Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 3713 - 3727, 2005

AN APPROACH-BASED CRASH ANALYSIS AND APPRAISAL MODEL FOR SAFETY DESIGN AND MANAGEMENT

Kevin P. HWANG
Associate Professor
Department of Transportation, Communication and Management Science
National Cheng-Kung University
Former Tainan City Transportation Commissioner
No. 1, University Boulevard, Tainan, 70101 Taiwan, Republic of China
Tel: +886-6-2757575-53228
Fax: +886-6-2753882
Email: hwangis@mail.ncku.edu.tw

Kai-Yuan WEI
Staff
Operations Strategy Dept., Operations & Maintenance Division
Taiwan High Speed Rail Corporation
2nd Floor, No. 100, Sec. 5, Hsin Yi Road, Taipei,
10105, Taiwan, Republic of China
Tel: +886-2-87892000-3207
Fax: +886-6-37252300
Email: kaiyuan_wei@thsrc.com.tw

Abstract: This research focuses on examining the interacting relationship between vehicle crashes and road engineering factors. The back-propagation artificial neural network (BPN) method with feed-forward structure is used to develop the model. Each intersection is decomposed into several approach sets to establish the proposed microscopic relationship with vehicle crashes. A total of 1,225 records at 69 primary/secondary intersections in Tainan, Taiwan, were used to calibrate the bi-level models of 11 crash types. First is a distinguish model to tell if an approach set will have crash; the second is a prediction model to forecast crash numbers by type for each approach set. Error rates for all distinguish models are under 15% with an average of 6.1%. Error rates for all prediction models are under 25% with an average of 6.7%. An overall RMS is 0.1617. Practical field test also proves model's validity.

Key Words: Crash, Approach set, Distinguish model, Prediction model

1. INTRODUCTION

Many factors contribute to the occurrence of vehicle crashes. They are commonly divided into three aspects: human, vehicle, and road environment. Their interactions are very complicate and their effects on different types of crashes are also very sophisticated. To improve road environment by physical engineering work normally depends heavily on experts and their experiences. They come to investigate a site, making judgment, and suggesting improvement countermeasures. But often experts are not available or unacquainted with traffic pattern of such a site. Consensus countermeasure is thus difficult to achieve. Therefore an appraisal model to preview any improvement effect has its merit.

Tradition linear or multiple regression analysis has proved its weakness in explaining the interrelationship between crashes and their contributing factors. Multivariate statistical analysis or Poisson regression has merit in establishing such an interrelationship. Artificial neural network (ANN) has already proved its ability in parallel calculation, high memory, error acceptance, and self-leaning. This research thus adopts ANN to solve the proposed approach set based microscopic model. With condition and collision diagrams, the model tries to find the engineering factors which will affect the occurrence of intersection crashes. And through application of the prediction model in countermeasure appraisal, different

improvement suggestion can be pre-evaluated before a site check by experts.

2. LITERATURE REVIEW

To establish the proposed ANN model with the aforementioned microscopic approach, literature about collision diagram, vehicle crash analysis, and application of ANN method in highway safety are reviewed.

2.1 Collision Diagram

A condition diagram presents a graphic display of a road section or an intersection on which geometric design, traffic control devices, and environmental features are shown, such as Figure 1 and Figure 2. A collision diagram uses representative symbols to illustrate crash patterns on a condition diagram. Through such a crash pattern display, a traffic engineer can diagnose possible engineering crash causes for the site. Several systems capable of drawing collision diagram and performing crash pattern analysis are available in this field, such as the WACDS (Washington Automated Collision Diagramming System) from Washington State Transportation Center (Nyerges T.L. and Cihon, R.F. 1989).

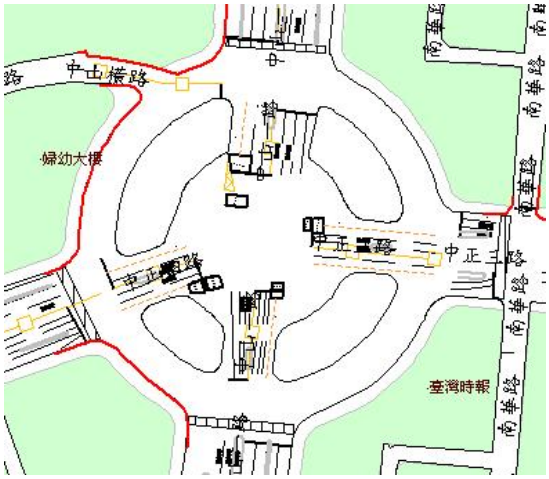

Figure 1. Condition Diagram

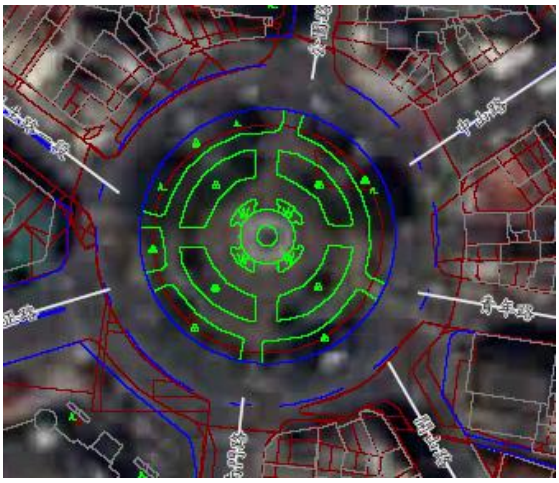

Figure 2. Condition Diagram overlapped with Aerial Photo

2.2 Crash Analysis

Before the model of this research is developed, environmental factors which may contribute to crash occurrence and the way to identify them are reviewed and summarized in Table 1 and 2.

Table 1 Analysis Methods and Findings

Researchers	Analysis Method	Findings
Poch, M. and Mannering, F. (1996)	Negative Binomial Regression	Model can be used to remove crash causing attributes when creating a new or improving an old intersection.
Miaou, S.P. and Lum, H. (1993)	Poisson Regression	Poisson regression is better than linear regression.
Turner, S. and Nicholson, A. (1998)	Poisson Regression	Prediction accuracy for suburban intersection is about 30%, for urban intersection is 40%.

Table 2 Analysis Methods and Findings from Taiwan

Researchers	Analysis Method	Findings
Xu, Tian-ben (1995)	Spot map analysis and risk analysis	The method can highlight the difference between crashes to improve individual intersection at a much more detailed level.
Xu, Tian-ben (1997)	Highway safety checking list	Checking list can be used at different stages such as at planning, design, appraisal, or improvement stage.
Shi, Feng-yu (1987)	Multiple Regression	Vehicle type and turning ratio are highly correlated with crashes. Traffic volume, pedestrian, and road features are less correlated with crash.
Zhan, Bing-yuan (1990)	Conflict Analysis	Motorcycles have great impact on traffic flow especially at low flow situation. However, conflicts turn less while traffic volume is close to capacity at lower speed.

2.3 ANN in Highway Safety Analysis

Chao (1996) uses 16 input variables, such as mile post, section length, time of day, weather, lane number, etc., to model vehicle crash risk and severity. The error rate is between 10%~13%. Mussone, Rinelli, and Reitani (1996) apply 12 attributes, including driver, vehicle, and environmental factors in their developed back propagations ANN model. They found that crashes do not homogenously occur on highway, thus output of the model can't be viewed as crash possibility. Sayed and Abdelwahab (1998) use 21 features, including alignment, slope, speed limit, pavement, weather, lighting, land use, time, crash occurring at an intersection or not, crash pattern, severity, traffic control device, volume/capacity ratio, vehicle type, driver characteristics, to calibrate a back propagation ANN model. In addition, performance of the model is compared with another model developed by the Fuzzy K-nearest Neighbors method. They reached conclusions that ANN model is slightly better than Fussy model in prediction, and ANN model in general has better calculation performance; namely faster to converge.

3. MODEL STRUCTURE

3.1 Approach Set

Because vehicle is a major entity of a crash, its moving direction is very critical when who has the right of way is considered in crash analysis. At an intersection, there are many approaches (or legs), flows from different approaches form movements and moving sequence which is primarily controlled by signal phase plan. Since there are many movements and a crash can occur at any movement, knowing which movement involving a crash is more important than just knowing an overall crash pattern. Thus, by displaying movement conflict through delicate collision diagrams to perform crash analysis is the basic concept of this research. Movement conflict is used to develop the proposed theoretical model to diagnose environmental features and to predict possible occurrence of crashes. To represent those dynamic movements, the static approach configuration of an intersection is used. Because movement conflict often determines crash type, therefore, the proposed model is further classified into a more detailed crash type level.

Because vehicles enter/leave intersection from/to different approaches and face different environments, to avoid ambiguity this research defines “approach” as “incoming upstream approach” to represent the movement (maneuver) involving crash. Therefore a T intersection has three approaches and a normal 4-leg intersection has four approaches, and when a vehicle or a pedestrian involves in a crash, an approach is assigned as its attribute.

Since a crash may occur when vehicles coming from the same approach or from different approaches, approaches (legs) of an intersection have to set up with intersection’s signal phase plan to form maneuver set or approach set which is called in this research. Therefore an intersection of N legs will normally have N(N+1)/2 approach sets, such as Figure 3 illustrates. As for a round-about, it is looked as several T intersections neighboring each other, as illustrated in Figure 4.

It is worthy reminding the features and benefit of using such a microscopic approach to conduct crash analysis and to build appraisal model for intersection safety improvement. First, crashes of the same type at an intersection may not belong to the same approach set (maneuver set). Different legs have different feature, safety improvement can be done with the whole intersection or can be confined to certain approach only. However, an improvement has to be very specific. Second, if analysis is done at intersection level, those examined intersections can only be classified into a few categories, but there actually exist great differences among them. For example, the 69 intersections analyzed in this research can be categorized into 6 groups only; namely 6 types of intersections, but significant variations still exist within each group. How to distinguish between two similar intersections belong to the same group and model their difference to identify crash causing factors is very difficult, and that is the purpose of this research.

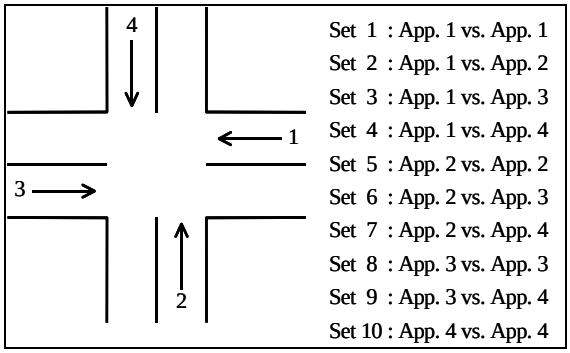

Figure 3. Definition of a 4-leg Approach Sets

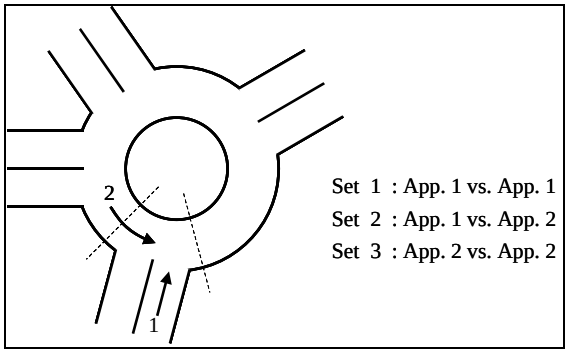

Figure 4. Definition of a Round-about Approach Sets

Therefore decomposing an intersection into N(N+1)/2 approach sets renders a possibility to conduct crash analysis at a more detailed and precise level. Those factors unable to be modeled as crash causing features can then be modeled. Such a research approach is also consistent with the application of collision diagram in crash analysis except the latter is executed at an intersection level and no model calibration is available (Nyerges, T.L. and Cihon, R.F. 1989).

3.2 Data Preparation

Before the model can be developed, there exist some issues need to be solved. For example, a single vehicle crash occurs on one approach therefore no approach set can be defined. To

overcome such a problem, single vehicle crash is grouped into crash type of hitting traffic device and given an extra pseudo approach which is the same as its incoming approach. As for multi-vehicle crash, only the most two critical vehicles are identified for analysis, the other vehicle(s) are excluded.

Of the 69 primary and secondary intersections of Tainan, Taiwan, selected for modeling, they are decomposed into 745 approach sets. In average each intersection is decomposed into 11 approach sets, which is greater than 10 ($11 > 4(4+1)/2=10$), 10 is a normal approach set number for a 4-leg intersection. The reason is that an exclusive left turn phase will add extra approach sets to a normal 2-phase 4-leg intersection.

3.3 A Bi-Level Model

Analysis of the source data shows that many approach sets (33%, 249 approach sets) of the 69 intersections have no records of crash experience. This implies that many approach sets can't be used to develop the prediction model; otherwise bias and error will exist within a reliable model. After several preliminary tests the aforementioned doubt is proved, thus this research takes a bi-level approach to handle such a situation. At the first level, a distinguish ANN model is trained to tell if an approach set will have crash experience. Thereafter, those approach sets identified to have crash experience are collected to train the second level model. It is a crash number prediction ANN model.

For the purpose of both model training and evaluation process, 596 (80%) approach sets were randomly selected to train the distinguish ANN model. After training, the other 149 (20%) approach sets were used to evaluate the accuracy of the ANN model. As the convergence threshold of error rate at model's training process, this study referred to the research of Chao (1996) and selected 15% as the threshold. The repetitive ANN training process will stop after an error rate below 15% is achieved.

After the distinguish model is developed, the model is used to screen all 745 approach sets to identify the approach sets having crash experience. Note some distinguishing error exists and purposely retained to keep the bi-level models consistent, which implies that not only the approach sets with crash experience are used to develop the second level model. Of those approach sets identified with crash, 80% of them are used to train the prediction ANN model. Thereafter, the other 20% approach sets are added to the model to evaluate it validity. From the research of Poch and Mannering (1996), they have achieved a 21.67% error rate in their frequency prediction model. This study thus selects 25% as an error rate threshold for model training. If the error rate of a trained model is above 25%, parameter of the ANN model is modified and retrained until an acceptable ANN model is achieved. What is the reason the two models apply different thresholds is explained as below.

Because at the first distinguish level, the output is yes (1) with crash experience or no (0) without crash experience; and at the second prediction level, the output is the predicted crash number by type, therefore the prediction model is much more difficult to achieve an acceptable error rate than the distinguish model. In addition, output from the first level model is used to develop the second level model, error of the first level model is embedded in the second level model, thus a lower criteria of 25% compared to 15% as an error rate threshold is selected in the training process of the second level model.

3.4 Variable Selection

Dependent Variable and Crash Type:

At the first level model, the dependent variable is a logic value 0 or 1: an approach set with crash will be given a “1” otherwise a “0”. For the second level model, crash numbers for approach sets is used as dependent variable. Since both distinguish and prediction models are built by crash types, crash numbers of different crash types have to be counted. This research classifies crash types as shown in Table 3. Because the detail of hitting fixed facilities or others is not the major concern of this study and their crash numbers are relatively small, those further detailed items are combined into a K category. In addition, because the size of pedestrian related crashes is small, those detailed items are also grouped into an L category. Thus, a total of 12 crash types are initially included in the modeling process.

Independent Variables:

Since this research is aimed to appraise the effectiveness of environmental improvement, only engineering and traffic related factors are considered as independent variable in the model. Through literature review, 14 engineering factors and 8 traffic factors are selected as initial independent variables. They are explained more detailed as follows:

Table 3. Crash Type Classifications

Road crash record table			This research	
Code	Class I	Class II	Code	Classifications
A	Crashes between cars	Head on	A	Head on
B		Opposite direction sideswipe	B	Opposite direction sideswipe
C		Same direction sideswipe	C	Same direction sideswipe
D		Rear-end	D	Rear-end
E		Backing	E	Backing
F		Different directions within intersection	F	Different directions within intersection
G		Angle	G	Angle
H		Others	H	Others
I	Single car crash	Overturn (Capsize)	I	Overturn (Capsize)
J		Out of road	J	Out of road
K	Hit fixed facilities or other things	Guardrail	K	Hit fixed facilities or other things
L		Sign / traffic light		
M		Toll-house		
N		Traffic island		
O		Non-fixed facilities		
P		Bridge / house		
Q		Trees / Telegraph pole		
R		Car parked		
S		Animals		
T		Roadwork		
U		Others		
1	Pedestrian crashes	Opposite direction	L	Pedestrian crashes
2		Same direction		
3		Crossing the road		
4		Playing on road		
5		Working on road		

6		Running into road		
7		Getting out of (or behind)car		
8		Standing on sidewalk		
9		Others		

Engineering Related Variables:

After literature and engineering design standard review, this research summaries the following factors to be considered in the model. They are listed in Table 4 and those input data are obtained with the assistance of a Tainan City geographic information system, such as Figures 5 and 6 show. In addition, a set of 1:1,000 digital electronic maps (similar to Figure 1) are also used to calculate input data such as road width, angle between roads, etc. As for density of manholes, number of signs, and other data unavailable from the digital maps, this research employed a video-logging system to collect the needed data.

Table 4. Codes of Engineering Factors

Factor	Code	Note
Central reserve style	NCEN	0: Traffic island, 1: Reflector, 2: Marking only
Division between fast lanes and slow lanes	LANE	0: No slow lane, 1: Reflector, 2: Marking only
Number of left turn lanes	LL	Number of lanes
Number of right turn lanes	RL	Number of lanes
Width of fast lane	NMLW	In 0.1meter precision
Width of road	MRW	In meters precision
Width of non-fast lane	MSW	In 0.1meter precision
Slope (Grade)	S	0: Non-slope (level road), 1: Slope
Density of manholes	NPDR	Number of manholes / 100 ㎡
Illumination	NLI	0: no light, 1: normal light, 2: enhanced light
Number of signs	SG	Number of signs/Approach
Minimum speed to cross intersection	NMWCH	Width of road / (Yellow time + All red time)
Angle between roads	NDA	In degrees from 1:1,000 digital map

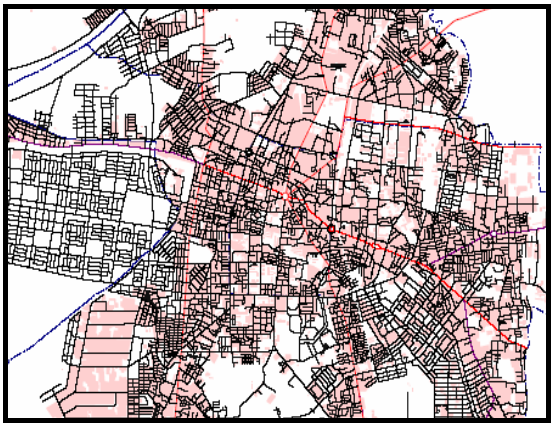

Figure 5. Tainan City GIS

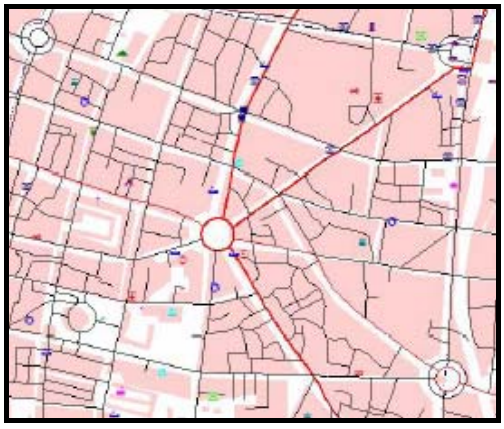

Figure 6. Enlarged Area Map

Traffic Factors:

Traffic conditions, namely vehicle exposures, have impact on the occurrence and number of crash. Those traffic factors considered are listed in Table 5.

Table 5. Codes of Traffic Conditions

Factor	Code	Note
Flow of intersection	NIF	In PCU, 1 decimal precision
Flow of approach set	NAF	In PCU, 1 decimal precision
Peak hour factor	PHF	0.00-1.00, 2 decimal precision
Left turn ratio	LTR	0.00-1.00, 2 decimal precision
Right turn ratio	RTR	0.00-1.00, 2 decimal precision
Heavy veh ratio	HVR	0.00-1.00, 2 decimal precision
Motorcycle ratio	MTR	0.00-1.00, 2 decimal precision
Lane blocked condition	NB	0: not blocked, 1: road side is blocked 2: slow lane is blocked, 3: fast lane is blocked

4. DATA SOURCE AND DESCRIPTION

4.1 Data Source

This research uses vehicle crash records from Tainan Police Bureau as a data base. During a period of 16 months, a total of 12,976 crashes are recorded with an average of 811 crashes per month. After checking each intersection with crash record to see if it has proper traffic data as required in Table 5, 69 primary and secondary intersections are selected for further analysis. There are 1,225 crashes belong to those 69 intersections, with an average of 17.8 crashes per intersection.

As for traffic volume, related survey data are used and data interpolation has to be performed to obtain proper data for the model.

4.2 Data Description

Of the 1,225 crash records, 39.72% of them belong to the 4-leg intersections; 54.22% happens within intersection, and the rest of them occur at spots approaching or leaving intersections. It is also noted that most crashes (61.79%) occurs between 8 a.m. and 7 p.m. However at morning or evening peak hours, no significant crash peak is observed. As for crash sites, they are mainly within 4-leg or multi-leg intersections. As for 3-leg intersection, angle crash is the major crash type; and for roundabout, sideswipe by opposite flow is the major crash type. In addition, it worth mentioning is that single vehicle crash or pedestrian related crash tends to be more serious than average crash.

5. Model Development and Application

5.1 Data Processing and Parameter Setting

Before modeling, each intersection is decomposed into several approach sets and crash

records are assigned to those 745 approach sets. Table 6 illustrates the basic statistics of the source data after decomposition and assignment. It is seen from Table 6, all 1,225 crashes can be attributed to 754 approach sets. However, since an approach set may have more than one type of crash, so crash-involving approach sets are much smaller than 754. Actually, those 1,225 crashes occur at 496 approach sets; that is 67% of the 745 approach sets. From Table 6, the number and percentage of crash type involving approach sets is also seen; it tells that angle crashes involve the most percentage of approach sets; 201 out of 745 (27%).

Table 6. Binary Outputs Basic Statistics

Crash Classifications	Number of Crash Approach Set	% of 745 Approach Set	Standard Error	Skewness Factor
Head on	21	3	0.17	5.7128
Opposite direction sideswipe	13	2	0.13	7.3855
Same direction sideswipe	129	17	0.38	1.7311
Rear-end	148	20	0.40	1.5136
Backing	20	3	0.16	5.8665
Different directions within intersection	188	25	0.43	1.1426
Angle	201	27	0.44	1.0394
Other	19	3	0.16	6.0318
Overturn (Capsize)	1	0	0.04	27.2947
Out of road	0	0	0.00	N/A
Hit fixed facilities or other things	7	1	0.10	10.1910
Pedestrian crashes	17	2	0.15	6.4041
Total of above	754	103	--	--
Sum	496	67	0.47	-0.7043

After data compilation, this research employs feed-forward structure and the Back-Propagation Network (BPN) method of the Supervised Learning Network to develop the proposed ANN model. Parameters and their initial values of the model are listed in Table 7.

Table 7. Network Variables' Initial Values

Network Variables	Initial Value
Hidden layers	1 layer
Input neuron	Chosen by linear variation decision model of software
Neurons in hidden layer	Decided by overall net searching model of software, but set 30 as the maximum.
Transfer function in hidden layer	Tanh
Output neuron	1
Output transfer function	Sigmoid
Learning rule	Adaptive gradient rule

5.2 Model Training and Result

For the proposed bi-level model, the first distinguish model includes 11 crash-type

sub-models. Since there has no type J (Out of road) crash record (see as in Table 6); it is excluded. After development of the distinguish model, prediction model is further developed. Because crash type I (Overturn) has only one record, it is also excluded at this stage. Therefore only 10 crash-type prediction sub-models are developed.

Table 8 illustrates the environmental contributing variables included in the 11 distinguish sub-models and in the 10 prediction sub-models. Table 9 shows the meaning of transformation for each variable in Table 8.

Table 8 Variables Chosen by Crash-type Models

Variables	Head on	Opposite direction sideswipe	Same direction sideswipe	Rear-end	Backing	Different directions within intersection	Angle	Other	Overturn (Capsize)	Hit fixed facilities or other things	Pedestrian crashes
NIF		△	△	△							
tanh(NIF)	△			△	○	∨	△				
NAF	△					△		△			○
tanh(NAF)	△	△								○	
NAF^2					○						
fzrgt(NAF)							○				
fzlft(NAF)								○	○		
MinPHF	○	△	○				△			○	
tanh(MinPHF)	○	○	△		△		○	∨	○		
LTR	△		△		○		△				
tanh(LTR)			△			△	○				
Exp(LTR)					△						
RTR		○	△	△		∨					
tanh(RTR)		○				∨				○	
In(RTR)								△			
HVR			∨	△		△	∨	△		△	
tanh(HVR)			△	∨		△					
Inv(HVR)					△						
MTR		○	△	△	○	△	△			○	
tanh(MTR)		○								○	
NDA	○		○	○		○	∨	△		△	△
NDA^2	○	○	○	∨	∨	○	○	△		○	
NPDR	∨				∨			○	○	○	∨
tanh(NPDR)			△			△				○	
NPDR^2							△				
NLI	○		△		○	△	○	○	○	∨	○
NLI^2				∨							
NMWCH										○	
tanh(NMWCH)			∨		○	∨		△		○	
NB		△			∨	△					△
NB^2			△								
ln(NB)	○				○						

Variables	Head on	Opposite direction sideswipe	Same direction sideswipe	Rear-end	Backing	Different directions within intersection	Angle	Other	Overturn (Capsize)	Hit fixed facilities or other things	Pedestrian crashes
ln(NB/(1-NB))							△			○	
NCEN		○		△	△	∨	○			○	○
ln(NCEN)	○		○				∨				○
NCEN^2			△	△							
Sqrt(NCEN)	△										
LANE	○	∨		○	△	∨		△		∨	
LANE^2	○		○								
LL			○	△	△	△				△	
LL^2	△										
Inv(LL)		○			○	○				○	○
Logical(LL)			△					△			
RL	△	△	∨		○	△		△			○
tanh(RL)				△						○	
RL^2	△										
NMLW	△				∨		△				
tanh(NMLW)					△	○	△				○
NMLW^2								△			
fzlft(NMLW)						△					
MRW			△		○	△		○	○	∨	
tanh(MRW)					∨	∨					
fzrgt(MRW)						○	○	○	○		
Inv(MRW^2)							△				
MSW			∨	○	○	○		○	○	○	
tanh(MSW)	○		○			○	∨	○	○		
tanh(S)											
S	∨			○	△						
S^2							△				
Inv(S)						△	○				
Logical(S)			△					△			
PH	○	○		△	△	△	△				△
tanh(PH)	△										
ln(PH)			△								
PH^2				○			△				
SG		△			○	○	△	△			∨
tanh(SG)						△					
SG^2			△	○		○	○				
fzrgt(SG)			○								
Number of ○	11	9	8	6	12	9	9	7	8	15	7
Number of △	10	6	16	10	9	15	13	12	0	3	3
Number of ∨	2	1	4	3	5	7	4	1	0	3	2
○+△+2∨	25	17	32	22	31	38	30	21	8	24	14

Note: ◯ indicates variable chosen by distinguish model; △ indicates variable chosen by prediction model; ∨ indicates variable chosen by both bi-level models。

From Table 8, it is found that the head-on, opposite direction sideswipe, same direction sideswipe, and rear end crash distinguish models, do not use intersection traffic volume (NIF) and approach set volume (NAF) as input variables. In addition, road width (MRW) which implies road capacity is also excluded by those distinguish model. It means the changes of intersection volume or approach set volume do not affect an approach set's crash experience. Nevertheless, in the prediction model, volume related variables are all included in the aforementioned distinguish models. It implies that volumes (NIF and NAF) are related with crash number.

Table 9. Variables Procession Codes

Procession Codes	Procession Method	Note
tanh(X)	$\frac{e^{x}+e^{-x}}{e^{x}-e^{-x}}$	
X^2	X^2	
fzrgt(X)	Fuzzy Right(X)	Fuzzy right function
fzlft(X)	Fuzzy Left(X)	Fuzzy left function
ln	ln(X)	
$\ln(\frac{X}{1-X})$	$\ln(\frac{X}{1-X})$	
Inv(X)	$\frac{1}{X}$	
Exp(X)	e^{X}	
Sqrt(X)	$\sqrt{X}$	
Logical	Logical analysis X	If X is bigger than threshold, output is the max X, else the minimum X.

After model training, performance of the first level distinguish models have their error rates between 2.01% and 14.77% with an average value of 6.07%. For the second level prediction models, their error rates are between 0% and 23.76% with an average value of 6.67%. If combining the bi-level models into one integral model, the performance is shown in Table 10. It is learned from Table 10 that the correlation coefficients between the actual crash number and predicted number for all 10 integral models are above 0.7 except for the overturn crash type which does not have prediction model and its coefficient can't be calculated either. It is also noted that with lower error rate, the correlation coefficient will be higher and the root mean square (RMS) will be lower.

Table 10. Integral (Distinguish and Prediction) Model Performance

Crash Types	Error Rate	Correlation Coefficient (Integral Model)	C.C. Prediction Only	RMS (Integral model)	RMS
Head on	0.54%	0.8973	0.8965	0.0733	0.0735
Opposite direction sideswipe	0.27%	0.9186	0.9200	0.0518	0.0516
Same direction	13.56%	0.7407	0.7716	0.5816	0.5365

sideswipe					
Rear-end	10.74%	0.8394	0.8468	0.4997	0.4777
Backing	0.40%	0.9201	0.9173	0.0635	0.0645
Different directions within intersection	23.76%	0.5796	0.6532	0.7226	0.6665
Angle	22.68%	0.6868	0.7174	0.7985	0.7376
Other	1.07%	0.7948	0.7934	0.1036	0.1036
Overturn (Capsize)	0.13%	N/A	-0.0073	0.0366	0.1323
Out of road	0.00%	1.0000	0.9999	0.0000	0.0020
Hit fixed facilities or other things	0.27%	0.8632	0.8632	0.0819	0.0819

5.3 Practical Appraisal and Application

After the 11 distinguish and 10 prediction crash type models were developed, several practical tests were conducted to appraise the usefulness of the proposed method. Since only one case was practically field investigated by an expert team, it is selected in this paper as the target intersection (Figures 7 and 8). It is a 5-legs intersection with an extraordinary 36 crash records observed during the same 16 months data collecting period which is much higher than the average crash number 17.8 in this study.

At experts' field investigation they found very few engineering work can be done except some management proposal. Nevertheless, some improvements finally come out after group discussion. They include: 1. remove a redundant truck-path sign, and 2. reinforce parking violation enforcement, and some other proposals which can't be evaluated by this model. This research thus evaluates the applicability and usefulness of the model. At the first stage, the prediction model estimates the intersection has 24 crashes compared to its original 36 crashes, which has an error rate of 33.3%. Secondly, the suggested improvements are appraised by the prediction model; it predicts 19 crashes will still occur after the improvement, which implies the suggested improvements can only achieve 29% effectiveness.

Although the result seems not convincing of model's applicability, it has to be noted that this target intersection is an outlier of the 69 intersections included to develop the model. And practical appraisal of the model is to see the usefulness of the model in evaluating effectiveness of the improvement proposal. As for prediction accuracy of the model it has been presented in Table 10. Truly, the 33.3% is higher than the highest number 23.76% in Table 10. However, it has two folds of meaning. It first implies that prediction error of each crash type will add up and it might finally turn out that intersection crash prediction error is higher than approach based crash type prediction error. On the contrary, it does imply that approach based crash type analysis is more accurate than intersection based crash analysis.

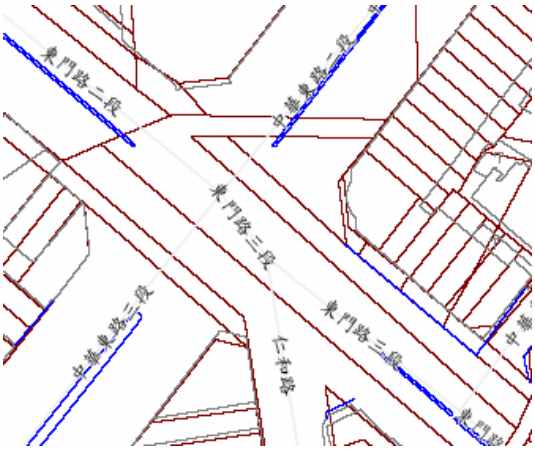

Figure 7. Target Intersection

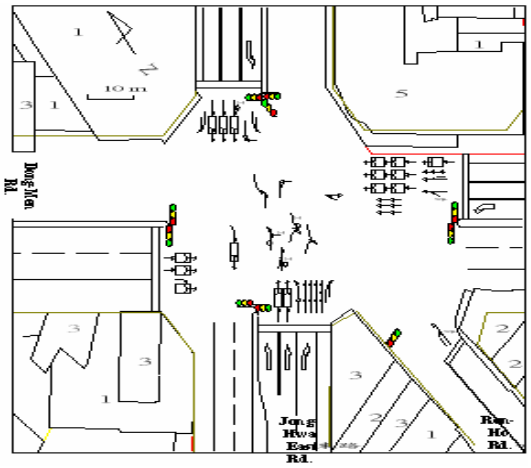

Figure 8. Target Intersection with Crash Records

6. CONCLUSION AND RECOMMENDATION

6.1 Conclusion

Through this research, several conclusions can be reached. They are

1. Decomposition of an intersection into several traffic maneuvering approach sets is a valid method to conduct vehicle crash analysis.
2. Crash analysis with the proposed microscopic approach can be supported by the ANN method. Though display of crash records by the proposed method is similar to the illustration of collision diagram, the method is not a judgmental or experience based method.
3. A bi-level crash type based model is developed. The first level is a distinguish model and the second level is a prediction model. Eleven crash type distinguish models and 10 crash type prediction models have been developed by the BPN method. An integral model to combine the bi-level model into a one step model is also developed.
4. The ANN distinguish model can be used to tell whether an approach will be crash free with its current engineering and traffic characteristics.
5. The ANN prediction model can be used to tell how many type-based crashes an approach set will have.
6. The developed bi-level model can be used to appraise proposal for engineering improvement before field implementation.

6.2 Recommendation

For this research, they are also some recommendations:

1. The proposed approach require intensive data to train the model, further sensitivity analysis may be helpful to simplify the model. In addition, keeping good and reliable traffic data is important.
2. Intersection configuration can be obtained through electronic digital maps or through aerial photo images; therefore, capability to use a GIS is important.
3. Further integration to simplify the bi-level model into a reliable one step model is recommended.

4. Neural network method behaves and serves like a black box, less detail can be unveiled during model training process. It is thus encouraged to conduct a comparative study by other method to reveal the importance of the crash contributing factor identified in this research.
5. The 33.3% error rate of the application case is above the 25% threshold of the prediction model. Although the former is at intersection level and the later is at approach crash type level, more cases with filed investigation is encouraged.

REFERCNCES

a) Journal papers

Poch, M. and Mannering, F. (1996) Negative binominal analysis of intersection-accident frequencies, Transportation Engineering, Vol. 122, No. 2, 105-113.

Miaou, S.P. and Lum, H. (1993) Modeling vehicle accidents and highway geometric design relationships, Accident Analysis and Prevention, Vol. 25, No. 6, 689-709.

Mussone, L., Rinelli, S. and Reitani G.. (1996) Estimating the accident probability of a vehicular flow by means of an artificial neural network, Environment and Planning B: Planning & Design, Vol. 23, No.6, 667-675.

Sayed, T. and Abdelwahab, W. (1998) Comparison of fuzzy and neural classifiers for road accidents analysis, Computing in Civil Engineering, Vol.12, No.1, 42-47.

Turner, S. and Nicholson, A. (1998) Intersection accident estimation: The role of intersection location and non-collision flows, Accident Analysis and Prevention, Vol.30, No.4, 505-517.

b) Papers presented to conferences

Xu, Tian-ben and Rao, Zhi-ping (1995) The application of spot accident analysis at signalized intersection, The 2nd Traffic Safety and Enforcement Conference, Taiwan, 5, Oct. 1995.

Xu, Tian-ben and Wang, Yi-chuan (1997) A research of traffic safety check lists, The 4th Transportation Safety Conference, Taiwan, 10, August 1997.

c) Other documents

Chao, Chung Jen, (1996) A research to distinguish dangerous section of highway by artificial neural networks, Master Thesis, Central Police University, Taiwan.

Nyerges, T.L. and Cihon, R.F. (1989) Automated collision diagram production, Report No: WA-RD a80.1: Contract/Grant Number: GC8286, Seattle.

Shi, Feng-yu (1987) Urban accidental spots distinguish and analysis model, Master Thesis, National Taiwan University, Taiwan.

Zhan, Bing-yuan (1990) Intersection analysis and improvement strategies developed by traffic conflict theory, Master Thesis, Central Police University, Taiwan.