Research on Knowledge Discovery in Database of Traffic Flow State Based on Attribute Reduction

Jia-lin WANG[1,2], Xiao-lu LI[1,2], Li WANG[1,2], Xi ZHANG[3], Peng ZHANG[3], Guang-yu ZHU[1,2]

1 MOE Key Laboratory for Transportation Complex Systems Theory and Technology, Bei-jing Jiaotong University, Beijing,China
2 Center of Cooperative Innovation for Beijing Metropolitan Transportation, Beijing,China
3 Municipality Key Laboratory of Urban Traffic Operation Simulation and Decision Sup-port, Beijing Transportation Research Center, Beijing,China

gyzhu@bjtu.edu.cn

Abstract. Recognizing and diagnosing the state of traffic flow is an im-portant research area, which is the basis of improving the level of traffic management and the quality of traffic information services. However, due to the increasing amount of traffic data collected, the traffic management sys-tem is facing the problem of "information surplus". After finishing several process, including data preprocessing, attribute reduction and rule acquisi-tion, finally obtained the knowledge rules of the traffic flow's state. Using the method of knowledge discovery can reveal some hidden, unknown and valuable information from the huge amount of traffic flow information, so as to provide rules and decision-making basis for traffic management de-partment.

Keywords: Rough Set, Knowledge Discovery in Database, State of Traffic Flow, Attribute Reduction, Rules Acquisition

1 Introduction

Faced with more and more serious traffic jams, many cities in recent years have carried out targeted transformation and optimization of road network, and the overall traffic capacity of roads has been increasing. But because of the surge in the number of motor vehicles, these measures can not effective alleviate traffic congestion. The premise of making these decisions is to correctly judge the state of traffic flow. In order to get traffic flow data for judging traffic flow status, and make more effective decisions for mitigating congestion, traffic management departments have applied advanced information technology to urban traffic management. By means of advanced road detection and information processing technology, traffic management departments have been able to get a lot of road traffic data scientifically and effectively, and sort them out and store them. However, with the increasing traffic volume data collected, traffic management system is also facing the problem of "information overload", just like other information systems.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

Knowledge Discovery in Database (KDD) is an important research method for obtaining important rules from a large number of information. The research field of KDD is closely related to artificial intelligence and machine learning. It also contains the methods of database and pattern recognition, which has become a very important research hotspot in the academic field [1]. Rough set theory is one of the classic methods in KDD. Golan.R and Ziarko.W applied rough set theory to the analysis of stock market [2], and found the dependence between stock price and economic index. The Canadian scholar Ziarko.W applied rough set theory to the water resource scheduling system, predicted the rules according to the rules obtained, and achieved good prediction results [3]. In pattern recognition, Kim Dai-jin has obtained the key attributes in the handwritten character recognition system by using the attribute reduction method of rough set, and obtained the recognition rule based on the key attribute [4]. At present, there is no relevant research to apply rough set theory to traffic flow state discrimination.

In this paper, based on rough set theory, knowledge discovery technology is used to acquire valuable knowledge and rules from mass traffic data, so as to correctly judge traffic flow status and provide basis for decision-making of traffic management department. The rest contents of this paper are arranged as following.

A traffic flow data preprocessing algorithm based on filtering algorithm is designed in part 2.

A brief introduction of rough set theory is introduced in part 3.

A decision table attribute reduction algorithm for traffic flow state based on rough set theory is designed in part 4. On the basis of the attribute reduction algorithm of the decision table of the classical differential matrix, the differential matrix is improved.

A case is studied to validate the improved attribute reduction algorithm for differential matrix in part 5.

2 Data Preprocessing

The traffic flow characteristic data used in this paper is based on the coil detec-tor. In the process of collecting traffic flow data by the coil detector, the damage of the coil, the failure of the communication equipment and the interference of the external environment will affect the accuracy and accuracy of the collected data. Therefore, the data collected by the coil detector can not be directly applied to this research, and the raw data need to be preprocessed. In view of the content of this paper, we should consider the microscopic randomness and macroscopic regularity of traffic flow data in preprocessing the data, and correct the fault data at the same time.

The sensors data record can be expressed as a 4 tuple structure [t, q, v, h]. Among them, t stands for time, q represents traffic volume, v stands for speed, h represents occupancy. Based on the above analysis, a traffic flow data prepro-cessing algorithm based on filtering algorithm is designed in this paper. The steps of the algorithm are as follows.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

Step1：Define the maximum traffic volume as Q_{max}, the maximum speed as V_{max}, and the maximum occupancy as H_{max}. If $q > Q_{max}$ or $v > V_{max}$ or $h > H_{max}$, identify the data as abnormal data.

Step2：When $h > H_1(H_1 = 95\%)$, If $v > V_1(V_1 = 5km/h)$, identify the data as abnormal data.

Step3：If $v = 0(km/h)$ and $q \neq 0$, identify the data as abnormal data.

Step4：When $q = 0$, if $v \neq 0 \,\&\&\, h \neq 0$, identify the data as abnormal data.

Step5：Calculate $AVEL$, if $AVEL \notin [1.5, 30]$, then remove it.

Step6：Replace the abnormal data with the nearest and normal data.

Step7：The method of first order differential operation is used to process the data. If the maximum change interval of the data should be reasonable order difference scores did not fall in the mean and variance by n data difference value determined by the situation(such as $\left[\bar{d} - \varepsilon_o * \sigma_d, \bar{d} + \varepsilon_o * \sigma_d\right]$), we can define the data for data distortion abrupt change, with the former a reasonable data record with $1/\varepsilon_o$ times the difference value of $p_{n-1} + 1/\varepsilon_o * d_n$ alternative.

The specific flow of the algorithm is shown in Fig. 1 as shown.

3 Rough Set Theory

Rough set theory is a new mathematical tool for dealing with fuzzy and uncertain knowledge. Knowledge is the object and subject of research. It can be understood as the summing up and induction of knowledge and rules in a certain field. The main idea of rough set theory is to obtain the decision or classification rules of the problem by means of knowledge reduction under the premise that the ability to classify the knowledge is constant [5].

Knowledge reduction is the core of rough set theory. Some knowledge and attributes are particularly important for the purpose of decision making, while some knowledge and attributes are irrelevant or even redundant. The meaning of the concept of knowledge reduction is to find out and eliminate redundant or unimportant knowledge in knowledge base, and make the classification ability of the knowledge base unchanged.

In the definition of knowledge reduction, reduction and core are the two most important concepts. The core of knowledge is the intersection of all knowledge reduction, that is to say that the core of knowledge exists in the reduction of any knowledge, and is the most important part of the knowledge reduction. Because knowledge core is the expression part of knowledge feature, it will never be eliminated in knowledge reduction. Attribute reduction is an important part of knowledge reduction. Attribute reduction is to remove the uncorrelated or unim-portant attributes under the condition of keeping the classification ability of knowledge base unchanged.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

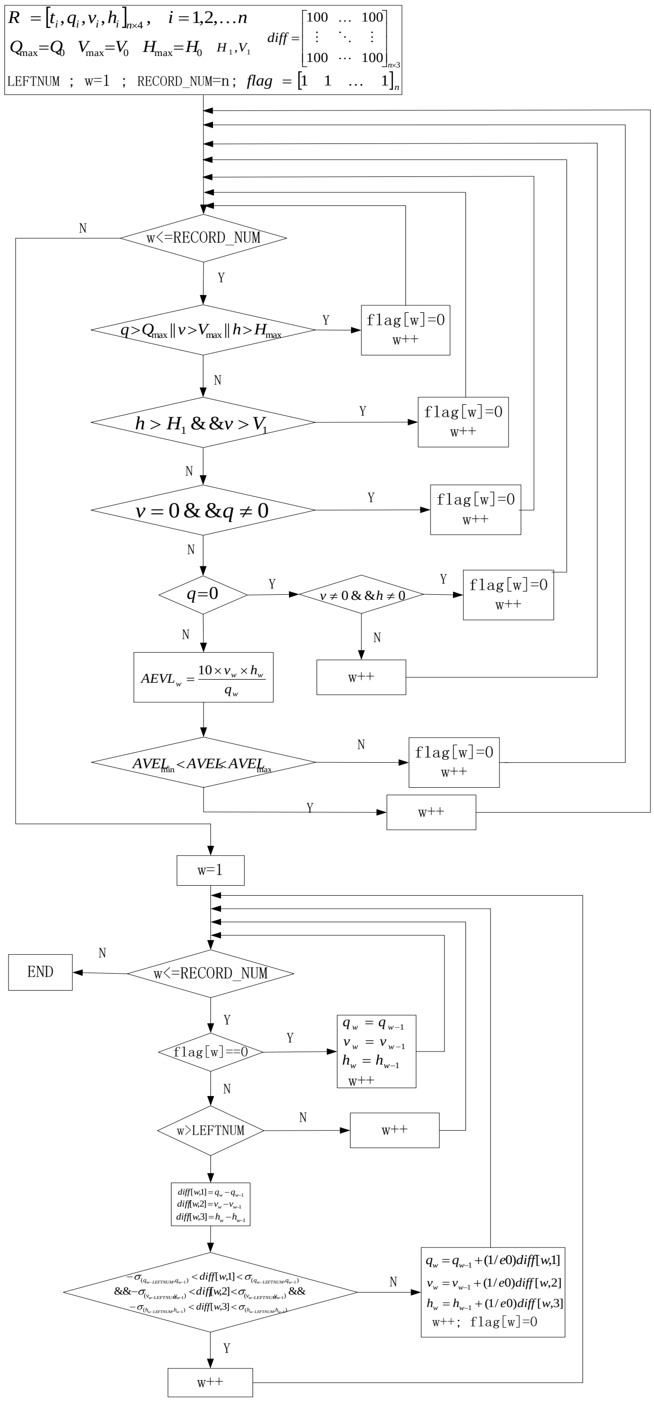

Fig. 1. Data preprocessing algorithm flow chart.

DOI: 10.1007/978-3-030-22744-9_21

4 Attribute Reduction of Decision Table Based on Rough Set Theory

4.1 Attribute Reduction Theory of Decision Table

The decision table contains a large number of data samples, from which many decision rules can be extracted. All the decision rules are summed up, and all decision rules of the decision problem can be obtained. However, the basic deci-sion sets obtained by this method are not practical, because many of them are not representative of the whole decision system. Therefore, it is necessary to reduce the attribute of the decision table to obtain the useful decision rules for the deci-sion making, so that the problem of knowledge discovery can be solved in the relevant practical fields.

As the core content of rough set theory, attribute reduction based on decision table has become a branch of the rough set theory that has been focused on. For different decision systems, we can use the attribute reduction algorithm to adapt to its decision environment, so as to achieve the goal of effective decision rules. Different attribute reduction algorithms have the same goal, which is to find an optimal reduction, that is, the reduction set containing the least condition attrib-utes.

The most commonly used attribute reduction algorithms are discernibility ma-trix method, attribute importance degree method, information entropy method, genetic algorithm and so on. The discernibility matrix method is intuitive and simple in dealing with general decision table problems, and its calculation pro-cess is simple and easy to understand, and its reduction result is very accurate.

4.2 Attribute Reduction Algorithm for Decision Table Based on Discernibility Matrix

In the existing algorithm of attribute reduction using rough set theory, the differ-ence matrix method proposed by A.Skowron in 1992 is very classic and effec-tive. The idea of the algorithm is that by defining a difference matrix, all the dif-ference items of the condition attribute and the decision attribute are recorded, and then the matrix operation of the difference matrix is performed to find the kernel of the decision table and to output the set of all attribute reduction.

Give any incompatible decision table ：

$DT=(U,C\cup D,V,f)$, $\forall x_i,x_j\in U$ and $|U|=n$. We can define the improved discernibility matrix $M_{n\times n}(DT)$.

$$M_{n\times n}(DT)=(c_{ij})_{n\times n}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ * & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & c_{nn} \end{bmatrix}=\begin{bmatrix} c_{11} & * & \cdots & * \\ c_{21} & c_{22} & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

This includes

$$\begin{cases} c_{ij}=\left\{\alpha \mid (\alpha\in C)\wedge\left(f_a(x_i)\neq f_a(x_j)\right)\right\}, & f_D(x_i)\neq f_D(x_j) \\ \varnothing, & f_D(x_i)\neq f_D(x_j)\wedge f_C(x_i)=f_C(x_j) \\ -, & f_D(x_i)=f_D(x_j) \end{cases} \tag{1}$$

The algorithm of attribute reduction based on discernibility matrix for decision table is shown below.

Input: a decision table $DT=(U,C\cup D,V,f)$

Output: $RED_C(D)$ of *DT*

Step1：According to the specific definition of the difference matrix, the form of the lower triangulation matrix is used to generate the difference matrix $M_{n\times n}(DT)$.

Step2：Ergodic the difference matrix, searching all of the elements in detail. If $\varnothing$ does not exist, then jump to Step 3, and if $\varnothing$ exists, then quit.

Step3：Ergodic each attribute element in the difference matrix, assign the value to $CORE_C(D)$ and output $CORE_C(D)=\left\{\alpha|(\alpha\in C)\wedge\left(\exists c_{ij},\left((c_{ij}\in M_{n\times n})\wedge(c_{ij}=\{\alpha\})\right)\right)\right\}$.

Step4：In all possible combinations of attributes, a combination of attributes containing relative kernel is solved and judged according to the following conditions: (1) $\forall c_{ij}\in M_{n\times n}(DT)$, if $c_{ij}\neq\varnothing$, $B\cap c_{ij}\neq\varnothing$? (ignore $c_{ij}\neq\varnothing\vee-$); (2) Is B independent?

If the two discriminant conditions are satisfied, Value to $RED_C(D)$ and ergodic a combination of attributes that contain all the relative kernel *D*.

Step5：Output $RED_C(D)$

The attribute reduction algorithm based on the differential matrix method above has been well applied to many practical problems. At the same time, it has a lot of advantages, especially for the relative *D* kernel and all the relative *D* re-duction calculations very convenient and accurate. However, the classical algo-rithm based on differential matrix is only suitable for a complete and compatible decision table, and there is a problem of "nuclear explosion" caused by the large amount of calculation in the relative reduction of the decision table [6].

4.3 An Improved Attribute Reduction Algorithm for Decision Table Based on Discernibility Matrix

The traditional discernibility matrix is improved so that it can be applied to the inconsistent decision table. Give any incompatible decision table $DT=(U,C\cup D,V,f)$, $\forall x_i,x_j\in U$ and define $d(x_i)=card\left(\{z\mid z=f_D(y),\forall y\in[x_i]c\}\right)$. It describes the total number of all the different decision values of objects, and these decision values must be equivalent classes of the complete set of conditional attributes, so we can define the improved discernibility matrix $M^*{}_{n\times n}(DT)$.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

$$M*_{n\times n}(DT)=(r_{ij})_{n\times n}=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ * & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & r_{nn} \end{bmatrix}=\begin{bmatrix} r_{11} & * & \cdots & * \\ r_{21} & r_{22} & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

This includes

$$\begin{cases} c_{ij}=\left\{\alpha\,|\,(\alpha\in C)\wedge\left(f_a(x_i)\neq f_a(x_j)\right)\right\}, & f_D(x_i)\neq f_D(x_j) \\ \varnothing, & f_D(x_i)\neq f_D(x_j)\wedge f_C(x_i)=f_C(x_j) \\ -, & f_D(x_i)=f_D(x_j) \end{cases}$$

$$\begin{cases} r_{ij}=c_{ij}, & \min\left\{d(x_i),d(x_j)\right\}=1 \\ \varnothing, & \min\left\{d(x_i),d(x_j)\right\}>1 \end{cases} \tag{2}$$

By adding the $\min\{d(x_i),d(x_j)\}$ condition, the traditional discernibility matrix is improved [7]. If a decision table is compatible, there will always be $\min\{d(x_i),d(x_j)\}=1$, and the improved discernibility matrix will be degenerated into a Skowron discernibility matrix. If the decision table is incompatible, $\min\{d(x_i),d(x_j)\}$ may be more than 1. At this point, we must use the improved discernibility matrix to compute the result of attribute reduction in the correct decision table.

5 Case Study

5.1 Overview of Traffic Flow Status

Collecting traffic flow characteristic data is the first step to identify traffic flow status. Then, the traffic flow data are pre processed scientifically and effectively to analyze the relationship between traffic flow data. Traffic flow data include vehicle flow, average speed, lane occupancy, vehicle density and other parame-ters. The characteristics of traffic flow data can reflect the running state of traffic flow. For example, when the average speed of a section is mainly distributed be-tween 40-60km/h, it shows that the traffic flow condition of this section is rela-tively smooth [8]. Generally speaking, the traffic flow state of road sections or intersections can be divided into four modes: smooth, normal, mildly congested and congested [9].

The key step of traffic flow status recognition is to classify the traffic flow pat-tern, extract the characteristics of traffic flow, and identify the traffic flow status by a specific pattern recognition algorithm. There are many algorithms in the study of traffic flow status recognition, which is not the research direction of this paper. Therefore, this paper selects the traffic flow characteristic data set of known traffic flow state. This paper divides traffic flow status into three grades: smooth, mildly congested and congested.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

5.2 Data Sources

The 48 hour traffic section data collected by the coil detector in Beijing on May 4, 2012 (Friday) numbered 03218 is taken as the research data. There are 4 lanes in the section, Lan1, Lan2, Lan3 and Lan4 respectively. The record includes 1 sets of traffic basic data: vehicle flow, average speed of vehicles, occupancy of lanes and traffic volume of long cars. The sampling interval is 2mins.

5.3 Attribute Reduction of Traffic Flow State Decision Table

First, the data collected by the coil detector is preprocessed by using the method of part 2, and then the preprocessed traffic flow characteristic data set is preprocessed. A group of four minutes of data was extracted from the time of 6:00-10:00 in the morning. A total of 60 sets of data were extracted as the representative set of training samples, representing the form of a decision table. The 12 dimension characteristic data of Lan1, Lan2, Lan3 and Lan4 traffic volume, speed and occupancy (q_1, v_1, o_1, q_2, v_2, o_2, q_3, v_3, o_3, q_4, v_4, o_4) are taken as the conditional attributes of decision table. Among them, the flow rate of Lan1 is q_1, the speed is v_1, and the occupancy rate is o_1. Similarly, traffic flow characteristic data of Lan2, Lan3 and Lan4 can be seen. Taking the running state of the traffic flow on the basis of priori knowledge as the decision attribute of the decision table, a decision table for the running state of the traffic flow is finally constructed. Rough set method needs to deal with discretized data, but the collected traffic flow data is continuous. So first, we must simply discretize the traffic flow data and make all continuous conditional attribute values discretized. After analyzing, sorting and extracting the relevant data, the discretization interval of the specific condition attributes is set as follows.

Traffic Volume(pcu/4 mins): 1/[0,75], 2/(75,100], 3/(100,105], 4/(105,+∞);

Occupancy(%):1/[0,0.2], 2/(0.2,0.4], 3/(0.4,0.5], 4/(0.5,1];

Traffic states(M): 0/Smooth, 1/Mildly Congested, 2/Congested;

The conditional attribute values of all 60 sets of data are discretized into {1, 2 3, 4}, and the decision attribute values are discretized into {0, 1, 2}. According to the discrete interval generated after discretization, the decision table of traffic flow state is constructed.

The attribute reduction algorithm of the decision table in the part 4 is realized by MATLAB programming, and the improved matrix is improved for the incompatible decision table, and the attribute reduction is completed in the decision table. After data preprocessing, data discretization, decision tables and decision tables, the final result of attribute reduction is obtained. There are 15 attribute reduction sets in the attribute reduction table, and the kernel of the decision table is {v_3}. It can be concluded from the calculation results that the number of characteristic attributes of the decision table is reduced from the initial 12 to 4. The attribute reduction can greatly reduce the computational complexity of the decision table by eliminating the redundant conditions, which provides the basis for the rule acquisition of traffic flow state.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

5.4 Rule Acquisition

15 feature reduction sets are generated by arranging the attribute reduction set of traffic flow status decision table generated by MATLAB programming as shown in Table 1.

Each set of attribute reduction sets corresponds to a decision information table after reduction. For one of the reduction sets, the decision rules can be obtained by analyzing their corresponding decision information tables. A group of characteristic attribute reduction sets $\{q_1, v_1, v_3\}$ are used as an example to analyze the knowledge discovery process of the traffic flow status in this section. By deleting the feature attributes of traffic flow other than the reduction set, and merging all the same rows in the decision table, a traffic state decision table based on reduction set $\{q_1, v_1, v_3\}$ is finally generated as shown in Table 2.

Table 1. Attribute reduction set of traffic flow state

	Reduction Set
1	$\{q_1, v_1, v_3\}$
2	$\{v_1, q_2, v_3\}$
3	$\{v_1, q_3, v_3, v_4\}$
4	$\{v_1, v_3, o_3, v_4\}$
5	$\{v_1, q_3, v_4, o_4\}$
6	$\{q_1, v_2, v_3\}$
7	$\{q_2, v_2, v_3\}$
8	$\{v_2, q_3, v_3, v_4\}$
9	$\{v_2, v_3, o_3, v_4\}$
10	$\{v_3, q_4, v_4\}$
11	$\{v_2, v_3, v_4, o_4\}$
12	$\{q_1, o_1, v_3, q_4\}$
13	$\{o_1, q_2, v_3, q_4\}$
14	$\{q_1, q_3, v_3, q_4\}$
15	$\{q_2, q_3, v_3, q_4\}$

Based on the above analysis, we know that knowledge discovery based on rough set theory can effectively extract all the smallest attribute reduction sets in decision tables, and we can extract corresponding traffic flow state decision rules based on deleted traffic attributes state decision tables. Based on Table 2, we can get 11 traffic flow status decision rules. Some of these rules are as follows.

Rule 1: If $\{ q_1, v_1, v_3\}=\{1, 4, 4\}$, that is, Lan1's traffic volume is in the [0,75] range, the average speed of Lan1 is in the range of $(60, +\infty)$, and the average speed of Lan3 is in the range of $(60, +\infty)$. The traffic state is smooth.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

Rule 2: If { q_1, v_1, v_3}={2, 4, 3}, that is, Lan1's traffic volume is in the (75,100] range, the average speed of Lan1 is in the range of (60, +∞), and the average speed of Lan3 is in the range of (45,60]. The traffic state is smooth.

Rule 3: If { q_1, v_1, v_3}={2, 3, 3}, that is, Lan1's traffic volume is in the (75,100] range, the average speed of Lan1 is in the range of (45,60], and the average speed of Lan3 is in the range of (45,60]. The traffic state is smooth.

Rule 4: If { q_1, v_1, v_3}={2, 3, 2}, that is, Lan1's traffic volume is in the (75,100] range, the average speed of Lan1 is in the range of (45,60], and the average speed of Lan3 is in the range of (30,45]. The traffic state is smooth.

Rule 5: If { q_1, v_1, v_3}={3, 2, 2}, that is, Lan1's traffic volume is in the (100,105] range, the average speed of Lan1 is in the range of (30,45], and the average speed of Lan3 is in the range of (30,45]. The traffic state is smooth.

Rule 6: If { q_1, v_1, v_3}={2, 2, 2}, that is, Lan1's traffic volume is in the (75,100] range, the average speed of Lan1 is in the range of (30,45], and the average speed of Lan3 is in the range of (30,45]. The traffic state is mildly congested.

Rule 7: If { q_1, v_1, v_3}={2, 2, 1}, that is, Lan1's traffic volume is in the (75,100] range, the average speed of Lan1 is in the range of (30,45], and the average speed of Lan3 is in the range of [0,30]. The traffic state is mildly congested.

Rule 8: If { q_1, v_1, v_3}={2, 1, 1}, that is, Lan1's traffic volume is in the (75,100] range, the average speed of Lan1 is in the range of [0,30], and the average speed of Lan3 is in the range of [0,30]. The traffic state is mildly congested.

Rule 9: If { q_1, v_1, v_3}={1, 1, 1}, that is, Lan1's traffic volume is in the [0,75] range, the average speed of Lan1 is in the range of [0,30], and the average speed of Lan3 is in the range of [0,30]. The traffic state is mildly congested.

Rule 10: If { q_1, v_1, v_3}={2, 2, 4}, that is, Lan1's traffic volume is in the (75,100] range, the average speed of Lan1 is in the range of (30,45], and the average speed of Lan3 is in the range of (60, +∞). The traffic state is smooth.

Table 2. Decision table of deleting the redundant attributes of traffic flow state

U	q_1	v_1	v_3	M
x_1	1	4	4	0
x_2	2	4	3	0
x_3	2	3	3	0
x_4	2	3	2	0
x_5	3	2	2	0
x_6	2	2	2	1
x_7	2	2	1	1
x_8	2	1	1	1
x_9	1	1	1	1
x_{10}	2	2	4	0
x_{11}	2	3	4	0

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

Rule 11: If { q_1, v_1, v_3}={2, 3, 4}, that is, Lan1's traffic volume is in the (75,100] range, the average speed of Lan1 is in the range of (45,60], and the average speed of Lan3 is in the range of (60, +∞). The traffic state is smooth.

As long as there is a set of traffic flow data at any time of this section of the road, the traffic states can be judged according to the above rules. Therefore, the recognition rule of traffic states can objectively describe the characteristics of traffic state, and provide a strong basis for traffic control guidance.

The above 11 rules are only the set of decision rules obtained from a set of traffic flow characteristic attributes reduction set { q_1, v_1, v_3}. In the same way, the other 14 groups of traffic flow characteristic reduction sets can also be used for rule acquisition. The same decision rules are merged to get the decision rule set of all traffic flow status of the section. Finally, the obtained decision rule set is stored in the corresponding rule knowledge base as the traffic flow state pattern classification knowledge of the section. In the future traffic states recognition, we only need to match the classification characteristics of the traffic flow data and the classification rules that are stored in the knowledge base, and we can identify the running the traffic states at this time. It can also predict and analyze the future traffic states based on the knowledge rule of traffic flow, so as to work out a plan for alleviating traffic jams.

6 Conclusion

This paper does not consider the influence of weather, road facilities, drivers' psychological state and other factors on traffic flow state. In the future study, these factors should also be taken into account in the identification and diagnosis of traffic state. However, in this paper, the attribute reduction of traffic flow state decision table is reduced to eliminate the redundant attributes, and the knowledge discovery of traffic flow state is realized. In the case analysis, we get the specific knowledge rules of the traffic flow status, and provide a scientific and effective decision-making basis for the traffic management department to formulate the correct traffic flow guidance measures.

References

1. L. Wang. Knowledge discovery of uncertain information based on rough set and its application in urban traffic management. Southwest Jiao Tong University. Chengdu, 2011.
2. Y. G. Jing. Research on attribute reduction algorithm based on rough set. Chengdu, Southwest Jiao Tong University Press, 2013.
3. D. Sen, S. K. Pal. “Generalized rough sets, entropy, and image ambiguity measure,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 1, pp. 117-128, August 2009.
4. W. Ziarko. “Variable precision rough set,” Journal of Computer and System Sciences. vol. 46, pp. 39-59, 1993.
5. W. X. Zhang, Z. W. Wu. Rough set theory. Beijing, Science Press, 2001.
6. A. Petrosino, A. “Ferone. Rough fuzzy set-based image compression,” Methematical and Computer Modeling. Vo10. 160, pp. 1485-1506, 2009.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21

7. M. Ningler, G, Stockmanns, G. Schneider, et al. “Adapted Variable Precision Rough Set Approach for EEG Analysis,” Artificial Intelligence in Medicine. Vo3. 47, pp. 239-261, 2009.
8. S. L. Pu, R. M. Li, Q. X. Shi. “Research on traffic flow recognition algorithm based on rough set fuzzy recognition technology,” Journal of Wuhan University of Technology. Vo6. 34, pp. 1154-1158, 2010.
9. Yang H, Bell M G H. “Models and algorithms for road network design: a review and some new developments,” Transport Review. Vo3. 18, pp. 257-278, 1998.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22744-9_21