An ANN model to study driver behaviour in urban environment

L. Mussone, G. Reitani, S. Rinelli

Department of Transport Systems and Mobility, Polytechnic of Milan, I-20133 Milan, Italy
Department of Land Engineering, University of Pavie, I-27100, Pavie, Italy

Abstract

The aim of the work is to set up a model of driver behaviour in certain specific urban traffic situations. Urban traffic is extremely variable and during the rush-hours is particularly sensitive to incorrect driver behaviour.

The proposed model describes behaviour in the presence of prescriptive signals such as traffic light signals. For this purpose a data field detection has been carried out in a city of Northern Italy during a period of about six months and a database of this data has been built. In particular red light and not giving way violations have been detected.

This model allows us to describe number of violations, subdivided into red light and not giving way violations, for each type and amount of traffic flow, for each phase duration of traffic light and for meteorological and day or night conditions.

An artificial neural network is use to carry out the model and a feedforward architecture with a backpropagation learning technique was used.

1 Introduction

Driver behaviour is a topic issue in traffic regulation and control: to understand how drivers react to some specific prescriptive signals it is important to improve the control chain of which the weakest link is precisely driver behaviour. This situation is due to difficulties in acquiring information and consequently in defining a model.

Simulation by computer tends to reduce the complexity of the algorithm by eliminating from the model irrational or random behaviour: for this reason it is not always satisfactory. Video techniques have recently been applied in field detection but they are still either expensive or unpractical in large scale...
The relation between environmental conditions and their effects on traffic flow, in particular on individual choices, can be seen as an interesting field of investigation. It is therefore chosen to collect data on traffic flow and reactions to outside conditions where it was easiest, i.e. at all access points to a certain number of intersections with traffic lights, as described in a previous work (Mussone, Reitani and Rinelli [11]).

The seventeen variables taken into consideration in building the model are related to: time of day (15 minute intervals), type of vehicle, weather conditions, type and number of violations, w/c ratio and vehicle flow. Other parameters have been detected such as sex and age of drivers and number plate but they are not used in this model.

In defining the model it is assumed that violations committed by drivers may be expressed as a non-linear function depending on time, w/c ratio, vehicle type, weather condition, data containing traffic flow variables.

2 Actual practice in driver behaviour

Research in the field of modelling and analysis of driver behaviour is now been carried out but at the moment there is not much material compared to the complexity of the problem. This field is undoubtedly of great interest for all applications of transport and particularly for real time control and planning of traffic flow. Advanced information systems for travellers (ATIS and DIS projects) is one aspect that has been explored by a number of researchers: the aim of these studies is to guide the user in his choices and make both transport systems and general mobility more efficient (Bhat et al. [3]), (Conquest et al. [4]), (Ben-Akiva et al. [2]).

In the field of the evaluation of user information, VMS or RDS, other studies compare lab experiments with field results (Schofer et al. [13]) or investigate human factors like ability, experience, physical condition, age, education (Dingus and Hulse [7]), visual conditions in particular are analyzed in (Armstrong and Upchurch [1]) and descriptive parameters of driving in (Delhomme [6]). Even fewer are studies on driver behaviour analysis using artificial neural net models (Dougherty and Joint [8]) and (Yang et al. [16]).

3 Data collection and analysis

The intersection taken into consideration in carrying out the model is a very busy one, especially during rush hours. It is already described in a previous work (Mussone, Reitani and Rinelli [11]). Formally it is an intersection of two roads although in fact there are more than four branches due to the implementation of the one way system. Width of access carriages ranges between 6 and 7.5m (19 and 25 ft). Surveys were taken only on statistically significant days, in spring 1994, in time slots during the morning, afternoon and evening rush hours and in different weather conditions.

Two distinct types of survey sheet were used: the first to note traffic flow
and composition in fifteen minute time slots; the second to note what violations were being committed and, in each case, the main characteristics of the driver and vehicle involved.

Three different types of violations were taken into account: crossing a red light, turning in a different direction from the one indicated by the lane markings, not giving way when required. The last two have been grouped subsequently in a single group ("not giving way") because they are not significant separately. Multiple violations by the same driver were noted and included in the results.

Data collected on the sheets were transferred to a spreadsheet. It includes reference coordinates, traffic flow and number of violations, classified by type of vehicle and type of violation. This sheet also gives an idea of the size of the database: 20,829 vehicles observed, 1223 violations noted with an average of 0.06 violations per vehicle.

The set of input data includes 5,760 different combinations of parameters, obtained by combining the ten classes of vehicle types defined with all the values of the other parameters such as weather conditions (three cases), environmental conditions (two), w/c ratio (four), traffic flow (six), survey stations (four).

4 Neural Network Application

Many different models are now included under the ANN heading, but this article refers only to feedforward and recurring nets. They can be used to solve identification problems, underlining any difficulty or problem with the algorithm, as is usually done with more traditional statistical techniques (Sjöberg et al. [14]). These models are referred to without inferring anything from their physical characteristics, using the "black-box" approach. It should be said that the attempt to deduce anything from the values of the connections appears a long and difficult operation which gives few useful results.

A lot of 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. The number of hidden neurons, and the number of layers needed to obtain the desired approximation is still being studied. Cybenko's (Cybenko [5]) formulation is the first and best known as investigated in a previous work by the author (Mussone [10]).

The learning paradigm used in learning nets was backpropagation (BP), rediscovered on many occasions and again recently (Rumhelart et al. [12]): it is a heuristic solution to the training problem. Many authors, such as (Weiss et al. [15]), (Sjöberg et al. [14]), (Masters [9]) underline the difficulties of training, and in particular the problems of overfitting or overtraining which adversely affect the performance of the neural net.

The artificial neural network the authors used, is a three layer structure (one hidden layer). Input to the model is an array of eleven data. Output is an array of numbers that represent the normalized number of expected violations, "red violations" and "not giving way". Both the input and the output data
referred to a fifteen minute interval.

The model is thus able to elaborate any potential situation, even if it has not actually been monitored during surveys, because, by suitably varying the input parameters, it is possible to create any scenario that may occur.

In the model used in the following paper a sample of 92 cases was used, divided at random into test and training sets, both comprising 46 elements. Great care has been dedicated to random extraction due to the relatively small size of the sample. In fact it was necessary to have test and train data sample that were as homogenous as possible. All data was normalized to the maximum value of the variable: for the number of violations the normalization factor was 25. The optimum network (in the sense of performance) is that shown in Fig. 1 which was obtained after 40000 learning iterations.

The models with only one hidden layer almost always give better results than those with two. The optimal transfer function for the hidden layer is the sigmoidal such as the output transfer function. Because the model has more than one output, error, absolute and rmse, is averaged on two outputs to give the total error to compare performance among models.

5 Results

The model proposed in this paper allows a clear and complete reading of driver behavioural processes at an urban intersection. So in the following discussion only some relevant results are shown.

Driver behaviour seems to be basically affected by flow value and phase duration (w/c ratio). The number of violations generally increases as flow increases or w/c ratio diminishes in each meteorological condition and for each branch of the intersection. It has to pointed out that the two type of violations have very different values in the same condition of flow (Fig. 2 ).

To investigate model characteristics some flow compositions have been considered: flow has been classified into 10 classes according to percentage of
vehicle composition. The first class has 65% of cars and the tenth class has 95% of cars; intermediate classes have an increasing value of percentage of cars from 65 to 95% and different percentages of other types of vehicles.

**Red light Violation - Type 1 (Day, Flow Class 1)**

**Not giving way Violation - Type 2 (Day, Flow Class 1)**

Figure 2: A comparison between the two types of violations as function of flow and meteorological conditions.

The relationship between flow versus violations strongly depends heavily on w/c ratio: when this ratio decreases, the slope of this curve increases and the maximum number of violations is reached for a lesser value of flow. In other words when w/c ratio decreases the number of violations is relevant also when the value of flow is low: for example with w/c = 0.2, day and flow class 1, the number of violations per hour is near 40 already for a flow of 800 veh./h (Fig. 2 (a)).

Flow composition plays an important role in behaviour: in particular when flow composition becomes more homogeneous (class 10) number of violation of
type 2 increase. Violations also increase when meteorological conditions are worse, both during day and night and both for type 1 and type 2 violations. It appears surprisingly, that when driving is more difficult because of rain or reduced visibility, the number of violations increases. There is on the other hand, a negligible difference between driving during day and night, even though for some combinations of input parameters the number of violations increases by very little.

It is interesting to point out that type 2 violations are almost irrelevant when flow is low but when flow is high they are more numerous than type 1 violations.

A great number of vehicles crosses red lights: the asymptotic value is more than 50 violations per hour when w/c ratio is 0.2, flow is higher than 1500 veh./h, and especially when limits to circulation are imposed by adverse

![Red light Violation - Type 1 (w/c = 0.6, Day)](image)

![Not giving way Violation - Type 2 (w/c = 0.6, Day)](image)

Figure 3: A comparison between the two type of violations as function of percentage of cars and flow.
meteorological conditions. In these conditions if the total of violations (including all type 2 violations) is considered, incorrect driver behaviour concerns 6% of total flow. These results and particularly the importance of w/c ratio show how attention must be paid in timing design to improve flow safety.

Another series of figures shows the number of violations subdivided into the two types of violations when meteorological conditions vary. In each figure there are 6 curves, one for each flow class. In general, for both types of violations (1 and 2), when meteorological conditions become worse, the number of violations increases when flow is high and it diminishes or stays constant when flow is low. Differences among the cases are more evident during the day than the night: the most significant increases are observed when calm and cloudy conditions are compared; the most significant reductions when cloudy and rainy conditions are compared.

Figure 3 shows the relationship between the number of violations versus flow classes (i.e. percentage of cars) in other words versus a more homogeneous flow composition. Type 1 violations decrease as flow class increases (high car percentage) especially when the w/c ratio is high (w/c=0.8) and when meteorological conditions are worse. When flow value is high (>1000 veh./h) this tendency changes (Fig. 3 (a)). Type 2 violations increase more rapidly when flow is high especially when it is calm; when flow is low they tend to decrease (Fig. 3 (b)).

Figure 4: Red light violations as function of w/c and meteorological conditions.

It is possible to assert that in calm weather drivers are guilty of more violations of type 2 as flow class increases; while when it is cloudy or rainy the number of these violations is almost insensitive to percentage of cars and it is almost constant. A low percentage of cars (low flow class) generally produces a higher number of violations with some exceptions: when flow is high, w/c ratio is low, type 1 violations don't depend on flow composition.

High w/c ratio (Fig. 4) represents a good point of reference in ensuring a more correct driver behaviour until flow increases to near saturation value and it determines in itself the incorrect behaviour.
References


