An Accident Analysis for Urban Vehicular Flow

L. Mussone , S. Rinelli

Department of Transport Systems and Mobility, Polytechnic of Milan, I-20133, Italy

e-mail: mussone@cdc8g5.cdc.polimi.it

Abstract

The paper deals with vehicular safety in urban road networks. The complexity of the vehicular flow process and accident dynamic is considerably high because of the large number of variables necessary to describe them. To obtain a good statistical significance only four branch intersections are considered for the present analysis.

Accident data were collected over a period of about two years in a medium-large city in the Northern part of Italy (with about 100.000 inhabitants) for a total amount of about 400 accidents distributed over 25 intersections. Geometrical characteristics of the intersections were collected, too: in particular, the number of lanes for each branch, the area of the intersection, traffic lights the presence of stop or give way signals.

Data were processed by means of using a feedforward neural network with backpropagation learning algorithm. Input variables are eleven (daytime/night-time, area of the intersection, type of accident, road bed condition, weather, type of vehicle, type of violation and flow); output is the accident index which is related to the total accident number of each intersection.

Results show non-linear relationships of the accident index with flow, area of intersection, meteorological conditions and type of violations.

1 Introduction

A lot of models and methods have been developed to describe urban accidents: for example, macroscopic and log-linear models, discriminant and correspondence factorial analysis. A topic aspect of those studies is that of working with qualitative variables which describe the accidents.

Among macroscopic models Oppe's model (Oppe, 1988)[14] has to be mentioned. In this model two distinct functions represent the accident rates with deaths and traffic flow evolution as a long term prediction.

The first function relates the ratio between the total number of fatalities for a given year, f, and the total veh./km value, v, to the respective year, t:

$$\ln\left(\frac{f}{v}\right) = \alpha t + \beta \qquad \alpha < 0 \tag{1}$$

where α and β are scale parameters to fit. This means that the decrease of the ratio f/v is proportional to time.

The second function is the S-shaped logistic and relates the already realised traffic, v, at time t to the volume potential to be realised in the future, v_{max}. A generalisation of this formulation results in the model:

$$\ln\left(\frac{v}{v_{max} - v}\right) = \alpha' t + \beta' \qquad \alpha' > 0 \tag{2}$$

where α' and β' are scale parameters to fit. The assumption is that the ratio between the actual traffic volume and its variation increases proportionately to time.

Another technique is discriminant analysis which is used to investigate how speed limit violation is influenced by a large number of variables (Kanellaidis et al., 1995) [7]. These variables can be grouped into three categories related to driver (experience and behaviour) and vehicle characteristics. Discriminant analysis extracts the variables which are important to distinguish between groups and its formula can be used to analyse new cases. The formulation is established by the linear discriminant function, one for each group, and has the following formula:

$$D_i = B_0 + B_{j\,1} X_1 + B_{j\,2} X_2 + ... + B_{j\,j} X_j \tag{3}$$

where D_i is the i-th disrciminant function; B_{jj} is the coefficient of the discriminant variable J for the function i, and X_j is the discriminant variable j. It must be noted that this formulation supposes a linear dependence of variables for the whole range of their variability.

Another study (Jadan and Nicholson, 1992) [6] is based on stratification of accidents occurred in each link and then on calculation of the best stratification. The aim is to identify significant relationships among variables such as amount of travel, certain combinations of carriageway and land-use types.

A different approach (Retting et al., 1995) [15] is based on the analysis of crashes to single out the potential countermeasures. To determine the gravity of an accident a particular scale (KABCO) has been used and its acronym is K for killed, A for incapacity injury, B for non-incapacity injury, C for possible injury and O for uninjured.

Other studies concern the evaluation of risk factors related to user behaviour (Shibata and Fukuda, 1994) [17], (Barjonet, 1995) [2] or age and experience of drivers (Levy, 1990) [11].

Road design is involved when it can induce a change in driver behaviour. Traffic regulation and calming are strategies which lead to reducing speed and enable the driver increase risk perception. Some techniques (Institute of Transportation Engineers, 1989) [10] are related to the use of woonerf, chokers, cul-de-sac, artificial humps, heightened intersections, and others.

2 Data Collection

Data were collected by the city police in a city in the neighbourhood of Milan over a period of two and a half years, from March 1993 to November 1995. Data refer to accidents that occurred from six a.m. to 12 p.m. because the city police works only in this range.

First the database is ordered according to the geometrical type of intersections (AASHTO, 1984) [1]: three-leg (unchannelised T, flared T, T with turning roadways, unchannelised Y), four-leg (unchannelised, flared and channelised), multileg and rotary intersections. For the present study only four-leg unchannelised intersections are considered because is the type most represented: resulting in 533 records. Intersections with the most significant number of accidents are extracted from that set leading to a database of 402 accidents subdivided over 25 intersections. The compromise of choosing more than one intersection is necessary because a single intersection is not sufficiently significant due to the relatively small number of accidents so it is necessary to group similar intersections together.

The fields of the database are: day, time, type of traffic control (signalisation or stop and give way), number of accidents of the intersection, area of the intersection, type of accident, type of violation, type of road bed, weather conditions, traffic flow, type of vehicles involved in the accidents, and sex of drivers.

The area of the intersection refers to the area of manoeuvre for the two interacting flow streams identifying all the possible points of vehicular conflicts. Compatibility between flow values and area has been calculated according to the Transportation Research Board (1985) [18] formulation both for signalised and not signalised intersections.

The type of vehicle subdivides vehicles into four categories: two wheels vehicles (D), low powered cars under 1600cc. (U), high powered cars over 1600cc. (C) and heavy vehicles (H).

The type of violation also identifies the regulamentary regime of the intersection (traffic light, stop or give way signal presence) when the accident occurred.

The number of accidents for the intersection, k_i, is used, according to (Nicholson, 1995) [13], to calculate the accident index which is:

$$s_i = k_i / M \qquad (4)$$

where M is equal to the number of accidents on that intersection with the highest number of accidents. It must be underlined that M is only a scale factor and therefore results must be used comparatively and not in an absolute way.

The ‘sex of driver’ variable is not used now in order to reduce the number of input variables of the model. Flow on each branch is calculated by a program which assigns O/D matrix to the urban network. The matrix has been prepared for the urban traffic plan purposes in due course during the accident data collection.

3 The Neural Network Model

3.1 Theory

The definition of a neural network model depends on the identification of the relationship between input and output for a limited number of known cases, so as to infer output values for new cases. For this purpose, starting in the 1980s, research in the field of neural nets has considerably increased. Many different models are now included under this heading, but this paper refers only to feedforward nets. These models are referred to without inferring anything from their physical characteristics, using a "black-box" approach.

Many theorems state that multi-layered feedforward neural nets (with at least one hidden layer), by using neuron transfer functions of a sigmoidal type and linear input combinations, can approximate any function which belongs to L^2 with a small margin of error (Cybenko, 1989) [3], (Hornik, 1991) [5].

The backpropagation technique (BP) applies only to multi-layered feedforward architectures, and is a heuristic solution to the training problem in feedforward nets, which, it should be remembered, is an NP-complete problem. The advantages of using neural networks rather than classical analytical methods in solving complex problems of physical processes have been well known since the 80s (Dougherty, 1995) [4].

As in a previous work (Mussone et al. 1995) [12] the approach is to reconstruct by means of neural nettwork approximation capability the relationships between some parameters (input data) and an accident index (output data).

3.2 The Model

Data are divided at random into two sets, the learning and the test sets which have the same number of elements. The accident index coefficient which ranges from 1 to 50, is normalised to 1. Input variables are 11 but the number of input neurons is 22 because some variables need more than one neuron to be classified. Almost 40 different architectures of network models are created and learnt through the Neural Network toolkit of Matlab and backpropagation learning, with a different number of hidden layers and neurons, different transfer functions (hyperbolic tangent, sigmoidal, sinus) and iterations. The best network in the sense of performance has only 1 hidden layer and 4 hidden neurons (Fig. 1).

The input, hidden and output layer have a linear, tangent hyperbolic and sigmoidal transfer function respectively. The number of learning iterations is 15,000 and the RMS error calculated on output by means of using the test and train set is 0.12699.

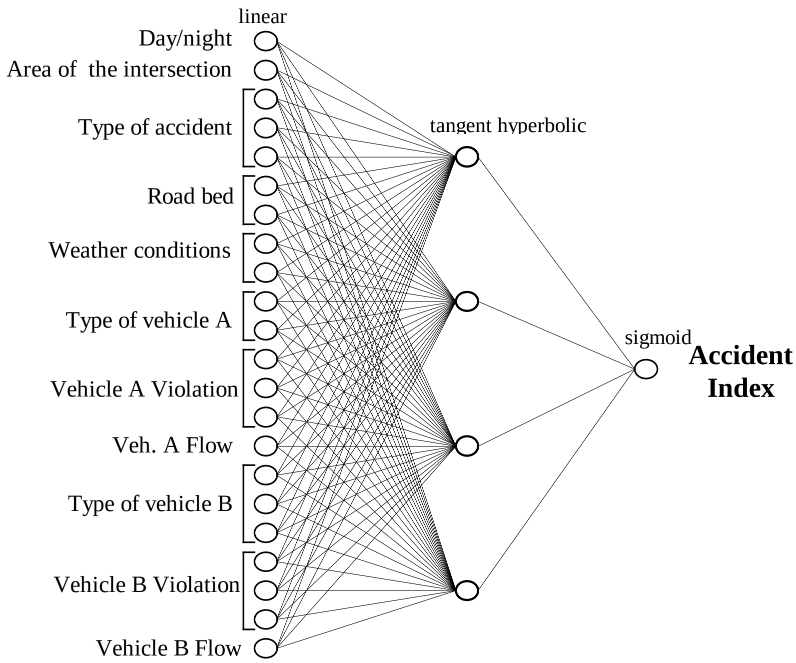

Figure 1: The optimal neural network model

4 Results

Data in figures 2 to 6 are obtained by providing the neural network model with some situations and observing its output. These cases are only a few of those possible and limited to those which the authors consider more interesting. The resolution of area of the intersection (100 m^2) has been chosen to reduce the number of drawn points without loosing in significance. There is nothing to prevent one drawing more points and obtaining smoother functions.

In all figures the Y axis identifies the accident according to eq 5). In figures from 2 to 5 the X axis is the area of the intersection and curves are drawn changing the flow of vehicle A and B, in some cases (Fig, 2 and 3) fixing the flow of vehicle B (which committed the violation, if any) and varying that of vehicle A (in legend) in the other cases (Fig. 4 and 5) varying both. This series of figures illustrates the accident index trend when different scenarios are considered: signalised or not signalised intersection, dry road bed and calm or rainy weather and type of accident, rear-side or collision.

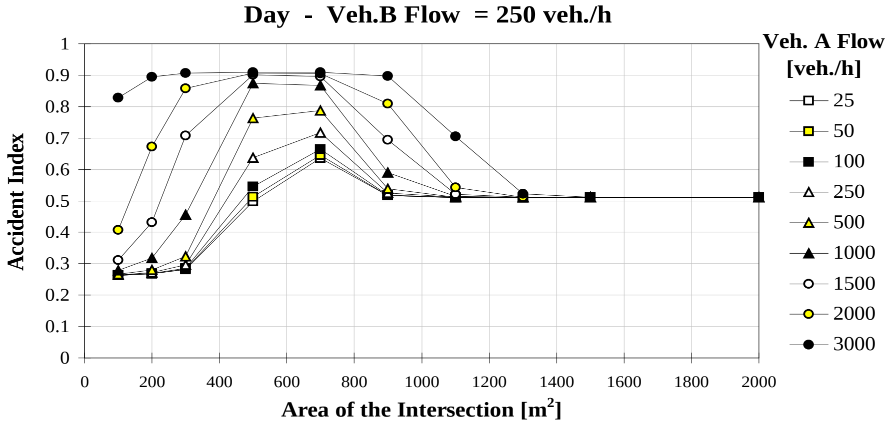

Figure 2: Rear-side accident in a “give way” intersection with a dry road bed and calm weather.

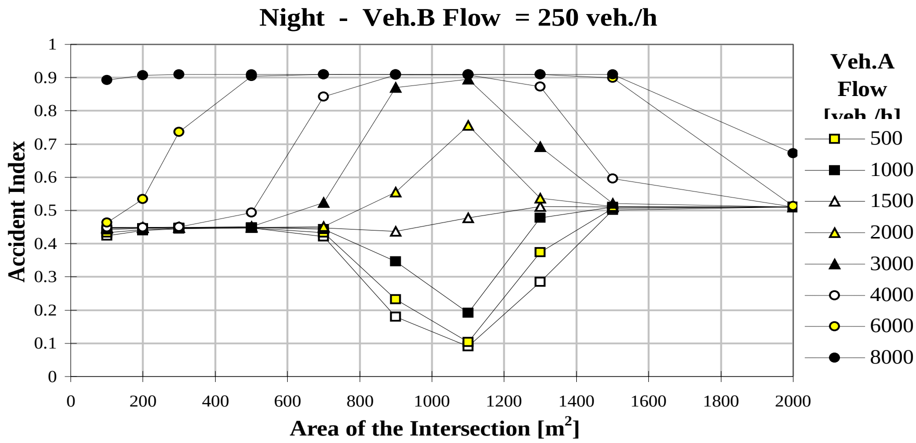

Figure 3: Rear-Side Accident for a signalised intersection with dry road bed and calm weather

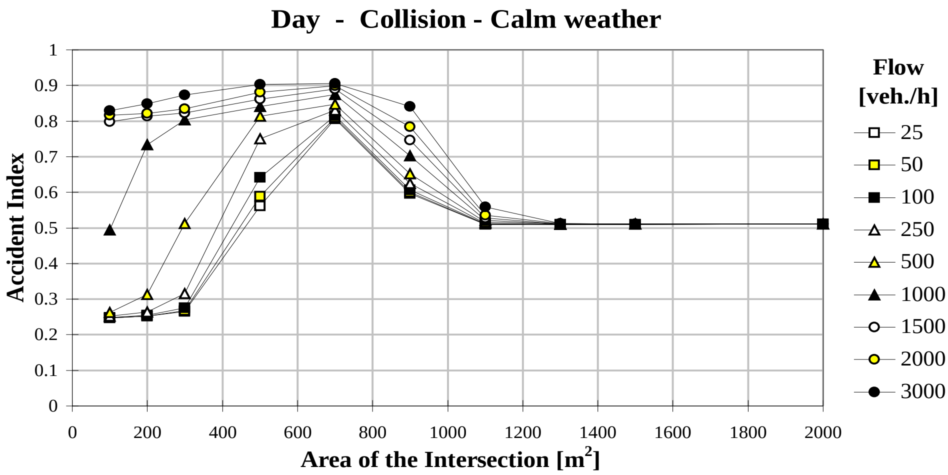

Figure 4: Collision with dry road bed and calm weather.

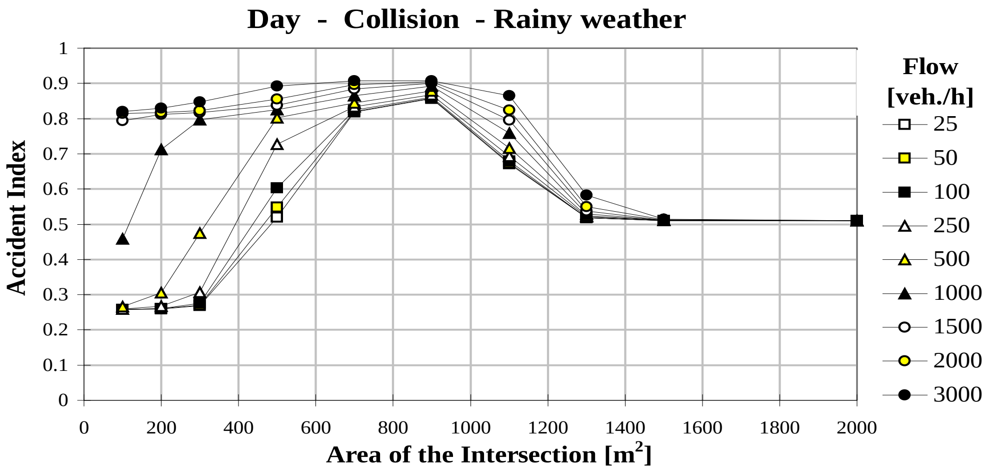

Figure 5: Collision with wet bed and rainy weather.

The first significant result is the influence of the parameter area of the intersection in determining the accident index values. In all figures accident index exhibits a maximum (or a minimum in Fig. 3) for a particular (and critical)value of area. This value changes according to changes in the scenario but it is somewhat limited in the interval 700÷1100 m^2. This maximum (or minimum) changes if different values of flow for vehicle A and B are considered. For a fixed value of flow for vehicle B, increasing flow for vehicle A means to increase the accident index and to widen the interval of maximum (Fig. 2, 3, 4 and 5). Also by increasing flow for vehicle A the maximum is reached for lesser values of area, meaning that the higher the flow the lesser the critical area.

The comparison between signalised and non-signalised intersections gives another significant result. When an intersection is signalised the accident index decreases and reaches its minimum for the critical value of the area. This effect disappears when flow for vehicle A becomes very high (over 1500 veh./h).

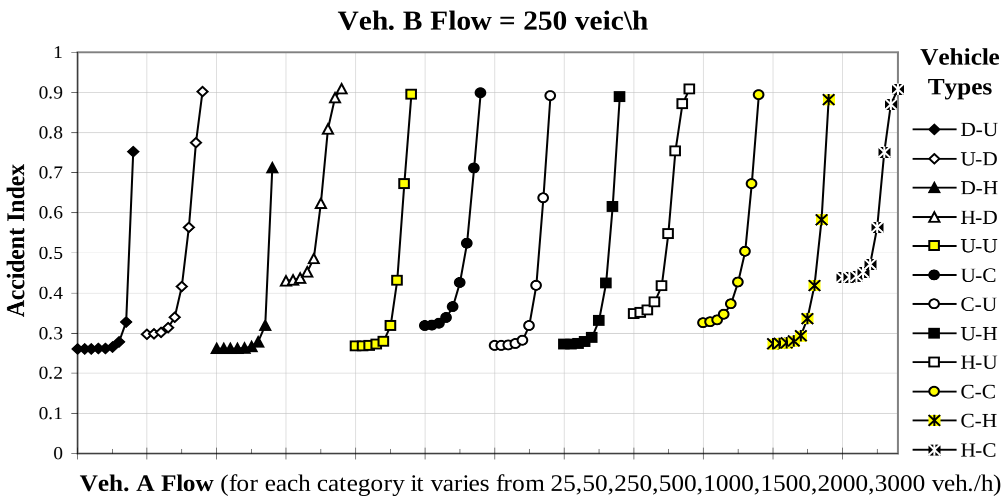

Figure 6: Rear-Side accident in a “Give way” intersection for different types of vehicle couples involved in the accident.

Daytime and night-time exhibit different trends and night-time has almost always higher accident index values than daytime and also different critical areas.

The weather conditions affect the trend and (as on night-time) bad weather conditions lead to higher accident index values; the critical area is wider, too (Fig. 4 and 5).

Difference are shown between figures considering the type of collision (Fig. 2 and 4) though in this case it is more difficult to identify a particular trend.

The last figure (Fig 6) shows how the accident index changes when the couple of involved vehicles changes (for legend see in chapter 2): it is assumed that the first vehicle of the couple committed the violation. For each couple, the flow of vehicle A varies discreetly from 25 to 3000 veh./h. It is interesting to note that the most dangerous conditions involve heavy vehicles (H-D, H-U and H-C couples) which did not commit the violation. An explanation may be that heavy vehicles have a somewhat different speed from that of other vehicles which tend to underestimate it and commit the violation leading to the accident.

5 Final Remarks

This paper is the first approach to urban accident analysis using an artificial neural network. A problem in this approach (like in any modelling problem) is to identify the most significant parameters and to synthesise some complex aspects into one variable.

In this sense the paper shows the positive significance of flow values, weather condition, type of involved vehicles, type of control and area of the intersection. In particular a critical area is shown in which the accident index reaches a maximum value. Signalised intersections show critical areas but with a minimum accident index confirming that signalisation may improve safety: there are some exceptions when flows are very high (greater than 2000 veh./h) or very low (less than 400 veh./h).

A neural network appears well suited to implement the model which suffers only in general terms because it is not known how these results can be applied to other intersections not used in the sample. For this reason future work will deal with a greater accident database in such a way as to consider not only four branch intersections but all other types of intersections.

The final aim of the work is to write a set of tables to calculate the level of safety of an intersection to use both for planning and for regulation.

Acknowledgements

Thanks are due to the students G.M. Lazzarini e P.G. Lugaresi who, for their degree, actively co-operated in preparing the data elaborated in a later stage by neural networks and organising the results.

References

1. American Association of State Highway and Transportation Officials (AASHTO) A policy on geometric design of highways and streets. AASHTO, Washington, D.C., 1984.
2. Barjonet P.E (1995) European drivers and traffic safety. 7th WCTR, 23-25 July, Sydney 1995.
3. Cybenko G. (1989). Approximation by superpositions of sigmoidal functions, *Mathematics control, signals and systems* , **2**, 303-314.
4. Dougherty, M. (1995) A review of neural networks applied to transport. *Transpn Res.-C*, 1995, **4**, 247-260.
5. Hornik K. (1991) Approximation capabilities of multilayer feedforward networks. *Neural Networks* **4**, 251-257.
6. Jadan K.S., Nicholson A.J. (1992) Relationships between road accidents and traffic flows in an urban network, *Traff. Engng Control*, September 1992, pp. 507-511.
7. Kanellaidis G., Golias J., Zarifopoulos K.(1995) A survey of driver attitudes toward speed limit violations. *Journal of safety research*, Vol. 26, No. **1**, pp.. 31-40, 1995.
8. Kim K., Nitz L., Richardson J., Li L. (1995a) Personal and behavioural predictors of automobile crash and injury severity. *Accident analysis and prevention*, Vol. 27, No. **4**, pp. 1-13, 1995.
9. Kim K., Nitz L., Richardson J., Li L. (1995b) Analyzing the relationship between crash types and injuries in motor vehicle collisions in Hawaii. *Transpn. Research Record* **1467**, pp. 9-13, 1995.
10. Salvatore F. (1992) trattamento statistico dei dati di incidente. *Autostrade*, 1992.
11. Shibata A., Fukuda K. (1994) Risk factors of fatality in mototr vehicle traffic accidents. *Accident analysis and prevention*, Vol. 26, No. **3**, pp. 391-397, 1994.
12. Transportation Research Board (1985) *Highway Capacity Manual*. Special Report **209.**