neural networks; transportation; prediction of road traffic parameters

Teresa PAMUŁA
Silesian University of Technology, Faculty of Transport
Krasiński 8, 40-019 Katowice, Poland
Corresponding author. E-mail: teresa.pamula@polsl.pl

NEURAL NETWORKS IN TRANSPORTATION RESEARCH – RECENT APPLICATIONS

Summary. Neural networks’ (NNs) capability of mapping the nonlinear functions of variables describing the behaviour of objects and the simplicity of designing their configuration favours their applications in transport. This paper presents representative examples in the scope of prediction of road traffic parameters, road traffic control, measurement of road traffic parameters, driver behaviour and autonomous vehicles, and transport policy and economics. The features of the solutions are examined. The review shows that feedforward multilayer neural networks are the most often utilised configurations in transportation research. No systematic approach is reported on the optimisation of the NN configurations to achieve a set level of performance in solving modelling tasks.

ZASTOSOWANIE SIECI NEURONOWYCH W BADANIACH W TRANSPORCIE – OSTATNIE OPRACOWANIA


1. INTRODUCTION

Models of transportation problems often contain complex and nonlinear relations between variables describing their mostly highly dynamic behaviour. The application of artificial intelligence methods, which combine elements of self-learning, adaptation and self-organisation, enables effective elaboration of such models. Artificial neural networks implemented in software or using hardware are the dominant representatives of processing algorithms related to artificial intelligence.

Neural networks were proposed in the 1960s of the XX century and introduced into transportation research in the 1990s. The significance of these is reported in the review by Dougherty done in 1995 [1]. The author distinguishes nine areas of NN applications - evaluation of driver behaviour,
determination of road traffic parameters and O-D matrices, evaluation of road surface, detection and classification of vehicles, traffic pattern analysis and incident detection, freight operations, traffic forecasting, road traffic control, and transport policy and economics.

Since the second half of the 1990s, about fifty new publications in journals and a similar number in conference proceedings annually report advances in the application of NNs in transportation (WoS 1049/738, Scopus 904/1103 publication records). M.G. Karlaftis and E.I. Vlahogianni in 2011 [2] compared the scope of using statistical and NN methods in solutions to transport problems. They propose to divide the areas of application into six groups: traffic operations, infrastructure management, maintenance and rehabilitation, planning, environment and transportation, and safety and human behaviour. This division is coincident with the presented earlier. It emphasises the bonds between problems, e.g. the determination of traffic parameters and the detection of vehicles and incidents. The introduction of integrated systems of managing and controlling traffic - ITS (Intelligent Transport Systems) forces the merging of research problems. The year 2008 brings about new research topics, that is, the problems of autonomous vehicle control and their participation in road traffic. This is highlighted by a significant increase in the number of WoS and Scopus publication records.

A slightly different classification of topics related to NNs in transport is proposed in this paper:
- traffic forecasting
- traffic control,
- evaluation of traffic parameters,
- maintenance of transport infrastructure,
- transport policy and economics,
- driver behaviour and autonomous vehicles.

Representative reports of sample applications for modelling and computation published in journals in the last five years will be examined in this paper.

2. PROPERTIES OF NEURAL NETWORKS

A neural network (artificial neural network) implemented as a software programme or in the form of a hardware device, processes data “computes” using a set of simple elements modelling the functions of neurons ordered into layers. The most outstanding property of NN is its ability to map nonlinear relations between variables describing the model’s behaviour. This mapping is obtained in the course of training of the network without the need for a thorough analysis of the properties of the variables. In the case of mapping dynamic behaviour, there is no need to assume the stationarity of behaviour. Stationarity is usually necessary when using other modelling methods, especially statistics. NNs are capable of modelling systems with very complex dynamics behaviour.

One can list the important properties of neural networks:
- classification and recognition—during the training the network learns the characteristics of the sequence of templates and elaborates the classification formula,
- approximation—the network computes the value of a multivariable nonlinear function based on learned properties of the function,
- association—the network memorises the set of templates and picks the most similar when presented with new data,
- data clustering—the network determines common characteristics of processed data,
- prediction—following or future values or statistic parameters are elaborated,
- low sensitivity to data “noise” or errors,
- ability to work when partly damaged [3].

The drawback of computations performed with NNs is the dependence of the results on the method of training of the network. Such phenomena as overfitting—the network remembers the training sequence; dependence on the choice of initial parameters—the order of the training samples changes the NNs outcome; necessity of inputs standardisation—heterogeneous features require a diverse scaling scheme—highly influence the training process. The proper selection of the training sequence and the appropriate choice of neuron parameters determine the success of modelling or computing.
There are no rules for constructing a network to achieve set-forth processing properties. In literature, networks are developed in the course of experiments on the basis of general premises derived from earlier tested configurations. The vast majority of applications utilise multilayer feedforward networks. The networks, besides the input and output layer, are built of one or several hidden layers. Most of the neural networks employ some form of gradient descent training algorithms, using backpropagation to compute the weights of neuron inputs. Rare solutions are based on recurrent networks or Kohonen nets (SOM self-organising maps).

3. APPLICATIONS OF NNs IN TRANSPORT

Neural networks are used in transport research to solve such problems or computational tasks as:
- classification and clustering,
- function approximation,
- time-series analysis and forecasting.

The proposed classification of research topics shows that each group will have its specific methodology of study – Table 1.

<table>
<thead>
<tr>
<th>Group of applications of NN – research topics in transport</th>
<th>Classification and clustering</th>
<th>Function approximation</th>
<th>Time-series analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>traffic forecasting</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>traffic control</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>evaluation of traffic parameters</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>maintenance of transport infrastructure</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>transport policy and economics</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>driver behaviour and autonomous vehicles</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

This classification emphasises topics related to the field of ITS, which currently receives the most attention from researchers working in transportation. The paper’s author selects and reviews applications from journal papers published in the years 2010-2015 (WoS 509, Scopus 473 publication records). Features of the representative NN applications are presented and discussed. The timeline is chosen to complement the previous review of M.G. Karlaftis and E.I. Vlahogianni done in 2011 [2].

3.1. Traffic forecasting

The tasks of forecasting traffic parameters are performed for local traffic control systems, for networks as well as for managing traffic. These are defined by the time horizon of predictions and the expected accuracy. NN-based predictions compete with methods founded on statistics.

NN-based forecasting methods differ in:
- the way the forecasting task is identified,
- the generated output of the network.

Identification covers the choice and resolution of parameters that are passed to the inputs of the network. Parameters such as length of trip, speed of travel and traffic flow are processed. The network, after training, generates the next values of the input variables or indicates the direction of change of these.

In the work [4] it is proposed to use speed profiles to predict the mean travel times in the road network of a town. The speed profiles are determined by clustering real traffic data. This operation changes the original feature space, lowering significantly its dimensionality and thus enabling the
reduction of the number of network inputs. Prediction of mean travel times is obtained in a ½-hour horizon.

Mean travel time is sensitive to traffic incidents, which are of random nature. Van Hinsbergen et al. [5] propose the modelling of this randomness with neural network committees. Bayesian techniques are used for training, which results in the capability of predicting “low-frequency trend”, that is, the values of the travel times. Feedforward single layer networks are used with an undetermined number of hidden nodes. The problem of optimisation of the network parameters is not discussed.

Forecasting is done usually with the aid of historical data acquired at regular intervals of time from stationary traffic measurement devices. Currently, Intelligent Transport Systems can utilise traffic data from other sources, for instance, those derived from GSM (Global System for Mobile Communications) data. Such sources provide data with irregular time stamps and with widely scattered interval values. In the work [6] authors use such data as inputs of an NN for predicting the travel speed on traffic routes. The neural network is an arbitrarily chosen feedforward back-propagation network with four inputs, a hidden layer of 12 neurons, and a single output neuron. Preliminary, forecasts did not provide sufficient accuracy and the data set was supplemented with acceleration values and neighbouring traffic routes flow values. This expansion gave expected results, which proves that measurements done at irregular time intervals are suitable for prediction but require auxiliary information to enhance their accuracy.

Traffic flow values are frequently predicted. For instance, K. Kumars et al. [7] predict flow values, using NNs, on the basis of a sequence of historic flow values, traffic speeds and the weekday on which the prediction is done. In order to achieve fewer errors, the traffic stream is treated as a set of vehicles of different categories. Each category is represented by a separate input variable. Good prediction results are reported for time horizons of 5-15 minutes. The trial and error approach is used to obtain the best performing network architecture. Ten multilayer perceptron (MLP) models with different numbers of hidden neurons were constructed and trained using the same set of training data.

A similar approach is presented in [8]. Time series of traffic flow values in mornings attributed to workdays and holidays are used for prediction of traffic flows in the afternoon hours, respectively, on workdays and holidays.

Prediction errors are frequently compared with errors of prediction based on statistical analyses. The application of NNs is not explicitly superior, so some authors develop hybrid solutions. K. Yan et al. [9] use exponential smoothing to preprocess traffic flow data before inputting the data to the NN. A three-layer MLP is used with nine hidden nodes. The number of nodes is related to the size of the training sets. The network is trained using the Levenberg-Marquardt algorithm, which improves the generalisation capabilities of the network. Results confirm the better prediction ability of the hybrid solution.

Hybrid solutions gain on popularity and their development becomes an important research topic.

3.2. Traffic control

Attempts are made to utilise computation methods based on NNs in the designs of road traffic controllers, systems of traffic network control, and traffic management systems. NN enables the merging of historical data with current identification parameters of the road situation for the elaboration of control decisions.

The authors of [10] overview the problems of computational intelligence methods in urban traffic signal control and stress the importance of NN in this field. A number of examples of system-wide traffic-adaptive controllers and local real-time signal controllers based on NNs are given. Solutions using fuzzy logic are of special interest. In this case the NN generates a fuzzy value, which maps the control decision. The developed algorithms improve the efficiency of traffic control by over ten percent in comparison with fixed time signalling. Network input variables represent the presence of vehicles on approaches to junctions and historical traffic flow values on these approaches. No discussion is provided on the design of the NNs.

The authors of [11] make explicit reference to the idea of neurons. The neural network used to model the traffic control system is trained by a live expert—a traffic engineer. The state-space of the
Neural networks in transportation research – recent applications

junction is identified and modified on the basis of the expert’s experience and knowledge. Simulations and comparison with the MOVA strategy prove the advantage of this solution. A three-layer MLP is used with a variable number of nodes in the hidden layer. The number of hidden nodes from the range (5-11) is selected based on the performance of the NN during simulations.

3.3. Evaluation of traffic parameters

The image of the traffic situation in the case of ITS is mapped with good accuracy by O-D matrices. Many works report advances in determining O-D matrices using incomplete or “noisy” traffic data from different measuring devices. An important problem for ITS functioning is the detection of traffic incidents that may destroy the equilibrium of the managed traffic system.

Lorentzo et al. [12] developed an NN-based solution for determining the O-D matrix using traffic flow values registered between junctions. Traffic data dimensionality is reduced with PCA and the results are placed on the inputs of a multilayered feedforward neural network. It is trained using as inputs the eigenvalues of the flow dataset. The number of hidden nodes is 50 and the number of neurons in the output layer is equal to the number of cells of the processed OD matrix. The NN topology is cross-validated in the course of several iterations. The trained NN is capable of generating values of O-D matrices for almost real-time traffic management. In [13] a solution, which uses congestion detector values, vehicle positions and time of travel for short-term prediction of travel speed, is proposed. The travel speed on a given route is determined using multiple NNs working with traffic sensor data. Additionally, the networks utilise historical data in the course of training. Three-layer MLPs are used with the number of hidden nodes determined in a similar way as in the previous solution. The range of hidden nodes considered is between 5 and 200.

The authors of the solution presented in [14] use the NN for merging data from fixed road sensors and mobile measurement stations in cars. Feed forward networks were tested with one (50 neurons) and two (30-30) hidden layers. NN inputs correspond to the number of sensors. The trained network correctly maps the road traffic parameters in the area covered by the sensors and moving vehicles. Good approximation properties of the network ensure that a comparatively small number of measuring cars adequately supplement the determination of traffic flows in a road network. The two hidden layer topology provided better accuracy of flow prediction.

Works [15, 16] deal with the problem of accident detection. Solutions, which evaluate the level of accident hazard in the monitored traffic area, are proposed. Temperature, humidity, weather conditions, and time of day favour hazards. These, together with historical data on the occurrence of accidents, are inputs of the NN. The capability of clustering data is successfully utilised for computing the accident hazard map. In the first work a simple perceptron is used for mapping accidents on the road network, while the second solution uses a three-layer MLP with 10 hidden nodes to predict accidents on a highway.

3.4. Maintenance of transport infrastructure

The subject of highest concern in the maintenance of transport infrastructure is pavement upkeeping. NNs are applied for prediction of pavement—condition and performance, management and maintenance strategies, and distress forecasting. H. Ceylan et al. [17] present an exhaustive survey of works in this field. Most of the applications use MLP for modelling the problems. The topology of used NNs is rarely discussed.

3.5. Transport policy and economics

In the group of transport policy and economics, topics such as assessment of the consequences of transport infrastructure expansions and evaluation of the effects of changing the scope of transport services dominate.

Authors in [18] present the solution to planning the expansion of container terminals. The configuration of an NN is proposed for the optimisation of the order of carrying out the expansion
objectives considering the investment resources. The input variables are—equipment to be installed, generated traffic, and changes to port transport infrastructure. Additionally, characteristics of previous expansion plans of similar investments are utilised. Closely related problem, but in the field of highway construction, is elaborated by M. Li and W. Chen [19]. The developed NN generates the assessment of the planned investment on the basis of construction costs, environmental impact, and applied management systems and policy. The obtained model is used for simulating different expansion strategies. In both cases MLP networks are used with arbitrarily chosen topology.

Effective provision of transport services requires the knowledge of the demand and proper management of the available supply. Representative examples of using NN for assessing demand are discussed in the works of Yingjun [20] and Gonzales [21]. Yingjun forecasts the number of taxicabs using a wavelet neural network. Network input variables are chosen to be the number of taxicabs and influence factors—service comfort, cost and speed. Gonzales undertakes the task of evaluating the transport preferences of residents on the basis of survey results and registered trip routes (using mobile phone applications). This vast amount of data is processed by NN to extract synthetic data describing the preferences. These, in turn, are used for transport policy planning.

Collected traffic data may be used to streamline the management of the available supply of transport services or resources. In [22] a reservation system based on an MLP with one hidden layer, which optimises its throughput and costs, is proposed. The NN works out entry decisions to the managed transport area. Input variables used are the size of vehicle and number of travellers, time of entry, and trip length.

3.6. Driver behaviour and autonomous vehicles

Perception of the road situation and judgement and decision-making by drivers are governed by many factors, and conventional modelling methods are hard to apply. The researches in this group of topics are concentrated on the problems of ensuring driver safety and developing vehicle control systems for driverless movement. The problem of ensuring safety is investigated from the position of the driver, that is, his capability to drive, as well as from the position of surmounting the traffic conditions, that is, hazards encountered while driving.

Research topics related to controlling autonomous vehicles cover the problems of vehicle movement identification and the chosen rules of driving, first of all: tracking the preceding vehicle and keeping the movement inside the road lane.

In the vast majority of works, the capability of driving is assessed on the basis of biological signals registered by sensors placed on the driver's body. In [23], as representative, to indicate the driver's level of stress, the time series of physiological signals—galvanic skin response and photoplethysmogram—are regarded. Authors carry out a comprehensive performance analysis of feedforward and recurrent NN for stress level detection. The statistical, structural and time-frequency changes observed in the recorded biosignals constitute the NN inputs. The level of stress is successfully estimated and can be used for providing safety alerts to drivers in real time. Recurrent NNs prove to give the best results.

M. Patel et al. [24] determine the level of fatigue on the basis of driver’s heart rate variability. The spectral image of power spectral density of the time series of electrocardiogram data is the input of a feedforward NN with two output neurons and no hidden layer. Different neuron activation functions are tested to obtain the desired error of fatigue classification. A similar value of error was attained as in the case of assessing the stress level. In [25] fatigue-based indicators derived from EEG time series, the head nodding angles are inputs of a three-layer neural network. An arbitrarily chosen 9-neuron hidden layer is used. Fatigue classification results are coincidental with [24]. Derivative to fatigue, the level of driver’s alertness is evaluated in much the same way by authors of the work [26]. Again, the EEG signal is used and additionally a five min forecast is done to ascertain whether it is safe to continue driving.

Without sensors on the driver’s body, the distraction level is determined in [27]. Head tracking data and long-term temporal context of driving are variables processed by a recurrent NN. This recurrent NN is further modified to include a long short-term memory, which is used to store and access
information over long periods to overcome the vanishing gradient problem. The topology of the solution is not discussed.

Chong et al. [28] develop a fuzzy neural network that generates driver’s behaviour features based on manoeuvre characteristics and lane keeping precision. In [29] the way the driver adapts the vehicle’s speed to traffic conditions is assumed as adequate to determine the driver’s behaviour. Gas and brake pedal actuation values constitute input variables of a Partly-Connected MLP. This modified MLP uses seven hidden nodes active at any time in the hidden layer. In both works the aim is to detect dangerous driver behaviour.

The difficult problem of identification of vehicle position relative to other road users or to infrastructure is solved by combining data from different sources. Authors in [30] propose to use a combination of video and GPS data to control vehicle movement. The video system provides synthetic information describing the content of the field of view; inaccessible parts are masked while the GPS receiver tracks the trip. The system is composed of a committee classifier consisting of six NNs for processing video data followed by an MLP for deriving the next manoeuvre action (the steering angle and velocity). A three-layer feedforward NN with a changing number of hidden nodes 3-15 is evaluated. A different approach is presented by Borenovic’s group [31]. Instead of GPS data, RSS (received signal strength) values obtained from GSM base stations are used for positioning. This eliminates the problems with reception of GPS signals in densely built-up areas. A multilayer feedforward network is used. A heuristic rule is proposed for determining the number of hidden nodes—the first hidden layer contains roughly two times the number of inputs or outputs, whichever is greater, and the following layers contain a decreasing number of nodes toward the number of outputs.

In [32] the relative positions of vehicles are determined using a fuzzy NN. The network receives processed data from a camera and from a scanner. The camera provides the description of the preceding vehicle’s shadow while the scanner determines the centre of the road lane. The NN generates alerts when the vehicle comes too close to the preceding object or when it leaves its lane. In a similar way the speed of a vehicle is controlled to prevent it from running into another road user [33]. The number of hidden layers of the NN coincides with the number of fuzzy rules; the neuron weights are determined by the fuzzy output linguistic variables.

The problem of supporting driving manoeuvres – navigating is examined in papers [34, 35]. The ability to generalise relations between data values is used to reduce the size of the solution space in the effort to find the shortest path in a road network. A minimal resource neural network framework is proposed for solving single-source shortest path problems of various graph types, which represent the road networks.

4. CONCLUSIONS

The features and computing capabilities of neural networks favour their wide application for solving transport problems. The presented examples illustrate the effectiveness of these solutions. Notable are applications in the field of prediction and evaluation of traffic parameters. Prediction tasks are influenced by a large number of variables, with interrelationships that are hard to model. Data clustering and association capabilities of the NNs have proved to be effective in elaborating these tasks. NNs enable the fusion of diverse kinds of data, which is especially useful in systems combining data from different sources, as is in the case of ITS. The examples of evaluation of traffic parameters utilise not only data from road infrastructure-based devices but also GPS and floating car data.

Noticeably lacking are works reporting research on the development and tuning of network configurations in order to achieve a set-forth level of performance. Mostly multilayer feedforward networks are applied with configurations based on some heuristics or literature records. In a number of examples, the topology of the hidden layer of the MLP is varied to improve its characteristic. The approach, which most effectively enhances the NN performance is to lower the dimensionality of the processed variables. Notable examples have been reported in the area of controlling autonomous vehicles. For instance, video data, used for determining the vehicle position, is reduced to a sequence of patches representing the observed road.
Research work is needed to develop effective ways of systematic design of NN configurations to resolve computation problems within defined accuracy bounds. The determination of the topology of hidden layers of MLP in relation to the complexity of input variables requires examination. Effects and results of using NNs prove that these tools will play an active role in future ITS development. Synergies of urban traffic control systems and advanced vehicle control systems are topics that will receive attention.

References


Received 12.02.2015; accepted in revised form 08.05.2016