

Article

Similarity Calculation of Sudden Natural Disaster Cases with Fused Case Hierarchy—Taking Storm Surge Disasters as Examples

Cheng Cheng * , Qingtian Zeng * , Hua Zhao  and Shansong Wang 

College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

* Correspondence: 5o2o@163.com (C.C.); qtzeng@sdu.edu.cn (Q.Z.)

Abstract: Sudden natural disasters have the characteristics of complexity, unpredictability and frequency. To better manage and analyze sudden natural disasters promptly with the help of historical natural disaster cases, this study adopts the method of fused case structure to calculate the similarity between sudden disaster cases. Based on the disaster information of historical natural disaster cases, this paper aims to perform similarity measures for sudden natural disaster cases that contain textual information, data information and geographic location information at the same time. Taking storm surge disasters as examples, we designed a hierarchical case structure of “vertex-edge-label” based on the characteristics of sudden natural disaster cases. Then, we calculated the case similarity based on three aspects of sudden natural disasters, which were “case scenario”, “disaster damage” and “case structure”. Finally, we aggregated multiple factors to obtain the similarity between storm surge cases and conducted experiments on the historical storm surge cases in China. The result verified the feasibility and effectiveness of the method and showed a higher accuracy of the established aggregated multifactor method compared with the benchmark method.



Citation: Cheng, C.; Zeng, Q.; Zhao, H.; Wang, S. Similarity Calculation of Sudden Natural Disaster Cases with Fused Case Hierarchy—Taking Storm Surge Disasters as Examples. *J. Mar. Sci. Eng.* **2022**, *10*, 1218. <https://doi.org/10.3390/jmse10091218>

Academic Editor: Han Soo Lee

Received: 25 July 2022

Accepted: 26 August 2022

Published: 31 August 2022

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



Copyright: © 2022 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/>).

Keywords: sudden natural disasters; domain ontology; similarity calculation; storm surge disasters

1. Introduction

With rising temperatures, global extreme weather events have increased, causing an upsurge in natural disasters worldwide. Natural disaster outbreaks are generally complex, unpredictable and frequent and have a serious impact on social and economic development [1–3]. One of the most serious marine natural disasters in the world is a storm surge disaster, which causes serious damage to the sea, nearshore and inland [4–6]. Storm surge disasters are transient, high-risk, complex and frequent, resulting in serious damage in a short time [7–9]. After the occurrence of sudden natural disasters, people need to respond promptly for an immediate analysis and handling, by comparing with historical natural disaster information. Furthermore, they need to make a qualitative damage determination in real time based on the actual situation of sudden natural disasters, so as to facilitate the subsequent adjustment and decision-making for effective rescue measures [10,11].

An important empirical resource for natural disaster emergency managers is the relevant historical disaster cases [6,12], but historical case information is widely recorded in texts and people need to obtain historical natural disaster information and knowledge from these texts to manage and analyze disasters [13]. As reference knowledge, historical natural disaster cases can provide a scientific basis and support for emergency decision-making to achieve an orderly and effective response over a short time. In recent years, scenario construction techniques have been proposed in the field of emergency management from home and abroad [14,15]. They can generate solutions to new problems by using source cases through case-based reasoning (CBR) techniques, and they complete the mapping of similar elements between source and target cases via reasoning [16,17]. Existing studies have carried out a large number of case similarity calculations for scenario building in the

areas of business [18–20], medical health [21–23] and disaster incidents [24,25]. However, the existing studies provide very limited help for emergency decisions during sudden natural disasters [2,26].

The paper [27] in the field of emergency management of sudden natural disasters proposed an ontology-based semantic similarity calculation model to calculate the similarity by quantifying the semantic similarity between concepts. The paper [28] proposed a case inference based on a differential evolution algorithm, which obtained a new differential evolution algorithm through a hybrid mutation operator and improved the adaptiveness of the model. The paper [29] proposed a modeling and analysis of disaster chains based on stochastic Petri nets, established the model elements of disaster chains and analyzed their control flow relations. The paper [30] proposed an emergency risk decision-making method based on fault tree analysis; they constructed fault trees by analyzing the evolution process of emergencies and described the logical relations between the conditions and factors leading to the evolution of emergencies.

The following problems and challenges are found when calculating the similarity between cases in the natural disaster field: (1) natural disaster information is usually recorded in unstructured texts, which generally contain a lot of noisy data (texts that are irrelevant to disaster descriptions); (2) texts contain a variety of information, such as text information, data information, geographic information, etc., which makes it difficult to calculate the similarity between cases by processing the corpus uniformly; (3) natural disaster cases contain many disaster events, and the values of the disaster-bearing body and damage information in disaster are not unified, which makes it more difficult to judge the similarity between disaster cases; (4) the temporal information between disaster events is not obvious as many natural disasters are transient in nature, resulting in a lack of temporal information in the text; (5) the spatial distribution of disaster events in natural disaster cases is wide and disasters are generally scattered in different locations.

It can be found that most of the existing studies focus on disaster analysis, numerical simulation, prediction and warning, damage assessment and emergency decision-making in the field of sudden natural disasters [31], while the comprehensive problems caused by sudden natural disasters are not considered globally. The distribution of disaster “dynamics” in the field of sudden natural disasters is widely varied, resulting in a large number of influencing factors. In this paper, we focus on the geographic distribution and the damage of natural disasters and build a multifactor, multiaspect and multihierarchy model to calculate the similarity between natural disaster cases. The main contributions of the study can be summarized as follows:

- We transform the case similarity calculation into a similarity calculation of the case structure and case node labels by using the “vertex-edge-label” case structure;
- We propose the similarity calculation of sudden natural disaster cases with a fused case hierarchy and we calculate the similarity in “case scenario”, “disaster damage” and “case structure” in a multihierarchy, multiaspect and multifactor way;
- The experimental results of the storm surge disaster cases show that the results of aggregating the three similarities of “case scenario”, “disaster damage” and “case structure” in this paper are better than other comparison methods.

2. Related Work

The intercase similarity calculations in traditional methods are mainly based on a case’s textual content [32–34] and case hierarchy [35,36].

In the similarity calculation method based on a case’s textual content, an ontology mapping is generally an effective means of multisource data fusion and an ontology enables the sharing, common understanding and reuse of the domain knowledge [27,37]. An ontology mapping obtains the source ontology that is similar to the target ontology by a similarity calculation method. It forms a mapping relationship, obtains a unified global representation and realizes the effective fusion of multisource data. The ontology-based similarity calculation methods can be broadly classified into five categories: semantic-

distance-based similarity calculation methods [38], information-content-based similarity calculation methods [33,39], concept-attribute-based similarity calculation methods [40], hybrid semantic similarity calculation methods [41] and deep-learning-based similarity calculation methods [42,43]. Among them, the calculation methods based on semantic distance have a low computational complexity, but the single consideration makes the results less accurate and unstable. For the calculation methods based on information content, the results are more accurate for a corpus with a high completeness and the algorithm does not have robustness. The calculation method based on concept attributes shows more accurate results but ignores the location and content information between concept nodes. Furthermore, the hybrid semantic similarity calculation method has more objective and accurate results but has a high computational complexity. Furthermore, the calculation method based on deep learning has accurate results but poor convergence. In this paper, we build a domain ontology for the domain of storm surge natural disasters and use the similarity calculation method of information content to calculate the similarity of the cases' textual content.

The graph structure is flexible and widely used in terms of similarity calculation methods for case hierarchies, and the graphs are available in the form of a Petri net model [44,45], BPMN network model [46], Bayesian network model [47], etc. a graph structure is used for the similarity measure of sudden natural disaster case hierarchies by methods such as the graph edit distance [48], maximum common subgraph [49] and graph isomorphism [50] for the similarity search and graph matching on the graph structure. With the development of neural networks, many studies have recently turned similarity estimation into a learning problem, and a graph neural network is a powerful tool for learning various structural graph representations [51,52]. However, since the extracted sudden natural disaster case structures have fewer graph structure nodes, shallower layers and smaller sizes, which makes the neural network [53] unable to learn effective features, we used the unsupervised graph edit distance method to learn the similarity between sudden natural disaster case hierarchies.

For the main problems of the similarity calculations between sudden natural disasters cases, to deal with sudden natural disaster cases timely and intelligently and to overcome the one-sidedness and limitations of the existing cases' text content and hierarchy, we propose a case similarity calculation with a fused case hierarchy.

3. Disaster Case Analysis

The current “scenario–response” model is the trend in the study of emergency natural disaster response, and managers need to study the “situation” of the occurrence and the development of that current emergency natural disaster [54]. The characteristics of the storm surge disaster “situation” are that the “situation” is spatially dispersed and temporally short-lived. Figure 1 shows the decomposition of multiple events in the spatial dimension of a storm surge disaster case, which shows the spatial distribution of the disaster conditions after the storm surge disaster $V_{M_1} \sim V_{M_8}$. The distribution of the disaster is generally expressed as the information and distribution of the storm-surge-disaster-bearing body, which is a manifestation of the set of “states”.

We performed the analysis from the perspective of the disposal model of emergency decision-making in sudden natural disasters, assuming that the fundamental goal of emergency decision-making is to control, reduce and eliminate the impact of the disaster around different disaster-bearing bodies. Based on the fundamental goal of emergency decision-making, we combined the definition and spatial analysis of sudden natural disaster texts. We believe that the damage of the disaster-bearing body (disaster damage) and the disaster scenario constitute the core of sudden natural disaster cases. Furthermore, we propose the concept of “sudden natural disaster cases”.

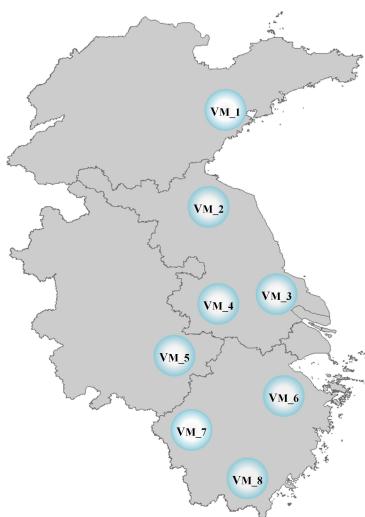


Figure 1. Decomposition of events in the spatial dimension in storm surge disaster.

This paper used 281 cases of storm surge disasters in China from 1949 to 2019. We analyzed the text of sudden natural disaster cases and found that they contained many case scenarios and disaster damage information. Therefore, in response to the spatial and temporal distribution characteristics of natural disasters, we constructed “sudden natural disaster cases” by converting case scenarios and disaster damage into case structure hierarchies according to the geographical location of natural disasters. We defined a case as a collection of information, which combined the disaster information suffered by a disaster-bearing body and the disaster-generating environment in a sudden natural disaster. A sample of sudden natural disasters from storm surges is shown in Figure 2.

Specifically, a sudden natural disaster case can be represented formally as a triplet $G = \{V, E, N\}$, where V is the set of all nodes of the graph, E is the set of all edges of the graph and N is the set of all node labels in the graph.

The set $V = \{V_E, V_I, V_P, V_M, V_L\}$ is the set of case nodes, $V_E = \{V_{E_1}, V_{E_2}, \dots\}$ is the set of case ID nodes, $V_I = \{V_{I_1}, V_{I_2}, \dots\}$ is the set of case damage nodes, $V_P = \{V_{P_1}, V_{P_2}, \dots\}$ is the set of case attribute nodes, $V_M = \{V_{M_1}, V_{M_2}, \dots\}$ is the set of disaster damage nodes and $V_L = \{V_{L_1}, V_{L_2}, \dots\}$ is the set of disaster geographic location nodes. The set of case edges $E = \{\langle V_i, V_j \rangle, \dots, \langle V_m, V_n \rangle, \dots\}$ is the set of directed edges, which is used to represent the order relation between the case nodes of sudden natural disasters.

The set of case node labels is $N = \{N_I, N_P, N_M, N_L\}$, where $N_I = \{Ha_I, Pr_I, Da_I\}$ is the set of case damage information and the statistical information on dead population, affected population, house loss, economic loss, crop damage and overalert tide value in sudden natural events, consisting of the set of case disaster-bearing bodies $Ha_I = \{Ha_{I_1}, Ha_{I_2}, \dots\}$, the set of attributes of the case disaster-bearing body $Pr_I = \{Pr_{I_1}, Pr_{I_2}, \dots\}$ and the set of data of the case disaster-bearing body $Da_I = \{Da_{I_1}, Da_{I_2}, \dots\}$; $N_P = \{Rep, Cop\}$ is the set of case attribute information, the meteorological attributes and time attributes, etc., at the landing of sudden natural cases, composed of the set of case attribute types $Rep = \{Rep_1, Rep_2, \dots\}$ and the set of case attribute data $Cop = \{Cop_1, Cop_2, \dots\}$; $N_M = \{Ha_M, Pr_M, Da_M\}$ is a disaster damage information set and the basic information of a sudden natural disaster event element, composed of a disaster-damage-bearing body set $Ha_M = \{Ha_{M_1}, Ha_{M_2}, \dots\}$, the set of attributes of the damage-bearing body $Pr_M = \{Pr_{M_1}, Pr_{M_2}, \dots\}$ and the data set of disaster-damage-bearing body $Da_M = \{Da_{M_1}, Da_{M_2}, \dots\}$; $N_L = \{Lo_L\}$ is a disaster geographic location information collection, which is the specific location of the event caused by sudden natural disasters and it consists of the set of geographic locations $Lo_L = \{Lo_{L_1}, Lo_{L_2}, \dots\}$. The sets Da_I, Cop, Da_M are numerical information and the other nodes are textual information.

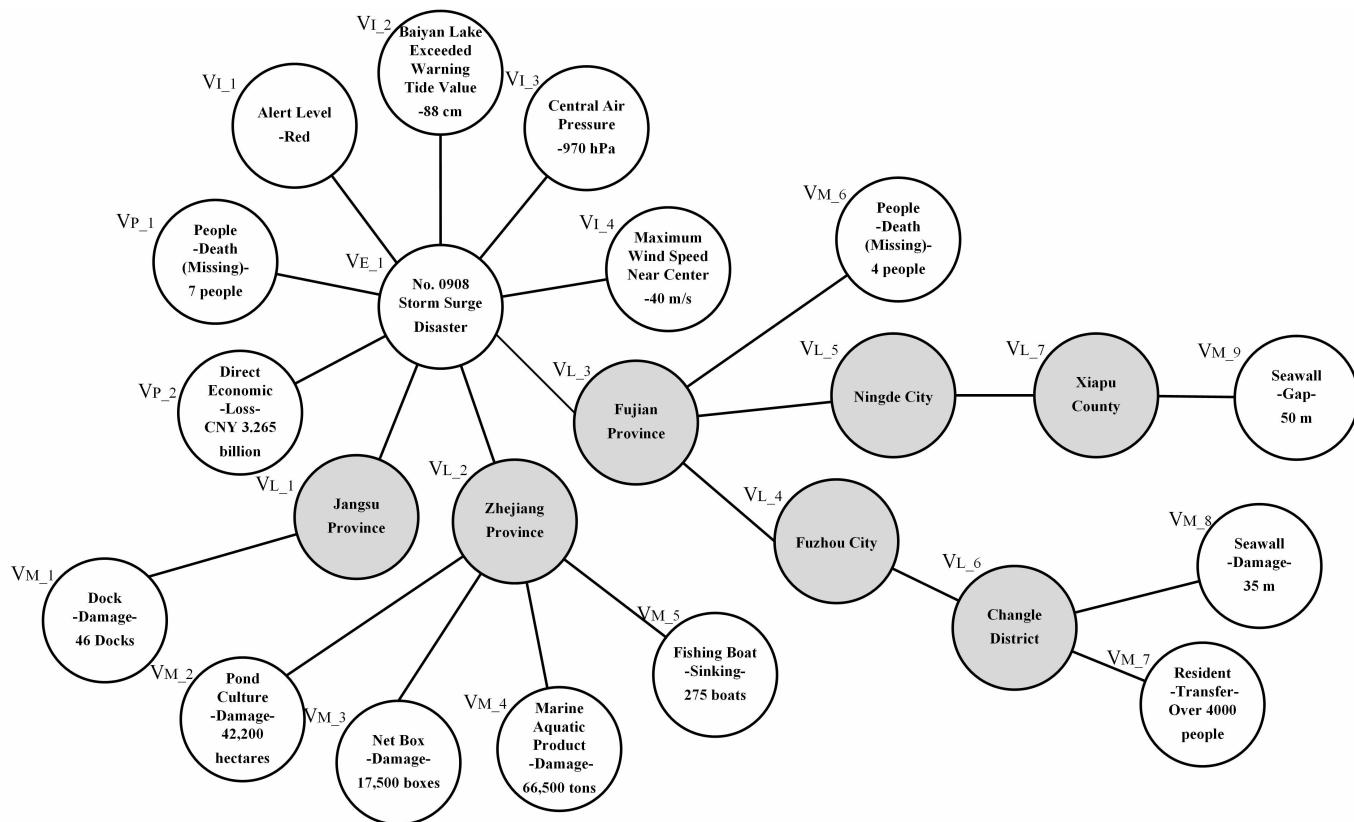


Figure 2. Sample of storm surge disaster cases.

4. Similarity Calculation between Cases

The similarity between cases was obtained by combining the similarity in “case scenario”, “disaster damage” and “case structure” between two cases. The similarity in “case scenario” in the case of sudden natural disasters was calculated from the label data between the case loss node and the case attribute node. The similarity in “disaster damage” was calculated from the textual information and data information in the label data. Furthermore, the “case structure” was the geographic location information of the event, which was transformed into a case hierarchy for the similarity calculation. In other words, the similarity in the label text and the case structure were calculated and analyzed in a multihierarchy, multiaspect and multifactor way for the label text and structure of the cases that constituted sudden natural disasters. The architecture of the similarity model between natural disaster cases is shown in Figure 3.

4.1. Similarity between Case Scenarios

The “case scenario” mainly contains the natural and disaster-forming attributes of the disaster [55]. The natural attributes of the storm surge disaster were mainly the statistical information about the disaster level, landfall time, landfall location, landfall wind speed, central air pressure and coastal water increase in the case; the disaster-forming attributes of the storm surge disaster were mainly the statistical information about the dead population, affected population, housing damage, economic loss, crop damage and excess tide value. The natural attributes and disaster-forming attributes corresponded to the case loss node V_I and the case attribute node V_P in the case, respectively.

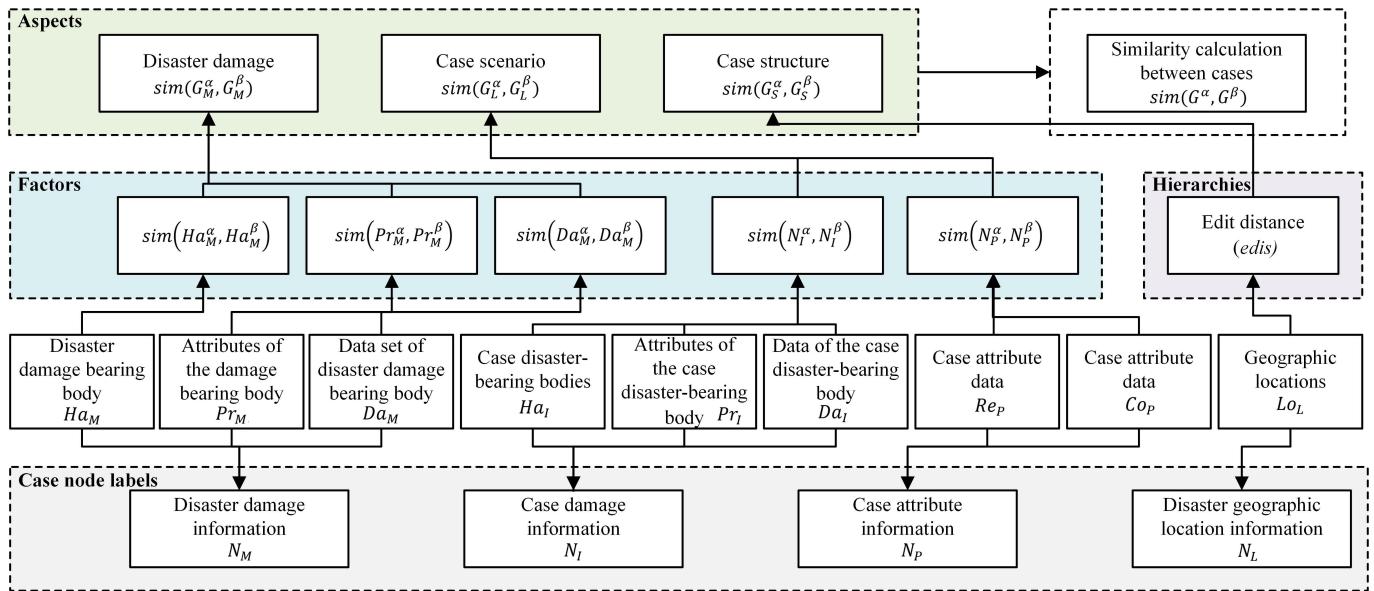


Figure 3. Architecture of the similarity model between natural disaster cases.

Specifically, we used the disaster level (*LL*), landfall wind speed (*WL*), central air pressure (*PL*) and coastal water increase (*IL*) information for the natural attributes' similarity calculation. For disaster-forming attributes followed the Statistical System of Natural Disaster Situation (Minfa [2016] No. 23) and Natural Disaster Statistics (Part 1): Basic Indicators (GB/T24438.1-2009), and for case damage we used the death population (*DC*), affected population (*PC*), house damage (*HC*), economic loss (*EC*), crop damage (*CC*) and exceeded warning tide values (*SC*) for the similarity calculation.

The similarity between the “case scenarios” was calculated by combining the information of natural attributes and disaster attributes, which was transferred to the similarity calculation of the label information in the V_I and V_P nodes of the disaster case. The formulas for the natural attributes in the “case scenario” were given by the similarity between disaster levels, similarity between landfall wind speeds and central air pressures and similarity between coastal water increases, respectively, and the formulas for the disaster attributes were for the similarity between case damages.

(1) Calculation of disaster level similarity: a storm surge disaster level P_{LL} is generally divided into four levels: red, orange, yellow and blue, corresponding to level I, level II, level III and level IV and corresponding to values 1~4. Thus, the disaster level similarity between case a and case B is calculated by the formula shown in Equation (1).

$$\text{sim}(P_{LL}^\alpha, P_{LL}^\beta) = 1 - \frac{|P_{LL}^\alpha - P_{LL}^\beta|}{\max(P_{LL}^\alpha, P_{LL}^\beta)} \quad (1)$$

(2) Landing wind speed and central pressure similarity calculation: landing wind speed P_{WL} and central pressure P_{PL} are both storm surge disaster environments; the more similar the landing wind speed and central pressure of cases a and B are to the environment, the higher their similarity is. The calculation formula is shown in Equations (2) and (3).

$$\text{sim}(P_{WL}^\alpha, P_{WL}^\beta) = 1 - \frac{|P_{WL}^\alpha - P_{WL}^\beta|}{\max(P_{WL}^\alpha, P_{WL}^\beta)} \quad (2)$$

$$\text{sim}(P_{PL}^\alpha, P_{PL}^\beta) = 1 - \frac{|P_{PL}^\alpha - P_{PL}^\beta|}{\max(P_{PL}^\alpha, P_{PL}^\beta)} \quad (3)$$

(3) Coastal water increase similarity calculation: the maximum water increase value P_{IL} in the storm surge disaster process was divided into five levels: extralarge, large, relatively large, medium and general, corresponding to level I, level II, level III, level IV and level V. The specific classification criteria for the levels are shown in Table 1. Coastal water increase counting in storm surge disasters involves multiple tide gauge sites. We first counted the number of tide gauge sites and tide gauge site water increase values in cases a and B; then, according to the corresponding values 1~5 of the grade, we averaged the site water increase values and calculated the similarity between coastal water increases in cases a and B. The calculation formula is shown in Equation (4).

$$\text{sim}(P_{IL}^{\alpha}, P_{IL}^{\beta}) = 1 - \frac{|P_{IL}^{\alpha} - P_{IL}^{\beta}|}{\max(P_{IL}^{\alpha}, P_{IL}^{\beta})} \quad (4)$$

Table 1. Storm water increase grade standard

Grade	Grade I	Grade II	Grade III	Grade IV	Grade V
Water increase value	≥ 251 cm	201~250 cm	151~200 cm	101~150 cm	50~100 cm

In conclusion, we combined these four calculations results of the disaster level, landfall wind speed, central air pressure and coastal water increase to derive the case attribute similarity $\text{sim}(N_P^{\alpha}, N_P^{\beta})$, and the formula is shown in Equation (5).

$$\text{sim}(N_P^{\alpha}, N_P^{\beta}) = \frac{\text{sim}(P_{LL}^{\alpha}, P_{LL}^{\beta}) + \text{sim}(P_{WL}^{\alpha}, P_{WL}^{\beta}) + \text{sim}(P_{PL}^{\alpha}, P_{PL}^{\beta}) + \text{sim}(P_{IL}^{\alpha}, P_{IL}^{\beta})}{4} \quad (5)$$

(4) Case damage similarity calculation: death (missing) people I_{DC} , affected population I_{PC} , house collapse I_{HC} , direct economic loss I_{EC} , crop damage area I_{CC} and exceeded warning tide value I_{SC} were divided into four grades (extralarge, large, relatively large, general) and the specific classification criteria are shown in Table 2. The case damage similarity between cases a and B was calculated according to the values 1~4 corresponding to their rank, where the formula for calculating the number of fatalities is shown in Equation (6) and $\text{sim}(I_{PC}^{\alpha}, I_{PC}^{\beta}), \text{sim}(I_{HC}^{\alpha}, I_{HC}^{\beta}), \text{sim}(I_{EC}^{\alpha}, I_{EC}^{\beta}), \text{sim}(I_{CC}^{\alpha}, I_{CC}^{\beta})$ and $\text{sim}(I_{SC}^{\alpha}, I_{SC}^{\beta})$ also can be calculated by the same formula.

$$\text{sim}(I_{DC}^{\alpha}, I_{DC}^{\beta}) = 1 - \frac{|I_{DC}^{\alpha} - I_{DC}^{\beta}|}{\max(I_{DC}^{\alpha}, I_{DC}^{\beta})} \quad (6)$$

Table 2. Criteria for disaster-forming attribute grades in case scenarios.

Grade	Grade I	Grade II	Grade III	Grade IV
Death (missing) people	≥ 100 people	31~100 people	10~30 people	less than 10 people 0.1~0.5 million people
Affected population	≥ 2 million people	1~2 million people	0.5~1 million people	less than 100,000 houses
House collapse	$\geq 300,000$ houses	200,000~300,000 houses	100,000~200,000 houses	<CNY 1 billion
Direct economic loss	CNY ≥ 5 billion	CNY 2~5 billion	CNY 1~2 billion	less than 100,000 hectares
Crop damage area	$\geq 500,000$ hectares	200,000~500,000 hectares	100,000~200,000 hectares	0~30 cm
Exceeded warning tide value	≥ 151 cm	81~150 cm	31~80 cm	

In conclusion, the six calculation results of death (missing) people, affected population, collapsed houses, direct economic loss, crop damage area and exceeded warning tide value were combined to derive the case damage similarity $\text{sim}(N_I^\alpha, N_I^\beta)$, and the formula is shown in Equation (7).

$$\text{sim}(N_I^\alpha, N_I^\beta) = \frac{\text{sim}(I_{DC}^\alpha, I_{DC}^\beta) + \text{sim}(I_{PC}^\alpha, I_{PC}^\beta) + \text{sim}(I_{HC}^\alpha, I_{HC}^\beta) + \text{sim}(I_{EC}^\alpha, I_{EC}^\beta) + \text{sim}(I_{CC}^\alpha, I_{CC}^\beta) + \text{sim}(I_{SC}^\alpha, I_{SC}^\beta)}{6} \quad (7)$$

4.2. Similarity in Disaster Damage

Normally, a disaster case contains multiple “disaster damage” pieces of information. For example, a storm surge disaster case may contain events such as “46 docks destroyed in Jiangsu Province”, “275 fishing boats sunk in Zhejiang Province”, etc., where the “disaster damage” corresponds to the node V_M in the case of a sudden natural disaster, and the label information of the node V_M is “46 docks destroyed” and “275 fishing boats sunk”; the label information contains the disaster’s textual information “docks” and “fishing boats” as the “disaster damage”-bearing body Ha_M ; “destroyed” and “sunk” are the “disaster damage”-bearing body attributes Pr_M ; “46 docks” and “275 boats” are the “disaster damage”-bearing body data Da_M .

All the “disaster damage” elements in cases a and B contributed to the similarity calculated by the textual information of Ha_M , Pr_M and the data information of Da_M .

4.2.1. Similarity in Textual Information

Both disaster-bearing body information Ha_M and disaster-bearing body attribute information Pr_M in the “disaster damage” information are concepts of the mentioned disaster domain ontology. The similarity calculation of Ha_M and Pr_M can be transformed into the similarity calculation between concepts of the ontology. In our previous work, we constructed the storm surge disaster domain ontology by comprehensively referring to the existing domain ontologies for sudden natural disasters and the National General Emergency Response Plan for Public Emergencies issued by the State Council and other emergency management departments [56,57]. Figure 4 shows the hierarchy among some concepts in the storm surge disaster domain ontology.

In this paper, the similarity between the disaster-bearing body Ha_M^α of storm surge disaster case a and the Ha_M^β of case B was taken as an example for the calculation of textual information similarity, which mainly contained the following three steps.

(1) Construct the disaster similarity matrix: Ha_M : the sets of Ha_M of the two cases are $Ha_M^\alpha = \{Ha_{M_1}^\alpha, Ha_{M_2}^\alpha, \dots, Ha_{M-m}^\alpha\}$, $m = (1, 2, \dots)$ and $Ha_M^\beta = \{Ha_{M_1}^\beta, Ha_{M_2}^\beta, \dots, Ha_{M-n}^\beta\}$, $n = (1, 2, \dots)$. Then, the following $m \times n$ disaster-bearing body similarity matrix M_{Ha_M} can be established based on the two Ha_M sets, and the calculation formula is shown in Equation (8).

$$M_{Ha_M} = \begin{bmatrix} \text{Sim}\left(Ha_{M_1}^\alpha, Ha_{M_1}^\beta\right) & \text{Sim}\left(Ha_{M_1}^\alpha, Ha_{M_2}^\beta\right) & \dots & \text{Sim}\left(Ha_{M_1}^\alpha, Ha_{M-n}^\beta\right) \\ \text{Sim}\left(Ha_{M_2}^\alpha, Ha_{M_1}^\beta\right) & \text{Sim}\left(Ha_{M_2}^\alpha, Ha_{M_2}^\beta\right) & \dots & \text{Sim}\left(Ha_{M_2}^\alpha, Ha_{M-n}^\beta\right) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Sim}\left(Ha_{M-m}^\alpha, Ha_{M_1}^\beta\right) & \text{Sim}\left(Ha_{M-m}^\alpha, Ha_{M_2}^\beta\right) & \dots & \text{Sim}\left(Ha_{M-m}^\alpha, Ha_{M-n}^\beta\right) \end{bmatrix} \quad (8)$$

where $\text{Sim}(Ha_{M_i}^\alpha, Ha_{M_j}^\beta)(i = 1, 2, \dots, m; j = 1, 2, \dots, n)$ is the similarity between the disaster-bearing body $Ha_{M_i}^\alpha$ in the set Ha_M^α and $Ha_{M_j}^\beta$ in the set Ha_M^β .

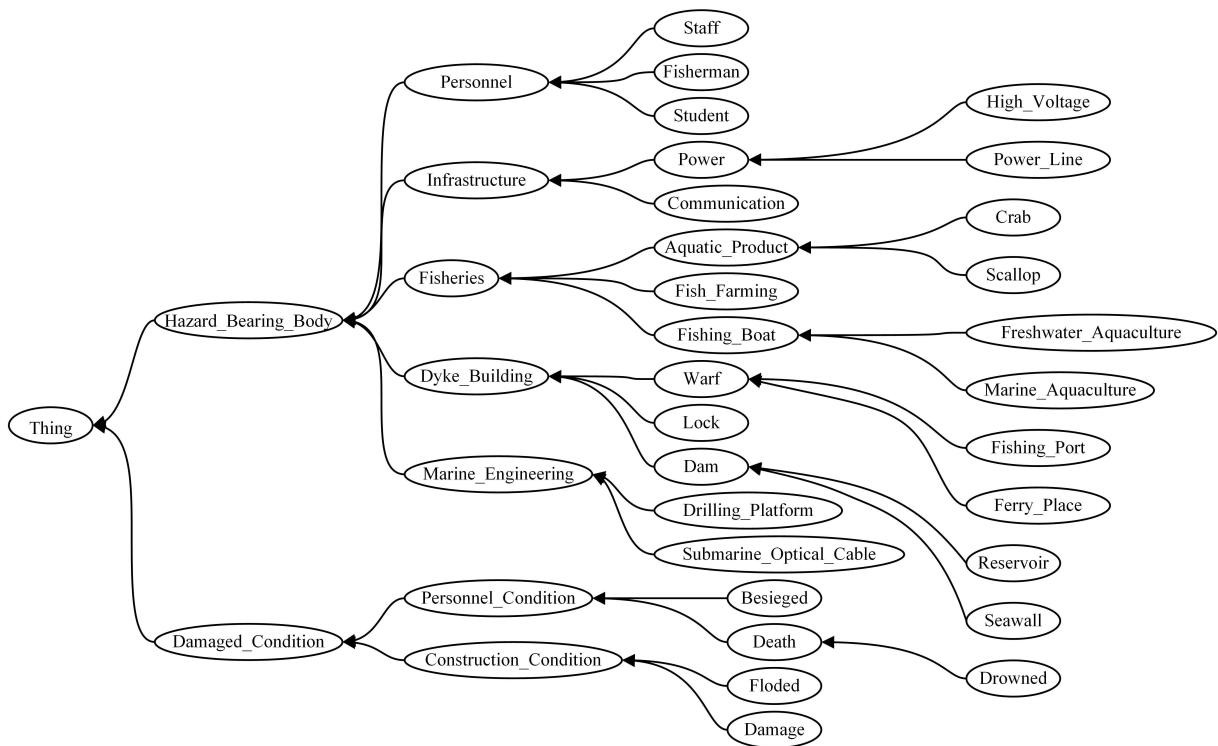


Figure 4. The concept of storm surge disaster domain ontology.

(2) Calculate the value of each element in the similarity matrix: each element of the matrix M_{Ha_M} represents the similarity between two concepts, and they can be calculated by any concept hierarchy method in the domain ontology. The similarity calculation method based on the domain ontology concept was used to calculate the concept similarity matrix M_{Ha_M} for the content based on the amount of concept information within the ontology. The elements in the matrix M_{Ha_M} were calculated by the formula shown in Equation (9).

$$Sim(Ha_{M_i}^{\alpha}, Ha_{M_j}^{\beta}) = \frac{2IC(Ha_{M_i}^{\alpha} \cap Ha_{M_j}^{\beta})}{IC(Ha_{M_i}^{\alpha}) + IC(Ha_{M_j}^{\beta})} \quad (9)$$

where $Ha_{M_i}^{\alpha} \cap Ha_{M_j}^{\beta}$ is the commonality between two words, which is expressed as the nearest neighbor common ancestor concept of concepts $Ha_{M_i}^{\alpha}$ and $Ha_{M_j}^{\beta}$ in the ontology structure tree, and $IC(Ha_{M_i}^{\alpha})$ and $IC(Ha_{M_j}^{\beta})$ denote the information quantity of concepts $Ha_{M_i}^{\alpha}$ and $Ha_{M_j}^{\beta}$, respectively.

The IC value illustrates the amount of information provided by a concept when it appears in context [58], and the IC value of the function used the method reported in [59] to calculate the information content of a concept. The value of the function increases incrementally with the depth of the ontology, which means a more abstract concept has a smaller IC value and a more concrete concept has a larger IC value. This approach ensures that the value of the concept decreases as it moves from the leaf node to the root node of the hierarchy, and the calculated value is greater for the more concrete concept. To calculate the concept information content (IC), the formula shown in Equation (10) was used:

$$IC(Ha_{M_k}) = -\log\left(\frac{|leaves(Ha_{M_k})| / |subsumers(Ha_{M_k})| + 1}{sum_{leaves} + 1}\right) \quad (10)$$

where $IC(Ha_{M_k})$ is the information content of the concept Ha_{M_k} , and Ha_{M_k} refers to a specific concept. $|leaves(Ha_{M_k})|$ is the number of the set composed of all leaf concepts

under Ha_{M_k} , $|subsumers(Ha_{M_k})|$ is the number of the set composed of all leaf concepts and all their parent concepts Ha_{M_k} , and sum_leaves is the number of all leaf concepts in the storm surge disaster domain ontology.

(3) Calculate the similarity of textual information: after calculating the value of each element in M_{HaM} , we selected the element with the largest value $sim(Ha_M^\alpha, Ha_M^\beta)$ and added it to the set \mathcal{M} . Then, we removed all the elements in row i and column j from M_{HaM} . Repeating the above procedure until the number of elements in the set \mathcal{M} gets to $|\mathcal{M}| = \min(|Ha_M^\alpha|, |Ha_M^\beta|)$, assuming that the set $\mathcal{M} = \{sim_1, sim_2, \dots, sim_N\}$, the similarity between the set of disaster-bearing bodies can be calculated by the sum of all elements of the normalized set \mathcal{A} . The corresponding formula is shown in Equation (11).

$$sim(Ha_M^\alpha, Ha_M^\beta) = \frac{(|Ha_M^\alpha| + |Ha_M^\beta|) \sum_{k=1}^N sim_k}{2|Ha_M^\alpha| \times |Ha_M^\beta|} \quad (11)$$

In the same way, we can calculate the similarity $sim(Pr_M^\alpha, Pr_M^\beta)$ between the set of attributes Pr_M in cases a and B.

4.2.2. Similarity in Data Information

Considering the different data types in the data set Da_M of the disaster-bearing body, we classified the data information into three types of cases, namely, exact values, definite intervals and fuzzy intervals, for the calculation of the similarity.

(1) Similarity between exact values: the data of exact value type can be continuous or discrete, and we used the evolution formula based on the Hamming distance formula [60] to calculate the similarity of value attributes in this paper. The formula is shown in Equation (12), where a, b are the specific data information values of concepts C_1, C_2 and S denotes the range of these data information values.

$$sim(C_1, C_2) = 1 - \frac{|a - b|}{S} \quad (12)$$

(2) Similarity between definite intervals: a definite interval is a set of real numbers in a closed interval with definite upper and lower bounds and the data information interval value of the concept C_1 is (c, c') and for the concept C_2 , it is (d, d') . The formula for calculating the similarity between them is shown in Equation (13).

$$sim(C_1, C_2) = \frac{(c, c') \cap (d, d')}{\max(c', d') - \min(c, d)} \quad (13)$$

(3) Similarity between fuzzy intervals: fuzzy intervals have no definite upper and lower bounds, the values of such data information are usually sets and the elements in the sets are fuzzy concepts corresponding to the concept variables x . In this paper, the intervals were divided by the trapezoidal affiliation function [61], and the affiliation function was as in Equation (14), shown in Figure 5. The similarity between the concepts C_1 and C_2 was set to one in the optimal value interval, and the other intervals were expressed by the affiliation function relation. The two fuzzy intervals and their overlapping interval areas were calculated according to the responding affiliation function, and the rate of area overlapping was the similarity between the fuzzy intervals. The formula is shown in Equation (15), where g, k are the fuzzy intervals of concepts C_1 and C_2 .

$$f(x) = \begin{cases} \frac{x - \min(y)}{\max(y) - \min(y)} & \forall x \in [\min(y), \max(y)] \\ 1 & \forall x \in [k_1, k_2] \\ \frac{\max(y) - x}{\max(y) - \min(y)} & \forall x \in (k_2, \max(y)) \\ 0 & \text{other} \end{cases} \quad (14)$$

$$\text{sim}(C_1, C_2) = \frac{g \cap k}{g - k - (g \cap k)} \quad (15)$$

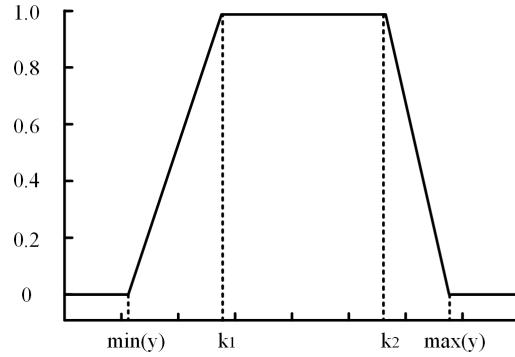


Figure 5. Fuzzy interval affiliation function.

The steps to calculate the similarity matrix of the data information were the same as those of the text information and took the calculation of the similarity $\text{sim}(Da_M^\alpha, Da_M^\beta)$ of the set Da_M as an example. We needed to calculate the value of each element in M_{Da_M} at first and select the element with the largest value to add to the set \mathcal{M} . Then, we standardized the sum of all elements of the set \mathcal{M} to obtain the final “disaster damage” set Da_M of cases a and B, and the formula is shown in Equation (16).

$$\text{sim}(Da_M^\alpha, Da_M^\beta) = \frac{(|Da_M^\alpha| + |Da_M^\beta|) \sum_{k=1}^N \text{sim}_k}{2|Da_M^\alpha| \times |Da_M^\beta|} \quad (16)$$

4.3. Similarity between the Case Structures

In this paper, we converted the geographic location in disaster cases into a hierarchy for a multihierarchy similarity calculation, and a case hierarchy sample is shown in Figure 6. The difference in two cases’ structures was expressed as a sequence of editing operations, and the sequence of editing operations was the editing operations required when changing one case to another, such as deleting nodes, inserting nodes, deleting edges and inserting edges. Then, the case structure similarity between cases a and B was $\text{sim}(G_S^\alpha, G_S^\beta)$, which was calculated as shown in Equation (17).

$$\text{sim}(G_S^\alpha, G_S^\beta) = 1 - \frac{\text{edis}(G_S^\alpha, G_S^\beta)}{\max(|G_S^\alpha|, |G_S^\beta|)} \quad (17)$$

$$\text{edis}(G_S^\alpha, G_S^\beta) = w_{\text{subV}} \times \text{subV} + w_{\text{skipV}} \times \text{skipV} + w_{\text{skipE}} \times \text{skipE} \quad (18)$$

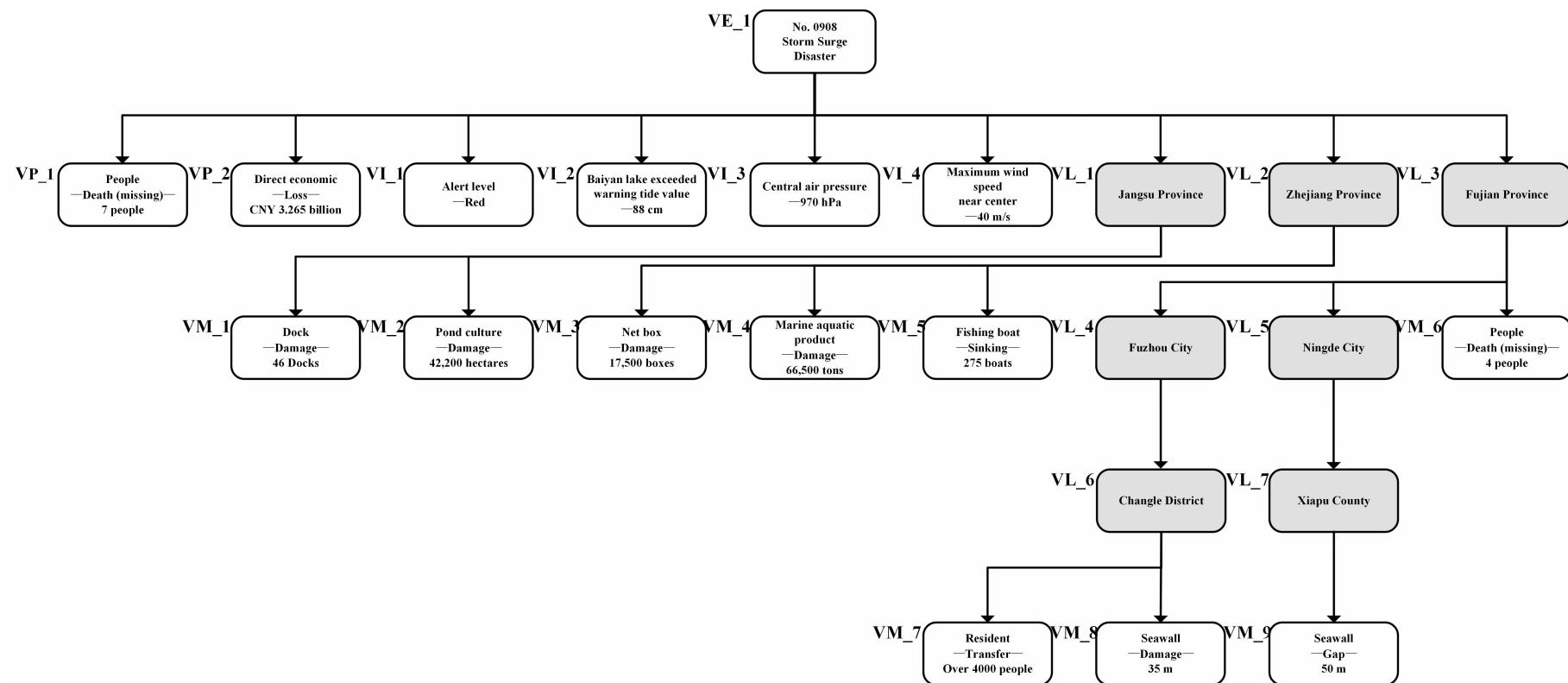


Figure 6. Sample of storm surge hazard case hierarchy.

In Equation (18), $\text{edis}(G_S^\alpha, G_S^\beta)$ is the edit distance between graph G_S^α and graph G_S^β , $|G_S^\alpha| = |V_S^\alpha| + |E_S^\alpha|$, $|V_S^\alpha|$ denotes the number of nodes in the graph G_S^α , $|E_S^\alpha|$ denotes the number of edges in the graph G_S^α , $\text{sub}V$ denotes the set of all nodes to be replaced, $\text{skip}V$ denotes the set of all nodes to be inserted or deleted, $\text{skip}E$ denotes the set of all edges to be inserted or deleted, $w_{\text{sub}V}$ denotes the cost weight assigned to the set of replacement nodes, $w_{\text{skip}V}$ denotes the cost weight assigned to the set of nodes to be added or deleted and $w_{\text{skip}E}$ denotes the cost weight assigned to the set of edges associated with the nodes to be added or deleted. For this group of parameters, we used the experimental best values $w_{\text{sub}V} = 1$, $w_{\text{skip}V} = 1$, $w_{\text{skip}E} = 0.5$.

4.4. Similarity Calculation between Cases

The similarity between two cases can be obtained by aggregating the label and structure information of the cases in a multihierarchy, multiaspect and multifactor way. Specifically, we set the source case a as $G^\alpha = (G_L^\alpha, G_M^\alpha, G_S^\alpha)$ and the target case B as $G^\beta = (G_L^\beta, G_M^\beta, G_S^\beta)$. The similarity between the two cases was calculated by the similarities of the “case scenario” G_L , “disaster damage” G_M and “case structure” G_S .

The “case scenario” similarity G_L can be obtained by combining the multifactor case attribute similarity $\text{sim}(N_P^\alpha, N_P^\beta)$ and the case damage similarity $\text{sim}(N_I^\alpha, N_I^\beta)$. The “disaster damage” similarity G_M can be obtained by combining the multifactor “disaster damage”-bearing body similarity $\text{sim}(Ha_M^\alpha, Ha_M^\beta)$, “disaster damage”-bearing body attribute similarity $\text{sim}(Pr_M^\alpha, Pr_M^\beta)$ and “disaster damage”-bearing body data similarity $\text{sim}(Da_M^\alpha, Da_M^\beta)$. Then, the similarity $\text{sim}(G^\alpha, G^\beta)$ between G^α and G^β was calculated as shown in Equation (19).

$$\text{sim}(G^\alpha, G^\beta) = w_1 \text{sim}(G_L^\alpha, G_L^\beta) + w_2 \text{sim}(G_M^\alpha, G_M^\beta) + w_3 \text{sim}(G_S^\alpha, G_S^\beta) \quad (19)$$

$$\text{sim}(G_L^\alpha, G_L^\beta) = w_{L1} \text{sim}(N_I^\alpha, N_I^\beta) + w_{L2} \text{sim}(N_P^\alpha, N_P^\beta) \quad (20)$$

$$\text{sim}(G_M^\alpha, G_M^\beta) = w_{M1} \text{sim}(Ha_M^\alpha, Ha_M^\beta) + w_{M2} \text{sim}(Pr_M^\alpha, Pr_M^\beta) + w_{M3} \text{sim}(Da_M^\alpha, Da_M^\beta) \quad (21)$$

where $\text{sim}(G^\alpha, G^\beta)$ is the overall similarity between the two cases, w_1 , w_2 and w_3 are the weights of the aggregation method, w_{L1} and w_{L2} are the weights in the “case scenario”, w_{M1} and w_{M2} are the weights in the “disaster damage”, and they satisfy the conditions of $0 \leq w_{L1}, w_{L2} \leq 1$, $0 \leq w_{M1}, w_{M2} \leq 1$, $0 \leq w_1, w_2, w_3 \leq 1$ and $w_{L1} + w_{L2} = 1$, $w_{M1} + w_{M2} = 1$, $w_1 + w_2 + w_3 = 1$.

5. Experimental Evaluation

5.1. Experimental Setup

The sudden natural disaster data were the collected storm surge disaster cases in China from 1949 to 2019, containing 281 storm surge disaster cases with a total of 39,340 case pairs, including four storm surge disaster grades: red, orange, yellow and blue. a total of 1155 case damage nodes, 1075 case attribute nodes, 5253 disaster damage nodes and 627 disaster geography nodes were extracted.

Eleven domain experts were invited to estimate the similarity between each storm surge disaster case pair; the domain experts included five domain-expert teachers and six PhD and graduate students in the field of sudden natural disasters. The domain experts made judgments based on the values of similarity between cases in four levels of no similarity [0 0.25], low similarity [0.25 0.50], medium similarity [0.50 0.75] and high similarity [0.75 1.0], respectively. We eliminated one maximum and one minimum value from the 11 estimation results and then averaged the remaining 9 estimation results of each case pair to get the final manual estimation similarity.

To better evaluate the performance of the aggregation method in this paper, we used 31,472 case pairs as the training set to determine the best weights for the method performance results and we used 7868 test case pairs as the input data. In this paper,

we used a regression analysis [62] to calculate the weights of the case similarity, and we used 31,472 case pairs and their final manually estimated similarity as input to calculate the maximum value of the Pearson correlation coefficient between the similarity of the aggregation method and the similarity of the manual estimation. The formula is shown in Equation (22).

$$\text{sim}(G^\alpha, G^\beta) = w_0 + w_1 \text{sim}(G_L^\alpha, G_L^\beta) + w_2 \text{sim}(G_M^\alpha, G_M^\beta) + w_3 \text{sim}(G_S^\alpha, G_S^\beta) + e \quad (22)$$

where w_0 is the constant term and e is the residual term; we finally obtained the best values of the weights from the experiment as $w_1 = 0.5972$, $w_2 = 0.2857$ and $w_3 = 0.1171$. The formula for calculating the intercase similarity is Equation (23), and the weight coefficients in Equations (20) and (21) can be calculated similarly, where $w_{L1} = 0.3063$, $w_{L2} = 0.6937$, $w_{M1} = 0.1627$ and $w_{M2} = 0.8373$.

$$\text{sim}(G^\alpha, G^\beta) = 0.5972 \text{sim}(G_L^\alpha, G_L^\beta) + 0.2857 \text{sim}(G_M^\alpha, G_M^\beta) + 0.1171 \text{sim}(G_S^\alpha, G_S^\beta) \quad (23)$$

In this paper, we used the maximum value, mean, standard deviation and Pearson correlation coefficient of the difference between the similarity of the comparison experiment algorithm and the similarity of the manual evaluation to compute the performance test experiments. Furthermore, we assessed the correlation between the similarities derived from various algorithms and the results obtained from the manual evaluation. The Pearson correlation coefficient was used to characterize the correlation between two variables, with values ranging from 0 to 1, and a higher value meaning a greater correlation between the two. The Pearson correlation coefficient was used as a reference indicator and the formula is shown in Equation (24).

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \cdot \sqrt{\sum (y - \bar{y})^2}} \quad (24)$$

where $x \in X$, $y \in Y$, \bar{x} and \bar{y} are means, the numerator is the covariance of X and Y , and the denominator is the product of the standard deviation of X and Y .

We used precision, recall and F1 scores to perform accuracy testing experiments on groups of storm surge disaster levels and compared the performance of this aggregation method with 10 benchmark methods.

5.2. Analysis of Experimental Results

Based on the fused case structure, the core of the similarity calculation of storm surge disaster cases had two points, which were the similarity calculation between cases by fusing multiaspect and multifactor information and the similarity calculation between cases by fusing the multihierarchy information of case structure. Therefore, the focus in the experimental design phase was to verify whether adding different aspects, different factors and multihierarchy information would bring gains. We experimentally tested and verified the textual-information-based similarity calculation method (sim1), data-information-based similarity calculation method (sim2), case-scenario-based similarity calculation method (sim3), disaster-damage-based similarity calculation method (sim4), case-structure-based similarity calculation method (sim5) and the similarity calculation methods based on case scenario + disaster damage (sim6), case scenario + case structure (sim7) and disaster damage + case structure (sim8). Furthermore, we designed two neural network learning methods, the Word2vec [63] method (sim10), which can model document information, and the graph neural network (GNN) [53] method (sim11). These two methods were compared with the aggregation method (sim9) from this paper, and the comparison results are shown in Tables 3 and 4.

5.2.1. Multifactor Analysis

From the experimental results shown in Table 3, it can be seen that the Pearson correlation coefficients of the aggregation method *sim9* proposed in this paper are all higher than those of the other 10 methods. The similarity calculation methods *sim1* based on textual information and *sim2* based on data information are all single-factor methods, and the similarity calculation methods *sim3* based on case scenario, *sim4* based on disaster damage and *sim5* based on case structure are all multi-factor methods that only consider one side. It is observed that the Pearson correlation coefficients of the single-factor methods are lower than those of the other methods, indicating that the multifactor methods can represent storm disaster cases better than the single-factor methods.

Table 3. Experimental results of similarity performance test in storm surge disaster cases

Calculation Method of Similarity	Expression	Max Error	Mean Error	Standard Deviation	Pearson Correlation Coefficient
Textual information	<i>sim1</i>	0.7861	0.2466	0.1804	0.7410
Data information	<i>sim2</i>	0.1827	0.3447	0.2785	0.9046
Case scenario	<i>sim3</i>	0.2370	0.6120	0.4995	0.9323
Disaster damage	<i>sim4</i>	0.3436	0.6321	0.5059	0.9352
Case structure	<i>sim5</i>	0.5436	0.1857	0.1320	0.5623
Case scenario + disaster damage	<i>sim6</i>	0.2162	0.3429	0.3131	0.9829
Case scenario + case structure	<i>sim7</i>	0.4832	0.1207	0.8550	0.7945
Disaster damage + case structure	<i>sim8</i>	0.1666	0.4882	0.3119	0.9659
Aggregation method (ours)	<i>sim9</i>	0.1760	0.4528	0.3290	0.9886
Word2vec	<i>sim10</i>	0.5233	0.1564	0.1157	0.6170
GNN	<i>sim11</i>	0.4843	0.1610	0.1362	0.5901

sim1 has a much lower Pearson correlation coefficient than *sim2* and the main reason is that the information extracted from *sim1* is the conceptual information about the disaster-bearing bodies and their attributes in the document. It does not directly reveal the severity of the disaster and only shows that those disaster-bearing bodies have been damaged by the storm surge. On the contrary, the data information in *sim2* can directly show the severity of the disaster and *sim2* is more effective than *sim1*, for *sim1* only uses textual information. However, since the data information in the data-based similarity calculation method does not reflect the specific disaster-bearing body object, it leads to better results when the two are combined in the disaster-damage-based similarity calculation method.

From Tables 3 and 4, we can see that the similarity experimental results of *sim5* are worse than those of other methods. The reason is that *sim5* uses structural information, which is calculated by the editing distance. It cannot verify whether the disaster events in two cases are similar or not by directly using the editing distance. In actual storm surge disaster cases, there are many nodes with different information contents. For example, 139 people died in case a and 2 people died in case B represent one node; the editing operation of the node operation is similar, but the similarity between the substantial information contents of the two nodes is very low. Therefore, the combination of structural and textual information in the *sim7* similarity experiment could not lead to a significant improvement in the effect, as it lacks information about the severity of the storm surge disaster cases. The results of the experiments are consistent with human cognitive logic: when evaluating the similarity manually, experts will first judge the similarity of keywords and data in the text between two cases before judging the similarity between the scenarios and structures.

From the accuracy test's experimental results in Table 4, we can see that the similarity experimental results of red level and blue level have higher accuracy, where the textual information factor and data information factor can bring obvious accuracy improvement for the methods. Different disaster levels are classified differently, and the experimental results

show that the classification method does not distinguish the orange level from the yellow level significantly, and the effect is worse compared to the accuracy performance of the red level and the blue level. This is due to the special nature of the dichotomous classification problem, and adding too many factors to the method does not bring an improvement in the accuracy rate.

The experimental comparison shows that the improvement of similarity experimental results of aggregation method *sim9* is not caused by a single factor individually, but the combination of multiple factors leading to the resulting enhancement. The textual information factor and the data information factor are the keys to significantly improving the experimental results. The role of the textual information factor and the data information factor is to provide the basic feature information for the method to ensure that the method can successfully learn the similarity-related information. While the case scenario information and case structure information further enhance the aggregation ability, the basic feature information is sufficient for the method to distinguish the similarity between different data. We can see that when multiple factors are available at the same time, the method *sim9* obtains the globally optimal result. Furthermore, the result will fall back when any factor has been removed, so we believe that multiple factors are essential.

Table 4. Experimental results of similarity accuracy test in storm surge disaster cases.

Calculation Method of Similarity	Expression	Red			Orange			Yellow			Blue		
		P.	R.	F.	P.	R.	F.	P.	R.	F.	P.	R.	F.
Textual information	<i>sim1</i>	71.97	54.52	62.04	63.38	51.73	56.97	68.45	57.59	62.55	73.80	63.63	68.34
Data information	<i>sim2</i>	77.21	60.46	67.82	72.24	54.78	62.31	75.11	60.48	67.01	78.32	62.85	69.74
Case scenario	<i>sim3</i>	74.71	63.80	68.83	66.16	63.39	64.75	78.10	63.92	70.30	77.65	65.57	71.10
Disaster damage	<i>sim4</i>	81.61	67.31	73.77	73.33	66.89	69.96	76.98	65.55	70.81	75.27	67.92	71.41
Case structure	<i>sim5</i>	57.34	39.24	46.59	56.61	38.48	45.82	61.31	43.62	50.97	64.94	59.83	62.28
Case scenario + disaster damage	<i>sim6</i>	91.79	71.02	80.08	81.83	71.01	76.04	82.58	73.88	77.99	89.46	73.01	80.40
Case scenario + case structure	<i>sim7</i>	71.70	58.95	64.70	70.68	59.52	64.62	70.93	52.85	60.57	78.98	62.80	69.97
Disaster damage + case structure	<i>sim8</i>	90.73	69.39	78.64	80.03	69.27	74.26	84.84	70.12	76.78	86.87	69.53	77.24
Aggregation method (ours)	<i>sim9</i>	92.54	79.42	85.48	82.53	74.64	78.39	85.41	77.78	81.42	90.94	76.20	82.92
Word2vec	<i>sim10</i>	86.71	64.21	73.78	80.32	60.17	68.80	76.68	61.24	68.10	77.35	63.18	69.55
GNN	<i>sim11</i>	68.72	44.61	54.10	59.41	40.94	48.48	66.54	45.10	53.76	73.24	57.34	64.32

5.2.2. Neural Network Analysis

From Tables 3 and 4, it can be seen that *sim11*, a method based on a graphical neural network, has poor experimental results in calculating the similarity between cases. Comparing the results of *sim11* and *sim5*, it can be seen that the graphical neural network can indeed learn the structural information in the cases, and the reasons for its poor accuracy are similar to the reasons of *sim5*. The case structure information extracted by *sim11* can only show the hierarchical relation between each disaster event in the case. This hierarchical relation can only show the geographical location of each disaster damage event and the relation between them, and it is hard to learn the damage data, casualty data and other data information factors. The previous experiments through multifactor learning and analysis showed that only when the text information, data information, scenario information and case structure information were combined in multiaspect, multifactor and multihierarchy ways could we get better results. This explains why the *sim11* experiments were less effective.

The similarity experimental results in the neural network method *sim10* based on word vector are much better than those of the method *sim11*. We think the reason is that the features of word vector Word2vec are made from the extracted textual information, which is the feature word vector obtained by a secondary calculation, and it is not the textual feature of the original document. Word vector Word2vec shows more features of the text itself; if the feature space is large enough, the model can learn certain knowledge based on these features to make predictions. In this paper, each type of feature had

a separate dimension in the feature matrix, and each word in the word vector approach was a node with 200-dimensional features. When we pooled the word vectors, the information contained in the word vectors themselves was masked, and the pooled word vectors of a document were not able to fully characterize the document. The information in the documents that played a key role in the similarity calculation between cases was masked a lot by pooling, which led to a worse similarity experiment result of Word2vec than the aggregation method *sim9* with multifactor learning.

6. Example Analysis

Taking storm surge disasters no. 1 and no. 2 as examples, we show the process of similarity calculation between the two disasters. The relevant information in the two cases was extracted automatically from case text by the natural language processing technology, which included the label information of V_P , V_I , V_M , V_L and the structure information of the cases. Figure 7 shows the case of storm surge disasters no. 1 and no. 2.

The label information V_P of the storm surge disaster cases is shown in Tables 5 and 6, each node label N_P of a case attribute contains two attribute values of case attribute type Rep and case attribute data Cop . The label information V_I of the storm surge disaster cases is shown in Tables 7 and 8, the label N_I of case damage node in each case contains three attribute values of case disaster-bearing body Ha_I , case disaster-bearing body attribute Pr_I and case disaster-bearing body data Da_I . The label information V_M and V_L of the storm surge disaster cases are shown in Tables 9 and 10, the label of disaster damage node N_M in each case contains three attribute values of disaster-damage-bearing body Ha_M , disaster-damage-bearing body attribute Pr_M and disaster-damage-bearing body data Da_M , and the geographical location node label N_L contains the attribute value of geographical location Lo_M .

Based on the above results, the “case scenario” of storm surge disaster no. 1 contains 11 nodes of V_P and V_I , its “disaster damage” contains 12 nodes of V_M , and the “case structure” has four layers; the “case scenario” of storm surge disaster no. 2 contains 10 nodes of V_P and V_I , its “disaster damage” contains 16 nodes of V_M , and the “case structure” has three layers. The similarity in “case scenario”, “disaster damage” and “case structure” in both cases was calculated from the contents shown in Tables 5–10, and the aggregated similarity multifactor was 0.8922, using Equation (23).

Table 5. Label information V_P of storm surge disaster case no. 1.

Case Attribute	Rep	Cop
N_{P_1}	Disaster level	Yellow
N_{P_2}	Landing wind speed	35 m/s
N_{P_3}	Central air pressure	970 hPa
N_{P_4}	Coastal water increase	100 cm

Table 6. Label information V_P of storm surge disaster case no. 2.

Case Attribute	Rep	Cop
N_{P_1}	Disaster level	Yellow
N_{P_2}	Landing wind speed	23 m/s
N_{P_3}	Central air pressure	990 hPa
N_{P_4}	Coastal water increase	132 cm

Table 7. Label information V_I of storm surge disaster case no. 1.

Case Damage	Ha_I	Pr_I	Da_I
N_{I_1}	People	Dead	Four people
N_{I_2}	Population	Affected	42,900 people
N_{I_3}	House	Loss	3280 rooms
N_{I_4}	Economy	Loss	CNY 1.2106 billion
N_{I_5}	Crops	Affected	203,180 hectares
N_{I_6}	Old Town Station	Exceeds the warning tide value	2 cm
N_{I_7}	Quarry Bay	Exceeds the warning tide value	69 cm

Table 8. Label information V_I of storm surge disaster case no. 2.

Case Damage	Ha_I	Pr_I	Da_I
N_{I_1}	People	Dead	Twelve people
N_{I_2}	Population	Affected	168,330 people
N_{I_3}	House	Loss	321 rooms
N_{I_4}	Economy	Loss	CNY 0.230253 billion
N_{I_5}	Crops	Affected	47,333 hectares
N_{I_6}	Xiuying Station	Exceeds the warning tide value	30 cm

Table 9. Label information V_M and V_L of storm surge disaster case no. 1.

Disaster Damage	Ha_M	Pr_M	Da_M	Lo_L
N_{M_1}	People	Affected	3,667,800 people	Guangdong Province
N_{M_2}	Population	Emergency transfer	42,900 people	Guangdong Province
N_{M_3}	House	Collapse	3280 rooms	Guangdong Province
N_{M_4}	Aquaculture	Affected	48,930 hectares	Guangdong Province
N_{M_5}	Fishing boat	Damage	More than 120 boats	Huilai County, Jieyang City, Guangdong Province
N_{M_6}	Fishing boat	Sank	Eight boats	Huilai County, Jieyang City, Guangdong Province
N_{M_7}	Direct economy	Loss	CNY 1.2106 billion	Guangdong Province
N_{M_8}	Direct economy	Loss	CNY 0.557 billion	Shanwei City, Guangdong Province
N_{M_9}	Direct economy	Loss	CNY 0.417 billion	Jieyang City, Guangdong Province
N_{M_10}	Direct economy	Loss	CNY 0.1215 billion	Chaozhou City, Guangdong Province
N_{M_11}	People	Hurt	Five people	Hong Kong
N_{M_12}	Boat	Damage	Two boats	Hong Kong

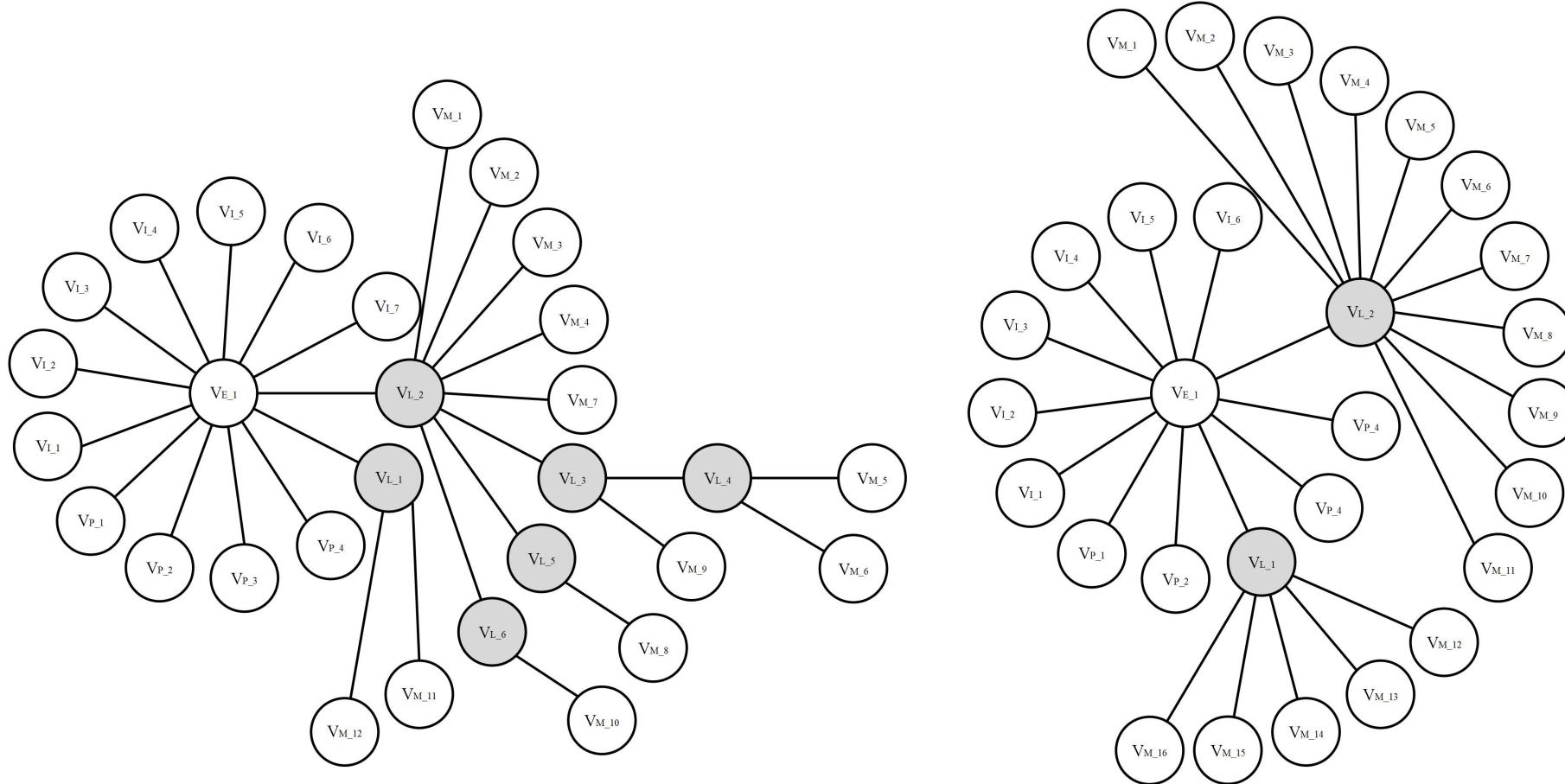


Figure 7. Storm surge disaster cases no. 1 and no. 2.

Table 10. Label information V_M and V_L of storm surge disaster case no. 2.

Disaster Damage	Ha_M	Pr_M	Da_M	Lo_L
N_{M_1}	People	Dead	Four people	Hainan Province
N_{M_2}	Population	Affected	167,590 people	Hainan Province
N_{M_3}	House	Damage	321 rooms	Hainan Province
N_{M_4}	Boat	Damage	34 boats	Hainan Province
N_{M_5}	Farmland	Affected	45,700 hectares	Hainan Province
N_{M_6}	Cage culture	Damage	155 cages	Hainan Province
N_{M_7}	Aquaculture	Loss	940 tons	Hainan Province
N_{M_8}	Breakwater	Damage	246 m	Hainan Province
N_{M_9}	Bank protection	Damage	48 m	Hainan Province
N_{M_10}	Road	Damage	900 m	Hainan Province
N_{M_11}	Direct economy	Loss	CNY 0.23 billion	Hainan Province
N_{M_12}	People	Dead	8 people	Guangxi Zhuang Autonomous Region
N_{M_13}	Population	Affected	380 people	Guangxi Zhuang Autonomous Region
N_{M_14}	Fishing boat	Damage	2 boats	Guangxi Zhuang Autonomous Region
N_{M_15}	Farmland	Flooded	1633 hectares	Guangxi Zhuang Autonomous Region
N_{M_16}	Direct economy	Loss	CNY 0.03 billion	Guangxi Zhuang Autonomous Region

7. Summary

Cases of sudden natural disasters contain textual information, data information and geographic location information, which are difficult to read and understand by computers. In this paper, we took a holistic approach to consider the comprehensive problems associated with sudden natural disaster disasters and calculate the similarity between sudden natural disaster cases by aggregation methods in multihierarchy, multiaspect and multifactor ways. Furthermore, we comprehensively considered the text content and the geographical distribution in cases and solved the one-sidedness and limitation of similarity calculation between cases with case text content and case hierarchy. Meanwhile, we converted the text information of sudden natural disaster cases into case label information and converted the disaster damage events into a case structure based on geographic location information. This can help calculate the similarity of sudden natural disasters and provide corresponding automated processing methods and reference solutions for natural disaster emergency managers. The results were verified by storm surge disaster domain experiments, which showed that the aggregated multihierarchy, multiaspect, multifactor method worked better compared to other single-factor methods. The case similarity calculated by the aggregated method was closer to the manually estimated similarity than other methods.

The case similarity calculation method proposed in this paper still needs to be further refined in the future; the sudden natural disaster domain database needs to be continuously expanded and more different features need to be found. We will improve the case similarity calculation by optimizing the sudden natural disaster domain ontology, learning matrix and weights, to provide a more intelligent and automated disaster analysis and disaster decision-making.

Author Contributions: Conceptualization, Q.Z., H.Z. and C.C.; methodology, C.C. and S.W.; software, C.C. and S.W.; validation, H.Z. and C.C.; formal analysis, Q.Z. and C.C.; investigation, C.C. and S.W.; resources, Q.Z. and C.C.; data curation, C.C.; writing—original draft preparation, C.C.; writing—review and editing, Q.Z., H.Z. and C.C.; visualization, C.C.; supervision, Q.Z.; project administration, H.Z. and C.C.; funding acquisition, Q.Z. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number U1931207 and 61702306; Science and Technology Development Fund of Shandong Province of China, grant number ZR202102230289, ZR202102250695 and ZR2019LZH001; Humanities and Social Science Research Project of the Ministry of Education, grant number 18YJAZH017; Taishan Scholar Program of Shandong Province, grant number ts20190936 and tsqn201909109; Shandong Chongqing Science

and technology cooperation project, grant number cstc2020jscx-lyjsAX0008; Science and Technology Development Fund of Qingdao, grant number 21-1-5-zlyj-1-zc; Shandong University of Science and Technology Research Fund, grant number 2019KJN024 and 2015TDJH102; Natural Science Foundation of Shandong Province of China, grant number ZR2021MG038.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Huang, D.; Wang, S.; Liu, Z. a systematic review of prediction methods for emergency management. *Int. J. Disaster Risk Reduct.* **2021**, *62*, 102412. [[CrossRef](#)]
- Wang, D.; Wan, K.; Ma, W. Emergency decision-making model of environmental emergencies based on case-based reasoning method. *J. Environ. Manag.* **2020**, *262*, 110382. [[CrossRef](#)]
- Evelpidou, N.; Tzouxanioti, M.; Gavalas, T.; Spyrou, E.; Saitis, G.; Petropoulos, A.; Karkani, A. Assessment of Fire Effects on Surface Runoff Erosion Susceptibility: The Case of the Summer 2021 Forest Fires in Greece. *Land* **2021**, *11*, 21. [[CrossRef](#)]
- Liu, H.; Luo, N.; Zhao, Q. Research on the Construction of Typhoon Disaster Chain Based on Chinese Web Corpus. *J. Mar. Sci. Eng.* **2022**, *10*, 44. [[CrossRef](#)]
- State Oceanic Administration of China. Bulletin of China Marine Disaster. 2021. Available online: <http://www.nmdis.org.cn/hygb/zghyzhgb/2021nzghyzhgb/> (accessed on 10 May 2022).
- Wang, K.; Yang, Y.; Reniers, G.; Li, J.; Huang, Q. Predicting the spatial distribution of direct economic losses from typhoon storm surge disasters using case-based reasoning. *Int. J. Disaster Risk Reduct.* **2022**, *68*, 102704. [[CrossRef](#)]
- Pan, A. Study on the decision-making behavior of evacuation for coastal residents under typhoon storm surge disaster. *Int. J. Disaster Risk Reduct.* **2020**, *45*, 101522. [[CrossRef](#)]
- Gao, J.; Xu, Z.; Liang, Z.; Liao, H. Expected consistency-based emergency decision making with incomplete probabilistic linguistic preference relations. *Knowl.-Based Syst.* **2019**, *176*, 15–28. [[CrossRef](#)]
- Helderop, E.; Grubacic, T.H. Streets, storm surge, and the frailty of urban transport systems: A grid-based approach for identifying informal street network connections to facilitate mobility. *Transp. Res. Part D Transp. Environ.* **2019**, *77*, 337–351. [[CrossRef](#)]
- Kou, G.; Ergu, D.; Shi, Y. an integrated expert system for fast disaster assessment. *Comput. Oper. Res.* **2014**, *42*, 95–107. [[CrossRef](#)]
- Chen, N.; Liu, W.; Bai, R.; Chen, A. Application of computational intelligence technologies in emergency management: A literature review. *Artif. Intell. Rev.* **2019**, *52*, 2131–2168. [[CrossRef](#)]
- Smith, J.; Yazdanpanah, F.; Thistle, R.; Musharraf, M.; Veitch, B. Capturing expert knowledge to inform decision support technology for marine operations. *J. Mar. Sci. Eng.* **2020**, *8*, 689. [[CrossRef](#)]
- Chen, S.; Yi, J.; Jiang, H.; Zhu, X. Ontology and CBR based automated decision-making method for the disassembly of mechanical products. *Adv. Eng. Inform.* **2016**, *30*, 564–584. [[CrossRef](#)]
- Chen, C.; Yang, Y.; Wang, M.; Zhang, X. Characterization and evolution of emergency scenarios using hybrid Petri net. *Process. Saf. Environ. Prot.* **2018**, *114*, 133–142. [[CrossRef](#)]
- De Nicola, A.; Melchiori, M.; Villani, M.L. Creative design of emergency management scenarios driven by semantics: An application to smart cities. *Inf. Syst.* **2019**, *81*, 21–48. [[CrossRef](#)]
- Bannour, W.; Maalel, A.; Ben Ghezala, H.H. Emergency Management Case-Based Reasoning Systems: A Survey of Recent Developments. *J. Exp. Theor. Artif. Intell.* **2021**, *1*–24. [[CrossRef](#)]
- Hadj-Mabrouk, H. Application of Case-Based Reasoning to the safety assessment of critical software used in rail transport. *Saf. Sci.* **2020**, *131*, 104928. [[CrossRef](#)]
- Song, B.; Yan, W.; Zhang, T. Cross-border e-commerce commodity risk assessment using text mining and fuzzy rule-based reasoning. *Adv. Eng. Inform.* **2019**, *40*, 69–80. [[CrossRef](#)]
- Okudan, O.; Budayan, C.; Dikmen, I. a knowledge-based risk management tool for construction projects using case-based reasoning. *Expert Syst. Appl.* **2021**, *173*, 114776. [[CrossRef](#)]
- Khosravani, M.R.; Nasiri, S. Injection molding manufacturing process: Review of case-based reasoning applications. *J. Intell. Manuf.* **2020**, *31*, 847–864. [[CrossRef](#)]
- Bentaiba-Lagrid, M.B.; Bouzar-Benlabiod, L.; Rubin, S.H.; Bouabana-Tebibel, T.; Hanini, M.R. a case-based reasoning system for supervised classification problems in the medical field. *Expert Syst. Appl.* **2020**, *150*, 113335. [[CrossRef](#)]
- El-Sappagh, S.; Elmogy, M.; Riad, A. a fuzzy-ontology-oriented case-based reasoning framework for semantic diabetes diagnosis. *Artif. Intell. Med.* **2015**, *65*, 179–208. [[CrossRef](#)]
- Perez, B.; Lang, C.; Henriet, J.; Philippe, L.; Auber, F. Risk prediction in surgery using case-based reasoning and agent-based modelization. *Comput. Biol. Med.* **2021**, *128*, 104040. [[CrossRef](#)] [[PubMed](#)]
- Feng, Y.; Xiang-Yang, L. Improving emergency response to cascading disasters: Applying case-based reasoning towards urban critical infrastructure. *Int. J. Disaster Risk Reduct.* **2018**, *30*, 244–256. [[CrossRef](#)]

25. Qingsong, Z.; Junyi, D.; Yu, G.; Peng, L.; Kewei, Y. a scenario construction and similarity measurement method for navy combat search and rescue. *J. Syst. Eng. Electron.* **2020**, *31*, 957–968. [[CrossRef](#)]
26. Zhang, Y.J. Research on emergency aid decision-making model for environmental emergency based on case-based reasoning. In *Proceedings of the Applied Mechanics and Materials*; Trans Tech Publ.: Bach, Switzerland, 2014; Volume 675, pp. 206–212. [[CrossRef](#)]
27. Yang, Y.; Ping, Y. an ontology-based semantic similarity computation model. In Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (Bigcomp), Shanghai, China, 15–17 January 2018; pp. 561–564.
28. Yu, X.; Li, C.; Zhao, W.X.; Chen, H. a novel case adaptation method based on differential evolution algorithm for disaster emergency. *Appl. Soft Comput.* **2020**, *92*, 106306. [[CrossRef](#)]
29. Chen, Y.; Zhang, J.; Zhou, A.; Yin, B. Modeling and analysis of mining subsidence disaster chains based on stochastic Petri nets. *Nat. Hazards* **2018**, *92*, 19–41. [[CrossRef](#)]
30. Liu, Y.; Fan, Z.P.; Yuan, Y.; Li, H. a FTA-based method for risk decision-making in emergency response. *Comput. Oper. Res.* **2014**, *42*, 49–57. [[CrossRef](#)]
31. Jiang, B.; Chen, T.; Yuan, H.; Fan, W. Emergency decision-making method for rainstorm disasters based on spatiotemporal scenario analyses. *Tsinghua Sci. Technol.* **2022**, *62*, 52–59.
32. Chandrasekaran, D.; Mago, V. Evolution of semantic similarity—A survey. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–37. [[CrossRef](#)]
33. Harispe, S.; Ranwez, S.; Janaqi, S.; Montmain, J. Semantic similarity from natural language and ontology analysis. *Synth. Lect. Hum. Lang. Technol.* **2015**, *8*, 1–254.
34. Lastra-Díaz, J.J.; Goikoetxea, J.; Taieb, M.A.H.; García-Serrano, A.; Aouicha, M.B.; Agirre, E. a reproducible survey on word embeddings and ontology-based methods for word similarity: Linear combinations outperform the state of the art. *Eng. Appl. Artif. Intell.* **2019**, *85*, 645–665. [[CrossRef](#)]
35. Ma, G.; Ahmed, N.K.; Willke, T.L.; Yu, P.S. Deep graph similarity learning: A survey. *Data Min. Knowl. Discov.* **2021**, *35*, 688–725. [[CrossRef](#)]
36. Frisoni, G.; Moro, G.; Carlssare, G.; Carbonaro, A. Unsupervised event graph representation and similarity learning on biomedical literature. *Sensors* **2021**, *22*, 3. [[CrossRef](#)]
37. Lee, C.H.; Wang, Y.H.; Trappey, A.J. Ontology-based reasoning for the intelligent handling of customer complaints. *Comput. Ind. Eng.* **2015**, *84*, 144–155. [[CrossRef](#)]
38. Rada, R.; Mili, H.; Bicknell, E.; Blettner, M. Development and application of a metric on semantic nets. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 17–30. [[CrossRef](#)]
39. Aouicha, M.B.; Taieb, M.A.H. Computing semantic similarity between biomedical concepts using new information content approach. *J. Biomed. Inform.* **2016**, *59*, 258–275. [[CrossRef](#)]
40. Likavec, S.; Lombardi, I.; Cena, F. Sigmoid similarity—a new feature-based similarity measure. *Inf. Sci.* **2019**, *481*, 203–218. [[CrossRef](#)]
41. Lu, J.; Xue, X.; Lin, G.; Huang, Y. a new ontology meta-matching technique with a hybrid semantic similarity measure. In *Advances in Intelligent Information Hiding and Multimedia Signal Processing*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 37–45.
42. Pilehvar, M.T.; Navigli, R. From senses to texts: An all-in-one graph-based approach for measuring semantic similarity. *Artif. Intell.* **2015**, *228*, 95–128. [[CrossRef](#)]
43. Zhou, P.; El-Gohary, N. Semantic information alignment of BIMs to computer-interpretable regulations using ontologies and deep learning. *Adv. Eng. Inform.* **2021**, *48*, 101239. [[CrossRef](#)]
44. Zeng, Q.; Lu, F.; Liu, C.; Duan, H.; Zhou, C. Modeling and verification for cross-department collaborative business processes using extended Petri nets. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *45*, 349–362. [[CrossRef](#)]
45. Zeng, Q.; Liu, C.; Duan, H. Resource conflict detection and removal strategy for nondeterministic emergency response processes using Petri nets. *Enterp. Inf. Syst.* **2016**, *10*, 729–750. [[CrossRef](#)]
46. Guo, W.; Zeng, Q.; Duan, H.; Yuan, G.; Ni, W.; Liu, C. Automatic extraction of emergency response process models from Chinese plans. *IEEE Access* **2018**, *6*, 74104–74119. [[CrossRef](#)]
47. Li, S.; Chen, S.; Liu, Y. a method of emergent event evolution reasoning based on ontology cluster and Bayesian network. *IEEE Access* **2019**, *7*, 15230–15238. [[CrossRef](#)]
48. Chen, X.; Huo, H.; Huan, J.; Vitter, J.S. an efficient algorithm for graph edit distance computation. *Knowl.-Based Syst.* **2019**, *163*, 762–775. [[CrossRef](#)]
49. Wallis, W.D.; Shoubridge, P.; Kraetz, M.; Ray, D. Graph distances using graph union. *Pattern Recognit. Lett.* **2001**, *22*, 701–704. [[CrossRef](#)]
50. Dijkman, R.; Dumas, M.; García-Bañuelos, L. Graph matching algorithms for business process model similarity search. In *Proceedings of the International Conference on Business Process Management*, Ulm, Germany, 8–10 September 2009; Springer: Berlin/Heidelberg, Germany, 2009; pp. 48–63.
51. Bai, Y.; Ding, H.; Bian, S.; Chen, T.; Sun, Y.; Wang, W. Simgnn: A neural network approach to fast graph similarity computation. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, Melbourne, VIC, Australia, 11–15 February 2019; pp. 384–392.
52. Hamilton, W.L. Graph representation learning. *Synth. Lect. Artificial Intell. Mach. Learn.* **2020**, *14*, 1–159.
53. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Philip, S.Y. a comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4–24. [[CrossRef](#)]

54. Li, M.; Cao, P. Extended TODIM method for multi-attribute risk decision making problems in emergency response. *Comput. Ind. Eng.* **2019**, *135*, 1286–1293. [[CrossRef](#)]
55. Yu, F.; Dong, J.; Ye, L. *Historical Materials of Storm Surge Disasters in China: 1949–2009*; China Ocean Press: Beijing, China, 2015.
56. Jain, S.; Mehla, S.; Mishra, S. an ontology of natural disasters with exceptions. In Proceedings of the 2016 International Conference System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 25–27 November 2016; pp. 232–237.
57. Joshi, H.; Seker, R.; Bayrak, C.; Ramaswamy, S.; Connelly, J.B. Ontology for disaster mitigation and planning. In Proceedings of the 2007 Summer Computer Simulation Conference, San Diego, CA, USA, 15–18 July 2007; pp. 1–8.
58. Meymandpour, R.; Davis, J.G. a semantic similarity measure for linked data: An information content-based approach. *Knowl.-Based Syst.* **2016**, *109*, 276–293. [[CrossRef](#)]
59. Lin, D. an Information-Theoretic Definition of Similarity. In Proceedings of the Fifteenth International Conference on Machine Learning, Madison, WI, USA, 24–27 July 1998; Volume 98, pp. 296–304.
60. Norouzi, M.; Fleet, D.J.; Salakhutdinov, R.R. Hamming distance metric learning. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Volume 25.
61. Khairuddin, S.H.; Hasan, M.H.; Hashmani, M.A.; Azam, M.H. Generating clustering-based interval fuzzy type-2 triangular and trapezoidal membership functions: A structured literature review. *Symmetry* **2021**, *13*, 239. [[CrossRef](#)]
62. Peng, C.Y.J.; Lee, K.L.; Ingersoll, G.M. an introduction to logistic regression analysis and reporting. *J. Educ. Res.* **2002**, *96*, 3–14. [[CrossRef](#)]
63. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.