Artificial intelligence techniques to predict urban traffic indicators
Lorenzo Mussone
Department of Transport Systems and Mobility
Polytechnic of Milan, P.za L. Da Vinci, 32, 20133 Milano
e-mail: mussone@cdc8g5.cdc.polimi.it

Abstract

The paper deals with the application of neural network techniques to the supervision of urban traffic. This field is strictly related to the evaluation of vehicular traffic flow and its time prediction which are the main object of numerous international researches.
To carry out a traffic control by using traffic signal control it is better to refer to a flow state rather than flow values consequently it is a good idea to define a set of rules to evaluate flow states.
The synthesis introduced by the definition of state allows to explain the complex relationships describing flow with an immediate functional classification easy applicable to a road link or a more complex urban network.

1 Introduction

The field of urban flow network control and evaluation has been tackled by more authors with different aims and methods; a quite comprehensive survey is in (Bielli et al., 1994) [3]. From the future directions of traffic light control proposed by (Bell, 1992) [2] some years ago, it was possible to understand traffic control should be applied to wide urban network; now this concept is generally accepted and control strategies apply to network level and no more to single intersections. In spite of this there are no works on criteria definition or on flow state classification to evaluate the state of the network.

The change in perspective, from single intersection to wide networks, is unavoidable when interactions between road branches become strong and their mutual effects can not be neglected. New road capacity constraints, such as acoustic and atmospheric pollution ( whose causes and effects are not bounded in space) confirm more the network approach. First works in traffic light control at network level are UTOPIA (Mauro and DiTaranto, 1989) [15], (Sauthier and

All instruments allowing prediction of flow traffic are strictly connected to control methods. Among works on short time prediction of urban flows there are (Hamed et al., 1995) [11], (Dochy et al, 1994) [6] and (Clark et al., 1993) [4]. (Hamed et al., 1995) [11] carries out a 1 minute prediction by using the Box-Jenkins ARIMA model with different orders. The interesting aspect consists in the possibility to describe flow by a model which required a continuous (or linearized) process. This is true when the urban network is not congested such as that referred to the paper (the city of Amman in Jordan).

In (Dochy et al, 1994) [6] and (Clark et al., 1993) [4] neural networks are used to predict flow at short time (1 hour) with particular care to model performance evaluation. The comparison of neural network approach with conventional ones (such as Box-Jenkins model) is tackled in (Clark et al., 1993) [4] where qualities and deficiencies of two approaches are pointed out.

Neural network role in transport, particularly for the specific field of traffic control, is further discussed with a greater extent in (Dougherty, 1995) [7] and (Mussone, 1995) [16]. Other approaches deals with prediction by using statistical methods to factorize the different components present in traffic flow (Aldrin, 1995) [1]. The aim of these approaches is to describe flow variations with the local, daily, weekly, seasonal and holiday day components.

Another interesting work is (Ghali e Smith, 1994) [9] dealing with the problem of defining indicators to evaluate performances of control systems. At last, in (Roberg, 1994) [17] a simulation model is proposed to study dispersion of congestion in wide area urban network.

2 The data collection scenario

The cases to which to apply the proposed methodology have been singled out first from cities in which traffic signal control based on intelligence distributed architecture systems are installed, and where strategy control may be on local or central CPU.

In the following list the systems taken into account with their dimension and field surveys made in 1995 are described:
- **Morbegno**: 29 flow collection sections, each with two magnetic loops (for speed and vehicle type classification), linked to 3 dynamic traffic control systems with calculated traffic light plan and working in a unique traffic basin;
- **Padova**: about 140 flow collection sections, a lot of which with two magnetic loops. Traffic data are transmitted to 47 dynamic traffic control systems with calculated traffic light plan. Control is exerted on 4 basins or strong interaction areas;
• **Parma**: 76 flow collection sections, some of which with two magnetic loops, linked to 29 dynamic traffic control systems with calculated traffic light plan and working in a unique traffic basin (strong interaction area);

• **Piacenza**: 76 flow collection sections, each with two magnetic loops linked to 3 dynamic traffic control systems with calculated traffic light plan and working in a unique traffic basin;

• **Terni**: 65 flow collection sections, each with two magnetic loops, linked to 20 dynamic traffic control systems with calculated traffic light plan and working in a unique traffic basin.

Some surveys on the urban networks allowed us to understand the real structure of these monitored traffic networks. After some preliminary surveys to build up flow relationships, it clearly results that for the networks of Parma, Piacenza and Terni cities which have only one traffic basin it is not easy to define traffic directrices and strong interaction areas.

On the other hand, the network of Padua city, divided in different traffic basins according to its traffic directrices, and that of Morbegno city, with only one traffic directrix, allow us a better identification of traffic basin characteristics for which a more linear shape of flow curves in respect of other urban situations is expected.

The choice of focusing experimentation on the network of Morbegno city follows from the previous considerations and particularly from two aspects:

• the prevailing traffic is characterised by movements using the central axis (the traffic directrix, S.S.38) to cross the city and only low traffic local roads insert on this axis; this allows us to well study the directrix and the spatio time distribution of flow;

• the system has a lot of magnetic loops for traffic data collection in respect of the length of the network; particularly, a lot of sensors are installed also on exit lanes of intersections and this allows us a greater accuracy in the flow data collection phase.

### 3 The methodology

Methodology used to work out the models is based on artificial neural networks, more precisely on feedforward neural networks with backpropagation learning paradigm. This type of neural network is well known in technical papers and a lot of theorems, the first and the most famous is in Cybenko’s paper (Cybenko, 1989) [5] and in other papers (Hornik et al., 1990) [12], (Hornik, 1991) [13], (Girosi et al, 1991) [10], (Leshno et al, 1993) [14], state the capability of approximating with a little error any function belonging to $L^2$ space. Applications into transport, planning and control fields are numerous and well studied (Dougherty, 1995) [7], (Mussone, 1995) [16].

Essentially the methodology to build up models consists of four phases. The first consists in a random extraction from the whole data set (which must be
exhaustive of all necessary knowledge about the process to be modelled) a data subset the dimension of which may vary from few hundreds to one thousand data. There are no rules on the exact dimension of the subset but it must satisfy at least the law of large numbers; a very wide dimension may not lead to significant positive results in learning when indeed it may slow learning procedure. When data can be grouped into classes, that is the model should recognize a particular pattern classification, data extraction should be homogenous for each class: this imply that an equal number of data must be randomly extracted for each class.

In the second phase the data subset is randomly subdivided into two sets, the train and test set, which will be used respectively to learning and validation of the neural network.

The third and fourth phase consist in learning and testing of neural network, iteratively until the optimal network is not singled out where optimal means that it minimizes model error.

The number of inputs and outputs of models has not theoretical limits. But the relationships between input/output vector dimensions and the dimension of optimal network are not known yet, so as it is not known how many data are necessary for the learning phase. It is obvious that when the number of inputs and especially the number of outputs increases the complexity of the process increases and by consequence more complex networks and more data are necessary.

The heuristic approach in four phases previously reported tackles and solves this problem though computing times needed to work out the optimal network are quite long. If the heuristic procedure fails to consider some network configurations leading to better performance, results obtained by means of other networks remains however valid.

In practice, the heuristic procedure is carried out by using the neural network toolbox of Matlab to single out the optimal network and the best ten network configurations are in-depth analyzed by NeuralWorks (a neural network shell) to single out the best model after an exhaustive evaluation of their performance.

4 The models

Data collected in the 29 sections of Morbegno refer to four standard days and the collection time interval is 6 minutes. Data have been prepared normalizing input and output values to 1.

Two feedforward not recurrent completely connected models are worked out. The first calculates the flow state vector, the second predicts it in at one step prediction. The first model is preliminary to the second and has the task of pattern recognition of flow conditions. The second model uses the same data but outputs are time shifted by 6 minutes. It must be underlined that application of this methodology to other scenarios need only to consider a different number
of data collection sections and, therefore, a different number of inputs of the model.

4.1 The model of traffic state classification

This model has been thought to recognize flow state of an urban network after it has been classified into four increasing value classes of congestion. Flow data collected by each magnetic loop are analyzed and then classified according to their speed values. Speed intervals are calculated by statistical analysis of flow data archives considering the impact that these speeds has on the whole network.

Traffic quality indicators are classified into free, smooth, slowed or congested flow. Indicators are calculated for each of 29 points and finally a 4 dimension vector can be used to describe the state of the network.

The function used to calculate the indicator value is

\[
\text{if } v \geq v_1 \Rightarrow \text{the state is Free Flow (I state);}
\]

\[
\text{if } v_2 < v \leq v_1 \Rightarrow \text{the state is Smooth Flow (II state);}
\]

\[
\text{if } v_3 < v \leq v_2 \Rightarrow \text{the state is Slowed Flow (III state);}
\]

\[
\text{if } v \leq v_3 \Rightarrow \text{the state is Congested Flow (IV state);}
\]

where, \( v \), is the detected speed. Summing the occurrences for each state the state vector is obtained. \( v_1, v_2 \) and \( v_3 \) depend on loop position in the network.

In learning the network also flow value has been introduced though it is not used directly by the classification; it is author’s opinion that since flow is very related to speed it can help model to recognize the flow state. The number of active loops as input variable is necessary because not always all the loops are active and this can change the total number of one state.
RMS error is calculated for each state and it is 0.03216, 0.04523, 0.08968 and 0.07796, which lead to an average overall error of 0.06126 (in Tab. 1 results for each state with different number of learning iterations are reported). The optimal network has the least average RMSE but this does not necessarily coincide with the least RMSE value of each state.

**Tab.1 :** RMSE on test data for the optimal network of the first model

<table>
<thead>
<tr>
<th>RMSE</th>
<th>1 state</th>
<th>2 state</th>
<th>3 state</th>
<th>4 state</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.091871</td>
<td>0.087383</td>
<td>0.287443</td>
<td>0.290804</td>
<td>0.189375</td>
</tr>
<tr>
<td>No. learning iterations</td>
<td>10000</td>
<td>0.033943</td>
<td>0.045983</td>
<td>0.093601</td>
<td>0.074721</td>
</tr>
<tr>
<td></td>
<td>20000</td>
<td>0.032121</td>
<td>0.04462</td>
<td>0.09287</td>
<td>0.085936</td>
</tr>
<tr>
<td></td>
<td><strong>40000</strong></td>
<td><strong>0.03216</strong></td>
<td><strong>0.04523</strong></td>
<td><strong>0.08968</strong></td>
<td><strong>0.07796</strong></td>
</tr>
<tr>
<td></td>
<td>50000</td>
<td>0.029979</td>
<td>0.042177</td>
<td>0.093715</td>
<td>0.082984</td>
</tr>
</tbody>
</table>

### 4.2 The model of traffic state prediction

The second model predicts the state vector in at one step prediction. The prediction of a vehicular flow state for a traffic network is a necessary operation for any traffic control strategy. Time horizon of control is strictly related to the control strategy adopted.

In this application a 6 minute time interval ($\Delta t$) is used. In successive works performance at more steps and with longer time horizon will be evaluated.

Data to learn and test the neural network are prepared by using the same data used for the previous model: to each record (at instant t) the state vector of the following record (at instant t+ $\Delta t$) and the number of working detectors is added. Records having all detectors out of use are rejected, so as those records preceding them and those at the day end, because in these cases prediction is without sense (all states are null). Records remaining are 932 subdivided into to sets of 466 records.

As in the previous model the number of active loops as input variable is necessary because not always all the loops are active and this can change the total number of one state. The actual state vector (at instant t) in input improves prediction.

RMSE of this model is slight higher than that of the previous model (0.08205 to be compared to 0.06126). Also for this model the highest error is for the third state (0.127) and for the fourth (0.116), the lowest for the first (0.038) and second one (0.048). Prediction of the third state is more difficult
both because has the highest frequencies and because speed interval to be recognized is quite narrower than the others.

In Table 2 the errors of the optimal network are reported and in Fig. 2 the optimal network scheme is drawn.

**Tab.2** : RMSE on test data for the optimal network of the second model

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 state</td>
<td>2 state</td>
<td>3 state</td>
<td>4 state</td>
<td>average</td>
<td></td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.09077</td>
<td>0.082086</td>
<td>0.19881</td>
<td>0.281221</td>
<td><strong>0.16322</strong></td>
<td></td>
</tr>
<tr>
<td>No. learning iterations</td>
<td>5000</td>
<td><strong>0.0382</strong></td>
<td><strong>0.0476</strong></td>
<td><strong>0.1267</strong></td>
<td><strong>0.1157</strong></td>
<td><strong>0.0821</strong></td>
</tr>
<tr>
<td>3000</td>
<td>0.035643</td>
<td>0.047703</td>
<td>0.128213</td>
<td>0.124011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6000</td>
<td>0.039041</td>
<td>0.049587</td>
<td>0.12521</td>
<td>0.115608</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7000</td>
<td>0.03933</td>
<td>0.047632</td>
<td>0.126607</td>
<td>0.115073</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td>0.035505</td>
<td>0.048584</td>
<td>0.128085</td>
<td>0.120248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20000</td>
<td>0.038539</td>
<td>0.05464</td>
<td>0.129355</td>
<td>0.119897</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30000</td>
<td>0.03788</td>
<td>0.056341</td>
<td>0.127983</td>
<td>0.121131</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Optimal network for the second model
(in squared brackets the number of neurons for the layer is shown).

In Fig. 3, 4 and 5 the prediction of state vector (smooth, slow and congested flow) for the test set data is shown. Free flow state prediction is not reported here because it is quite similar to smooth flow state and not significant for the discussion. When the state vector has very low values the network predicts only a mean value (in this case it means a great percentage error but small absolute error) but when the state vector becomes greater the capability is
very good (a small percentage error). The slow flow and congested flow cases show a very good degree in prediction accuracy; there is only a slight filtering effect in peak data and a little time shift.

![Figure 3: Prediction for smooth flow conditions.](image)

![Figure 4: Prediction for slow flow conditions.](image)

![Figure 5: Prediction for congested flow conditions.](image)
5 Conclusions

The advantage of the proposed methodology is that it requires no a priori hypotheses about the process under study so it adapts itself well to every types of non linearity present in a process. This property is reflected by operating method of building up models which must be necessarily based on a survey of field data of the process. Another interesting aspect to be underlined consists in using this method as explorative tool since it is quite easy to include initially in the model a lot of variables and then to choose the most significant after having analyzed the results of the model.

The application of this method to the traffic network analysis has been divided into two directions: firstly it is checked if a neural network is capable of recognizing the traffic flow state of an urban network after those states were classified according to speed values. Secondly, the most interesting problem of time prediction of flow states has been tackled: where prediction is at one step of six minutes. Results allow us to think that similar good performance can also be obtained for more step prediction or when time horizon is lengthened (with time step of 12 or 18 minutes).

A further improvement in performance and in methodology use could become by inserting in the models meteorological conditions and flow composition (percentage of heavy vehicles on the total flow). Flow composition would allow to evaluate how rerouting heavy vehicles on other paths flow conditions are bettered. Meteorological conditions, instead, allow to evaluate how the induced road capacity reduction affect flow dynamic.

Acknowledgements

Thanks are due to Prof. L. Marescotti for his profitable discussions and to Arch. V.M. Curti who prepared traffic data to be used by neural networks. This paper is funded by CNR-PFT2 contract no. 96.00183.PF74.

References

