

Research on node importance evaluation of railway logistics distribution network based on Rough Set

LIU Xiu-miao

{LiuxiuMiao22@163.com}

HEBEI VOCATIONAL COLLEGE OF RAIL TRANSPORTATION, Shijiazhuang 050000, China

Abstract: The traditional evaluation method of the node importance of railway logistics distribution network has not eliminated redundant data, resulting in errors in the evaluation results. Therefore, the evaluation method of the node importance of railway logistics distribution network based on rough set is proposed. This method uses the rough set theory to eliminate redundant node data, obtain the collection and distribution uncertainty index, establish an importance evaluation index, obtain the node importance evaluation weight, and achieve the higher accuracy of the network node importance evaluation. The experimental results show that compared with the traditional evaluation method, the evaluation results of network node importance based on rough set are closer to the expected results. It can be seen that the performance of this method is superior.

Keywords: rough set; collection and distribution network; node importance

1 Introduction

With the continuous development of national economy and the continuous optimization of industrial structure adjustment, China's railway transportation industry has grown rapidly. In order to improve the transportation efficiency and save the transportation cost, the railway logistics collection and distribution requires the importance evaluation of the network nodes, so as to set up the railway transportation scheme.

According to this requirement, the traditional evaluation method directly evaluates the node importance of railway logistics with complex road conditions and overlapping lines, and the railway transportation department formulates the collection and distribution scheme according to the evaluation results. However, as the length and complexity of the route become longer, there are some drawbacks in the scheme under the design of the evaluation results. The elimination of redundant data is not complete, resulting in errors in the evaluation results.

Therefore, in view of the shortcomings of the traditional methods, this paper puts forward the evaluation method of node importance of railway logistics collection and distribution

network based on rough set [1]. As a reliable technology to deal with uncertainty, rough set theory is of great significance to remove redundant nodes in the network. This method not only solves the existing problems of the traditional evaluation method, but also improves the accuracy of the evaluation results to a new level, ensures the feasibility of the implementation of the railway logistics collection and distribution scheme, and provides strong technical support for the development of the national railway transportation industry.

2 Evaluation of node importance of railway logistics distribution network based on Rough Set

2.1 obtaining uncertainty index of collection and distribution based on Rough Set

On the premise of using rough set theory to keep the classification ability unchanged, the redundant attributes of railway logistics network nodes are eliminated, and the hidden potential laws in the data are explored. Given an information system dimension $IS=(W, K)$, where W is the domain and K is the set of conditional attributes on the domain W . For any given conditional attribute $k \in K$, there exists a function $k:W \rightarrow T_k$, T_k expressed as a value field with attribute k . Each element in W is called an individual, an object, or a row.

For any given attribute subset $I \subseteq K$ and any given individual $x \subseteq K$, the following information functions are satisfied:

$$Inf_I(x) = \{(k.k(x)): k \in I\} \quad (1)$$

Suppose that a set of conditional attributes on the domain W is K , if there are $I \subseteq K$ and $I \neq \emptyset$, then $I \cap K$ is an equivalent relation on domain W , which is called the indiscernible relation on I , and it is recorded as $IND(I)$. According to the above, we can know the concept $X \subseteq W$ of network node and the attribute subset $I \subseteq K$. Then we can define the upper and lower approximate domain of the concept X :

$$\begin{cases} \bar{I}(X) = \bar{I}(IS, X) = \{x | (x \in W) \wedge ([x]_I \cap X \neq \emptyset)\} \\ \underline{I}(X) = \underline{I}(IS, X) = \{x | (x \in W) \wedge ([x]_I \subseteq X)\} \end{cases} \quad (2)$$

It can be concluded that the set $IN_I(X) = \bar{I}(X) - \underline{I}(X)$ is a I boundary region of X ; the set $POS_I(X) = \underline{I}(X)$ is a I positive region of X ; the set $NEG_I(X) = W - \bar{I}(X)$ is a I negative region of X [3]. Figure 1 below shows the

upper and lower approximation domain, boundary domain and negative region control diagram of set X under the control of rough set.

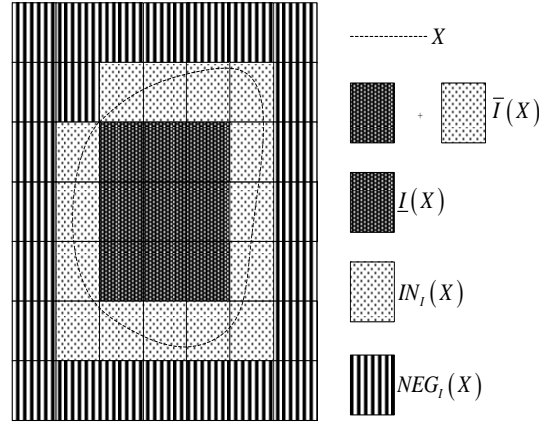


Fig. 1. Schematic diagram of upper and lower approximation domain, boundary domain and negative region

Using the parameters in the figure above, the attributes of railway logistics collection and distribution network nodes are reduced. Through the upper and lower approximation domain, boundary domain and negative region control of the set, the final region is delineated in the display area of the graph, reducing the possibility of nodes' disorderly connection, so as to reduce the attributes of network nodes. Attribute reduction needs to keep the classification ability of database unchanged and delete unnecessary redundant attributes. Suppose that given a database $M = (W, S)$ and an equivalent relation family $G \in S$ in the database, for any $H \in G$, if there are:

$$IND(G) = IND(G - \{H\}) \quad (3)$$

When the above formula holds, it is unnecessary to call logistics data H as G , otherwise it is necessary to call logistics data H as G . If it is necessary for each $H \in G$, H to be G , G is called independent, otherwise G is called dependent or not independent, so as to realize attribute reduction of logistics network nodes [3].

Rough set is a method to deal with the uncertainty problem, so the uncertainty index is measured for the uncertainty of the nodes in the distribution network of iron flow logistics. It is known that in the information system $IS = (W, K)$, there is a condition $X \subseteq W$, then under the condition of attribute subset $I \subseteq K$, the membership calculation expression

of $x \in W$ relative to concept $X \subseteq W$ is:

$$\varphi(I, X, x) = \frac{|[x]_I \cap X|}{|[x]_I|} \quad (4)$$

Where $|\cdot|$ represents the potential of the set. If there is a conceptual condition $X \subseteq W$ in the logistics collection and distribution decision system $DS = (W, K, a)$, which is called decision attribute a depends on the condition attribute set K with degree $h(0 \leq h \leq 1)$, then:

$$h = \lambda(K, a) = \frac{|POS_K(a)|}{|W|} \quad (5)$$

Where λ is the degree of dependence. According to the formula and formula (4), the uncertainty index is obtained :

$$\beta = (I, X) 1 - \frac{\varphi}{\ln h} \times \left| \kappa \cdot \frac{I(X)}{I(X)} \right| \quad (6)$$

In the formula, κ represents the influence factor of the uncertainty degree index; β is the uncertainty index of the logistics network node [4].

2.2 Establishment of importance evaluation index

According to the structural characteristics of railway logistics collection and distribution network, the evaluation indexes of the node importance are established based on the central characteristics and structural hole characteristics of network nodes. According to the idea of constructing the contribution matrix of node importance, a new method of node importance ranking is proposed. This method uses the importance index value of the structure hole and the k core of the adjacent nodes to get the importance contribution relationship between the nodes. At the same time, the k core importance of the nodes is used to describe the global position information of the nodes. Based on the analysis of the local importance of the network, combined with the global importance, the importance evaluation index of the nodes is more comprehensive and the evaluation result is more accurate True. From a sociological point of view, structural holes exist in the gap between data without redundant connections, as shown in Figure 2 [5].

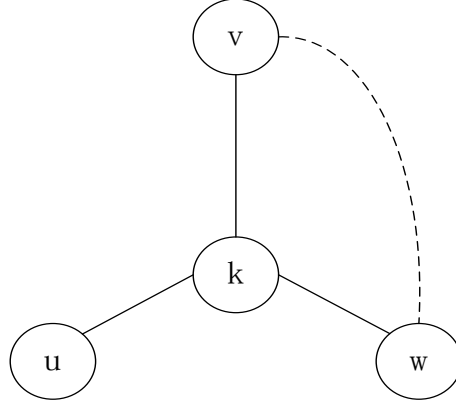


Fig. 2. Schematic diagram of structural hole

After the rough set processing in the previous step, there is no redundant connection between u and v in the figure. The logistics data with structural holes between them can bring cumulative rather than overlapping network route planning. As can be seen from Figure 2, node k has more network revenue than its neighbor nodes u , v and w , that is, in the network, node k is more important than other nodes. However, if nodes v and w are connected, the network revenue of node k will be reduced. From the point of view of complex network, the more structural holes a node has, the more conducive to information dissemination.

Let $Q = (A, B)$ be a logistics collection and distribution undirected network without self loop, which has n nodes and m edges, of which $A = \{a_1, a_2, \dots, a_n\}$ is the set of all nodes in the network and $B = \{b_1, b_2, \dots, b_m\}$ is the set of edges between nodes. The adjacency matrix is recorded as $C_{n \times n} = (d_{ij})_{n \times n}$, where:

$$d_{ij} = \begin{cases} 1, & i \text{ connected to } j \\ 0, & i \text{ is not connected to } j \end{cases} \quad (7)$$

According to the above formula (7), the degree of node i can be recorded as:

$$f(i) = \sum_{j \in Q} d_{ij} \quad (8)$$

According to the degree of node, the adjacency degree of node i can be expressed as:

$$h(i) = \sum_{\gamma \in \Upsilon(i)} k(\gamma) \quad (9)$$

In the formula, $\Upsilon(i)$ represents the i set of adjacent nodes, and γ represents the adjacency influence parameter. It is known that the input energy of node evaluation index i to evaluation node j can be shown in Figure 3 [6].

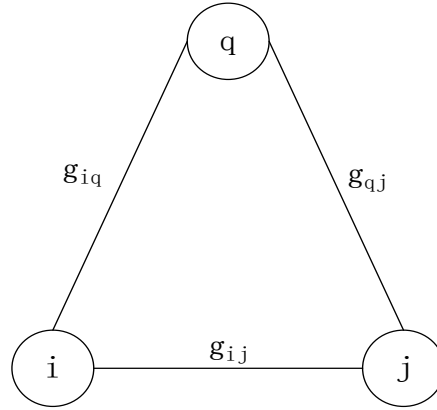


Fig. 3. Schematic diagram of evaluation index input energy

According to the above figure, the network constraint coefficient can measure the constraints of adjacent nodes when the nodes in the network form the structural hole, and it is an indicator to measure the structural hole. The smaller the constraint coefficient is, the greater the degree of structural hole is, and the more important the nodes are. Then the constraint coefficient can be expressed as:

$$RC_i = \sum_{j \in \Upsilon(i)} \left(g_{ij} + \sum_q g_{iq} g_{qj} \right)^2 \quad (10)$$

Where: q represents the number of common neighbors of node i and node j . Considering the influence of the node degree and the topology of the number of neighbors on the node, the input energy of node i to node j can be recorded as g_{ij} :

$$g_{ij} = \frac{RC_i Q(j)}{\sum_{s \in \Upsilon(i)} Q(s)} \quad (11)$$

Where: $Q(j)$ represents the separation function of node j ; $Q(s)$ represents the separation function of the desired node similar to the attribute of node j . According to the formula, the evaluation indexes of importance are as follows:

$$p_i^e = g_{ij} (p_i^z + \alpha p_i^m) \quad (12)$$

In the formula: p_i^e represents the evaluation index with the evaluation degree of e ; p_i^z represents the evaluation standard index with the constraint amount of z ; p_i^m represents the evaluation reference value with the update times of m ; α represents the controllable parameters of the node to achieve the accurate establishment of the evaluation index of the node importance [7].

2.3 Determine the evaluation weight and evaluate the importance of network nodes

According to the regional characteristics of the railway logistics network, taking the average degree of nodes, the centrality of mediators and the rate of message forwarding of the whole network as the attributes of the evaluation key nodes, and combining the evaluation importance index, the evaluation weight of the network nodes is calculated, so as to realize the evaluation of the importance of the network nodes.

According to the subjective information of logistics transportation, AHP is used to weight the importance evaluation indexes of each attribute. On the basis of determining the decision-making objectives and evaluation attributes, a suitable hierarchical structure model is built for the next calculation; Establish the judgment matrix, assume $f = \{f_1, f_2, f_3, \dots, f_n\}$ is the set of decision attributes, D is the goal, s_{ij} is the value of f_i for the quantitative expression of the relative importance of f_j [8]. By comparing the relative importance of the attributes between the two, the judgment matrix is established as follows:

$$S = \begin{bmatrix} 1 & s_{12} & s_{13} & \cdots & s_{1m} \\ s_{21} & 1 & s_{23} & \cdots & s_{2m} \\ s_{31} & s_{12} & 1 & \cdots & s_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & s_{m3} & \cdots & 1 \end{bmatrix} \quad (13)$$

In the above formula: s represents the node attribute in the logistics collection and distribution network. According to the above matrix, calculate the attribute weight of the evaluation node. The judgment matrix S is calculated, and the accumulation of elements in each row is R_j :

$$R_j = \prod_{i=1}^m s_{ij} \quad (14)$$

Where, $j=1,2,\dots,m$. The m power root of R_j is calculated to obtain w_j :

$$w_j = \sqrt[m]{R_j} \quad (15)$$

According to the above formula, the weight \bar{w}_j corresponding to attribute f_j is obtained by normalizing $w = (w_1, w_2, w_3, \dots, w_m)$ vectors:

$$\bar{w}_j = \frac{w_j}{\sum_{c=1}^m w_c^2} \quad (16)$$

Where: \bar{w}_j represents the weight of the evaluation of the importance of the network nodes under the control of the boundary constraint c [9]. Use the consistency index test method to test the consistency of the weight. The following formula (17) is the consistency index test result:

$$\begin{cases} CI = \frac{\zeta_{\max} - m}{m - 1} \\ CR = \frac{CI}{RI} \end{cases} \quad (17)$$

In the above formula: ζ_{\max} represents the maximum consistency of the weight index; CR represents the consistency index test coefficient; RI represents the average consistency index of the weight, and the value of this index can be determined according to table 1.

Table 1. RI list of values

Judgment matrix dimension	<i>RI</i>	Judgment matrix dimension	<i>RI</i>
1	0.00	6	1.24
2	0.00	7	1.32
3	0.58	8	1.41
4	0.96	9	1.45
5	1.12	-	-

Because the judgment matrix is generated based on subjective judgment, and there may be inconsistencies in the judgment matrix, CR is introduced to check whether the judgment matrix is consistent. If $CR = 0.1$ is calculated, it is considered that the matrix is completely consistent; if $CR < 0.1$, it is considered that the matrix basically meets the requirements; if $CR \geq 0.1$, it is considered that the matrix does not meet the consistency requirements, and the above calculation must be repeated until the consistency requirements are met [10]. At this point, the weight can be used to evaluate the importance of railway logistics network nodes based on rough set.

3 Experimental test

The evaluation method proposed in this paper is compared with the traditional evaluation method, and the evaluation effect of the importance degree of railway logistics collection and distribution network nodes is compared through test experiments. Set up the experimental test platform, select a section of area as the basic environment for the experimental test, using MSE to measure the difference between test results and expectations. And Figure 4 below is the schematic diagram of the network topology of the railway logistics collection and distribution area selected for the experimental test.

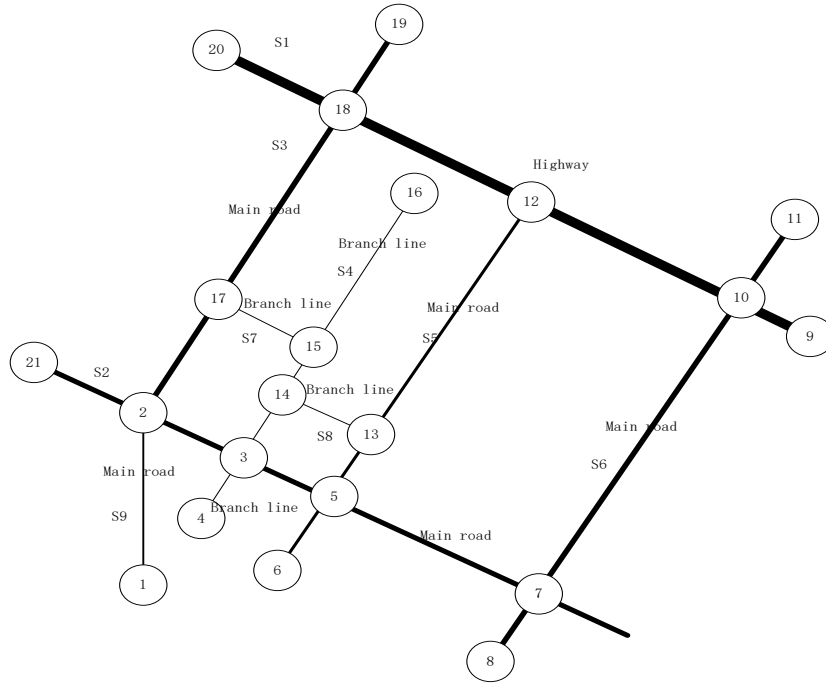


Fig. 4. Network topology of experimental test area

According to the railway collection and distribution line in the figure above, the road area network control nodes are set to connect the basic information of the road with each node, as shown in Table 2 below.

Table 2. Basic information of sections in the area

Link number	Target	Road grade	Link length
S1	5	Highway	700
S2	5	Main road	700
S3	4	Secondary road	650
S4	5	Secondary road	600
S5	4	Secondary road	645
S6	4	Main road	710
S7	2	Branch road	220

S8	2	Branch road	240
S9	2	Branch road	350

It is known that there are 22 nodes in the above regional network. In this experiment, nodes 1 to 10 are taken as the importance evaluation objects. The data in Table 3 below is the basic road data of these nodes.

Table 3. Road data information at nodes

Node number	Traffic	Capacity	Saturation
1	5301	8943	0.62
2	4049	6950	0.67
3	5486	9528	0.58
4	6792	9516	0.73
5	12485	16827	0.79
6	11397	12795	0.82
7	2495	4910	0.5
8	1046	1795	0.65
9	779	1204	0.68
10	11834	6133	0.68

Based on the above data as the experimental basic data, it is assumed that the expected value of the evaluation result is shown in Figure 5 below.

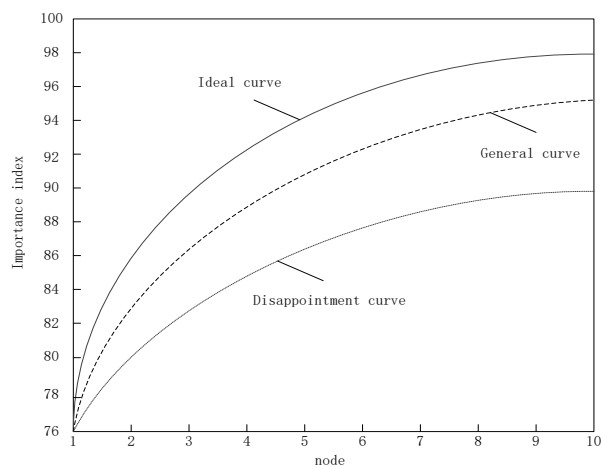


Fig. 5. Expected test results

Taking the data in Figure 5 as the experimental reference value, two evaluation methods are used to evaluate the importance of network nodes, among which the test results of the proposed evaluation method are group A, the test results of the traditional evaluation method are group B, and Figure 6 is the comparison results of the five tests of this experiment.

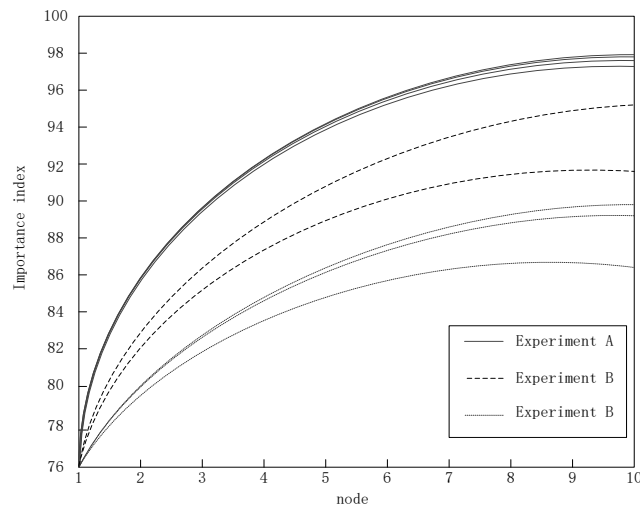


Fig. 6. Experimental test comparison results

According to the above figure, compared with the expected test results, the test results of the proposed evaluation method are very similar to the expected curve; while the evaluation results of the traditional evaluation method are similar to the general curve twice and the disappointed curve thrice, which shows that the evaluation effect of the network node importance evaluation method based on rough set is better.

4 Conclusion

Based on the evaluation method of rough set, the redundant data in the network node is eliminated by rough set theory, and the evaluation weight is obtained according to the established importance evaluation index, so as to realize the more regional and comprehensive importance evaluation of railway logistics network node. The test results of the evaluation method are very similar to the expected curve. However, this evaluation method is proposed for the collection and distribution of railway logistics, which may not be suitable for the evaluation of the importance of other transport network nodes, so it can not be used blindly.

Reference

[1] Xu Fei, Ren Yongtai, Yi Ke-han.: Study of Evaluative Method of Graduate Training Quality Based on

- Interval-valued Fuzzy Rough Set Theory. *Mathematics in Practice and Theory*, Vol. 48, no. 11, pp. 1-8 (2018)
- [2] Li Linxiao, Dong Yu.: Track Circuit Fault Diagnosis Model Based on Principal Component Heuristic Algorithm. *Railway Standard Design*, Vol. 63, no. 9, pp. 169-173 (2019)
- [3] Yang Rui, Su Dan, Yan Lei, et al.: Quick Fault Diagnosis Method for Power Information Communication System Based on Rough Set and Bayesian Network. *Bulletin of Science and Technology*, Vol. 34, no. 7, pp. 96-100 (2018)
- [4] Tu Min, Zhou Zhiyuan.: Node Location of Container Underground Logistics System. *Logistics Technology*, Vol. 38, no. 6, pp. 61-68 (2019)
- [5] Yin Rongrong, Yin Xueliang, Cui Mengdi, et al.: Node Evaluation Method Based on Importance Contribution in Scale-free Networks. *Journal of Software*, Vol. 30, no. 6, pp. 1875-1885 (2019)
- [6] Sun Houfei, Sun Xuefeng, Han Jianli.: Evaluation of Importance of Networked Air Defense Nodes Based on Rough Sets. *Ship Electronic Engineering*, Vol. 39, no. 6, pp. 34-37+60 (2019)
- [7] Mo Zhenchun, Fu Lihua, Peng Yaohui, et al.: Identification of Ecological Space Network Key Nodes Based on a Comprehensive Importance Evaluation. *Journal of Hunan University of Technology*, Vol. 32, no. 2, pp. 64-69 (2018)
- [8] Li Guiju, Huang Hongguang, Shu Qin.: Evaluation method for node importance in power telecommunication network based on interdependent network theory. *Power System Protection and Control*, Vol. 47, no. 11, pp. 143-150 (2019)
- [9] Feng Yu, Ge Ju.: Structure Nodes of Xinjiang Air Transport Network Using Entropy Weight and TOPSIS. *Logistics Sci-tech*, Vol. 42, no. 6, pp. 106-109 (2019)
- [10] Liang Jingjing, Wang Hongyu.: Simulation of the Importance of Information Nodes in Wireless Communication Networks. *Computer Simulation*, Vol. 35, no. 7, pp. 443-446 (2018)