Research Article

Predicting Severity and Duration of Road Traffic Accident

Fang Zong,[1] Huiyong Zhang,[1,2] Hongguo Xu,[1] Xiumei Zhu,[2] and Lu Wang[3]

[1] *College of Transportation, Jilin University, 5988 Renmin Street, Changchun, Jilin 130022, China*
[2] *School of Management, Jilin University, 5988 Renmin Street, Changchun, Jilin 130022, China*
[3] *China Academy of Civil Aviation Science and Technology, Beijing 100028, China*

Correspondence should be addressed to Huiyong Zhang; jlu.zf@163.com

This paper presents a model system to predict severity and duration of traffic accidents by employing Ordered Probit model and Hazard model, respectively. The models are estimated using traffic accident data collected in Jilin province, China, in 2010. With the developed models, three severity indicators, namely, number of fatalities, number of injuries, and property damage, as well as accident duration, are predicted, and the important influences of related variables are identified. The results indicate that the goodness-of-fit of Ordered Probit model is higher than that of SVC model in severity modeling. In addition, accident severity is proven to be an important determinant of duration; that is, more fatalities and injuries in the accident lead to longer duration. Study results can be applied to predictions of accident severity and duration, which are two essential steps in accident management process. By recognizing those key influences, this study also provides suggestive results for government to take effective measures to reduce accident impacts and improve traffic safety.

1. Introduction

Traffic accidents are a significant source of deaths, injuries, property damage, and a major concern for public health and traffic safety. Accidents are also a major cause of traffic congestion and delay. Effective management of accident is crucial to mitigating accident impacts and improving traffic safety and transportation system efficiency. As two major steps of the accident response program (shown in Figure 1), severity prediction and duration estimation are, therefore, of great importance. Accurate predictions of severity and duration can provide crucial information for emergency responders to evaluate the severity level of accidents, estimate the potential impacts, and implement efficient accident management procedures.

To the authors' knowledge, most of the previous studies examined accident severity and duration separately, although they were found to have correlation between each other. Moreover, only one or two of the three aspects of accident severity, that is, number of fatalities, number of injuries, and property damage, were investigated by existing researchers. Therefore, the present study is aimed at developing a model system to estimate both accident severity and duration. Furthermore, three indicators for accident severity will be set, which represents number of fatalities, number of injuries, and property damage, respectively. In doing so, we will provide crucial information for emergency responders to take effective management measures.

The remainder of this paper is organized as follows. In Section 2, we present the literature review on predictions of severity and duration in general. The data are described in Section 3. Following is accident severity modeling in Section 4 and duration forecasting in Section 5. The paper concludes with a summary and directions for future research.

2. Existing Literature

As two major factors in accident analysis, severity and duration have long been important topics for research. Most of the previous studies examined only one of severity and duration. For example, with respect to severity analysis, Chang and Mannering [1] studied the relationship between injury severity and vehicle occupancy using Washington State accident data. Mannera and Wünsch-Ziegler [2] investigated
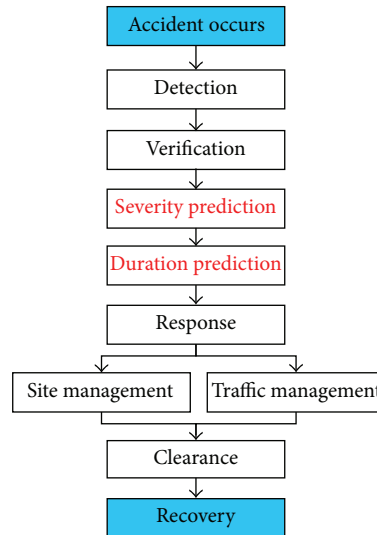
FIGURE 1: Accident response procedure.

accident severity and determined the important effects of related factors. As for duration, Chung [3] modeled accident duration with freeway accident data collected in Korea. Anastasopoulos et al. [4] presented a Bayesian network model that can be used to learn emerging patterns and predict accident clearance time. Nevertheless, accident severity was found to have influence on duration time by some researchers. For instance, Nam and Mannering [5] revealed that whether there is fatality or injury in accident impacts accident duration. Besides, as shown in Figure 1, severity prediction and duration estimation are connected procedures in the accident management system. Therefore, the two indicators should be considered together and combined in one model system.

Concerning severity analysis, which includes mainly three aspects, that is, number of fatalities, number of injuries, and property damage, most of the existing researchers investigated it as one comprehensive indicator; for example, Mannera and Wünsch-Ziegler [2]took accident severity as one independent variable with four alternatives, namely, fatal, severe injury, light injury, and property damage. Milton et al. [6] defined severity levels as property damage only, possible injury, and injury. Malyshkina and Mannering [7] modeled severity by using three alternatives, that is, fatality, injury, and property damage only. In addition, a number of researchers considered only one or two of the three aspects of severity. For instance, Stone and Broughton [8] and Sze and Wong [9] considered only the aspect of fatality by defining two levels of severity, that is, fatal and nonfatal accident. Delen et al. [10] defined injury severity levels as no injury, probable injury, nonincapacitating, incapacitating and fatality. Similarly, Ballesteros et al. [11] and Roudsari et al. [12] considered only number of fatalities and injuries but not property damage. In fact, different types of losses as well as the amount of losses lead to different response measures and last possibly for disparate amount of time. For example, either an accident resulting in $167–5000 property damage or an accident leading to 1–3 injuries will be defined as level 2

accident in Zhang's study [13]. However, the latter one needs rescue services but the former one does not. This indicates that any of the three indicators, that is, number of fatalities, number of injuries, and property damage, is crucial to making accident response decision and is therefore recommended to be modeled separately in order to provide more detailed information about accident management.

As mentioned above, most of the previous studies examined accident severity and duration separately, although they were found to have correlation between each other. Moreover, only one or two of the three aspects of accident severity, that is, number of fatalities, number of injuries, and property damage, were investigated by the existing studies. Therefore, the present work is aimed at developing a model system to estimate both accident severity and duration. Furthermore, three indicators for accident severity will be investigated, which represent number of fatalities, number of injuries, and property damage, respectively.

3. Data and Modeling Framework

The dataset for the study contains police-reported traffic accident records for Jilin province, China, in 2010. With records containing missing values eliminated, our final dataset consists of 3,914 cases, in which, 1,280 (32.70%) cases were pedestrian involved accidents and 387 (9.89%) cases were non-motor-vehicle-involved accidents. In addition to severity information, the data contains information regarding accident duration, accident characteristics (vehicle fire, crash type, accident occurrence time, and number of lanes affected), emergency services (police services, fire and rescue services, tow services, and emergency medical services), vehicle characteristics (vehicle type involved, debris involved, hazardous material involved, and disabled vehicles involved), environmental factors (weather conditions and visibility distance) and road conditions (number of lanes,
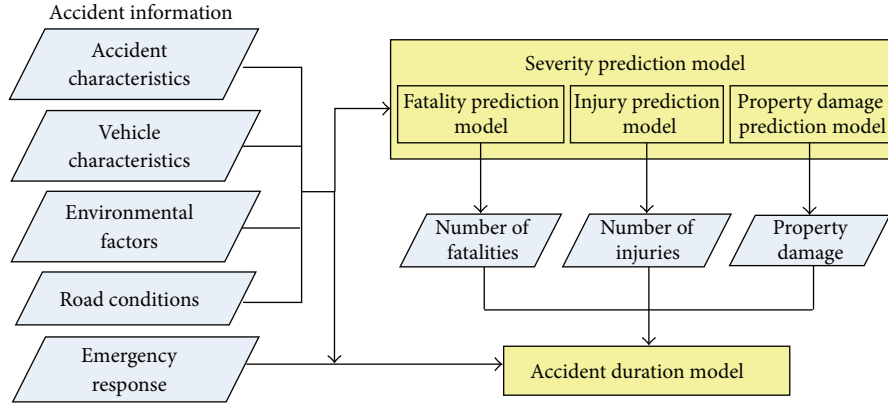
Figure 2: Accident severity and duration modeling framework.

pavement condition, road geometrics, and roadway surface condition, etc.).

Based on a preliminary correlation test, 4 independent variables and 26-candidate dependent variables were selected from the dataset, as shown in Table 1.

With *Nof*, *Noi*, and *Pd* as independent variables, three separate severity prediction models will be developed. Then, duration modeling will be conducted by taking accident severity as input. The modeling framework is shown in Figure 2.

4. Severity Modeling

Besides the Ordered Probit model [14], which is often used in discrete choice modeling, SVM will be introduced in this paper and be compared with the Ordered Probit model according to the prediction accuracies.

4.1. Ordered Probit Model. As shown in Table 1, the alternatives of the severity related dependent variables are all ordered. Since multinomial logit (MNL) model, which is commonly used in discrete choice modeling, would fail to account for the ordinal nature of the dependent variable and have the problem of Independence from irrelevant alternatives (IIA) [15], this study will employ Ordered multiple choice model for severity modeling.

The Ordered multiple choice model assumes the relationship:

$$\sum_{j=1}^{J} P_n(j) = F\left(\alpha_j - \beta_j X_n, \theta\right), \quad j = 1, \ldots, J-1,$$

$$P_n(J) = 1 - \sum_{j=1}^{J} P_n(j),$$

(1)

where $P_n(j)$ is the probability that alternative j happens in accident n ($n = 1, \ldots, N$), α_j is an alternative specific constant, X_n is a vector of the attributes of accident n, β_j is a vector of estimable coefficients, and θ is a parameter that controls the shape of probability distribution F. Therefore,

F, can have various shapes of distribution based on different value of θ.

The Ordered Probit model, which assumes standard normal distribution for F is the most commonly used Ordered multiple choice model [16]. The Ordered Probit model has the following form:

$$P_n(1) = \Phi\left(\alpha_1 - \beta_j X_n\right),$$

$$P_n(j) = \Phi\left(\alpha_j - \beta_j X_n\right)$$

$$- \Phi\left(\alpha_{j-1} - \beta_j X_n\right), \quad j = 2, \ldots, J-1,$$

(2)

$$P_n(J) = 1 - \sum_{j=1}^{J-1} P_n(j),$$

where $P_n(j)$ is the cumulative standard normal distribution function. For all the probabilities to be positive, we must have $\alpha_1 < \alpha_2 < \cdots < \alpha_{J-1}$.

4.2. Support Vector Machine Model. Support vector machine (SVM) is a type of learning algorithms based on statistical learning theory, which can be adjusted to map the input-output relationship for the nonlinear system [17–19]. SVM has been widely used in transportation modeling; for example, Bolbol et al. [20] employed SVM classification in travel behavior analysis, Apatean et al. [21] used it in road obstacle classification, and Abdel-Aty and Haleem [22] applied it to analyze angle crashes at unsignalized intersections. Previous studies indicate that SVM can conduct discrete choice modeling with acceptable accuracy. Therefore, it is chosen to be employed to model accident severity in this paper.

Given a set of input-output data pairs $D = (x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)$ ($x_i \in X \subseteq R^m$, $y_i \in Y \subseteq R^n$, and l being the number of training samples, that are randomly and independently generated from an unknown function, SVM estimates the function using the following equation [23]:

$$f(x) = w \cdot \Phi(x) + b \quad w, x \in R^m, \ b \in R^n,$$

(3)

TABLE 1: Variables and statistics based on survey data.

Factors	Variables	Values	Percentage (%)	Variables	Values	Percentage (%)
Accident severity	Number of fatalities: Nof	0	89.59	Number of injuries: Noi	0	9.86
		[1, 2]	10.38		[1, 3)	85.89
		More than 3	0.03		[3, 11)	4.14
	Property damage (yuan): Pd	Less than 1000	61.18		Over 11	0.11
		[1001, 30000)	37.19			
		Over 30000	1.63			
Duration	Duration (Continuous value)	Mean (min) 192.95	Standard deviation 111.63			
Accident characteristics	Motor-vehicle-only accident: Mvoa	Yes	57.41	Vehicle fire: Vf	Yes	8.93
		No	42.59		No	91.07
	Head-on type collision: Hotc	Yes	8.93	Weekend or festival: Wof	Yes	38.60
		No	91.07		No	61.40
	Rear-end type collision: Retc	Yes	19.64	Vehicle rollover: Vr	Yes	26.79
		No	80.36		No	73.21
	Time of day: Tod	[00:00, 6:00)	6.24	Number of lanes blocked: Nolb	0	3.57
		[6:00, 18:00)	69.12		1	62.50
		[18:00, 24:00)	24.64		over 1	33.93
Vehicle characteristics	Bus involved: Bi	Yes	16.07	Hazardous material involved: Hmi	Yes	1.79
		No	83.93		No	98.21
	Truck involved: Ti	Yes	89.29	Disabled vehicles involved: Dvi	Yes	27.27
		No	10.71		No	72.73
	Debris involved: Di	Yes	53.57			
		No	46.43			
Environmental factors	Weather conditions: Wc	Sunny	89.48	Visibility distance (meter): Vd	Less than 50	8.90
		Fog	0.23		[50, 100)	22.70
		Sleet	5.97		[100, 200)	19.86
		Other	4.32		Over 200	48.54
Road environment factors	Number of lanes in each direction: Nol	2	33.92	Accident location (horizontal): Alh	Motor vehicle lanes	71.68
		3	51.79		Bike lane	6.60
		Over 3	14.29		Shared motor vehicle and bike lane	13.71
	Pavement condition: Pc	Asphalt	96.95		Sidewalk	2.22
		Cement	2.85		Crosswalk	3.42
		Sand and gravel	0.07		Other	2.37
		Soil	0.07	Accident location (vertical): Alv	Regular road section	60.01
		Other	0.06		Four-way intersection	20.43
	Roadway surface condition: Rsc	Dry	85.16		Other road sections (narrow carriageway and tunnel, etc.)	1.09
		Wet	6.38		Other intersections	18.47
		Slippery (snowy or icy conditions)	6.76	Road geometrics: Rg	Flat and straight	98.57
		Other	1.70		Hill or bend	1.43
	Traffic signal control: Tsc	Yes	17.46			
		No	82.54			
Emergency services	Police services: Ps	Yes	71.43	Fire and rescue services: Frs	Yes	16.07
		No	28.57		No	83.93
	Tow services: Ts	Yes	98.21	Emergency medical services: Ems	Yes	33.93
		No	1.79		No	66.07

where $\Phi(x)$ represents the high-dimensional feature spaces which are nonlinearly mapped from the input space x, w denotes a parameter vector, and b is the threshold [24, 25].

If the domain of output space y only takes category values, that is, -1 and $+1$, the learning problem then refers to support vector classification (SVC) [26].

For classification about the training data D, SVM's linear soft-margin algorithm is used to solve the following primal quadratic programming problem:

$$
\begin{aligned}
\min_{w,b,\xi} \quad & \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{l}\xi_i \\
\text{s.t.} \quad & y_i\left(w^T x_i + b\right) \geq 1 - \xi_i, \\
& \xi_i \geq 0, \quad i = 1, 2, \ldots, l,
\end{aligned}
\tag{4}
$$

where C is a penalty parameter and ξ_i are the slack variables. The goal is to find an optimal separating hyperplane,

$$
w^T x_i + b = 0,
\tag{5}
$$

where $x \in R^m$. The Wolfe dual, that is, (4), can be expressed as

$$
\begin{aligned}
\max_{\alpha} \quad & \sum_{j=1}^{l}\alpha_j - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}y_i y_j\left(x_i \cdot x_j\right)\alpha_i\alpha_j \\
\text{s.t.} \quad & \sum_{i=1}^{l}y_i\alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \ i = 1, \ldots, l,
\end{aligned}
\tag{6}
$$

where $\alpha \in R^l$ are lagrangian multipliers. The optimal separating hyperplane of (5) can be given by

$$
w = \sum_{i=1}^{l}\alpha_i^* y_i x_i, \quad b = \frac{1}{N_{sv}}\left(y_j - \sum_{i=1}^{N_{sv}}\alpha_i^* y_i\left(x_i \cdot x_j\right)\right), \tag{7}
$$

where α^* is the solution of (6) and N_{sv} represents the number of support vectors such that $0 < \alpha < C$. A new sample is classified as $+1$ or -1 according to the finally decision function $f(x) = \text{sgn}((w \cdot x) + b)$.

In order to conduct multiclass classification (as SVC model is originally designed for binary classification), one-against-one method will be employed in this paper [27, 28].

4.3. Estimation Results. By using Stata and Matlab, the severity prediction models based on Ordered Probit and SVM are estimated, respectively. The estimation results as well as the prediction accuracies are shown in Table 2.

The last row shows the hit ratio for all the models. In general, higher value of hit ratio represents higher goodness-of-fit of the model. As all the hit ratio values of the Ordered Probit models are higher than that of SVM models, Ordered Probit-based models are chosen as the severity prediction models.

The results indicate that hazardous material involved in the accident, weather, and accident location are significant in all the three models. According to the estimation results,

hazardous material involved will increase the probability of high property damage. The reason is that hazardous material will increase the probability of occurrence of fire or even explosion, which leads to high damage to the vehicles and goods.

Some of the variables have impact on only one or two indicators. For example, bus involved, truck involved, time of day, and traffic signal control are crucial to number of fatalities and injuries. The more buses or trucks are involved, the more fatalities and injuries the accident will cause. In addition, the factors of road geometrics, vehicle fire, and vehicle rollover are important for number of fatalities, while roadway surface condition has effect on number of injuries. The results also indicate that disabled vehicles involved, debris involved, visibility distance, pavement condition, and motor-vehicle-only accident are significant for property damage. The more disabled vehicles or debris is involved in the accident, the more property damage the accident will lead to. As for motor-vehicle-only accident, the results reveal that accidents with only vehicles involved cause more property damage than that with pedestrian or non-motor-vehicles involved.

5. Accident Duration Modeling

5.1. AFT Model and KM Estimator. As suggested by Nam and Mannering [5] and Stathopoulos and Karlaftis [29], hazard-based duration models have an advantage in that they allow the explicit study of duration effects of accidents (i.e., the relationship between how long an accident has lasted and the likelihood of it ending soon). Thus, hazard-based duration models, in particular the accelerated failure time (AFT) metric, were utilized in this study to model the accident duration. The reason that we choose AFT model is that, compared with other forms of hazard-based model, AFT model is predominately fully parametric; that is, a probability distribution is specified and it is also less affected by the choice of probability distribution [30, 31], and the results of AFT model are easily interpreted [32].

Let T be a nonnegative random variable representing the accident duration. The hazard at time t on the continuous time-scale $h(t)$ is defined as the instantaneous probability that the duration under study will end in an infinitesimal time period Δt after time t, given that the duration has not elapsed until time t. A mathematical definition for the hazard function is as follows:

$$
h(t) = \lim_{\Delta \to 0^+}\frac{P(t \leq T < t + \Delta \mid T > t)}{\Delta}.
\tag{8}
$$

Let $f(\cdot)$ and $F(\cdot)$ be the density and cumulative distribution function for T, respectively. Then the probability of ending in an infinitesimal interval of range Δt, after time t is $f(t)\Delta t$. And the probability that the process lasts for at least t is given by the survival equation

$$
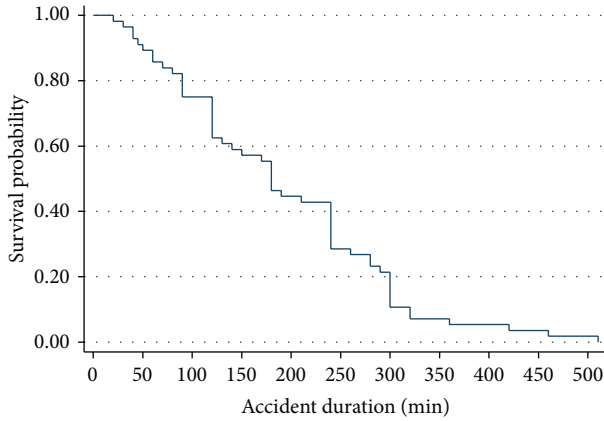S(t) = P(T > t) = 1 - F(t).
\tag{9}
$$

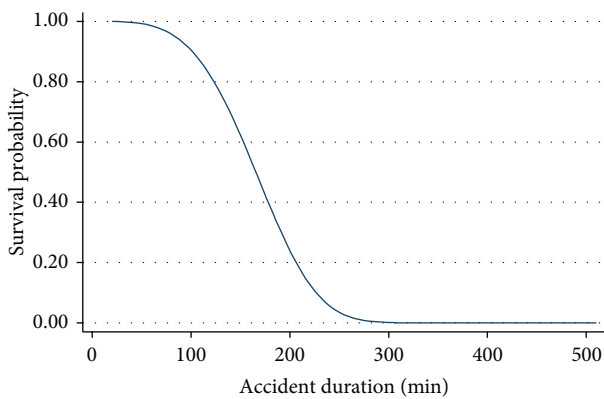FIGURE 3: Survival curve of accident duration.

FIGURE 4: The estimated survival curve of accident duration.

Thus, the hazard function can be further expressed as

$$h(t) = \frac{f(t)}{S(t)} = \frac{dF(t)/dt}{S(t)} = \frac{-dS(t)/dt}{S(t)} = \frac{-d\ln S(t)}{dt}. \quad (10)$$

The distribution of the hazard can be assumed to be one of many parametric forms or to be nonparametric. Because the distribution of the accident duration is unknown, one of the nonparametric methods, the Kaplan-Meier (KM) product limit estimator, is conducted to explore the covariates effects and the potential distribution.

As a nonparametric method, the KM estimator, produces an empirical approximation of survival and hazard but hardly takes any covariate effects into consideration. It is similar to an exploratory data analysis. Denoting the distinct failure times of individuals n as $t_1 < t_2 < \cdots < t_m$, the KM estimator of survival at time t_i is computed as the product of the conditional survival proportions:

$$S_{KM}(t_i) = \prod_{k=1}^{i} \frac{r(t_k) - d(t_k)}{r(t_k)}, \quad (11)$$

where $r(t_k)$ is the total number of accidents at risk for ending at t_k and $d(t_k)$ is the number of accidents stopping at t_k.

By using the KM estimator, the survival function curves of the accident duration are estimated, which are shown in

Figure 3. The results indicate that the survival probability decreases with duration, which implies an accelerated failure time model with Weibull or Exponential distribution should be employed. Therefore, the AFT model is developed to examine the linkages between duration and covariates relative to accident information.

The AFT model permits the covariates to affect the duration dependence. Its survival function is given as

$$S(t) = S_0\left[t \cdot \exp\left(-\beta' X\right)\right], \quad (12)$$

where $S_0(\cdot)$ is the baseline survival function. The corresponding hazard function is

$$h(t) = \frac{-\partial S(t)/\partial t}{S(t)} = h_0\left[t \cdot \exp\left(-\beta' X\right)\right] \exp\left(-\beta' X\right). \quad (13)$$

The AFT model can be expressed as a log-linear model:

$$\ln t = \beta' X + \varepsilon. \quad (14)$$

Assuming that the random error ε follows either a Weibull distribution or an Exponential distribution, one can get two kinds of AFT models, and both of them are often used in duration analysis.

5.2. Estimation Results. Assuming that the random error in (14) follows a Weibull distribution and an Exponential distribution, respectively, the accident duration models are established. The models are estimated by employing maximum likelihood estimation (MLE), and the estimation results are shown in Table 3.

The Mean absolute percentage error (MAPE), which looks at the average percentage difference between predicted values and observed ones, is adopted to examine the accuracy of the developed duration predication model. MAPE is calculated as

$$\text{MAPE} = \frac{1}{n}\sum_{i-1}^{n} \left|\frac{A_i - P_i}{A_i}\right|, \quad (15)$$

where A_i is the observed value and P_i is the predicted value for observation i.

The MAPE value of Weibull distribution (0.22) is less than that of the Exponential distribution (0.23), indicating that the values predicted by the AFT model with the Weibull distribution is more close to the actual accident duration [3]. Therefore, the Weibull distribution function is chosen.

The estimation results indicate that most of the results were consistent with the theoretical expectation. According to the results, the variables with respect to accident severity significantly affect accident duration: the more fatalities and injuries occur in the accident, the longer duration it will lead to. This supports the necessity of combining predictions of accident severity and duration in one model system. Besides, accident type is revealed to be crucial to duration: comparing with other types of accidents, the duration of rear-end type collision is 37% shorter, while that of rollover is 28% longer. The results also show that the duration of accident involving bus, truck, debris, or hazard material is 60%, 58%, 55%,

TABLE 2: Estimation results of severity prediction models.

| Variables | Fatality model | | | Injury model | | | Property damage model | | |
| | SVM | Ordered Probit | | SVM | Ordered Probit | | SVM | Ordered Probit | |
		Coef.	Z-stat.		Coef.	Z-stat.		Coef.	Z-stat.
Dvi	—	—	—	—	—	—	$\sqrt{}$	0.23	1.99
Bi	$\sqrt{}$	0.75	8.82	$\sqrt{}$	0.28	4.37	—	—	—
Ti	$\sqrt{}$	0.64	7.93	$\sqrt{}$	0.29	4.61	—	—	—
Di	—	—	—	—	—	—	$\sqrt{}$	0.11	2.00
Hmi	$\sqrt{}$	0.04	1.34	$\sqrt{}$	0.04	1.65	$\sqrt{}$	0.21	1.96
Tod	$\sqrt{}$	0.12	2.94	$\sqrt{}$	0.04	1.21	—	—	—
Wc	$\sqrt{}$	0.12	2.3	$\sqrt{}$	−0.05	−2.3	$\sqrt{}$	0.04	2.46
Vd	—	—	—	—	—	—	$\sqrt{}$	0.10	5.92
Tsc	$\sqrt{}$	0.03	2.31	$\sqrt{}$	−0.02	−1.56	—	—	—
Alh	$\sqrt{}$	−0.03	−1.44	$\sqrt{}$	0.04	2.71	$\sqrt{}$	−0.03	−2.12
Alv	$\sqrt{}$	−0.11	−6.88	$\sqrt{}$	0.06	5	$\sqrt{}$	−0.13	−12.16
Rsc	—	—	—	$\sqrt{}$	0.11	4.03	—	—	—
Rg	$\sqrt{}$	−0.26	−1.71	—	—	—	$\sqrt{}$	−0.18	−2.01
Pc	—	—	—	—	—	—	$\sqrt{}$	−0.18	−2.09
Vf	$\sqrt{}$	0.73	7.70	—	—	—	$\sqrt{}$	0.13	1.98
Vr	$\sqrt{}$	0.04	1.32	—	—	—	—	—	—
Mvoa	—	—	—	—	—	—	$\sqrt{}$	0.45	11.66
α_1	—	0.63	—	—	−0.68	—	—	0.06	—
α_2	—	2.79	—	—	2.36	—	—	1.99	—
α_3	—	—	—	—	3.71	—	—	—	—
Hit ratio (%)	89.21	89.59		86.50	86.89		59.57	62.66	

TABLE 3: Estimation results of accident duration model.

| Variables | Weibull distribution | | Exponential distribution | |
	Coef.	z-stat.	Coef.	z-stat.
Constant	5.12	13.99	4.71	11.76
Nof	0.51	4.14	0.51	1.43
Noi	0.33	4.45	0.34	1.28
Pd			−0.13	−1.01
Retc	−0.37	−2.62	—	—
Vr	0.28	2.06	—	—
Nolb	0.24	4.48	0.25	1.64
Bi	0.60	4.01	0.41	1.07
Ti	0.58	3.12	—	—
Di	0.55	5.28	0.45	1.35
Hmi	0.88	2.89	—	—
Wof	−0.14	−1.49	—	—
Alv	−0.57	−4.55	−0.43	−1.06
Nol	−0.18	−2.81	—	—
Ts	0.38	1.35	—	—
γ (shape parameter)	0.26	—	—	—
Prob > χ^2	0		0.0067	

TABLE 4: Goodness of fit index and estimated distribution statistics of accident duration model.

Model statistics	Mean (min)	Variance	Maximum (min)	Minimum (min)	MAPE value
Observed value	192.95	111.63	510	20	0.22
Predicted value	188.38	84.52	327.14	53.03	

the results reveal that the accident occurs at regular road section or 4-way intersection results in longer duration than that occuring at other locations. The reason may be that the traffic volume is higher at regular road section or intersection. Regarding emergency services, the accident which needs tow services has longer duration. Moreover, as the number of lanes occupied in the accident increases, duration increases.

By using the accident duration model, the survival curve of duration is estimated, which is shown in Figure 4. Comparing with observed value, the prediction accuracy of accident duration model is shown in Table 4.

6. Conclusions

In this paper, a severity prediction model system was constructed by employing Ordered Probit model, and a duration prediction model was established by applying Hazard model. Accident severity, including number of fatalities, number of injuries, and property damage, as well as accident duration was forecasted with the models.

or 88% longer than that of other accidents, respectively. Besides, according to the results, the accident which occurs in weekend or festival is found to be associated with shorter duration. The reason is that the traffic volume in nonworking day is lower than that in working day. As for accident location,

Study results can be applied to severity and duration prediction, which are essential steps in accident response process. By comparing SVM and Ordered Probit model, it also makes a methodological contribution in enhancing prediction accuracy of severity estimation. In addition, by identifying the key effects of related factors on accident severity and duration, the results provide useful clues for government to take effective measures in order to reduce accident impacts and improve traffic safety.

One limitation of current study is that some factors, such as characteristics of the driver, passenger and pedestrian, and traffic condition, which have potential effects on accident severity and duration, are not considered because of the lack of suitable data. Further study should be done to collect the related information and investigate the impacts of these factors.

Acknowledgments

References

[1] L.-Y. Chang and F. Mannering, "Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents," *Accident Analysis and Prevention*, vol. 31, no. 5, pp. 579–592, 1999.

[2] H. Mannera and L. Wünsch-Ziegler, "Analyzing the severity of accidents on the German Autobahn," *Accident Analysis and Prevention*, vol. 57, no. 8, pp. 40–48, 2013.

[3] Y. Chung, "Development of an accident duration prediction model on the Korean Freeway Systems," *Accident Analysis and Prevention*, vol. 42, no. 1, pp. 282–289, 2010.

[4] P. C. Anastasopoulos, V. N. Shankar, J. E. Haddock, and F. L. Mannering, "A multivariate tobit analysis of highway accident-injury-severity rates," *Accident Analysis and Prevention*, vol. 45, pp. 110–119, 2012.

[5] D. Nam and F. Mannering, "An exploratory hazard-based analysis of highway incident duration," *Transportation Research A*, vol. 34, no. 2, pp. 85–102, 2000.

[6] J. C. Milton, V. N. Shankar, and F. L. Mannering, "Highway accident severities and the mixed logit model: an exploratory empirical analysis," *Accident Analysis and Prevention*, vol. 40, no. 1, pp. 260–266, 2008.

[7] N. V. Malyshkina and F. L. Mannering, "Markov switching multinomial logit model: an application to accident-injury severities," *Accident Analysis and Prevention*, vol. 41, no. 4, pp. 829–838, 2009.

[8] M. Stone and J. Broughton, "Getting off your bike: cycling accidents in Great Britain in 1990–1999," *Accident Analysis and Prevention*, vol. 35, no. 4, pp. 549–556, 2003.

[9] N. N. Sze and S. C. Wong, "Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes," *Accident Analysis and Prevention*, vol. 39, no. 6, pp. 1267–1278, 2007.

[10] D. Delen, R. Sharda, and M. Bessonov, "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks," *Accident Analysis and Prevention*, vol. 38, no. 3, pp. 434–444, 2006.

[11] M. F. Ballesteros, P. C. Dischinger, and P. Langenberg, "Pedestrian injuries and vehicle type in Maryland, 1995–1999," *Accident Analysis and Prevention*, vol. 36, no. 1, pp. 73–81, 2004.

[12] B. S. Roudsari, C. N. Mock, R. Kaufman, D. Grossman, B. Y. Henary, and J. Crandall, "Pedestrian crashes: higher injury severity and mortality rate for light truck vehicles compared with passenger vehicles," *Injury Prevention*, vol. 10, no. 3, pp. 154–158, 2004.

[13] H. Y. Zhang, *Analyzing Traffic Accident Situation with Bayesian Network [Doctor Dissertation]*, Jilin University, 2013.

[14] J. L. Bowman and M. E. Ben-Akiva, "Activity-based disaggregate travel demand model system with activity schedules," *Transportation Research A*, vol. 35, no. 1, pp. 1–28, 2000.

[15] P. Ray, "Independence of irrelevant alternatives," *Econometrica*, vol. 41, pp. 987–991, 1973.

[16] M. A. Quddus, R. B. Noland, and H. C. Chin, "An analysis of motorcycle injury and vehicle damage severity using ordered probit models," *Journal of Safety Research*, vol. 33, no. 4, pp. 445–462, 2002.

[17] B.-Z. Yao, C.-Y. Yang, J.-B. Yao, and J. Sun, "Tunnel surrounding rock displacement prediction using support vector machine," *International Journal of Computational Intelligence Systems*, vol. 3, no. 6, pp. 843–852, 2010.

[18] B. Yu, Z. Z. Yang, and B. Z. Yao, "Bus arrival time prediction using support vectormachines," *Journal of Intelligent Transportation Systems*, vol. 10, no. 4, pp. 151–158, 2006.

[19] F. Zong, H. Lin, and X. Pan, "Daily commute time prediction based on genetic algorithm," *Mathematical Problems in Engineering*, vol. 2012, Article ID 321574, 20 pages, 2012.

[20] A. Bolbol, T. Cheng, I. Tsapakis, and J. Haworth, "Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification," *Computers, Environment and Urban Systems*, vol. 36, no. 6, pp. 526–537, 2012.

[21] A. Apatean, A. Rogozan, and A. Bensrhair, "Visible-infrared fusion schemes for road obstacle classification," *Transportation Research C*, vol. 35, pp. 180–192, 2013.

[22] M. Abdel-Aty and K. Haleem, "Analyzing angle crashes at unsignalized intersections using machine learning techniques," *Accident Analysis and Prevention*, vol. 43, no. 1, pp. 461–470, 2011.

[23] K.-K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39–51, 1998.

[24] J. B. Yao, B. Z. Yao, L. Li, and Y. L. Jiang, "Hybrid model for displacement prediction of tunnel surrounding rock," *Neural Network World*, vol. 22, no. 3, pp. 263–275, 2012.

[25] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[26] Y. M. Li, S. G. Gong, and H. M. Liddell, "Support vector regression and classification based multi-view face detection and recognition," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 300–305, 2000.

[27] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[28] B. Yu, W. H. K. Lam, and M. L. Tam, "Bus arrival time prediction at bus stop with multiple routes," *Transportation Research C*, vol. 19, no. 6, pp. 1157–1170, 2011.

[29] A. Stathopoulos and M. G. Karlaftis, "Modeling duration of urban traffic congestion," *Journal of Transportation Engineering*, vol. 128, no. 6, pp. 587–590, 2002.

[30] P. Lambert, D. Collett, A. Kimber, and R. Johnson, "Parametric accelerated failure time models with random effects and an application to kidney transplant survival," *Statistics in Medicine*, vol. 23, no. 20, pp. 3177–3192, 2004.

[31] N. Keiding, P. K. Andersen, and J. P. Klein, "The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates," *Statistics in Medicine*, vol. 16, no. 1–3, pp. 215–224, 1997.

[32] R. Kay and N. Kinnersley, "On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: a case study in influenza," *Drug Information Journal*, vol. 36, no. 3, pp. 571–579, 2002.