

Diagnosis of causes for high railway traffic based on Bayesian network

Yingchao Zhou, Pinchao Meng*

College of Science, Changchun University of Science and Technology, Changchun 130022, China

Corresponding Author Email: mengpc@cust.edu.cn

<https://doi.org/10.18280/mmep.060118>

ABSTRACT

Received: 17 September 2018

Accepted: 10 December 2018

Keywords:

rail traffic, high traffic, Bayesian network, fuzzy set

To intuitively analyze and diagnose why high railway traffic often occurs, this paper builds a Bayesian network on a rail transit system, for instance, the Shanghai Metro L1, to explore it from different angles such as holidays, mega-events, bad weather and sudden accidents. The fuzzy set theory is also integrated to quantify the conditional probability of some events using the fuzzy languages. It turns out that the Bayesian network constructed can well infer the probability of massive passenger traffic, and diagnose the dominant factors that may cause it. Hereby, the rail transit management authorities can take proper measures against it to reduce the risks the high rail traffic may impose.

1. INTRODUCTION

As rail transit features large capacity, safe and efficient operation, in the last decade, China's rail construction has boomed. Rail transit gives the people facility on work and life but there is still a risk of huge passenger traffic. As a large-scale communal facility, the rail transport system is complex and has a closed installation. Dense population and special geography are always found around there, hence there is a variety of factors that may cause high passenger traffic in rail transit trains, so that different response measures adopted by the rail transit management authorities vary. Thus obviously, it is particularly important to timely diagnose the causes for the occurrence of huge passenger traffic. Since the Bayesian network is not only a model based on probabilistic reasoning but also a risk quantification method for managing the risks, it enables the inference from incomplete, inaccurate or uncertain knowledge or information. For this purpose, this paper introduces the principle of Bayesian network to construct a Bayesian network for diagnosing why large passenger traffic occurs in rail transit systems.

In the last decade, the study of the Bayesian network has directed to solve practical problems based on the real conditions. Ozgur [1] investigated to what extent the subjective mental health of the UK public is subjected to the air pollution, income maintenance, housing benefits, and household income with the Bayesian network. Ying [2] developed a Bayesian network regarding investment portfolios to analyze how the new product development performance evolve with the different portfolios. Kathrin [3] applied a Bayesian network to simulate the cyclic pressure test system from which semiconductor lifetime data could be available. Nepal [4] et al. quantified the procurement risks with the Bayesian network and incorporated the risks into expected costs to determine the optimal suppliers for products. Laetitia [5] applied the Bayesian network to weigh all decisions that could facilitate the sustainable development of aquaculture. Marsh [6] developed a Bayesian network for railway operation accidents to analyze why the railway accidents occur. Jonathan

Agner et al. [7] introduced the Bayesian network to estimate the life expectancy of bone graft patients, thereby determining whether or not to have a bone graft surgery and which surgical procedure would be performed on patients.

When the Bayesian network is used to perform the probabilistic reasoning, the prior probabilities are first required [8]. In many studies involving the Bayesian networks, the prior probabilities of events are often identified. When it is unknown, the common method as used currently is the expert evaluation that will inevitably leads to subjective errors [9-11]. As a result, it is particularly important for the Bayesian network to define these unknown prior probabilities and minimize subjective errors. For this purpose, this paper incorporates the fuzzy set theory [12] into the Bayesian network, and applies it to determine the unknown prior probability.

This paper first builds a Bayesian network to explore why the large passenger traffic occurs. Next, the fuzzy set theory is introduced for the prior probability unknown in Bayesian networks. Then, a positive analysis is carried out on the real railway systems, for example, the Shanghai Metro L1.

2. BAYESIAN NETWORK CONSTRUCTION

There are 5 factors that may lead to high passenger traffic in rail transit: morning-evening peaks, holidays, major events, bad weather and sudden accidents, each of which will function separately. However, these factors are not independent of each other. Among them, holidays can lead to the host of various great events, and bad weather can cause the sudden accidents. The huge passenger traffic can result in the congestion, which may in turn raise the stampede accidents. Since it is certain there will be huge passenger traffic in the morning and evening rush hours, here the four factors, i.e. holidays, major events, bad weather and sudden accidents, are chosen for constructing a Bayesian network to dig out the root of high passenger traffic in rail transit, as shown in Figure 1.

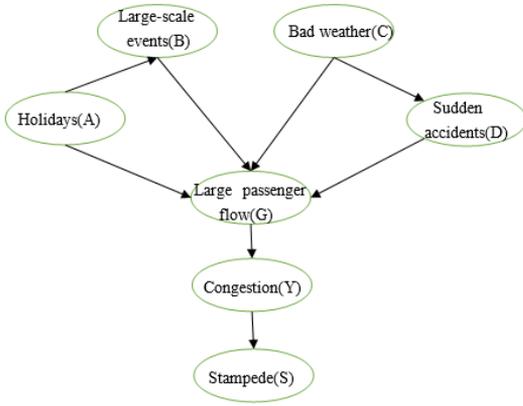


Figure 1. Bayesian network for the cause of huge passenger traffic

can determine the probability of an event. In 1992, Wickens [13] classified the probability of event occurrence into seven natural languages in the *Engineering psychology and human performance*, they are Very High (VH), High (H), Fairly High (FH), Medium (M), Fairly Low (FL), Low (L) and Very Low (VL). This paper replaces the natural languages with triangle or trapezoid fuzzy numbers, as shown in Fig. 2. Triangle fuzzy number is represented as $A = (a, b, c)$; trapezoid fuzzy number is $A = (a, b, c, d)$; the membership functions are expressed respectively

$$A(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ \frac{c-x}{c-b}, & b < x \leq c \\ 0, & x > c \end{cases}$$

$$A(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & b < x \leq c \\ \frac{d-x}{d-c}, & c < x \leq d \\ 0, & x > d \end{cases}$$

3. PROBABILISTIC ANALYSIS AND DETERMINATION BASED ON FUZZY SET THEORY

3.1 Processing fuzzy number for expert language description

Expert evaluation is one of the commonly used methods that

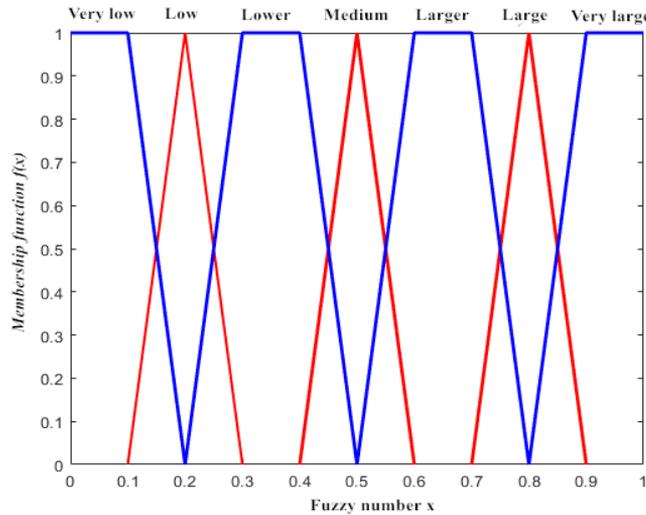


Figure 2. Natural language description of fuzzy numbers

The forms of its fuzzy numbers and λ cut set are shown in Table 1.

Table 1. Forms of fuzzy numbers and λ cut set

Fuzzy number form	λ cut set
$f_{VL} = (0.0, 0.1, 0.2)$	$f_{VL} = [0.1\lambda + 0, -0.1\lambda + 0.2]$
$f_L = (0.1, 0.2, 0.3)$	$f_L = [0.1\lambda + 0.1, -0.1\lambda + 0.3]$
$f_{FL} = (0.2, 0.3, 0.4, 0.5)$	$f_{FL} = [0.1\lambda + 0.2, -0.1\lambda + 0.5]$
$f_M = (0.4, 0.5, 0.6)$	$f_M = [0.1\lambda + 0.4, -0.1\lambda + 0.6]$
$f_{FH} = (0.5, 0.6, 0.7, 0.8)$	$f_{FH} = [0.1\lambda + 0.5, -0.1\lambda + 0.8]$
$f_H = (0.7, 0.8, 0.9)$	$f_H = [0.1\lambda + 0.7, -0.1\lambda + 0.9]$
$f_{VH} = (0.8, 0.9, 1.0)$	$f_{VH} = [0.1\lambda + 0.8, -0.1\lambda + 1.0]$

In order to minimize the error, this paper uses the arithmetic mean method to synthesize the evaluation results of multiple experts. The comprehensive evaluation of n experts can be expressed as

$$Q_i = \frac{\sum_{j=1}^m f_{ij}}{n} = \frac{f_{i1} + f_{i2} + \dots + f_{im}}{n} \quad j = 1, 2, \dots, m; i = 1, 2, \dots, n \quad (1)$$

where Q_i represents the fuzzy occurrence probability of the event i ; f_{ij} represents the fuzzy value of the event i evaluated by the expert j ; m is the number of events; n is the number of experts.

3.2 Defuzzification of fuzzy numbers

To obtain the prior or conditional probabilities of events at each node of the Bayesian network, it is particularly important to convert the fuzzy value of each event into a certain value. Currently, the commonly used methods for defuzzification include the integral value, center of gravity, weighted mean, etc. [14-15], among which, the integral value method is most concise and convenient, and easy to understand, so that this

paper takes this method, and its formula is

$$I(Q) = (1 - \varepsilon)I_R(Q) + \varepsilon I_L(Q) \quad (2)$$

where $I(Q)$ is the defuzzification value of fuzzy number Q ; ε is the optimistic coefficient, $\varepsilon \in [0,1]$, when $\varepsilon=0$ or 1 , it respectively corresponds to the upper and lower bounds of defuzzification value for fuzzy number Q , and when $\varepsilon=0.5$, it is the representative value of the defuzzification of fuzzy number Q ; $I_L(Q)$ and $I_R(Q)$ are the integral values of the inverse membership functions to the left and right of the fuzzy number, respectively; for the trigonometric function, $I_L(Q)$ and $I_R(Q)$ can be represented with λ cut sets

$$I_R(Q) = \frac{1}{2} [\sum_{\lambda=0.1}^1 \lambda_R(Q) \Delta\lambda + \sum_{\lambda=0}^{0.9} \lambda_R(Q) \Delta\lambda] \quad (3)$$

$$I_L(Q) = \frac{1}{2} [\sum_{\lambda=0.1}^1 \lambda_L(Q) \Delta\lambda + \sum_{\lambda=0}^{0.9} \lambda_L(Q) \Delta\lambda] \quad (4)$$

where $\lambda_R(Q)$ and $\lambda_L(Q)$ are the upper and lower bounds of λ cut set of the fuzzy number Q ; $\lambda = 0, 0.1, 0.2, \dots, 1$; $\Delta\lambda = 0.1$.

4. POSITIVE ANALYSIS

This paper takes the Shanghai Metro L1 as a test analysis case to build a Bayesian network for digging out the root of huge passenger traffic in the subway.

4.1 Determine the prior or conditional probabilities of event

With reference to relevant statistics of the Shanghai Metros, the reports of major news media on events such as huge passenger traffic in metros, sudden accidents, the meteorological report of Shanghai Meteorological Bureau and the mega events announcement posted by relevant authorities in Shanghai, this paper makes a statistics on relevant data from massive passenger traffic and sudden accidents occurred in Shanghai Metro L1 from 2013 to 2016, filtered and calculated with the MATLAB software. Then the prior and partially conditional probabilities of event occurrence at each node are available, see below for details. It is TRUE if there is a huge passenger traffic, abbreviated T, otherwise it is FALSE, abbreviated F:

①Holiday (A)		
A	T	F
P(A)	0.1205	0.8795
②Mega-event (B)		
B	T	F
P(B)	0.4219	0.5781
③Bad weather (C)		
C	T	F
P(C)	0.0329	0.9671
④Sudden accident (D)		
D	T	F
P(D)	0.0164	0.9836
⑤Holiday (A)-Mega-event (B)		
B	P(B A=T)	P(B A=F)
T	1	0.3178
F	0	0.6822
⑥Bad weather (C)-sudden accident (D)		
D	P(D C=T)	P(D C=F)

T	0.1667	0.0113
F	0.8333	0.9887

With the expert evaluation, 4 experts in metro safety operation are chosen to evaluate dependent events, thereby the descriptive natural language can be available for the fuzzy probability values of relevant events, as shown in Table 2. Based on the advice of four experts, the arithmetic mean is taken to obtain a comprehensive fuzzy probability value.

Table 2. Language description of related events

Expert	Event X	Event Y	Event M	Event N
1	Very high	Very high	Very high	Lower
2	Very high	Very high	High	Low
3	Very high	High	High	Lower
4	High	High	High	Low

Among them, the event X represents the conditional probability of occurrence G when A and B all occur; the event Y represents the conditional probability of occurrence of G when A occurs but B does not occur; the event M denotes the conditional probability of occurrence of G when A does not occur but B occurs; the event N means the conditional probability of occurrence of G when A and B all do not occur.

According to the formula (1), the average fuzzy numbers of events X, Y, M, and N can be available

$$Q_X = [0.1\lambda + 0.775, -0.1\lambda + 0.975]$$

$$Q_Y = [0.1\lambda + 0.75, -0.1\lambda + 0.95]$$

$$Q_M = [0.1\lambda + 0.725, -0.1\lambda + 0.925]$$

$$Q_N = [0.1\lambda + 0.15, -0.1\lambda + 0.4]$$

After the average fuzzy numbers of events X, Y, M, and N are available, when $\varepsilon=0.5$, according to the formulas (2) ~ (4), the probabilities of event X, Y, M, and N can be obtained, respectively, that is, conditional probability of G relative to A and B. The results are shown in Table 3. Similarly, the conditional probability of G relative to C and D can be available. The results are shown in Table 4.

Table 3. Conditional probability of G relative to A and B

A	B	P(G=T A,B)	P(G=F A,B)
T	T	0.875	0.125
T	F	0.85	0.15
F	T	0.825	0.175
F	F	0.275	0.725

Table 4. Conditional probability of G relative to C and D

C	D	P(G=T C,D)	P(G=F C,D)
T	T	0.9	0.1
T	F	0.85	0.15
F	T	0.875	0.125
F	F	0.275	0.725

4.2 Inference of Bayesian network

After determining the prior or conditional probabilities of

event at each node, the probability of occurrence of the node event G can be obtained by the constructed Bayesian network, i.e.

$$\begin{aligned}
 P(G = T) &= \sum P(A; B; C; D; D = T) \\
 &= P(A = T, B = T, C = T, D = T, G = T) \\
 &+ P(A = T, B = T, C = F, D = T, G = T) + \dots \\
 &+ P(A = F, B = F, C = T, D = F, G = T) \\
 &+ P(A = F, B = F, C = F, D = F, G = T)
 \end{aligned} \tag{5}$$

The descriptive expression and sorting order of whether events A, B, C, D, G occur in formula (5) are shown in Table 5.

Table 5. Descriptive expression and sorting order of whether events A, B, C, D, G occur

Serial number	A	B	C	D	G
1	T	T	T	T	T
2	T	T	F	T	T
3	T	F	T	T	T
4	T	F	F	T	T
5	T	T	T	F	T
6	T	T	F	F	T
7	T	F	T	F	T
8	T	F	F	F	T
9	F	T	T	T	T
10	F	T	F	T	T
11	F	F	T	T	T
12	F	F	F	T	T
13	F	T	T	F	T
14	F	T	F	F	T
15	F	F	T	F	T
16	F	F	F	F	T

With the constructed Bayesian network, a joint probability formula can be available for the node events A, B, C, D, G, i.e.

$$P(A, B, C, D, G) = P(A) \cdot P(B|A) \cdot P(G|A, B) \cdot P(C) \cdot P(D|C) \cdot P(G|C, D) \tag{6}$$

Substitute (6) into (5), then

$$\begin{aligned}
 P(G = T) &= \sum P(A; B; C; D; D = T) \\
 &= P(A = T) \cdot P(B = T|A = T) \cdot P(G = T|A = T, B = T) \\
 &\cdot P(C = T) \cdot P(D = T|C = T) \cdot P(G = T|C = T, D = T) \\
 &+ \dots + \\
 &P(A = F) \cdot P(B = F|A = F) \cdot P(G = T|A = F, B = F) \\
 &\cdot P(C = F) \cdot P(D = F|C = F) \cdot P(G = T|C = F, D = F)
 \end{aligned} \tag{7}$$

The probability of event at each node, as given above, and the conditional probability obtained are substituted into the formula (6), the joint probabilities of A, B, C, D, G can be calculated under different conditions, as shown in Table 6. According to the formulas (6) and (7), the probability that G occurs, $P(G=T)=0.1476$, and the probability that G does not occur, $P(G = F) = 0.8524$, can be calculated.

Table 6. Joint probability under different conditions

Serial number	A	B	C	D	G	P(A,B,C,D,G)
1	T	T	T	T	T	0.0001
2	T	T	F	T	T	0.0001
3	T	F	T	T	T	0
4	T	F	F	T	T	0
5	T	T	T	F	T	0.0025
6	T	T	F	F	T	0.0277
7	T	F	T	F	T	0
8	T	F	F	F	T	0
9	F	T	T	T	T	0.0001
10	F	T	F	T	T	0.0022
11	F	F	T	T	T	0.0001
12	F	F	F	T	T	0.0016
13	F	T	T	F	T	0.0054
14	F	T	F	F	T	0.0606
15	F	F	T	F	T	0.0038
16	F	F	F	F	T	0.0434

4.3 Diagnosis of the cause for huge passenger traffic

If the event G has occurred, that is, there is a huge passenger traffic in the rail transit, then the constructed Bayesian network can reversely calculate the posterior probability that other events occur based on the prior probability for relevant event and conditional probability as obtained above. The posterior probability is based to diagnose the most likely cause that G occurs. Therefore, the probability that the event A occurs in the case where the known event G occurs can be obtained, that is, $P(A=T|G=T)=0.206$. Similarly, $P(B=T|G=T)=0.6687$, $P(C=T|G=T)=0.0813$, $P(D=T|G=T)=0.0285$.

After the posterior probabilities of 4 events are compared, it is known that:

(1) If the massive passenger traffic occurs during the off-peak period, the most likely factors that may cause it are the mega events, holidays, bad weather, and sudden accidents. The rail transit manager only needs to diagnose it in the above sequence. According to the diagnosis results, the appropriate measures can be taken.

(2) If the large passenger traffic occurs at the rush hour, although the Bayesian network structure gets more complex, the general diagnosis theory is alike. Since there will be a huge passenger traffic in rush hours, the first factor should be considered as the peak periods, followed by the hosting of mega events, holidays, bad weather, and unexpected accidents.

5. CONCLUSION

This paper introduces the Bayesian network into the field of rail transit, and builds it for exploring why there is a higher rail traffic. The fuzzy set theory is also integrated to quantify the conditional probability of some events. Now it is proved that the constructed Bayesian network can well infer the probability that huge passenger traffic occurs, and timely diagnose the dominant factors that may cause such traffic in rail transit. It is indeed significant in safeguarding the smooth operation and imperfecting the risk management of rail transit.

ACKNOWLEDGEMENTS

Project supported by National Natural Science Foundation of China (51378076).

REFERENCE

- [1] Ozdamar O, Giovanis E. (2017). The causal effects of income support and housing benefits on mental well-being: an application of a Bayesian network. *Metroeconomica* 68(3): 398-424. <https://doi.org/10.1111/meca.12131>
- [2] Yang Y, Xu DL. (2017). A methodology for assessing the effect of portfolio management on NPD performance based on Bayesian network scenarios. *Expert Systems* 34(2): 1-10. <https://doi.org/10.1111/exsy.12186>
- [3] Kathrin P, Olivia B, Jurgen P. (2015). Bayesian network model with application to smart power semiconductor lifetime data. *Risk Analysis* 35(9): 1623-163. <https://doi.org/10.1111/risa.12342>
- [4] Nepal B, Prakash OY. (2015). Bayesian belief network-based framework for sourcing risk analysis during supplier selection. *International Journal of Production Research* 53(20): 6114-6135. <https://doi.org/10.1080/00207543.2015.1027011>
- [5] Laetitia H, Marie S, Cecile B. (2013). Capturing Ecosystem Services, Stakeholders' preferences and trade-offs in coastal aquaculture decisions: a bayesian belief network application. *PLoS ONE* 8(10): 1-12. <https://doi.org/10.1371/journal.pone.0075956>
- [6] Marsh W, Bearfield G. (2012). Using bayesian networks to model accident causation in the UK railway industry [EB/OL].[2012-02-10].[http://www.eecs.qmul.ac.uk/William/PSAM7/accident causation PSAM7.pdf](http://www.eecs.qmul.ac.uk/William/PSAM7/accident%20causation%20PSAM7.pdf).
- [7] Jonathan AF, John E, Patrick JB, Rikard W, John HH. (2011). Estimating survival in patients with operable skeletal metastases: An Application of a Bayesian Belief Network. *PLoS ONE* 6(5): 1-7. <https://doi.org/10.1371/journal.pone.0019956>
- [8] Neda T, Mandy K, Christopher K. (2019). Predicting ecosystem components in the Gulf of Mexico and their responses to climate variability with a dynamic Bayesian network model. *PLoS ONE* 14(1): 1-23. <https://doi.org/10.1371/journal.pone.0209257>
- [9] Giovanni AG, Auguste R, Eric B. (2016). Improving users' product acceptability: an approach based on Bayesian networks and a simulated annealing algorithm. *International Journal of Production Research* 54(17): 5151-5168. <https://doi.org/10.1080/00207543.2016.1156183>
- [10] Yan DQ, Zhou Q, Wang JZ, Zhang N. (2016). Bayesian regularisation neural network based on artificial intelligence optimisation. *International Journal of Production Research* 55(8): 2266-2287. <https://doi.org/10.1080/00207543.2016.1237785>
- [11] Niloofar P. (2013). Performance evaluation of imputation based on Bayesian Network. *Sankhya B* 5(1): 90-111. <https://doi.org/10.1007/s13571-013-0061-8>
- [12] Zhu B, Xu ZH. (2018). Probability-hesitant fuzzy sets and the representation of preference relations. *Technological and Economic Development of Economy* 24(3): 1029-1040. <https://doi.org/10.4172/2169-0316.1000e115>
- [13] Wickens CD. (1992). *Engineering psychology and human performance*. New York: Harper Collins Publishers Inc 211-257.
- [14] Walid M, Wassim S, Kamel BO. (2017). Uncertainties handling in safety system performance assessment by using fuzzy Bayesian networks. *Journal of Intelligent & Fuzzy Systems* 33(2): 995-1006. <https://doi.org/10.3233/jifs-162296>
- [15] Li ZH, Wei FY. (2017). The logarithmic operational laws of intuitionistic fuzzy sets and intuitionistic fuzzy numbers. *Journal of Intelligent & Fuzzy Systems* 33(6): 3241-3253. <https://doi.org/10.3233/JIFS-161736>