Application of Attribute Weight Determination and Attribute Similarity Calculation Based on Case Reasoning in Urban Rail Transit Cases

Jinbing Ha[a], Jing Cai[b, *], and Lianjie Xiao[c]

Department of Information Management, Nanjing University of Science &Technology, Nanjing 210094, China

[a]1802772568@qq.com, [b]905160572@qq.com, [c]1061939301@qq.com

*Corresponding author

Keywords: urban rail transit, emergency, case reasoning, attribute weight, similarity

Abstract: In the event of urban rail transit emergencies, mastering the scheduling experience of train operations will play an important role. The emergency command center database stores a relatively rich emergency response plan. When an emergency occurs, the on-site staff often lack sufficient experience. They cannot guarantee a complete and correct emergency response plan for the first time after receiving the police. This delays the train to some extent. During the adjustment period, it is easy to cause a large area of trains to be delayed, traffic lines are blocked, and the quality of rail transit operations services is degraded. In this paper, the theory of case-based reasoning is introduced, and the combined feature weight determination method, the case similarity calculation method based on Euclidean distance membership function and the similarity calculation method determined by weighted nearest neighbor algorithm are introduced in the urban rail transit emergency case database.

1. Introduction

Case-based reasoning is a new field of research that was originally proposed in the study of cognitive science. From the perspective of cognitive science, human beings can transmit the perceived content information to the brain for memory storage to form a modular knowledge structure. This knowledge structure provides a reference for solving new identical or similar problems encountered precedent. Therefore, the establishment of case-based reasoning is based on a summary of the understanding of the objective world: similar problems have the same or similar solutions.

2. Urban rail transit emergency case base representation

The urban rail transit emergency case library uses the framework representation to present case representations. Frame representation has the advantages of wide application scope, clear logical structure and flexible reasoning, and is one of the most common case representation methods. The framework usually adopts the <point-slot-value> ternary representation structure in the semantic network. A frame can be composed of several slots, and the slot can be decomposed into several sides as needed. Table 1 shows the types of slots, side items, side values and side values of some emergencies in urban rail transit operations.

3. Feature attribute weight determination

Feature attribute weights represent the order of importance of attributes in a case. Determining the feature attribute weight is critical to solving the local similarity and case global similarity of a single attribute indicator, which will affect the accuracy of the similarity calculation and the final matching result.

Table 1. Description of some urban rail transit emergency case libraries based on frame representation

Serial number	Slot	Side item	side value	Value type
1	Slot A: Incident Basic Information I	Side A_1: Incident Attributes	A_{11}: Time of occurrence	Value
			A_{12}: Site of occurrence	Enumerate
			…	…
		…	…	…
2	Slot B: Incident handling process P	Side B_1: Disposal schedule	B_{11}: Disposition start time	Value
			B_{12}: Time required for disposal	Value
			…	…
		…	…	…
…	…	…	…	…

From the historical research, there are more than ten methods for calculating the weight coefficient of common cases. According to the different data sources needed to calculate the weights, the weight determination method can be divided into subjective weight determination, objective weight determination and subjective objective combination to determine the weight. This article will focus on the method of determining the combined weights.

In this paper, the combined weight determination method is based on the weights determined by the AHP method, the entropy method and the gray correlation method to combine the weights. ω_{sub} and ω_{obj} are used to represent the subjective and objective weights respectively, and $\omega=\alpha\omega_{obj}+(1-\alpha)\omega_{sub}$ is used to determine the combined weight of the case attributes. Which indicates selection preference. Whenα >0.5, the weight of the case attribute is biased toward the objective weight when α<0.5, and the weight of the case attribute is biased toward the subjective weight sub. In this paper, α=0.5.

Using AHP and bribe method combined weight calculation method to calculate index weights, get the type, duration, influence range, train type, grouping, maximum speed, full load rate, station distance, curve radius, slope, speed limit, door, shielding Door, vehicle ATP/ATO, ground ATP/ATO attribute weights are: 0.28, 0.29, 0.021, 0.004, 0.003, 0.003, 0.003, 0.015, 0.004, 0.009, 0.003, 0.251, 0.132, 0.041, 0.024.

Using AHP and gray correlation degree combination weight calculation method to calculate index weights, get the type, duration, influence range, train type, editor, maximum speed, full load rate, station distance, curve radius, slope, speed limit, door, Screen door, vehicle ATP/ATO, ground ATP/ATO attribute weights are: 0.201, 0.345, 0.036, 0.003, 0.031, 0.014, 0.007, 0.018, 0.007, 0.004, 0.054, 0.148, 0.78, 0.030, 0.023.

4. Similarity calculation of eigenvalues

Case similarity calculation is the key in case retrieval. Designing a reasonable and effective similarity calculation method can effectively improve the speed and efficiency of case retrieval. The basic idea of solving the similarity between emergencies in this paper is to calculate the local similarity (Locality) and then weight the local similarity to obtain the global similarity (Global Similarity) according to the preset similarity. The threshold value is selected to select the most similar cases.

Table 2 is a partial case base of urban rail transit emergencies. In the case, 15 attributes of urban rail transit emergencies are selected. The attribute value types include numbers, intervals and enumerations. Each case in the case library corresponds to a train operation adjustment plan. There is a target case T_1, as shown in Table 3. The similarity between the target case T_1 and Case1 is calculated by combining the corresponding similarity calculation formula.

(1) The calculation of global similarity: in order to quantitatively analyze the degree of similarity between cases, it is necessary to comprehensively consider the global similarity of the four attributes. Different importance is different in the case for different attributes, so different weight values ω_i, i=1,2,⋯,15,$\omega_i\in[0,1]$ are assigned to each attribute. The global similarity function is expressed as follows:

$$\text{Global similarity(T, S)} = \sum_{i=1}^{n} Local\ \ Similarity(T_i, S_i) * \omega_i \quad (1)$$

Table 2. Partial Case Base of Urban Rail Transit Emergency

Case /Serial number	Scene type	Duration / s	Scope of influence / station	Train type	Grouping /section	Maximum speed / km / h	Full load rate/percentage
Case1	Signal device failure	[1,120]	[3,5]	A	6	100	0.88
Case2	Train failure	[120,300]	[4,6]	A	8	100	0.55

Case / Serial number	Station distance / m	Curve radius / m	Slope / degree	Speed limit / km / h	Door / performance	Vehicle ATP/ATO	Ground ATP/ATO
Case1	2000	489	0.02	80	0	0	1
Case2	2000	490	0.03	80	0	1	0

Table 3. Target Case T1

Case	Scene type	Duration	Scope of influence	Train type	Grouping	Maximum speed	Full load rate
T_1	Train failure	[300,600]	[4,6]	B	6	100	0.88

Case	Station distance	Curve radius	Slope	Speed limit	Door	Vehicle ATP/ATO	Ground ATP/ATO
T_1	1000	305	0.045	60	1	0	0

(2) Calculation of local similarity:

① Digital similarity calculation: Let the target case T and the source case S each contain x digital side values, the target case T has a digital side value a_i, and the source case S has a digital side value b_i, and $i=1,2,\cdots,x$, $a_i,b_i\in[U,V]$, U, V represents the lower and upper bounds of the i-th side value.Then the similarity between a_i and b_i is expressed as:

$$\text{Local Similarity}(a_i, b_i) = 1 - \frac{|b_i - a_i|}{V-U} \tag{2}$$

Use this formula to calculate the similarity between the target case T_1 and the "grouping", "maximum speed", "full load rate", "station distance", "curve radius", "slope", and "speed limit" in Case1.$Sim_{bz}(6,6)=1$,$Sim_{zdsd}(80,100)=0$,$Sim_{mzl}(0.72,0.88)=0.73$,$Sim_{zj}(1000,2000)=0.38$,$Sim_{qxbj}(305, 489)=0.22$, $Sim_{pd}(0.045,0.02)=0.72$, $Sim_{xs}(60,80)=0$.

② Interval similarity calculation:

Set the interval side value $x\in[a_1, a_2]$ in the target case T, interval type side value $y\in[b_1, b_2]$ in source case S, and a_1, a_2 ,b_1, $b_2\in[U^0, V^0]$, U^0, V^0indicates the lower and upper bounds of the side value x and the side value y. Then the interval similarity can be expressed as:

$$\text{Local Similarity}([a_1, a_2], [b_1, b_2]) = 1 - \frac{\iint_{a_1\le x\le a_2, b_1\le y\le b_2} |y-x|\, dy\, dx}{(a_2-a_1)(b_2-b_1)(V^0-U^0)} \tag{3}$$

According to this formula, the similarity between the "duration" and the "influence range" in the target case T_1 and Casel is calculated. $Sim_{cxsj}([300,600], [1,120])=0.36$, $Sim_{yxfw}([4,6], [3,5])=0.64$.

③ Enumeration similarity calculation:

Let all the enumerated side values in the target case T form the symbol sequence Q_T, and all the enumerated side values in the case library source case S form the symbol sequence Q_S, then the similarity between Q_T and Q_S is expressed as:

$$\text{Local Similarity}(Q_T, Q_S) = \frac{\sum_{i=1}^{y} p_i}{y} \tag{4}$$

$$p_i = \begin{cases} 0, q_i^T \neq q_i^S \\ 1, q_i^T = q_i^S \end{cases} \tag{5}$$

In the formula, q_i^T represents the i-th symbol of the Q_T corresponding symbol sequence, and q_i^S represents the i-th symbol of the Q_S corresponding symbol sequence.

The similarity between "scenario type", "train type", "carport", "shield port", "vehicle ATP/ATO", and "ground ATP/ATO" in target case T_1 and Case1 is calculated according to this formula.

Sim_{cjlx}(train failure, Signal device failure)= Sim_{lclx}(B,A)= Sim_{cm}(1,0)= Sim_{pbm}(0,1)=0.

$Sim_{czATP/ATO}$(0,0)= $Sim_{dmATP/ATO}$(0,0)=1.

Combining the nearest neighbor algorithm, AHP, AHP and entropy method combined weight calculation method, AHP and gray association weight calculation method are used respectively, and the similarity between target case T_1 and Case1 is calculated as: $Sim_{AHP}(T_1$,Case1)=0.26, $Sim_{AHP+entropy}(T_1$,Case1)=0.17, $Sim_{AHP+gray}(T_1$,Case1)=0.25.

Similarly, using three different weight determination methods, the three cases with the highest similarity to the target case T_1 in the case base can be solved. The three cases using the AHP method were Case77, Case90 and Case94, and the similarities were 0.84, 0.82, and 0.81, respectively. The three cases obtained by using AHP and the entropy method combined weight calculation method are Case77, Case90 and Case53, and the similarities are 0.83, 0.82, and 0.80, respectively. The three cases obtained by AHP and gray correlation weight calculation method are Case95, Case90 and Case74, and the similarities are: 0.87, 0.86, and 0.84.

The definition of precision is the ratio of "retrieving the number of cases with similarity greater than or equal to 0.8" and "retrieving the total number of cases". Select 100 cases to input 10 times as a batch of 10 cases as the initial case in the case library, set the similarity threshold lower bound to 0.7, select different weight determination methods, and search and compare. The comparison results are shown in Table 4.

Table 4. Comparison of search efficiency when the number of cases is 100

Weight determination method	Case similarity ≥0.8	the total number of retrieved cases	precision	Maximum similarity
AHP	4	7	0.57	0.84
AHP + Entropy	3	7	0.43	0.83
AHP + Gray correlation	5	8	0.63	0.87

It can be seen that the three weight calculation methods can retrieve the case with the highest similarity from the case base. The AHP and gray association weight calculation method not only retrieves the similarity of the case, but also effectively reduces the missed detection.

5. Conclusion

This paper mainly introduces the basic principles of case-based reasoning, analyzes the case description strategy in case-based reasoning, focuses on the key techniques in case-based reasoning, and builds urban rail transit on the basis of case representation of urban rail transit emergencies. Issue the event case base model. In the case of case attribute weight design, the combined weight method is used to calculate the case attribute weight value, and the case-based reasoning of the target emergency case is beneficial to the urban rail transit operation system to make organizational adjustments to the train operation more quickly and effectively. It is more theoretical and practical to normalize the transportation of the line.

References

[1] Jovanovic D, Harker P T. DECISION SUPPORT SYSTEM FOR TRAIN DISPATCHING: AN OPTIMIZATION-BASED METHODOLOGY [J]. Transportation Research Record Journal of the Transportation Research Board, 1991, 1314: 31-40.

[2] Ashley K D. Arguing by Analogy in Law: A Case-Based Model [M] // Analogical Reasoning. Springer Netherlands, 1988.

[3] Li J, Jiang C, Ma F, et al. A New Translation Strategy Based-on Case-Based Reasoning in Machine Translation System [C] // World Congress on Intelligent Control & Automation. IEEE, 2006.

[4] Nishikant M, Sanja P, Santhanam S. A self-adaptive case-based reasoning system for dose planning in prostate cancer radiotherapy [J]. Medical Physics, 2011, 38 (12): 6528-6538.

[5] Goh Y M, Chua D K H. Case-based reasoning for construction hazard identification: Case representation and retrieval [J]. Journal of Construction Engineering & Management, 2009, 135 (11): 1181-1189.

[6] Lu Y, Li Q, Xiao W. Case-based reasoning for automated safety risk analysis on subway operation: Case representation and retrieval [J]. Safety Science, 2013, 57 (Complete): 75-81.

[7] Weber B, Mateas M. Conceptual neighborhoods for retrieval in case-based reasoning [C] // International Conference on Case-based Reasoning: Case-based Reasoning Research & Development. Springer-Verlag, 2009.