ANLTERNATIVE APPROACHES OF IDENTIFYING ACCIDENT PRONE LOCATION

By

CHARLENE WANG

B.Sc., Hohai University, 1989

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF CIVIL ENGINEERING

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

APRIL 1999

ABSTRACT

Despite the many advances in highway design, traffic engineering, automobile manufacturing, and police enforcement technologies, many countries around the world still suffer from an ever-increasing problem of traffic accidents. Therefore, there has been a keen interest throughout the world in developing Road Safety Improvement Programs (RSIPs) aimed at: 1) identifying accident-prone locations; 2) diagnosing their problems; and 3) suggesting proper countermeasures. While the success of these programs has varied considerably from one jurisdiction to another, the overall performance of these programs has been less than satisfactory in terms of number of accidents eliminated. The main reason for that is believed to relate to the inadequacy of procedures adopted in the execution of these programs. Most importantly, the faulty identification of accident-prone locations (i.e., identification of locations that are not really accident-prone) seems to be the primary reason behind the lack of success of these programs.

This thesis discusses the problems encountered in developing and implementing RSIPs from an engineering perspective. It also describes the efforts in developing new techniques to make these programs more effective in identifying and treating accident-prone locations. Sayed (1995) has described two new techniques for the identification process. The first one is the Modified RSIP that alters the definition of accident-prone locations and introduces the concept of correctable accidents. The second one is the Countermeasure-Based RSIP that starts by identifying prevailing accident patterns and then suggest proper engineering countermeasures, thus effectively reversing the normal flow of procedures in traditional RSIPs. This thesis introduces further refinements to

these new techniques. The classification process of accidents in the first method is refined using artificial neural networks and neuro-fuzzy models. Accident prediction models are used to identify over-represented accident patterns in the second method. These refinements significantly improve the results of the two methods. Examples of real-life applications are given and their results are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENT

It's been a tough year for me last year with the birth of my baby girl and my thesis writing going on at the same time. With much help, I have finally finished. There are many people I would like to acknowledge for contributing to my study.

First, I would like to thank my supervisor, Dr. Tarek Sayed, for his careful guidance and supervision throughout my study, for his encouragement and constructive suggestions, and for all the time he spent in reviewing my thesis.

Second, I would like to thank my classmate, Luis Felipe Rodriguez, for the time he spent in explaining his accident prediction models that I referenced in my thesis.

Finally, I would like to thank my husband Jian Huang for his support, encouragement and understanding. Specially, I would like to thank my soon-to-be seven year old daughter Michelle for watching her little baby sister Alyssa while I was busy writing my thesis and for her understanding for not being able to spend enough time with her mother. I'm so grateful to have their support and love, I could not possibly to finish my thesis without them.

CHAPTER 1

INTRODUCTION

1.1 Background

There has been a serious concern about traffic safety since the start of the automobile age, approximately eleven decades ago. In spite of this concern, traffic safety problems have prevailed over the past century causing enormous economic and social costs. In 1995, there were 534 fatalities, about 50,000 injuries and 180,000 property damage only accidents. The direct annual cost to the province exceeded 2.0 billion dollars (ICBC 1995 Annual report, Vancouver, B.C.). Recognizing these safety problems and the need to reduce the social and economic costs associated with them, road safety authorities have established Road Safety Improvement Programs (RSIPs). The objective of these programs is to monitor traffic conditions, collect and analyze accident data, locate trouble spots with abnormally high accident frequencies and implement appropriate and effective countermeasures in order to improve the safety potential of these sites. The RSIPs adopt a procedure which has the following three phases:

Detection: identification of accident-prone locations;

Diagnosis: assessment of the causes of the safety problems at the accident prone locations; and

Remedy: recommendation of the most effective countermeasure(s) to alleviate the safety problems at the identified locations.

While the success of these programs has varied considerably from one jurisdiction to another, the overall performance of these programs has been less than satisfactory in terms of number of accidents eliminated. The main reason for that is believed to related to the inadequacy of procedure adopted in the execution of these programs. Most importantly, the faulty identification of accident-prone locations (i.e., identification of locations that are not really accident-prone) seems to be the primary reason behind the lack of success of these programs.

In traditional RSIPs programs locations are identified as accident-prone on the basis of the total number of accidents. Traditionally, locations that exhibit a higher accident occurrence than an established "norm" are deemed to be hazardous. However, this criterion provides no consideration of the factors contributing to the accidents and provides little or no insight into whether the safety at these locations can be enhanced by road improvements. These traditional methods may result in the identification of locations that are not truly hazardous from a road safety authority perspective, and consequently, may lead to misapplication of safety improvement funds. Sayed (1995) proposed two alternative approaches to identify hazardous locations.

The first is the "modified black spot program" which is implemented to identify hazardous locations on the basis of their contributing factors and causes. This approach uses a fuzzy pattern recognition algorithm to classify accidents, according to their contributing factors and causes, into one or a combination of the three main highway system elements: the driver, the vehicle and the road. Statistical techniques such as the Empirical Bayes approach are then used to identify hazardous locations from a road, driver or vehicle point of view (Sayed, Abdelwahab and Navin, 1995). This thesis will introduce a further refinement to the classification process: artificial neural networks and neuro-fuzzy models. A comparative evaluation of the classifiers is carried out and their relative advantages are discussed.

The second approach is "the countermeasure-based program" which argues that a location with a given number of accidents with well-defined patterns can be treated more effectively than a location with a larger number of accidents with poorly defined patterns. Traditional approaches start with a problem (high accident occurrence) and attempt to find solutions (countermeasures). The countermeasure-based program reverses the traditional process by first identifying main accident patterns that can be targeted by specific countermeasures and then searching for locations which have over-representation of these patterns. Sayed (1995) assessed the over representation of accident patterns by the likelihood that the ratio of the number of accidents of a particular pattern to the total number of accidents at the location is higher than usual. This thesis will introduce a further refinement to the identification process using accident prediction models. The advantages of using these models will be discussed and their usefulness demonstrated using case studies.

1.2 Thesis Structure

This thesis is divided into seven chapters. Chapter One provides an overview of the thesis and its structure. Chapter Two summarizes previous work on identifying hazardous locations and briefly discusses the two new identification approaches suggested by Sayed (1995): the Modified RSIP and the Countermeasure-Based RSIP. Chapter Three introduces a refinement to accident classification in the Modified RSIP using artificial neural networks. The advantages of using artificial neural networks in accident classification are discussed. Chapter Four presents a further improvement to the classification process using neuro fuzzy techniques. A comparative evaluation of artificial neural networks and the neuro fuzzy techniques is carried out and the results discussed. Chapter Five describes the application of the classification process for the identification of accident-prone locations from a broader perspective than the engineering improvement. Chapter Six deals with using accident prediction models to improve the identification of treatable locations in the countermeasure-based program. Chapter Seven provides suggestions for further research and the summary and conclusion of the thesis.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

As mentioned earlier, recognizing the importance of reducing the social and economic costs of road accidents, the majority of road authorities have established Road Safety Improvement Programs (RSIPs). The main goal of these programs is to identify locations that may have safety problems "hazardous locations[1]" and establish countermeasures to correct them. In broad terms, the development of RSIPs involves the following functions:

- Location identification, or detection (i.e. which locations are considered hazardous),
- Problem identification, or diagnosis (i.e. what causes the identified locations to be hazardous) and
- Solution identification, or remedy (i.e. given these locations and their problems, what countermeasures are effective to alleviate the problem)

The first phase (detection) defines the scope and the size of the safety problem. The following is a description of it.

[1] The terms: hazardous, accident-prone, black spot are used interchangeably in road safety analysis.

2.2 Black Spot Programs

Programs to identify hazardous locations are usually called "Black Spot Programs". The basic assumption of these programs is that the road design plays a contributory role in the occurrence of many road accidents. Therefore, improving the engineering elements of hazardous locations can avert a significant proportion of accidents. Problem locations may be defined as specific sites (intersections or short road sections) or they may have broader definitions such as routes and areas, the latter is normally reserved for residential areas. These locations are usually considered in subcategories of the road system. Different categorization criteria are often used which include, whether the location in an urban or rural area, and the road class(e.g. freeway, arterial, collector, etc.).

In black spot programs, a hazardous location is defined as any location (section or intersection) that exhibits a higher potential for accidents than an established "norm". The higher potential for accidents can be expressed in terms of any accident measure such as accident frequency, rate, severity or a combination thereof. The following are a description of the measures most commonly used to identify hazardous locations.

2.2.1 The Frequency Measure

The accident frequency measure (AF) is defined as the number of accidents per location during a specific time period. If the observed AF meets or exceeds a predefined value, the location is considered hazardous. Several different section lengths and/or years of data are often used. The predefined frequency criterion usually varies by area type (urban/rural) or other variables such as highway class.

The use of an accident pin map has been one of the earliest methods of identifying hazardous locations. Each accident is represented by a pin on the map. Different pin colors and sizes can be used to indicate accident types and severity. From the map, locations that have clusters (high frequency) of accidents can be easily identified. Also, the use of two pin maps (one for the current year and the other for the preceding year) can be useful for comparison purposes. This process can be automated using geographical information system (GIS) software.

Proponents of using the frequency measure to identify hazardous locations argue that locations identified by this method have a high number of accidents and consequently have a higher potential for accident reduction. The problem with using the frequency method, however, is that it does not account for the effect of traffic exposure. For example, 10 accidents per km may be considered "high" for a section that carries 15,000 veh/day, and "low" for another section that carries 40,000 veh/day.

2.2.2 The Rate Measure

The accident rate measure (*AR*) is defined as accidents per million-vehicle-kilometers (mvk) for sections, and accidents per million-entering-vehicle (mev) for intersections:

Sections:

$$AR = \frac{N \times 10^6}{L \times AADT \times t \times 365} \tag{2.1}$$

Intersections:

$$AR = \frac{N \times 10^6}{AADT \times t \times 365} \tag{2.2}$$

where

N = Number of accidents observed during period t,

L = Segment length (km),

$AADT$ = Average Annual Daily Traffic volume (veh/day), counting all approaches in the case of intersection, and,

T = observation period (years).

The calculation of accident rates requires the availability of a traffic volume file. The volume file should be formatted by a compatible location reference method as the accident file. Locations that meet or exceed a predefined accident rate are then identified as hazardous.

The advantage of using accident rates is that it allows comparisons to be made between sites with similar characteristics but with different levels of exposure. However, although the use of AR addresses the exposure effect, it introduces another bias in the identification of hazardous locations when applied to lower volume roads.

For example, two accidents per year may be considered low from a frequency point of view. However, on a low volume road, it may result in a high accident rate, i.e., for 1 km section and 1 year period, 2 accidents will result in AR greater than 2.0 if the traffic volume is less than 2700 veh/day. Therefore, identifying hazardous locations based solely on accident rates can be misleading.

2.2.3 The Frequency Rate Method

To address the weakness of using either the rate or the frequency measures, several researchers such as Zegeer and Deen (1977) suggested using both AF and AR to identify hazardous locations. Usually, locations that meet the frequency criteria are first selected and then ranked using the rate criteria. However, some agencies use the rate to select locations and the frequency for ranking. Other agencies define a double criterion where a location must meet both predefined accident frequency and rate.

2.2.4 The Severity Method

The severity method uses the Accident Severity Index (ASI), defined as the weighted sum of fatal (F), injury (I), and property-damage-only (PDO) accidents.

$$ASI = 100\times F + 10\times I + PDO \tag{2.3}$$

Because all accidents are weighted against the PDO accident, the ASR is also known as Equivalent PDO ($EPDO$). For example, if the portion of F, I, and PDO accidents at some location is 1%, 33% and 66%, the ASR value would be equal to 4.96.

Various jurisdictions use different weights than the 100, 10, and 1 shown above. However, the ASR is an arbitrary ratio that relates the proportions of accidents with various severity levels at a given location and any convenient weighting will do.

Locations are ranked by their *EPDO* and those that meet or exceed a certain threshold value are selected. In some jurisdictions an *EPDO* rate is calculated by dividing the *EPDO* by the traffic volume to account for exposure.

2.2.5 Issues in Hazardous Locations Identification

2.2.5.1 Time Period and Section Length

Selecting the section length and the time period for which accident data are accumulated to calculate accident measures is a controversial issue. For example, as the length of road gets very small, the probability of zero or one accident tends toward unity. As the length of road gets very large, the effect of isolated hazards will be submerged and lost. Zegeer (1982) states that "accident rates in accidents per million vehicle miles become unstable and of questionable value for highway segments of short length (i.e. less than 0.3 miles) even when several years of accidents and volume data are used." Nicholson (1980) recommended the avoidance of sections shorter than one kilometer.

The choice of time is also controversial. The shorter the time period, the greater the probability of quickly detecting sudden changes in the accident occurrence. However, statistical reliability considerations indicate that a longer time period is required, while longer periods would prevent the quick detection of changes in accident rates. A time period of one to three years is commonly used in most road jurisdictions (Zegeer, 1982).

2.2.5.2 Observed and Expected Accidents

The accident frequency or rate at a particular location is a random variable whose true value can not be predicted with absolute accuracy. This causes the process of identifying hazardous locations based on the accident history to be subject to uncertainty. Therefore, accident measures should be calculated based on the *expected* number of accidents at a location. The problem, however, is that this expected number is generally not known, and statistical techniques should be utilized for its estimation.

2.2.6 Classical Statistical Techniques to Identify Hazardous Locations

High frequency of accidents may not necessarily mean that a particular location is truly hazardous. This high frequency may be solely caused by random variations of accident occurrence. An optimal identification technique would only identify locations that are truly hazardous and would not identify any non-hazardous locations. To address this concern, many classical statistical techniques have been developed and used to identify accident-prone locations based on historical accident data. Typically, a location will be identified as hazardous if its observed accident measure exceeds some critical level. The following is a discussion of some statistical techniques for identifying hazardous locations.

2.2.6.1 The Confidence Interval Technique

The simplest statistical technique to identify hazardous locations is the confidence interval technique that is based on the assumption that the observed accident frequency or rates are normally distributed. The technique involves calculating a critical threshold that is equal to the sample mean frequency or rate plus a multiple of the sample standard deviation. The multiple coefficients depend on the degree of confidence desired.

A location is considered hazardous if:

$$C_i > \mu + k\sigma \tag{2. 4}$$

where C_i is the accident frequency or rate at the location, μ is the mean frequency or rate of the population of similar locations, σ is the standard deviation of the population and k is obtained from the normal distribution function (k = 1.645 for 95% confidence level.) The reliability of this technique has been questioned since it is apparently very sensitive to the sample mean and standard deviation of the population accident frequencies or rates. Another problem with this technique is the normal distribution assumption that does not account for the special nature of accidents as rare and random events.

2.2.6.2 The Statistical Quality Control Technique

The rate quality control technique (Nordon 1956), which is based on statistical quality control procedures seems to be the most widely used statistical technique among highway agencies to identify hazardous locations. The technique defines a location as hazardous if the observed number of accidents exceeds a critical number or if the observed accident rate exceeds a critical accident rate. The main assumption of the technique is that the number of accidents occurring at a given location during a given time period can be approximated by the Poisson distribution. This assumption is widely accepted among safety researchers and has been investigated many times and turns out to be supported by a vast body of empirical evidence (Oppe 1982; 1992).

Based on the Poisson assumption, then we can write:

$$P(n) = \mathbf{e}^{-a}\,(a)^n\,/\,n! \tag{2.5}$$

where

$P(n)$ = probability that n accidents will occur at a given location during the given time period, and

a = expected number of accidents at the given location during the given time period.

Equation (5) can also be written as:

$$P(n) = \mathbf{e}^{-\lambda m}(\lambda m)^n / n! \tag{2.6}$$

where

λ = expected accident rate in accidents per million vehicle kilometers and

m = number of vehicle kilometers in million.

The value for a *and* λ are taken to be the average number/rate across all similar locations in a specified region. Based on Equation (2.6), an upper control limit U can be calculated such that:

$$\text{Probability } (X \geq U) = P \tag{2.7}$$

where

X = the observed number of accidents,

P = predefined probability limit.

The upper control limit (the critical limit) can then be calculated using a table of Poisson distribution. However, calculating the upper control limit from these tables involves double interpolation (for a and for X). Nordon (1956) obtained satisfactory approximations to determine the critical rate or number by using:

$$CN = a + k\sqrt{a} + 0.5 \tag{2.8}$$

$$CR = \lambda + k\sqrt{\lambda / m} + \frac{1}{2m} \tag{2.9}$$

where k is a constant related to the Level of Significance as follows:

Level of Significance	*K*
0.01%	3.719
0.05%	3.290
0.1%	3.090
0.5%	2.576
1%	2.326
5%	1.645

Locations that have a higher accident number or accident rate than the *CN* or the *CR* are considered to be accident-prone since their deviations from their expected means can not be reasonably attributed to the random fluctuation in accident occurrence.

Given the value of λ for a group of similar locations (e.g. signalized intersections), a graph can be plotted relating the critical accident number or rate to the traffic exposure at these locations. This graph can then be used to identify hazardous locations. For example, Figure 2-1 shows the critical accident rate curve for the urban signalized intersections category based on three years of accident data (1989-1991) from the B.C. Ministry of Transportation and Highways.

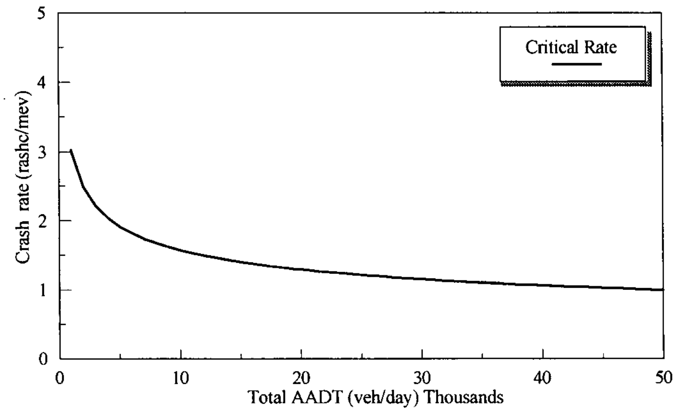

Figure 2-1 Critical Accident Rates for Urban Signalized Intersections in B.C.

2.2.7 Selection of the Identification Method

Most jurisdictions utilize more than one method to identify hazardous locations. It is common to use the rate quality control method in conjunction with the accident frequency method. Although the accident frequency method does not account for exposure, it is useful to exclude locations with low numbers of accidents before applying the rate quality control method and calculating critical rates. It is also important to use the accident severity method as a supplemental method (Zegeer and Deen, 1977) since the existence of severe accidents (injury and fatal) should justify a further analysis of locations than property damage only accidents.

A treble-measure criteria which combines the accident frequency, rate and severity and their critical values was developed by Abdelwahab and Sayed, (1993). The measure identifies a location as hazardous if:

$$((AR>R_c\, or\ ASI>S_c)\, and\ AF>F_c) \tag{2.10}$$

where R_c, S_c, and F_c are the critical values for AR, ASR, and AF respectively.

2.3 Bayesian Identification of Accident-prone Locations

The number of accidents at a location is a random variable that fluctuates around some unknown mean. This randomness is the reason that historical accident data at a location does not always accurately reflect its long-term accident characteristics. For example, a location that has low accident frequency during long periods of time may have had high accident rates during portions of this period and vice versa (Higle and Witkowski, 1988). The regression to the mean effect is also inherent in ratio data. Because of this phenomenon, a higher than normal number of accidents at a location may be followed by a lower than normal number during a similar succeeding period even if no changes are introduced on the sites (and vice-versa). One type of analysis suggested to account for these random variations is the Empirical Bayes approach (Higle and Witkowski, 1988; Brude 1988). The advantage of Empirical Bayes analysis is that it regards the accident count at any location as a random variable. The technique combines regional accident characteristics (for a group of similar locations) with the location-specific accident history to estimate the probability that the location is accident-prone.

The technique is based on Bayes' theorem which can be mathematically described as:

$$P(\phi|x) = \frac{P(x|\phi) \times P(\phi)}{\sum P(x|\phi) \times P(\phi)} \tag{2.11}$$

where ϕ is a parameter such as the number of accidents at a location, $P(\phi)$ is the prior distribution of ϕ, $P(x|\phi)$ represents the probability of making x observations for a specific value of ϕ (observation distribution), and $P(\phi|x)$ is the posterior distribution of ϕ which represents the resolution of the prior distribution given the observations.

Typically, the observation distribution will be assumed to be a Poisson or binomial distribution and the prior distribution will be a gamma or beta distribution. The main issue then is how to estimate the parameters of the prior distribution. In a pure Bayesian analysis, these parameters are usually assumed based on engineering judgment and past experience. While in the Empirical approach, the parameters are estimated using a sample of observations from population of similar locations (the same kind as the one being investigated).

Higle and Witkowski (Higle and Witkowski, 1988) have described an Empirical Bayes procedure to identify accident-prone locations. The procedure is based on the following two assumptions:

1) the actual number of accidents at any given location (N_i) follows a Poisson distribution such that at any given location, where the accident rate is known ($\tilde{\lambda}_i = \lambda$), and the expected value is given by λV_i, then the observation probability distribution is given by the following:

$$P[N_i = n | \tilde{\lambda}_i = \lambda, V_i] = \frac{(\lambda V_i)^n}{n!} e^{-\lambda V_i} \tag{2.12}$$

where: $\tilde{\lambda}_i$ = accident rate at location i (treated as a random variable).

V_I = number of vehicles passing through location i during specific time period.

2) the probability distribution of the regional accident rate (the prior distribution), $f_R(\lambda)$, follows a gamma distribution. Then the prior distribution can be given by the following:

$$f_R(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \tag{2.13}$$

where:

$f_R(\lambda)$ = probability density function associated with the accident rate across the region (the prior distribution),

α, β = parameters of the gamma distribution.

The parameters α and β of the prior distribution are estimated using the method of moments estimates (MME), where α and β are chosen so that the mean and variance associated with the gamma distribution are equal to the mean ($\bar{x}$) and variance (s^2) of the sample.

Consequently, in the MME method the α and β parameters can be selected according to the following:

$$\beta = \overline{x} / s^2 \tag{2.14}$$

$$\alpha = \beta \overline{x} \tag{2.15}$$

Morris (1988) showed that estimating α and β parameters based on equations (2.14) and (2.15) will lead to biased and inefficient estimation. He suggested using the following equation instead of equation (2.14):

$$\beta = \frac{V^* \, \overline{x}}{V^* \, s^2 - \overline{x}} \tag{2.16}$$

where:

V^* = the harmonic mean of $(V_i, \ldots\ldots)$

The next step is to combine the regional probability distribution (the prior distribution) with the location specific accident rate to obtain the location specific probability density function or the posterior distribution $f_i(\lambda | N_i, V_i)$. According to Berger (1985), the resulting probability distribution $f_i(\lambda | N_i, V_i)$ is a gamma distribution with the following parameters:

$$\alpha_i = \alpha + N_i \tag{2.17}$$

$$\beta_i = \beta + V_i \tag{2.18}$$

And thus, the probability density function associated with the accident rate at location *i* is given by:

$$f_i(\lambda|N_i,V_i)=\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}\lambda^{\alpha_i-1}e^{-\beta_i\lambda} \tag{2.19}$$

Finally, location i will be identified as accident-prone if there is a significant probability that the location's accident rate, $\overline{\lambda}_i$, exceeds the observed regional accident rate, X_R. Thus, location i is identified as accident-prone if:

$$P\{\overline{\lambda}_i > X_R|N_i,V_i\} > \delta \tag{2.20}$$

or equivalently if:

$$\left[1-\int_0^{X_R}\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}\lambda^{\alpha_i-1}e^{-\beta_i\lambda}\,d\lambda\right] > \delta \tag{2.21}$$

where δ represents the confidence level desired, such as 0.95, or 0.99 and $X_R=\frac{\sum_{i=1}^{m}N_i}{\sum_{i=1}^{m}V_i}$.

Sayed (1995) has added a small modification to Higle and Witkowski's method. Instead of using Equation 2.21 to calculate whether a location is accident-prone, a critical accident rate for each location, λ_{c_i}, which corresponds to a δ probability that the location accident rate, $\tilde{\lambda}_i$, exceeds the reference group accident rate, X_R, will be calculated by solving the following equation for λ_{c_i} (substituting $\alpha+\lambda_{c_i}V_i$ for α_i in Equation 2.21) :

$$\left[1-\int_0^{X_R}\frac{\beta_i^{(\alpha+\lambda_{c_i}V_i)}}{\Gamma(\alpha+\lambda_{c_i}V_i)}\lambda^{(\alpha+\lambda_{c_i}V_i-1)}e^{(-\beta_i\lambda)}\,d\lambda\right] = \delta \tag{2.22}$$

2.4 New Techniques to Identify Accident-prone Locations

In the traditional identification methods discussed previously, sites are identified for treatment based on their accident occurrence being greater than some norm. Accident correctability and targeted accident patterns are usually not considered. There are three components of the highway system: the *driver*, the *vehicle* and the *road environment.* Accidents occur due to a failure in any of these components or a combination of them as shown in Figure 2-2. The identification of potential treatment sites should be based on the factors which contributed to their accidents. For example, accidents which occur due to adverse road design should have greater influence for highway departments in identifying accident-prone locations than those which occur due to impaired driving. Similarly, if an accident could have been avoided by timely advice then technologies from a caring highway should be considered.

Sayed (1995) has presented two other methods to identifying potential treatment sites. The first is the "modified black spot program" which considers accident correctability and the degree of highway forgiveness. The second is the "countermeasure-based-program" which primarily considers locations that have well defined accident patterns that can be targeted by specific countermeasures. The following is a brief description of the two methods. However, they will be described in detail in Chapters Three, and Six.

2.4.1 The Modified Black Spot Program

The modified black spot program identifies potential treatment sites based on the assessment of accident contributing factors as described by Sayed et al (1995). The basic idea is to classify accidents according to their patterns and contributing factors into one or a combination of the three components of the highway system (the driver, the vehicle and the road). The method utilized fuzzy pattern classification. The advantage of fuzzy set theory is that a degree of membership in a set can be specified. This is very important in pattern recognition where the membership of an element in a certain group is usually not clear. The output from the classification process is three membership values for each accident, one value for each of the three highway system components. Each membership ranges from 0 to 1 and they sum to one. For example, a certain accident can have membership values of 0.6 in the driver component, 0.3 in the road environment component, and 0.1 in the vehicle component. These memberships will be used to modify the method of calculating accident measures.

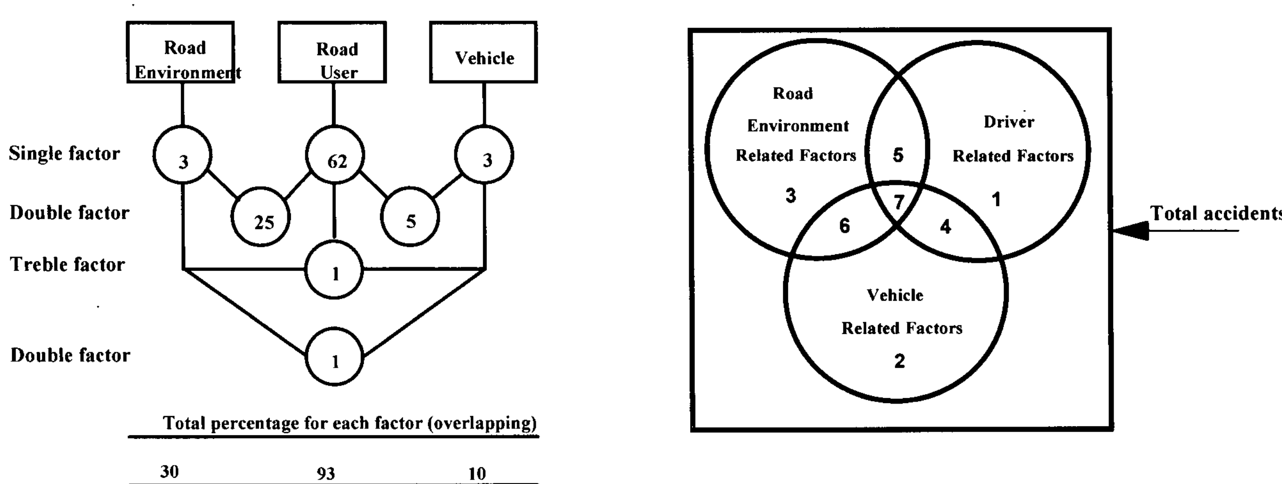

Figure 2-2 Accident Contributing Factors

An important application of the method is investigating the effects of the "caring highway" concept on accidents. For example, rumble strip with adequate area to provide recovery space for run-off the road vehicles can eliminate some of the single-vehicle off-road accidents, even though these crashes involved some driver error. As shown in Figure 2-2, highway agencies should devote their attention and spending on correcting accidents that belong to category 3. Accidents that belong to categories 5 and 7 are "somewhat correctable" by road improvements but perhaps are correctable using the concept of "caring highways". The degree of "caring" can be assessed by the degree (W_i) with which accidents belong to the driver group.

2.4.2 The Countermeasure-based Program

The countermeasure-based method first identifies the main accident patterns that can be targeted by specific countermeasures and then searching for locations which have over-representations of these patterns. In the countermeasure-based approach, a location is identified as accident-prone if it has over-representation of a particular accident type in the total number of accidents.

To illustrate, consider, for example, a signalized intersection with a total of 35 accidents in three years. Similar signalized intersections with the same operating environment (reference group) have an average of 50 accidents in three years. Traditional black spot programs as accident-prone will not identify this intersection. However, if it is known that 25 out of the 35 accidents involved left turn collisions, then implementing a single countermeasure (e.g., addition of a left-turn lane, phasing, etc.) can be very cost effective.

2.5 Accident Prediction Models

Developing accident prediction models, which provide reliable safety estimates of road segments and intersections can enhance the success of RSIPs. These safety estimates can be used in identifying accident-prone locations and evaluating the effectiveness of remedial measures. The models examine the traffic and road-related factors that appear to underlie the occurrence of accidents and explain, in statistical sense, the occurrence of accidents as a function of these factors.

When estimating the model parameters using regression, there are two main options: the conventional linear regression approach assuming a Normal distribution error structure and the generalized linear modeling (GLIM) approach assuming a non-Normal error structure (usually Poisson or negative binomial). Several researchers (e.g. Jovanis and Chang, 1986, Saccomanno and Buyco 1988, Miaou and Lum 1993) have shown that conventional linear regression models lack the distributional property to adequately describe random, discrete, non-negative, and typically sporadic events which are all characteristics of traffic accidents. Recognizing these limitations, it was decided to use the GLIM approach.

2.5.1 Generalized Linear Regression Models (GLIM)

Assuming that Y is a random variable which describes the number of accidents at an intersection (or segment) in a specific time period, and y is the observation of this variable during a period of time. The mean of Y is Λ which can also be regarded as a random variable. Then for $\Lambda = \lambda$, Y is Poisson distributed with parameter λ:

$$P(Y=y \mid \Lambda=\lambda)=\frac{\lambda^{y}e^{-\lambda}}{y!};\ E(Y \mid \Lambda=\lambda)=\lambda;\ \mathrm{Var}(Y \mid \Lambda=\lambda)=\lambda \tag{2.23}$$

Hauer et al. (1988) have shown that Λ follows a gamma distribution (with parameters κ and κ/μ), where κ is the shape parameter and μ is the mean of the distribution:

$$f_{\Lambda}(\lambda)=\frac{(\kappa/\mu)^{\kappa}\lambda^{\kappa-1}e^{-(\kappa/\mu)\lambda}}{\Gamma(\kappa)} \tag{2.24}$$

with an expected value and a variance of

$$E(\Lambda)=\mu;\ \mathrm{Var}(\Lambda)=\frac{\mu^{2}}{\kappa} \tag{2.25}$$

Kulmala (1995) has also shown that the point probability function of Y based on (2.24) and (2.25) is given by the negative binomial distribution:

$$P(Y=y)=\frac{\Gamma(\kappa+y)}{\Gamma(\kappa)y!}\left(\frac{\kappa}{\kappa+\mu}\right)^{\kappa}\left(\frac{\mu}{\kappa+\mu}\right)^{y} \tag{2.26}$$

with an expected value and variance of:

$$E(Y)=\mu;\ \mathrm{Var}(\mathrm{Y})=\mu+\frac{\mu^{2}}{\kappa} \tag{2.27}$$

As shown in equation (2.27), the variance of the observed number of accidents is generally larger than its expected value. The only exception is when $\kappa \to \infty$, where the distribution of Λ is concentrated at a point and the negative binomial distribution is identical to the Poisson distribution (Kulmala, 1995).

As described earlier, for the GLIM approach, the error structure is usually assumed to be Poisson or negative binomial. The main advantage of the Poisson error structure is the simplicity of the calculations (the mean and the variance are equal). However, this advantage is also a limitation. It has been shown (Kulmala and Roine 1988, and Kulmala, 1995) that most accident data will likely be overdispersed (the variance is greater than the mean) which indicates that the negative binomial distribution is the more realistic assumption. The only problem with the negative binomial assumption is in the determination of the value of κ. Several methods to calculate κ have been proposed (Hauer, 1988; Kulmala, 1995). The most of these methods is the maximum likelihood estimate (NAG, 1996).

2.6 Conclusion

Highway Safety Improvement programs involve three main processes: the identification of accident prone locations; the analysis of the causes and contributing factors of the safety problem and the selection of appropriate countermeasures. The identification process is undertaken using different statistical techniques (classical and Bayesian) to identify locations which have higher potential for accidents than an established "norm". This traditional criterion provides no consideration of the factors contributing to the accidents and provides little or no insight into whether the safety at thes locations can be enhanced by road improvements. Accident contributing factors and causes are not considered in the identification process. These traditional methods may result in the identification of locations that are not truly hazardous from a road safety authority perspective, and consequently, may lead to misapplication of safety improvement funds.

Sayed (1995) proposed two alternative approaches to identify hazardous locations. The first is the "modified black spot program" which is implemented to identify hazardous locations on the basis of their contributing factors and causes. This approach uses a fuzzy pattern recognition algorithm to classify accidents, according to their contributing factors and causes, into one or a combination of the three main highway system elements: the driver, the vehicle and the road. Statistical techniques such as the Empirical Bayes approach are then used to identify hazardous locations from a road, driver or vehicle point of view. The second approach is "the countermeasure-based program" which argues that a location with a given number of accidents with well-defined patterns can be treated more effectively than a location with a larger number of accidents with poorly defined patterns. Traditional approaches start with a problem (high accident occurrence) and attempt to find solutions (countermeasures). The countermeasure-based program reverses the traditional process by first identifying main accident patterns that can be targeted by specific countermeasures and then searching for locations which have over-representation of these patterns. Finally, accident prediction models were developed to provide reliable safety estimates of road segments and intersections.

CHAPTER 3

ACCIDENT CLASSIFICATION USING ARTIFICIAL NEURAL NETWORKS

3.1 Background: Accident Classification

As described earlier, a hazardous location is defined as any location that exhibits a higher potential for accidents than an established "norm". The higher potential for accidents is usually expressed in terms of any accident measure such as accident frequency, rate, severity or a combination thereof. In traditional Black Spot programs, these accident measures are calculated from the total population of accidents. No consideration is given to whether these accidents were caused by road deficiencies or can be treated by road improvements. Results of these traditional improvement programs can, therefore, be disappointing and may lead to misallocation of safety funds. To overcome these deficiencies, an alternative approach was suggested by Sayed (1995). The approach, labeled modified black spot program, identifies accident-prone locations based on an assessment of accident contributing factors. The basic idea is to classify accidents according to their patterns and causes into one or a combination of the three road system components: the driver, the vehicle and the road environment as shown in Figure 2-2. For example, to identify road related accident-prone locations, more emphasis should be given to accidents which belong to category 3, followed by accidents which belong to categories 5, 6 and 7.

The classification process was carried out using a fuzzy K-nearest neighbors algorithm. The classification results were compared to those of two experts and were shown to produce satisfactory results (Sayed, 1995). The present thesis introduces further refinement to the classification process: using artificial neural networks. A comparative evaluation of the fuzzy k-nearest neighbors and a multilayer feed-forward back-propagation neural network classifier is carried out. The theoretical backgrounds of both classifiers are presented and their relative advantages are discussed.

3.2 Pattern Recognition and Fuzzy Sets

Pattern Recognition can be generally defined as the allocation of objects to classes so that individual objects in one class are as similar as possible to each other and as different as possible to objects in other classes. The classification process can be performed with or without labeled data (data of known classification). If labeled data is available the process is usually referred to as supervised learning. In this process, the algorithm is given a set of patterns with known classification (i.e. labels) and is required to classify an unknown object based on the information acquired from the labeled data. In the case of unsupervised learning (also called cluster analysis), prior information about classes are not available and the clustering has to be based on an established similarity criteria (Bezdek and Pal, 1992). The results of this process are greatly influenced by the choice of both the number of clusters (c) and the similarity criterion.

The theory of fuzzy sets was introduced by Zadeh (1965). It is based on the simple idea of introducing a degree of belonging of an element to a specific set. The theory deals with a subset A of the universe of discourse U, where the transition between membership and non-membership is gradual rather than sharp. A fuzzy subset A of a universe of discourse U is characterized by a membership function $u_A(x)$, which associates each element $x \in U$ a membership $u_A(x)$ in the interval [0, 1] that represents the grade of membership in A.

The relationship between fuzzy sets and classification is based on the fact that most real-world classes are fuzzy in nature. Thus, given an object p and a cluster C, the basic question is not whether p is a member of C, but the degree to which p belongs to C, i.e. grade of membership of p in C (Kandel, 1982). In conventional classification techniques an object is assigned to one and only one of the classes, with a degree of membership equal to one, assuming well defined boundaries between classes, while in fuzzy pattern recognition an observation can belong to more than one class with different degrees. This is very important in pattern recognition where the membership of an element in a certain group is usually not clear.

3.2.1 The Fuzzy K-Nearest Neighbor Algorithm

The fuzzy K-Nearest Neighbor algorithm is considered one of the most accurate algorithms in pattern recognition (Keller et al., 1985; Bezdek et al., 1986). The classical (crisp) K-NN algorithm classification rule assigns an input sample vector y, which is of unknown classification to the class which is represented by a majority amongst its K-nearest neighbors (Duda and Hart, 1973). The K-nearest neighbors are chosen from a labeled data sample (data of known classification). The fuzzy K-NN algorithm assigns class membership to a sample observation based on the observation distance from its *K*-nearest neighbors and their memberships (Keller et al., 1985).

If $=\{x_1, x_2, ..., x_n\}$ is the set of n labeled samples and u_{ij} is the membership of the *j*th labeled data in the *i*th class, then the fuzzy *K*-NN algorithm is simply described as follows (Keller et al., 1985):

BEGIN

Input *y*, of unknown classification.

Set $K, 1 \leq K \leq n$.

Initialize $i = 1$.

DO UNTIL (*K*-nearest neighbors found)

Compute distance from *y* to x_i

IF $(i \leq K)$ THEN

Include x_i in the set of *K*-nearest neighbors

ELSE IF (x_i is closer to *y* than any previous nearest neighbor) THEN

Delete farthest in the set of *K*-nearest neighbors.

Include x_i in the set of *K*-nearest neighbors

END IF

Increment i.

END DO UNTIL

Initialize $i = 1$

DO UNTIL (y assigned membership in all classes)

$$\text{Compute } u_i(y) = \frac{\sum_{j=1}^{K} u_{ij}(1/\|y - x_j\|^{2/(m-1)})}{\sum_{j=1}^{K}(1/\|y - x_j\|^{2/(m-1)})} \tag{3.1}$$

Increment i.

END DO UNTIL

END

As shown in equation (3.1), the assigned memberships of observations are influenced by the class memberships of the K-nearest neighbors u_{ij} and the inverse of the distance to the K-nearest neighbors. The best value of the integer K is usually data dependent. For the results shown in this paper, a K value of 10 was found to give the best results. The memberships of the labeled sample can be assigned in several ways such as using fuzzy cluster analysis or based on expert opinions. The distance between observations can be represented by any distance measure ($\|\bullet\|$) such as the Euclidean distance, defined as (Bezdek, 1981):

$$d_{yx_i} = \sum_{v=1}^{p}(y_v - x_{iv})^2 = (y - x_i)'(y - x_i) \tag{3.2}$$

where p is the number of variables for observation i. With this distance, the variables are given equal weights. Another form of distance measure is the Mahalanobis' distance in which the correlation between variables are taken into account (Bezdek, 1981):

$$d_{yx_i}^2 = (y - x_i)' \Sigma^{-1} (y - x_i) \tag{3.3}$$

where Σ is the sample covariance matrix of x. The Euclidean distance is usually used when the variables are statistically independent, while the Mahalanobis' distance mitigates the effects of statistical dependence between pairs of variables. The variable m in equation (3.1) defines how heavily the distance is weighted when calculating each neighbor's contribution to the membership value (Keller et al., 1985). When m=2, the contribution of neighbors is weighted by the reciprocal of its distance from the point being classified. As m increases the contribution of each neighbors is more evenly weighted. Usually, practitioners use $m = 2$.

3.3 Artificial Neural Networks

An Artificial Neural Network (ANN) attempts to mimic, in a very simplified way, the human mental neural structure and functions (Hsieh, 1993). It can be characterized as a massively parallel interconnection of simple neurons that function as a collective system. The network topology consists of a set of nodes (neurons) connected by links and is usually organized in a number of layers. Each node in a layer receives and processes weighted input from nodes in the previous layer and transmits its output to nodes in the following layer through links. Each link is assigned a weight that is a numerical estimate of the connection strength. The weighted summation of inputs to a node is converted to an output according to a transfer function (typically a sigmoid function). Most ANNs have three layers or more: an input layer, which is used to present data to the network; an output layer, which is used to produce an appropriate response to the given input; and one or more intermediate layers, which are used to act as a collection of feature detectors.

The ability of a neural network to process information is obtained through a learning process which is the adaptation of link weights so that the network can produce an appropriate output. In general, the learning process of an ANN will reward a correct response of the system to input by increasing the strength of the current matrix of nodal weights. Therefore, the likelihood of producing similar output when the same inputs are entered in the future will increase. An incorrect response from the system is discouraged by adjusting the nodal weights so that the system will respond differently when it encounters similar inputs in the future (Hsieh, 1993). The learning process may be supervised or unsupervised based on the availability of target output. In the supervised learning inputs proceed through the network and produce an output. The difference between this output and the target output represents an error that is then propagated back through the network to "train" it. In unsupervised learning, the network automatically detects important features and organizes the input data into classes based on these features. More information about neural networks can be found in Lawrence (1991).

There are several neural network models which can be used in pattern recognition (both supervised and unsupervised). For supervised pattern recognition, the most commonly used ANN is the feed-forward network trained using the back-propagation algorithm (Rumelhart and McClelland, 1986; Jones and Hoskins, 1987) which is adopted in the present study. The back-propagation algorithm can be described in three equations. First, weight connections are changed in each learning step (k) with:

$$\Delta w_{ij}^{[s]}{}_{(k)} = \eta(t) \, . \, \delta_{pj}^{[s]} \, . \, x_i^{[s-1]} + m \, . \, \Delta w_{ij}^{[s]}{}_{(k-1)} \tag{3.4}$$

Second, for output nodes it holds that:

$$\delta_{pj}^{[o]} = (d_j - o_j) \, . \, f_j'(I_j^{[s]}) \tag{3.5}$$

and third, for the remaining nodes it holds that:

$$\delta_{pj}^{[s]} = f_j'(I_j^{[s]}) \, . \sum_k \delta_{pk}^{[s+1]} \, . \, w_{jk}^{[s+1]} \tag{3.6}$$

where

$x_j^{[s]}$ = actual output of node j in layer s

$w_{ij}^{[s]}$ = weight of the connection between node i *at layer* (s-1) and node j at layer (s)

$\delta_{pj}^{[s]}$ = measure for the actual error of node j

$I_j^{[s]}$ = weighted sum of the inputs of node j in layer s

$\eta(t)$ = time dependent learning rate

$f()$ = transfer function

m = momentum factor (between 0 and 1)

d_j, o_j = desired and actual activity of node j (for output nodes only)

3.4 The Accident Data

The data for this investigation came from the provincial accident database files consisting of all police reported accidents in the Province of British Columbia. The data sample consisted of 994 accidents, each assigned a membership into the three classes defined as: the driver, the vehicle and the road. The assignment of memberships was made by two safety experts from the British Columbia Ministry of Transportation and Highways, and was made based on an examination of the accident records and the apparent contributing factors associated with each accident. During the selection of the labeled data, every effort was made to include as many different accident types as possible. Each input vector included 21 variables. The first 15 variables and their levels are described in Table 3-1. In addition to the variables listed in the figure, another six variables describing accident contributing factors as assigned by the police officers investigating the accident were also included. The first three contributing factors were associated with the first vehicle in the crash and the other three with the second vehicle (for multi-vehicle accidents). The contributing factors range from driver related, such as alcohol involvement and driving without due care, to vehicle related such as engine or brake failure, to road related, such as road obstructions and pavement surface deficiencies. These contributing factors are listed in Table 3-2.

Table 3-1 Selected Variables and Their Levels

Variable	Levels	Variable	Levels
1. Degree of curvature	1. Straight 2. Single Curve 3. Sharp Curve	9. Accident location	1. At intersection 2. Not at intersection
2. Road grade	1. Flat 2. Some grade 3. Steep grade 4. Hillcrest/Sag	10. Accident type	1. Single vehicle-fixed object 2. Single vehicle-other 3. Multiple vehicle-head on 4. Multiple vehicle-side/angle 5. Multiple vehicle-rear-end 6. Pedestrian 7. Animal
3. Speed limit	1. 50-60 kph 2. 70-80 kph 3. 90-110 kph	11. Severity	1. Property damage only 2. Injury 3. Fatal
4. Surface condition	1. Dry 2. Wet 3. Ice/Snow	12. Traffic control device	1. Exist 2. None
5. Weather condition	1. Clear/Cloudy 2. Raining 3. Smog/Fog 4. Ice/Snow	13. Use of a restraint device	1. Device used 2. Vehicle equipped but device not used 3. Vehicle not equipped
6. Lighting conditions	1. Day light 2. Dark/Full illumination 3. Dark/Some illumination 4. Dark/No illumination	14. Volume/capacity ratio	- Value
7. Land use	1. Undeveloped/Agriculture Area 2. Rural residential area 3. Urban residential 4. Central Business District	15. Vehicle type	1. Passenger cars only 2. At least one van or pickup 3. At least one truck or bus
8. Accident time	1. Non-rush hour 2. Rush hour		

Table 3-2 Apparent Contributing Factors

1. Alcohol involvement	20. Cutting in	63. headlights defective
2. Driver inexperience	21. Driving on wrong side of road	64. Turn Signal defective
3. Drugs	22. Improper turning	65. Oversize vehicle
4. Extreme fatigue	23. Failing to yield right of way	66. Steering failure
5. Fell asleep	24. Ignoring traffic control device	67. Tires failure inadequate
6. Illness	25. Pedestrian error confusion	68. Two hitch failure
7. Sudden loss of consciousness	40. Obstruction on Road	69. Driverless vehicle
8. Pre-existing physical disability	41. pavement surface defective	70. Windshield defective
9. Prescribed medication	42. Road maintenance/construction	71. Engine failure
10. Attempted suicide	43. Sign obstruction	72. Suspension defect
11. Driving without due care	44. Insufficient traffic control	73. restraint system
12. Failing to signal	45. Road/intersection design	74. Insecure load
13. Ignoring officer	46. Roadside/hazard	75. Dangerous good
14. Previous traffic accident	47. Wild animal	76. Vehicle modification
15. Following too closely	48. Weather	77. Glare artificial
16. Improper passing	49. Visibility impaired	78. Glare sunlight
17. Unsafe speed	60. Accelerator defective	79. Domestic animal
18. Avoiding vehicle/ pedestrian/ cycle	61. Brakes defective	99. Others
19. Backing unsafely	62. Brake lights out	

The data was standardized using the following function (Romesburg , 1984):

$$Z_{ij} = \frac{x_{ij} - c_{\min j}}{c_{\max j} - c_{\min j}} \tag{3.7}$$

where $c_{\max j}$ and $c_{\min j}$ are the maximum and minimum of the *j*th variable in all observations. The main reason for standardizing the data matrix is that the variables are usually measured in different units. By standardizing the variables and recasting them in dimensionless units, the arbitrary effects of similarities between objects are removed.

3.5 The Neural Network Model

In general, the development process of a feed-forward back-propagation neural network involves the following steps (Lawrence, 1991):

1. <u>Defining the network architecture (number of layers and nodes in each layer)</u>. The input layer for the network consisted of 21 nodes representing the input variables with their levels as described in Table 3-1 (15 input variables plus 6 input variables representing contributing factors). The output Layer had three nodes representing the membership of an accident into the three highway system components: the road, the vehicle and the road environment. After several trials, it was found that one hidden layer with five nodes gives the best results. (Figure 3-1)

2. <u>Training the network</u>. Two-thirds of the labeled data (600 input vectors with their memberships) was presented to the network. After experimenting with different transfer functions, learning and momentum rates, the following parameters were

selected: a sigmoid transfer function, a learning rate of 0.2 and a momentum rate of 0.8. The threshold error was set at 0.05. The network training stopped after approximately 3400 epochs (an epoch is one complete pass through a set of input and target patterns while training the network). The error graph is shown in Figure 3-2.

3. Testing the network. The trained network was used to run a test data (300 new, untrained input vectors). The outcomes (three memberships) were compared with the memberships assigned by the experts. The three memberships produced by the network did not necessarily add to one as in the case of the experts and the fuzzy K-NN algorithm. Therefore the results were scaled to add to one.

A misclassification measure between the network output and the expert's was defined as:

$$E = (\sum_{i=1}^{c} (u_{ik} - u'_{ik})^2)^{0.5} \tag{3.8}$$

where:

E = a measure of classification error, c = the number of classes (3 in this case), u_{ik} = is the membership value of the kth observation in the ith class as assigned by the experts , and u'_{ik} = the membership value of the kth observation in the ith class as estimated by the network.

A value of E equal to zero indicates identical experts and network membership values, while a value of 1.41 indicates a complete disagreement.

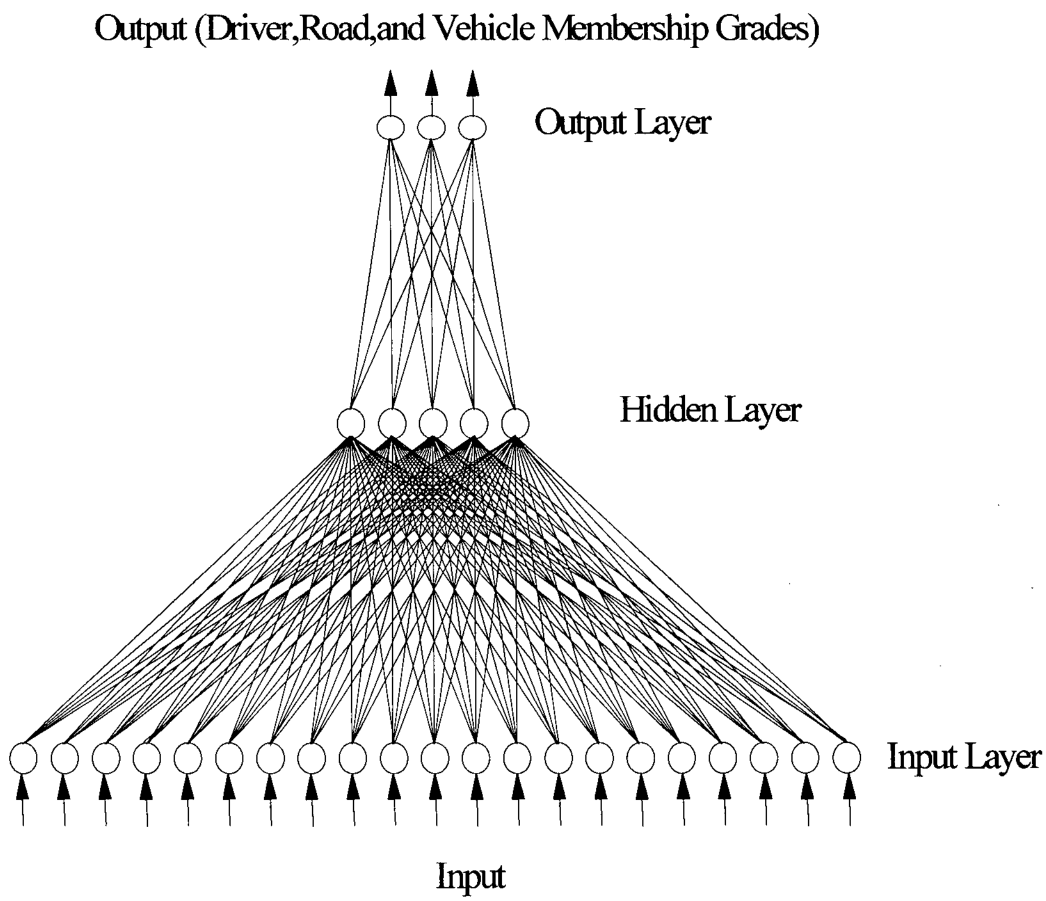

Figure 3-1 Feed-forward Back-propagation Neural Network Accident Classification Model

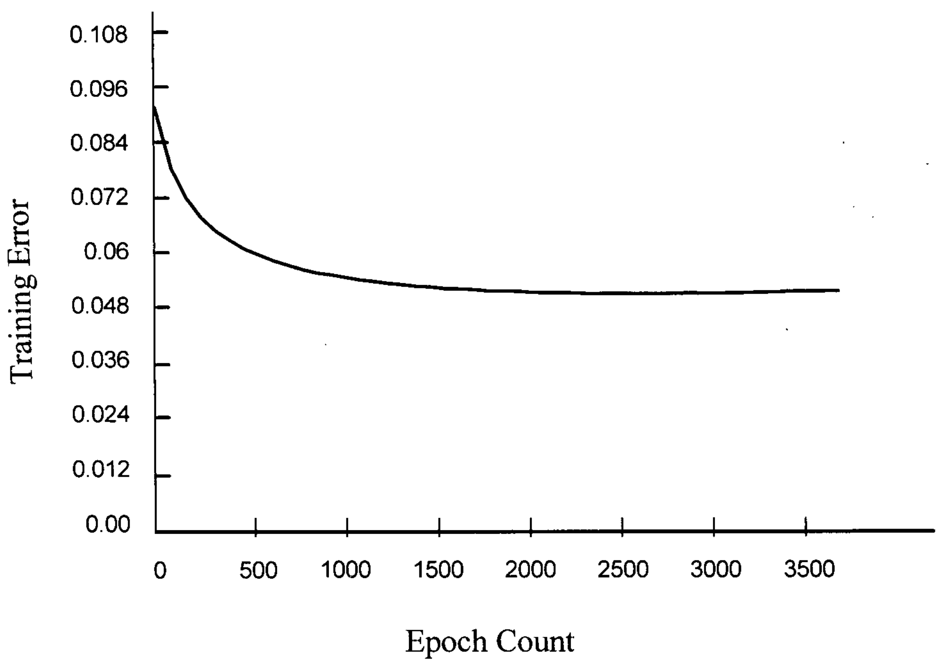

Figure 3-2 The Training Error Graph

3.5.1 Applications and Results

To allow for comparison between the Fuzzy K-NN and the ANN methods, results are presented to show the difference in the magnitude of classification error compared to expert classification. The classification error was calculated by the Euclidean distance formula of equation (3.8). Table 3-3 shows the results of the comparison. In general, the ANN classifier performed slightly better and produced more consistent results than the fuzzy K-NN classifier. The average classification error was 0.107 in the case of the ANN classifier compared to 0.142 in the case of the fuzzy K-NN classifier. The standard deviation was 0.165 in the case of the ANN classifier compared to 0.224 for the fuzzy K-NN classifier, which indicates higher consistency of the ANN classification. The 25, 50, and 75 percentile of error were smaller in the case of ANN classifier compared to the fuzzy K-NN classifier.

To further check the consistency of the classifier results, the kappa statistics (Cohen 1960; Fleiss 1971) was employed. Each accident was classified into one of the seven categories shown in Figure 2-2, using the memberships assigned by the experts and the two classifiers. An α-cut operator of 0.15 was applied to the membership values produced by the classifiers (observations with memberships less than 0.15 in any class were assigned zero membership in that class). The α-cut value of 0.15 was selected based on the minimum membership value (greater than zero) assigned by the experts to any accident in the three classes. Table 3-4 shows the agreement and disagreement between the expert classification and the ANN classifier results.

Table 3-3 Distribution of Classification Error for ANN and Fuzzy K-NN Classifiers

Classifier	Average classification error	Standard deviation of the classification error	25 percentile classification error	50 percentile classification error	75 percentile classification error
ANN	0.107	0.165	0.024	0.048	0.120
Fuzzy K-NN	0.142	0.224	0.037	0.079	0.145

Table 3-4 Agreement Between the Experts Classification and the ANN Classifier Results

	ANN Classifier							
Experts	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7	Total
Category 1	168	0	0	1	14	0	0	183
Category 2	0	7	0	2	0	1	0	10
Category 3	0	0	2	0	4	0	0	6
Category 4	7	1	0	2	0	0	0	10
Category 5	32	0	0	0	48	0	2	82
Category 6	0	0	0	0	0	1	0	1
Category 7	0	0	0	0	1	0	7	8
Total	207	8	2	5	67	2	9	300

As shown in Table 3-4, the number of accidents assigned by experts and ANN classifier to category 1 is 168, the number of accidents assigned by the experts to category 1 and by the ANN classifier to category 5 is 14 and so on. The *kappa* statistic is defined as:

$$kappa = \frac{\overline{P} - \overline{P}_e}{1 - \overline{P}_e} \tag{3.9}$$

where:

$\overline{P}$ is the overall percent agreement;

$\overline{P}_e$ is the overall percent agreement expected by chance.

A positive value for *kappa* indicates agreement, a value of zero indicates an agreement that can be expected by chance, and a negative value indicates disagreement. For the data in Table 3-4, $\overline{P}$ is calculated by the sum along the diagonal divided by the total number of cases. Thus, $\overline{P}$ =235/300=0.783. The overall percent agreement expected by chance may be calculated from the percentage of assignment of each expert to the seven categories, i.e.

$$\begin{aligned}\overline{P}_e &= \frac{207}{300}\times\frac{183}{300}+\frac{8}{300}\times\frac{10}{300}+\frac{2}{300}\times\frac{6}{300}+\frac{5}{300}\times\frac{10}{300}+\frac{67}{300}\times\frac{82}{300}+\frac{2}{300}\times\frac{1}{300}+\frac{9}{300}\times\frac{8}{300}\\ &= 0.484\end{aligned} \tag{3.10}$$

Therefore, using Equation 3.9, *kappa* = 0.58. The variance of *kappa* is calculated using (Fleiss, 1971):

$$\mathrm{Var}(kappa) = \frac{1}{N} \times \frac{\sum_j p_j^2 - (\sum_j p_j^2)^2}{(1 - \sum_j p_j^2)^2} \tag{3.11}$$

where:

N is the total number of cases;

$j = 1, \ldots, k$ is the seven categories of classification;

p_j is the proportion of all assignments of the *j*th category

For example, $p_1 = 0.5(207/300 + 183/300) = 0.65$. And therefore, using Equation 3.11, Var(*kappa*) = 0.003. Under the hypothesis of no agreement beyond chance and using the central limit theorem, the value $kappa / \sqrt{\mathrm{Var}(kappa)}$ may be approximately distributed as a standard normal variant (Fleiss, 1971). Therefore, $kappa / \sqrt{\mathrm{Var}(kappa)} = .552 / \sqrt{.003} = 10.31$, which greatly exceeds the critical Z value of 2.32 at the 99% significance level, indicating strong agreement between the experts and the ANN classifier results. A similar comparison between the experts results and the fuzzy K-NN classifier results produced a value of 0.54 for *kappa*, and a value of 9.80 for $kappa / \sqrt{\mathrm{Var}(kappa)}$ which also indicate a strong agreement with the experts results.

3.5.2 Discussion

As shown from the comparison of results, the ANN classifier overall performance is slightly better than the K-NN classifier. The results of both classifiers, however, are considered good. One of the most important attributes of both classifiers is their ability to spot patterns in data that classical pattern recognition systems may not be able to detect. Therefore, both classifiers are recognized as ideal tools for dealing with environments that are highly unstructured and that may involve incomplete or noisy data (such as the case of traffic accident data).

Table 3-5 provides a comparison of the relative advantages of both classifiers. The results have indicated that both classifiers are efficient and produce good results. However, the ANN classifier produces slightly better results and has a greater computational efficiency as compared to the fuzzy K-NN classifier.

Table 3-5 Comparison of ANN and Fuzzy K-NN Classifiers

Comparison Criteria	ANN Classifier	Fuzzy K-NN Classifier
Selection of parameters	To achieve good classification results, several parameters such as the learning rate, the momentum rate and the number of nodes in hidden layers have to be carefully selected.	Only one parameter has to be carefully selected (the parameter *K*).
Optimality (with regard to classification error)	Higher than the fuzzy K-NN classifier.	Slightly lower than the ANN classifier.
Computational Efficiency	Very efficient. Once the network is trained, the classification is completed very fast.	Less efficient. All the labeled data has to be permanently stored. So that for each input classification, a comparison of this input to all of the labeled data is carried out.
Fault tolerance (ability to deal with incomplete or corrupt data)	High	High
Ability to deal with highly unstructured problems	High	High
Adaptivity (the ability to adjust when given new data)	Adaptive through retraining	Highly adaptive. The K-NN will adapt to the new data as soon as it is stored. (no training is required)

3.6 Conclusion

Traditional methods of identifying accident prone locations, or black spots do not account for accident contributing factors or causes. This leads to the identification of locations which may not be truly hazardous or accident prone from a road environment perspective. This Chapter described an alternative approach suggested by Sayed (1995). The approach labeled modified black spot program, identifies accident prone locations based on accident contributing factors. Two classification processes (Fuzzy K-nearest neighbors algorithm, Artificial Neural Networks) was carried out and the results were compared. The results showed that Artificial Neural Networks have inherent advantages that make them slightly superior to fuzzy K-nearest neighbors algorithm classification techniques. These advantages related to the computational efficiency and, for the investigation presented in this Chapter, a higher classification accuracy of ANNs compared with fuzzy classifiers. These advantages increase their applicability to the field of accident data analysis. Better accident data classification allows for more accurate detection of safety deficiencies, which, in turn, leads to more effective allocation of safety funds.

CHAPTER 4

ACCIDENT CLASSIFICATION USING NEURO-FUZZY MODELS

4.1 Background

In Chapter 3, it was shown that Artificial Neural Networks (ANNs) could be successfully used for accident classification. ANNs produced higher predictive accuracy of accident classification compared with the K-NN algorithm. In addition, ANNs were believed to have higher computational efficiency than the fuzzy K-NN algorithm.

However, despite the flexibility and the good performance of ANNs in modeling nonlinear relationships, they have often been criticized for acting as "black boxes". The knowledge contained in the ANN model is kept in the form of a weight matrix that is hard to interpret and can be misleading at times.

The efficiency of ANN models is also another point of concern. Since it is not always possible to determine the significance of the input variables in advance, any potential candidate may be included in the model. It is therefore important to identify and exclude those input variables that do not have a significant contribution. This would lead to a more efficient model. This is particularly important in cases where a large number of potential input variables exist but only a subset of them would actually affect the output. Conventional techniques of variables selection such as stepwise modeling (Masters, 1993) may not produce optimal results (Masters, 1993; Abdelwahab and Sayed, 1998).

It is therefore desired to develop efficient models that are able to correctly simulate the input/output relationship using a minimal set of input variables while providing insight into the decision-making process. The models should also adopt approaches that use as few constraining assumptions as possible so that they can be successfully used for a wide variety of applications. It is also beneficial to have the ability to insert any available knowledge or expertise into the model if necessary.

In this Chapter, the merits of adopting the neuro-fuzzy approach will be discussed. It will be shown that the proposed approach has the potential to provide similar or better accuracy compared with ANNs while overcoming their shortcomings.

4.2 The Neuro-Fuzzy Approach

4.2.1 Fuzzy Logic and Fuzzy Sets

Fuzzy logic is another area of AI that has been successfully applied to an increasing number of applications. The concept of fuzzy logic was introduced by Zadeh (1965). Basically, it is a superset of conventional (Boolean) logic that has been extended to handle imprecise data and the concept of partial truth. In fuzzy logic, variables are "fuzzified" through use of so called "membership functions" that define the membership degree to fuzzy sets, or as they are often called "linguistic variables" (e.g., inexpensive). A fuzzy subset A of a universe of discourse U is characterized by a membership function $u_A(x)$, which associates each element $x \in U$ a membership $u_A(x)$ in the interval $[0,1]$ that represents the grade of membership in A.

To each variable, a small number of fuzzy sets (linguistic variables) are assigned whose membership functions overlap and cover the necessary range of variation for that variable. Consequently, statements that contain linguistic variables (e.g., Transit fare is inexpensive) may take "truth values" of other than "true" or "false", represented by a number in the range [0,1]. This provides an opportunity to define logical operators such as "AND" or "OR".

One way of representing fuzzy sets is by using B-spline functions. B-spline functions (Figure 4-1) are piecewise polynomials of order k which have been widely used in surface fitting applications. The order of the functions determines their smoothness. They can be used to implement crisp fuzzy sets (k=1) or the standard triangular fuzzy membership functions (k=2) or other smoother representations. A univariate B-spline function of order k is non-zero only over k intervals, which are generated by a (k+1), knots. A Multivariate B-spline function can be formed by taking the tensor product of n univariate functions (Brown and Harris, 1995).

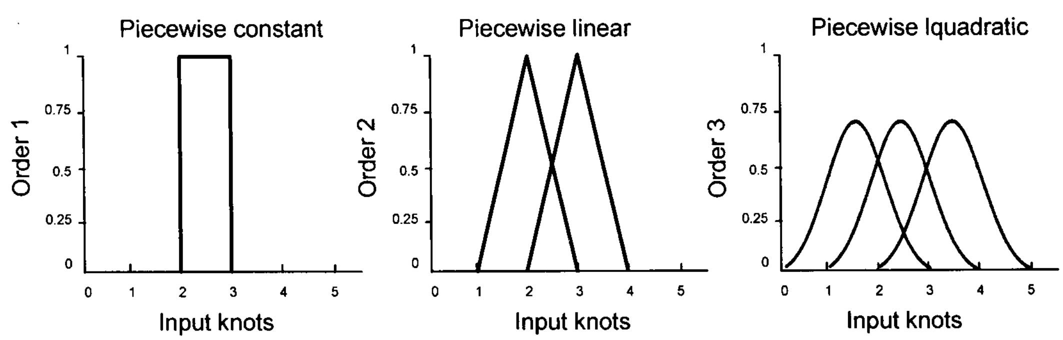

Figure 4-1 B-spline Fuzzy Membership Functions

4.2.2 Fuzzy Systems

Fuzzy systems use fuzzy implications or "IF THEN rules" to process information. One such rule may look like this:

```
IF "Transit Fare" is INEXPENSIVE and "Transit Waiting Period" is
Short
THEN "Preferred Mode of Transportation" is Transit
```

The part of the rule between the "if" and "then" is the rule's premise or antecedent. This is a fuzzy logic expression that describes to what degree the rule is applicable. The part of the rule following the "then" is the rule's conclusion or consequent. The consequent part of the rules may result in a fuzzy or crisp variable. In cases where the consequent is in fuzzy form but a crisp number is desired, an inverse operation called "defuzzification" needs to be done. Fuzzy systems are defined by a number of fuzzy rules, a number of membership functions, and mechanisms to apply logical operators. There are numerous successful applications of fuzzy systems in control and modeling. They are most suitable for situations where an exact model of a process is either impractical or very costly to build but an imprecise model based on the existing human expertise can do the job. In such cases, fuzzy systems are considered the best alternative, though they do not perform optimally.

4.3 Neuro-Fuzzy Models

The knowledge contained in fuzzy systems are transparent to the user but can not be acquired directly from data. ANNs on the other hand, have the ability to learn the knowledge from a set of data, but the knowledge gained is hidden from the user. The concept of neuro-fuzzy systems has emerged in recent years as researchers have tried to combine the transparent, linguistic representation of a fuzzy system with the learning ability of an ANN (Brown & Harris, 1994). A neuro-fuzzy system uses an artificial neural network learning algorithm to determine its parameters (i.e. fuzzy sets and fuzzy rules) by processing data samples. Therefore it can be trained to perform an input/output mapping, just as with an ANN, but with the additional benefit of being able to provide the set of rules on which the model is based. This gives further insight into the process being modeled.

Several merger types of ANNs and fuzzy systems have been reported in the literature. They include various representations and architectures and therefore are suitable for different applications. Methods proposed by Kosko (1992) and Jang (1993) are among many variations that combine neural networks and fuzzy system. While the former uses competitive networks to generate rules for fuzzy systems, the latter proposes a hybrid back-propagation/least square learning method to tune the parameters of a so called adaptive network-based fuzzy inference system (ANFIS). However, these methods suffer from what is often referred to as the "curse of dimensionality" (Brown and Harris, 1994).

To illustrate this "curse", one may consider a fuzzy system with N input variables each of which having M membership functions. In such a system, as many as M^N combinations (or potential fuzzy rules) would exist. This exponential growth of fuzzy rules with number of inputs makes it impractical to use most existing architectures for problems of high dimensionality (such as the problem in this study). In reality, many of these rules would be redundant for modeling purpose, and therefore a suitable technique should start from a simple architecture and build on it as necessary.

One such approach uses ANalysis Of VAriance (ANOVA) (Brown and Harris, 1995) for decomposition of the output function of dimension *n* as;

$$f(X) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{i}^{n-1} \sum_{j=i+1}^{n} f_{i,j}(x_i, x_j) + \cdots + f_{1,2,\cdots,n}(X) \tag{4.1}$$

in which f_0 represents a constant (the function bias) and other terms represent univariate, bivariate, and other subfunctions respectively. In many cases a majority of the terms, particularly those of higher dimensions, are zero or negligible. For such cases, the ANOVA representation allows the input/output mapping to be approximated using a limited number of, as they are often called, "subnetworks" of much lower dimensions. An example of this decomposition is shown in Figure 4-2, where a five-dimensional function is decomposed into a one–dimensional and two two-dimensional sub-networks.

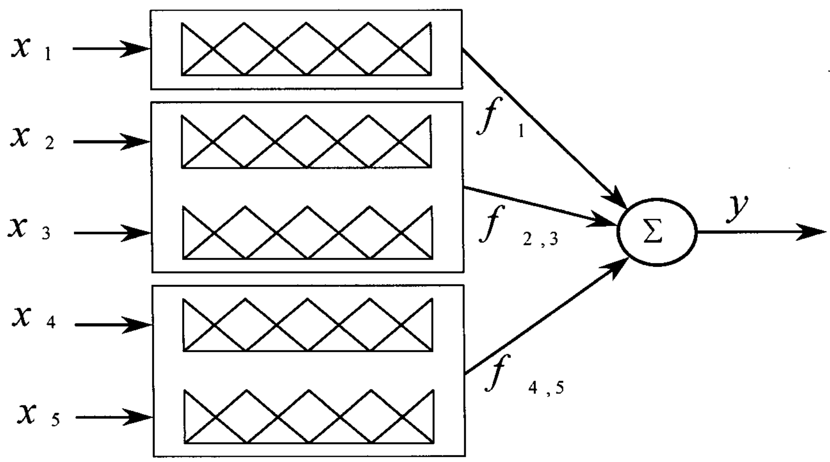

Figure 4-2 An Additive ANOVA Decomposition of A Neuro-fuzzy Rule Base (Brown and Harris 1995)

A fuzzy rule within each sub-network may have the form:

IF (x_1 is large AND x_2 is small) THEN (y is small) with confidence c,

Where c is the rule confidence or weight. A rule confidence of zero indicates that the rule is not contributing to the output while a rule confidence of one indicates that the rule is completely true. Values between zero and one allow the rules to partially fire. The number of fuzzy rules used in each sub-network depends on the number of membership functions that are used to fuzzify the inputs of that sub-network. In the above example, assuming that five membership functions are used for each variable, the first, second and third sub-network consist of 5, 25 and 25 fuzzy rules respectively. The consequent part of all the rules will be ORed (i.e., summed in this case) together.

ASMOD or adaptive spline modeling of observation data (Kavli, 1994) is an algorithm that uses the above decomposition to calculate parameters for a spline representation of an n-dimensional function. Denoting a B-spline sub-model by $s_u(.)$ and its associated input vector x_u, $u = 1,\ldots U$, the overall network output is given by:

$$y(x) = \sum_{u=1}^{U} s_u(x_u) \tag{4.2}$$

And if w_u denotes the weight coefficient vector for each submodel and $w = \bigcup_{u=1}^{U} w_u$, and for the basis function output vectors $a = \bigcup_{u=1}^{U} a_u$, the network output is then given by:

$$y(x) = \sum_{i=1}^{p} a_i(x) w_i \tag{4.3}$$

In its most general form, the B-splines could be of any dimension. For the above example, the first five splines will be one-dimensional while the rest of them will be two-dimensional. The membership functions for the first variable are equivalent to the one-dimensional splines but the membership functions for the second and third variables are equivalent of one-dimensional splines whose tensor product has produced the mentioned two-dimensional splines. The elements in the weight vector represent the rule confidences. It should also be noticed that the output function is linear with respect to the weight vector and thus can be solved easily. In this representation, algebraic product is used for fuzzy AND and sum is used for fuzzy OR.

In ASMOD algorithm, for any model structure (i.e., specific combination of subnetworks, the number and location of splines), one can use the training data to calculate the mean square error (MSE) of the output. The algorithm starts from the simplest structure (e.g., only first variable in one subnetwork with two triangular splines) and iteratively refines its structure until some stop criteria is satisfied. In each step among a number of potential (single) changes to the structure, the one with the best performance is selected and the process continues. Addition of a new input, combining an existing input to a subnetwork, splitting a subnetwork, and deleting an input are all possible changes to the structure.

Adding splines in the middle of existing ones, deleting splines, and changing the order of splines are also some other changes that are considered by the algorithm. Some measure of statistical significance is used as stopping criteria. Among many such measures, Bossley et al. (1995) state that, for noisy data, the Bayesian statistical significance measure appears to perform well and therefore is used in this study. The Bayesian statistical significance measure is given by (Brown and Harris, 1994):

$$K = L\ln(J) + p\ln(L) \tag{4.4}$$

where K is the performance measure, p is the size of current model, J is the MSE and L is the number of data pairs used to train the network.

4.4 Results

To compare the training results between ANN and the neuro-fuzzy model, the same training data (as described in Chapter 3) is used. A misclassification measure between the ANN or neural-fuzzy output and the expert's was define as:

$$E = (\sum_{i=1}^{c}(u_{ik} - u'_{ik})^2)^{0.5} \tag{4.5}$$

Where:

E = a measure of classification error

c = the number of classes (3 in this case)

U_{ik} = is the membership value of the kth observation in the ith class as assigned by the experts

U'_{ik} =the membership value of the kth observation in the ith class as estimated by the network.

A value of E equal to zero indicates identical experts and ANN or neural-fuzzy model memberships, while a value of 1.41 indicates a complete disagreement. Table 5-1 shows the results of the comparison. In general, the neuro-fuzzy classifier performed slightly better and produced more consistent results than the ANN classifier. The average classification error was 0.032 in the case of the neuro-fuzzy classifier compared to 0.107 in the case of the ANN classifier. The standard deviation was 0.055 in the case of the neuro-fuzzy classifier compared to 0.165 for the ANN classifier that indicates higher consistency of the neuro-fuzzy classification. The 25, 50, and 75 percentile of error were smaller in the case of neuro-fuzzy classifier compared to the ANN classifier.

Table 4-1 Membership Classification Error with the Expert Rating

	ANN classification	Neural-fuzzy classification
Average error	0.107	0.032
Standard deviation	0.165	0.055
25% error	0.024	0.011
75% error	0.12	0.03

Among the 21 variables, the neural-fuzzy model selected only 7 variables. The variables selected are Road curve, weather, road surface, contributing factor C11, C12, C13 and C21. Contributing factors C11, C12 and C13 are the three contributing factors assigned to vehicle 1. The police usually designate the driver of vehicle 1 as the driver in fault. C21 is the first contributing factor associated with vehicle 2. In many cases these two contributing factors will be left blank as the driver in vehicle two is assumed not to be in fault. The road curve, weather and road surface are important factors which define the contribution of the road environment to the accident.

4.5 Discussion

A comparison of the prediction results of ANN and Neuro-fuzzy models shows that the performance of the neuro-fuzzy approach is slightly better than ANN. In addition, the neuro-fuzzy approach uses the smallest number of input variables to obtain the slightly better performance. Therefore, it could be considered as the most efficient among the two approaches. The transparency of the knowledge gained by the neuro-fuzzy approach is also an advantage. The use of linguistic variables makes it relatively easy to interpret the rules and if necessary change them. Table 4-2 provides a comparison of the two approaches.

Table 4-2 Comparison of the Two Approaches

Comparison Criteria	ANN	Neuro-fuzzy
Constraining assumptions present	no	no
Ability to model nonlinearity	yes	yes
Ability to model multivariability	yes	yes
Automatic exclusion of irrelevant inputs	no	yes
Transparency of The model	not transparent	very good
Ability to insert expert knowledge in the model	no	yes

4.6 Conclusion

Despite the flexibility and the good performance of ANNs in modeling nonlinear relationships, they have often been criticized for acting as "black boxes". The knowledge contained in the ANN model is kept in the form of a weight matrix that is hard to interpret and can be misleading at times. The efficiency of ANN models is also another point of concern. Since it is not always possible to determine the significance of the input variables in advance, any potential candidate may be included in the model. It is therefore important to identify and exclude those input variables that do not have a significant contribution. This would lead to a more efficient model—Neuro-fuzzy model. It is shown that the neuro-fuzzy approach has the potential to provide similar or better accuracy compared with ANNs while overcoming their shortcomings. A comparison of the prediction results of ANN and Neuro-fuzzy models shows that the performance of the neuro-fuzzy approach is slightly better than ANN. In addition, the neuro-fuzzy approach uses the smallest number of input variables to obtain the slightly better performance.

CHAPTER 5

USING ACCIDENT CORRECTABILITY TO IDENTIFY ACCIDETN-PRONE LOCATIONS

5.1 Background

In the previous two chapters three methods were described that utilize the knowledge of safety experts in classifying accidents into one or a combination of the three highway system components: the driver, the vehicle and the road. This chapter describes the application of the classification process for the identification of accident-prone locations from a broader perspective than the engineering improvement.

5.2 Application of the Method to the Identification of APLs

In traditional black spot programs, locations are identified as accident-prone if they exhibit significant number of accidents above an established norm. Accident causes are not considered. The concept of "accident correctability" considers the factors which contributed to the accident (i.e. road-related, driver-related, or vehicle-related). In order to be identified as accident-prone, locations must exhibit a significant number of *correctable* accidents. In the following sections, the APLs will be identified as road-related, driver-related and vehicle-related respectively.

5.2.1 Redefining Accident Frequency and Rate

The identification of accident-prone locations involves a comparison of a certain accident measure, usually the accident rate, with an established norm. Traditionally, the total number of accidents is used to calculate this accident measure. The classification process described in the previous two chapters will be used to modify the method of calculating accident measures. As previously described, the output of the classification process is three membership values (one representing each component of the road system). Each value ranges between zero and one. For example, from a road improvement (engineering countermeasures) perspective, the accident frequency and accident rate can be calculated as follows:

$$Accident\ Frequency = \sum_{i=1}^{N} W_i \tag{5.1}$$

$$Accident\ Rate = \frac{\sum_{i=1}^{N} W_i}{Exposure\ Measure} \tag{5.2}$$

where:

W_i = the degree with which the ith accident belongs to the road environment group (ranges between 0 and 1).

N = the total number of accidents at the location during a certain time period.

Exposure is usually measured in million vehicle kilometers for road sections and million entering vehicles for road intersections.

5.2.2 Examples

To illustrate the application of the modified black spot program, the method was applied to the South Coast Region of the Province of British Columbia, Canada. The South Coast Region is one of six regions of the provincial highway system. There are approximately 3000 kms of primary highways (two lanes, multi-lane, and freeways), and 275 signalized intersections in this region. Using police reported accident data from 1989 through 1991, accident-prone locations were identified by both the traditional black spot method and the modified method proposed in this thesis. Results from the two methods were compared and presented below. Figure 5-1 and Figure 5-2 show the results of applying the two methods to identify accident-prone locations for the signalized intersection category. As discussed above, and as a result of weighting accidents by their road correctability, the number of accident-prone locations using the modified method was reduced. In addition to increasing the potential effectiveness of safety improvement projects, limiting the list of accident-prone locations by the Modified method eliminates the frustration of diagnosing the nature and causes of the problem at the identified accident-prone locations. A traditional accident-prone location is not guaranteed to exhibit a recognizable pattern of accidents for which a solution can be readily found. An accident-prone location identified by the modified method, on the other hand, has a higher chance of showing some pattern of accidents for which some engineering countermeasures can be proposed.

As indicated in the introduction to this section, the modified method not only reduces the total number of accident-prone locations (as discussed above), but also alters their ranking. This is very important in situations when the road authority has resources to address only a limited number of black spots, it is important to focus on those with the highest potential of accident reduction. It is logical that the more a location deviates from the typical (i.e., norm or average) accident rate, the higher the potential for accident reduction.

To illustrate this point, Table 5-1 shows a list of accident-prone signalized intersections in the South Coast Region ranked by the traditional method (showing only the top 19 to economize on space). For comparison purposes, the corresponding ranking using the proposed (modified) is also given along with the difference between the two ranks. The criteria used for ranking accident-prone locations is the ratio of observed accident rate to the critical accident rate. In the modified method, the observed accident rate was adjusted by the algorithm as explained before. The difference in rank between the two methods seems to increase as the one approaches the bottom of the list. Another important point is that many of the intersections that ranked highly in the traditional method were not accident-prone in the modified method.

Table 5-1 Accident-prone Intersections Using the Traditional Method

(showing top 19 only)

Intersection number		Traditional Rank	Modified Rank	Difference (Trad-Mod)
Segment No.	Kilometer Mark			
3183	6.1	1	2	-1
2510	18.9	2	1	1
770	3.7	3	3	0
3183	4.5	4	4	0
770	13.4	5	7	-2
2730	0.6	6	6	0
2510	12.0	7	5	2
3183	3.1	8	8	0
2710	12.7	9	9	0
2510	14.0	10	10	0
3183	7.7	11**	23	-12
2730	12.8	12	11	1
2510	15.6	13	19	-6
2510	17.2	14	16	-2
770	15.4	15	12	3
2714	16.3	16	13	3
770	0.9	17**	24	-7
770	0.0	18**	27	-9
2510	20.5	19**	34	-15

** Indicates intersection is not accident-prone by the modified method.

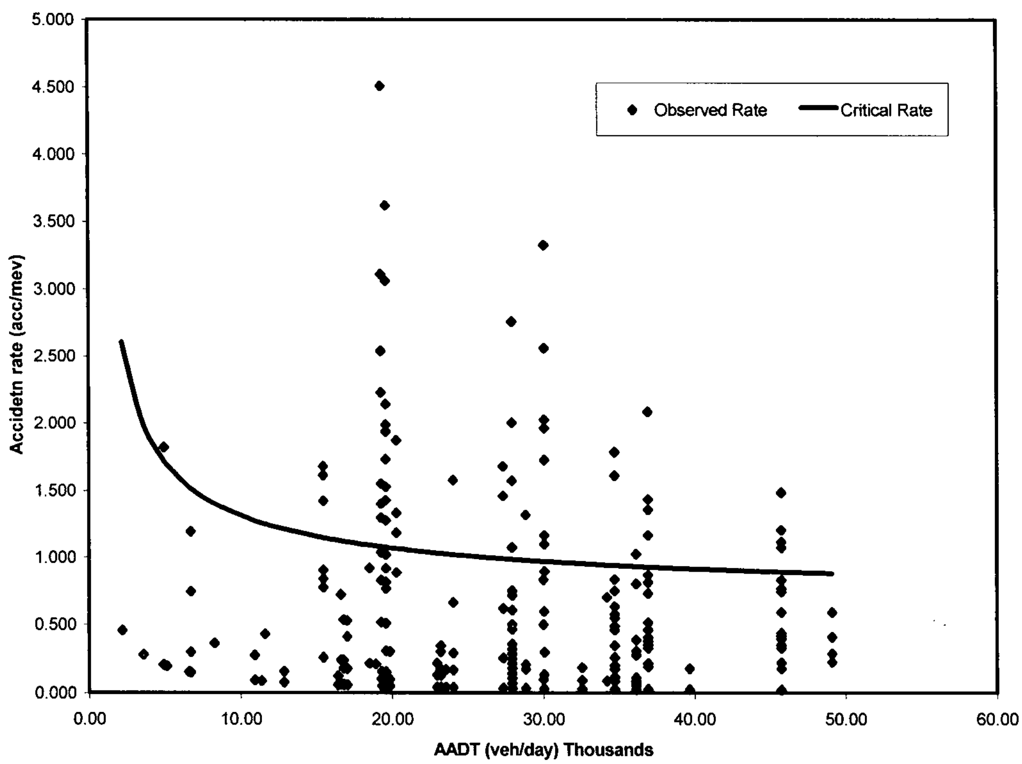

Figure 5-1 APLs for Signalized Intersections Using traditional Method (#APLs=55)

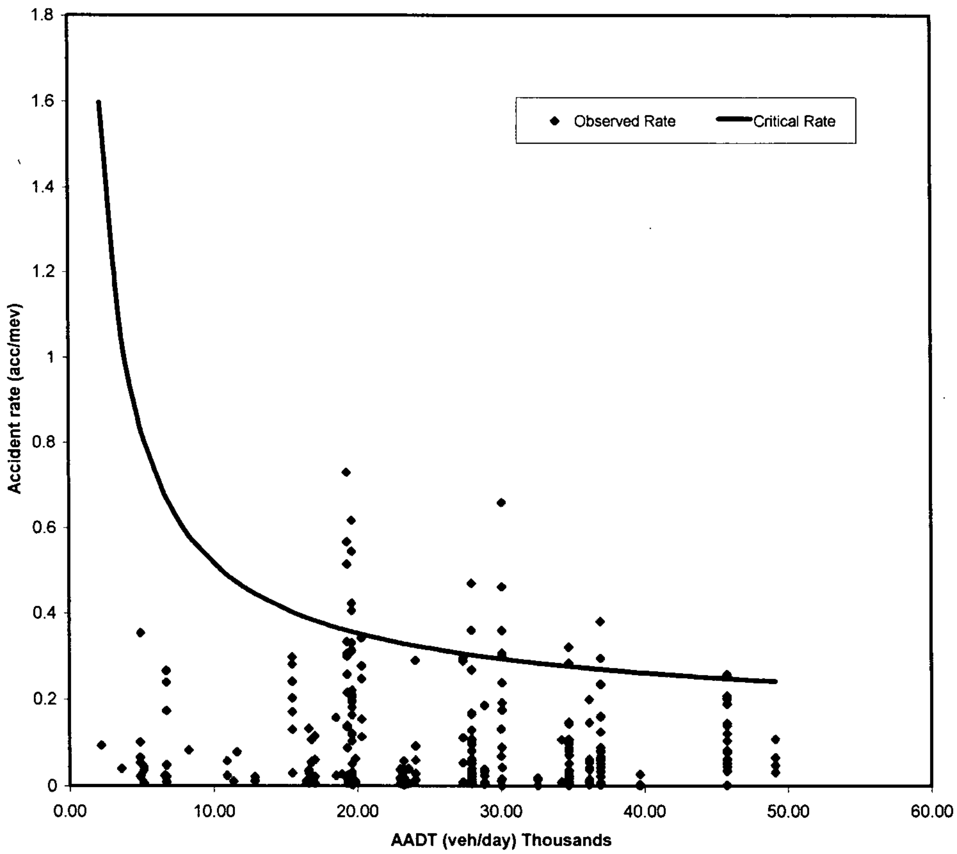

Figure 5-2 APLs for Signalized Intersections Using ANN Classification Output

(#APLs=19)

5.2.3 Driver-Related Accident-prone Locations

Another goal of this study is trying to find driver-related factors involved in the accident and identifying driver-related accident-prone locations. By applying the ANN technique, the three-road system membership values were generated and this gave the opportunity to identify the driver-related accident-prone locations. The number of driver-related accident-prone locations that provided by the ANN rating technique is 48, and it's shown in Figure 5-3.

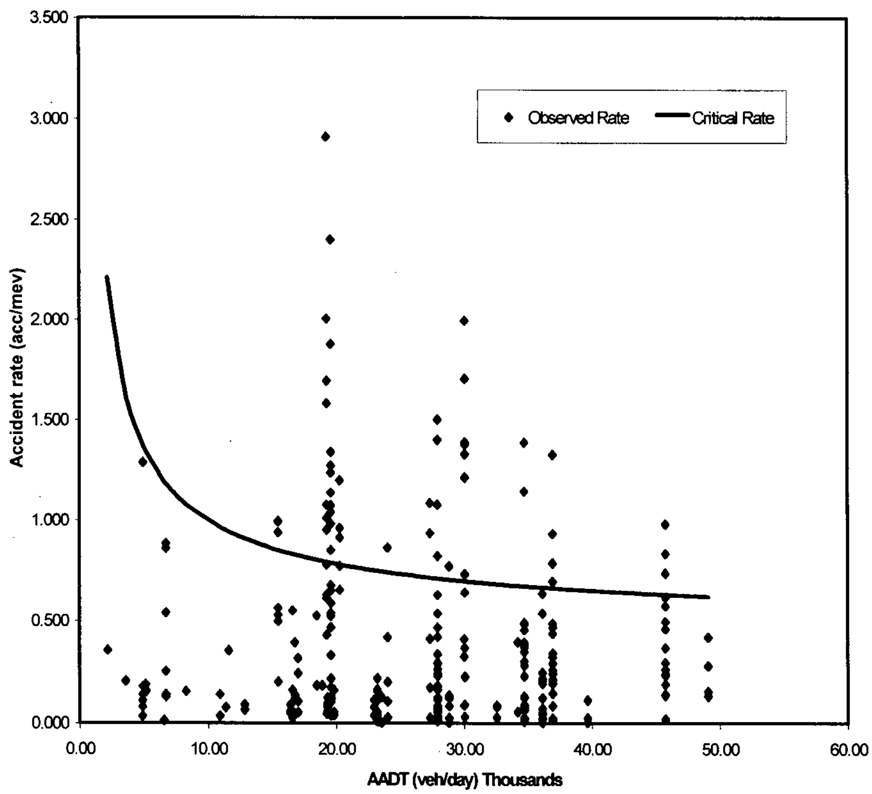

Figure 5-3 APLs for Signalized Intersections Using ANN Rating (#APLs=48)

Compared to the traditional method, not only the number of accident-prone locations was different, but also the rankings were different, the results were listed in Table 5-2.

Interseciton		Traditional	ANN	Difference
Seg no	kilo Mark	Rank	Rank	(Trad-ANN)
3183	6.1	1	1	0
2510	18.9	2	3	-1
770	3.7	3	2	1
3183	4.5	4	4	0
770	13.4	5	6	-1
2730	0.6	6	8	-2
2510	12	7	5	2
3183	3.1	8	11	-3
2710	12.7	9	10	-1
2510	14	10	12	-2
3183	7.7	11	7	4
2730	12.8	12	14	-2
2510	15.6	13	13	0
2510	17.2	14	15	-1
770	15.4	15	19	-4
2714	16.3	16	9	7
770	0.9	17	17	0
770	0	18	20	-2
2510	20.5	19	16	3
3120	3.2	20	22	-2
2714	13.4	21	18	3
3190	0	22	23	-1
720	26.3	23	21	2
770	2.8	24	27	-3
770	14.8	25	30	-5
2730	14.6	26	24	2
3110	6.7	27	41	-14
2710	20.2	28	26	2
3190	2.3	29	31	-2
3184	13.1	30	40	-10
2710	11.9	31	36	-5
3183	5.3	32	28	4
770	9.5	33	25	8
3184	3.2	34	38	-4
720	25.1	35	29	6
3172	11.6	36	44	-8
770	4.4	37	33	4
3183	7	38	32	6
2710	0	39	48	-9
720	23.1	40	39	1
3120	4.7	41	37	4
3120	5.2	42	34	8
3184	12.3	43	43	0
3183	1.5	45	35	10
2510	7.2	47	47	0
770	29.5	49	45	4
2510	5.000	50	46	4

Table 5-2 APLs Ranking Difference in Both Method

5.3 Conclusion

In traditional black spot programs, locations are identified as accident prone if they exhibit significant number of accidents above an established norm. Accident causes are not considered. The concept of "accident correctability" considers the factors which contributed to the accident (i.e. road-related, driver-related, or vehicle related). In order to be identified as accident prone locations, locations must exhibit significant number of *correctable* accidents. The APLs should be identified as road-related, driver-related and vehicle-related respectively. The new approach redefines accident frequency and rate to reflect accidents correctability and then uses ANNs classification and Empirical Bayes technique to identify APLs. The method was applied to identify accident prone locations in the South Coast Region of the Province of British Columbia and the results were compared with the traditional approach. The results indicated that the new approach has two main benefits over the traditional approach. First, accidents are considered as road related, driver-related, or vehicle related, resulting in fewer number of accident prone locations. Secondly, this new technique eliminates locations which are not correctable from a road authority perspective, and it also identifies locations which are most likely correctable from enforcement and driver education, and thus increasing the potential effectiveness of road safety improvement programs.

CHAPTER 6

CONTERMEASURE-BASED APPROACH USING ACCIDENT PREDICTION MODELS

6.1 Introduction

In traditional black spot programs, for a location to be identified as accident-prone, it must exhibit a higher accident occurrence than an established "norm". However, a problem that arises when analyzing accident-prone locations identified by the traditional Black Spot program is that many of these locations do not have well-defined accident patterns for which countermeasures can be developed. And consequently, treating these locations may not be cost effective. To address this problem, Sayed (1995) described another approach for the identification process which is based on identifying locations that have over-representation of particular accident patterns. This facilitates the selection of countermeasures and identifies locations that are good candidates to be cost effectively treated.

Traditional Black Spot programs start with a problem (high accident occurrence) and attempt to find solutions (countermeasures), but many of these locations identified by this method may not have well-defined accident patterns for which countermeasures can be developed. Consequently, treating these locations may not be cost effective. The countermeasure-based approach reverses the traditional process of linking problems with solutions by first identifying main accident patterns that can be targeted by specific countermeasures and then searching for locations which have over-representation of these patterns. Sayed (1995) assessed the over-representation of a particular accident type in the total number of accidents using the ratio of the accident pattern to the total number of accidents. This chapter describes an alternative approach using accident prediction models. Two types of accident predictions model are used and the results are compared.

6.2 The Countermeasure-based Approach Using APMs

There are two methods can be applied to identify accident-prone locations using accident prediction models. Both methods use accident prediction models to calculate the predicted accident frequency (mean) and its variance, method one uses EB refinement, and method two does not use EB refinement. Only the first method will be discussed in this thesis.

6.2.1 The Accident Prediction Models

Accident prediction models can provide reliable safety estimates of road segments and intersections and can be used in identifying accident-prone locations. The recently developed accident prediction models (Sayed et al., 1998) for urban signalized intersections in the Greater Vancouver Regional District (GVRD) utilize the generalized linear regression modeling (GLIM) approach (as described in Chapter 2, section 2.5.1), which addresses and overcomes the shortcomings associated with the conventional linear regression approach.

The accident prediction models use sample accident, traffic and intersection design data corresponding to urban signalized intersections located in Greater Vancouver. The data set contained a total of 139 intersections from the City of Vancouver and the City of Richmond. The source of the data is the MV104 accident reporting form, the British Columbia's accident police report. The data set included a total of 12,325 accidents occurred during the 1993-1995 period.

Model for the Total Number of Accidents (equation 6.1) and Model for Specific Accident Type (Left Turn Accident, equation 6.2) are described below:

Model for the Total Number of Accidents:

$$Accidents\,/\,year = 1.8718 \times \left(\frac{AADT_{Major\ road}}{1000} \right)^{0.36} \times \left(\frac{AADT_{Minor\ road}}{1000} \right)^{0.4669} \qquad (6.1)$$

where $k = 9.823$

Model for the Left Turn Accidents:

$$Accidents\,/\,year = 0.2879 \times \left(\frac{AADT_{Major\ road}}{1000} \right)^{0.4568} \times \left(\frac{AADT_{Minor\ road}}{1000} \right)^{0.5685} \qquad (6.2)$$

where $k = 3.42$

6.2.2 Empirical Bayes (EB) Approach

There are two clues to the safety of a location: its traffic and road characteristics, and its historical accident data (Hauer, 1992, Brude and Larsson, 1988). The Empirical Bayes (EB) approach makes use of both kinds of clues. The EB approach is used to refine the estimate of the expected number of accidents at a location by combining the observed number of accidents at the location with the predicted number of accidents obtained from the GLIM model to yield a more accurate, location-specific safety estimated. The EB estimated number of accidents for any intersection can be calculated using (Hauer 1992):

$$EB\ safety\ estimate = \alpha \cdot pred + (1-\alpha) \cdot count, \quad where \quad \alpha = \frac{1}{1+\frac{var(pred)}{pred}} \tag{6.3}$$

where

count = observed number of accidents at the location

pred = predicted number of accidents as estimated from the GLIM model

var(pred) = the variance of the GLIM estimates

Since $var(pred) = \frac{(pred)^2}{\kappa}$ (Sayed et al., 1998), equation 6.3 can be rearranged as:

$$EB\ safety\ estimate = \left(\frac{\kappa}{\kappa + pred}\right) pred + \left(\frac{pred}{\kappa + pred)}\right) count \tag{6.4}$$

In addition, the variance of the EB estimate can be calculated using (Kulmala, 1995):

$$var\ (EB\ safety\ estimate) = \frac{\kappa\ pred^2}{(\kappa + pred)^2} + \left(\frac{pred}{\kappa + pred}\right)^2 count \tag{6.5}$$

As showed in equation 6.4, the EB approach combines both the individual accident history of the location and GLIM model prediction. In addition to combining the two safety clues and providing site-specific safety estimates, it has also been shown that the

EB method significantly reduces the regression to the mean effects that are inherent in observed accidents count (Brude and Larsson, 1988). Regression to the mean is a statistical phenomenon that refers to the tendency of extreme events (high number of accidents) to be followed by less extreme values (a lower number of accidents) event if no change has occurred that in the underlying mechanism which generates the process.

6.2.3 Identification of Accident-prone Locations

Accident-prone locations (APLs) are defined as the locations that exhibit a significant number of accidents compared to a specific norm. Because of the randomness inherent in accident occurrence, statistical techniques that account for this randomness should be used when identifying APLs. The EB refinement method can be used to identify APLs according to the following process (Belanger, 1994):

Estimate the predicted number of accidents and its variance for the intersection using the appropriate accident model. This follows a gamma distribution (the prior distribution) with parameters α_1 and β_1, where:

$$\beta_1 = \frac{E(\Lambda)}{Var(\Lambda)} = \frac{\kappa}{E(\Lambda)} and \alpha_1 = \beta_1 \cdot E(\Lambda) = \kappa \tag{6.6}$$

Determine the appropriate point of comparison base on the mean and variance values obtained in step (1). Usually the 50th percentile, P_{50} is used as a point of comparison. P_{50} is calculated such that:

$$\int_0^{P_{50}} \frac{(\kappa/E(\Lambda))^{\kappa} \cdot \lambda^{\kappa-1} \cdot e^{-(\kappa/E(\Lambda))\lambda}}{\Gamma(\kappa)} d\lambda = 0.5 \tag{6.7}$$

Calculate the EB safety estimate and its variance from equations (6.4) and (6.5) respectively. This is also a gamma distribution (posterior distribution) with parameters α_2 and β_2:

$$\beta_2 = \frac{EB}{Var(EB)} = \frac{\kappa}{E(\Lambda)} + 1 and \alpha_2 = \beta_2 \cdot EB = \kappa + count \tag{6.8}$$

Then, the probability density function of the posterior distribution is:

$$f_{EB}(\lambda) = \frac{(\kappa/E(\Lambda)+1)^{(\kappa+count)} \lambda^{\kappa+count-1} e^{-(\kappa/E(\Lambda)+1)\lambda}}{\Gamma(\kappa+count)} \tag{6.9}$$

Identify the location as accident-prone if there is a significant probability that the intersection's safety estimate exceeds the P_{50} value. Thus, the location is identified as accident-prone if:

$$\left[1 - \int_0^{P_{50}} \frac{(\kappa/E(\Lambda)+1)^{(\kappa+count)} \lambda^{\kappa+count-1} e^{1(\kappa/E(\Lambda)+1)\lambda}}{\Gamma(\kappa+count)} d\lambda \right] \geq \delta \tag{6.10}$$

where δ represents the confidence level desired (usually 0.95).

To illustrate, consider a signalized intersection No. 28 with observed accident rate of 39.7 acc/yr, applying the model for the total number of accidents, the expected value is 20.17 acc/yr with a variance of 41.40 (acc./yr)^2. The P_{50} value for the Gamma distribution can be calculated as 19.49 acc/yr. The EB safety and its variance calculated from equations 5.3 and 5.4 are 33.28 acc/yr and 22.38 (acc./yr)^2 respectively. The posterior distribution is also shown in Figure 6-1. From the figure, it can be shown that the probability of having accidents less than P_{50} is 0.03 percent. This means that there is a significant probability

(99.7%) of exceeding the P_{50} value and the intersection can be considered accident-prone. For the same intersection, the observed left turn accident rate is 8.33 acc/yr, the expected accident rate is 5.53 acc/yr with a variance of 8.95 $(acc/yr)^2$. The P_{50} value is 5.00 acc/yr, the EB safety and its variance are 7.26 acc/yr and 4.49 $(acc/yr)^2$ respectively. The posterior distribution is shown in Figure 6-2. From the figure, it can be shown that the probability of having accidents less than P_{50} value is 14 percent. This means this intersection will not be considered as left turn accident-prone location, so adding a left turn lane or applying other left turn related engineering improvements may not be cost effective to reduce the number of accident.

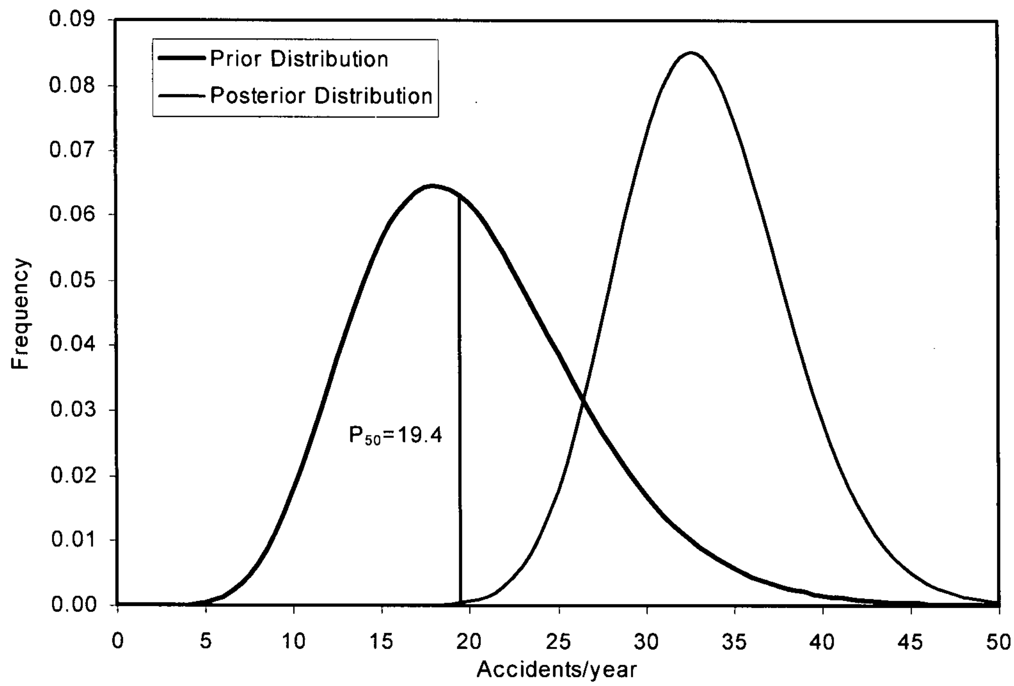

Figure 6-1 Identification of APL for Intersection No.28 (Total Accident)

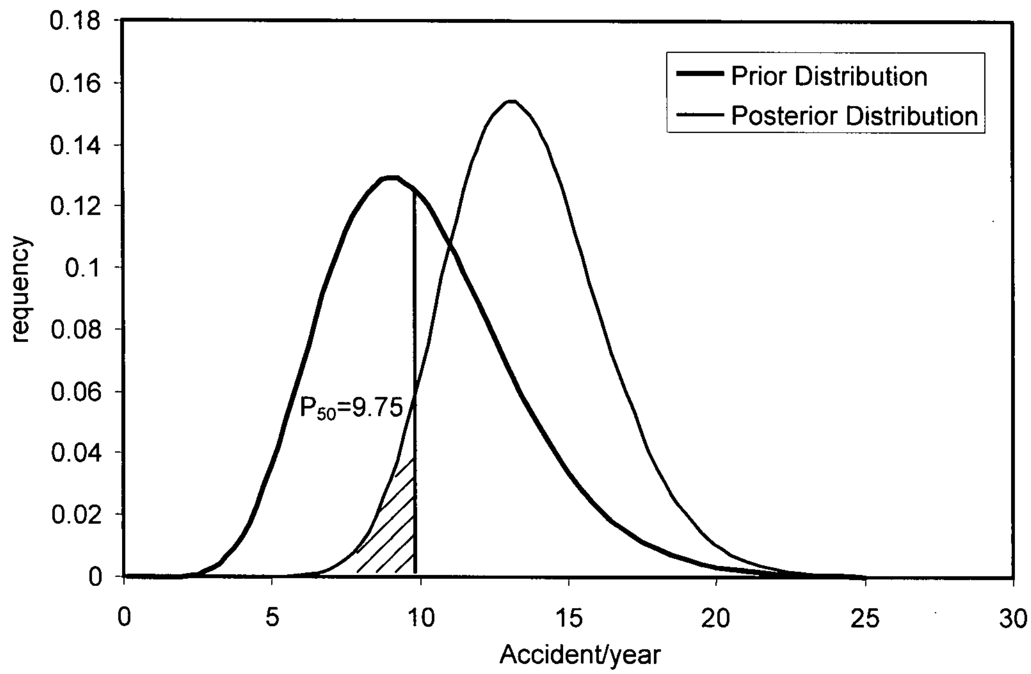

Figure 6-2 Identification of APL for Intersection No.28 (Left-turn Accident)

On the other hand, consider another signalized intersection NO. 82 with an observed total accident rate 35.3 acc/yr and left turn accident rate 14.67 acc/yr. By applying the same method, it was found that this intersection is not accident-prone in terms of total accident number. But it is accident-prone in terms of left turn accident. So adding a left turn lane or applying other left turn related engineering improvement will likely be cost effective to reduce the number of accident.

To investigate the effectiveness of the this method, it is applied to the set of 139 signalized intersections from the City of Vancouver and the City of Richmond using accident data from 1993 to 1995. As an illustration, left turn accidents are considered.

Table 6-1 shows intersections that identified by black spot program and intersections that identified by countermeasure-based program. There are 15 intersection identified as accident-prone locations by traditional black spot program and 15 intersection identified as accident-prone locations by countermeasure-based program (left turn accidents).

As shown in Table 6-1, seven out of the 15 intersections that exhibited over-representation of left turn accidents are not identified as accident-prone by the traditional black spot program, the ranking of the accident-prone location is different too. The ranking criterion used in this table is to calculate the difference between the EB estimated and the predicted frequency (as obtained from the accident prediction model) for the accident-prone locations identified by this method.

Table 6-1 Comparison between the Traditional Black Spot and the Countermeasure-Based Programs (Left Turn Accident)

Intersection Number	Countermeasure-Based Program (Left Turn Accident)			Black Spot Program		
	EB-Pred*	Rank	Left-turn accident-prone	EB-Pred*	Rank	Accident-prone location
36	5.27	1	Yes	8.55	6	Yes
112	5.21	2	Yes	8.39	9	Yes
121	5.21	3	Yes	7.02	12	Yes
82	4.38	4	Yes			No
132	4.37	5	Yes	8.42	8	Yes
7	4.24	6	Yes	12.72	2	Yes
18	3.88	7	Yes	8.62	5	Yes
27	3.72	8	Yes	11.97	3	Yes
64	3.63	9	Yes			No
99	3.11	10	Yes			No
79	2.93	11	Yes			No
69	2.80	12	Yes			No
76	2.33	13	Yes			No
86	2.26	14	Yes			No
128	1.76	15	Yes			No
28			No	13.11	1	Yes
46			No	11.06	4	Yes
52			No	8.45	7	Yes
129			No	7.95	10	Yes
15			No	7.42	11	Yes
53			No	5.00	14	Yes
5			No	3.79	15	Yes

6.3 Conclusion

This chapter described the alternative technique for identifying accident-prone locations by applying the accident prediction model and the countermeasure-based approach. The countermeasure-based approach reverses the traditional process of linking problems with solutions by first identifying main accident patterns that can be targeted by specific countermeasures and then searching for locations which have over-representation of these patterns. The techniques were applied to identify accident-prone location of 139 signalized intersections in Vancouver and Richmond. The results indicated that many locations identified in the countermeasure-based approach because of their well-defined accident patterns are not identified as accident-prone locations according to the traditional approach. The application of this technique should facilitate the selection of countermeasures and improve chances of getting better return for the money spent in highway safety improvement program.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATION

7.1 Conclusions

Given the increasing cost of road accidents, many road authorities have established road safety improvement programs. While the success of these programs has varied considerably from one jurisdiction to another, the overall performance of these programs has been less than satisfactory in terms of number of accidents eliminated. The main reason for that is believed to related to the inadequacy of procedure adopted in the execution of these programs. Most importantly, the faulty identification of accident-prone locations (i.e., identification of locations that are not really accident-prone) seems to be the primary reason behind the lack of success of these programs.

This thesis has reviewed the traditional concepts of Road Safety Improvement Programs and proposed two new concepts and ideas for enhancing the performance of these programs. First, the definition of an accident prone location has been altered to the causes and the factors that contributed to the accidents. This required the development of sophisticated mathematical procedures that are believed to be currently viable and will become more viable with the advancement in the technologies of high-speed computing. Using the artificial neural network and neuro-fuzzy approaches, a correctable accident was defined as a real number between zero and one that reflects the degree with which the road component (or other components) has contributed to the occurrence of this accident. It is believed that a RSIP based on the newly-defined accident prone location list will not only be more successful, but also more cost-effective than traditional RSIPs. Second, for jurisdictions where these concepts cannot be implemented for any reason, an alternative approach for identifying candidate safety improvement projects was introduced. The Countermeasure-Based approach relies on the selection of sites with recognizable patterns of accidents for which engineering treatments can be easily found. Finally, it is believed that a primitive condition for the successful application of any RSIP is the establishment of a reliable traffic accident data collection program at the national level. Of particular importance to this effort is the establishment of a road location referencing system (preferably GIS-based) whereby accidents can be accurately related to the characteristics of the location where they occur.

7.2 Recommendations for Further Research

This section presents a series of improvements, which can enhance and strengthen the methodologies described in this thesis.

7.2.1 The Obstacles Associated with Road Safety Data

The variables used in the classification process and the countermeasure-based approach come from the provincial accident database. Accident data in the province has been degrading in recent years. The main reason for the degradation in data is a reduction in resources and shifting priorities within provincial enforcement agencies. The reliability of accident data that is required to evaluate road safety management programs is often suspected. Even the highest level of accident reporting data (that data collected by enforcement officials) has at times, been found to be unreliable. The sources of unreliability in the data are many, but the principle sources of errors are the mistakes made by officials at the scene, either by a mis-interpretation or a simple coding error. Errors in judgment and recollection are also made in the self-reporting of accidents (reported by those involved in the crash). More problems are introduced when the data is entered into the data warehouse by clerical staff. Still another problem with the collision data in British Columbia is the accuracy of the data. There are numerous data fields that are required to be completed and many are either subjective or difficult to determine with great accuracy.

A much-needed research would be to develop a new approach which utilizes the supplemental safety data to evaluate road safety performance. A risk index, based on supplemental data, can be developed and demonstrated in the assessment of road safety and to support planning and design decisions. A comparison with the results from the classical approaches should be performed and guidelines on when this approach should be undertaken should be established.

7.2.2 Proactive Road Safety

The techniques provided in this thesis are part of a reactive approach to road safety. There is an inherent obstacle in delivering road safety in a reactive manner. To be effective, significant road safety problems, evidenced by a high frequency of accidents, must exist before hazardous locations can be identified and remedial actions taken to improve safety performance. Allowing an accident problem to develop and then reacting to that problem is costly compared to an approach that attempts to prevent accidents before a facility is opened. Thus, a proactive approach to delivering road safety is expected to overcome this obstacle.

In recent years, attention to the management of road safety has surfaced in the area of road planning. Planning initiatives often operate within a road authority's capital program, in contrast to the black-spot program that normally operates within a rehabilitation program. The intention of introducing a focus on road safety early in the planning process is to prevent accidents from occurring once a new facility is opened. Consequently, this is a proactive approach to road safety.

In the past, road safety was only considered in an implicit manner, such that if the road design standards were met, then it was assumed that all safety concerns were satisfied. Unfortunately, without explicit and focused attention to road safety issues, the selection of minimum design standards will often occur, resulting in a less than satisfactory level of safety for a new facility.

However, one obstacle associated with the delivery of proactive road safety measures is the lack of process and opportunity to explicitly consider road safety issues. There is also a lack of the necessary tools to evaluate road safety in a proactive manner. A research that targets these areas is surely needed.

REFERENCE

Abdelwahab, W. and Sayed, T. (1993). "Some observations on the use of accident rates in the identification of accident-prone locations", Internal Report, Highway Safety Branch, Ministry of Transportation and Highways, British Columbia, Canada.

Abdelwahab, W. and Sayed, T. (1998). "Artificial Neural Networks: A New Approach to Behavioral Mode Choice Modeling", Proceedings of the 8th World conference on Transport Research, Antwerp, Belgium, July 12-17.

Belanger, C. (1994). "Estimation of safety of four-leg unsignalized intersections"' *Transportation Research Record,* 1467, National Research Council, Washington D. C., pp. 23-29.

Berger, J.O., (1995). "Statistical decision theory and bayesian analysis", 2nd ed. Springer Verlag, New York.

Bezdek, J. C., (1981). "Pattern recognition with fuzzy objective function algorithm", *Plenum Press*, New York, N. Y.

Bezdek, J. C.; Chuali, S. K., and Leep, D., (1986). "Generalized K-nearest neighbor rules", *Fuzzy Sets and Systems* 18, pp. 237-256.

Bezdek, J. C. and Pal, S. K., (1992). "Fuzzy models for pattern recognition", *IEEE Press*, New York, N.Y.

Bossley K., Brown M., and Harris C., (1995). "Parsimonious Neurofuzzy Modeling", Technical report, University of Southampton, Department of Electronics and computer Science.

Brown, M., and Harris, C., (1994). "Neurofuzzy Adaptive Modeling and Control", Prentice Hall, New York.

Brown, M., and Harris, C., (1995). "A Perspective and Critique of Adaptive Neurofuzzy Systems Used for Modeling and Control Applications", *International Journal of Neural Systems*, Vol. 6(2), pp. 197-220.

Brude, U., and Larsson, J. (1988). "The use of prediction models for eliminating effects due to regression-to the-mean in road accident data", *Accident Analysis and Prevention*, 20(4): 299-310.

Cohen, J. A., (1960). "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement* 20, 37-46.

Duda, R. A. and Hart, P. E. (1973). "Pattern classification and scene analysis", John Wiley, New York, NY.

Fleiss, J. L. (1971). "Measuring nominal scale agreement among raters", *Psychological Bulletin*, 76(5), 378-382.

Hauer, E., Ng J. C. N. and Lovell J., (1988). "Estimation of safety at signalized intersections", *Transportation Research Record*, 1185, Transportation Research Board, National Research Council, Washington D. C., pp. 48-61.

Hauer, E. (1992). "Empirical Bayes approach to the estimation of 'unsafety': The multivariate regression method", *Accident Analysis and Prevention*, vol24, No 5, pp. 457-477.

Higle, J.L. and Witkowski, J.M., (1988). "Bayesian Identification of Hazardous Locations", *Transportation Research Record*, 1185, Transportation Research board, National Research council, Washington D. C., pp. 24-31.

Hsieh, C. (1993). "Some Potential Applications of Artificial Neural Systems in Financial Management", Journal of System Management, April, pp. 12-15.

Jang, J., (1993). "ANFIS: Adaptive-Network-Based Fuzzy Inference System", *IEEE Transaction on Systems, Man and Cybernetics*, Vol. 23, pp. 665-685.

Jones, W. and Hoskins, J., (1987). "Back-propagation, A generalized Delta Learning Rule", Byte magazine, October, pp. 10-15.

Jovanis, P. P. and Chang, H. L., (1986). "Modeling the relationship of accidents to miles traveled", *Transportation Research Record*, 1068, Transportation Research board, National Research Council, Washington D. C., pp. 42-51.

Kandel, A. (1982). "Fuzzy techniques in pattern recognition", John Wiley & Sons, New York, NY.

Kavli, T., (1993), "ASMOD – An Algorithm for Adaptive Spline Modeling of Observation Data", *International Journal of Control*, Vol. 58(4), pp. 947-967.

Keller, J. M.; Gray, M. R. and Givens, J. A. (1985). "A fuzzy K-nearest neighbor algorithm", *IEEE Trans. Syst., Man, Cybern.,* 15(4), 580-585.

Kosko, B., (1992). "Neural Networks and Fuzzy Systems", Englewood Cliffs, N.J.: Prentice-Hall.

Kulmala, R., (1995). "Safety at rural three- and four-arm junctions. Development of accident prediction models", *Espoo* 1995, Technical Research Centre of Finland, VTT 233.

Kulmala, R. and Roine, M., (1988). "Accident prediction models for two-lane roads in Finland", *Conference on traffic safety theory and research methods proceedings*, April, Session 4: Statistical analysis and models. Amsterdam: SWOV, pp. 89-103

Lawrence, J., (1991). "Introduction to Neural Networks", 3rd edition, California Scientific Software, Grass Valley.

Master, T., (1993). "Practical Neural Network Recipes in C++", *Academic Press*, San Diego.

Miaou, C. and Lum, H., (1993). "Modeling vehicle accident and highway geometric design relationships", *Accidents Analysis and Prevention*, Vol. 25, No 6, pp. 689-709.

Morris, C. N., (1988). "Determining the Accuracy of Bayesian Empirical Bayes Estimators in the Familiar Exponential Families", *Statistical Decision Theory and Related Topics, IV*, Vol. 1, New York, 1988, pp. 251-263.

Numerical Algorithms Group (NAG), (1994). "The GLIM system. Release 4 manual", The Royal Statistical Society.

Nicholson, A. J. (1980). "Identification of hazardous locations", *PRU Newsletter* No. 66, Ministry of Transport, Newzealand.

Nordon, M., Rrlansky, J., and Jacobs, H., (1956). "Applications of statistical quality-control techniques to analysis of highway accident data", *Bulletin* 117, Highway Research Board, National Research Council, Washington, D.C., pp. 17-31.

Oppe, S., (1982). "Detection and analysis of black spots with even small accident figures"' R82-15, SWOV, Leidschendam, The Netherlands.

Oppe, S., (1992). "A comparison of some statistical techniques for road accident analysis", *Accident Analysis and Prevention*, 24(4), pp. 397-423.

Rodriguez, L. F., (1998). "Accident Prediction Models for Unsignalized Intersections", Master Thesis, Department of Civil Engineering, The University of British Columbia, Vancouver, Canada.

Romesburg, H. C., (1984). "Cluster analysis for researchers", Lifetime learning Publications, Belmont, California, USA.

Rumelhart, D. E. and McClelland, J. (eds) (1986). "Parallel distributed Processing", MIT Press, Cambridge, Vol. 1.

Saccomanno, F. and Buyco, C., (1988) "Generalized loglinear models of truck accident rates", Paper presented at Transportation Research Board 67th annual meeting. Washington, D. C.

Zadeh, L. A. (1965). "Fuzzy Sets", *Information Control* 8, pp. 338-353.

Sayed, T.; Abdelwahab, M.; and Navin, F. (1995). "Identifying accident-prone locations using fuzzy pattern recognition", *Journal of Transportation Engineering*, 121(8): 352-358.

Sayed, T. (1995). "A Highway Safety Expert System: A new Approach to Safety Programs", Ph.D. thesis. Department of Civil Engineering, University of British Columbia.

Sayed, T., Feng, C., and Rodriguez, F. (1998). "Accident Prediction Models for Signalized Intersections In British Columbia", presented at the ITE 68th Annual Meeting, August 9-12, Toronto.

Zegeer, C.V. and Deen, R.C. (1977). "Identification of hazardous locations on city streets", *Traffic Quarterly* 31, pp.549-570.

Zegeer, C.V. (1982). "Highway accident analysis systems", topic panel, Philip Brinkman…[et al.], Washington, D.C.: Transportation Research Board, National research Council.