Article

An Ontological Metro Accident Case Retrieval Using CBR and NLP

Haitao Wu ¹, Botao Zhong ^{1,*}, Benachir Medjdoub ², Xuejiao Xing ¹ and Li Jiao ¹

¹ School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan 430074, China; haitao_w@hust.edu.cn (H.W.); d201780870@hust.edu.cn (X.X.); jiaoli@hust.edu.cn (L.J.)

² School of Architecture, Design and the Built Environment, Nottingham Trent University, Nottingham NG1 4FQ, UK; benachir.medjdoub@ntu.ac.uk

* Correspondence: dadizhong@hust.edu.cn

Received: 11 June 2020; Accepted: 16 July 2020; Published: 31 July 2020

Abstract: Metro accidents are apt to cause serious consequences, such as casualties or heavy economic loss. Once accidents occur, quick and accurate decision-making is essential to prevent emergent accidents from getting worse, which remains a challenge due to the lack of efficient knowledge representation and retrieval. In this research, an ontological method that integrates case-based reasoning (CBR) and natural language processing (NLP) techniques was proposed for metro accident case retrieval. An ontological model was developed to formalize the representation of metro accident knowledge, and then, the CBR aimed to retrieve similar past cases for supporting decision-making after the accident cases were annotated by the NLP technique. Rule-based reasoning (RBR), as a complementary of CBR, was used to decide the appropriate measures based on those that are recorded in regulations, such as emergency plans. A total of 120 metro accident cases were extracted from the safety monthly reports during metro operations and then built into the case library. The proposed method was tested in MyCBR and evaluated by expert reviews, which had an average precision of 91%.

Keywords: metro accident; ontology; CBR; NLP; accident response

1. Introduction

The metro has recently become a popular means of public transportation. As of 5 September 2016, in China, 44 cities are preparing to construct metro systems and 27 cities are already operating metro systems [1]. However, as more and more metro lines are put into service, catastrophic accidents occasionally occur during the metro operation, such as turnout failures, signal failures, and terrorist attacks. For instance, in the serious Daegu metro accident (18 February 2003), 192 persons died and another 151 were injured [2]. In the metro collision accident in Washington, about 80 persons were injured (22 June 2009) [3]. In the rear-end accident of line ten in Shanghai (27 September 2011), 40 persons were injured and the line operation was interrupted for more than six hours [4].

Metro accidents are apt to cause serious consequences, such as casualties or heavy economic loss, due to complex underground structures, complicated facilities, narrow spaces, and crowded passengers [2,5]. Therefore, this sets a higher demand for accident prevention and response. Once accidents occur, a quick response and accurate decision-making are essential to prevent emergent accidents from getting worse, which is still a challenge to managers in metro operation companies and was the research starting point of this study.

Traditionally, the manager tends to retrieve the related knowledge to support decision-making, such as previous similar accident records, regulations, and emergency plans. Historical cases are

deemed the most effective resources for supporting emergency decision-making as part of a time-limited response strategy plan [6]. It is feasible for managers to obtain support through an intelligent knowledge system that can integrate the valuable knowledge of experts and real cases. Case-based reasoning (CBR) is prevailing as a decision-support approach in construction safety management, which applies the experience of similar past accidents to the new accident. For example, Lu, et al. [7] proposed a CBR-based method to identify safety risks by analyzing the precursors installed on the subway system. Goh and Guo [8] developed a CBR-based online system that can support the design and selection of an active fall protection system. Yu et al. used the CBR method to retrieve similar historical cases to support decision-making in the response to risks connected with urban water supply networks [4]. However, in the metro operation stage, knowledge retrieval is still a huge challenge to the current CBR methods. Two critical issues need to be solved to improve the performance of knowledge retrieval: one is how to effectively represent the domain knowledge and the other is how the knowledge can be retrieved quickly and accurately from the database.

Considering the former obstacle, most knowledge concerning metro accidents is in various unstructured forms (e.g., accident records, expert experience, and emergency plans). For instance, in China, the majority of the accident records and more than 80% of the emergency plans of metro operations are plain-text documents or rigid electronic files. Non-structured or semi-structured knowledge restricts the automation of decision-making progress regarding metro accidents. Ontology is the solution to this obstacle. Ontologies provide a way to formally represent the domain knowledge by defining the concepts, relations, and axioms [9]. Ontologies provide non-ambiguous semantic representation to domain knowledge. However, there are still some gaps in knowledge concerning the combination of ontology technology and CBR for decision-making regarding metro accident responses.

Regarding the second issue, the annotation of past accidents is a critical step for quick CBR-based retrieval from the database. As part of the routine inspection of metro operations, new accident reports continually need to be added into databases where accident records are written in everyday language using different expressions. However, manually annotating accident records is time-consuming and labor-intensive work. A natural language process (NLP), as a technique in computer science, enables the processing of textual documents, such as the automatic analysis of injuries reports [10], nature hazards [11], and information extraction from construction contracts [12]. It can be seen that NLP can assist with quick case retrievals in CBR by facilitating accident annotation. However, there are very few research efforts in the metro operational domain.

This study proposed an improved CBR-based framework by integrating ontology and NLP for efficient knowledge representation and retrieval, which can be used to support decision-making in the metro accident responses. An ontological model was established to formalize the domain knowledge of metro accidents. NLP techniques were used to extract the textual information from collected textual accident reports and semi-automate the annotation of accident cases into a corpus defined in the ontology model. Then, the semantic similarity of accident cases was calculated to find the most similar cases using the CBR. Additionally, to support decision making, RBR (rule-based reasoning) was applied to complement CBR when there was a lack of relevant cases, which can refer to the appropriate measures in guidance documents, such as emergency plans. A case study in a metro company was implemented in the Protégé and MyCBR, to verify the feasibility and effectiveness of the framework. The proposed framework could not only improve the emergent accident handling capabilities of metro lines but it also formalized the domain knowledge through the developed ontology model, which can be used for the further intelligent operation of metros.

2. Literature Review

This study proposed a framework to support decision-making regarding metro accident responses, which integrates the ontology and CBR techniques. Previous works of relevant aspects are reviewed in the following parts.

2.1. Safety Management of Metro Projects

The metro, regarded as a modern means of transport with a lot of benefits, such as large transportation capacity, high speed, and low pollution [13], has quickly been adopted by various countries. However, catastrophic accidents may occasionally occur during the process of rapid urbanization, which would cause huge economic losses. The safety management of metro systems has raised a lot of attention from scholars, which mainly contains risk identification, analysis, and response [14]. For example, Liu, et al. [15] proposed a systematic method by combining exploratory factor analysis (EFA) and a structural equation model (SEM) to identify risk factors in the metro construction. Ding, et al. [16] developed a safety risk identification system (SRIS) for metro construction based on construction drawings, which can be used to identify potential risks during the pre-construction stage. Liu, et al. [17] identified the critical success factors for safety management in the metro construction using expert interviews and interpretive structural models (ISM). Yan, et al. [18] developed a data envelopment analysis (DEA)-based model to evaluate the risk of crashing and trampling accidents in metro systems. Li, et al. [13] established a metro operation hazard network (MOHN), in which 28 hazards and 48 interrelations between hazards were identified from 134 metro accidents. Recently, some studies have adopted digital technologies, such as Building Information Modeling (BIM) and the Internet of things (IoT), to identify and analyze safety risks in metro systems. Ding, et al. [19] proposed a real-time safety early warning system to prevent potential accidents in underground construction, which integrates a fiber Bragg grating (FBG) sensor system and a radio frequency identification (RFID)-based labor tracking system. Li, et al. [20] proposed a BIM-based process for the automated safety risk recognition for underground construction at the pre-construction stage. In the construction field, a computer vision technique was used to identify safety risks by detecting unsafe behavior and statuses [21].

In general, previous studies related to the safety management of metro systems mainly focused on risk identification and analysis. Few studies have focused on how to support decision-making to respond to accidents during metro operations, which is also an essential factor in reducing damage after metro accidents occur.

2.2. Ontology Technology for Safety Management

Safety management is a knowledge-intensive process in the construction industry, where information is scattered across various systems; due to the lack of common representation, the heterogeneous information cannot be shared and re-used effectively between different systems. Ontology, as an essential semantic technique, provides a precise, formal specification of a shared conceptualization of a domain [22]. Compared with the database schema, ontology technology can provide a way to present knowledge with explicit and rich semantics, which support efficient knowledge reasoning and query [23].

Many industries have developed ontologies for efficient knowledge management, such as medicine, computer science, and biology. For instance, using ontology engineering, Koo, et al. [24] proposed a semantic framework for integrating the processing model and dataset in the domain of biorefining. In the construction industry, Ding, et al. [25] established an ontology-based semantic network to produce a construction risk knowledge map, in which the ontology was used to standardize the description of each aspect of risk knowledge and facilitate the knowledge reasoning and retrieval. Wang, et al. [26] built an emergency plan system ontology to promote communication and sharing between different plan systems. Wu, et al. [27] developed a domain ontology to represent knowledge of scenario-based hazard evaluation. Guo and Goh [28] established an ontology model to formalize the knowledge of active fall protection system design, which can facilitate the knowledge reuse and sharing among professional engineers.

2.3. CBR-Based Decision-Making

CBR generally refers to the process of solving new problems based on experience with similar past cases [29]. It is suggested by Schank [30] that the experiences and cases can be reused to facilitate decision-making when a manager faces similar problems. CBR is memory-based and stimulates the human thinking process, which has been always applied in problem-solving fields [31]. As a decision support tool, CBR is well suited to construction safety management considering the construction industry's large amount of historical experiences. For example, Guo, et al. [32] proposed a CBR system to provide design support for an injection mold design. Xie, et al. [33] proposed a CBR system for a hydro-generator design and developed a case base library in this system to facilitate the design process. Yu, et al. [6] used the CBR method to analyze the response to risks connected with the urban water supply network. Additionally, rule-based reasoning (RBR) is usually added as the complementary method of CBR in the decision-making process, considering that CBR may fail when no relevant case is available [34]. Ferrara and Baumgartner [35] discussed the corresponding advantages and disadvantages of CBR and RBR, and pointed out that their integration is useful. Shi and Barnden [36] presented a hybrid approach combining CBR with RBR for diagnosing multiple faults. Goh and Guo [8] presented a web-based CBR–RBR system to support the design of an active fall protection system.

Typically, CBR contains four main processes: retrieve, reuse, revise, and retain, where retrieve is the most important process in any CBR system [7]. Finding an effective method for case representation ensures that domain knowledge can be acquired accurately and easily, thus laying a good foundation for efficient retrieval [32]. However, data regarding metro accident cases are usually stored in plain-text documents or rigid electronic files that are in unstructured formats, such as PDF files, Word processing files, or HTML pages. The non-structural data do not follow a machine-readable format, thereby leading to a poor semantic understandability of traditional CBR [32]. One potential solution to this problem is the development of a semantic framework based on ontologies. Ontology technology used by a CBR system has several functions, such as representing cases and facilitating similarity calculations. Tung, et al. [37] combined CBR, RBR, and ontology to develop a solution retrieval system for expert searches and problem diagnosis. Bouhana, et al. [38] proposed an ontology-based CBR information retrieval system that can improve the accuracy of case retrieval and facilitate the search process. Xu, et al. [39] developed a knowledge model based on ontology and then used CBR and RBR to support automated decision-making for the disassembly of mechanical products. Maalel, et al. [40] applied an ontology-based CBR approach to developing a system that can help to build and operate historical railroad accidents.

Previous studies paved the way to recognize the advantage of the combination of CBR and ontology techniques in the construction industry. However, few efforts have been conducted to build an ontological CBR framework to facilitate knowledge representation and retrieval for decision-making during metro operations.

Therefore, by combining ontology with CBR, a research framework for supporting decision-making in the metro accident response was constructed, and the feasibility of the framework was verified using a metro case study. The main contributions of this study are as follows: first, an ontology model was developed for comprehensively formalizing the domain knowledge and thus providing a good foundation for metro accident retrieval. Second, we used the NLP techniques to semi-automate the annotation of metro accident cases into keywords defined in the ontology model for the following similarity calculation. Third, the Semantic Web Rule Language (SWRL) was used to represent rules for reasoning through the response measures recorded in regulations, such as emergency plans for metro operations, which can support decision-making when there are no similar cases in the case library.

3. The Overall Ontological CBR Framework for Metro Accident Responses

The proposed ontological CBR method aimed to improve the performance of knowledge retrieval for supporting decision-making during metro operations. Additionally, the proposed method can facilitate the development of an online knowledge system for metro accident responses. Figure 1 shows

the overall structure of the system, which consists of four layers: information acquisition, ontology development, semantic processing, and application service. In the information acquisition layer, heterogeneous information provided by different data sources, including regulations, emergency plans, and accident records, are collected and usually stored in several types of forms. For the information acquisition layer, specific ontologies are developed to describe the knowledge and information of the different domains. NLP techniques were used to achieve the semi-automated annotation of metro accidents. For the semantic processing layer, rule containers, reasoning engines, and an annotated case base were created to support the management and application of the ontology-described knowledge. In the application service layer, multiple service applications, including knowledge sharing and retrieval, case-based reasoning, and rule-based reasoning, were provided through the programmable interfaces for users. The detailed knowledge retrieval process of the proposed method is shown in Figure 2.

Figure 1. Overall architecture of the ontological case-based reasoning (CBR) system (adapted from Tao, et al. [41]). RBR: Rule-based reasoning.

In the proposed method, the domain knowledge of metro operational accidents (e.g., basic concepts, relations, rules, etc.) was identified. Then, the ontology model was established to semantically represent the domain knowledge, in which the concepts were represented by classes and individuals, while the relations between different concepts were represented by properties. The complex knowledge reasoning could be achieved through the semantic rules in the ontological model. By combining NLP techniques and the ontology model, the specific knowledge stored in historical accident records was processed and semantically annotated with the concepts in the ontology. This enabled the accident reports to be organized in the structure of the given ontology model, semantically retrieved, and reasoned via their annotation information. Thus, based on the semantic representation, these annotated cases could be understood by computational reasoning, which supported the decision-making process by calculating the similarities between accident cases.

At last, as shown in Figure 2, the knowledge base for the metro accident response was developed by integrating the fact base, rule base, and case base. The integrated knowledge base could support further rule-based reasoning (RBR) and case-based reasoning (CBR).

Figure 2. Workflow of knowledge retrieval for supporting decision-making.

3.1. Knowledge Source for Metro Accidents

Knowledge related to metro accident responses is contained in the various documents, such as regulations, emergency plans, and accident records. These documents were selected as the knowledge source that can be divided into general knowledge (from regulations, emergency plans, and decision rules of the metro enterprise) and specific case knowledge (from historical accident records). The related knowledge supported the development of the ontology model. The related documents were introduced as follows:

- (1) Regulations on metro operational accidents: Chinese government departments have issued regulations related to metro operations, such as the Regulations for the Operation Management of Urban Rail Transit (GB/T 30012-2013), Standard for the Safety Assessment of Existing Metro (GB/T 50438-2007), and Approaches for Operation Management of Urban Rail Transit. Other relevant regulations provide professional terms related to metros, such as the Code for Design of Metro (GB 50157-2013) and Technical Code of Urban Rail Transit (GB 50490-2009).
- (2) Metro emergency plans: Chinese government departments and metro companies have issued metro emergency plans that cover all possible accidents during operations. These plans serve as manuals that describe procedures for managing all types of emergencies [42], such as the Emergency Plan for National Urban Rail Transit and the Emergency Plan for Wuhan Metro Operation.
- (3) Decision rules of metro companies: Metro companies have their own decision rules, such as the Operating Accident Processing Rules of the Wuhan Metro Company, which is based on the actual situation and managerial demands.
- (4) Accident records: The historical accident records contain the operation-checking records and problem-solving records. The operation-checking records include the contents and results of operational checks while the problem-solving records contain the solution and the results of the problem.

3.2. Knowledge Representation in the Ontology Model

In terms of knowledge representation, the proposed method aimed to comprehensively formalize the domain knowledge and thus provide a good foundation for knowledge retrieval. An ontology

model was developed to attain this goal. The ontology provides the common vocabulary/concepts and relationships for the accident knowledge and their context description. The concepts and relations extracted from the domain knowledge were used to build the ontological model.

3.2.1. Taxonomy of Metro Accidents

The taxonomy of the basic concepts of metro accidents is a foundational step in developing the ontology model, which means that all concepts are hierarchically divided into categories or sub-categories. Studies of applying taxonomy for analyzing the specification information can be found in the literature. For example, Anumba, et al. [43] proposed a synthesized taxonomy development method for constructing contractual semantics, which can be utilized to support ontology modeling. The taxonomy of metro accident concepts was determined based on related regulations, emergency plans, decision rules, and accident records, which was used to facilitate the building of an ontological model of metro accidents. The majority of accidents during metro operations were represented by equipment failures, such as signal system failures and metro vehicle failures. The accidents of signal system failures were selected as specific objects for case retrieval in this paper. For example, Table 1 shows the taxonomy of a specific accident during metro operations.

Table 1. The taxonomy of a metro signal system failure.

Category	Detailed Problem of Metro Signal System Failure		
Signal system infrastructure failure	System management center (SMC) equipment failure	Vehicle control center (VCC) equipment failure	Station controller (STC) equipment failure
	Automatic transfer switch (ATS) failure	Communication loop failure	Annunciator failure
	Wayside equipment failure	Track circuit failure	Switch machine failure
	Turnout	Turnout indication loss	Turnout equipment failure
	Axle counting	Axle counting equipment failure	Axle counting red tape occupancy
	Auxiliary equipment failure	Station departure indicator failure	Emergency stop equipment failure
Onboard signal system failure	Onboard signal equipment failure	Train management system (TMS) failure	Remote control interface unit failure
	Onboard signal loss	Vehicle onboard controller (VOBC)	Two VOBC failures
	Onboard signal loss	Communication loss	Single VOBC failure
	Onboard signal loss	Input failure	Train overshoot

3.2.2. Ontology Model Development

The ontological model includes different classes, properties, and individuals. These classes and individuals were created based on a taxonomy of metro accidents, while the properties were created based on the relations between these concepts. The ontological model was established as shown in Figure 3. This ontological model acts as the data structure to systematize the representation of accident cases, which contains three parts: accident characteristics, accident responses, and post-accident assessment.

Figure 3. The top-level ontological model for metro accidents.

- (1) **Accident characteristic** The accident characteristic category refers to basic information relating to metro operating accidents, such as the accident location, fault characteristics, and the accident type. The main accident locations are on the train, in the station, in the running section, or in the yard. The fault characteristic is the core element of the accident characteristic and the external manifestation of an accident. The managers usually recognize the accident through the identification of the accident characteristics. Different accidents have different fault characteristics. The accident types are determined through the fault characteristic. In this study, the accident types in metro operations were terrorist attacks, natural disasters, emergencies, passenger service failure, equipment system failure, and other accidents.
- (2) **Accident response** Accident responses are essential to reducing accidental damage. The accident response category refers to the specific procedures and details for handling accidents during metro operations. This category includes staff responses, resource responses, and measure responses.
- (3) **Post-accident assessment** The post-accident assessment refers to the treatments applied after a metro accident to avoid the same accident from happening again. This process includes the accident impact, accident causes, and corrective measures. The accident impact includes casualties, economic loss, and traffic delays.

3.2.3. Accident Cases Formalization

The aforementioned ontological model provides a unified semantic structure for the standardized expression of metro accident cases. Based on the semantic framework above, historical accident cases can be represented in a standard semantic format to facilitate computer analysis and processing.

Based on the developed ontology model, the metro accidents can be described by a set of attributes that can be divided into three main categories, namely, accident characteristics, accident response, and post-accident assessment. The name of these attributes is defined in the classes and properties of the ontology model, while the value of these attributes is based on the individuals of the ontology model. The details of the main attributes are shown in Table 2.

Table 2. The details of the main accident attributes.

Category	Attribute	Type	Meaning
Accident characteristic	Operational line	Integer	The metro operation line where the accident happened
	Station	String	The metro station where the accident happened
	Section	String	The metro section where the accident happened
	Train number	String	The code of the train that had the accident
	Fault characteristic	String	The representation of the train's abnormal operating state caused by the accident
Accident response	Response staff	String	The staff member who takes the accident response measures for the accident
	Response measure	String	The response measure taken by the accident staff for troubleshooting
	Response resource	String	The response equipment and supply for assisting the implementation of response measures
Post-accident assessment	Death number	Integer	The number of deaths caused by the accident
	Injury number	Integer	The number of injuries caused by the accident
	Economic loss	Float	The economic loss caused by the accident
	Delay time	Date	The delay time of the train caused by the accident
	Delayed trains	Integer	The number of delayed trains caused by the accident
	Lost trains	Integer	The number of lost trains caused by the accident
	Empty trains	Integer	The number of empty trains caused by the accident
	Accident level	String	The severity of the accident
	Accident type	String	The accident type according to the accident cause
	Fault cause	String	The main reason for the accident
Corrective measure	Corrective measure	String	The corrective measure according to the accident cause

Therefore, a case can be expressed as a three-tuple, $C = (C_c, C_r, C_p)$, where C_c , C_r , and C_p refer to a set of attributes that describe the accident, its accident response, and its post-accident assessment, respectively. For a given case base C_b , C_i is the sequence number of a case in C_b ($1 \leq i \leq n$, where n is the total number of cases in C_b). C_{ic} , C_{ir} , and C_{ip} represent the accident characteristic, accident response, and post-accident assessment of case C_i , respectively.

Web Ontology Language (OWL) was selected to encode the metro accident knowledge. OWL enables the knowledge to be linked together and represented semantically in a semantic network. The accident knowledge is exported into an “OWL” file that can be managed by the ontology management tool. Based on the concepts in the developed ontology model, textual accidents can be represented in a standard format, which can be used for further retrieval. In addition to the case retrieval, the corresponding responses to the accidents should also be obtained. Although the ontology was developed to represent domain knowledge, it cannot express the complex constraint knowledge that typically occurs in the form of rules. Therefore, the expression of complex constraint knowledge requires other technologies to be implemented, such as SWRL (Semantic Web Rule Language). Additionally, based on the ontology model, the metro accident cases can be annotated for further retrieval.

4. Automated Annotation of Metro Accidents

As shown in Figure 4, natural language processing (NLP) techniques were applied to annotate the accident records, which include word segmentation and stop word removal. The NLP can further support the semi-automated annotation of accident records by extracting the keywords of the accident records such that they can be semantically matched with the concepts in the ontology. Thus, the similarity calculation between the query sentences and past accidents can be quickly achieved after the annotation of accidents.

Figure 4. The annotation process of accident records.

(1) Word segmentation

Word segmentation divides a string of written language into its component words. A typical feature of the Chinese language is that there is no blank space between characters in a sentence, which results in a variety of ways for word segmentation and leads to different interpretations. For example, “钢筋混凝土 (reinforced concrete)” could be interpreted as “钢筋 (bar)” and “混凝土 (concrete).” Inaccurate word segmentation can lead to ambiguities and thus hinder the interpretation of the information. A function provided by *Jieba* (a Chinese word segmentation software based on Python) can overcome this difficulty by enabling users to create their own dictionaries. In this study, the concepts defined in the ontology model can be used as an additional dictionary for Chinese word segmentation. With the help of the ontology model, words can be segmented accurately with a word-group such that sentences in the accident reports can be interpreted correctly.

(2) Stop words removal

Stop words are extremely common words with less importance, such as prepositions, pronouns, and conjunctions. Through removing these meaningless words from the text, the retrieval efficiency can be improved. This paper used the stop list released by the Language Technology Platform (LTP), which contained 1208 words and punctuation marks [44].

After the data pre-processing procedure, including word segmentation and stop word removal, has been conducted, these accidents can be annotated and represented by a set of keywords defined in the ontology model, which can be used for similarity calculations in the case retrieval.

5. Similarity Measurement of the Accident Cases

The mechanism of the proposed framework for decision support is as follows: first, the query sentences were transformed into formalized concepts in the developed ontology model; then, the similarity between the query concepts and concepts of stored cases needs to be measured; finally, the most similar case is identified for reuse. Because the case base usually contains a large number of cases, it is crucial to evaluate the similarity between cases and ultimately find a similar case for reference.

Each case has specific attributes (numerical and symbolic). In modeling cases with ontology, the similarity between cases was calculated by the local–global approach. The local similarity focuses on the specific attributes of a case, such as the similarity of fault characterization. After the calculation of all local similarities, the global similarity is measured to reveal the overall similarity between cases, which considers the weight of each attribute.

5.1. Local Similarity

There are various attributes of cases in different domains. As far as the metro accident domain is concerned, the attributes can be divided into explicit numerical attributes and explicit symbolic attributes. Take the signal failure accident as an example; Table 3 shows the attributes of signal failure accidents.

The proposed similarity measure methods for the attributes of signal failure accidents are presented as follows.

5.1.1. Similarity of Numerical Attributes

For numerical attributes, the similarity calculation is typically based on the absolute distance of the two values to be compared. The similarity of numerical attributes is calculated using Equation (1):

$$\text{Sim}_n(X_i, Y_i) = 1 - \text{dis}(X_i - Y_i) = 1 - \frac{|x_i - y_i|}{\text{Max}_i - \text{Min}_i}, \quad (1)$$

where $\text{Sim}_n(X_i, Y_i)$ denotes the similarity of attribute i of cases X and Y , and the $\text{dis}(X_i - Y_i)$ reveals the absolute distance. In Equation (1), X_i and Y_i denote the values of attribute i of cases X and Y , respectively.

Table 3. Attributes of signal system failure accidents.

Category	Attribute Name	Type	Value
Accident characteristic	Train number	Symbol	A01, D03, E28, etc.
	Vehicle fault characterization	Symbol	Bogie failure, out of order, emergency braking, etc.
	Signal failure characterization	Symbol	SMC equipment failure, indication loss in turnout, etc.
Accident response	Staff response	Symbol	Train dispatcher, station watch-keeper, etc.
	Driving mode	Symbol	Automatic (ATO) driving pattern, Protective Manual (PM) driving pattern, Restricted Manual (RM) driving pattern
	Steering order	Symbol	Clean off the train, standby state, back to the parking lot, etc.
	Response measure	Symbol	Re-input, reactivation, equipment changes, etc.
Post-accident assessment	Time delay	Integer	Numerical information
	Train delay	Integer	Numerical information
	Train failure	Integer	Numerical information
	Clean off the train	Integer	Numerical information
	Accident cause	Symbol	Textual information
	Corrective action	Symbol	Textual information

5.1.2. Similarity of Symbolic Attributes

Given the ontology knowledge structure of the fault characteristic, for the symbolic attribute, the method aims to measure the comprehensive semantic similarity adopted from Dong [45].

The computation of a comprehensive similarity between symbolic attributes consists of two parts: the tree similarity and the similarity of the upper attributes. The tree similarity based on the semantic tree reveals the direct similarity of the attributes, while the similarity of the upper attributes reveals the indirect similarity of the concepts. Both of them are measured by considering the ancestor–child and parent–child relationships in similar maximal paths. After determining the similarities based on the semantic tree and based on the upper attributes, the comprehensive similarity can be obtained based on the weighted average of these similarities.

(1) Similarity measure based on the semantic tree

Sim_t denotes the similarity based on the semantic tree, which is the direct similarity between two concepts [46]. As shown in Equation (2), the similarity measure is based on the level of the attributes and the distance between them:

$$\text{Sim}_t(X_i, Y_i) = \begin{cases} \frac{\alpha \times [dl(X_i) + dl(Y_i)]}{[dis(X_i, Y_i) + \alpha] \times 2 \times \max dl \times \max(|dl(X_i) - dl(Y_i)|, 1)}, & \alpha > 0, X_i \neq Y_i, \\ 1, & X_i = Y_i \end{cases} \quad (2)$$

In Equation (2), the $dl(X_i)$ and $dl(Y_i)$ reveal the semantic level of the attributes, the $\text{dis}(X_i, Y_i)$ is the minimum distance between X_i and Y_i in the hierarchy, and the $\max dl$ is the maximum number of hierarchy conceptual structures. Additionally, α is an adjustable parameter that is the semantic distance of similarity to 1, where α is equal to 0.5 for a large number of experiments. Based on these parameters, the $\text{Sim}_t(X_i, Y_i)$ can be calculated through Equation (2). When the value of $\text{Sim}_t(X_i, Y_i)$ is 1, an equivalent relationship exists between the corresponding attributes X_i and Y_i .

(2) Similarity measure based on the upper attributes

Sim_s denotes the similarity of X_i and Y_i based on the upper attributes. It reveals the indirect similarity of X_i and Y_i . Equation (2) is used to calculate the similarity of X_i and Y_i based on the upper attributes.

(3) Comprehensive similarity measure

Sim_c denotes the comprehensive similarity, which refers to the final value of similarity between target attributes X_i and Y_i . As shown in Equation (3), this similarity is obtained by calculating the average value of the similarities based on the semantic tree and the upper attributes:

$$Sim_c(X_i, Y_i) = \beta Sim_t(X_i, Y_i) + \gamma Sim_s(X_i, Y_i), \beta + \gamma = 1. \quad (3)$$

In Equation (3), β denotes the weight value of the similarity based on the semantic tree, and γ denotes the weight value of the similarity based on the upper attributes. The values of β and γ range from 0 to 1. In this paper, β and γ were set to 0.5, which suggests that the similarities based on the semantic tree and the upper attributes had equal influences. For example, the semantic concept tree of signal failure illustrated in Figure 5 supports the similarity measure between the signal failure attributes.

Figure 5. Semantic concept tree of signal failure.

An example is given to show the calculation process of the similarity using the aforementioned methods. The similarity between “turnout failure” (abbreviated as T) and “STC equipment failure” (abbreviated as S) was calculated as follows. First, as shown in (4), we needed to measure the tree similarity based on the developed semantic model using Equation (2). Then, the tree similarity of the upper attribute was calculated in (5). The upper attribute of “turnout indication failure” and “STC equipment failure” was “wayside equipment failure” (abbreviated as W) and “signal system infrastructure” (abbreviated as I), respectively. Finally, the comprehensive similarity between T and S can be measured in (6) using Equation (3), where the value was 0.133. Thus, as shown in Table 4, the similarity values between the attributes of signal failure were obtained.

$$dl(T) = 4, dl(S) = 3, maxdl = 5, dis(T, S) = 3, |dl(T) - dl(S)| = 1, \quad (4)$$

$$Sim_t(T, S) = \frac{0.5 \times (3 + 4)}{(3 + 0.5) \times 2 \times 5 \times 1} = 0.100$$

$$dl(W) = 3, dl(I) = 2, dis(W, I) = 1, Sim_t(W, I) = \frac{0.5 \times (2 + 3)}{(1 + 0.5) \times 2 \times 5 \times 1} = 0.166 \quad (5)$$

$$Sim_c(T, S) = \frac{(Sim_t(T, S) + Sim_s(T, S))}{2} = \frac{(0.100 + 0.167)}{2} = 0.133 \quad (6)$$

Table 4. Similarity between the semantic concepts of partial signal failures.

Semantic Concept	STC Equipment Failure	Annunciator Failure	Relay Failure	Axle Counting Failure	Turnout Indication Failure	Single VOBC Failure	Two VOBC Failures	TMS Failure	Communication Loss/Cutout
STC Equipment Failure	1								
Annunciator Failure	0.218	1							
Relay Failure	0.218	0.65	1						
Axle Counting Failure	0.133	0.302	0.302	1					
Turnout Indication Failure	0.133	0.302	0.302	0.293	1				
Single VOBC Failure	0.07	0.166	0.166	0.174	0.174	1			
Two VOBC Failures	0.07	0.166	0.166	0.174	0.174	0.7	1		
TMS Failure	0.14	0.155	0.155	0.166	0.166	0.302	0.302	1	
Communication Loss/Cutout	0.14	0.155	0.155	0.166	0.166	0.302	0.302	0.207	1

5.2. Global Similarity

The global similarity reflects the degree of similarity between cases. We used the weighted Euclidean distance for the measure of global similarity. The weight of each attribute is determined by discussing the relative importance of each attribute in the overall evaluation when trying to find a similar case for solving a certain problem.

Assuming that the case representation consists of n features with feature weights w_i , the similarity between target cases X and Y can be computed as follows:

$$\text{Sim}(X, Y) = \sum_{i=1}^n w_i \times \text{Sim}_i(X_i, Y_i). \quad (7)$$

In Equation (7), $\text{Sim}_i(X_i, Y_i)$ denotes the local similarity and derives from $\text{Sim}_n(X_i, Y_i)$ or $\text{Sim}_c(X_i, Y_i)$, w_i denotes the weight of attribute i of cases X and Y , and Sim denotes the global similarity.

6. Accident Case Development and Implementation for the Accident Response—A Case Study

To provide a proof-of-concept implementation for the proposed method, a case study was conducted. The case study focused on modeling actual metro accident knowledge and illustrating the use of the proposed method within future metro accident analyses. Protégé 4.3 was used to develop the ontological model and MyCBR was used for creating the CBR prototypes. Protégé is a free, open-source ontology editor that can provide a visual environment to create and edit an ontology model [47]. MyCBR allows for preliminary modeling of similarity measures and supports the simulation of a case retrieval [48].

6.1. Case Database

The metro accident cases in a metro company in Wuhan used to develop the accident case database mainly came from the following sources:

- (1) Checking records of metro operations, which included both the contents and results of operational checks.
- (2) Safety production monthly reports of metro operations, which kept a record of all accidents that have occurred over the previous month and their corresponding details, including the accident process description, relevant accident response measures, accident impact, cause analysis, and corrective measures.

For example, a turnout failure accident saved in the safe production monthly report of the metro operations was introduced as follows, which consists of the accident characteristics, accident response, and post-accident assessment. At 6:12 a.m. on 14 July 2016, the indication of the 16th and 18th turnouts were lost, thereby affecting the metro operation and subsequently resulting in train delays. At 6:15 a.m., the operational company decided to change the train route. At 6:47 a.m., the fault was fixed, and the train route was changed back to the original line. The cause of the accident and the corrective measures as determined after the investigation. The turnout failure was attributed to the omission of overhaul work on the 16th and 18th turnouts due to bad weather. However, it is worthy to note that the accident records, such as the one shown in Figure 6, were unstructured and expressed in natural language. This means that they were difficult to process digitally. These documents would be processed into structured data using NLP techniques.

	Case ID	Train number	Station	Fault characteristic	Response measure	Driving mode	Steering order	Corrective action	Time delay	Train failure	Train delay
1	caseID	列车号	车站	故障表征	司机应急处理措施	司机驾驶模式	行调指令	整改方案	延误	掉线	晚点
2	1	A03	汉口北	cutout	换端	PM模式	检查车辆状况	车辆部门应在车辆检修作业中加强接线检查, 同时加强对乘务员执行《列车出库标准》情况的监督检查。	4	2	4
3	2	A03	兴业路	cutout	重新投入	RM模式	越过故障区段	通号部联合乘务员尽快找到故障的具体原因	3	0	0
4	3	A04	东吴大道	cutout	重新投入	PM模式	前往投入点	检修人员对该处松动线缆重新连接	4	0	0
5	4	A08	汉口北	cutout	重新投入	PM模式	前往投入点	通号部和车辆一部要相互配合, 查明原因	4	0	0
	5	A09	东吴大道	单VOBC故障	重新投入	PM模式	前往投入点	变更1号线一期车辆ATC速度传感器的安装位置, 以减少因打滑或空转而引起VOBC死机	3	0	0

Figure 6. The annotated metro accident cases.

A total of 120 accidents were extracted from the safety monthly reports on metro operations. With the support of NLP techniques, these accidents were analyzed and automatically annotated. The classes, properties, and individuals of the abovementioned ontological model were used to assist the annotation of metro accidents, as shown in Figure 6, such as the “train number,” “station,” and “fault characteristics.” These annotations belong to different attributes, which can be used in the similarity calculations. These cases were stored in CSV files [49].

Based on the statistical analysis of the metro operating accidents, more than 80% of accidents were caused by the poor state of equipment and were represented by equipment failures, such as signal system failures and metro vehicle failures. Thus, in this study, the “signal system failure” accidents were selected as the concrete example to test the proposed method. As shown in Figure 7, the classes and individuals of the “signal system failure” were developed using Protégé 4.3.

Figure 7. The screenshot of classes and individuals of the “signal system failure.”

6.2. Accident Cases Retrieval

In this case, MyCBR followed the local–global approach, which divided the similarity definition into a set of local similarity measures for each attribute, a set of attribute weights, and a global similarity measure for calculating the final similarity value. In this case, for an attribute-value-based case representation consisting of n attributes, the similarity between an input case X and a known case Y may be calculated by using Equation (4) in Section 5.2.

The similarity measures of the attributes of accidents were calculated in MyCBR, where various modes for measuring similarity are provided, including the standard, table, and taxonomy modes. The standard mode measures the similarity of numeric attributes using a basic formula, while the table and taxonomy modes measure the similarity of symbolic attributes. Specifically, the table mode is based on the user-defined concept similarity degree, while the taxonomy mode is based on the hierarchical structure of the attribute value.

The attributes of accident cases were divided into numerical attributes and symbolic attributes. The “standard mode” was adapted to measure the similarity of partial numerical attributes, including “time delay,” “train delay,” “train failure,” and “clean off the train,” as shown in Figure 8. The “table mode” was adopted to measure the similarity of the symbolic attributes. The core algorithm of the “standard mode” and “table mode” was the similarity measure method of the numerical attributes and symbolic attributes given in Section 5.1.

Figure 8. Similarity measures for numerical attributes in standard mode.

Then, as shown in Figure 9, the weights of each attribute were determined using the Delphi method based on the importance of attributes when describing the accident knowledge. Following the development of the case database and the definition of the attribute weights, the query sentences were inputted into the system. The search engine then computed the similarity between the query sentences and all cases stored in the case database.

Figure 9. Setting interface of the case attributes weights.

6.3. Testing and Evaluation

6.3.1. Testing

Given that the number of accidents caused by signal system failures was considerably larger than the number of accidents due to other causes, here, the fault characteristics of the signal system failure were chosen as the index to retrieve similar cases. Assume that a metro has experienced the following problem:

The signals of the metro are cut out during the metro operation stage, which may cause the loss of the metro position and then become rear-ended. As reported in the SMC (system management center) equipment, the metro integrity was lost.

“Cutout” was defined as the query sentence of the above case. Figure 10 presents the similar cases retrieved. Their similarity to the input query was given. The case attributes are shown in descending order based on their similarity to the corresponding attributes of the input case. For providing reliable support to managers, we have considered retrieved cases whose similarity scores were more than 0.7. In Figure 10, facing the signal failure accident with “cutout,” managers can give the following orders based on the retrieval of similar cases: the driver should change the driving mode into PM (protection of the artificial driving mode) and the train should go to the entry, where the signal malfunction is resolved and the operation is resumed. According to the historical cases, the delay time caused by the above accidents is roughly four minutes, and improving the information system was adopted as the corrective measure.

Query		KB_474610_Cl...	KB_474610_Cl...	KB_474610_Cl...	KB_474610_Cl...	KB_474610_Cl...	KB_474610_Cl...
		1	2	3	9	10	11
Similarity		0.96	0.96	0.96	0.82	0.82	0.73
事故原因	车钩天线阻值异常	故障代码显示模糊...	VOBC分析数据...	车钩设备接线端子...	列车完整性回路设...	雨天导致列车在环...	
列车号	undefined_ A03	A12	A07	A21	A33	C10	
司机应急处置措施	重新投入	重新投入	重新投入	重新投入	重新投入	重新投入	
司机驾驶模式	PM模式	PM模式	PM模式	PM模式	PM模式	PM模式	
延误	4	5	5	4	3	4	
掉线	0	0	1	0	0	0	
故障表征	CUTOUT	CUTOUT	CUTOUT	列车完整性丢失	列车完整性丢失	单VOBC故障	
整改方案	通号部要加强车钩...	通号部加强信号系...	通号部加强信号...	检修人员对该处松...	改造列车完整性回...	变更1号线一期车辆...	
晚点	1	0	1	0	0	2	
清客	0	1	1	0	0	1	
行调指令	前在投入点	前在投入点	退回车辆段	建站乘降	建站乘降	前在投入点	
车站名	汉口北站	汉口北站	汉口北站	黄陂路站	黄陂路站	东吴大路站	

Figure 10. Retrieval results of “cutout” in descending order based on their similarity.

6.3.2. Evaluation

The evaluation was performed to test the ability of the proposed method to accurately retrieve a suitable case from the case database. Metrics like “precision” and “recall” are usually used in the evaluation of performance in CBR research [50]. “Precision” and “recall” were computed by utilizing Equations (8) and (9):

$$\text{Precision} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false positives}} \quad (8)$$

$$\text{Recall} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} \quad (9)$$

where “true positive (TP)” represents a retrieved case that provided a feasible solution for the query case, “false positive (FP)” means that a solution of the recommended case was not reliable for the query case, and “false negative (FN)” reflects that the stored cases with suitable solutions for the query case were not retrieved from the library. In this study, the “precision” metric is more important than “recall.” “Precision” is the portion of retrieved cases that were deemed by the system to provide suitable solutions for the query cases, while “recall” is the portion of all cases in the case library with feasible solutions, including those that were not retrieved [8]. In the context of a metro accident response, the accurate retrieval of cases and avoiding the provision of unfeasible cases (high precision) is more important than retrieving all the relevant cases (high recall). The retrieved erroneous cases may affect the decision-making in the emergent response of metro accidents during operations and further lead to serious consequences, such as casualties or heavy economic loss. Therefore, “precision” was selected

as the evaluation metric since “precision” and “recall” are often negatively correlated. Additionally, considering the circumstance that in the real working environment, managers expect to get the required information within a limited amount of time, the case with the highest similarity to the input query has the most value to the users. Thus, in the evaluation process, we verified the suitability of the solution included in the retrieved case with the highest similarity to each given query through expert reviews.

A set of keywords (e.g., “cutout,” “single VOBC failure,” “TMS failure,” “ATC equipment failure”) that are relevant to metro operational accidents were defined for making up 10 testing queries. Retrieved cases with a similarity of more than 0.7 were selected as the retrieved results in this study, which were then evaluated via expert reviews. Then, five experts in a metro company in Wuhan (filed engineers and metro dispatchers with an average of 10 years of experience) were invited to evaluate the performance of the retrieval results because they were familiar with metro accident responses. The experts would discuss the retrieved results for each testing query with each other to judge whether the retrieved cases belonged to “TP” or “FP” and verify the feasibility of the highest similarity case. Table 5 shows the result of the expert review on the 10 testing queries. The mean precision of the proposed method was 91%. Additionally, the recommended solutions in the most similar case were feasible to solve the problems in the query case, which could help the manager to undertake better decision-making in a limited amount of time in the accident. The accuracy of the highest similarity case verified the feasibility of the proposed method in practice.

Table 5. Expert review results.

No	Testing Query	Number of Retrievals		Performance	
		TP	FP	Precision	Feasibility of Highest Similarity Case
1	Cutout	6	0	100%	Yes
2	Single VOBC failure	10	1	91%	Yes
3	STC equipment failure	8	1	89%	Yes
4	TMS failure	9	0	100%	Yes
5	SMC equipment failure	10	1	91%	Yes
6	ATS failure	8	1	89%	Yes
7	Metro door failure	9	1	90%	Yes
8	Loss of indication of a turnout	7	0	100%	Yes
9	Remote control interface unit failure	5	1	83%	Yes
10	Emergency braking cannot be alleviated	6	2	75%	Yes

6.4. Inference of Accident Response Rules in RBR

CBR has some advantages, such as incremental learning, easy acquisition, and easy maintenance. However, it may fail when no relevant case is available, especially in metro accidents in which the complete records are rare. Thus, to overcome shortages of the CBR method due to the lack of relevant cases, rule-based reasoning (RBR) is added as the complementary method in the decision-making process in this paper. The integration of CBR and RBR is useful in the decision-making process [51]. Cases reveal the knowledge or experiences obtained from specific situations and rules represent the general knowledge of a specific domain from documents, such as regulations and emergency plans.

In a realistic accident response process, managers need to identify the accident level first based on fault characteristics of the accident, and then take appropriate measures under the guidance of various documents. Figure 11 shows an example of an accident response process in the case of large passenger flow. Thus, in this study, RBR was used to infer the accident level and response measures. Semantic Web Rule Language (SWRL) was used to represent rules for reasoning the response measures. An SWRL rule contains an antecedent part and a consequent part, both of which were written in Web Ontology Language (OWL) classes, properties, individuals, and data values. As an example, the rules of a large passenger flow accident are presented as follows. Table 6 shows the response rules under

the circumstance of level 3 passenger flow. As shown in Table 6, different staff in a specified position should take different measures to minimize the impact of accidents.

Figure 11. The handling process of a large passenger flow accident.

Protégé was used to verify the feasibility of RBR. In Protégé, by iterating the SWRL rules listed in Table 6, the response measures to be taken were inferred. As seen in Figure 12, once the passenger flow level increased to level 3, setting the passenger flow divider to guide the passenger flow should be taken by the station attendant. Therefore, through encoding the SWRL rules, RBR can be used to support decision-making by deciding the response measures to the metro accident, which are recorded in documents like emergency plans.

Rule 1 of level 1
Rule 2 of level 1
Rule 1 of level 3
Rule 2 of level 3
Rule 1 of level 2
Rule 2 of level 2
Responding rule 1
Responding rule 2
Responding rule 3
Responding rule 4
Responding rule 5

Enabled	Name	Expression
☒	一级客流判定规则1	地铁运营应当预案(?x) ∧ 车站名称(?x, 中南路站) ∧ 站厅人数(?x, ?y) ∧ swrl:greaterThan(?y, 5100) → 有事件类型(?x, 一级客流)
☒	一级客流判定规则2	地铁运营应当预案(?x) ∧ 车站名称(?x, 中南路站) ∧ 站厅人数(?x, ?y) ∧ swrl:greaterThan(?y, 6800) → 有事件类型(?x, 一级客流)
☒	二级客流判定规则1	地铁运营应当预案(?x) ∧ 车站名称(?x, 中南路站) ∧ 站厅人数(?x, ?y) ∧ swrl:greaterThan(?y, 6700) → 有事件类型(?x, 二级客流)
☒	二级客流判定规则2	地铁运营应当预案(?x) ∧ 车站名称(?x, 中南路站) ∧ 站厅人数(?x, ?y) ∧ swrl:greaterThan(?y, 9800) → 有事件类型(?x, 二级客流)
☒	三级客流判定规则1	地铁运营应当预案(?x) ∧ 车站名称(?x, 中南路站) ∧ 站厅人数(?x, ?y) ∧ swrl:greaterThan(?y, 5900) → 有事件类型(?x, 三级客流)
☒	三级客流判定规则2	地铁运营应当预案(?x) ∧ 车站名称(?x, 中南路站) ∧ 站厅人数(?x, ?y) ∧ swrl:greaterThan(?y, 8200) → 有事件类型(?x, 三级客流)
☒	应急处理规则1	地铁运营应当预案(?x) ∧ 有事件类型(?x, 三级客流) ∧ 应急人员岗位(?x, 值班站长) → 采取响应措施(?x, 向站区和客运部申请)
☒	应急处理规则2	地铁运营应当预案(?x) ∧ 有事件类型(?x, 三级客流) ∧ 应急人员岗位(?x, 行车值班员) → 采取响应措施(?x, 加强与站厅、行调联系)
☒	应急处理规则3	地铁运营应当预案(?x) ∧ 有事件类型(?x, 三级客流) ∧ 应急人员岗位(?x, 客运值班员) → 采取响应措施(?x, 适当减少或关停临站客流)
☒	应急处理规则4	地铁运营应当预案(?x) ∧ 有事件类型(?x, 三级客流) ∧ 应急人员岗位(?x, 站务员A) → 采取响应措施(?x, 在电扶梯处进行乘客引导)
☒	应急处理规则5	地铁运营应当预案(?x) ∧ 有事件类型(?x, 三级客流) ∧ 应急人员岗位(?x, 站务员B) → 采取响应措施(?x, 在B、C、D出入口增加客流)

应急人员岗位: 站务员A

车站名称: 中南路站

站厅人数: 9740

站台人数: 6750

Accident level: 有事件类型: 三级客流 Level 3

Corresponding response: 采取响应措施: 在电扶梯处进行乘客引导及分流

setting the passenger flow divider to guide the passenger flow

Figure 12. The decision-making result as shown in the interface of Protégé.

Table 6. Accident response rules for level 3 passenger flow.

Decision Rule for the Response Measures of Different Roles	
Response rules	Metro accident(?x) $\wedge$ Accident level(?x, Level 3 passenger flow) $\wedge$ Response staff(?x, Station agent) $\rightarrow$ Response measure(?x, Apply for additional staff to the station area and passenger transport department)
	Metro accident(?x) $\wedge$ Accident level(?x, Level 3 passenger flow) $\wedge$ Response staff(?x, Driving attendant) $\rightarrow$ Response measure(?x, Strengthen the joint control between stations and trains, and set the order exemption inspection mode of import and export in the station)
	Metro accident(?x) $\wedge$ Accident level(?x, Level 3 passenger flow) $\wedge$ Response staff(?x, Passenger attendant) $\rightarrow$ Response measure(?x, Appropriately reduce or shut down the temporary ticketing points to limit the import passenger flow)
	Metro accident(?x) $\wedge$ Accident level(?x, Level 3 passenger flow) $\wedge$ Response staff(?x, Station operator A) $\rightarrow$ Response measure(?x, Guide and shunt the passengers at the escalator)
	Metro accident(?x) $\wedge$ Accident level(?x, Level 3 passenger flow) $\wedge$ Response staff(?x, Station operator B) $\rightarrow$ Response measure(?x, Add the isolated railing at entrances and exits of B, C, D, and E to conduct the bidirectional flow)

7. Conclusions

In this study, an ontological framework integrating CBR and NLP was proposed to facilitate knowledge retrieval and reasoning, which can be used to support decision-making in metro accident responses. First, an ontological model was developed to formalize the representation of domain knowledge, such as the regulations, emergency plans, and past accident records. Based on the ontology model, NLP techniques were adopted to automatically annotate accident cases into a set of keywords, which could enhance the efficiency of further case retrievals. Then, RBR was used as a complement to CBR because CBR may fail when there are few or no relevant cases. The combination of CBR and RBR can avoid this problem in the metro accident response by retrieving the specific knowledge found in a similar case or general knowledge found in regulations. SWRL rules were used in this study to represent the response measures in the regulations. Finally, the proposed framework was tested and evaluated through a case study via expert reviews. A total of 120 metro accidents in the operational phase were used to build a case library. Then, five experts in a metro operating company were invited to evaluate the performance of the proposed method in 10 testing queries based on the “precision” metric. Considering the demand for quick responses, users tend to adopt solutions in the case with the highest similarity to the input query. Thus, the accuracy of cases with the highest similarity was also verified in 10 testing queries. The results show the good performance of the proposed method.

However, this study still has some limitations that may be alleviated in future studies. The limitations are summarized as follows.

First, the ontology development is time-consuming and considerable initial work needs to be done. A significant manual effort is required to update the ontology model if the metro accident knowledge base is updated.

Second, the proposed framework is limited in terms of case retrieval within a small case database (120 accident records). The main purpose of this study was to develop a general framework based on ontological CBR and NLP for supporting decision-making when facing emergent accidents rather than a complete accident case database. However, the small database would affect the retrieval results. To apply the proposed framework in practice, the size of the database should be extended in the future. Moreover, with the increasing number of stored accidents in the database, the pattern and trends of historical accidents can be identified through statistical methods, which can be used to make possible predictions of future accidents.

Third, the proposed framework has not been validated in practice. To promote its application, a system with a convenient interface should be further developed.

Author Contributions: Conceptualization, B.Z. and H.W.; Methodology, H.W.; Software, L.J.; Validation, H.W. and L.J.; Writing—original draft preparation, H.W., B.Z., and X.X.; Writing—review and editing, H.W., B.Z. and B.M.; Funding acquisition, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is partly supported by the National Natural Science Foundation of China (grant nos. 51878311, 71732001, and 51978302)

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

References

1. Xing, Y.; Dissanayake, S.; Lu, J.; Long, S.; Lou, Y. An analysis of escalator-related injuries in metro stations in China, 2013–2015. *Accid. Anal. Prev.* **2019**, *122*, 332–341. [[CrossRef](#)] [[PubMed](#)]
2. Shi, C.; Zhong, M.; Nong, X.; He, L.; Shi, J.; Feng, G. Modeling and safety strategy of passenger evacuation in a metro station in China. *Saf. Sci.* **2012**, *50*, 1319–1332. [[CrossRef](#)]
3. Murray-Tuite, P.; Wernstedt, K.; Yin, W. Behavioral shifts after a fatal rapid transit accident: A multinomial logit model. *Transp. Res. Part F Traffic Psychol. Behav.* **2014**, *24*, 218–230. [[CrossRef](#)]
4. Dong, H.; Li, F.; Xu, R. Design of emergency preplan exercise management information system for URT. CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient. In Proceedings of the 12th COTA International Conference of Transportation Professionals, Beijing, China, 3–6 August 2012; pp. 2813–2823. [[CrossRef](#)]

5. Wang, J.; Fang, W. A structured method for the traffic dispatcher error behavior analysis in metro accident investigation. *Saf. Sci.* **2014**, *70*, 339–347. [\[CrossRef\]](#)
6. Yu, F.; Li, X.Y.; Han, X.S. Risk response for urban water supply network using case-based reasoning during a natural disaster. *Saf. Sci.* **2018**, *106*, 121–139. [\[CrossRef\]](#)
7. Lu, Y.; Li, Q.; Xiao, W. Case-based reasoning for automated safety risk analysis on subway operation: Case representation and retrieval. *Saf. Sci.* **2013**, *57*, 75–81. [\[CrossRef\]](#)
8. Goh, Y.M.; Guo, H. FPSWizard: A web-based CBR-RBR system for supporting the design of active fall protection systems. *Autom. Constr.* **2018**, *85*, 40–50. [\[CrossRef\]](#)
9. Lee, D.Y.; Chi, H.L.; Wang, J.; Wang, X.; Park, C. A linked data system framework for sharing construction defect information using ontologies and BIM environments. *Autom. Constr.* **2016**, *68*, 102–113. [\[CrossRef\]](#)
10. Tixier, A.J.P.; Hallowell, M.R.; Rajagopalan, B.; Bowman, D. Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports. *Autom. Constr.* **2016**, *62*, 45–56. [\[CrossRef\]](#)
11. Liu, X.; Guo, H.; Lin, Y.R.; Li, Y.; Hou, J. Analyzing spatial-temporal distribution of natural hazards in China by mining news sources. *Nat. Hazards Rev.* **2018**, *19*, 04018006. [\[CrossRef\]](#)
12. Lee, J.; Yi, J.S.; Son, J. Development of automatic-extraction model of poisonous clauses in international construction contracts using rule-based NLP. *J. Comput. Civ. Eng.* **2019**, *33*, 04019003. [\[CrossRef\]](#)
13. Li, Q.; Song, L.; List, G.F.; Deng, Y.; Zhou, Z.; Liu, P. A new approach to understand metro operation safety by exploring metro operation hazard network (MOHN). *Saf. Sci.* **2017**, *93*, 50–61. [\[CrossRef\]](#)
14. Sanchez-Cazorla, A.; Alfalla-Luque, R.; Irimia-Diéguez, A. Risk identification in megaprojects as a crucial phase of risk management: A literature review. *Proj. Manag. J.* **2016**, *47*, 75–93. [\[CrossRef\]](#)
15. Liu, W.; Zhao, T.; Zhou, W.; Tang, J. Safety risk factors of metro tunnel construction in China: An integrated study with EFA and SEM. *Saf. Sci.* **2018**, *105*, 98–113. [\[CrossRef\]](#)
16. Ding, L.; Yu, H.; Li, H.; Zhou, C.; Wu, X.; Yu, M. Safety risk identification system for metro construction on the basis of construction drawings. *Autom. Constr.* **2012**, *27*, 120–137. [\[CrossRef\]](#)
17. Liu, P.; Li, Q.; Bian, J.; Song, L.; Xiahou, X. Using interpretative structural modeling to identify critical success factors for safety management in subway construction: A China study. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1359. [\[CrossRef\]](#)
18. Yan, L.; Tong, W.; Hui, D.; Zongzhi, W. Research and application on risk assessment DEA model of crowd crushing and trampling accidents in subway stations. *Procedia Eng.* **2012**, *43*, 494–498. [\[CrossRef\]](#)
19. Ding, L.; Zhou, C.; Deng, Q.; Luo, H.; Ye, X.; Ni, Y.Q.; Guo, P. Real-time safety early warning system for cross passage construction in Yangtze Riverbed Metro Tunnel based on the internet of things. *Autom. Constr.* **2013**, *36*, 25–37. [\[CrossRef\]](#)
20. Li, M.; Yu, H.; Liu, P. An automated safety risk recognition mechanism for underground construction at the pre-construction stage based on BIM. *Autom. Constr.* **2018**, *91*, 284–292. [\[CrossRef\]](#)
21. Zhong, B.; Wu, H.; Ding, L.; Love, P.E.; Li, H.; Luo, H.; Jiao, L. Mapping computer vision research in construction: Developments, knowledge gaps and implications for research. *Autom. Constr.* **2019**, *107*. [\[CrossRef\]](#)
22. Anumba, C.; Issa, R.R.A.; Pan, J.; Mutis, I. Ontology-based information and knowledge management in construction. *Constr. Innov.* **2008**, *8*, 218–239. [\[CrossRef\]](#)
23. Xing, X.; Zhong, B.; Luo, H.; Li, H.; Wu, H. Ontology for safety risk identification in metro construction. *Comput. Ind.* **2019**, *109*, 14–30. [\[CrossRef\]](#)
24. Koo, L.; Trokanas, N.; Cecelja, F. A semantic framework for enabling model integration for biorefining. *Comput. Chem. Eng.* **2017**, *100*, 219–231. [\[CrossRef\]](#)
25. Ding, L.; Zhong, B.; Wu, S.; Luo, H. Construction risk knowledge management in BIM using ontology and semantic web technology. *Saf. Sci.* **2016**, *87*, 202–213. [\[CrossRef\]](#)
26. Wang, Z.; Li, H.; Zhang, X. Construction waste recycling robot for nails and screws: Computer vision technology and neural network approach. *Autom. Constr.* **2019**, *97*, 220–228. [\[CrossRef\]](#)
27. Wu, C.G.; Xu, X.; Zhang, B.K.; Na, Y.L. Domain ontology for scenario-based hazard evaluation. *Saf. Sci.* **2013**, *60*, 21–34. [\[CrossRef\]](#)
28. Guo, H.; Goh, Y.M. Ontology for design of active fall protection systems. *Autom. Constr.* **2017**, *82*, 138–153. [\[CrossRef\]](#)
29. Jonassen, D.H.; Hernandez-Serrano, J. Case-based reasoning and instructional design: Using stories to support problem solving. *Educ. Technol. Res. Dev.* **2002**, *50*, 65–77. [\[CrossRef\]](#)

30. Schank, R.C. *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*; Cambridge University Press: Cambridge, UK, 1983.
31. Chen, W.T.; Chang, P.Y.; Chou, K.; Mortis, L.E. Developing a CBR-based adjudication system for fatal construction industry occupational accidents. Part I: Building the system framework. *Expert Syst. Appl.* **2010**, *37*, 4867–4880. [\[CrossRef\]](#)
32. Guo, Y.; Hu, J.; Peng, Y. A CBR system for injection mould design based on ontology: A case study. *Comput. Des.* **2012**, *44*, 496–508. [\[CrossRef\]](#)
33. Lin, L.; Lin, L.; Zhong, S. Handling missing values and unmatched features in a CBR system for hydro-generator design. *Comput. Des.* **2013**, *45*, 963–976. [\[CrossRef\]](#)
34. 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. [\[CrossRef\]](#)
35. Ferrara, E.; Baumgartner, R. *Combinations of Intelligent Methods and Applications*; Springer: Berlin, Germany, 2011.
36. Shi, W.; Barnden, J.A. How to combine CBR and RBR for diagnosing multiple medical disorder cases. *Intell. Tutoring Syst.* **2005**, *3620*, 477–491.
37. Tung, Y.H.; Tseng, S.S.; Weng, J.F.; Lee, T.P.; Liao, A.Y.; Tsai, W.N. A rule-based CBR approach for expert finding and problem diagnosis. *Expert Syst. Appl.* **2010**, *37*, 2427–2438. [\[CrossRef\]](#)
38. Bouhana, A.; Zidi, A.; Fekih, A.; Chabchoub, H.; Abed, M. An ontology-based CBR approach for personalized itinerary search systems for sustainable urban freight transport. *Expert Syst. Appl.* **2015**, *42*, 3724–3741. [\[CrossRef\]](#)
39. Xu, F.X.; Liu, X.; Chen, W.; Zhou, C.; Cao, B. Ontology-based method for fault diagnosis of loaders. *Sensors* **2018**, *18*, 729. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Maalel, A.; Mejri, L.; Hadj-Mabrouk, H.; Ben Ghezela, H. Toward a knowledge management approach based on an ontology and Case-based Reasoning (CBR): Application to railroad accidents. In Proceedings of the Sixth International Conference on Research Challenges in Information Science (RCIS), Valencia, Spain, 16–18 May 2012; pp. 1–6.
41. Tao, M.; Ota, K.; Dong, M. Ontology-based data semantic management and application in IoT- and cloud-enabled smart homes. *Future Gener. Comput. Syst.* **2017**, *76*, 528–539. [\[CrossRef\]](#)
42. Canós-Cerdá, J.H.; Alonso, G.; Jaen, J. A multimedia approach to the efficient implementation and use of emergency plans. *IEEE MultiMed.* **2004**, *11*, 106–110. [\[CrossRef\]](#)
43. Anumba, C.; Pan, J.; Issa, R.; Mutis, I. Collaborative project information management in a semantic web environment. *Eng. Constr. Arch. Manag.* **2008**, *15*, 78–94. [\[CrossRef\]](#)
44. Che, W.; Li, Z.; Liu, T. LTP: A chinese language technology platform. In Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations Coling, Beijing, China, 23–27 August 2010; pp. 13–16. [\[CrossRef\]](#)
45. Dong, X.; Yong, Z.; Yang, C. Case Knowledge representation and conceptual similarity computation based on semantic information. *Comput. Sci. Eng.* **2010**, *12*, 34. [\[CrossRef\]](#)
46. Canós, J.H.; Borges, M.; Penadés, M.C.; Gómez, A.; Llavador, M.; Canós-Cerdá, J.H. Improving emergency plans management with SAGA. *Technol. Forecast. Soc. Chang.* **2013**, *80*, 1868–1876. [\[CrossRef\]](#)
47. Horridge, M.; Jupp, S.; Moulton, G.; Rector, N.A.; Stevens, R.; Wroe, C. A practical guide to building OWL ontologies using the protege 4 and CO-ODE tools. In *The Semantic Web: Research and Applications*; Springer: Berlin, Germany, 2014. [\[CrossRef\]](#)
48. Zilles, L. MyCBR Tutorial. 2009. Available online: <http://www.mycbr-project.net/tutorials.html> (accessed on 16 March 2020).
49. Stahl, A.; Roth-Berghofer, T.R. Rapid Prototyping of CBR Applications with the open source tool myCBR. *Intell. Tutoring Syst.* **2008**, *5239*, 615–629.
50. Nasiri, S.; Zahedi, G.; Kuntz, S.; Fathi, M. Knowledge representation and management based on an ontological CBR system for dementia caregiving. *Neurocomputing* **2019**, *350*, 181–194. [\[CrossRef\]](#)
51. Jim, P.; Ioannis, H. Categorizing approaches combining rule-based and case-based reasoning. *Expert Syst.* **2010**, *24*, 97–122. [\[CrossRef\]](#)