CONGESTION PREDICTION ON MOTORWAYS: A COMPARATIVE ANALYSIS

Giovanni Huisken and Martin van Maarseveen
Department of Civil Engineering & Management, University of Twente
P O Box 217 - 7500 AE ENSCHEDE - The Netherlands
Tel: (31) 53 489 2543 - Fax (31) 53 489 4040 - E-mail: g.huisken@sms.utwente.nl

SUMMARY

The paper reports on the evaluation of the performance of various short-term congestion prediction methods, i.e. multi-linear regression, time series analysis, multi-layer perceptrons, radial basis function networks, self-organising systems, and fuzzy logic. Data were gathered through dual induction loops on a A10 motorway section in the Netherlands during a four-week period. The data consist of 1 minute aggregated time bins of volume, occupancy, speed, and both a reliability and a congestion indicator. The method's results are similar, except for multi-linear regression. Self-organising systems were omitted due to huge error production.

INTRODUCTION

Dynamic Traffic Management (DTM) measures intend to increase infrastructure capacity by making more efficient use of the infrastructure resulting in spatial and/or temporal traffic flow effects. Unfortunately, it is widespread policy to activate these measures *after* congestion has already occurred and the traffic flow has broken down resulting in a time consuming traffic flow regeneration. It is therefore important to have some knowledge about traffic conditions in advance so anticipating DTM measures could be taken to keep the traffic flow going. The objective of this paper is to give some clues regarding the best choice of methodology in the particular situation to predict short-term congestion on motorways.

METHODOLOGY

The section reports on the methodologies that have been chosen to predict congestion. All methods can be described as data driven methods, meaning that the development of the models is heavily dependent on the data set that has to be analysed. In order to compare the prediction performances of the methods they should provide results in the same quantities. The chosen quantity is a binary one: 'congestion' or 'no congestion'. It has also advantages in pre-processing the data set because the data acquisition system includes information about the state of traffic in the form of a congestion indicator. The indicator is used as a target.

Multi-Linear Regression

Multi-linear regression (1) is a fairly simple method to describe observations Y that are linearly depending on variables θ and white noise e with mean zero and variance σ^2 (Gaussian disturbances) and can be mathematically noted as

$$Y_i = \theta_0 + c_{1i} \cdot \theta_1 + c_{2i} \cdot \theta_2 + \ldots c_{ki} \cdot \theta_k + e_i \qquad (1)$$

Using the ordinary least squares estimate the parameters *c* are determined. These can then be used to estimate observations given the variables.

Time Series Analysis

The Auto Regression Moving Average (ARMA) time series analysis method (2) is widely used as a prediction method. Given F observations X_0, X_1, …, X_{F-1}, ARMA(f, g) time series processes can be written in the form:

$$X_t = \mu + a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_f X_{t-f} + b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_g \varepsilon_{t-g} \qquad (2)$$

for $t = f, f + 1, \ldots$ and with ε is white noise with mean zero and variance σ^2. The parameters μ, a_1, a_2, …, a_f, b_0, b_1, …, b_g, and σ^2 are to be estimated by deriving least squares and maximum likelihood estimators. If we rewrite (2) as

$$RX = M\varepsilon + PX^0 + \mu I \quad , \quad X = \begin{bmatrix} X_f \\ X_{f+1} \\ \cdots \\ X_{F-1} \end{bmatrix} \quad X^0 = \begin{bmatrix} X_0 \\ X_1 \\ \cdots \\ X_{f-1} \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_{f-g} \\ \varepsilon_{f-g+1} \\ \cdots \\ \varepsilon_{F-1} \end{bmatrix} \qquad (3)$$

and *R*, *M*, and *P* representing matrices containing the *a* and *b* parameters and *I* being the unity vector we can minimise (3) to obtain least squares estimation and maximise (3) to obtain maximum likelihood estimation.

Multi-Layer Perceptrons

Multi-layer perceptrons (MLPs) (3) are neural networks that generally exist of one layer of input neurons, one or more layers of hidden neurons and a layer of output neurons whereas the subsequent layers are fully connected. In mathematical terms, a neuron *k* can be described by the following equations:

$$y_k = \varphi(u_k - \theta_k) \quad with \quad u_k = \sum_{j=1}^{p} w_{kj} \cdot x_j \qquad (4)$$

where x_1, x_2, …, x_p are the input signals, w_{k1}, w_{k2}, …, w_{kp} are the synaptic weights of neuron *k*, u_k is the linear combiner output, θ_k is the threshold that can be looked upon as an external parameter, $\varphi(\cdot)$ is the activation function (a sigmoid function such as the hyperbolic tangent function), and y_k is the output signal of the neuron.

The MLP network is a supervised learning network meaning that during the training phase all inputs are mapped on desired outputs. The error i.e. the difference between the actual and the desired output is a criterion that is used to adjust the weights of the neurons iteratively so that the total error of all input-output pairs is minimised. The algorithm responsible for this method is called the learning rule and the most commonly used one is the back-propagation algorithm. The delta rule is defined by:

$$\Delta w_{kj}(n) = \eta \cdot \delta_k(n) \cdot x_j(n) \quad with \quad \delta_k = -e_k \cdot \varphi'(u_k) \qquad (5)$$

where e is the error, η is the rate of learning, and Δw represents the change of weight. However, the above-described change in weights holds only for neurons belonging to the output layer. Weights belonging to the hidden layer(s) are adjusted backwards according to:

$$\Delta w_{ji}(n) = \eta \cdot \delta_j(n) \cdot x_j(n) \quad with \quad \delta_j = \varphi'(u_k) \cdot \sum_k \delta_k \cdot w_{kj} \tag{6}$$

Radial Basis Function Networks

Radial basis function (RBF) networks (4) are supervised neural networks that seem similar to MLPs. The difference, however, can be found in the construction of the usually three-layered network. The input layer is made up of neurons with a linear transfer function. The second layer (the hidden layer) which has to be of a high enough dimension has a different purpose than that of the MLP; this can be seen in that the transfer function is not sigmoid but e.g. Gaussian. The output layer supplies the response of the network to the activation patterns applied to the input layer. RBF networks are most often used to deal with approximation problems in a multidimensional space.

Self-Organising Maps

Self-organising maps (SOMs) (5) are unsupervised neural network systems that intend to optimise their free parameters according to statistical regularities of the input (training) data and usually map them onto a two-dimensional lattice structure. After the network is tuned and new input data are offered the data are mapped onto the lattice area that has the most statistical similarity to the training data.

Fuzzy Logic

Fuzzy Logic (6) is an expansion of the classic set theory using uncertainties and probabilities. In fuzzy set theory an element has a probability *between* 0 and 1 of belonging to a certain set whereas in classic set theory it is either a member (probability 1) or not (probability 0). Fuzzy logic has emerged as a tool to translate linguistic into mathematical information and is preferably used as a control and/or decision tool.

DATA ACQUISITION

The field data set contains information gathered from a part of the outer western roadway section (southbound) of the A10 - the beltway around Amsterdam (figure 1). The data were collected from 35 traffic lane induction loops during the period from 13.38 January 7th until 23.59 February 5th 1999 through the MONICA - MONItoring CAsco - data management system into one minute aggregated time bins. The data consist of information about volume, mean speed, standard deviation of speed, occupancy, and an indication of congestion. These parameters with the exception of the latter two were given for each of three categories: vehicles up to 5 meters, vehicles with a length between 5 meters and 12.5 meters, and vehicles over 12.5 meters.

Due to technical reasons there were four gaps in the data collection: the period of 02.49 January 10th until 09.04 January 11th, the period of 02.51 until 05.58 of January 16th, the entire day of January 19th and the period of 03.17 until 08.23 of January 23rd. In the remaining data set 9.9% of data was excluded because of unreliability.

The total data set was divided into four equally sized subsets with each a specific percentage of congestion time (see figure 2). The total congestion time percentage averaged over the 4 subsets is 9,8%.

MODEL DEVELOPMENT

In order to compare the results the data set was divided into 4 subsets of equal size. The method's performance was evaluated with one subset as input when the remaining 3 subsets were used to describe or train and test the method's parameters or weights and this procedure was carried out for every subset. Finally the results of the subsets were averaged.

Input Features

The input features include volume, mean speed, occupancy and the standard deviation of speed within the 1-minute time bin which can be regarded as an indicator of the chaos of the traffic. In this research we only used data of the first category, i.e. vehicles up to 5.1 meters because of architectural consequences for the neural networks would we have used the complete data set.

The time series analysis method used temporal information of the 2 target detectors as input, whereas the other methods used spatial information. It turned out that the extra information of the additional (to the target loops) 33 dual loops did not significantly improve the performance of the MLP method. For other methods the additional information could not be used due to the limitations in computer memory capacity.

Output Features

The output features or targets consisted of binary congestion indicators of the target detectors (figure 1) and shifted in time over 5, 10, and 15 minutes in order to estimate the predictive performance of a look ahead period of respectively 5, 10, and 15 minutes. Since there are two target detectors there are 2 (detectors) * 3 (indicators) = 6 output features.

Performance Measures

The outputs produced through each model by input of each subset were combined and after using the hard limit function compared with the binary congestion indicator. If errors occurred they were categorised as *false alarm* (falsely predicting congestion) or just *error* (falsely not predicting congestion). Then the performance was measured by summing both errors and averaging them over the subsets.

RESULTS

For reasons of clarity only the final results per method per prediction horizon are given and the results per subset are omitted.

The results (see figure 3a-c) indicate that both supervised neural network methods (the MLP and the RBF network methods), the ARMA time series analysis method and the Fuzzy Logic method outperform the MLR method. The results of the SOM method were abandoned due to the huge errors that occurred and were not taken into consideration. The performance of all methods - with exception of MLR - is almost equal, however, if one method has to be chosen as a winner the RBF method is the one.

CONCLUSIONS

All methods described above have in some form been used to predict traffic conditions. It is rare, however, to find studies where the performances of these methods are compared – *if* a comparison has been done it usually is limited to 2 methods. In this study, which is part of a larger project, we compared 6 methods (effectively 5 due to large error productions of the SOM method) on a rather simple infrastructure network (a road section in one driving direction). It turned out that ARMA time series analysis, MLP neural networks, RBF neural networks, and Fuzzy Logic gave similar results and outperformed the MLR method.

Subsequent project studies will involve method comparison on more complex infrastructure networks to find out if these networks will produce similar results.

One remark has to be made: it concerns the use of these methods in case of an incident. It is very likely that the performance of all methods will worsen. It is therefore advisable to use an incident detection method parallel to a congestion prediction method.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the University of Twente for providing financial funds and the Transport Research Centre residing under the Dutch Ministry of Transport, Public Works and Water Management for providing the field data.

REFERENCES

(1) P.J. Bickel and T. Jackson. (1977) Mathematical Statistics: Basic Ideas and Selected Topics, Holden-Day, Inc., Oakland CA.

(2) G.E.P. Box and G.M. Jenkins. (1970) Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, USA.

(3) J.L. McClelland and D.E. Rumelhart. (1986) Parallel Distributed Processing, Volume 1, MIT Bradford Press.

(4) M.J.D. Powell. (1988) Radial basis function approximations to polynomials, in: *Numerical Analysis 1987 Proceedings*, pp. 223-241, Dundee, UK.

(5) T. Kohonen. (1982) "Self-organized formation of topological correct feature maps", *Biological Cybernetics* **43**, 59-69.

(6) L.A. Zadeh. (1965) "Fuzzy sets", *Information and Control* **8**, 338-353.

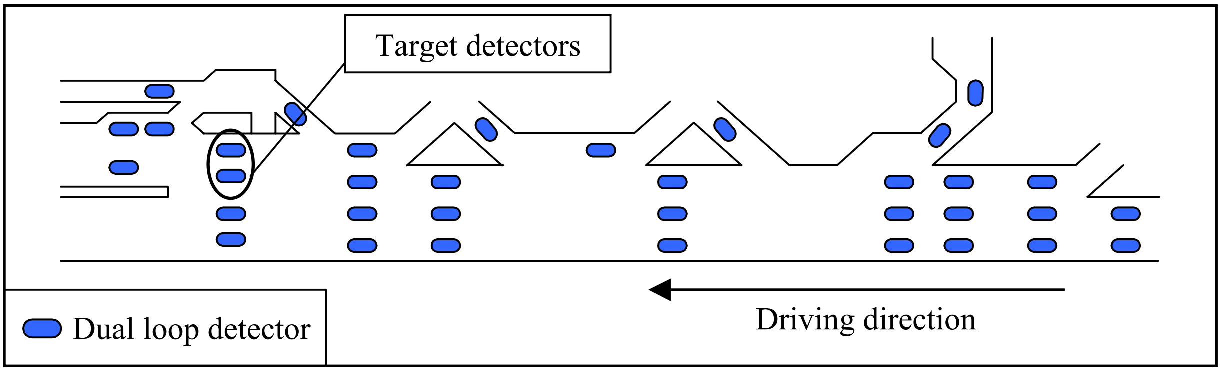

Figure 1: Schematic overview of the data acquisition site, part of the A10.

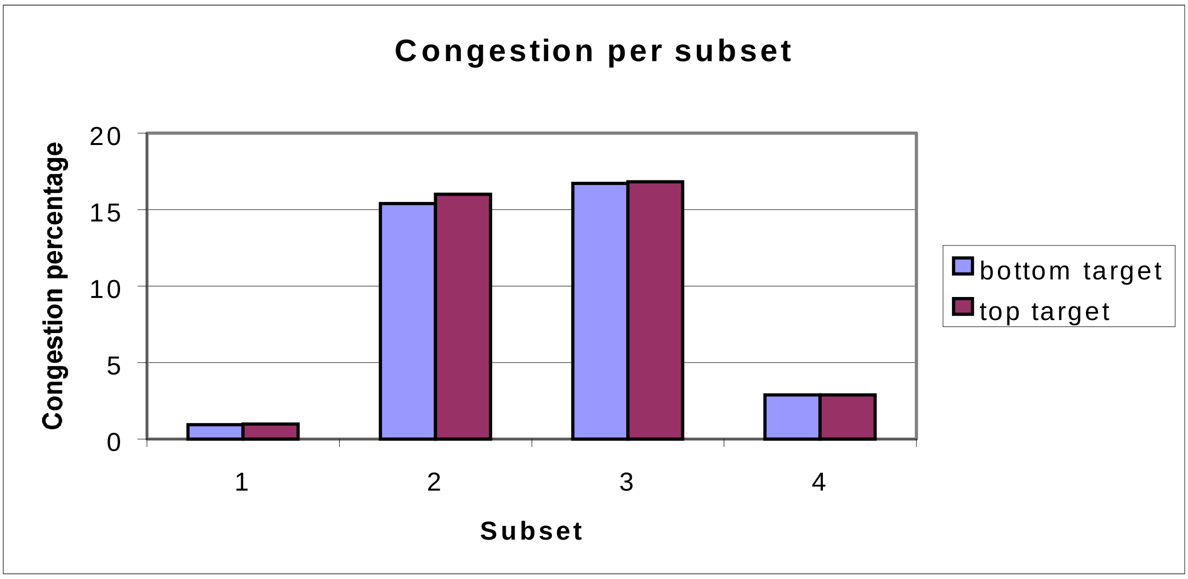

Figure 2: Congestion percentage (in time) per subset.

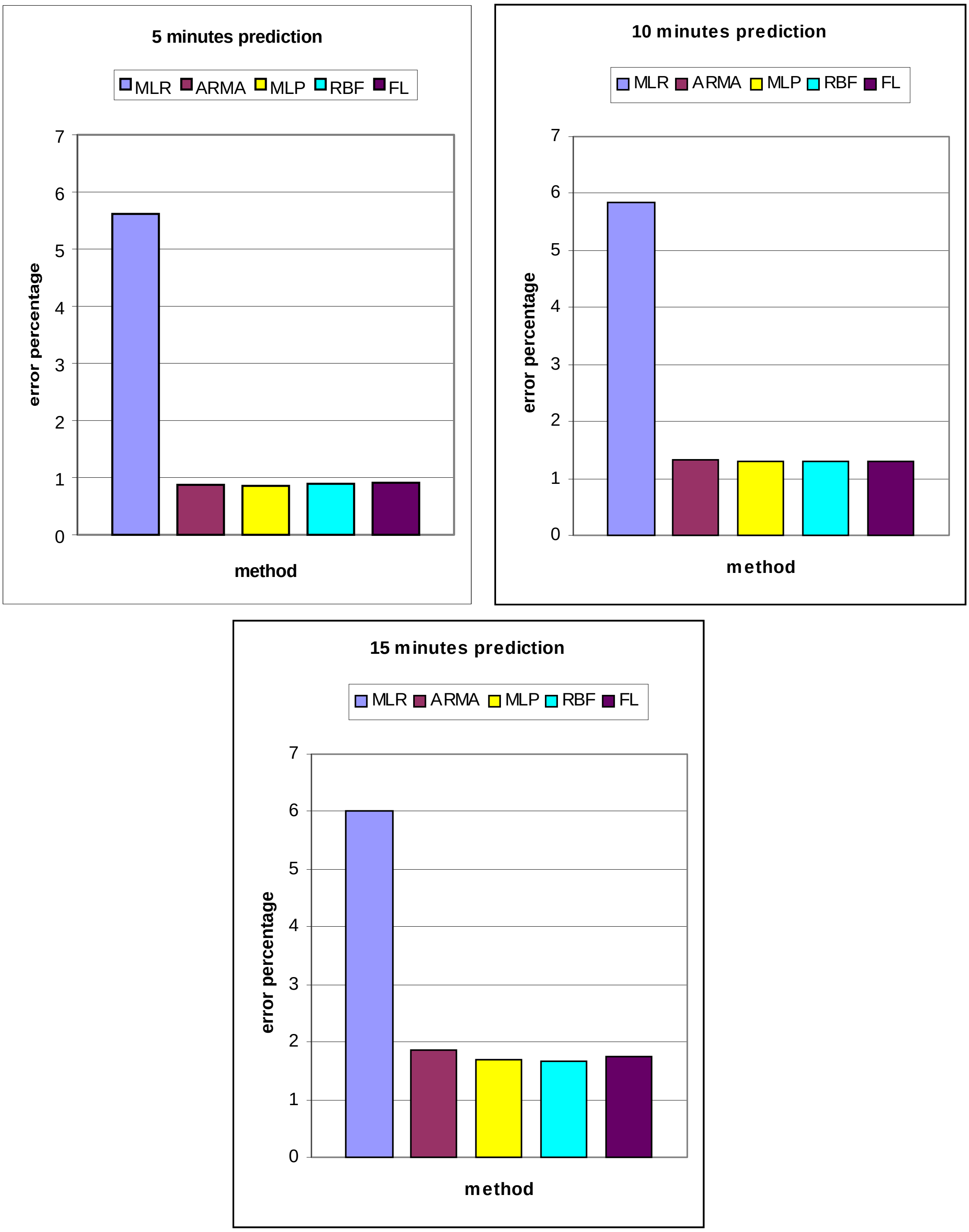

Figure 3a-c: Performance of the methods for the prediction horizon of a) 5 minutes, b) 10 minutes, and c) 15 minutes