

A Real-time Breakdown Prediction Method for Urban Expressway On-ramp Bottlenecks

Yingjun Ye¹, Guoyang Qin¹, Jian Sun^{1,*}, and Qiyuan Liu^{1,2}

¹Department of Traffic Engineering & Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China

²Huachuan Transportation Technology CO. LTD, Suzhou, 215500, China

*Corresponding author e-mail: sunjian@tongji.edu.cn

Abstract. Breakdown occurrence on expressway is considered to relate with various factors. Therefore, to investigate the association between breakdowns and these factors, a Bayesian network (BN) model is adopted in this paper. Based on the breakdown events identified at 10 urban expressways on-ramp in Shanghai, China, 23 parameters before breakdowns are extracted, including dynamic environment conditions aggregated with 5-minutes and static geometry features. Different time periods data are used to predict breakdown. Results indicate that the models using 5-10 min data prior to breakdown performs the best prediction, with the prediction accuracies higher than 73%. Moreover, one unified model for all bottlenecks is also built and shows reasonably good prediction performance with the classification accuracy of breakdowns about 75%, at best. Additionally, to simplify the model parameter input, the random forests (RF) model is adopted to identify the key variables. Modeling with the selected 7 parameters, the refined BN model can predict breakdown with adequate accuracy.

1. Introduction

Urban expressways, as the backbone of urban transport network, undertake a large amount of urban motorized traffic. Bottlenecks are considered as the major cause of traffic congestion on urban expressways. Among them, congestion at recurring bottlenecks occurs periodically [1]. Thus, it's essential to understand the traffic flow features at recurring bottlenecks to alleviate traffic congestion. Regarding traffic breakdown, numerous studies have been conducted on breakdown identification [2], stochastic nature of capacity [3,4], breakdown probability model [4,5], early-onset breakdown phenomenon and mechanism [1,6], and so on. However, the real-time prediction of traffic breakdown events is absent. Since traffic breakdown is caused by the spatial-temporal evolution of traffic conditions, it is suggested to be closely related with previous environment factors.

For these reasons, the objectives of this paper are to investigate: (1) the viability of using dynamic data for real-time breakdown prediction; (2) the universality of breakdown prediction models (i.e. to validate the validity of one unified model comparing with separate models at different sites). (3) the identification of critical factors of breakdown occurrence from various variables.

To address above issues, firstly, with respect to 3183 breakdown events and 21720 non-breakdown events collected from 10 isolated recurring bottlenecks on two expressways in Shanghai, China,



corresponding traffic flow data and facilities features are collected. Then, an effective classification algorithm, BN model is adopted to predict the breakdown. Separate models for each bottleneck are developed and compared with the unified model. What's more, among various variables, critical factors are selected by RF model to avoid high computational complexity and over-fitting.

This paper is organized as follows: Section 2 describes the study site and data used in this study. The methodology used in this study is introduced in section 3. Then, the results and discussions are presented in section 4. Section 5 concludes the paper and provides suggestions for future research.

2. Site Description and Data Preparation

2.1. Study Sites

10 isolated recurring bottlenecks from the Inner Ring Expressway and Yan'an Expressway in Shanghai, China are selected for the study. Table 1 presents the detailed geometry features.

Table 1. Geometry Features of Study Sites

Static feature	No. of main lane	Average lane width (m)	Left-side L.C. (m)	Right-side L.C. (m)	Ramp density (/km)	Merging length (m)	Curvature ($\times 10^{-4}$)
Symbol	No _m	Wid	LC _L	LC _R	RD	Len	Cur
Wuning Rd. S.	2	3.24	0.16	0.27	1.5	157	3.1
Wuning Rd. N.	2	3.22	0.24	0.30	1.5	49	6.98
Guangzhong Rd. N.	3	3.21	0.13	0.28	1	256	7.37
Zhoujiazui Rd. N.	2	3.86	0.24	0.28	1	86	0
Wanpingnan Rd. S.	2	3.2	0.22	0.25	1	140	6.59
Maoming Rd. N.	4	3.76	0.35	0.35	1.5	145	3.1
Jiangsu Rd. S.	3	3.75	0.36	0.36	1	145	8.92
Hongxu Rd. S.	3	3.76	0.34	0.34	1	137	0.39
Hongjing Rd. S.	3	3.76	0.36	0.33	0.5	215	0.78
Huashan Rd. S.	3	2.9	0.06	0.32	1	94	51.58
min	2	2.9	0.06	0.25	0.5	49	0
max	4	3.86	0.36	0.36	1.5	256	51.58
mean	2.7	3.47	0.25	0.31	1.1	142.4	8.88
stdev	0.68	0.34	0.11	0.04	0.32	60.26	15.34

NOTE: No. = number; L.C. = lateral clearance; stdev = standard deviation.

2.2. Traffic Data and Breakdown Identification

Aggregated 5-min traffic flow data are collected from the dual-loop detectors (i.e. volume, speed, occupancy, traffic composition). In addition, the weather data, and several time variables are also collected. All the dynamic variables used in the study are listed in Table 2.

Table 2. Dynamic Variables for Prediction Model

Dynamic feature	Symbol	Dynamic feature	Symbol
Upstream volume	$Q_u(t)$	Downstream occupancy	$O_d(t)$
Upstream speed	$V_u(t)$	Large vehicle ratio	R_l
Upstream occupancy	$O_u(t)$	Medium vehicle ratio	R_m
On-ramp volume	$Q_o(t)$	Small vehicle ratio	R_s
On-ramp speed	$V_o(t)$	Merging ratio	MR
On-ramp occupancy	$O_o(t)$	Day of week	DoW
Downstream volume	$Q_d(t)$	Time of day	ToD
Downstream speed	$V_d(t)$	Weather	Wth

In accordance with previous studies [1,6], we define the breakdown occurrence with speed drops below 45km/h and occupancy exceeds 25% lasting for at least 15 minutes. Moreover, non-breakdowns are also collected to present normal conditions. The detail identifications are listed as follows.

1) Find the data satisfy two thresholds within three consecutive 5-min intervals:

$$\{\max\{V_{t_{BD}}, V_{t_{BD}+5min}, V_{t_{BD}+10min}\} \leq 45\text{km/h}\} \wedge \{\min\{O_{t_{BD}}, O_{t_{BD}+5min}, O_{t_{BD}+10min}\} \geq 25\%\} \quad (1)$$

2) Identify non-breakdowns before $t_{BD} - 15min$ per day with volume not less than 1000 veh/h/ln as an additional constraint to improve the practicability of predictive model which satisfy:

$$\{\min\{V_{t_{NB}}, V_{t_{NB}+5min}, V_{t_{NB}+10min}\} \geq 45\text{km/h}\} \vee \{\max\{O_{t_{NB}}, O_{t_{NB}+5min}, O_{t_{NB}+10min}\} \leq 25\%\} \\ \wedge \{\min\{Q_{t_{NB}}, Q_{t_{NB}+5min}, Q_{t_{NB}+10min}\} \geq 1000\text{veh/h/ln}\} \quad (2)$$

where the t_{BD}/t_{NB} are the starting time of the breakdown/non-breakdown, respectively.

Three previous consecutive 5 min intervals are used for prediction of breakdowns, that is 5~10/10~15/15~20 minutes period data before t_{BD} and t_{NB} , respectively. As a result, 10 bottlenecks with four interval data come into a 10×4 sets of data. The separate models for each bottleneck are built with each set of data named by $Model_s^t$, where s represents the number of bottleneck, $s \in [1,10]$ and t stands for the time interval, $t \in \{t_{BD}, t_{BD} - 5min, t_{BD} - 10min, t_{BD} - 15min\}$.

3. Methodology

BN, also known as belief network, is a statistical inference approach based on Bayesian decision and graph theory. Since the predictive modelling of breakdown is in essence a classification problem, the BN, which has been used previously to deal with crash prediction [7,8], is adopted in this paper. The key of RN, Bayesian theory, can be denoted as:

$$P(C_i|\mathbf{x}) = P(\mathbf{x}|C_i)P(C_i)/P(\mathbf{x}) \quad (3)$$

Two tasks are important for BN:(1) Structure learning, i.e. finding the optimal network topology from all possible structures. (2) Parameter learning, i.e. calculating parameters of the conditional probability distribution. In this paper, the K2 algorithm and Bayes update algorithm are adopted for structure learning and parameter learning. Figure 1 presents the whole procedure of the prediction.

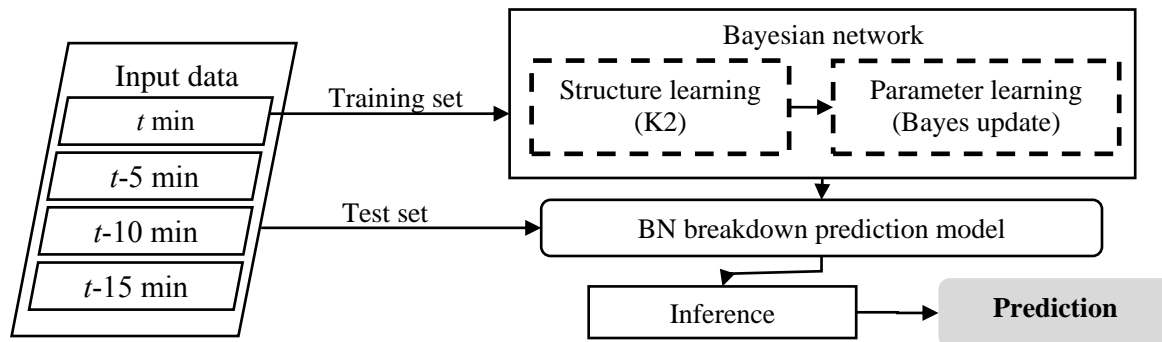


Figure 1. Procedure of Breakdown Prediction

4. Results

4.1. Evaluation Metrics

Due to the small ratio of breakdowns, to perform an comprehensive evaluation, four additional indicators based on the confusion matrix as shown in Table 3 are used in this study.

Table 3. Confusion Matrix

	Predicted breakdowns	Predicted non-breakdowns
Real breakdowns	True positive (<i>TP</i>)	False negative (<i>FN</i>)
Real non-breakdowns	False positive (<i>FP</i>)	True negative (<i>TN</i>)

Therefore, the five metrics are defined and stated mathematically as follows:

(1) Overall accuracy (OA), it presents the ratio of correct prediction over all samples;

$$OA = (TP + TN) / (TP + FN + FP + TN) \quad (4)$$

(2) Breakdown accuracy (BD Acc.), it denotes the accuracy of predicting breakdowns;

$$BD\ Acc. = TP / (TP + FN) \quad (5)$$

(3) Non-breakdown accuracy (NB Acc.), it denotes the accuracy of predicting non-breakdowns;

$$NB\ Acc. = TN / (FP + TN) \quad (6)$$

(4) Averaging BD Acc. and NB Acc. is the balanced accuracy, or BAC;

$$BAC = (BD\ Acc. + NB\ Acc.) / 2 \quad (7)$$

(5) F-value (F) denotes ability of model to detect breakdowns.

$$Precision = TP / (TP + FP) \quad (8)$$

$$Recall = TP / (TP + FN) \quad (9)$$

$$F = 2 \cdot Precision \cdot Recall / (Precision + Recall) \quad (10)$$

4.2. Results of BN Models

Followed by the identification of breakdown in Section 2.2, a total of 3183 breakdown events and 21720 non-breakdown events are collected from 10 isolated recurring bottlenecks. Ten separate BN models and one unified model were developed for prediction. Results are shown in Table 4.

Table 4. Prediction Results of all models (in percentage)

Prediction	Indicator	Modelling separately											All
		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Mean	
<i>Model_{BD}^S</i>	OA	86.89	95.18	96.17	94.66	97.78	95.11	89.52	97.27	97.41	94.32	94.43	93.02
	BD Acc.	100.00	95.29	92.52	90.22	100.00	95.65	100.00	90.10	100.00	97.00	96.08	100.00
	NB Acc.	84.07	95.14	98.06	95.21	97.01	95.06	86.77	98.76	97.00	93.58	94.07	91.48
	BAC	92.04	95.22	95.29	92.71	98.50	95.36	93.39	94.43	98.50	95.29	95.07	95.74
	F-value	72.93	91.01	94.29	78.67	95.89	77.19	79.84	91.92	91.46	88.18	86.14	83.83
<i>Model_{t_{BD}-5min}^S</i>	OA	63.93	75.60	72.38	71.53	74.45	68.58	81.76	76.45	75.88	75.98	73.65	75.27
	BD Acc.	69.07	62.35	75.93	66.30	81.90	67.92	83.84	79.21	83.52	68.00	73.80	75.37
	NB Acc.	62.83	80.16	70.53	72.17	71.85	68.65	81.22	75.88	74.65	78.21	73.61	75.25
	BAC	65.95	71.26	73.23	69.24	76.88	68.29	82.53	77.54	79.08	73.11	73.71	75.31
	F-value	40.36	56.68	65.34	33.70	62.32	29.75	65.61	53.69	49.03	55.28	51.18	52.44
<i>Model_{t_{BD}-10min}^S</i>	OA	56.57	70.48	67.30	68.56	66.58	54.71	78.20	76.28	67.68	74.02	68.04	72.52
	BD Acc.	60.82	68.24	64.22	82.61	74.29	66.67	79.80	77.23	61.96	65.00	70.08	68.76
	NB Acc.	55.65	71.26	68.93	66.84	63.91	53.39	77.78	76.08	68.62	76.54	67.90	73.36
	BAC	58.24	69.75	66.58	74.73	69.10	60.03	78.79	76.66	65.29	70.77	68.99	71.06
	F-value	33.15	54.21	57.61	36.45	53.42	22.71	60.31	52.88	34.97	52.21	45.79	47.60
<i>Model_{t_{BD}-15min}^S</i>	OA	64.85	63.86	63.26	70.70	71.43	53.42	73.17	70.48	65.49	72.71	66.93	70.50
	BD Acc.	60.82	70.59	78.70	81.52	60.95	61.11	67.68	68.32	47.78	58.00	65.55	70.79
	NB Acc.	65.71	61.54	55.12	69.37	75.08	52.57	74.60	70.93	68.33	76.82	67.01	70.44
	BAC	63.27	66.06	66.91	75.45	68.02	56.84	71.14	69.62	58.05	67.41	66.28	70.61
	F-value	37.94	50.00	59.65	37.78	52.46	20.75	51.15	44.37	27.65	48.13	42.99	46.55

For the separate models, *Model_{BD}^S* are built for reference with highest accuracies. The results show the viability of BN models for breakdowns prediction. *Model_{t_{BD}-5min}^S* performs better among other models comparing their OAs, BACs and F-values. Namely, the model can provide the best prediction using data collected 5-10 min prior to breakdowns. It should be pointed out that the unified model for all bottlenecks performs better than most separate models, with a best classification accuracy of breakdowns about 75%.

4.3. Key Variable Identification

Previous results suggested that the model can provide the best prediction using data collected 5-10 min prior to breakdowns. However, the model is developed with a total of 23 static and dynamic variables in Table 1 and 2 which may cause high computational complexity. Thus, simplify the model input, random forests model, is adopted to identify critical variables for breakdown prediction. Among all 23 variables, the top seven most influential variables are Od, Vd, Qd, Qo, Vo, Vu and ToD. Then, the new unified BN model built with data 5-10 min before breakdowns is compared with the previous BN model. As shown in Figure 2, the results indicate that the breakdown prediction of the refined model is maintained with much fewer variables.

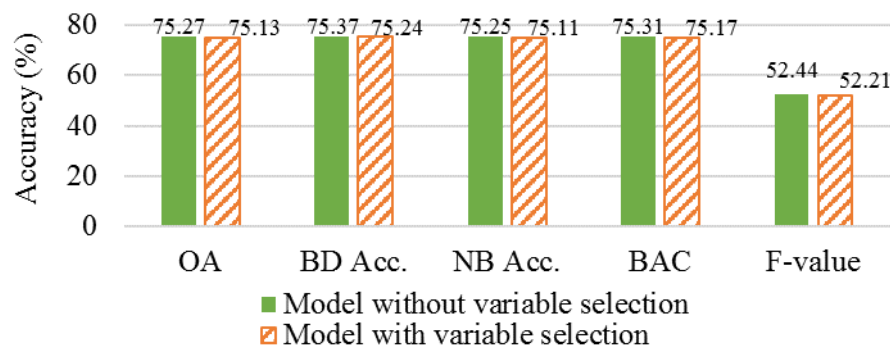


Figure 2. Comparison of Models with/without Variable Selection

5. Conclusion

This paper aims to investigate the relationships between breakdowns and various influential factors. With the real-time prediction of breakdowns, traffic management strategies for prevention can be implemented in advance. Therefore, based on 3183 breakdowns and 21720 non-breakdowns events, corresponding traffic data are collected from 10 isolated recurring bottlenecks in Shanghai, China, BN models are developed for prediction. The main conclusions are as follows:

1. With three time-interval data prior to breakdowns/non-breakdowns, the model $Model_{t_{BD}-5min}^s$, that is model using the 5~10 min data prior to events, performing best among other models, whose prediction accuracy indicators all higher than 73%.
2. One unified model for all bottlenecks is also built to validate the universality of breakdown prediction model. By comparing the results with separated models, it shows that the unified model works better than most separated models, with a best classification accuracy of breakdowns about 75%. Therefore, the feasibility of one model for all bottlenecks is verified.
3. The RF model is developed to select the most important variables for the establishment of the breakdown prediction model. Among 23 dynamic and static variables, 7 key factors are selected. Results show that the refined BN model with 7 significant variables can predict breakdown with adequate accuracy.

Acknowledgments

The authors would like to thank the Natural Science Foundation of China (51278362, 51422812), the New Century Excellent Talents in University (NCET13-0425) and the Fundamental Research Funds for the Central Universities for supporting this research.

References

- [1] Sun, J., J. Zhang, and H. M. Zhang. Investigation of Early-onset Breakdown Phenomenon at Urban Expressway Bottlenecks in Shanghai, China. *Transporter B: Transport Dynamics*, Vol. 2, No.3, 2013, pp. 215–228.
- [2] Cassidy, M. J., and R. L. Bertini. Some Traffic Features at Freeway Bottlenecks. *Transportation Research Part B*, Vol. 33, No. 1, 1999, pp. 25–42.
- [3] Elefteriadou, L., R. P. Roess, and W. R. McShane. The Probabilistic Nature of Breakdown at Freeway-merge Junctions. In *Transportation Research Record: Journal of the Transportation Research Board*, NO. 1484, Transportation Research Board of the National Academies, Washington, D.C., 1995, pp. 80–89.
- [4] Brilon, W., J. Geistefeldt, and M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proc., 16th Int. Symp. on Transportation and Traffic Theory*, 2005, pp. 125–144.
- [5] Sun, J., and J. Zhang. Survival Analyses of Traffic Flow Breakdown at Urban Expressway Bottlenecks. *Journal of Tongji University (Natural Science)*. Vol. 41, No. 4, 2013, pp. 530–

- 535.
- [6] Sun, J., L. Zhao, and H. M. Zhang. Mechanism of Early-onset Breakdown at On-Ramp Bottlenecks on Shanghai, China, Expressways. In *Transportation Research Record: Journal of the Transportation Research Board*, NO. 2421, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 64–73.
 - [7] Sun, J., and J. Sun. A Dynamic Bayesian Network Model for Real-time Crash Prediction Using Traffic Speed Conditions Data. *Transportation Research Part C Emerging Technologies*, Vol. 54, 2015, pp. 176–186.
 - [8] Oña, J., G. López, R. Mujalli, and F. J. Calvo. Analysis of Traffic Accidents on Rural Highways Using Latent Class Clustering and Bayesian Networks. *Accident Analysis and Prevention*, Vol. 51, 2013, pp. 1–10.