RESEARCH ON IDENTIFICATION OF FREEWAY BLACK SPOTS AND ACCIDENT INDUCED FACTORS

MENG Xianghai, Ph.D., Prof., School of Transportation Science and Engineering, Harbin Institute of Technology, Email:mengxianghai100@126.com

LI Mei, Graduate student, School of Transportation Science and Engineering, Harbin Institute of Technology,Email:li_mei_123@163.com

GUAN Zhiqiang, Graduate student, School of Transportation Science and Engineering, Harbin Institute of Technology, Email: sunangic@163.com

ABSTRACT

In order to improve the objectivity and fairness of identifying black spots and elements, three correlative methodologies are developed. Firstly, an approach for segment division based on dynamic clustering algorithm is proposed, which can be used to separate a highway into a series of subsections according to the appearance of accidents. Then, a self-organizing neural network model is established to identify the black spots from these subsections. Lastly, a methodology based on discrete multi-variable algorithm combined with rough set theory is presented to determine the accident induced factors of these black spots. The results show that (*1*) the segment division based on dynamic clustering algorithm can describe objectively the concentration and dispersion of accidents, (*2*) the neural network model can quickly identify the black spots and the corresponding results are reasonable, and (*3*) the methodology to identify the prominent accident causes can be used to establish a set of evaluation criteria and then to determine the accident induced factors of a black spot.

Keywords: segment division, black spot identification, accident induced factor identification, dynamic clustering algorithm, self-organizing neural network, rough set theory

1 INTRODUCTION

The first work to identify the black spots of an operational freeway is to separate it into a series of subsections according to the appearance of accidents or geometric conditions they own. In many cases, this is finished by manual work instead of by some algorithm, in which the basic assessment sites are determined by the researchers with their personal knowledge on the concentration or dispersion of accidents. Obviously, the results will be more or less different among different researchers. Furthermore, with the rapid development of new technologies and calculation methodologies, the way of manually dividing segment cannot meet the requirements of modern research work.

Some basic methodologies of expected frequency method, accident rate method, frequency rate method and quality control method are widely used in practice in the field of black spot identification. But, some limitations of them still exist *(1)*. Although in recent years some new approaches based on the accident indexes of frequency and cost *(2)*, double variables filtration algorithm *(3)*, fuzzy evaluation method *(4)*, BP neural network and some techniques based on GIS or GPS are developed or introduced to solve such problem *(5)*, they still need to be improved in the future work to promote the feasibility of identifying the black spots.

When black spots are identified, the next work is to determine the accident induced factors. Logistic model, logarithmic transformation linear regression model, variance analysis method and classified regression tree method are firstly used in finding the possible accident causes *(6) (7)*. Then, catastrophe theory in non-linear science, discrete multi-variable algorithm, fuzzy evaluation theory, SOFM neural network model and rough set theory are introduced lately *(8)~(11)*. They can make some kinds of decision for a highway or a specified black spot from one or more aspects of accident, whose fundaments and key issues will have signification reference to the problem of identifying the accident induced factors.

2 SEGMENT DIVISION BASED ON DYNAMIC CLUSTERING ALGORITHM

2.1 Purpose of segment division and selection of methodologies

Segment division is a work to separate a highway into a series of subsections based on the appearance of accidents. These subsections are taken as the basic sites, which will be evaluated. Those owning accidents exceed a critical value will be identified as black spots. Obviously, the purpose of segment division is to obtain the basic black spot evaluation units.

There are several ways to solve such problem like this, which divide a set into several sub-sets firstly and then cluster them into some new groups. Considering the facts that the sample size is large and the distribution of accidents along the route is discrete, the dynamic clustering algorithm is found suitable to divide segment. When it comes to the application of dynamic clustering algorithm, samples is each accident case and attribute variable is the highway station where each accident occurred. From the result of clustering analysis, a set of subsections is given, each specified by a starting station and an ending one.

2.2 Process of dividing segment based on dynamic clustering algorithm

The basic steps to implement segment division based on dynamic clustering analysis, in which the K-means algorithm is adopted, are given as follows:

Step 1.*Select the initial clustering centre*

Set *n* accident cases into *k* categories. Then calculate the average station $x^{(0)}$ of accidents belonging to each group. Taken the average station as the corresponding clustering centre, the initial clustering centre set $L^{(0)}$ can be obtained, which is $L^{(0)} = \{x_1^{(0)}, x_2^{(0)}, \cdots, x_k^{(0)}\}$.

Step 2.*Calculate the initial classification*

Calculate the distances between each accident case and all initial clustering centres to determine which group this case will be. For accident case *x*, if its distance to clustering centre $x_i^{(0)}$ is smaller than that to centre $x_j^{(0)}$, it belongs to cluster $G_i^{(0)}$, which is $G_i^{(0)} = \{x : d(x, x_i^{(0)}) \le d(x, x_j^{(0)}), j = 1, 2, \cdots k, j \ne i\}, i = 1, 2, \cdots, k$ .Where, $d(x\ , x_i^{(0)})$ is the Euclidean distance between case *x* and centre $x_i^{(0)}$. When all accident cases are grouped, the initial classification set $G^{(0)}$ is obtained, which is $G^{(0)} = \{G_1^{(0)}, G_2^{(0)}, \cdots, G_k^{(0)}\}$.

Step 3.*Determine the final classification by iterative calculation*

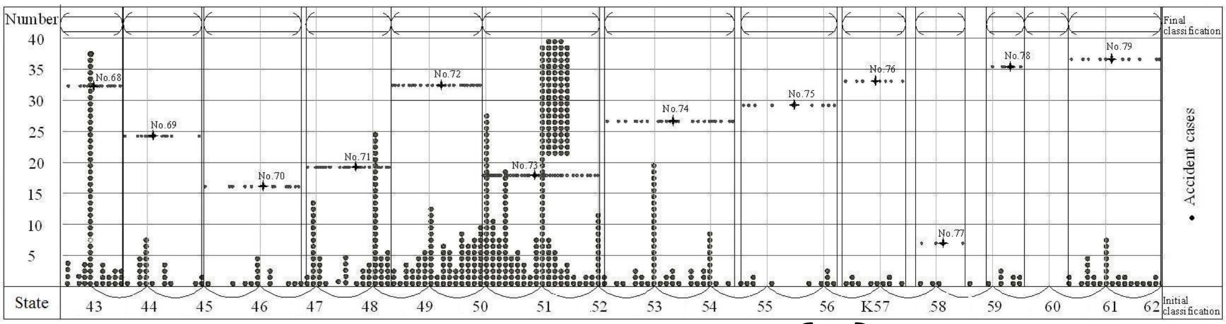
Calculate the new clustering centre set $L^{(1)} = \{x_1^{(1)}, x_2^{(1)}, \cdots, x_k^{(1)}\}$, in which the element $x_i^{(1)}$ can be obtained by $x_i^{(1)} = (1/n_l)\sum x_l,\ x_l \in G_i^{(0)}$ .Where, n_l is the number of accident cases in $G_i^{(0)}$.

Go to Step 2 repeatedly to obtain a new classification for each iterative process. When $G^{(m+1)} = \left\{G_1^{(m+1)}, G_2^{(m+1)}, \cdots, G_k^{(m+1)}\right\}$ and $G^{(m)} = \left\{G_1^{(m)}, G_2^{(m)}, \cdots, G_k^{(m)}\right\}$ are in accordance with prescriptive precision, then, $G^{(m)}$ is the final segment division set.

2.3 Case study for segment division

The segment of Beijing-Zhuhai freeway starting from Xiaotang in Hunan province and ending at Gantang in Guangdong province, which is 109 kilometres and has a 13-kilometer continuous long steep downgrade, is known as “China’s the most challenging mountainous freeway construction project” because of its critical terrain and complex climate conditions. Since first opened to traffic in 2003, more than 3766 accidents occurred on it, resulting in enormous property damage and casualties.

The southbound of the segment is taken as the case to illustrate the application of dynamic clustering algorithm for segment division. It is firstly divided into 109 subsections which are evenly spaced to have approximate one kilometre. Then, the accidents occurred on each subsections are summarized and the centre station of each subsection is selected as the initial cluster centre. Lastly, based on dynamic clustering algorithm, the final segment classification is obtained. Part of the results, sta.42+500 to sta.62+500, are listed in Table 1 and Figure 1.



Note：● —*accident case,* + —*cluster centre,* —*initial classification，* —*final classification*

Figure 1 – Accident samples and results of segment division

Table 1 – Initial and final classification for segment of sta.42+500 to sta.62+500

Initial classification			*Final classification*		
No.	*Subsection*	*Clustering centre*	*No.*	*Subsection*	*Clustering centre*
44	Sta.43- sta.44	Sta.43+500	68	Sta.42+590- sta.43+500	Sta.43+045
45	Sta.44- sta.45	Sta.44+500	69	Sta.43+550- sta.44+900	Sta.44+225
46	Sta.45- sta.46	Sta.45+500	70	Sta.45+000- sta.46+650	Sta.45+825
47	Sta.46- sta.47	Sta.46+500	71	Sta.46+800- sta.48+250	Sta.47+525
48	Sta.47- sta.48	Sta.47+500	72	Sta.48+300- sta.49+870	Sta.49+085
49	Sta.48- sta.49	Sta.48+500	73	Sta.49+900- sta.52+000	Sta.50+950
50	Sta.49- sta.50	Sta.49+500	74	Sta.52+100- sta.54+300	Sta.53+200
51	Sta.50- sta.51	Sta.50+500	75	Sta.54+500- sta.56+100	Sta.55+300
52	Sta.51- sta.52	Sta.51+500	76	Sta.56+300- sta.57+300	Sta.56+800
53	Sta.52- sta.53	Sta.52+500	77	Sta.57+650- sta.58+500	Sta.58+075
54	Sta.53- sta.54	Sta.53+500	78	Sta.58+900- sta.59+400	Sta.59+150
55	Sta.54- sta.55	Sta.54+500	79	Sta.60+300- sta.61+900	Sta.61+100
56	Sta.55- sta.56	Sta.55+500			
57	Sta.56- sta.57	Sta.56+500			
58	Sta.57- sta.58	Sta.57+500			
59	Sta.58- sta.59	Sta.58+500			
60	Sta.59- sta.60	Sta.59+500			
61	Sta.60- sta.61	Sta.60+500			

The number and the average length of subsections in initial and final classification are 18 versus 12 and 1 km versus 1.398 km, respectively.

Obviously, the results for segment division based on the approach presented in this paper are more reasonable, because the final cluster centres can describe very well the concentration degree of accidents for each subsection and the ranges of subsections can reflect objectively the dispersion situation of accidents.

3 IDENTIFICATION OF BLACK SPOTS BASED ON SELF-ORGANIZING COMPETITIVE NEURAL NETWORK

3.1 Limitations of basic black spot identification methodologies

The four basic methodologies to identify the hazard locations are expected frequency method, accident rate method, frequency rate method, and quality control method.
Expected frequency method assumes that the accident follows the Poisson distribution, then calculate a critical frequency based on a specific statistical confidence of level. A site is considered to be hazardous when its accident frequency exceeds the critical value. The disadvantage of no taking into account the exposures limits the application of this method. With respect to the sites be evaluated, an acceptable critical rate is determined to decide each site hazardous or not. This is the ***accident rate method***. Although taking into account the influence of traffic volume, the definition of critical accident rate has no sufficient theoretical foundation. ***Frequency rate method*** locates each site in a matrix according to its frequency and rate, and then determines a critical frequency as well as a critical rate. The site is selected as the black spot only if its frequency and rate are all bigger than the corresponding critical values. The limitation of frequency rate method is that there is too much subjective consideration on the selection of critical values.
Quality control method (QCM) calculates the critical accident rates based on statistical test, which can be used to evaluate the safety situation of each site. Because of having a

theoretical foundation, it is widely used to identify the black spots. The disadvantage of this method is that it can be only used under such situation that all sites be evaluated should have the same traffic or geometric condition. The statistical test in QCM is based on the assumption that the occurrence of accident can be approximated by Poisson distribution, but some recent researchers have found that the Negative Binomial distribution and the Bayes distribution may be more suitable than the Poisson distribution in some cases.

Although the methodologies recommended above can be used to identify the black spots, they all more or less have some limitations. The results obtained from different methods are not always the same because of the different indexes and criteria they take. Furthermore, it should deserve in-depth research to develop some better ways to identify objectively the hazard locations.

3.2 Identification methodology based on self-organizing competitive neural network

In essence, it is similar to the problem of pattern recognition to identify the black spots, and it is difficult to specify the fitting function and its conditions to be subject to when more identifying indexes are involved in. Obviously, it happens to be the advantage of the neural network technology to solve such a problem.

Perceptrons and BP models are the two widely used neural network for pattern recognition, but they all need clearly the target output for network training and testing. When we perform the task to identify the black spots, the safety situations of these sites are always unclear. And the disadvantage of no samples available limits greatly the application of these two networks in black spot identification.

Self-organizing competitive neural network which adopts the unsupervised learning rule can also be used in classification. Due to the fact that it can cluster automatically the samples into several categories without the needs of given target output, it is more suitable to solve the problem of black spot identification. Hence, an approach based on self-organizing competitive neural network is developed.

It contains an input layer and a competitive layer. The input vector is consists of two variables of accident frequency and accident rate described by accidents per 100 million vehicle kilometres. The outputs are the three safety situations of hazardous, normal or safe sites. The designed network architecture is shown in Figure 2.

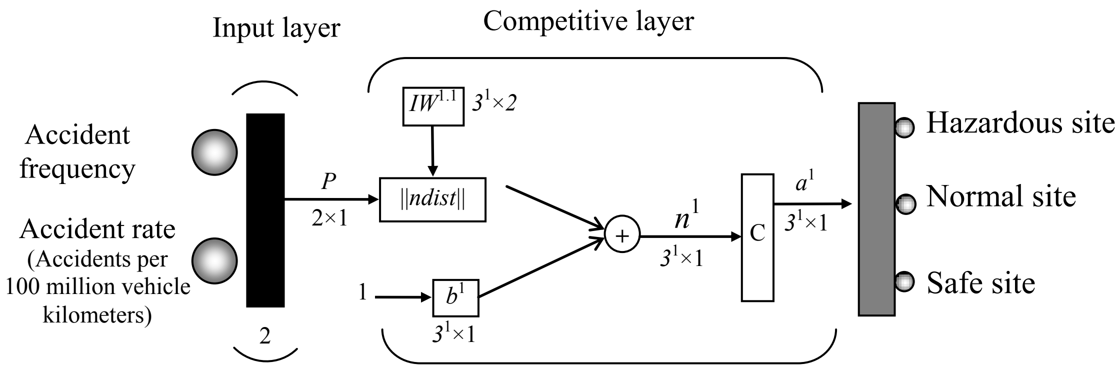


Figure 2 – Structure of self-organizing competitive neural network

3.3 Case study for black spot identification

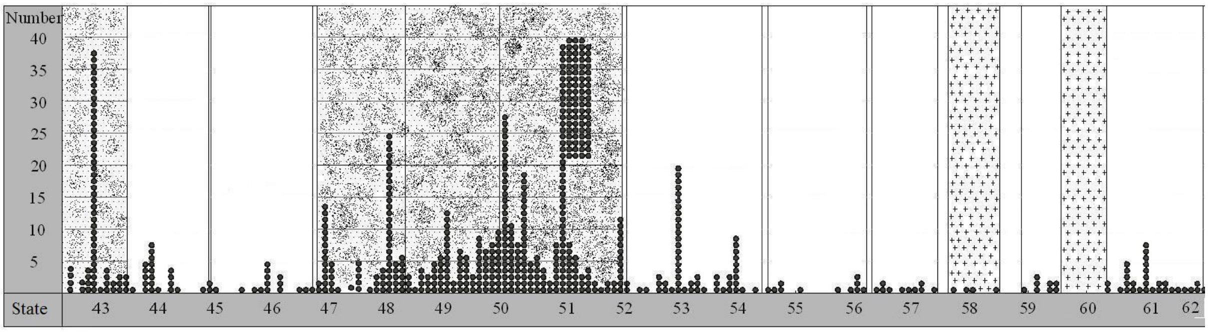
The safety situations of the 109 sites in the southbound of Beijing-Zhuhai freeway, which are divided by dynamic clustering algorithm, are evaluated by self-organizing competitive neural network. There are 10 black spots, which takes 9.17 percent of the total sites. The hazard sites are listed in Table 2 and the safety situations of each site in segment of sta.42+500 to sta.62+500 are shown in Figure 3.

It is meaningful to consider this problem from another point of view. The safety situations of the same segments are evaluated by expected frequency method, accident rate method, frequency rate method and quality control method. The corresponding black spots identified by the four methodologies are also listed in Table 2.

Table 2 – Results of black spot identification

Site	*Accident indexes*		*Self-organizing competitive neural network*	*Frequency rate method*	*Quality control method*	*Accident rate method*	*Expected frequency method*
	Frequency /total	*Rate per year*					
Sta.42+590-sta.43+500	61	333	☆	☆	☆	☆	☆
Sta.48+300-sta.49+870	87	352	☆	☆	☆	☆	☆
Sta.49+900-sta.52+000	219	694	☆	☆	☆	☆	☆
Sta.76+780-sta.78+700	122	315	☆	☆	☆	☆	☆
Sta.1+600-sta.2+000	34	422	☆	☆	☆	☆	※
Sta.10+900-sta.11+200	26	341	☆	☆	☆	☆	※
Sta.22+900-sta.23+000	25	1243	☆	☆	☆	☆	※
Sta.46+800-sta.48+250	72	228	☆	☆	※	※	☆
Sta.80+900-sta.82+900	66	164	☆	※	※	※	☆
Sta.102+150-sta.103+400	44	175	☆	※	※	※	※
Sta.9+970-sta.10+030	4	331	※	※	☆	☆	※
Sta.20+800- sta.21+000	12	298	※	※	☆	※	※
Sta.87+700-sta.88+600	39	215	※	☆	※	※	※

Note：Icon ☆ represents the black spot，and※ represents the normal or safe sites.



Note： —*black spots*， —*normal sites*, —*safe sites*

Figure 3 – Identification results based on self-organizing competitive nerual network

By analyzing the results of black spot identification, it can be found that the result of neural network is much close to that of frequency rate method, and the closeness degree is up to 76.9%. Although this still cannot prove that the self-organizing competitive neural network is superior to the basic methodologies, after all, it can quickly identify the black spots and the corresponding results are reasonable.

4 DETERMINING ACCIDENT INDUCED FACTORS OF BLACK SPOTS BASED ON DISCRETE MULTI-VARIABLE ALGORITHM COMBINED WITH ROUGH SET THEORY

4.1 Principles and processes

Having identified the hazardous locations, the next stage is to determine the accident induced factors for each black spot, which is very useful to develop the alternative countermeasures.
When taking into account the first and secondary causes of each accident, the types of accident causes summarized from all accidents of a highway would be enormous. Obviously, many of them are less important because they cause few accidents. Hence, in order to simplify the calculation, these unimportant causes should be distinguished and then be neglected.
As far as the remaining accident causes is concerned, the average accident frequency induced by each kind of causes can be calculated based on entire accidents, and the corresponding critical frequency can be determined based on a specified statistical confidence of level. With respect to a black spot, if the accident frequency induced by one accident cause is greater than the critical value, this cause can be treated as an accident induced factor of this black spot.
Hence, the process of determining accident induced factors of a black spot should follows the following three steps:

Step 1. *Select the initial accident causes (i.e., select the variables)*

Step 2. *Establish the evaluation criteria*

Step 3. *Determine the accident induced factors*

4.2 Variable selection based on discrete multi-variable algorithm

In terms of all possible accident causes (*i.e.*, independent variable), each of them is cross classified with the accident frequency (*i.e.*, dependent variable). Hence, a series of random event binary tables can be summarized. Calculate the Poisson statistical values of each table and choose the lowest P value of the variable as the first variable retained, while eliminate a non- significant variable. For the variables neither selected nor eliminated, each of them is also cross classified with the variable retained and the dependent variable. Then a series of random event triple tables are obtained. Calculate statistics, and then select one of the most significant variables, at the same time eliminate a non-significant variable. Repeat the above process, each time add a new variable, until all variables are selected or deleted. This algorithm is able to ensure that the selected variables have some significant effects on accident and can avoid the correlation problem between variables.

4.3Establishement of evaluation criteria based on rough set theory

The variables selected from step 1 can be subdivided into two categories according to the value of them. One is defined as the scalar variable whose value is the linguistic variable. For example, the accident cause of *speeding* is a scalar variable whose value is a linguistic variable of *yes* or *no*. The other is real-number variable whose value is a specified real number, such as the variable of *radius of horizontal curve*. Scalar variables are used for grouping, while the real-number variables are used as the calculation indexes, and then a cross-classification table can be obtained.

For each real-number variable *i* in each scalar variable *j* (*i.e*., group *j*), calculate the mean E_{ij} (*i.e*., the average accident frequencies) and standard deviation σ_{ij}, and then the critical accident frequency r_{ij} can be given as $r_{ij} = E_{ij} + k\sigma_{ij}$. Where, *k* is the number of standard deviations corresponding to the required confidence level.

When all r_{ij} are obtained, the evaluation criteria is established at last, which can be described by a series of vectors. For example, the vector $\mathbf{R}_k$ which contains the evaluation criteria for group *k* can be given as $\mathbf{R}_k = (r_{1k}, r_{2k}, \cdots, r_{lk})$. Where, *l* is the number of real-number variables in group *k*.

4.4 Identification of accident induced factors

When given a black spot, determine firstly the group that it is affiliated with, assuming that it belongs to the *k* group. Then, summarize or calculate the values q_{ik} of its real-number variables to get the target vector $\mathbf{Q}_k$, which is $\mathbf{Q}_k = (q_{1k}, q_{2k}, ..., q_{lk})$.

Finally, compare the corresponding elements of vector $\mathbf{Q}_k$ and $\mathbf{R}_k$. If index value q_{nk} in $\mathbf{Q}_k$ is greater than r_{nk} in $\mathbf{R}_k$, then for this black spot, the *n*th factor is the induced factor. Summarize all induced factors can obtain an accident induced factor set.

4.5 Case study for accident induced factor identification

The ten black spots determined by self-organizing competitive neural network are taken as the cases for accident induced factor identification. The geometric information of each site is listed in Table 3. By the application of discrete multi-variable algorithm, 10 accident elements are selected from the 99 causes summarized from the entire accidents. They are neglectful and careless, excessive speed, unfamiliarity with road conditions, fatigue, restricted sight distance, reduced pavement friction, small radius of horizontal curve (*i.e*., sharp curve), steeper grades, insufficient space between vehicles, and bad weather conditions.

Among these causes, neglectful and careless, excessive speed, the degree of familiarity with road conditions, fatigue and weather conditions are scalar variables, which are used for grouping. The other causes can be described by real-number, which are used as the evaluation indexes. Finally, based on rough set theory, a set of evaluation criteria is established. Comparing the accident indexes of the ten black spots with the corresponding standard, the accident induced factors for each black spot are obtained, which are listed in Table 4.

Table 3 – Horizontal and vertical alignment of black spots

Site	*Horizontal alignment*			*Vertical alignment*		
	Type	*Radius of Horizontal curve (m)*	*External angel(° ′ ″)*	*Type*	*Radius of Vertical curve (m)*	*Average gradient (%)*
Sta.42+590-sta.43+500	Transition	760	58°10′21.3″ (Right hand)	Level	∞	-2.314 (Downgrade)
Sta.48+300-sta.49+870	Circular curve	750	19°40′57.0″ (Right hand)	Crest	12000	-5.0 (Downgrade)
Sta.49+900-sta.52+000	S curves	620/600	35°28′03″(R) 49°14′38″(L)	Sag/ crest	22000/ 22000	-2.8/-3.9/-2.6
Sta.76+780-sta.78+700	Circular curve	1000	67°08′39″ (Left hand)	Level	∞	-2.0 (Downgrade)
Sta.1+600-sta.2+000	Tangent	∞	0	Sag	35000	-0.6 (Downgrade)
Sta.10+900-sta.11+200	Circular curve	7000	10°18′11.9″ (Left hand)	Crest	20000	-0.3 (Downgrade)
Sta.22+900-sta.23+000	Circular curve	2500	20°32′15.9″ (Left hand)	Level	∞	+2.5 (Upgrade)
Sta.46+800-sta.48+250	Circular curve	1000	35°10′58.0″ (Right hand)	Level	∞	-3.85 (Downgrade)
Sta.80+900-sta.82+900	C curves	1560/881	32°02′41″(R) 47°40′14″(R)	Crest	13685	-1.97/+1.24
Sta.102+150-sta.103+400	Circular curve	2000	8°54′11″ (Left hand)	Sag	30000	+0.9/-0.52

Table 4 – Accident induced factors of black spots

Site	*Neglectful and careless*	*Excessive speed*	*Unfamiliarity with road conditions*	*Fatigue*	*Restricted sight distance*	*Reduced pavement friction*	*Small radius of horizontal curve*	*Steeper grades*	*Insufficient space between vehicles*	*Bad weather*
Sta.42+590-sta.43+500										
Sta.48+300-sta.49+870										
Sta.49+900-sta.52+000										
Sta.76+780-sta.78+700										
Sta.1+600-sta.2+000										
Sta.10+900-sta.11+200										
Sta.22+900-sta.23+000										
Sta.46+800-sta.48+250										
Sta.80+900-sta.82+900										
Sta.102+150-sta.103+400										

Note: —*accident induced factors*.

5 CONCLUSIONS

Three correlative methodologies related to the identification of freeway black spots and accident induced factors are developed, which are (*i*) the methodology to divide the segment based on dynamic clustering algorithm, (*ii*) the self-organizing neural network model to

identify the black spots, and (*iii*) the approach to identify the accident induced factors of black spot based on discrete multi-variable algorithm combined with rough set theory.
The results show that (*i*) the segment division based on dynamic clustering algorithm can describe objectively the concentration and dispersion of accidents, (*ii*) the neural network model can quickly identify the black spots and the corresponding results are reasonable, and (*iii*) the methodology to identify the prominent accident causes can be used to establish a set of evaluation criteria and then to determine the accident induced factors of a black spot.

ACKNOWLEDGMENT

This project, *Research on Countermeasures to Improve the Safety Situations of Beijing-Zhuhai Freeway*, performed at the School of Transportation Science and Engineering of Harbin Institute of Technology, was sponsored by Guangdong Provincial Freeway Co. Ltd. and Guangdong Provincial Department of Transportation. The opinions, findings, and conclusions expressed in this paper are those of authors and not necessarily those of Guangdong Provincial Freeway Co. Ltd. and Guangdong Provincial Department of Transportation. The authors acknowledge engineers from the company and officers from the department for their continuous support throughout the project.

REFERENCES

1. MENG X H, SHENG H F, CHEN T E. (2008). Essence and an approach to identify the hazardous locations of urban arterial link based on BP artificial neural network. Journal of Highway and Transportation Research and Development, 25, 124-129.
2. Tarkoap, Kanodiam. (2004). Effective and fair identification of hazardous locations. Transportation Research Board.
3. ZHANG T J, TANG C C, ZHANG W H. (2006). Accident prone location identification on highway using double variable s filtration method. Journal of Highway and Transportation Research and Development, 23, 139-142.
4. PEI Y L, DAI T Y. (2005). Fuzzy evaluating method to distinguish black spot. Journal of Highway and Transportation Research and Development, 121-125.
5. Ngks, Hungwt, Wongwt. (2002). An algorithm for assessing the risk of traffic accident. Journal of Safety Research, 33, 387-410.
6. Zegeer C.V., Reinfurt D.W., Hunter W.W., Hummer J., Stewart R. and Herf L. (2004). Accident effects of sideslope and other roadside features on two-lane roads. Transportation Research Board.
7. Stamatiadis N., Jones S., Aultman-Hall L. Causal factors for accidents on south eastern low-volume rural roads. Transportation Research Board
8. MENG X H. (2000). Model applied to analyze the traffic accident's causes on black-spot of freeway. Journal of Harbin University of Architecture, 33, 113-116.
9. LI J L, SUN W H. (2005). Cause analysis of traffic accidents on express highway and study on their countermeasures. China Safety Science Journal, 15, 59-62.
10. LI D S, LIU K, ZHAO C. (2005). Application of SOFM neural network to categorized evaluation of road traffic accident. China Safety Science Journal, 15, 88-91.
11. YAO Z S, SHAO C F, LONG D L. (2005). Cause analysis of traffic accidents on road block spots based on rough set theory. China Safety Science Journal, 15, 107-109.