. Nguyen, and D.-N. Le, "Efficient dual-cooperative bait detection scheme for collaborative attackers on mobile ad-hoc networks," *IEEE Access*, vol. 8, pp. 227962–227969, 2020.

[38] A. A. Hamad, A. S. Al-Obeidi, E. H. Al-Taiy, O. I. Khalaf, and D. Le, "Synchronization phenomena investigation of a new nonlinear dynamical system 4d by gardano's and lyapunov's methods," *Computers, Materials & Continua*, vol. 66, no. 3, pp. 3311–3327, 2021.

[39] O. Wisesa, A. Adriansyah, and O.I. Khalaf, "Prediction analysis sales for corporate services telecommunications company using gradient boost algorithm," in *Proceedings of the 2nd International Conference on Broadband Communications, Wireless Sensors and Powering, BCWSP 2020*, pp. 101–106, Yogyakarta, Indonesia, September 2020.

[40] A. F. Subahi, Y. Alotaibi, O. I. Khalaf, and F. Ajesh, "Packet drop battling mechanism for energy aware detection in wireless networks," *Computers, Materials and Continua*, vol. 66, no. 2, pp. 2077–2086, 2020.

[41] X. Xiang, Q. Li, S. Khan, and O. I. Khalaf, "Urban water resource management for sustainable environment planning using artificial intelligence techniques," *Environmental Impact Assessment Review*, vol. 86, p. 106515, 2021.

[42] O. I. Khalaf and G. M. Abdulsahib, "Energy efficient routing and reliable data transmission protocol in WSN," *International Journal of Advances in Soft Computing and Its Application*, vol. 12, no. 3, pp. 45–53, 2020.

[43] O. I. Khalaf, G. M. Abdulsahib, and B. M. Sabbar, "Optimization of wireless sensor network coverage using the bee algorithm," *J. Inf. Sci. Eng.*vol. 36, no. 2, pp. 377–386, 2020.

[44] S. K. Prasad, J. Rachna, O. I. Khalaf, and D.-N. Le, "Map matching algorithm: real time location tracking for smart security application," *Telecommunications and Radio Engineering*, vol. 79, no. 13, pp. 1189–1203, 2020.

[45] J. Yang, D. Parikh, and D. Batra, "Joint unsupervised learning of deep representations and image clusters," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5147–5156, Las Vegas, NV, USA, June 2016.

[46] O. Kilinc and I. Uysal, "Learning latent representations in neural networks for clustering through pseudo supervision and graph-based activity regularization," 2018, https://arxiv.org/abs/1802.03063.

[47] C. C. Hsu and C. W. Lin, "CNN-based joint clustering and representation learning with feature drift compensation for large-scale image data," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 421–429, 2017.

[48] M. Poongodi, A. Sharma, M. Hamdi, M. Maode, and N. Chilamkurti, "Smart healthcare in smart cities: wireless patient monitoring system using IoT," *The Journal of Supercomputing*, pp. 1–26, 2021.

[49] X. Xu, L. Li, and A. Sharma, "Controlling messy errors in virtual reconstruction of random sports image capture points for complex systems," *International Journal of System Assurance Engineering and Management*, pp. 1–8, 2021.

[50] G. K. Sodhi, S. Kaur, G. S. Gaba, L. Kansal, A. Sharma, and G. Dhiman, "COVID-19: role of robotics, artificial intelligence, and machine learning during pandemic," *Current Medical Imaging*, 2021.

[51] Y. Liu, Q. Sun, A. Sharma, A. Sharma, and G. Dhiman, "Line monitoring and identification based on roadmap towards edge computing," *Wireless Personal Communications*, pp. 1–24, 2021.

[52] M. Fan and A. Sharma, "Design and implementation of construction cost prediction model based on SVM and LSSVM in industries 4.0," *International Journal of Intelligent Computing and Cybernetics*, 2021.

[53] H. Sun, M. Fan, and A. Sharma, "Design and implementation of construction prediction and management platform based on building information modelling and three-dimensional simulation technology in industry 4.0," *IET Collaborative Intelligent Manufacturing*, 2021.

[54] X. Ren, C. Li, X. Ma et al., "Design of multi-information fusion based intelligent electrical fire detection system for green buildings," *Sustainability*, vol. 13, no. 6, p. 3405, 2021.

[55] M. Kaur, D. Singh, and V. Kumar, "Color image encryption using minimax differential evolution-based 7D hyper-chaotic map," *Applied Physics B*, vol. 126, no. 9, pp. 1–19, 2020.

[56] M. Kaur and D. Singh, "Multiobjective evolutionary optimization techniques based hyperchaotic map and their applications in image encryption," *Multidimensional Systems and Signal Processing*, pp. 1–21, 2020.

[57] Z. Ali and T. Mahmood, "Complex neutrosophic generalised dice similarity measures and their application to decision making," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 2, pp. 78–87, 2020.

[58] T. Sangeetha and G. M. Amalanathan, "Outlier detection in neutrosophic sets by using rough entropy based weighted density method," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 2, pp. 121–127, 2020.

[59] C. Zhu, W. Yan, X. Cai, S. Liu, T. H. Li, and G. Li, "Neural saliency algorithm guide bi-directional visual perception style transfer," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 1, pp. 1–8, 2020.

[60] M. Safa, M. Ahmadi, J. Mehrmashadi et al., "Selection of the most influential parameters on vectorial crystal growth of highly oriented vertically aligned carbon nanotubes by adaptive neuro-fuzzy technique," *International Journal of Hydromechatronics*, vol. 3, no. 3, pp. 238–251, 2020.