


Article

# HAZOP Ontology Semantic Similarity Algorithm Based on ACO-GRNN

Yujie Bai, Dong Gao \* and Lanfei Peng

College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China; 2019210438@buct.edu.cn (Y.B.); 2019210491@buct.edu.cn (L.P.)

\* Correspondence: gaodong@mail.buct.edu.cn

**Abstract:** Hazard and operability (HAZOP) is an important safety analysis method, which is widely used in the safety evaluation of petrochemical industry. The HAZOP analysis report contains a large amount of expert knowledge and experience. In order to realize the effective expression and reuse of knowledge, the knowledge ontology is constructed to store the risk propagation path and realize the standardization of knowledge expression. On this basis, a comprehensive algorithm of ontology semantic similarity based on the ant colony optimization generalized neural network (ACO-GRNN) model is proposed to improve the accuracy of semantic comparison. This method combines the concept name, semantic distance, and improved attribute coincidence calculation method, and ACO-GRNN is used to train the weights of each part, avoiding the influence of manual weighting. The results show that the Pearson coefficient of this method reaches 0.9819, which is 45.83% higher than the traditional method. It could solve the problems of semantic comparison and matching, and lays a good foundation for subsequent knowledge retrieval and reuse.

**Keywords:** HAZOP; safety; ontology; semantic similarity; ant colony algorithm; neural networks



**Citation:** Bai, Y.; Gao, D.; Peng, L. HAZOP Ontology Semantic Similarity Algorithm Based on ACO-GRNN. *Processes* **2021**, *9*, 2115. <https://doi.org/10.3390/pr9122115>

Academic Editors: Yiyang Dai, Li Zhou and Xu Ji

Received: 11 October 2021  
Accepted: 21 November 2021  
Published: 24 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

HAZOP analysis studies the hazards and operability of the system by exploring the impact of the deviation which means that process factors off normal process design conditions (including equipment failure, human misoperation, etc.) [1]. When one analyzes the possible causes, consequences, safeguard procedures and safety tips of a deviation, then one obtains its propagation path, which constitutes the HAZOP analysis results. This method not only considers the inherent danger of the equipment, but also considers the dynamic danger when the state parameters deviate in the process flow, comprehensively predicts the danger of the process, and has been widely used in safety analysis. The complexity of industrial application scenarios requires efficient, low-cost, and reusable HAZOP analysis with computer assistance, reducing the reliance on experts and repetitive manual labor. Therefore, the objective of this work is to propose a semantic similarity algorithm applicable to the HAZOP knowledge, so as to provide similar cases as reference for HAZOP analysis. In this work, an HAZOP ontology was constructed, and neighborhood coincidence degree method and ACO-GRNN are proposed to calculate the synthetical semantic similarity.

The remainder of the paper is constructed as follows: Section 2 introduces the background and a critical comparison of our contribution with respect to the existing literature. Section 3 clarifies what are the current issues in HAZOP analysis and the value added by an ontology. Section 4 briefly summarizes the phases of the approach. Section 5 describes the construction method of HAZOP ontology. Section 6 presents the semantic similarity algorithm based on the HAZOP ontology proposed in this paper. Section 7 describes the techniques used, data sets, and parameters used. Section 8 analyzes the experimental results and describes the application scenarios of this method. Finally, some conclusions and future work are shown in Section 9.

## 2. Related Work

Ontology is an explicit formal specification which shares a conceptual model. Semantic similarity refers to the possibility that two words or sentences can be replaced without changing their semantic structure under the premise that their form is not exactly the same. Ontology-based semantic similarity refers to the semantic similarity of two concepts in ontology [2]. At present, there are four types of ontology-based semantic similarities: (1) Algorithm based on semantic distance: in an ontology tree, each concept is a node and the similarity of concepts is measured according to the distance between nodes [3–6]. The disadvantage of this method is that many other inheritance relationships are ignored when applied to large-scale ontologies. Other factors affecting semantic similarity are not considered in this method. (2) In ontology, the more information shared by concepts, the higher their similarity [7–11]. The disadvantage of this method is that it is greatly influenced by corpus. (3) Each concept node has its own property set, which reflects the characteristics of the concept. The higher the degree of attribute coincidence of concepts, the more similar they are [12–14]. The advantage of this approach is that it can solve the problem of semantic similarity across ontology. However, the disadvantage is that it is more suitable for processing large ontology with rich semantic knowledge, and not suitable for small ontology. (4) Hybrid method: the above three algorithms are considered simultaneously, i.e., semantic distance, information content, property characteristics and other factors are selected for comprehensive calculation [15–22]. The hybrid method considers more factors than the single method, but it mostly relies on expert experience and adopts the method of manual weight assignment to formulate the weight factors of each element. The literature [20,21] introduced a method based on improved back propagation (BP) neural network, in which the method of Xu Feixiang et al. [20] improved on the method of Han Xueren et al. [21]. Although subjective influence of manual weighting assignment was avoided, the calculation accuracy still needs to be improved. The method proposed by Sun [22] models the ontology as a heterogeneous graph network, but it has no great advantage in dealing with small ontologies. This method is also a hybrid method. Compared with other methods, the advantages of this method are high accuracy, low computational complexity, and suitability for small ontologies with rich semantic information.

## 3. Motivating Example

The traditional HAZOP analysis method is experts brainstorming, which requires a lot of time and labor cost. In order to solve the above problems, many scholars began to use computers to improve the efficiency of security analysis and proposed computer-assisted or computer-automatic HAZOP analysis methods [23–25], which reduced human workload to a certain extent. However, HAZOP data storage formats of different teams are also different, which makes data sharing and reuse still difficult. For example, in a HAZOP analysis project for oil synthesis systems, a total of 30 experts participated and spent more than a month meeting and analyzing. The resulting analysis was nearly 500 pages in tabular form. This data is hard to reuse by other teams, thus it is a huge waste of resources. In order to standardize knowledge and facilitate sharing and reuse, some studies combine knowledge ontology with HAZOP analysis. Knowledge ontology can help people to systematically and formally describe domain knowledge. Ontology is easy for computers to identify and process, thus it is easy for different teams to reuse and share. Kuraoka [26] first introduced the ontology into HAZOP analysis and used ontology to describe the structural description of common HAZOP factors. Wu Chongguang [27] proposed the scenario object model (SOM) for the propagation path of dangerous events based on knowledge ontology. On this basis, Zhang Beike [28] developed a graphic tool and applied it to actual HAZOP analysis to realize SOM graphic expression. Later, Gao Dong [29] proposed a standardization framework based on knowledge ontology, and realized the automatic standardization of HAZOP analysis results by BiLSTM neural network. However, HAZOP knowledge has not been fully exploited. Semantic similarity based on HAZOP ontology can be used in HAZOP semantic retrieval to promote the intelligentization of HAZOP.

#### 4. Approach Overview

Two problems have been identified above: (1) Existing studies have solved the problem about HAZOP analysis data sharing, but these data are not fully utilized. (2) HAZOP ontology belongs to the domain ontology and has the characteristics of strong professionalism and small sample size. However, most similarity algorithms are aimed at the general ontology with large amount of data, which is difficult to apply to HAZOP ontology.

To solve these problems, this paper proposes a comprehensive calculation method of ontology semantic similarity based on ACO-GRNN. In this method, the concept meaning, ontology semantic distance, and neighbor node coincidence degree are considered comprehensively. Then, ACO-GRNN is used to train the weight of each element, so that the final comprehensive calculation results are close to expert's scores. The innovation of this paper is to propose the calculation method of neighbor node coincidence degree, and choose GRNN to calculate the comprehensive semantic similarity and improve it with ACO. The advantage of this method is that the neighbor node coincidence degree algorithm can fully mine the object relationships in HAZOP ontology. In addition, since the computation of GRNN is small, HAZOP ontology is also smaller than general ontology, thus GRNN has a good effect on HAZOP data processing. Experimental results show that the calculation accuracy of semantic similarity under this method is improved obviously.

#### 5. Construction of HAZOP Ontology

Ontology is an effective way to realize knowledge reuse because of its powerful semantic function and theoretical support of mathematical logic. At present, the Chinese knowledge ontology used for semantic similarity calculation mainly includes HowNet and Word Net. However, the generic ontology contains fewer professional words and cannot be used for semantic similarity calculation in chemical safety field. Therefore, it is necessary to use expert analysis materials to construct chemical safety domain ontology.

##### 5.1. Ontology Constructing Method

The thematic area of this ontology is HAZOP. Its goal is to support knowledge exchange and promote HAZOP analysis to be intelligent. Hence, it focuses on establishing strict logical relationships between individuals. HAZOP ontology mainly stores HAZOP concepts, relationships and HAZOP individuals. In this paper, ontology is described by ontology web language (OWL). For the process of constructing ontology, the creation of concepts and relationships is done manually, and the creation of individuals is done by automated procedures. The pseudo code of the automatic program is shown below.

```
Load HAZOP table
Load ontology model
WHILE The next row of the table is not empty
DO {
    Read the next row from the table
    FOR each cell of the row as Object {
        Get a property with Object as the domain
        Get a cell as the Subject, which is the range of the property
        Building a triple, [Object, property, Subject]
        Add the triple to the ontology model
    }
}
Save the ontology
END
```

A part of the HAZOP ontology described by OWL is shown in Figure 1. The basic elements of OWL allow us to build a simple ontology. For example, owl: Class means to

define a class, and `rdfs: subClassOf` indicates the relationship between a subclass and its parent class, and `rdf:type` means to define an individual of a class.

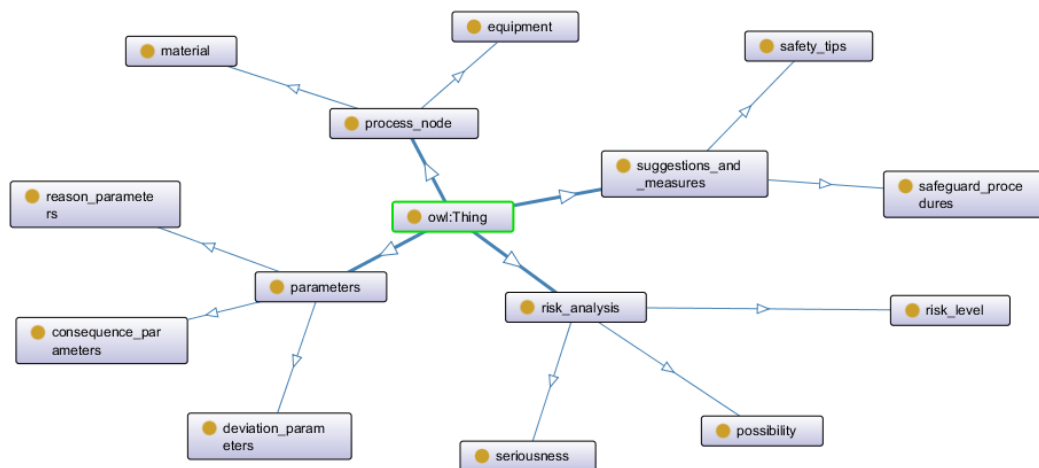
```
<owl:Class rdf:about="http://www.semanticweb/result#deviation_parameters">
  <rdfs:subClassOf rdf:resource="http://www.semanticweb/result#parameters"/>
</owl:Class>

<owl:NamedIndividual rdf:about="http://www.semanticweb/result#R5611101">
  <rdf:type rdf:resource="http://www.semanticweb/result#equipment"/>
  <conn-because rdf:resource="http://www.semanticweb/result#no_flow"/>
  <pathid rdf:datatype="http://www.semanticweb/result#integer">9</pathid>
</owl:NamedIndividual>
```

**Figure 1.** A part of HAZOP ontology described by OWL.

For relationships between concepts, it is preferred to reuse relationships defined in ontology, such as `rdf: type`. For relationships that are not included in the ontology, use the properties of ontology to customize them, such as `conn-because`. In ontology, properties include object properties and data properties, which can not only describe the common characteristics of the classes, but also explain the specific characteristics of the individuals. The domain can be regarded as the starting point of the property relationship, and the range is the end point of the property relationship. Properties connect the domain to the range. Object properties establish relationships between classes. Data properties establish relationships between classes and data. It is a quantitative description of the characteristics of a class or a name-specific reference.

In this paper, the HAZOP ontology is constructed by the top-down method. Domain concepts are listed first, then properties and relationships are defined, and finally the individuals are filled in. For the concepts of domain, HAZOP standardized information model is listed in this paper with reference to literature [29]. It divides the concepts of HAZOP ontology into three layers, and the structure diagram is shown in Figure 2, where, ● represents the class node of ontology.



**Figure 2.** HAZOP ontology hierarchy chart.

The first layer is the root node of the ontology. Secondly, parameters contain three subclasses: reason parameters, deviation parameters, and consequences parameters. The process node contains subclass material and equipment. The suggestions and measures have children safety tips and safeguard procedures. The risk analysis contains possibility, seriousness, and risk level.

### 5.2. Relationship Construction of HAZOP Ontology

In the process of building HAZOP ontology, 13 properties were customized, and the domains and ranges of each property were defined. See Table 1.

**Table 1.** Properties of HAZOP ontology.

Type	Name	Domains	Ranges
ObjectProperty	conn-act	process node	safeguard procedures
ObjectProperty	conn-because	process node	reason parameters
ObjectProperty	conn-dev	process node	deviation parameters
ObjectProperty	hasaction	reason parameters	process node
ObjectProperty	hascons	deviation parameters	process node/consequence parameters
ObjectProperty	hasreason	deviation parameters	process node/reason parameters
ObjectProperty	hasrisk-I	reason parameters	risk level
ObjectProperty	hasrisk-P	reason parameters	possibility
ObjectProperty	hasrisk-S	reason parameters	seriousness
ObjectProperty	hassug	reason parameters	process node
ObjectProperty	conn-lead	process node	consequence parameters
ObjectProperty	conn-suggest	process node	safety tips
DataProperty	pathid	-	-

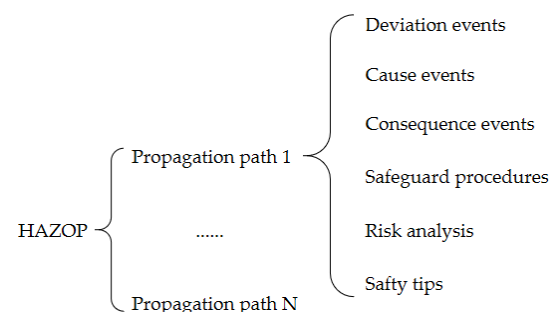
In the table, pathid is a data property, indicating the number of the deviation propagation path. All elements in the same deviation propagation path have the same pathid. One node can have multiple pathids at the same time.

Object property describes two types of relationships: relationships within events and relationships between events.

The relationships within events include conn-act, conn-dev, conn-because, conn-lead, and conn-suggest. The two nodes connected by such a relationship constitute an event. The relationships between events include hasaction, hascons, hasreason, hasrisk-I, hasrisk-P, hasrisk-S, and hassug. The two nodes connected by this kind of relationship represent two events on the propagation path. A complete propagation path is given in Section 5.3, showing examples of these relationships.

### 5.3. Propagation Paths in HAZOP Ontology

HAZOP information is composed of a large number of deviation propagation paths. The propagation path includes the deviation involved, cause events, consequence events, safeguard procedures, risk analysis, safety tips, etc., which constitute all elements of HAZOP analysis. See Figure 3.



**Figure 3.** Propagation paths of HAZOP.

HAZOP ontology is constructed according to the deviation propagation path to reflect the relationship between concepts intuitively. The data volume of HAZOP ontology is smaller than that of general ontology, and there is strict causal logic between concepts. A part of an HAZOP report is shown in Figure 4a. A row of records is a propagation path. Sort out the important information in Figure 4a to get Figure 4b. Convert the sorted

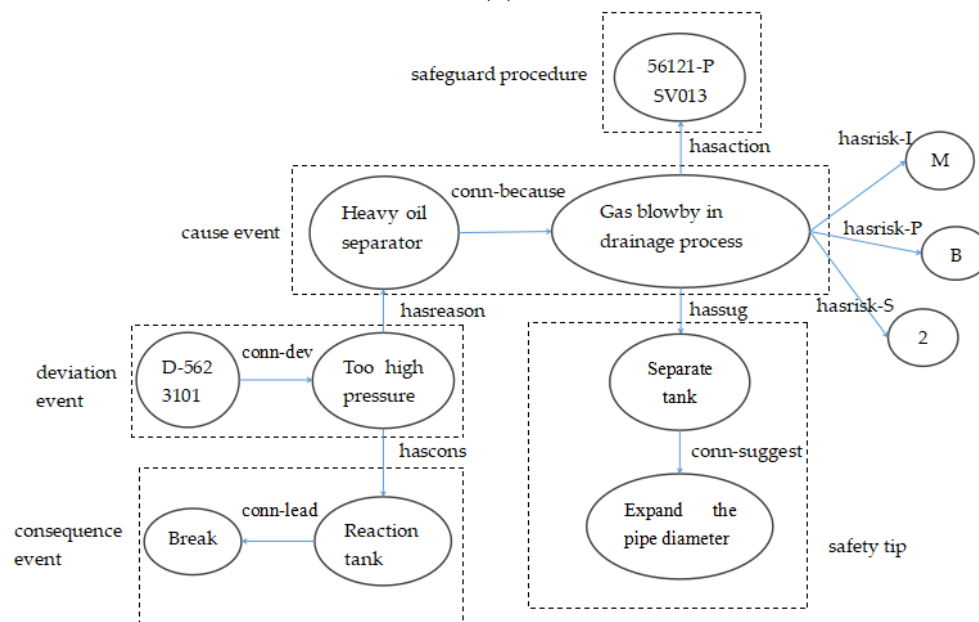
HAZOP record into a graph, as shown in Figure 4c. The dotted box represents the event as a whole.

XXX HAZOP Analysis Meeting Records										
Drawing number			Meeting date		20XX.XX.XX					
XXX-XXX-XXXX			Participants		XXX, XXX					
NO.	Process node	Deviation	Cause	Consequence	Risk analysis			Safeguard procedures	Safety tips	Remarks
					Seriousness	Possibility	Risk level			
3	D-5623101	Too high pressure	Reaction tank break	Gas blowby in drainage process of heavy oil separator	2	B	M	56121-PSV013	Expand the pipe diameter of separate tank	

(a)

NO.	Process node	Deviation parameter	Process node	Cause parameter	Process node	Consequence parameter	Risk analysis			Process node	Safeguard procedures	Process node	Safety tips	Remarks
							Seriousness	Possibility	Risk level					
3	D-5623101	Too high pressure	Reaction tank	Break	Heavy oil separator	Gas blowby in drainage process	2	B	M	56121-PSV013		Separate tank	Expand the pipe diameter	

(b)



(c)

**Figure 4.** An example of converting HAZOP records into a diagram. (a) A part of an HAZOP report. (b) The sorted HAZOP record. (c) A deviation propagation path.

It can be seen from Figure 4 that the nodes in the propagation path are strongly correlated. This is one of the characteristics of HAZOP ontology. In Figure 4, D-5623101 and too high pressure are connected by conn-dev, and they are labeled with a dotted box to indicate a deviation event. As shown in Figure 3, deviation events are part of the propagation path. As shown in Table 1, the domain of conn-dev is a process node, and the range is a deviation parameter. That is to say that a process node and a deviation parameter are connected by conn-dev to represent a deviation event.

## 6. Ontology Semantic Similarity Calculation Model Based on ACO-GRNN

Semantic similarity plays an important role in semantic retrieval, knowledge acquisition and other fields, and is a necessary step to realize intelligence of HAZOP analysis. The comprehensive semantic similarity based on ontology usually considers many factors, calculates their similarity and gives the weight, and finally obtains the weighted comprehensive semantic similarity. In the current research, the method of manually giving weights is difficult to avoid subjective influence, while the method of using neural network to train

weights avoids subjective influence, but its accuracy needs to be improved, and a large number of samples are needed to achieve better results. Therefore, GRNN with a small amount of computation is selected for weight calculation in this paper, and a calculation method of neighbor node coincidence degree is proposed. This paper is based on HA-ZOP ontology. Calculate the contextual information, semantic distance and coincidence degree of neighborhood nodes, and input them into ACO-GRNN. It then outputs semantic similarity close to the expert's scores.

### 6.1. Edit Distance Calculation Method

Edit distance is the minimum number of operands required to convert one string to another. The semantic similarity of concept pairs to be compared is inversely proportional to the editing distance. The larger the editing distance is, the smaller the semantic similarity is. The calculation formula is shown in Equation (1):

$$Sim_{Dis}(A, B) = 1 - \frac{ED_{A,B}}{\max(L_A, L_B)} \quad (1)$$

The semantic similarity calculated by editing distance can be used as one of the elements of comprehensive semantic similarity to explain the semantic association between textual information.

### 6.2. Ontology Node Distance Calculation Method

The similarity calculation method based on semantic distance of ontology nodes represents the difference between concepts by the length of connection path in ontology. Rada [5] proposed that the length of the shortest path between two nodes is used to represent the conceptual distance between nodes. The more similar two concepts are, the smaller the conceptual distance between them will be. In this method, the weight of all edges in the ontology is assumed to be 1. The calculation formula is shown in Equation (2):

$$Sim_{Node}(A, B) = \frac{1}{dis(A, B)} \quad (2)$$

where  $dis(A, B)$  is the sum of the distances between nodes  $A$  and  $B$  and the nearest public parent node. The computational complexity of this algorithm is relatively small, and does not involve unstable parameters, thus this algorithm is selected in this paper to calculate the semantic distance similarity. Using semantic distance similarity of ontology as one of the elements of comprehensive semantic similarity can explain the semantic association of concepts at the level of ontology structure.

### 6.3. Calculation Method of Neighbor Node Coincidence Degree

Zhang Zhongping [19] proposed an algorithm to comprehensively calculate the three elements of attribute name, attribute data type, and attribute value. The calculation method is as follows:

Assume that the concepts to be compared are  $A$  and  $B$ , and let the attribute of  $A$  be  $a_i$  and the attribute of  $B$  be  $b_j$ . The similarity between attribute  $a_i$  and  $b_j$  is  $ASim(a_i, b_j)$ , and its calculation formula is shown in Equation (3):

$$\begin{aligned} ASim(a_i, b_j) &= \omega_1 \times Sim(a_{i_{name}}, b_{j_{name}}) \\ &+ \omega_2 \times Sim(a_{i_{datatype}}, b_{j_{datatype}}) \\ &+ \omega_3 \times Sim(a_{i_{value}}, b_{j_{value}}) \end{aligned} \quad (3)$$

where  $Sim$  is the similarity of string,  $\omega_1$ ,  $\omega_2$ , and  $\omega_3$  are the weight of attribute name, attribute data type and attribute value respectively.  $\omega_1 + \omega_2 + \omega_3 = 1$ . The calculation formula of attribute similarity of  $A$  and  $B$  is shown in Equation (4):

$$Sim_{attr}(A, B) = \frac{\sum_{k=1}^m \omega_k \times ASim(a_i, b_j)}{\sum_{k=1}^m \omega_k} \quad (4)$$

where  $\omega_k$  is the weight of each  $ASim$ .

It is known that attributes in ontology can be divided into object attributes and data attributes. This algorithm uses data attributes to measure semantic similarity. However, there are plenty of neighbor node relationships in the ontology, and semantic information may be missing if object attributes are not considered. Besides, HAZOP ontology focuses on the relationships between nodes. Hence, it improves on the basis of this algorithm and uses object attributes to calculate the neighbor node coincidence degree. If two nodes are connected by an object attribute, they are neighbor nodes.

Assuming that there are concepts  $A$  and  $B$ , the rules for calculating their neighbor node coincidence degree are as follows:

Rule 1: For two nodes  $A$  and  $B$  of the ontology, the more same object attributes they have, the higher their similarity.

Rule 2: For the common object property  $o$  of nodes  $A$  and  $B$ , assume that the neighbor node of  $A$  connected by  $o$  is  $a$ , and the neighbor node connected with  $B$  is  $b$ . The more similar the concept names of  $a$  and  $b$  are, the higher the similarity between  $A$  and  $B$  will be.

Rule 3: Since concept names are usually represented by character text, use the method in Section 6.1 to determine their similarity.

Rule 4: Assume that the character similarity of  $a$  and  $b$  is  $Sim_{ab}$ , and consider  $Sim_{ab}$  as the contribution value of the common object property  $o$  to the similarity of  $A$  and  $B$ . When  $o$  connects multiple neighbor nodes, the largest  $Sim_{ab}$  is taken as the contribution value of  $o$  and it would be defined as  $KSim$ .

According to the above rules, the more common object attributes of concepts, the higher their similarity. When the common object attribute sets are the same, the similarity is judged by the similarity of concept names of neighbor nodes. For example, in Figure 5,  $node-A$ ,  $node-B$ , and  $node-C$  represent ontology nodes to be compared,  $attribute1$ ,  $attribute2$ , and  $attribute3$  represent object attributes, and  $node-a$ ,  $node-b$ , and  $node-c$  represent neighbor nodes. According to Figure 5, draw the venn diagram of object attributes of  $node-A$ ,  $node-B$ , and  $node-C$ , as shown in Figure 6. Since  $node-A$  and  $node-C$  have the fewest common object attributes, according to Rule 1,  $node-A$  and  $node-C$  have the lowest similarity. After the set of common object attributes is obtained, the venn diagram of neighbor nodes connected by common object attributes is drawn, as shown in Figure 7. According to rule 2–4, the final similarity relationship between  $node-A$ ,  $node-B$ , and  $node-C$  is:  $Sim_{AB} > Sim_{BC} > Sim_{AC}$ .

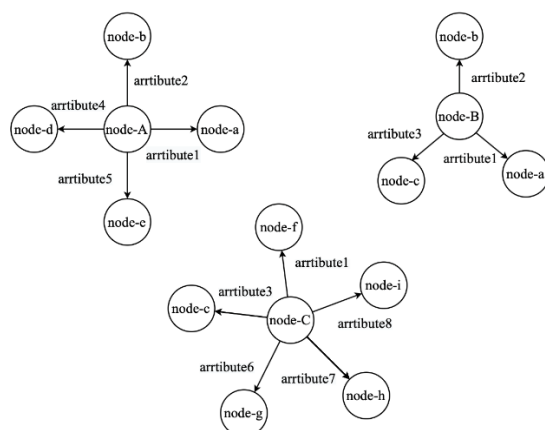


Figure 5. Nodes diagram.

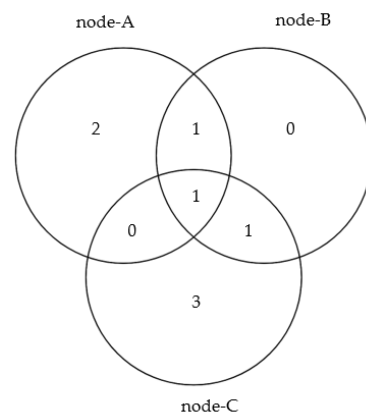


Figure 6. Venn diagram of node object properties.

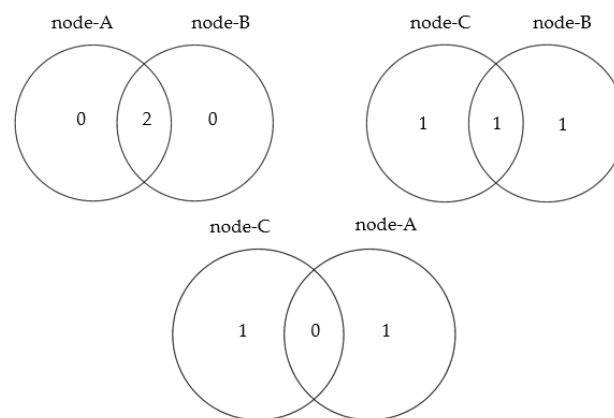


Figure 7. Venn diagram of common object properties.

After sorting out the above rules, the improved method to calculate node similarity based on neighbor node coincidence degree is as follows:

Assuming that there are concepts  $A$  and  $B$ , find all their object properties and corresponding neighbor nodes. Calculate the intersection of object properties of  $A$  and  $B$  to get  $E_{AB}$ . For an object property  $o$  in  $E_{AB}$ , assume that the neighbor node of its connection  $A$  is  $a_i$ , and the neighbor node of its connection  $B$  is  $b_j$ . Since concept names are usually strings, they are computed using the method of Section 6.1. The contribution of  $o$  to the similarity of  $A$  and  $B$  is shown in Equation (5):

$$KSim = \max Sim_{Dis}(a_i, b_j) \quad (5)$$

Since a node may have multiple neighboring nodes, the maximum value is taken as the contribution of  $o$  to simplify the calculation. Finally, the similarity calculation formula of concept  $A$  and  $B$  is shown in Equation (6):

$$Sim_{Pro}(A, B) = \frac{\sum_{l=1}^k N_l \times KSim_l}{N(A \cup B)} \quad (6)$$

$K$  is the number of common object attributes,  $N_l$  is the total number of occurrences of the  $l$ th common object properties, and  $N(A \cup B)$  is the total number of occurrences of all object properties on nodes  $A$  and  $B$  (the total number of neighbor nodes of  $A$  and  $B$ ).

$Sim_{Pro}$  calculated according to the method of neighbor node coincidence degree conforms to the following characteristics:

1.  $Sim_{Pro} \in [0,1]$ . In extreme cases, when  $k \times l = N(A \cup B)$  and each  $Ksim$  is 1, i.e., when concept  $A$  and concept  $B$  overlap completely,  $Sim_{Pro}$  is 1. When  $k = 0$ , i.e.,  $A$  and  $B$  have no common properties,  $Sim_{Pro}$  is 0.
2. Every node except the root node will have at least one parent node, so there is no need to consider the denominator of 0.
3. The value of  $Sim_{Pro}$  is positively correlated with the number of common object properties.

#### 6.4. Comprehensive Semantic Similarity

After obtaining the above editing distance similarity  $Sim_{Dis}$ , node semantic distance similarity  $Sim_{Node}$  and neighbor node similarity  $Sim_{Pro}$ , the three are weighted and the final comprehensive semantic similarity is  $Sim_{AB}$ , whose calculation formula is shown in Equation (7):

$$Sim_{AB} = \gamma Sim_{Dis} + \beta Sim_{Node} + \alpha Sim_{Pro} \quad (7)$$

where,  $\gamma, \beta, \alpha$  are the weights, representing the influence degree of each factor on the final calculation results, and  $\gamma + \beta + \alpha = 1$ . In order to avoid the subjective influence of artificial weights on the results of semantic similarity calculation, ACO-GRNN was used to train the weights, and finally the comprehensive semantic similarity was obtained.

#### 6.5. Comprehensive Semantic Similarity Calculation Based on ACO-GRNN

Because GRNN has the characteristics of rapid convergence and small computation, it is suitable for processing small sample data. Thus, GRNN is used for weight training of the above elements. The selection of the superparameter  $\delta$  in the pattern layer of GRNN usually affects the model performance, which can be set manually during the initialization of the model or determined by the optimization algorithm. The optimization of  $\delta$  is a discrete problem. Ant colony algorithm can converge quickly and is suitable for solving discrete problems. Therefore, the ant colony algorithm is used to improve GRNN to obtain the optimal value of  $\delta$ , so as to achieve the best performance of the model.  $\delta$  is a multi-dimensional vector, and every available value in each dimension is regarded as a city when the ant colony algorithm is used for optimization. The optimization process is carried out iteratively, and the best value for the model is selected from the solution set. The error function is shown in Equation (8).

$$e = \frac{\sum_{i=1}^n (y_i - Y_i)^2}{n} \quad (8)$$

where  $y_i$  is the predicted output value,  $Y_i$  is the expected output value, and  $n$  is the total number of predicted output values. The smaller  $e$  is, the closer the predicted output value is to the expected output value, and the better the model effect is. The ant colony algorithm was used to optimize  $\delta$ , and the minimum error value  $e$  corresponding to  $\delta$  was obtained after the iteration. The computational flow of comprehensive semantic similarity based on the GRNN network improved by ant colony algorithm is as follows, and the algorithm flow chart is shown in Figure 8.

Input: HAZOP ontology, training concept set  $E\{e_1, e_2, \dots, e_n\}$ , expert score set  $M$ , the test concepts  $A$  and  $B$ .

Step 1: Take any pair of concepts  $[e_i, e_j]$  from the training concept set  $E$ , or take concepts  $A$  and  $B$ , calculate their text similarity using Formula (1), and store the result  $Sim_{Dis}(e_i, e_j)$  into the matrix  $R$ .

Step 2: Obtain the ontology parent node sets  $L_i$  and  $L_j$  of  $e_i$  and  $e_j$  respectively. By comparing  $L_i$  and  $L_j$ , the nearest common parent node of  $e_i$  and  $e_j$  is obtained. If at least one of  $e_i$  and  $e_j$  has no parent node set, i.e., at least one does not exist in the ontology, they are regarded as irrelevant nodes by default, and the node distance is the farthest node distance in the ontology. The ontology semantic distance similarity of  $e_i$  and  $e_j$  is calculated by using Formula (2), and the result  $Sim_{Node}(e_i, e_j)$  is stored into the result matrix  $R$ .

Step 3: Find all properties of node  $e_i$  and corresponding properties and form them into the set  $S_1$ . All properties and corresponding attribute values of node  $e_j$  are searched to form  $S_2$ . Equations (5) to (6) are used to calculate attribute similarity, and the result  $Sim_{Pro}(e_i, e_j)$  is stored into the result matrix  $R$ .

Step 4: Repeat steps 1 to 3 until all concept pairs in concept set  $E$  and concepts  $A$  and  $B$  have been calculated.

Step 5: Use ACO to improve GRNN, and then train it. Taking the result matrix  $R$  and expert score set  $M$  as the input of ACO-GRNN.

Step 6: ACO-GRNN calculated the comprehensive semantic similarity  $Sim_{AB}$  of  $A$  and  $B$ .

Output:  $Sim_{AB}$  which is comprehensive semantic similarity of  $A$  and  $B$ .

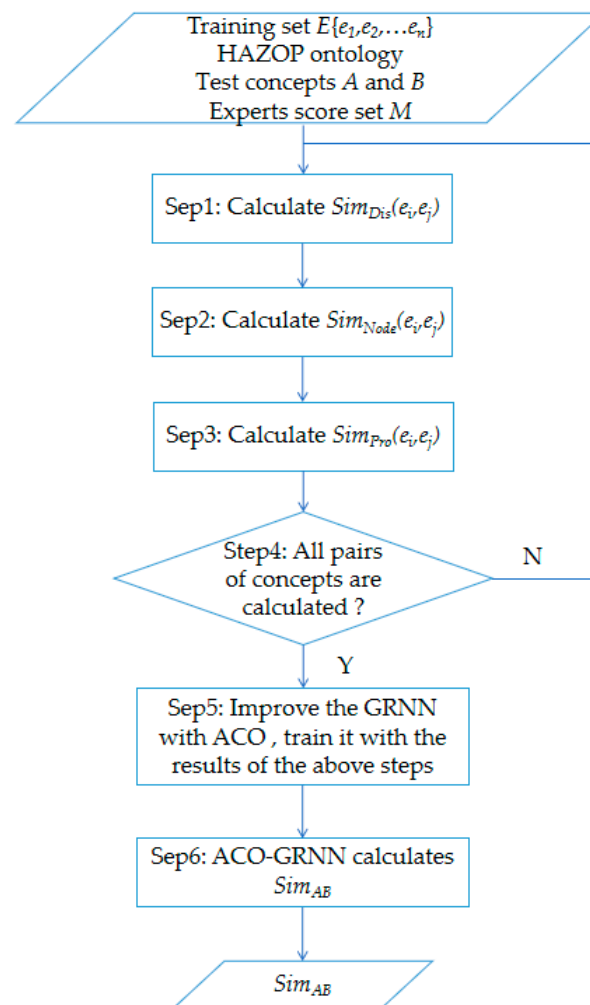


Figure 8. Algorithm flow chart.

## 7. Preparation of the Experiment

### 7.1. Tools and Techniques

The tools and techniques used in this experiment are as follows. Operating system: Windows 10. Programming languages: Java 1.8, Python 3.7 and OWL. Software: Protégé 5.5.0, Pycharm 2017.1, Eclipse and MySQL.

The architecture and the software module employed are as follows.

- Sort out the original HAZOP records and extract important information with PyCharm software and Python language, then store it in Excel.

- Use Protégé and OWL to construct ontology. Use Eclipse and Java programs for ontology automatic population.
- Use Eclipse and Java programs to parse ontologies into triples and store them persistently in MySQL.
- Use Eclipse and Java programs to calculate edit distance, node semantic distance and coincidence degree of neighbor nodes, then use a Python program to calculate the comprehensive semantic similarity.

### 7.2. Dataset

The experimental dataset is from HAZOP analysis report of an oil synthesis device. The key information of the report is extracted by the Chinese named entity recognition method proposed by Li [30], and then added to the HAZOP ontology as individuals. In order to test the effect of the model, 500 pairs of concepts were selected, among which the first 400 pairs were used as the training set and the last 100 pairs as the test set, and were input into the semantic similarity calculation model proposed in this paper.

The expert's scores required for model calculation were scored manually by 7 experts. Experts rely on their own expertise to give scores between [0,1]. 0 indicates that concepts are not similar at all, and the higher the score is, the more similar the concepts are. When an expert gives a score, both the definition of the concept in the dictionary and the role of the concept in HAZOP analysis (e.g., it is a device or an action) should be considered. To reduce decimal oscillations, experts are expected to keep the score to two decimal places. The scores are subjective. To make the data as reasonable as possible, we averaged their scores.

In order to verify the model, some synthetic algorithms are selected as comparative experiments. The comparative experimental methods are selected from the linear weighted similarity algorithm of Li Wenqing [18], the ACWA synthesis algorithm of Zhang Zhongping [19], the SA-BP synthesis algorithm of Xu Feixiang [20], and the synthesis algorithm based on graph convolution network (GCN) [22].

### 7.3. Evaluation Index

Pearson correlation coefficient can measure the correlation of two variables and is used to evaluate the consistency of model results and expert scoring results under the same test set. Given two variables  $X$  and  $Y$ , if there are no coincident elements in the set of values  $X$  and  $Y$ , then  $\rho$  can be used to represent the correlation when they move in the same direction. The value range of  $\rho$  is  $[-1,1]$ . When  $\rho$  is positive,  $X$  and  $Y$  are positively correlated; when 0, they are not correlated; when negative, they are negatively correlated. The stronger the correlation is, the greater the  $\rho$  value is.

The calculation formula of Pearson coefficient is shown in Equation (9):

$$\rho = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (9)$$

### 7.4. Experimental Parameters

In ant colony algorithm, there are three important adjustable parameters. *Alpha* represents the importance of pheromones, *Beta* represents the importance of heuristic factors, and *Rho* is the evaporation coefficient of pheromones. According to previous experience, when the value of *Alpha* in [1,3], the value of *Beta* is no more than 5, and the value of *Rho* in [0.2,0.5], the comprehensive performance of the model is good. Adjust the parameters and observe the error value. The smaller the error value is, the better the model effect is. The final parameter values are: *Alpha* = 2.5, *Beta* = 2.0, *Rho* = 0.3. Set the number of iterations to 20. At this point, the model error change line graph is shown in Figure 9. It can be seen from the figure that the final error is stable at 0.028.



Figure 9. Variation of error line.

## 8. Results

### 8.1. Experiment and Result Analysis

The results of this method and the comparison methods are shown in Table 2. The bar chart is shown in Figure 10.

Table 2. The comparison of Pearson coefficients of different methods.

Algorithm Name	Pearson
Linear weighting algorithm	0.6773
SA-BP algorithm	0.8473
ACWA algorithm	0.9455
GCN algorithm	0.9638
ACO-GRNN algorithm	0.9819

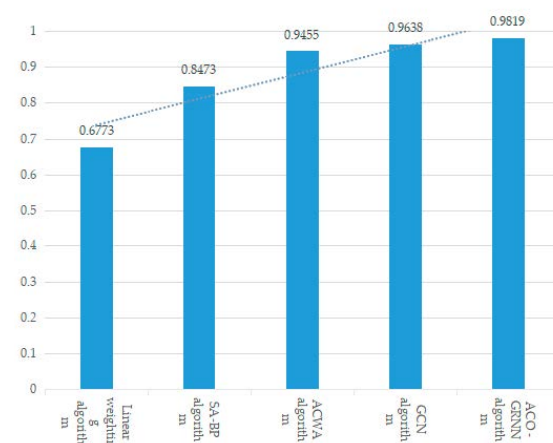


Figure 10. Results histogram comparison.

The comparison of calculation results shows that the Pearson coefficient of the proposed algorithm is improved by 45.83% compared with the linear weighting algorithm, 15.89% compared with the SA-BP algorithm, 3.85% compared with the ACWA algorithm, and 1.88% compared with the GCN algorithm.

Ten pairs of test concepts were randomly selected as the results to be displayed. The comparison between the results and experts scores is shown in Table 3 (all calculation results are taken to the fourth decimal point). The line comparison between the algorithm results and experts scores is shown in Figure 11.

**Table 3.** Experimental results of conceptual semantic similarity calculation.

Concept A	Concept B	SA-BP Algorithm	ACWA Algorithm	Linear Weighting Algorithm	GCN Algorithm	ACO-GRNN Algorithm	Expert Score
Inadequate reaction	Incomplete reaction	0.7474	0.6473	0.5927	0.9201	0.8436	0.9643
D-5611115	Decarbonization gas enters D-5611114	0.5178	0.6646	0.5703	0.8815	0.8438	0.7413
Too high liquid level	Too low liquid level	0.5643	0.3540	0.5927	0.7734	0.6868	0.7013
Equipment Trouble	Frozen block, damage, leak	0.7538	0.7254	0.5975	0.8961	0.9315	0.9130
LIC-0511	FIC-0511	0.7862	0.5258	0.5703	0.5914	0.7002	0.6741
Overpressure	Frozen block, damage, leak	0.3604	0.1464	0.5927	0.2823	0.3103	0.3100
Clean gas	Compressor C-5611101	0.4402	0.0989	0.3267	0.2507	0.1916	0.1794
Clean gas	Light oily water separator	0.4361	0.1048	0.3267	0.1173	0.2264	0.1989
Self protection interlock	L interlock	0.6545	0.4285	0.5975	0.6949	0.6961	0.7340
Carrier fluid	Cause hazard in severe case	0.3672	0.1472	0.5927	0.1387	0.3051	0.2999

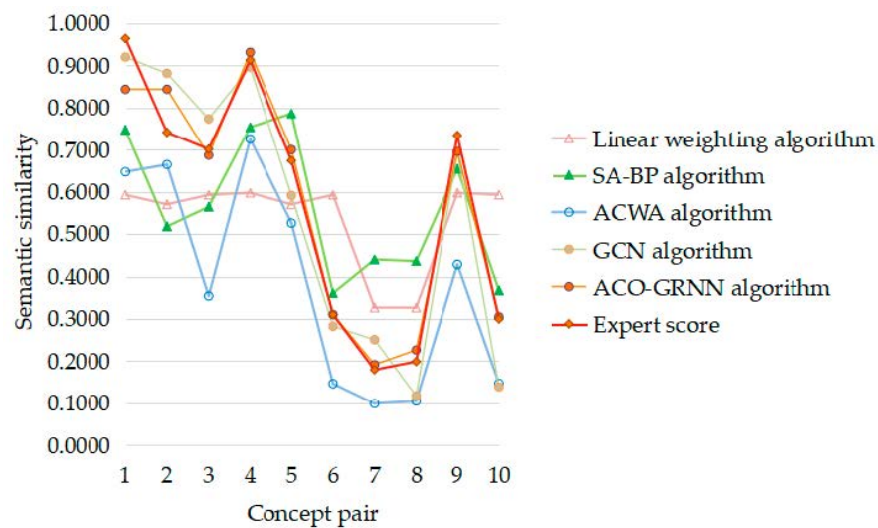
**Figure 11.** Algorithm result line graph.

Table 3 and Figure 11 show that:

1. The model has a high score for concepts that are obviously similar, such as inadequate reaction and incomplete reaction. For concepts with slightly lower character similarity and ontology level similarity, such as clean gas and compressor C-5611101, the calculated score is lower. There is no extreme value in the calculation results, and the whole is relatively flat.
2. In the 3 line of Table 3, the two concepts of too high liquid level and too low liquid level are semantically opposites. However, from the auxiliary analysis point of view, they both reveal the abnormal liquid level and are obviously highly correlated, thus the result can provide reference value.
3. It can be seen from Figure 11 that the algorithm in this paper is closer to experts score, with consistent trends and small data differences.

4. Among the comparison methods, the linear weighting method has the biggest difference with the expert rating, and the result is the most inaccurate.

The linear weighting method adopts the way of manual weighting and depends on experience, so the weight is often lack of scientific nature and the accuracy is low. The principal component analysis method used in ACWA aims at dimensionality reduction. Only when the data conform to Gaussian distribution, the variables obtained by dimensionality reduction are optimal. Therefore, the results of similarity calculation may be affected. SA-BP algorithm only considers the number of child nodes when calculating node similarity, which lacks in-depth comparison of node content. It is not optimal due to incomplete information. The principle of GCN algorithm is to map nodes to graph structures and carry out convolution calculation for neighbor nodes. According to the experimental results, there is no great advantage in using this model to deal with domain ontology with small data volume.

In this paper, an improved GRNN for computing small data sets is selected, and the approximation effect is outstanding, which is one of the reasons for the better results of the algorithm. In addition, in HAZOP ontology relationships, neighbor node relationships account for a large proportion, thus the analysis of neighbor node relationships plays an important role in judging concept similarity. In this paper, the algorithm improves the calculation of neighbor node similarity and considers more comprehensive factors, which contributes a lot to the comprehensive semantic similarity calculation. The characteristics of these methods are summarized in Table 4.

**Table 4.** Summary of different methods.

Algorithm Name	Advantages	Disadvantages	Whether to Use Neural Networks
Linear weighting algorithm	Low calculation	The way setting weights is not scientific	No
ACWA algorithm	Low calculation	High requirements for data	No
SA-BP algorithm	Stability	Lack of in-depth comparison of node content	Yes
GCN algorithm	Great for working with graph data	Slowly convergence, depend on the ontology structure	Yes
ACO-GRNN algorithm	Suitable for ontology with complex relationships, low computation	Depend on the ontology structure	Yes

Experimental results show that the proposed method is more accurate than other semantic similarity algorithms, and the computation results are closer to experts' scores. Therefore, the comprehensive semantic similarity calculation method proposed in this paper can solve the problems of low accuracy and complex calculation of the existing semantic similarity algorithm, and provide help for semantic retrieval and promote the intelligence of HAZOP analysis.

### 8.2. Application Scenario

The application value of this method is to assist HAZOP analysis and promote the intellectualization of HAZOP analysis. This method can help to increase the accuracy and comprehensiveness of analysis, save time, and improve the quality of analysis. In the future, based on this method, a software will be designed. There are many utilisation scenarios possible. For example, when HAZOP analysis is implemented in the movement stage of project, the software can be used to retrieve the deviation and obtain the same or similar deviations and propagation paths. They can be added to the HAZOP analysis results and high-risk cases will be alerted. The prototype GUI is shown in Figure 12.

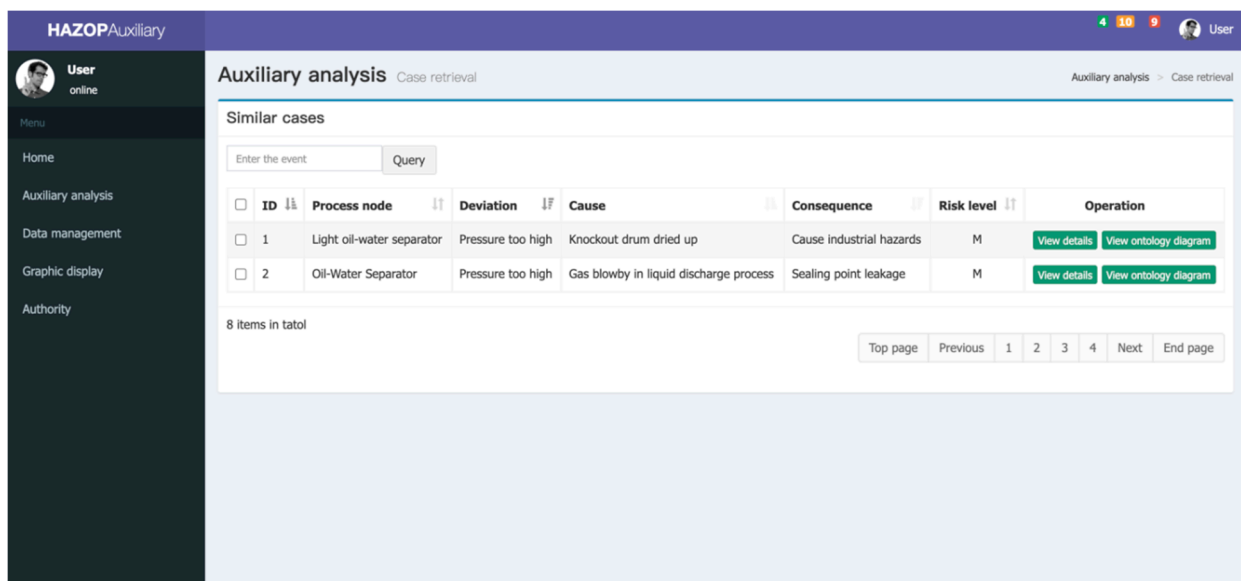


Figure 12. Prototype GUI.

In order to evaluate the concrete benefit in conducting the HAZOP analysis with the ontology, the following three indicators are proposed:

- Accuracy. HAZOP analysis needs to identify all potential risks. However, there may be a lot of invalid information in the analysis, and not all risks are serious. This software can give tips for major hazards and help analysts focus on them.
- Comprehensiveness. For example, when analyzing a deviation, this software can search out similar cases to supplement the analysis results, so as to improve the comprehensiveness of the analysis.
- Time. HAZOP analysis process has a lot of repetitive work, and the software can help save time.

## 9. Conclusions

HAZOP is widely used in chemical industry, which is an important mean to eliminate potential safety hazards and reduce risk levels. In order to make full use of the expert knowledge in HAZOP analysis report and promote the intelligentification of HAZOP analysis, an ontology semantic similarity comprehensive calculation method based on ACO-GRNN is proposed in this paper. First, the domain ontology was constructed according to HAZOP analysis report, and the GRNN was improved by ant colony algorithm. Then, the weight training was carried out on the three elements of string semantic similarity, node semantic distance similarity, and neighbor node similarity, and finally the comprehensive semantic similarity close to experts' score was obtained. When calculating comprehensive semantic similarity, the more semantic information the factors contain, the easier the result is to be close to the experts score. Based on the method of property coincidence degree, a calculation model of neighbor node coincidence degree is proposed, which has more semantic information and more accurate measurement results. The experimental results show that the comprehensive similarity method proposed in this paper has a great improvement in accuracy compared with the traditional semantic similarity algorithm, and the Pearson correlation coefficient reaches 0.9819, which is very similar to the expert scoring result. This algorithm solves the problem that the accuracy of traditional semantic similarity algorithm is not high. Applying the method proposed in this paper to the semantic retrieval process can promote the intelligence of HAZOP analysis, solve the problem of HAZOP knowledge sharing and reuse, and promote the construction of HAZOP talent team. The limitation of this method is that it is highly dependent on the richness of HAZOP ontology individuals. The future work will be to study ontology autonomous learning methods, so

that HAZOP ontology can be expanded autonomously, so as to solve the limitations of this paper. Then, the next step will be to implement a HAZOP auxiliary analysis software based on this algorithm.

**Author Contributions:** Conceptualization, D.G. and L.P.; project administration, D.G.; writing—original draft preparation, Y.B.; writing—review and editing, Y.B. and D.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation (NNSF) of China (61703026).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Special thanks to D.G. and L.P. for their efforts.

**Conflicts of Interest:** We declare that we have no financial or personal relationships with other people or organizations that could inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, “HAZOP Ontology Semantic Similarity Algorithm Based on ACO-GRNN”.

## References

- Dunjo, J.; Fthenakis, V.; Vilchez, J.A.; Arnaldos, J. Hazard and operability (HAZOP) analysis. A literature review. *J. Hazard. Mater.* **2010**, *173*, 19–32. [[CrossRef](#)] [[PubMed](#)]
- Sánchez, D.; Batet, M.; Isern, D.; Valls, A. Ontology-based semantic similarity: A new feature-based approach. *Exp. Syst. Appl.* **2012**, *39*, 7718–7728. [[CrossRef](#)]
- Hirst, D.G. Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms. *Lect. Notes Phys.* **1997**, *728*, 123–149.
- Leacock, C.; Chodorow, M. Combining local context and WordNet similarity for word sense identification. In *Massachu-Setts*; The MIT Press: Cambridge, MA, USA, 1998; pp. 265–268.
- Rada, R.; Mili, H. Development and application of a metric on semantic nets. *IEEE Trans. Syst. Man Cybern. Syst.* **1989**, *19*, 17–30. [[CrossRef](#)]
- Powers, D. Measuring Semantic Similarity in the Taxonomy of Word Net. *J. Struct. Biol.* **2007**, *159*, 36–45.
- Meng, L.L.; Gu, J.Z.; Zhou, Z.L. A new model of information content based on concept's topology for measuring semantic similarity in Word Net. *Int. J. Grid Distrib. Comput.* **2012**, *5*, 81–94.
- Seddiqui, M.H.; Aono, M. *Metric of Intrinsic Information Content for Measuring Semantic Similarity in an Ontology*, Proceedings of the 7th Asia-Pacific Conference on Conceptual Modelling, Brisbane, Australia, 1 January 2010; Australian Computer Society, Inc.: Darlinghurst, Australia, 2010; pp. 89–96.
- Sánchez, D.; Batet, M.; Isern, D. Ontology-based information content computation. *Knowl. Based Syst.* **2011**, *24*, 297–303. [[CrossRef](#)]
- Zhang, X.; Sun, S.; Zhang, K. An information Content-Based Approach for Measuring Concept Semantic Similarity in WordNet. *Wirel. Pers. Commun.* **2018**, *103*, 117–132. [[CrossRef](#)]
- Verschaffelt, P.; Bossche, T.; Gabriel, W.; Burdukiewicz, M.; Mesuere, B. MegaGO: A Fast Yet Powerful Approach to Assess Functional Gene Ontology Similarity across Meta-Omics Data Sets. *J. Proteome Res.* **2021**, *20*, 2083–2088. [[CrossRef](#)]
- Tversky, A. Features of Similarity. *Psychol. Rev.* **1977**, *84*, 222–226. [[CrossRef](#)]
- Petrakis, E.; Varelas, G.; Hliaoutakis, A.; Raftopoulou, P. X-Similarity: Computing Semantic Similarity between Concepts from Different Ontologies. *J. Inf. Manag.* **2006**, *4*, 233–237.
- Meng, X.; Yan, L.; Ma, Z.; Zhang, F.; Wang, X. An Adaptive Query Relaxation Approach for Relational Databases Based on Semantic Similarity. *Chin. J. Comput.* **2011**, *34*, 812–824. [[CrossRef](#)]
- Li, S.; Abel, M.-H.; Negre, E. Ontology-based Semantic Similarity in Generating Context-aware Collaborator Recommendations. In Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Dalian, China, 5–7 May 2021; pp. 751–756. [[CrossRef](#)]
- Wang, L.; Zhang, F.; Du, Z.; Chen, Y.; Liu, R. A Hybrid Semantic Similarity Measurement for Geospatial Entities. *Microprocess. Microsyst.* **2021**, *80*, 103526. [[CrossRef](#)]
- Li, Y.; Bandar, Z.A.; Mclean, D. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Trans. Knowl. Data Eng.* **2003**, *15*, 871–882.
- Li, W.; Sun, X.; Zhang, C.; Feng, Y. A Semantic Similarity Measure between Ontological Concepts. *Acta Autom. Sin.* **2012**, *38*, 229–235. [[CrossRef](#)]
- Zhang, Z.; Tian, S.; Liu, H. Compositive Approach for Ontology Similarity Computation. *Comput. Sci.* **2008**, *35*, 142–145+182.

20. Xu, F.; Ye, X.; Li, L.; Cao, J.; Wang, X. Comprehensive Calculation of Semantic Similarity of Ontology Concept Based on SA-BP. *Comput. Sci.* **2020**, *47*, 199–204.
21. Han, X.; Wang, Q.; Guo, Y.; Cui, X. Geographic Ontology Concept Semantic Similarity Measure Model Based on BP Neural Network Optimized by PSO. *Comput. Eng. Appl.* **2017**, *53*, 32–37.
22. Sun, L.; Wei, Y.; Wang, B. Similarity Calculation Method of Multisource Ontology Based on Graph Convolution Network. *Chin. J. Netw. Inf. Secur.* **2021**, *7*, 149–155. Available online: <http://kns.cnki.net/kcms/detail/10.1366.TP.20210610.1503.004.html> (accessed on 2 November 2021).
23. Zhao, J.; Zhao, L.; Cui, L.; Chen, M.; Qiu, T.; Chen, B. Case Based Reasoning Framework for Automating HAZOP Analysis. *CIESC J.* **2008**, *59*, 111–117.
24. Wang, F.; Gao, J.; Liu, W. Hazard and Operability and Explosion Accidents Analysis Technology in Chemical Processes. *CIESC J.* **2008**, *59*, 3184–3190.
25. Wang, F.; Gao, J.; Zang, B.; Zhang, X. Computer aided HAZOP analysis technology based on AHP. *Chem. Ind. Eng. Prog.* **2008**, *12*, 2013–2018.
26. Kiyoshi, K.; Rafael, B. *An Ontological Approach to Represent HAZOP Information*; Tokyo Institute of Technology: Tokyo, Japan, 2003; Available online: <http://ise.me.tut.ac.jp/members-e/rbp/pubs/ontological-approach-to-represent-hazop.pdf> (accessed on 2 November 2021).
27. Wu, C.; Xu, X.; Na, Y.; Zhang, W. Standardized Information for Process Hazard Analysis Based on Ontology. *CIESC J.* **2012**, *63*, 1484–1491.
28. Zhang, B.; Xu, X.; Na, Y.; Wu, C. Hazard Analysis and Application Based on Graphical Scenario Object Model. *CIESC J.* **2013**, *64*, 2511–2519.
29. Gao, D.; Xiao, Y.; Zhang, B.; Xu, X.; Wu, C. HAZOP Information Standardization Framework Based on Knowledge Ontology. *Chem. Ind. Eng. Prog.* **2020**, *39*, 2510–2518.
30. Li, F.; Zhang, B.; Gao, D. HAZOP Knowledge Graph Construction Method. *Chem. Ind. Eng. Prog.* **2021**, *40*, 4666–4677.