ARTIFICIAL NEURAL NETWORKS

CHAPTER - 3

ARTIFICIAL NEURAL NETWORKS

3.0 INTRODUCTION

A simulation model consists of a number of models that describe the various component of the simulated traffic system. The logic of manoeuvring individual vehicles through the simulated system forms the core of the traffic flow simulation models. The models used to represent the vehicle movement in the present day simulation models are mostly rule based and are derived by empirical methods from pre-processed data. As the driver behaviour is stochastic, the rule based models fail to represent the real life situation. The techniques, which have a performance close to human behaviour, are likely to represent the process better than rule-based models. For many reasons, neural computing offer a more realistic tool for modelling the highly probabilistic behaviour of drivers. An overview of artificial neural networks and their applications in the field of traffic engineering is the subject matter of this chapter.
3.1 APPLICATIONS OF NEURAL NETWORKS FOR TRAFFIC ENGINEERING STUDIES

Many researchers have reported that the neural networks offer a potential alternative tool for traffic engineering studies. Dougherty et al. (1993) used neural network to predict and classify traffic congestion. They made use of the data collected via an on line computer link to the SCOOT traffic control system and concluded that neural computing is well suited for analysis of large volumes of data. Hsio et al. (1994) successfully applied fuzzy logic and neural networks to predict the traffic congestion due to roadway incidents. Ledoux (1997) proposed a cooperation based neural network traffic flow model that is being integrated into a real time adaptive urban traffic control system. The model is expected to simulate the flow at an isolated intersection surrounded by neighbouring intersections. Fair performance of neural networks to predict the queue lengths for the next 60 sec was reported. Dharia and Adeli (2003) presented a neural network model for forecasting the freeway link travel time using the counter propagation neural network. The performance of the model was compared with that of a model using the back propagation algorithm. The model based on the counter propagation algorithm was reported to be faster than the back propagation network. The proposed freeway link travel-forecasting model finds use for real-time advanced travel information and management systems.

Pant and Balakrishnan (1994) studied the gap acceptance behaviour of drivers at stop-controlled intersections using neural network models. They compared the results of neural network model with that of the Binary-Logit model and concluded that the neural networks predicted the accepted or rejected gaps better than the Binary-Logit model.

Mussone et al. (1999) described a method based on the use of artificial neural networks (ANN) in order to work out a model that relates to the analysis of vehicular accidents in Milan. They used feed-forward neural networks with a back-propagation learning paradigm. The degree of danger of urban intersections under different scenarios was quantified by using ANN model. The main difficulty of including into a regression model the class variables such as day/night, type of intersection, class of roadbed, weather and, above all, type of accident can be easily overcome by using neural network models. Srinivasan et al (2004) evaluated the performance of a neural network model, originally developed for a freeway site in Singapore, for detection of incidents on a freeway site in California and observed that the neural network models are portable.

Awad (2004) developed models for estimating capacity of weaving segments of freeways using regression and neural networks. The multi-layer feed-forward neural network model was trained using an improved back propagation algorithm with momentum and adaptive learning rate. Although, linear regression technique showed satisfactory results, neural network technique outscores linear regression in the prediction performance, and generalization ability. The trained neural network architecture represented by weight and bias values for each layer can be simply used to predict capacity for weaving segments under new conditions.

Jiang and Zhang (2001) applied ANN technology to map the relationship of the traffic flow and average-space speed using the data from an arterial road in
Changchun, China. They concluded that this method makes the predictions more reliable, more cost-efficient and easier.

Artificial neural networks (ANNs) have been extensively studied and used in time series forecasting. Yasdi (1999) designed and trained a neural network based on recurrent Jordan architecture popular in the modelling of time series to predict the future values of the traffic time series using its past values. He considered three types of forecasting: weekly, daily and hourly predictions as long-term, mid-term and short-term predictions respectively. The weekly prediction relies on the historic pattern stored in the knowledge base, while for daily prediction, for each day of the week and for each predicted event, a reference pattern is stored which can be used for corresponding predictions. The hourly prediction, however, is a short time prediction, which gives a view of the current traffic situation. The trained values were compared with the corresponding actual values, and they were found to be in close agreement. These results were reported to be better than that of the compared methods. The model improved the forecasting by about 20% for the road traffic flow. The neural network method requires a database to perform the learning task successfully. Therefore, the database has to be updated frequently.

Messai et al (2002) proposed a new short term traffic flow prediction model based on a feed forward neural network and observed good performance of the model by comparing with simulated and real life data. Xiao and Wang (2003) and Xiao (2004) applied radial basis function neural network with feedback for traffic flow modelling and concluded that ANN has the excellent capability of describing the highly non-linear dynamic characteristics of the traffic flow.

Zhang (2003) proposed a hybrid methodology that combines both ARIMA and
ANN models to take advantage of the unique strength of ARIMA and ANN models in linear and non-linear modelling. The major advantage of neural networks is their flexible non-linear modelling capability. With ANNs, there is no need to specify a particular model form. Rather, the model is adaptively formed based on the features presented from the data. This data-driven approach is suitable for many empirical data sets where no theoretical guidance is available to suggest an appropriate data generating process. Chen et al (2003) proposed Grey Neural Network, a novel method that combines grey system theory with neural network, to forecast highway traffic flow and concluded that the GNN model outperformed the other methods of traffic flow forecasting. Hence, neural networks have great potential for application in traffic engineering.

3.2 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks consist of large number of simple, inter-connected processing units or nodes called neurons. The neuron consists of three basic elements as described below:

i. A set of synapses or connecting links, each of which is characterised by a weight or strength of its own. A signal at the input of the synapse connected to the neuron is multiplied by its synaptic weight. The weight is positive if the associated synapse is excitatory; it is negative if the synapse is inhibitory.

ii. An adder for summing the input signal, weighted by the respective synapses of the neuron.

iii. An activation function for limiting the amplitude of the output of neuron
The model of a neuron is presented in Fig. 3.1. A neuron takes in a set of inputs and computes an output according to a transfer function. The inter neuron connections will be of varying strength; each connection having a weight associated with it. The output state of a neuron will be either 'off' or 'on'. A change from one state to the other is triggered when the sum of the inputs (weighted by the strength of their respective connections) crosses some threshold. This threshold is usually represented by a transfer function. The neurons will be generally arranged in layers, with nodes in adjacent layers connected to each other, either partially or fully. A neural network thus has an input layer, an output layer and possibly one or more hidden layers. Thus, a neural network is a massive parallel-distributed processor that has a natural propensity for storing experimental knowledge and making it available for use.

The properties of Neural Networks viz., the non-linearity, the input-output mapping, the adaptivity, the fault tolerance, the evidential response, the
neurobiological analogy and others make it possible to solve complex problems using this approach.

3.3 NETWORK ARCHITECTURE

The manner in which the neurons of a neural network are structured is known as the network architecture. The network architecture can be broadly classified as: single layer feed forward networks, multi layer feed forward networks, recurrent networks and lattice structures. Multi-layer feed-forward networks are the most widely used.

3.3.1 Multi-layer Feed Forward Networks (MFNN)

This class of a feed forward network distinguishes itself by the presence of one or more intervening layers between the input and output layers called hidden layers. By adding one or more hidden layers the network acquires a global perceptive despite its local connectivity by virtue of the extra set of synaptic connections and the extra dimensions of neural interactions. The MFNN is said to be fully connected if each node in each layer of the MFNN is connected to every other node of adjacent forward layer. If some of the synaptic connections are missing the network is said to be partially connected.

3.3.2 Recurrent Neural Networks (RNN)

A RNN distinguishes itself from a feed forward neural network, in that it has at least one feedback loop. The feedback connections can originate from hidden layers as well as outer layers or from either of these layers. The presence of feedback loops has a profound impact on the learning capability of the network and on its performance. Moreover, the feedback loops involve the use of particular branches composed of unit delay elements, which result in a non-linear
dynamic behaviour by virtue of the non-linear nature of neurons. Partial recurrent neural networks are those which have feedback loops from only one of the layers. Of the common partial recurrent neural networks used, Elman network is popular. Elman network is effective in representing both temporal and spatial patterns.

3.4 LEARNING PROCESS

Learning is a process by which the free parameters of neural network are adapted through a continuing process of stimulation by the environment in which the network is embedded. The learning process involves the following steps:

   i. The neural network is stimulated by an environment.
   ii. The neural network undergoes changes as a result of this stimulation.
   iii. The neural network responds in a new way to the environment, because of the changes that have occurred in its internal structure.

The learning paradigms are basically classified as supervised learning, reinforcement learning and self organised learning. The supervised learning is performed under the supervision of an external teacher. Reinforcement learning involves the use of a critic that evolves through a trial and error process. Unsupervised learning is performed in a self-organised manner in that no external teacher or critic is required to instruct synaptic changes in the network. The back-propagation algorithm, which is a supervised learning algorithm, is the most successful and widely used algorithm for the design of multi-layer feed-forward networks used to represent large-scale systems. Fig. 3.2 shows the architecture of a feed-forward network with two hidden layers and one output layer.
3.4.1 Concept of Back-Propagation Algorithm

In multi layer feed forward networks, the input signal propagates through the network in a forward direction, on a layer by layer basis. The error back-propagation process consists of two passes through different layers of network: a forward pass and a backward pass. In the forward pass, an activity pattern is applied to the sensory nodes of the network, and its effect propagates through the network layer by layer. Finally, a set of outputs is produced as the actual response of the network. During the forward pass the synaptic weights of the network are all fixed. The actual response of the network is compared with the desired response and error signal is calculated. The error signal is then propagated backward through the network. The synaptic weights are adjusted so as to make the actual response of the network being closer to the desired response. During the backward process, the synaptic weights are all adjusted in
accordance with the error-correction rule. Detailed description of back-
propagation algorithm is available in many text books (Wasserman (1989),
Haykin (1994)).

The correction $\Delta w_{ij}(n)$ to be applied to the weight $W_{ij}$ is defined by the delta rule as

$$\Delta w_{ij}(n) = -\eta \frac{\partial \xi(n)}{\partial w_{ij}(n)}$$

Eqn. 3.1

where: $\eta$: rate of learning.

$\xi(n)$: instantaneous sum of error squares.

$W_{ij}$: weight of the synapse connecting the neuron 'i' and neuron 'j' at
iteration 'n'.

3.4.2 Rate of Learning

The training algorithm provides an approximation to the trajectory in weight space
computed by the method of steepest descent. The smaller the learning rate
parameter '$\eta$', the smaller will be changes to the synaptic weights in the network
from one iteration to the next and the smoother will be the trajectory in weight
space. However, this is at the cost of slower rate of learning. If, on the other
hand, the learning rate parameter '$\eta$' is too large so as to speed up the learning,
the resulting changes in the synaptic weights assume such a form that the
network may become unstable. A simple method of increasing the rate of
learning and yet avoiding the danger of instability is to modify the delta rule by
including a momentum term.

Considering its frequent use and record of success, a multi layer feed-forward
neural network, trained using back propagation algorithm has been selected in
the present study.
3.5 CONCLUSIONS

A review of the applications of neural networks has revealed that neural networks offer a potential alternate method of modelling traffic flow in mixed mode environment. An overview of the neural networks was presented in this chapter. It is suspected that neural networks might not be able to provide total solutions to complex systems, such as mixed traffic flow, as they need prior exposure in the form of training. Rather, they appear to be more useful for representing some of the sub-systems in the simulation model. Hence, it is proposed to use neural networks for modelling some of the component blocks of the traffic flow simulation model.