

# Research and Application on Case Representation and Retrieval Methods of Coal Mine Accidents

Wenxia Pan

School of Resource and Safety Engineering,  
China University of Mining & Technology,  
Beijing, 10083, China

**Abstract**—This paper applies case-based reasoning method to emergency decision-making research field, and mainly researches case representation and case retrieval method. At first, this paper analyzes character of coal mine emergency rescue work, and describes the coal mine accident case using frame representation. Then this paper adopts the secondary retrieval strategy based on clustering analysis, and designs comprehensive similarity algorithm based on two-layer construct combining structure similarity calculation and attribute similarity calculation. At last, the paper uses retrieval algorithm to show the application of case-based reasoning in emergency decision-making research field taking coal mine gas accident cases as examples.

**Keywords**-coal mine accident; case representation; case retrieval; case-based reasoning

## I. INTRODUCTION

The coal industry is the pillar industry of our country. In recent years, situation of safety production of the coal enterprises is generally stable in China, but it is not optimistic. Compared with other industries, security situation of the coal industry is still the most serious. Currently, accidents still happen frequently in coal industry, therefore it is urgent and necessary to strengthen the safety management of coal mine enterprise.

At present, there are still many problems existing in coal mine emergency management and rescue work in china, such as it is hard to take timely and effective emergency response strategy because lack of scientific, unified emergency rescue plan and mechanism; When facing accidents, many coal mine enterprises still implement emergency rescue work relying on traditional means such as paper plan, which is difficult for them to guarantee the efficiency of the real time dispatch and command when accident happens. Therefore, it is necessary to construct establish coal mine emergency rescue command management information system by using computer technology<sup>[1]</sup>. Case-based reasoning means taking advantage of similar history cases to solve current problems<sup>[2]</sup>. Its main step is to retrieval similar cases from cases library, then get one or more cases, and reuse or modify solution of the cases to solve current problems<sup>[3]</sup>. The process of case-based reasoning mainly includes five steps: case representation, case retrieve, case reuse and case revision and case retain.

This paper mainly studies case representation and case retrieval.

## II. CASE REPRESENTATION OF COAL MINE ACCIDENTS

Coal mine accident characters are as follows: (1) the kind of coal mine accident is various. (2) coal mine accident contains a series of states. (3) coal mine accident may derive other accident.

Case contents generally includes three parts: problem descriptions, solutions, result evaluation<sup>[4]</sup>. Problem descriptions are the attributes information of the accident. Solutions are case specific disposal method. Result evaluation refers to the evaluation of result that implementing solutions lead to. In this paper, description information of coal mine accident cases can be divided into five parts: basic information, environmental information, problem information, solution information, decision results description information. Among them, basic information includes case number, case time, accident type, etc. Environmental information includes area, terrain, etc. Problem information includes gas volume fraction, explosion scale, ventilation system damage, etc. Solution information includes the number of emergency rescue team, the number of medical rescue team, the number of logistics support team, etc. Decision results description information includes information of rescued personnel, the loss assessment of accident, etc.

This paper adopts frame representation method to represent case. Frame representation method is a kind of organized and structured knowledge representation method<sup>[5]</sup>. A frame is usually composed of frame name and a number of slots.

## III. CASE RETRIEVAL OF COAL MINE ACCIDENTS

The nearest neighbor algorithm is commonly used algorithm when retrieving case, but the nearest neighbor algorithm is unable to calculate the similarity between the cases having missing attribute values, and in practice measuring the similarity between history case and current problem may not be comprehensive and accurate enough only using attribute similarity, because relationships between case attributes are complex. In order to solve the problem, this paper proposes comprehensive similarity algorithm

based on two-layer construct combining structure similarity calculation and attribute similarity calculation, which calculate comprehensive similarity based on structural similarity and attribute similarity of history case and current case. The algorithm may solve the problem that traditional means can't calculate similarity because attribute value is missing, which can improve efficiency and accuracy of case retrieval.

#### A. Two-level Retrieval Strategy Based on Cluster Analysis

This paper uses two-level retrieval strategy based on cluster analysis: Firstly, we use cluster analysis method to sort cases in case library, made cases of the same kind have high similarity, while cases of different kind have lower similarity. Each cluster is represented by a typical case, thus we get several typical cases, and each typical case corresponds to many specific cases of the same kind.

When we begin to retrieval case, firstly, we should calculate the similarity between each typical case and current case, and choose the typical case with the highest similarity, the second step is to retrieval case from specific cases that the typical case corresponds to, then we can find out one or more specific cases which are more similar to current case. The flaw of the retrieval strategy is the solution may be a relatively satisfactory solution but not the optimal solution. So, in practice we can use the method if time is short, and we can use the method or retrieve in case library if time permits.

#### B. Structure Similarity calculation

Because the non-empty attributes of current case may not the same as that of historical case, so at first we should calculate structural similarity. in this paper, the collection of all problem variables is set to be  $P$ , the non-empty attributes set of the current case is set to be  $X$ , the non-empty attributes set of the historical case is set to be  $Y$ . Each attribute of one case is assigned different weight according to the level of importance of different attribute in decision-making process, the sum of weights of all attributes of one case is 1.

According to the different level of current case and historical case' importance, there are two method to calculate structural similarity. one method to view the current case the same important as historical cases, another method is to view the current case more important<sup>[6]</sup>.The calculated method of the two structural similarity is as follows:

1)  $Sim(X, Y)$  equals the sum of the attribute weights of the intersection of  $X$  and  $Y$  divides the sum of the attribute weights of the union of  $X$  and  $Y$ , the formula is shown as follows:

$$Sim(X, Y) = \frac{W_{X \cap Y}}{W_{X \cup Y}} = \frac{\sum_{i=1}^m w_i}{\sum_{j=1}^n w_j} \quad (1)$$

$Sim(X, Y)$  is structural similarity of  $X$  and  $Y$ , and  $w_i$  is the weight of the  $i$  th attribute in  $X \cap Y$ ,  $w_j$  is the weight of the  $j$  th attribute in  $X \cup Y$ ,  $m$  is the number of attributes in  $X \cap Y$ ,  $n$  is the number of attributes in  $X \cup Y$ .

2)  $Sim(X, Y)$  equals the sum of the attribute weights of the intersection of  $X$  and  $Y$  divides the sum of all attribute weights of  $Y$ , the formula is shown as follows:

$$Sim(X, Y) = \frac{W_{X \cap Y}}{W_Y} = \frac{\sum_{i=1}^m w_i}{\sum_{k=1}^l w_k} \quad (2)$$

$w_k$  is the weight of the  $k$  th attribute in  $X$ ,  $l$  is the number of attributes in  $X$ .

The scarcity of attributes refers to the degree of loss of the attribute values in the case, its calculation method is shown as follows:

$$\theta = \frac{W_{P-X}}{W_P} = 1 - W_X \quad (3)$$

In the formula,  $\theta$  is the scarcity of attributes;  $W_{P-X}$  is the sum of the attribute weights of difference sets of  $P$  and  $X$ ;  $W_P$  is the sum of the attribute weights of  $P$ , and the value is 1;  $W_X$  is the sum of the attribute weights of  $X$ .

#### C. Attribute Similarity calculation

In this paper, case attributes are divided into three types: determined numeric attributes, determined symbolic attributes, fuzzy attributes<sup>[7]</sup>. According to the different attribute types, we can use different attribute similarity calculation algorithms.

- Determined numeric attributes

Determined numeric attributes value is accurate number, in general it's value is real number or integer. There are a lot of methods to calculate similarity, this paper used the modified Hamming distance formula to calculate the attributes similarity of numeric attributes, the formula is:

$$Sim(x_i, y_i) = 1 - dist(x_i, y_i) = 1 - \frac{|x_i - y_i|}{\max_i - \min_i} \quad (4)$$

In the formula,  $Sim(x_i, y_i)$  is the attribute similarity of the  $i$  th number of  $X$  and  $Y$ ;  $x_i$  and  $y_i$  are respectively the attribute values of the  $i$  th number of  $X$  and  $Y$ ;  $\max_i$  and  $\min_i$  are respectively the maximum and minimum values of the  $i$  th number.

- Determined symbolic attributes

Determined symbolic attributes value are usually expressed in explicit terms which list all the possible values of the attribute, the value is a kind of enumerated values. determined symbolic attributes expressed value by terms, relationship between the attribute value doesn't exist quantity relations. similarity calculation formula is:

$$Sim(x_i, y_i) = \begin{cases} 1 & x_i = y_i \text{ or } x_i \subset y_i \\ 0 & x_i \neq y_i \end{cases} \quad (5)$$

- Fuzzy attributes

In this paper, fuzzy attributes include fuzzy concept attribute, fuzzy numbers attribute and fuzzy interval attribute. This paper uses fuzzy attribute similarity algorithm to calculate attributes similarity<sup>[8]</sup>. This paper uses trapezoidal fuzzy set to simulate fuzzy attributes, the shape function is shown as follows:

$$L(x) = R(x) = \max\{0, 1 - x\} \quad (6)$$

Membership function of Fuzzy set M is shown as follows:

$$\mu_M(x) = \begin{cases} L\left[\frac{m-\mu}{p}\right] & x \leq \underline{m} \\ 1 & \underline{m} \leq x \leq \bar{m} \\ R\left[\frac{x-\bar{m}}{q}\right] & x \geq \bar{m} \end{cases} \quad (7)$$

In the formula,  $\bar{m}$ ,  $\underline{m}$ ,  $p$ ,  $q$  are parameters. In the triangle fuzzy set,  $\bar{m}$  equals  $\underline{m}$ . The values of  $p$  and  $q$  vary according to different attributes. To the fuzzy concept attribute, the values of  $p$  and  $q$  are determined by domain experts; to the fuzzy numbers or fuzzy interval attribute, the value of  $p$  and  $q$  are usually taken as  $\frac{c\underline{m}}{p}$  and  $\frac{c\bar{m}}{q}$ , the default value of  $c$  is 0.1.

In this paper, relative area method is used to calculate the similarity between two fuzzy attribute, this method will calculate overlapping rate of union area of two membership functions, its advantage is accurate and simple. The formula is shown as follows:

$$Sim(x_i, y_i) = \frac{A(x_i \cap y_i)}{A(x_i \cup y_i)} = \frac{A(x_i \cap y_i)}{A(x_i) + A(y_i) - A(x_i \cap y_i)} \quad (8)$$

In this formula,  $A$  represents areas of the corresponding membership function of fuzzy set,  $x_i \cap y_i$  represents the intersection of two fuzzy set,  $x_i \cup y_i$  represents the union set of two fuzzy set. Fuzzy attribute similarity can be calculated according to the fuzzy attribute similarity algorithm.

#### D Comprehensive similarity calculation

As mentioned above, each attribute of one case is assigned different weight, the sum of weights of all attributes of one case is 1. Because attributes of  $X$  may be different from that of  $Y$ , when we calculate the similarity, we can only rely on their shared attributes, so we need to redistribute weight according to weight proportion of shared attributes, then calculate the weighted similarity of all attributes according to the similarity of redistribution weights and the similarity of each attribute. The formula is shown as follows:

$$Sim(X, Y) = \sum_{i=1}^n \left( \frac{w_i}{W_{X \cap Y}} Sim(x_i, y_i) \right) \quad (9)$$

In this formula,  $Sim(X, Y)$  represents similarity of  $X$  and  $Y$ .  $n$  represents the attributes number of intersection of  $X$  and  $Y$ .  $w_i$  represents the  $i$  th attribute weight of intersection of  $X$  and  $Y$ .  $W_{X \cap Y}$  represents the sum of attribute weights of intersection of  $X$  and  $Y$ .  $Sim(x_i, y_i)$  represents the  $i$  th attribute similarity between current case and historical case.

#### IV. APPLICATION OF EXAMPLES

This paper selects two coal mine gas explosion accident cases as examples to illustrate the specific application of case retrieval algorithm and to test its effectiveness. Firstly we set the current case as  $X$ , and set the historical case as  $Y$ , and calculate the structure similarity between  $X$  and  $Y$ , this paper considers current case as important as historical case. the sum of the weights of intersection attributes of  $X$  and  $Y$  is 0.9, and the sum of the weights of union attributes set of  $X$  and  $Y$  is 1. After being calculated, the weighted attribute similarity between case  $X$  and case  $Y$  is 0.883. comprehensive case similarity between case  $X$  and case  $Y$  is 0.795. Attribute information and related calculation results are shown in Table I.

TABLE I. THE INFORMATION OF GAS EXPLOSION ACCIDENT CASE ATTRIBUTES AND CALCULATION RESULTS

Case Attribute	Weight	Attribute Value of Case X	Attribute Value of Case Y	Attribute Similarity
accident level	0.1	major accident	large accident	0.6
gas volume fraction	0.15	14%	10%	0.9
CO volume fraction	0.1	0.3%		
CO2 volume fraction	0.1	11%	9%	0.9
collapse situation	0.1	local collapse	local collapse	1
ventilation condition	0.1	local interrupt	interrupt thoroughly	0.7

Communication condition	0.05	good	good	1
number of evacuated people	0.1	150~200	150~200	1
number of endangered people	0.1	32	47	0.9
whether cause fire or not	0.1	no	no	1

In this paper, threshold value is set 0.7. If case similarity value is higher than the threshold value, they are regarded as similar cases. At last, we compare case similarity value and pre-ordained threshold value, and we can see that similarity value of X and Y is higher than threshold value, so case X and case Y are regarded as similar cases, and decision makers can adopt solution of historical case Y as aid decision solution for current case X.

#### V. CONCLUSION

This paper mainly studies coal mine accident case representation and retrieval method. Firstly, this paper presents frame representation for emergency case according to the analysis of coal mine accident character. Then this paper proposes two level retrieval strategy based on cluster analysis, and designs two-layer case retrieval algorithm based on the structure similarity and attribute similarity, which improve the accuracy of case retrieval. Finally, the paper demonstrates realization process of case retrieval algorithm taking the coal mine gas accident cases as examples. This paper shows that it is better to retrieval similar cases from the history cases adopting case representation and retrieval method used in this paper, and the method is helpful for decision makers to solve current emergency decision problem.

#### REFERENCES

- [1] L. Wang, R. Zhang, Z. Zhao, W. Zhang and Y. Liu, "Emergency rescue command and management information system for coal mine," *Journal of Liaoning Technical University*, vol.25, No.5, Oct. 2006, pp. 655-657.
- [2] Z. Shi and W. Wang, *Artificial Intelligence*, National Defense Industry Press, 2007, pp. 212-218.
- [3] Z. Shi. *Advanced Artificial Intelligence*, Science Press, 2006, pp. 124-160.
- [4] J. Li, X. Chen, H. Liu, and X. Xie, "Researches on Case-Representation in Case-Based Reasoning System," *Journal of Hefei University(Natural Sciences)*, vol. 17, No. 3, Aug. 2007, pp. 26-29.
- [5] G. Zhu and Y. Xia, "Research and practice of frame knowledge representation," *Journal of Yunnan University(Natural Sciences)*, vol. 28, 2006, pp. 154-157.
- [6] Q. Zhong, S. Guo, X. Ye, and Y. Zhang, "Research on method of case representation and retrieval in emergency aid decision-making," *Journal of Dalian University of Technology*, vol. 51, No. 1, Jan. 2011, pp. 137-142.
- [7] Y. Zhang, Q. Zhong, X. Ye, and X. Qu, "Research on method of emergency aid decision-making based on CBR," *Application Research of Computers*, vol. 26, No. 4, Apr. 2009, pp. 1412-1415. Zhang and Y. Yu, "Hybrid Similarity Measure for Retrieval in Case-based Reasoning System," *Systems Engineering-theory & Practice*, Mar. 2002, pp. 131-136.