Train Operation Adjustment Decision-making Based on Rough Sets and Information Entropy Theory

WEN Chao[1], PENG Qi-yuan[1], CHEN Yuhong[1], CHEN Dalian[2]

1. Dept. of Traffic and Transportation, Southwest Jiaotong University, China, Chengdu, 610031
E-mail:wenchao0601@126.com

2. Sichuan JingAn Vocational College, China, Chengdu, 610031
E-mail：chenlian8281133@yahoo.com.cn

Abstract: Train operation Parameters were defined by using rough sets theory. The train operation decision table was established on the bases of the real-time data which was acquired from the actual train operation graphs in the train operation dispatching system. The information entropy theory was used in the process of analyzing each object of the decision table, and the function of each condition attribute for reducing decision-making uncertainty was studied. Then, the importance ranking of the train operation attributes were determined. The work can provide a reference to the dispatchers during train operation adjustment.

Key Words: Train operation adjustment, Rough sets, Information entropy, Decision table, Uncertainty

1 INTRODUCTION

There will be many stochastic and paroxysmal events during the process of train operation that lead to trains warping from transportation scheme. With the extending of running distance, the probability of being interfered will increase and it often appears to increase train operation delay rate. Then, in order to return to the normal operation and realize "running trains according to diagrams" in Chinese railway system, the real-time train operation adjustment measures need to be taken by the train operation organizers.

The studies about train operation adjustment abroad nearly focus on the topic of optimization of train graph, which deals with the problem of train operation in the way of offline（C.W. Tsang, T. K. Ho, 2003. J, Yuan, I. A. Hasen. 2002） In China, there is much attention paid to the real-time train operation model. The optimization objective such as maximum train punctuality rate(CAO Jiaming,1995), minimum total train delay time(WANG Zhengbin, DU Wen, 2004), the increase of average travel speed(LUO Qing,JIN Fucai, HU Siji, 2004), the increase of satisfaction degree of passengers(CHEN Yanru, PENG Qiyuan, JIANG Yang-sheng, 2003) and many optimization algorithms are designed.

The level of automatic traffic control in Chinese railway system is not high enough at present, moreover, the traffic density is large and there are many kinds of trains running on the lines. So the dispatchers need to consider much irregular decision conditions during train operation adjustment decision-making. The existing studies can not resolve the key issues of train operation adjustment and decision-making will become a complicated variable because of the diverse skill level of the train operation commanders.

The article (WANG Minghui, 2004) quantized the main real parameters of train operation and defined the condition attribute and decision attribute using rough sets theory. In this paper, a decision-making table for train operation is established and the function of each condition attribute for reducing decision-making uncertainty was studied by using information entropy theory. Then, the importance ranking of the train operation attributes were determined. The dispatchers may make train operation adjustment scheme in according with the need of train operation adjustment.

2 ROUGH SETS EXPRESSIONS OF THE TRAIN OPERATION ADJUSTMENT DECISION-MAKING PARAMETERS

Train operation adjustment is a process that makes a new decision based on existing experiences and trains' real-time running parameters such as train's rank, train's punctually degree and performance level of trains' journey and so on. As the rough sets theory is a new method to deal with incomplete and uncertain data, the procedure of using it to process data is very similar to that for the dispatchers to make decision during train operation commanding. This article presents seven conditional attributes which include relative train priority R_{AP}, the relative train punctuality R_{RT}, the relative train-covered distance R_{TC}, the train section margin G_B, the station train-stop margin G_S, the follow-on train's new departure delay D_D, the follow-on train's new arrival delay D_A. Then a decision attribute is given A_G. Giving the range of each parameter, there are $R_{AP} \in \{0,1,2\}$, $R_{RT} \in \{1,2,3,4,5\}$, $R_{TC} \in \{0,1,2\}$, $G_B \in \{0,1,2\}$, $G_S \in \{0,1,2,3,4\}$, $D_D \in \{0,1,2,3\}$, $D_A \in \{0,1,2,3,4\}$, $A_G \in \{0,1\}$.

This work is supported by National Natural Science Foundation of China under grant 60776827 and Railway Ministry of China under grant 2007X001-D.

3 DECISION-MAKINGTABLE FOR TRAIN OPERATION ADJUSTMENT

Train operation adjustment is a decision-making issue in management science. So, a Train operation problem can be manifested by a two-dimensional decision-making table. And a knowledge system S can be an equation expression form of the two-dimensional decision-making table.

$$S=(U,\ C,\ D,\ V,\ f) \quad 1$$

Where $U=\{x_1, x_2, x_i \cdots x_n\}$ is a non-empty finite set of objects called the universe and x_i is an object that corresponds an independent train operation record. The non-empty finite set C is a conditional attribute that consists of train operation attributes and there is $C=\{a_1, a2\ldots an\}$. D is the set of decision attributes and $D=\{d1, d2\ldots d_n\}$ is the decision results of train operation. C together with D forms the attributes of a two-dimensional decision-making table, and $C\cap D=\Phi$. V refers to the values of conditional attributes and decision attributes. And f defines an information function that specifies a value for xi in set U. In the decision-making table, each train operation record is divided into different decision class by conditional attributes and decision attributes. A train operation adjustment decision-making table is showed as table 1 based on the discreted and normalized data which was acquired from the actual train operation graphs in the train operation dispatching system.

Table1. Train operation adjustment decision-making table

U	R_{AP}	R_{RT}	R_{TC}	G_B	G_S	D_D	D_A	A_G
x_1	0	0	2	1	0	1	0	0
x_2	0	0	0	1	0	2	1	0
x_3	0	2	2	0	0	0	0	1
x_4	2	3	1	2	0	2	1	1
x_5	2	2	2	0	1	1	1	1
x_6	0	4	0	0	0	1	0	0
x_7	0	4	0	1	1	0	2	0
x_8	0	4	0	0	1	0	1	1
x_9	0	4	0	1	1	1	2	1
x_{10}	0	4	2	1	2	1	0	1
x_{11}	2	1	2	1	0	2	1	0
x_{12}	0	4	2	0	1	2	1	1
x_{13}	2	1	2	0	1	0	2	0
x_{14}	2	1	0	0	1	1	0	0
x_{15}	2	1	2	1	0	0	1	0
x_{16}	1	2	2	0	1	1	2	1
x_{17}	1	0	2	1	2	2	1	1
x_{18}	1	0	2	1	1	0	2	1
x_{19}	0	4	2	1	0	0	1	1
x_{20}	1	0	2	0	1	1	1	1

4. IMPORTANCE RANKING OF TRAIN OPERATION CONDITION ATTRIBUTES

The concept of information entropy was proposed by Shannon in 1948, and it is an effective method that dealing with statistical measure of uncertain information. Information entropy H can be expressed as follows.

$$H=-k\sum_{i=1}^{m} p_i \log_m(p_i) \quad 2$$

Where

The value of constant k is 1, and p_i stands for the probability of state i of the system.

If a system is divided into K (K>1) parts that one part stands for a sub-system, the expected value of information entropy can be expressed as follows.

$$H=\sum_{j=1}^{K} p_j H_j \quad 3$$

Where

p_j denotes the occurred probability of sub-system j;

H_j denotes the entropy of sub-system j;

Due to the classification subset of A_G is $D=\{D_1,D_2\}$, the occurred probability of class "0" is $p_0=8/20=0.4$, and the occurred probability of class "1" is $p_1=12/20=0.6$, the information entropy of decision attribute A_G can be calculated.

$$\begin{aligned} H_D &= -\sum_{i=1}^{m} p_i \log_m(p_i) \\ &= -(0.4\log_2 0.4 + 0.6\log_2 0.6) \\ &= 0.972 \end{aligned}$$

Where

$D_1=\{x_1, x_2, x_6, x_7, x_{11}, x_{13}, x_{14}, x_{15}\}$;

$D_2=\{x_3, x_4, x_5, x_8, x_9, x_{10}, x_{12}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}\}$.

Considering of the train operation adjustment conditional attribute of "relative train priority R_{AP}", it can be divided into 3 sub-systems by the attribute value. The expected value of information entropy that bases on attribute R_{AP} can be expressed as follows.

$$H_{R_{AP}}=\sum_{j=1}^{3} p_j H_j = p_1H_1+p_2H_2+p_3H_3$$

The parameters $p_1=0.5$, $p_2=0.2$ and $p_3=0.3$ respectively refer to the probability that each train operation record belongs to the equivalence class due to the value of attribute "R_{AP}". H_1、H_2、H_3 are respectively the entropy of the 3 sub-systems.

So

$$\begin{aligned} H_1 &= -\sum_{i=1}^{2}(D_i|R_{AP1})\log_m(D_i|R_{AP1}) \\ &= -\left(\frac{4}{10}\log_2(\frac{4}{10})+\frac{6}{10}\log_2(\frac{6}{10})\right) \\ &= 0.972 \end{aligned}$$

$$\begin{aligned} H_2 &= -\sum_{i=1}^{2}(D_i|R_{AP2})\log_m(D_i|R_{AP2}) \\ &= -\left(0\log_2 0+1\log_2 1\right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} H_3 &= -\sum_{i=1}^{2}(D_i|R_{AP3})\log_m(D_i|R_{AP3}) \\ &= -\left(\frac{4}{6}\log_2(\frac{4}{6})+\frac{2}{6}\log_2(\frac{2}{6})\right) \\ &= 0.918 \end{aligned}$$

Thus

$$H_{R_{AP}} = p_1 H_1 + p_2 H_2 + p_3 H_3$$
$$= 0.5\times 0.972 + 0.2\times 0 + 0.3\times 0.918$$
$$= 0.761$$
$$H_D - H_{R_{AP}} = 0.972 - 0.761 = 0.211$$

The condition attribute "R_{AP}" can reduce the decision-making uncertainty by 0.211 for classifying U. In the same way, the function of other condition attributes for reducing train operation decision-making uncertainty can be summed up in table 2. H_{Ci} in the table is the information entropy of each train operation conditional attribute.

Table2. Function of each condition attribute for reducing train operation decision-making uncertainty

	R_{AP}	R_{RT}	R_{TC}	G_B	G_S	D_D	D_A
H_1	0.972	0.972	1	0.918	0.954	0.985	0.972
H_2	0	0	0	1	0.778	0.954	0.881
H_3	0.918	0	0.891	0	0	0.972	0.972
H_4	—	0	—	—	—	—	—
H_5	—	0.863	—	—	—	—	—
H_{Ci}	0.761	0.545	0.879	0.913	0.771	0.969	0.926
H_D-H_{Ci}	0.211	0.427	0.093	0.059	0.201	0.003	0.046

The smaller the information entropy is, the larger of the function of reducing decision-making uncertainty will be. And the conditional attribute will be more important. From table 2, there will be an importance ranking of all conditional attributes during train operation adjustment decision-making.

$$R_{RT} \succ R_{AP} \succ G_S \succ R_{TC} \succ G_B \succ D_A \succ D_D$$

Where

$A \succ B$, A is more important than B.

It can be derived that the relative train punctuality is the most important attribute in the decision-making procedure of train operation adjustment. Then come to the relative train priority, the station train-stop margin, the relative train-covered distance, the train section margin and the follow-on train's new arrival delay. The attribute of the follow-on train's new departure delay plays the least function.

5. CONCLUSIONS

This paper works by using rough sets and information entropy theories. A train operation adjustment decision-making table is established and the information entropy of conditional attributes is studied. The importance of each attribute for reducing decision-making uncertainty is compared and the importance ranking of the attributes is given. This is a job of knowledge mining and discovering. The research result shows that, the 3 attributes named relative train punctuality, relative train priority, the station train-stop margin make great contribute to train operation adjustment decision-making. It proves that the dispatchers should seriously concern the train operation principles of "delay trains should let go of punctual trains" and "lower rank trains should not overtaking higher rank trains" (NIE Lei, ZHANG Xing-chen, ZHAO Peng, YANG Hao, HU An-zhou, 2001). The station train-stop margin G_S is the difference between minimal stopping time and scheduled stopping time at stations. It manifests the flexibility of train diagram. So, the station train-stop margin is one of the key points that should be considered carefully by train operation commander.

It is also obvious that we should especially pay attention to the attributes of relative train punctuality and relative train priority when modeling for train operation adjustment and designing for it. The importance ranking of the train operation adjustment conditional attributes can provide a reference for determining the importance degree of each optimization objective when a multi-objective optimization model is established for train operation adjustment. At the same time, there is a great gap to solve the problem of train operation and develop an intelligent train operation adjustment expert system by knowledge mining and discovery.

REFERENCES

[1] C.W. Tsang and T. K. Ho (2003). "The conflict resolution at connected railway junctions". Computational methods in engineering and science. Lu et al. (eds), Swets&Zeitlinger, Lisse, 987-993, 2003.

[2] J,Yuan, I. A. Hasen (2002). "Punctuality of train traffic in dutch railway stations". In Kelvin CP Wang & Guiping Xiao (Eds.), *Traffic and transportation studies*: Proceedings of ICTTS Guilin, China, 23-25, 2002.

[3] CAO Jiaming. Optimization model for adjusting diagram on double-track rail lines. *Journal of Southwest Jiaotong University.* Vol.30,No.5,520-526,1995.

[4] WANG Zhengbin, DU Wen. Train operation regulation model for auto-block double line and study on its heredity algorithm. *Railway transport and economy*. No. 6, 61-63. 2004.

[5] LUO Qing, JIN Fucai, HU Siji. Decomposition coordination computing model for train operation adjustment. *Journal of Beijing University*. Vol.28, No. 6, 87-89, 94, 2004.

[6] CHEN Yanru, PENG Qiyuan, JIANG Yangsheng. Research on a model for adjusting train diagram on double-track railway with satisfactory optimization. *Journal of China railway society*. Vol. 25, No.3, 8-12, 2003.

[7] WANG Minghui. Study on the Application of Rough Set in Railway Dispatching System. *China railway science.* Vol.25, No.4, 103-107, 2004.

[8] NIE Lei, ZHANG Xing-chen, ZHAO Peng, YANG Hao, HU An-zhou. Study on the strategy of train operation adjustment on high speed railway. *Journal of the China railway society.* Vol.23, No.8, 1-6, 2001.