Accident Analysis and Prevention 130 (2019) 160–166

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Modeling when and where a secondary accident occurs

Junhua Wang[a], Boya Liu[a], Ting Fu[b], Shuo Liu[a,*], Joshua Stipancic[b]

[a] School of Transportation Engineering, Tongji University, 4800 Cao'an Highway, Shanghai 201804, China
[b] Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, QC H3A 0C3, Canada

ARTICLE INFO

Keywords:
Secondary accident
Spatiotemporal Gap
BP neural network
LSSVM
Incident management

ABSTRACT

The occurrence of secondary accidents leads to traffic congestion and road safety issues. Secondary accident prevention has become a major consideration in traffic incident management. This paper investigates the location and time of a potential secondary accident after the occurrence of an initial traffic accident. With accident data and traffic loop data collected over three years from California interstate freeways, a shock wave-based method was introduced to identify secondary accidents. A linear regression model and two machine learning algorithms, including a back-propagation neural network (BPNN) and a least squares support vector machine (LSSVM), were implemented to explore the distance and time gap between the initial and secondary accidents using inputs of crash severity, violation category, weather condition, tow away, road surface condition, lighting, parties involved, traffic volume, duration, and shock wave speed generated by the primary accident. From the results, the linear regression model was inadequate in describing the effect of most variables and its goodness-of-fit and accuracy in prediction was relatively poor. In the training programs, the BPNN and LSSVM demonstrated adequate goodness-of-fit, though the BPNN was superior with a higher CORR and lower MSE. The BPNN model also outperformed the LSSVM in time prediction, while both failed to provide adequate distance prediction. Therefore, the BPNN model could be used to forecast the time gap between initial and secondary accidents, which could be used by decision makers and incident management agencies to prevent or reduce secondary collisions.

1. Introduction

Freeway incidents not only cause severe traffic congestion and travel delays, but can also result in secondary accidents, the risk of which has been estimated to be six times greater than that of a primary accident (Tedesco et al., 1994). Therefore, more attention should be given to secondary accidents on freeways. Most studies to date have focused on models to identify and predict secondary accidents (Green et al., 2012; Hirunyanitiwattana and Mattingly, 2006; Imprialou et al., 2013; Karlaftis et al., 1999; Khattak et al., 2009; Moore et al., 2004; Raub, 1997; Vlahogianni et al., 2010; Vlahogianni et al., 2012; Wang et al., 2016a; Yang et al., 2013a,b; Yang et al.,2014a,b,c; Zhan et al., 2008; Zhan et al., 2009; Zhang and Khattak, 2011). However, very few of them have considered the spatiotemporal distribution of secondary accidents. Sun & Chilukuri (2010), Chung (2013), and Chimba & Kutela (2014) suggested methods to visualize the spatiotemporal distribution of secondary collisions. Chung (2013) found that the average distance between the primary and secondary accidents was 1.34 miles, while the average time gap was 65.81 min. Wang et al. (2016b) used a mixture of normal and Weibull distributions to express the spatiotemporal gaps between accidents. Mean gap time was 74 min, while distance was 4.518 miles, with standard deviations of 53 min and 5.34 miles respectively (Wang et al., 2016b). However, gaps of less than 1 mile and less than 10 min were observed in 19.4% and 26.5% of cases respectively. In contrast to these studies, Chimba (2014) concluded that a large portion of secondary accidents occurred within 20 min and 0.5 miles upstream from the occurrence of the primary incidents. Zhang and Khattak (2010) also showed that the average gap time was 20 min. Park and Haghani (2016) studied the occurrence and clearance of primary and secondary incidents in order to adjust stochastic capacity, demonstrating that microsimulation could accurately model these accident scenarios when calibrated on sufficient data.

Among the limited research modeling spatiotemporal gap between consecutive accidents, Chimba and Kutela (2014) used a linear regression model and determined that the number of vehicles in the primary accident, incident duration, presence of construction zones, AADT, incident detection by law enforcement agencies, and bad weather conditions increased the distance gap. However, incidents with

* Corresponding author.
E-mail addresses: benwjh@163.com (J. Wang), hatakeya@sohu.com (B. Liu), ting.fu@mail.mcgill.ca (T. Fu), avensonliu@163.com (S. Liu), joshua.stipancic@mail.mcgill.ca (J. Stipancic).

https://doi.org/10.1016/j.aap.2018.01.024
Received 21 October 2016; Received in revised form 24 November 2017; Accepted 17 January 2018
Available online 02 February 2018
0001-4575/

long queues, wide shoulders, large proportion of trucks, and high number lanes decreased distance gaps. Despite high goodness-of-fit and significance of most parameters, one shortcoming of the linear regression model is the possibility of predicting infinitely large gaps. Therefore, a better model is required. Additionally, the time gap and distance of the secondary accident should be investigated simultaneously as they are likely to be correlated. This paper aims to address these gaps to better understand the possible time and location of secondary accidents.

Machine learning techniques have become promising for data analysis and output prediction. Machine learning is defined as "an algorithm that can learn from data without relying on rules-based programing" (Akerkar and Sajja, 2016). Together with data modeling, machine learning can be categorized according to two different cultures within statistical modeling, namely the algorithmic modeling culture and the data modeling culture (Breiman, 2001). Though both focus on data analysis, machine learning and traditional data modeling have many differences. Machine learning is a subfield of computer science and artificial intelligence, which learns from data without relying on programmed instructions, while data modeling is a subfield of mathematics, which explores relationship between observable variables and outcomes based on fixed rules (Akerkar and Sajja, 2016). Rather than requiring a modeler to understand the relationship between variables and results and to make specific assumptions, machine learning relies on the actions of computers that are not explicitly programmed (Sinha, 2015). Data analysis methods using machine learning techniques automate analytical model building, and are therefore capable of investigating larger and more complex data sets while providing more accurate results (Dhage and Raina, 2016) compared to traditional statistical modeling methods. Besides, targeted prediction purposes (Kohavi and Provost, 1998), machine learning may outperform traditional data modeling which focus on the "formalization of relationships between variables" (Johnson, 2013), in predicting secondary accidents. In other words, machine learning algorithms are potential solutions to the limitations of traditional statistical models for predicting secondary accidents, though this has yet to be explored.

This paper explored the application of machine learning algorithms for investigating occurrence of secondary accidents in time and space. Two machine learning algorithms, including the back-propagation neural network (BPNN) and the least squares support vector machine (LSSVM), were established to predict the spatiotemporal distribution of secondary accidents. BPNN and LSSVM models not only relieve the constraint of linear relationships between the variables and outcomes, but also allow for the simultaneous evaluation of the time and space gaps.

2. Secondary accident identification

Accident data was collected using the California Statewide Integrated Traffic Record System (SWITRS) from California interstate freeways over three years from January 2010 to December 2012. According to data provided by the Caltrans Performance Measurement System (PeMS), a total of 49,753 accidents occurred across the eight Caltrans districts and were used for analysis. Accident data and related traffic data were combined for each accident observation using time and location indicators. The PeMS loop data used in this paper included traffic volume, average occupancy, and average speed. The data were recorded every five minutes from loops upstream of the recorded accident location.

This study used the shock wave boundary filtering (SWBF) method (Wang et al., 2016a) to identify secondary accidents, illustrated in Fig. 1. The abscissa refers to the time after the occurrence of the initial accident, the ordinate is the distance upstream from the accident location, and the slope represents the speed of the shock wave. The impact of an initial accident is comprised of three stages where three unique shock waves are formed. In stage one, the initial accident creates a bottleneck at the accident location reducing speed and increasing density, thereby generating the first shock wave propagating upstream. When rescue personnel, police, or road agencies intervene, the second stage occurs as the traffic condition worsens. An increase in the slope of the shock wave represents the start of the second stage. The third stage begins when the bottleneck is removed and congestion begins to dissipate, creating a third shock wave. Thus, of the three shock waves generated, two are congestion propagating shock waves and one is a traffic relief shock wave. The distance affected by the initial accident is determined by the propagation of the traffic shock wave. The total influence of an initial accident, the gray area, can also be calculated using the shock waves. As in the figure, the entire period from when the primary accident occurs until the final shockwave ends is defined as the duration of the accident. Although this figure shows one secondary accident, the black point, an initial accident may create multiple secondary collisions. The data set for this study contained 204 primary accidents and 209 secondary accidents.

3. Linear regression based secondary accident spatiotemporal occurrence model

3.1. Variables

Detailed accident information available from the SWITRS is listed in Table 1. Upstream traffic volume for the 5 min preceding an accident was collected from the loop system, and shock wave speed was calculated using the SWBF procedure. 21 variables were incorporated in the models, including crash severity (fatal, severe), violation category (alcohol or drug, unsafe speed, following too closely, improper turning, improper driving), weather (clear and cloudy, raining), tow away, road surface (dry, wet), lighting (daylight, dusk, dark-streetlights), parties involved, volume, shock wave speed (1, 2, 3), and duration generated by the primary accident. 180 primary accidents were used for modelling and 24 primary accidents were used for testing.

3.2. Modelling

The time gap and distance of the accident pairs was first modelled using linear regression, results of which are reported in Table 2. Most of the parameters fail to explain the time gap and distance gap between the primary and secondary accidents. The correlation analysis in Table 2 shows that most independent variables are not significant, indicating that linear models are likely unsuitable for spatiotemporal gap prediction. Nonetheless, some parameters still correlate with spatiotemporal gaps. Primary accidents caused by alcohol or drug use, following too closely, and presence of rain may lead to a secondary accident in a short period of time. Another significant parameter is daylight, which increases the spatiotemporal gap between the primary accident and the secondary accidents. Higher speed of the first shock wave and higher traffic volumes tend to cause longer space gaps.

Table 3 shows that the correlation coefficient (CORR) as a measure of goodness-of-fit and mean square error (MSE) as a measure for residual examination are 0.433 and 0.653 h^2 for modeling time gap, and 0.491 and 22.197 $mile^2$ for modeling distance gap. This demonstrates that the linear model is not adequate for modelling the spatiotemporal gaps between primary and secondary accidents. In testing the model to predict secondary accidents, the linear regression model resulted in a CORR of 0.356 and a MSE of 0.439 h^2 for time gap prediction, and a CORR of 0.436 and a MSE of 20.662 $mile^2$ for distance prediction Prediction using linear regression is inadequate.

4. BPNN and LSSVM based secondary accident spatiotemporal occurrence model

As linear regression performed poorly in modeling the spatiotemporal gaps between primary and secondary accidents, machine

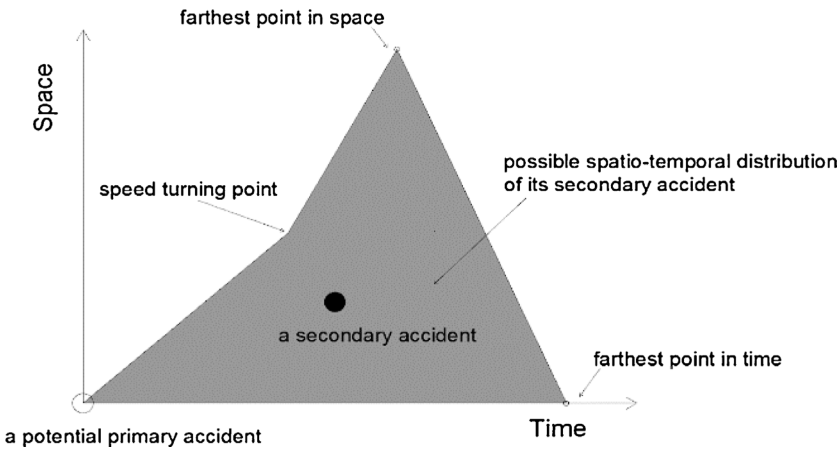

Fig. 1. Spatiotemporal secondary accident matching.

learning approaches may provide a better model fit and more accurate predictions. Two machine learning algorithms, the BPNN model and the LSSVM model, were therefore introduced.

4.1. Introduction of the models

4.1.1. Introduction of BPNN model

The BPNN is a feedforward network that can be trained using the back propagation learning algorithm, one of the best known and most widely applied methods for training neural networks. The technique stores ample relations between input and output and need not describe the relationship before the end of learning. The learning rule is the gradient descent method. The sum of square errors is minimized through back propagation by constantly adjusting network weights and threshold. BPNN is more adaptable and precise compared to more traditional models, is therefore widely used, and has achieved satisfactory results in recognition, fitting function relations, intelligent control, and other fields. Fig. 2 shows a typical structure of three-layer BPNN.

BPNN was chosen for this study for two main reasons. First, the time and distance gap between the primary accident and secondary accident should be analyzed simultaneously as they occur simultaneously and are likely correlated. BPNN can predict time gap and distance gap simultaneously as multiple outputs. Second, BPNN models can better describe nonlinear relationships. As described herein, distance gap follows a mixture distribution instead of normal distribution, and the correlation analysis in Table 2 shows that most of the independent variables are not significant in the linear regression model.

4.1.2. Introduction of LSSVM model

A Support Vector Machine (SVM) is a supervised machine learning method widely applied for nonlinear problems including pattern

Table 1
Specification of variables.

Field	Variable Type	Variable Explanation (percentage in secondary accidents)			Correlation test			
					Time gap		Distance gap	
					Correlation	Sig	Correlation	Sig
Crash severity	Discrete	Fatal (4%)	–	–	−0.098	0.162	−0.059	0.400
		Severe (3.9%)	–	–	−0.070	0.319	0.020	0.771
		Other (92.1%)	–	–	–	–	–	–
Violation category	Discrete	Unsafe Speed (65.3%)	–	–	0.003	0.965	0.096	0.171
		Alcohol or Drug (4.6%)	–	–	−0.247	0.000	−0.220	0.002
		Following Too Closely (1.4%)	–	–	−0.154	0.028	−0.020	0.781
		Improper Turning (13.7%)	–	–	0.089	0.204	0.051	0.470
		Improper Driving (13.6%)	–	–	0.112	0.111	−0.014	0.841
		Other (1.4%)	–	–	–	–	–	–
Weather	Discrete	Raining (9.2%)	–	–	−0.127	0.071	−0.037	0.595
		Clear and Cloudy (90.5%)	–	–	0.127	0.071	−0.011	0.871
		Other (0.3%)	–	–	–	–	–	–
Tow away	Discrete	Tow yes (73.7%)	–	–	0.007	0.924	0.058	0.414
		Other (26.3%)	–	–	–	–	–	–
Road surface	Discrete	Dry (83.1%)	–	–	0.079	0.261	0.025	0.727
		Wet (16.7%)	–	–	−0.085	0.226	−0.044	0.533
		Other (0.2%)	–	–	–	–	–	–
Lighting	Discrete	Daylight (78.4%)	–	–	0.226	0.001	0.218	0.002
		Dusk (2.4%)	–	–	−0.031	0.655	0.039	0.580
		Dark-streetlights (12.1%)	–	–	−0.154	0.028	−0.166	0.018
		Other (7.1%)	–	–	–	–	–	–
–	–	–	Mean	Std. Dev	–	–	–	–
Parties involved	Discrete	Number of parties involved in the primary accident.	2.31	0.95	0.073	0.298	0.095	0.175
Volume (Veh/5min)	Continuous	5 mins traffic volume before the secondary accident.	125.16	153.89	0.132	0.060	0.180	0.010
Shock wave1 (km/h)	Continuous	Shock wave generated by the primary accident.	36.18	61.05	−0.024	0.728	0.256	0.000
Shock wave2 (km/h)	Continuous	When rescue personnel, police or road agencies intervene.	135.91	156.29	−0.034	0.634	0.094	0.179
Shock wave3 (km/h)	Continuous	Dissipation shock wave after accident has been transacted.	273.30	175.27	−0.046	0.514	0.152	0.030
Duration (h)	Continuous	Duration of the accident.	1.45	1,22	0.100	0.181	0.045	0.546

Table 2
Estimation results - linear regression model.

Model	Time gap					Distance gap				
	Non-standard coefficient		Standard coefficient	t	Sig.	Non-standard coefficient		Standard coefficient	t	Sig.
	B	Standard deviation				B	Standard deviation			
Constant	64.694	8.284	–	7.809	.000	.572	.875	–	.654	.514
Alcohol or Drug	−42.429	16.799	−.189	−2.526	.012	–	–	–	–	–
Following Too Closely	−75.237	36.204	−.147	−2.078	.039	–	–	–	–	–
Raining	−26.306	11.250	−.166	−2.338	.020	–	–	–	–	–
Daylight	23.043	9.082	.189	2.537	.012	2.141	.897	.175	2.389	.018
Volume	–	–	–	–	–	.005	.002	.176	2.404	.017
Wave1	–	–	–	–	–	.021	.005	.287	4.126	.000

Table 3
Comparison of the models.

Modeling				Testing in prediction			
Correlation coefficient (CORR)		Mean square error (MSE)		Correlation coefficient (CORR)		Mean square error (MSE)	
Time	Distance	Time (h^2)	Distance ($mile^2$)	Time	Distance	Time (h^2)	Distance($mile^2$)
0.433	0.491	0.653	22.197	0.356	0.258	0.439	20.662

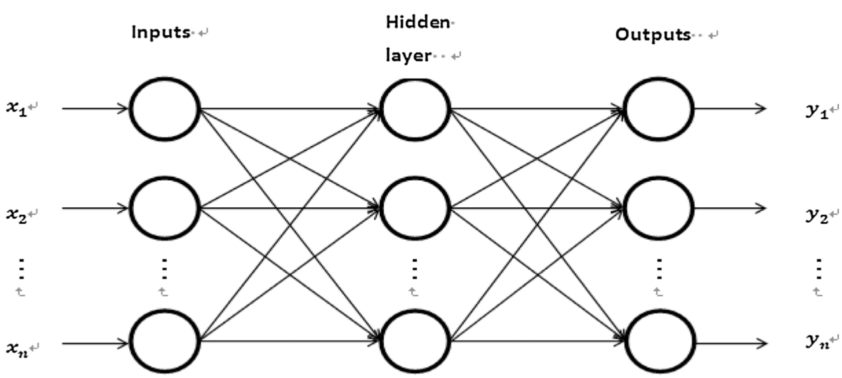

Fig. 2. A typical structure of three-layer BPNN.

Table 4
Comparison of BPNN architectures (21-*N*-2).

Hidden nodes (N)	Training			Testing in prediction		
	Correlation coefficient (CORR)	Mean square error (MSE)		Correlation coefficient (CORR)	Mean square error (MSE)	
		Time (h^2)	Distance ($mile^2$)		Time (h^2)	Distance ($mile^2$)
10	0.944	0.110	3.461	0.345	0.472	26.026
12	0.976	0.041	1.808	0.417	0.465	22.556
14	0.985	0.015	1.653	0.218	0.374	27.823
16	0.985	0.016	1.643	0.403	0.550	22.623
18	0.984	0.020	1.431	0.580	0.482	20.057
20	0.985	0.010	1.871	0.434	0.566	24.521
22	0.985	0.018	1.458	0.370	0.482	23.168
24	0.984	0.022	1.343	0.385	0.463	23.436

recognition, classification, and regression analysis. The LSSVM proposed by Suykens and Vandewalle (1999) is a reformulation of the standard SVM. This paper uses LSSVM for prediction of spatiotemporal gap of secondary accidents as it addresses some limitations in Artificial Neural Network (ANN) models. ANN models do not describe the relationship and importance of parameters, the model structure is difficult to understand and describe, and may be prone to over-fitting. Slow speed of convergence, arriving at local minimum, and less generalizable performance are additional shortcomings (Samui, 2011). SVM is usually less vulnerable to over fitting and suitable for small-sample sizes. With a quadratic loss function instead of the quadratic program, LSSVM transforms inequality constraints in traditional SVMs to equality constraints (Suykens et al., 2002). Moreover, it takes square error and loss function as loss experience of the training set, reducing computational resources and decreasing convergence time.

4.2. Model training and testing

4.2.1. BPNN based model training

The BPNN used the same variables as the linear regression model, with 180 primary accidents used for modelling and 24 used for testing. The topological structure of the BPNN included one input layer, one

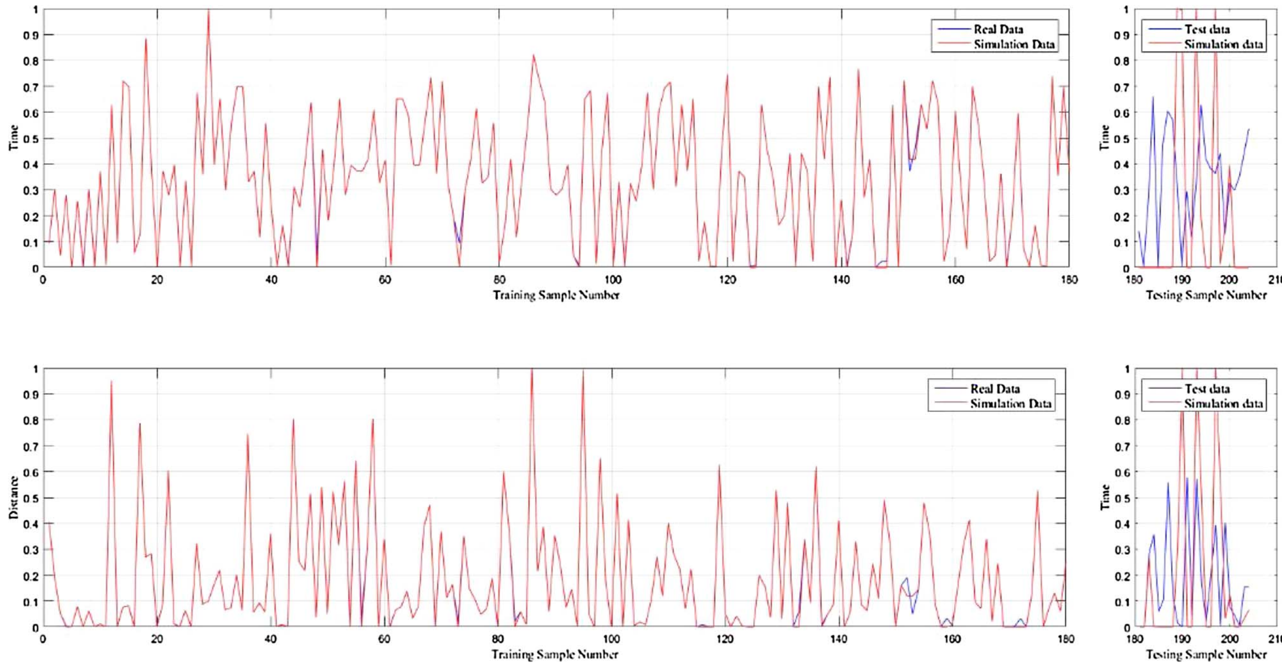

Fig. 3. BPNN simulation results.

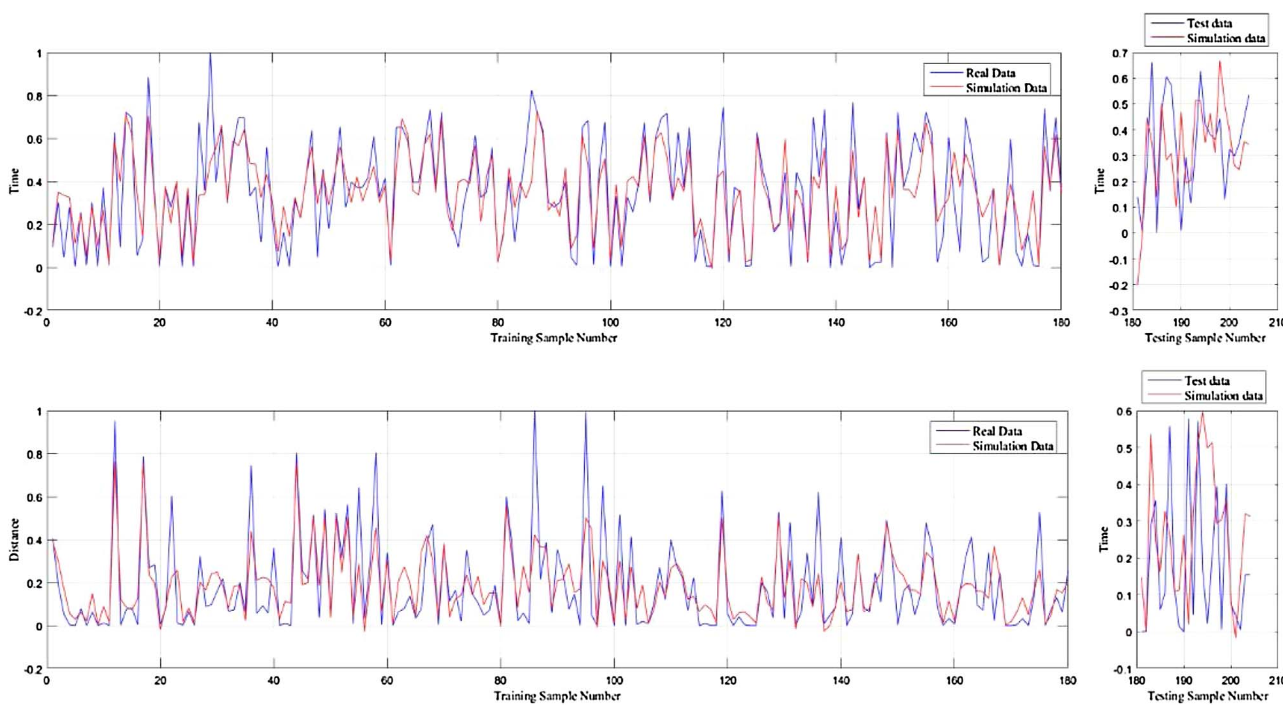

Fig. 4. LSSVM simulation results.

Table 5
Comparison of the models.

Models	Training				Testing in prediction			
	Correlation coefficient (CORR)		Mean square error (MSE)		Correlation coefficient (CORR)		Mean square error (MSE)	
	Time	Distance	Time (h^2)	Distance ($mile^2$)	Time	Distance	Time (h^2)	Distance ($mile^2$)
BPNN	0.984		0.020	1.431	0.580		0.482	20.057
LSSVM	0.861	0.812	0.057	3.573	0.401	0.236	7.144	0.004

output layer, and one hidden layer. Results for a single hidden layer with various numbers of nodes (from 10 to 24) are compared for training and testing samples in Table 4. The performance of BPNN models is validated based on MSE and CORR. Among all tested networks, the 21-18-2 structure (21-input neurons, 18-neurons in the hidden layer and 2-output neurons) resulted in the lowest MSE and good correlation between the targeted and predicted output for training and testing data sets. For training, the values of time MSE, distance MSE, and CORR are 0.020, 1.431 h^2 and 0.984 $mile^2$ respectively, demonstrating high reliability and precision. For testing, the values of time MSE, distance MSE and CORR are 0.482 h^2, 20.057 $mile^2$ and 0.58 respectively. The output of the time gap model fits the observed data well, but distance gap for testing does not. The visual comparison between simulated and real data is presented in Fig. 3. Such visualization proves the goodness-of-fit, as the polylines of the trained and tested data mainly overlaps well.

4.2.2. LSSVM based model training

Inputs, training, and testing data sets were identical for the LSSVM. LSSVM performance depends on parameter selection as each specific model is uniquely determined by its parameters. An RBF kernel function was used in this study, and the kernel function parameter (gamma) and penalty factor (C) were the key parameters. A particle swarm optimization (PSO) algorithm was used to determine the optimal parameter values. A comparison of simulated and real data is presented in Fig. 4. The MSE of time gap for training and testing are 0.057 h^2 and 7.144 h^2 respectively, and the associated CORR is 0.861 and 0.401. The MSE of distance gap for training and testing are 3.573 $mile^2$ and 0.004 $mile^2$ respectively, and the associated CORR is 0.812 and 0.236.

5. Comparison between models

Differing from traditional data modeling methods which rely on fixed rules including assumptions and predefined mathematical models, machine learning trains computer to “think and find” its own logic from the variable inputs to the outputs. Having different a focus related to data analysis and being unique in the way they are conducted, calibrated, and validated and the applied parameters (Breiman, 2001), makes it difficult to compare machine learning to more traditional statistical models. Despite this, as both are applied in this paper, their performance in modeling and prediction can be validated and compared.

Training and prediction performance results of the machine learning models are presented in Table 5. Results for the performance validation included in Table 3 and Table 5 are intuitive. Machine learning approaches with much higher MSE and better goodness-of-fit outperformed the linear regression approach in uncovering the relationship between the variable inputs and outputs of a secondary accident. Meanwhile, machine learning with its sole target on prediction, resulted in better prediction accuracy.

Comparison can also be made between the two machine learning models applied. The sample size for these models was relatively small (180 training accidents and 24 testing accidents). LSSVM is suitable for small-sample sizes, while there are no clear limitations for BPNN sample size. According to the model form and results, the sample data were adequate for training and testing. The comparison of BPNN and LSSVM models is shown in Table 5. Both methods have good fit for the training data set. The BPNN model has the higher CORR value of 0.984 and the lower MSE values for time and distance gaps compared to LSSVM model. For testing, LSSVM has the lower distance MSE value of 0.004 $mile^2$ for distance gap, while BPNN model has the lower time MSE value of 0.482 h^2. Some of the individual test samples had great differences in the LSSVM according to Fig. 4. Therefore, the BPNN model is determined to outperform the LSSVM in time gap prediction, while both methods perform poorly for distance prediction.

6. Conclusions

This research aimed to investigate the spatiotemporal gaps between primary and secondary accidents on freeways. The SWBF method was applied to identify secondary accidents. Linear regression was conducted to model the time and distance gaps of the accident pairs. The results show the linear model is not well suited for spatiotemporal gap prediction. Primary accidents caused by alcohol or drug use, following too closely, and rain may lead to a secondary accident in a short time. During day time, the spatiotemporal gap between the primary accident and the secondary accident is longer. Higher initial shock wave speed and higher traffic volume tend to create longer distance gaps.

To improve prediction, machine learning algorithms that offer the possibility of better prediction for non-linear problems, including BPNN and LSSVM models, were developed to investigate when and where a secondary accident occurs. The validation work for training shows that both BPNN and LSSVM have high goodness-of-fit. The time MSE, distance MSE and CORR values for the BPNN model are 0.020 h^2, 1.431 $mile^2$ and 0.984, respectively, indicating higher reliability and precision for training the spatiotemporal gap between accident pairs compared to the LSSVM model. For testing, the BPNN model had the highest CORR value of 0.580 and a lower time MSE value of 0.482 h^2. Therefore, the BPNN is better for time prediction. Although the LSSVM distance testing model had the lowest MSE value of 0.004 $mile^2$, some of the individual test samples had large differences. Therefore, the BPNN model outperforms the LSSVM in time prediction, while both methods perform poorly for distance prediction.

7. Discussion and future work

This paper provides three models for predicting the location and time of a secondary accident. The main contribution of this paper is to introduce machine learning approaches to effectively predict secondary accidents on freeways, a topic which has drawn much attention from both researchers and practitioners, though has been studied infrequently due to the limitations of traditional methods. The study and its results can provide decision makers and incident management agencies reliable information for real-time incident monitoring and secondary accident prevention. Using these models, secondary accidents can be targeted and prevented in certain time and space. Like most loop data systems, the health rate of loops in PeMS is about 55%. There are likely more secondary accidents than those identified in this study. Loop health also affects the shock wave calculations. Therefore the sample size for modelling is limited to 180 observations. The models can be further improved if other advanced detectors, such as radar, are used in the future.

Acknowledgement

This research is supported by the National Natural Science Foundation (71301119) and the Fundamental Research Funds for the Central Universities.

References

Akerkar, R., Sajja, P.S., 2016. Intelligent Techniques for Data Science.
Breiman, L., 2001. Statistical modeling: the two cultures. Stat. Sci. 16 (3), 199–231.
Chimba, D., Kutela, B., 2014. Scanning secondary derived crashes from disabled and abandoned vehicle incidents on uninterrupted flow highways. J. Saf. Res. 50, 109–116.
Chung, Y., 2013. Identifying primary and secondary crashes from spatiotemporal crash impact analysis. Transp. Res. Rec. J. Transp. Res. Board 2386, 62–71.
Dhage, S., Raina, C.K., 2016. A review on machine learning techniques. Int. J. Recent Innovation Trends Comput. Commun. 4 (3), 395–399 March.
Green, E.R., Pigman, J.G., Walton, J.R., McCormack, S., 2012. Identification of secondary crashes and recommended countermeasures to ensure more accurate documentation. Transportation Research Board 91st Annual Meeting. Compendium of Papers CD-ROM Washington, D.C.
Hirunyanitiwattana, W., Mattingly, S.P., 2006. Identifying secondary crash characteristics

for California highway system. Transportation Research Board 85th Annual Meeting.
Imprialou, M.I.M., Orfanou, F.P., Vlahogianni, E.I., Karlaftis, M.G., 2013. Methods for defining spatiotemporal influence areas and secondary incident detection in freeways. J. Transp. Eng. 140, 70–80.
Johnson, R.A., 2013. Stat. Princ. Methods.
Karlaftis, M.G., Latoski, S.P., Richards, N.J., Sinha, K.C., 1999. ITS impacts on safety and traffic management: an investigation of secondary crash causes. J. Intell. Transp. Syst. 5, 39–52.
Khattak, A., Wang, X., Zhang, H., 2009. Are incident durations and secondary incidents interdependent? Transportation research record. J. Transp. Res. Board 2099, 39–49.
Kohavi, R., Provost, F., 1998. Glossary of terms. Mach. Learn. 30, 271–274.
Moore, J.E., Giuliano, G., Cho, S., 2004. Secondary accident rates on Los Angeles freeways. J. Transp. Eng. 130, 280–285.
Park, H., Haghani, A., 2016. Stochastic capacity adjustment considering secondary incidents. IEEE Trans. Intell. Transp. Syst.
Raub, R.A., 1997. Occurrence of secondary crashes on urban arterial roadways. Transp. Res. Rec. J. Transp. Res. Board 1581, 53–58.
Samui, P., 2011. Application of least square support vector machine (LSSVM) for determination of evaporation losses in reservoirs. Engineering 3, 431–434.
Sinha, H., 2015. Machine Learning vs. Statistical Modeling: The Real Difference. Inforgix, Inc.. https://www.infogix.com/blog/machine-learning-vs-statistical-modeling-the-real-difference/.
Sun, C.C., Chilukuri, V., 2010. Dynamic incident progression curve for classifying secondary traffic crashes. J. Transp. Eng. 136, 1153–1158.
Suykens, J.A.K., Gestel, T.V., Brabanter, J.D., Moor, B.D., Vandewalle, J., 2002. Least Squares Support Vector Machine. World Scientific, Singapore.
Suykens, J.A.K., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Process. Lett. 9, 293–300.
Tedesco, S.A., Alexiadis, V., Loudon, W.R., Margiotta, R., Skinner, D., 1994. Development of a 40 model to assess the safety impacts of implementing IVHS user services, moving toward deployment. Proceedings of the IVHS America Annual Meeting 343–352 Atlanta Georgia.
Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., Halkias, B.M., 2010. Freeway operations, spatiotemporal-incident characteristics, and secondary-crash occurrence. Transp. Res. Rec. 1–9.
Vlahogianni, E.I., Karlaftis, M.G., Orfanou, F.P., 2012. Modeling the effects of weather and traffic on the risk of secondary incidents. J. Intell. Transp. Syst. 16, 109–117.
Wang, J., Xie, W., Liu, B., Fang, Se, Ragland, D.R., 2016a. Identification of freeway secondary accidents with traffic shock wave detected by loop detectors. Saf. Sci. 87, 195–201.
Wang, J.H., Liu, B.Y., Zhang, L.F., Ragland, D.R., 2016b. Modeling secondary accidents identified by traffic shock waves. Accid. Anal. Prev. 87, 141–147.
Yang, H., Bartin, B., Ozbay, K., 2013a. Identifying secondary crashes on freeways using sensor data. Transp. Res. Record: J. Transp. Res. Board.
Yang, H., Bartin, B., Ozbay, K., 2013b. Investigating the characteristics of secondary crashes on freeways. Transportation Research Board 92nd Annual Meeting. Compendium of Papers CD-ROM Washington, D.C.
Yang, H., Bartin, B., Ozbay, K., 2014a. Mining the characteristics of secondary crashes on highways. J. Transp. Eng. 140, 04013024.
Yang, H., Morgul, E.F., Bartin, B., Ozbay, K., 2014b. Development of an on-line scalable approach for identifying secondary crashes. Transp. Res. Record J. Transp. Res. Board 14–1689.
Yang, H., Ozbay, K., Xie, K., 2014c. Assessing the risk of secondary crashes on highways. J. Saf. Res. 49, 143–149.
Zhan, C., Shen, L., Hadi, M.A., Gan, A., 2008. Understanding the characteristics of secondary crashes on freeways. Transportation Research Board 87th Annual Meeting.
Zhan, C.J., Gan, A., Hadi, M., 2009. Identifying secondary crashes and their contributing factors. Transp. Res. Rec. 68–75.
Zhang, H., Khattak, A., 2010. What is the role of multiple secondary incidents in traffic operations? J. Transp. Eng. 136, 986–997.
Zhang, H.B., Khattak, A., 2011. Spatiotemporal patterns of primary and secondary incidents on urban freeways. Transp. Res. Rec. J. Transp. Res. Board 19–27.