Proceedings of the Eastern Asia Society for Transportation Studies, Vol.10, 2015

Analysing the Behaviour and Performance of Neural Network Trip Distribution Models toward Different Hidden Layer and Node Numbers

Gusri YALDI

Civil Engineering Department, Padang State Polytechnic, Padang, Indonesia 25163
E:mail: gusri.yaldi@yahoo.com

Abstract: Neural Network (NN) approach is widely known as an intelligent computer system working based on human brain working system where important properties characterizing the NN performance are learning algorithm, activation function, number of layers, number of nodes inside each layer, and learning rate. It is difficult to find articles scrutinizing and reporting the behaviour of a NN model trained with different properties. However, former studies were found using different NN properties for different studies and suggesting different performances, including the hidden layer (HL) and node numbers. This paper is intended to analyse the behaviour and performance of a NN model trained with different properties for working trip distribution estimation, in this case the hidden layer and its node numbers. Therefore, this paper becomes essential and the findings are expected can be useful for those interested in using and developing NN as a modelling tool, especially for trip distribution modelling.

Keywords: Neural Network Model, Trip Distribution, Hidden Layer, Hidden Layer Node

1. INTRODUCTION

Neural Network (NN) approach is widely known as an intelligent computer system working based on the human brain working system. It is a forecasting method that specifies output by minimizing an error term indicated by the deviation between input and output through the use of a specific training algorithm and random learning rate (Black, 1995; Zhang et al, 1998).

Its applications for prediction and estimation in various disciplines keep growing. It is not only because its simple characteristics, but also its considerable ability in predicting and estimating pattern. A neural model is able to capture the relationship between independent and dependent data without knowing the distributional attributes of the variable used. Another big advantage of a neural model is that it has better capability in estimating trip interchange numbers without knowing the distance decay or impedance function, as found by the empirical results reported by Tillema et al. (2006). However, working with this intelligent approach is also like playing a puzzle game. The puzzle can be defined as the approach itself, while the component of the puzzle represents its properties. The NN model will perform satisfactorily when the properties are properly selected, or it will go to negative direction when wrong properties/parts of the puzzle are used in the training. Then, some modellers suggested that this approach is good for forecasting and some suggested it has poor generalization ability. These contradictive situations are reflected in formers studies, especially in the spatial movement modelling, including passenger, commodity and migration flow (Yaldi et al., 2009a, Murat Celik, 2004, Mozolin et al., 2000, Black, 1995). Therefore, the users of NN must first understand the nature of the problem, then select main properties of the network, before finally train and use the results.

Literature review suggests that the adoption of NN approach for fully constraint spatial movement model was initialized by Black (1995). NN approach was used to model the commodity and migration flow in USA. The traditional gravity model was used to construct the NN model structure and proposed as an alternative to the well-known doubly-constrained gravity model. The assessment of the NN model performance was firstly regarding its ability to satisfy the Trip production (P) and Trip Attraction (A) constraints, and then followed by the estimated flows. The same evaluation step was also applied by Mozolin et al. (2000) and Yaldi et al. (2010).

Important properties that characterize the NN performance are learning algorithm, activation function, number of layers, number of nodes inside each layer, and learning rate (Teodorovic and Vukadinovic, 1998, Dougherty, 1995). The amount of dataset and the ratio for training, validating and testing is also important for the NN fitting performance (Carvalho et al., 1998). It is difficult to find articles scrutinizing and reporting the behaviour of a NN model trained with different properties. However, former studies were found using different NN properties for different studies and suggesting different performances as reported in Table 1, including the hidden layer (HL) and node numbers. This paper is intended to analyse the behaviour of a NN model trained with different properties, in this case the hidden layer and its node numbers. Therefore, this paper becomes essential and the findings are expected can be useful for those interested in using and developing NN as a modelling tool.

Multilayer Feed forward Neural Network is used here in order to investigate the behaviour of neural models toward different hidden layer and node numbers. Other sub components such as training algorithm and activation function are discussed, but not thoroughly. Meanwhile, the number of nodes in both input and output layer can be seen in Figure 1. Three nodes will be used in the input layer, while output layer has only one node. Levenberg-Marquardt training algorithm is used to train the neural models. All of the layers, except the input layer, have the logsig as the activation function for all of nodes. The uniform activation function is used since the aim of these experiments is to look at the behaviour of neural models towards the hidden layer and node numbers. It would be more complicated when more variables are added in these experiments. Consequently, important behaviours and conclusions would be difficult to draw. This is the information for the model specification.

For the model training part, three datasets are prepared. The data split method is unique compared to the methods used by Mozolin et al. (2000), Yaldi et al. (2009b), and Shir-mohammadli et al. (2010). The mix linear data normalization method is used, the same as that used by Black (1995) and Yaldi et al. (2011). The batch training mode is used to train the model, while over training is avoided by using a validation dataset. Each initial weight configuration is randomly defined by the modelling tool, in this case the Nguyen-Widrow technique. Consequently, the same model will require multiple training or trials so that it is not sensitive and dependent upon a specific value of weight. For the model testing and performance sections, the relevant sub sections discussed in this paper are root mean square error and coefficient of determination.

2. MODEL SCNARIOS

The scenarios used in these experiments are reported in Table 2. As a black (grey) box approach, the application of NN in trip distribution modelling requires a number of experiments/trials. Empirical findings can help to improve the neural model performance and reduce criticism.

Table 1. Summary of former NN models in transport modelling

Author(s)	Subject	ZN	TA	HL#	[2]HL node#		AF		DS	DN	[3]Exp#	Data share (%)			Epoch #	PM	Comparison
					1st HL	2nd HL	HL	OL				Train	Val	Test			
Black (1995)	1 & 2	3, 9, & 9	BP	1			1	1	-	SN	-	100	-	-	Up to 150000	RMSE	GM
Dantas et al. (2000)	1	17	BP	2	15	7	2	3	1	LT	-	75		25	-	MSE	-
Mozolin et al. (2000)	1	15, 20, 345, & 507	Quickprop	1	vary	-	2	2	1	SN & LT	5	vary			Up to 100000	SRMSE	GM
Celik (Celik 2004b)	2	48	LM	1	[4]cond	-	1	1	-	-	-	100	-	-	-	ARV & SRMSE	Box Cox
Celik (2004a)	2	48	BP ("LM")	1	16	-	1	1	-	-	-	100	-	-	-	RMSE	Box Cox
Tillema et al. (2006)	1	15	-("BP")	1	Vary (1-20)	-	- ("1)	- ("1")		SN	-	100	-	-	100	RMSE	GM
Yaldi et al. (2009b)	1	36	VLR	1	Vary (1-20)	-	1	1	1	SN	1	vary	vary	vary	Up to 500	RMSE	GM
Yaldi et al. (2009a)	1	36	LM	1	10	-	1	1, 2, 3	-	SN	10	100	-	-	Up to 100	RMSE	GM
Yaldi et al. (2010a)	1	36	BP, VLR, LM	1	10	-	1	1	-	SN	10	100	-	-	Up to 100	RMSE & SRMSE	GM
Yaldi et al. (2010b)	2	36	BP, VLR, LM	1	10	-	1	1	-	SN	10	100	-	-	Up to 100	RMSE	-
Shir-mohammadli et al. (2010)	1	186	LM		10	-	1*& 2		1			60	15	25	-	MSE	Logit-Fratar
Yaldi et al. (2011a)	1	36	BP, VLR, LM	1	10	-	1	1	-	SN	10	100	-	-	-	RMSE	GM
Yaldi et al. (2011b)	1	36	BP, VLR, LM	1	10	-	1	1	2		30	40	30	30	Up to 100000	RMSE	GM

[1] = All model is MLFFN; [2] = based on the reported best performance; [3] = Number of experiment for each model structure/scenario,

SN = simple normalization; LT = linear transformation; Subject 1= people; Subject 2= commodity; ZN= number of zones; PM= performance measurement; RMSE= root mean square error; SRMSE= standardized RMSE; MSE= mean square error; ARV= average relative variance

TA= Training Algorithm; HL= Hidden Layer; AF= Activation Function; DN= Data Normalization; Exp= Experiment number; OL= Output layer

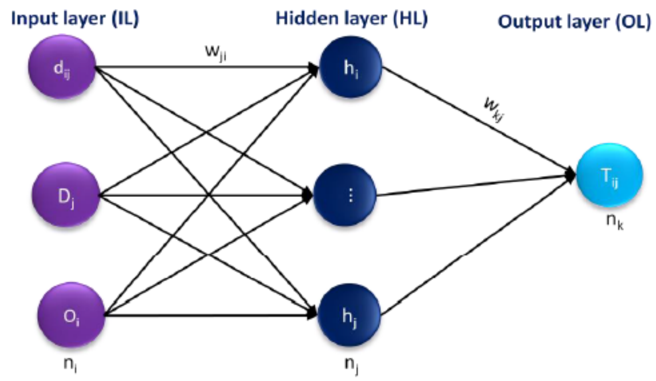

Figure 1 The trained NN model structure

Table 2. Model scenarios

Scenario number	HL number	HL node number(s)	Number of scenario
1	-	-	1
2	1	1-20	20
3	2	1-20	20
Total number of configurations			**41**
Total number of experiments			**41 × 30 = 1230**

In this paper, the behaviour of the neural model in the spatial interaction modelling is explored, for various conditions, namely:

- Random initial weights
  The network will be trained with different initial weights assigned to the neural models. There will be 30 experiments for different initial connection weights.
- The experiments are undertaken for different number of HL(s)
  There are three types of neural models based on the HL numbers. These are (1) The neural models without HL, (2) With one or single HL, and (3) With two or double HLs. The HL number is limited to a maximum of two layers. It is rare to find a study which has more than two HLs, especially in transport modelling. Zhang et al. (1998) found that many of the studies which deployed NN used either one or two HLs. Most studies in travel demand modelling have also used a single HL. Therefore, the maximum number of HL used in this study is two layers.
- The experiments are undertaken for different number of nodes in HL
  The HL node numbers vary from 0 to 20 nodes. There will be no node in HL when HL is not used in the model. It is expected that a moderate number of nodes in HL can be found through these experiments. A moderate number of hidden layer nodes will reduce the model complexity and can help to avoid over fitting.

The total number of trials related to the conditions above is more than 1200 trials. It is expected that the behaviour of neural models for spatial interaction toward different numbers of HL and nodes can be explored. It is also expected the neural models with sufficient node numbers in HL can be defined through these experiments.

Yaldi et al. (2009b) trained the neural models for a single HL, with its node number ranging from 1 to 20 nodes. It did not report the neural model without the HL. Mozolin et al.

(2000) tested for HL node numbers of 5, 20, and 50 nodes. Neither of these studies considered the neural models where there is no HL. In addition, the number of experiments for neural model architecture with different HL node numbers was limited to five. This is now considered inadequate to show the variation in the neural model testing performance, for two main reasons, namely:

- The initial weight configuration is defined randomly. It means that the neural model will have different performance for different initial connection weight configurations. The variation needs to be evaluated, so that the neural models can be freed from dependency on any specific value of initial connection weights.
- The training of the neural model will be stopped due to several conditions. One of these is when the minimum gradient is reached. Some neural models trained with LM appear to experience this condition. No one knows in which configuration of the initial connection weights the neural model will be stopped. The effect of gradient stop towards the estimations must be checked. Thus, the neural models are required to be trained with a sufficient number of initial connection weight configurations.

Exploring the behaviour of the neural models trained with different state in the HL will contribute to the development of NN application in not only the transportation domain, but also in other sectors. This study will also fill the gap from previous studies related to the configuration of neural models for different HL and HL node numbers.

3. MODEL DATA

The work trip data based on the 2005 home interview survey conducted in Padang City, West Sumatra, Indonesia is used in this section (see Figure 2 for Padang City map). This includes 36 zones as depicted by Figure 2. Therefore, there are 36 x 36 input sample patterns, consisting of 1296 inputs. Based on trip purpose, the work trip contributed about 16 per cent of total trips. Shopping and school trips were about 9 and 20 per cent respectively while other trip type shares more than half of the total trips (see Figure 3).

4. MODEL OUTPUT AND DISCUSSION

Results are available for training more than 1200 experiments. Analysis of the results seeks to explore the behaviour of neural models with different HL and nodes numbers. The best expected structure of neural models can then be derived from these experiments.

‘

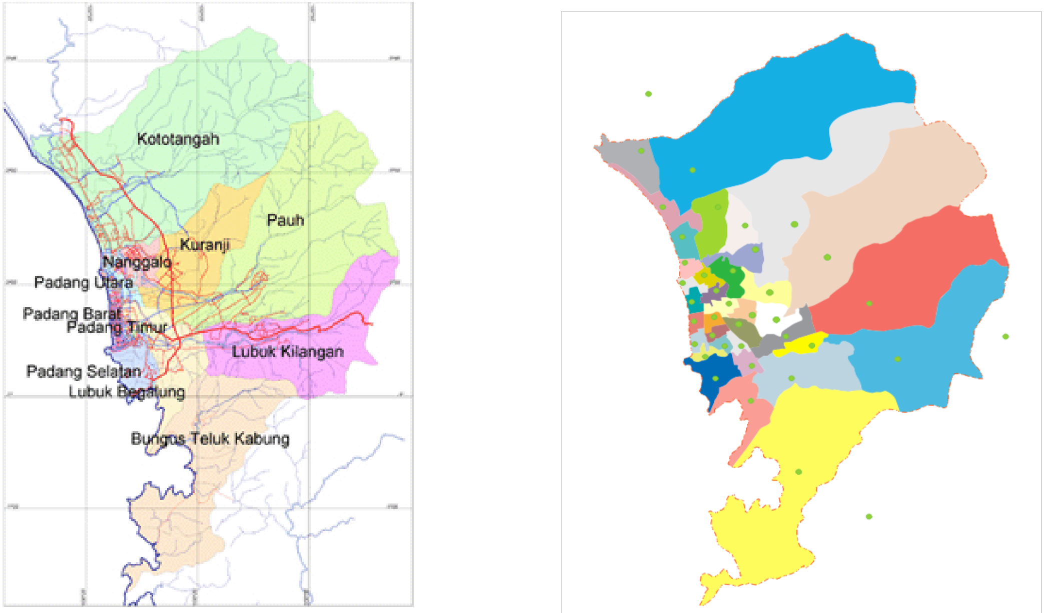

Figure 2 Padang city and districts, and traffic analysis zone

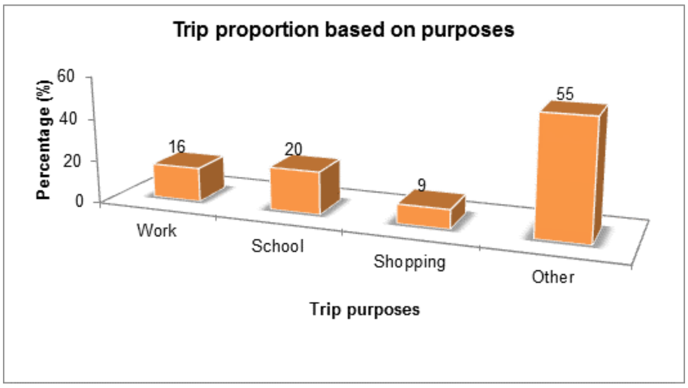

Figure 3 Trip shares for different trip purposes

4.1 Node and epoch number relationship

The number of stopped epoch for each neural model with different number of HL and nodes can be seen in Tables 3 and 4 show. The average number of stopped epoch or iterations for different number of nodes in the HL is reported in Figure 4. It represents the average stopped epoch number for the neural models with single and double HLs. Each HL has a different number of nodes, ranging from one to 20. The neural model without HL has an average epoch number of 11.

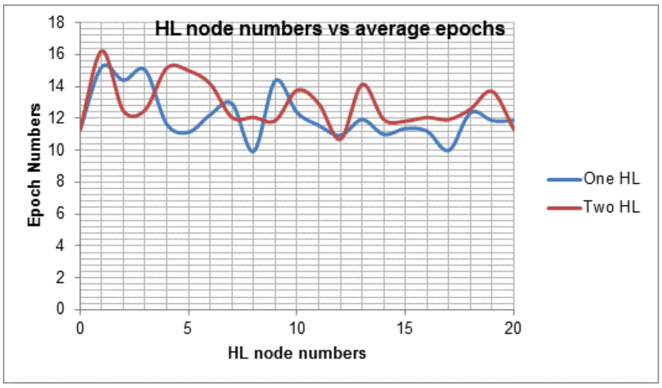

Figure 4 Hidden node number versus epoch number

Figure 4 shows that there is an irregular relationship between HL node and average stopped epoch numbers. It occurs for both single and double HLs. There are at least two results to be drawn from this figure. Firstly, a higher number of nodes in HL do not always result in a lower average stopped epoch number. Secondly, more HL in the neural model structure is also not followed by a faster convergence speed of the training.

The average stopped epoch is within the ranges of 10 to 15 and 11 to 16 epochs for single and double HL respectively. Thus double HL neural models require a slightly higher average stopped number than the single HL neural model. In fact, more than half of the neural models with single HL trained with less average stopped epoch number than the neural models with double HLs. However, there is likely to be a trend where the average number of stopped epoch drops when there are more nodes in HL. The average stopped epoch number for the neural model without HL is 11 iterations, the same as the double HL neural model with 20 nodes. This is slightly higher than the single HL neural model, also with 20 nodes.

The trend can be viewed in more detail by showing the number of stopped epoch for different HL and HL node numbers, instead of the average one as in Figures 5 and 6. The first figure displays the relationship between stopped epoch numbers with different node number for single HL. Figure 6 shows the double HLs. Both figures indicate that the stopped epoch number varies and fluctuates for different number of HL nodes, similar to the finding from Figure 4.

The figures only show specific number of HL nodes only. Neural models with 0, 5, 10, 15, and 20 HL nodes are selected for both single and double HL neural models. The graph would be too crowded if more details are plotted, potentially obscuring any relationships. However, the complete epoch number for each experiment can be seen in Tables 4 and 5.

The most extreme number of stopped epoch belongs to neural models with 20 and 5 nodes for single and double HLs respectively. It also can be observed from the figures that some training stopped at a very early stage, even after the first epoch. This situation occurs in almost all of the neural models configurations, except for the one without HL. An important information result to be drawn from the figures is that there are no linear relationships between the HL, HL nodes, and the stopped epoch numbers.

Table 3. Stopped epoch number for single HL

Experiment	Hidden layer node number																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	11	20	22	26	10	1	24	1	1	12	22	17	12	14	12	18	11	15	18	11	10
2	11	18	1	8	1	24	1	16	13	12	14	14	11	9	9	10	13	11	11	1	11
3	10	11	17	19	13	16	8	16	12	10	23	11	19	11	15	10	12	1	1	11	20
4	11	16	11	12	29	14	1	13	1	9	12	9	11	12	17	12	8	1	11	27	16
5	9	12	12	1	17	24	1	13	16	15	10	13	13	12	29	10	13	20	11	1	19
6	11	11	10	16	12	1	12	13	1	22	14	9	10	10	15	10	12	7	13	12	10
7	11	12	12	18	10	27	14	13	1	9	13	11	15	14	1	10	13	1	17	11	14
8	11	29	33	15	13	9	24	13	16	11	9	17	9	10	11	11	28	10	11	10	12
9	20	11	15	13	13	15	20	11	10	10	9	9	9	14	14	1	10	12	13	11	12
10	12	10	13	16	16	12	12	11	13	11	14	21	1	11	12	10	13	12	19	11	14
11	11	15	1	14	14	19	19	1	12	21	8	12	17	14	12	16	8	18	17	11	10
12	10	28	1	15	11	12	22	23	1	19	27	10	12	12	1	10	12	10	11	12	1
13	10	14	24	11	15	1	13	13	1	1	1	16	11	14	1	15	12	11	10	13	37
14	10	13	14	18	21	14	13	12	11	1	11	16	10	20	14	12	12	13	11	10	9
15	11	34	14	12	1	8	11	23	14	19	11	1	13	16	9	12	1	9	17	11	9
16	9	13	16	11	1	1	1	16	14	20	11	21	13	11	13	1	10	12	15	31	10
17	12	13	1	15	11	1	10	13	12	13	8	14	10	1	10	9	15	19	8	13	18
18	11	17	11	23	12	22	1	12	10	20	17	10	14	10	11	11	10	1	12	13	9
19	11	10	22	14	1	11	15	12	11	23	13	11	11	9	16	12	16	1	11	13	14
20	11	13	29	1	1	11	1	13	15	17	12	12	1	14	11	10	12	18	9	11	1
21	11	13	28	21	12	18	11	10	10	12	10	11	9	17	15	10	1	1	10	12	14
22	10	16	1	23	19	12	10	12	13	15	25	1	8	10	12	16	11	15	13	33	1
23	23	9	11	11	1	1	12	17	14	12	19	12	13	13	1	11	9	11	27	11	9
24	10	15	18	14	1	19	12	14	20	11	9	1	12	11	13	9	10	13	1	11	1
25	9	14	22	13	12	9	20	15	11	14	9	13	10	9	8	9	12	20	15	11	12
26	9	17	10	13	28	1	12	1	10	10	10	10	9	13	1	14	1	1	13	1	14
27	11	13	16	14	12	12	12	12	1	16	16	10	13	11	14	14	17	9	16	10	17
28	11	15	27	12	13	8	26	15	15	27	12	24	11	10	1	14	12	8	9	10	1
29	11	12	1	24	13	10	13	22	18	26	1	1	12	13	19	22	13	10	13	12	14
30	11	13	19	28	16	1	16	12	1	13	1	10	9	13	13	12	9	10	8	1	17

Table 4. Stopped epoch number for double HL

Experiment	Hidden layer node number																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	11	20	17	10	13	1	8	1	1	12	16	1	12	14	11	14	14	12	10	12	9
2	11	14	9	1	10	1	11	13	13	13	10	15	11	11	10	12	15	14	9	1	8
3	10	24	14	1	14	19	17	12	12	14	14	10	11	22	12	14	1	13	14	12	23
4	11	28	11	12	12	18	1	14	14	13	25	10	10	7	10	11	11	12	11	10	10
5	9	10	22	9	15	12	26	14	14	1	19	20	22	13	15	13	11	14	22	13	13
6	11	12	1	15	16	16	12	1	1	9	1	1	11	15	9	15	12	11	14	23	13
7	11	13	15	9	28	11	20	17	17	11	21	14	13	7	12	15	10	14	9	11	11
8	11	14	11	9	15	14	13	1	1	14	11	13	12	20	17	12	12	13	11	12	12
9	20	12	15	12	1	1	10	16	16	10	11	11	12	10	8	16	9	13	12	19	1
10	12	12	24	10	1	16	9	13	13	20	10	18	12	14	10	1	11	12	10	19	15
11	11	38	1	15	52	22	17	27	27	10	11	13	10	15	10	10	1	11	12	16	12
12	10	14	1	10	11	24	51	12	12	10	15	10	14	21	11	10	15	23	12	8	10
13	10	15	29	13	12	12	13	12	12	12	10	11	11	22	1	11	15	11	9	11	10
14	10	14	14	16	19	10	10	14	14	10	12	10	12	17	13	16	11	20	9	14	10
15	11	11	14	11	13	40	13	17	17	9	19	19	18	22	13	11	10	11	12	17	20
16	9	12	1	24	12	15	21	12	12	15	22	10	0	1	10	13	11	10	12	9	10
17	12	9	34	15	16	16	14	10	10	10	13	9	13	10	17	10	14	13	11	15	11
18	11	50	24	9	10	11	11	12	12	11	11	14	1	13	15	11	13	11	11	10	14
19	11	13	12	20	20	1	13	11	11	16	1	11	1	23	24	9	14	8	10	12	10
20	11	18	1	1	11	11	10	1	1	10	10	32	8	7	1	12	14	10	18	14	12
21	11	11	9	1	19	13	13	1	1	15	16	9	1	13	11	11	13	10	17	17	24
22	10	23	9	28	19	23	11	10	10	11	10	12	13	12	16	14	11	14	21	11	11
23	23	12	23	39	12	13	9	9	9	15	12	10	14	19	14	13	13	10	11	21	9
24	10	16	19	20	21	8	17	24	24	19	17	9	11	13	11	11	17	1	12	12	1
25	9	12	11	1	16	25	1	10	10	14	13	13	12	16	12	14	14	11	18	18	12
26	9	13	1	16	9	16	13	16	16	1	16	12	1	27	13	11	18	12	9	10	18
27	11	11	7	15	18	21	14	13	13	9	1	36	14	13	13	16	19	10	12	9	16
28	11	15	1	15	16	13	13	23	23	14	34	14	17	1	12	14	14	10	17	22	1
29	11	10	8	1	1	29	12	9	9	15	11	20	12	13	13	14	9	12	10	22	12
30	11	11	16	18	23	18	22	17	17	13	21	1	12	13	14	1	10	12	13	11	1

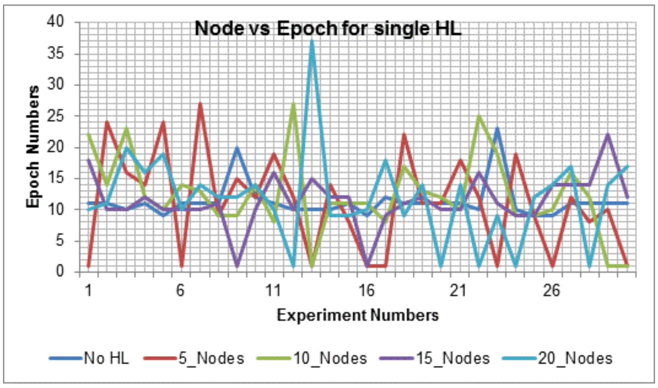

Figure 5 Hidden node number versus epoch number, single HL

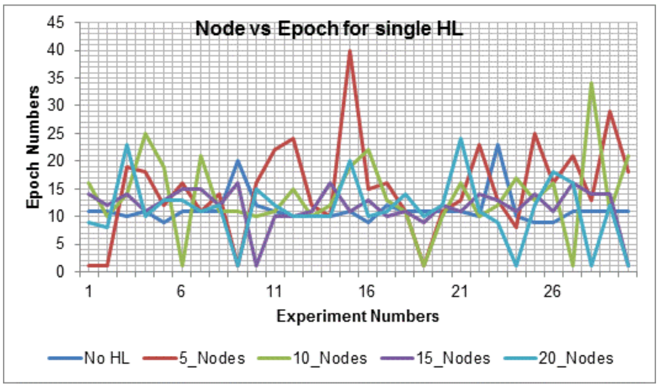

Figure 6 Hidden node number versus epoch number, single HL

4.2 Failed training

As indicated before, some neural model trainings were stopped at a very early stage. They were stopped after the first iteration, except for the neural models without HL. This situation is poor for neural model performance as the training cannot proceed to reach its best outputs. The training will produce its best performance for the testing when it is stopped due to the increasing error in the validation dataset. It is allowed to increase five times consecutively for this model. This stop is called validated stop. The model should perform better when trained with more epochs and stopped due to validation error increase.

Another condition for the training to stop is when the minimum gradient of the training is reached. When the training is stopped due to this condition, the testing performance will tend to be poor, as it is unlikely that the training reached the optimum configuration of the connection weights. This stop is called minimum gradient stop. The training stopped due to minimum gradient is reached is considered to have failed. The percentage of failed trainings for neural models with different HL and HL node numbers is showed in Figure 7 and also reported in Table 5. There are 20 neural models for single HL and double HLs respectively. Each of the neural models is trained for 30 trials.

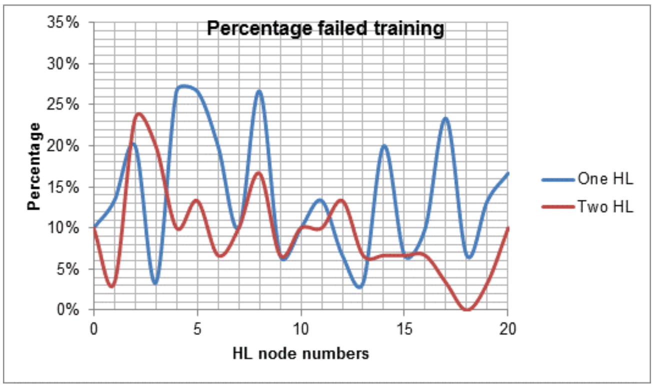

Figure 7 Percentage of failed training

Table 5 Percentage failed training and average epoch

HL node #	Percentage failed training (%)		Average epoch #	
	One HL	Two HLs	One HL	Two HLs
1	10	10	11	11
2	13	3	15	16
3	20	23	14	12
4	3	20	15	13
5	27	10	12	15
6	27	13	11	15
7	20	7	12	14
8	10	10	13	12
9	27	17	10	12
10	7	7	14	12
11	10	10	12	14
12	13	10	12	13
13	7	13	11	11
14	3	7	12	14
15	20	7	11	12
16	7	7	11	12
17	10	7	11	12
18	23	3	10	12
19	7	0	12	13
20	13	3	12	14

Failed training occurs for almost 97 per cent of the double HL neural models, while it occurs for all of single HL neural models. It can be seen in Table 5 that the percentage of failed training varies for different number of HL and HL nodes. On average, the single HL neural models have 14 per cent failed trainings, and double HLs have nine per cent. Only the double HL neural model with 19 hidden nodes has all the 30 trials successful. It appears that in average for 30 trials there will be more than four and almost three of the total trials will fail to converge with optimum configuration of connection weight, for single and double HL neural models respectively.

The neural models where there is no HL had an average of 10 per cent stopped training. From 30 experiments there are three trainings which failed. An indicator of failed training is the incidence of zero trip estimates as seen in Figures 8 and 9. These figures show the failed training results for some of single and double HL neural models. The results are indicated by the points touching zero estimation. The frequency of points at zero level along the "x" axis represents the number of failed training. The 'obs' line represents the observed trip or actual trip numbers.

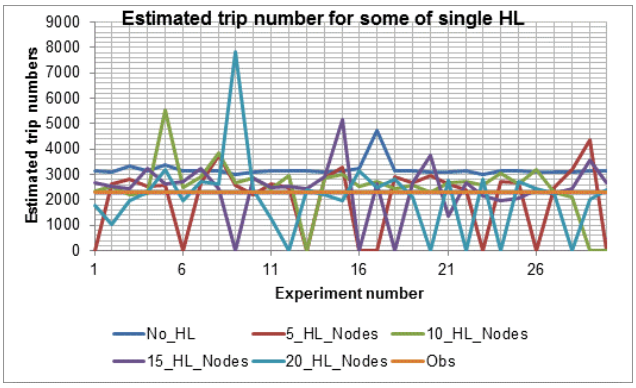

Figure 8 Impact of failed training_calibration_1HL

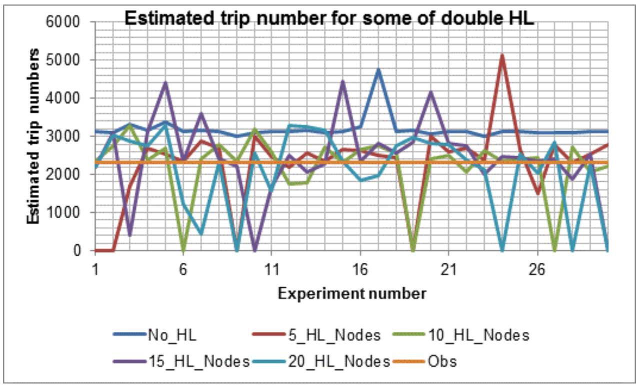

Figure 9 Impact of failed training_calibration_2HL

Since the initial weight configuration is randomly defined, nobody knows in advance which initial weight configuration of the neural model will fail to converge with optimum connection weight configuration. This is a serious problem and hence needs to be overcome.

4.3 Model performance comparison: Root mean square error

The behaviour of the neural models trained toward the number of stopped epoch, HL and HL nodes number was described in the previous section. The discussion now turns to the calibration and testing performances covering the root mean square error (RMSE) and the coefficient of determination (R^2).

The calibration and testing performances of the neural models for different configurations of HL and HL nodes are displayed in the Figures 10-13. The information drawn from these figures are (1) The common shape of both figures indicates that there is a greater variation among the neural models with single and double HLs than the model without

HL, and (2) The RMSE for neural models with single HL has a greater variation than the double HL neural models.

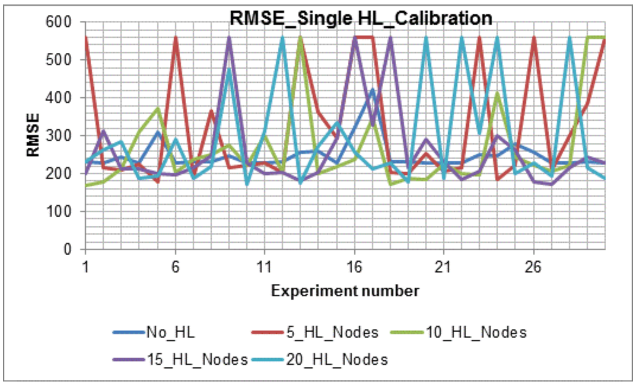

Figure 10 RMSE for single HL (Calibration)

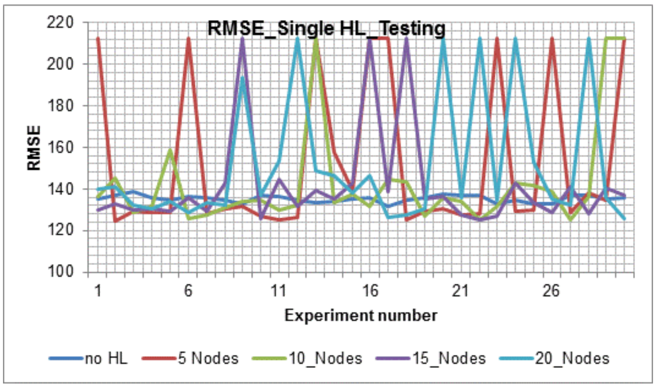

Figure 11 RMSE for single HL (Testing)

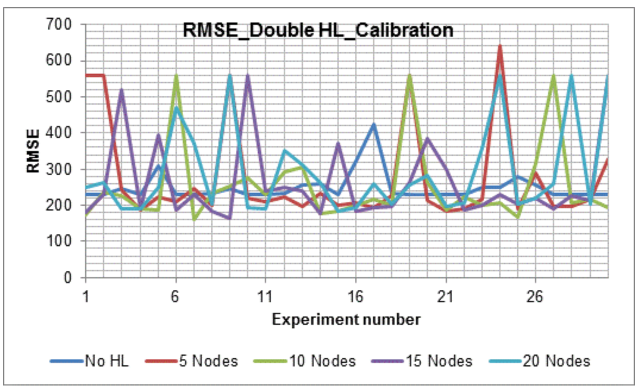

Figure 12 RMSE for double HL (Calibration)

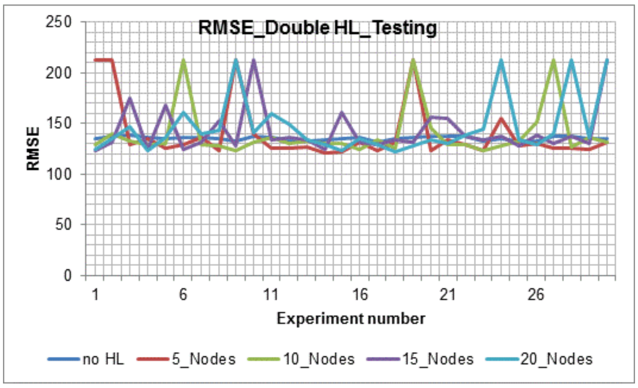

Figure 13 RMSE for double HL (Testing)

The RMSE varies from a maximum of 212 to a minimum of 120 trips. It can be seen that there are many points in the figures which reach the maximum RMSE. This value of RMSE belongs to all neural models where trainings stopped as its minimum performance gradient is reached (the minimum gradient is set to be 10^{-10}), i.e. failed trainings. Thus, it is obvious that the training stopped due to the minimum gradient condition will greatly affect neural model performance. The performance of calibration defines the performance of testing as indicated by the common trend in these figures, although the magnitudes are different (see also Table 6).

Using double HLs in the neural model architecture is sometimes believed to avoid the training reaching its minimum gradient too early. The experiments show that it is not always true. There are two indicators, namely (1) A single HL neural model with a specific HL node numbers which experiences early stop does not experience the same situation when double HL is used, and (2) The double HL neural model with a specific numbers of HL node experiences early stop, while it does not experience that situation only a single HL is used. For example is the neural model with 5 HL nodes, the sixth trial. It experiences early stop. See Figures 10 and 12 for calibration and Figure 11 and 13 for testing. It is also listed in Table 6. This table shows the relationship between stopped epoch number, RMSE, and R^2 for hidden layer with five nodes only. It can be seen from Figures 8 and Table 6 that the neural model with 5 nodes experiences gradient stop for the sixth experiment for single HL. However, the training was successful when double HLs are used (see Figure 9). Thus, the RMSE for the failed training is much higher than the successful one as depicted by Figures 10-13 for both calibration and testing levels.

Another example is the double HL neural model with 20 HL nodes which experience early stop, yet it did not experience that situation when there is only a single HL as illustrated by Figures 10 and 12, for the 30th trial. There are also neural models with single and double HLs experiencing early stop for the same number of HL nodes. This is HL node numbers of 20, the ninth experiments. Thus, adding extra HL may either prevent or trigger early stop.

All of the performance figures demonstrate that the RMSE for neural models without HL are relatively plateaued, unlike the neural models with HL, where the RMSE continuously fluctuates. RMSE for the zero HL neural models ranged from 132 to 138 trips. The maximum RMSE for this model is much below the neural models with HL. However, its minimum RMSE is much higher than the lowest RMSE for neural models with HL. The figures also reflect the impact of the existence of HL in the neural model architecture. It reveals that the educated usage of HL may reduce the total error to below that of the neural model without HL. This is important. Although the latter neural model has a much more relatively stable RMSE than the one with HL, its RMSE is probably still higher than the best RMSE calculated by different modelling procedures, such as the gravity model. Figures 10-13 indicate that there are many points where the RMSE is under the line representing the RMSE for the neural without HL. Performance should be even better if the early stop problem faced by both single and double HL neural models is overcome.

Table 6 Stopped epoch number, RMSE and R^2 (Five hidden layer node)

Trial #	Single HL					Double HL				
	Stopped epoch #	RSME Cal	RSME test	R^2 Cal	R^2 test	Stopped epoch #	RSME Cal	RSME test	R^2 Cal	R^2 test
1	1	560	212	0.017	0.004	1	560	212	0.386	0.036
2	24	215	124	0.814	0.509	1	560	212	0.018	0.032
3	16	210	129	0.835	0.461	19	244	129	0.793	0.533
4	14	226	129	0.799	0.495	18	188	135	0.849	0.410
5	24	177	129	0.865	0.462	12	225	126	0.791	0.500
6	1	560	212	0.000	0.054	16	209	129	0.824	0.504
7	27	190	128	0.845	0.472	11	245	136	0.766	0.408
8	9	365	130	0.771	0.466	14	200	123	0.836	0.513
9	15	215	132	0.805	0.442	1	560	212	0.043	0.000
10	12	222	127	0.821	0.490	16	219	140	0.796	0.381
11	19	230	125	0.810	0.517	22	211	126	0.815	0.502
12	12	204	126	0.831	0.518	24	222	125	0.803	0.529
13	1	560	212	0.003	0.018	12	197	127	0.841	0.474
14	14	362	158	0.454	0.260	10	232	120	0.813	0.539
15	8	294	140	0.677	0.378	40	201	122	0.829	0.524
16	1	560	212	0.007	0.044	15	206	131	0.819	0.456
17	1	560	212	0.005	0.112	16	194	123	0.841	0.508
18	22	202	125	0.825	0.499	11	227	132	0.809	0.435
19	11	199	129	0.842	0.458	1	560	212	0.000	0.042
20	11	255	130	0.774	0.453	11	214	123	0.826	0.516
21	18	207	127	0.817	0.485	13	183	134	0.858	0.428
22	12	217	128	0.828	0.479	23	192	129	0.842	0.463
23	1	560	212	0.168	0.184	13	216	123	0.809	0.516
24	19	185	129	0.854	0.471	8	642	154	0.689	0.300
25	9	222	130	0.799	0.497	25	190	127	0.846	0.476
26	1	560	212	0.000	0.001	16	288	130	0.658	0.456
27	12	208	129	0.828	0.503	21	197	126	0.834	0.508
28	8	302	138	0.731	0.394	13	198	125	0.835	0.518
29	10	386	135	0.762	0.4117	29	216	124	0.814	0.5086
30	1	560	212	0.002	0.010	18	329	131	0.752	0.451

Note: stopped epoch one represents the failed training, Cal = Calibration, Test= Testing

4.4 Model performance comparison: Determinant coefficient

Like the RMSE, the R^2 for the neural models with single and double HLs has an oscillatory pattern. It has some points under the line representing the R^2 for the neural model without HL as represented by Figures 14-17 for calibration and testing. Like the RMSE, the performance of calibration also determines the performance of testing. The figures also represent the impact of early stop due to the minimum gradient condition as described before. Therefore, it is expected that this situation can be improved if the early stop can be avoided.

The figures show the R^2 for single HL neural models is slightly lower than the double HL neural models. The maximum value of R^2 is 0.539 and 0.542 for single and double HL neural models consecutively. It is 0.492 for the neural model without HL. Therefore, these figures suggests that (1) There is more opportunity for the neural model with HL to perform better than that without HL if the early stop problem can be fixed, and (2) There is also more opportunity for the neural model with HL to perform at least at the same level as other modelling procedures when there are no early stop trainings.

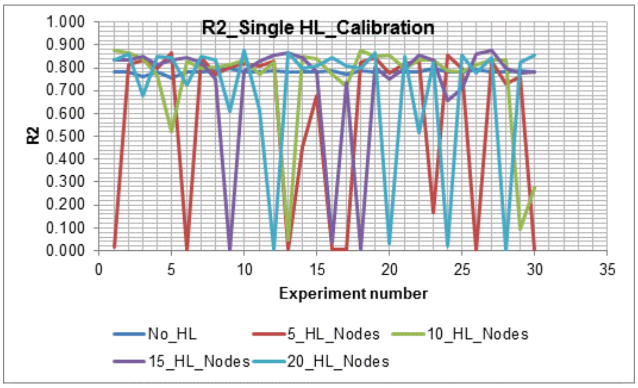

Figure 14 R^2 for single HL (Calibration)

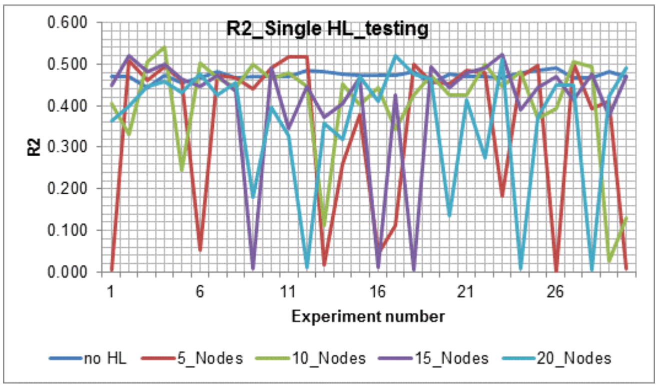

Figure 15 R^2 for single HL (Testing)

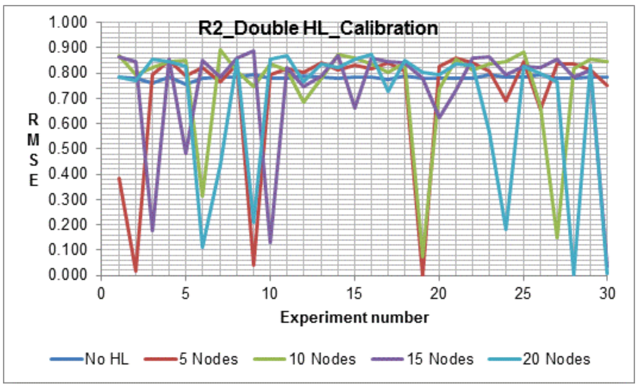

Figure 16 R^2 for double HL (Calibration)

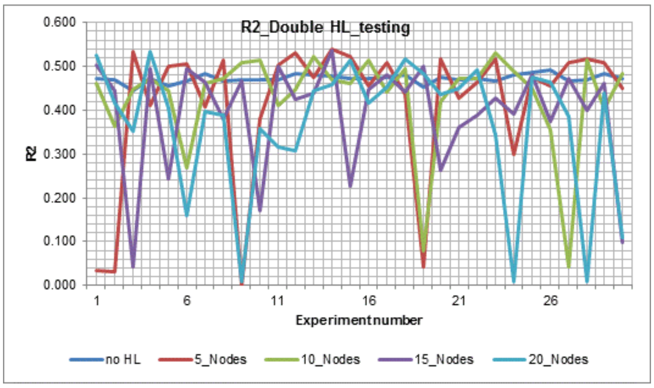

Figure 17 R^2 for double HL (Testing)

The consequences of failed training can be seen from the RMSE and R^2. Table 6 shows the example of failed training effects toward the model performance in term of RMSE and R^2. As previously described, the failed training indicated by the zero estimation produces much higher RMSE than the successful training. For R^2, it is indicated by the value reaching zero points along "x" axis, for both calibration and testing levels. Another finding is that the calibration outputs have better goodness-of-fit than the testing outputs, as expected. However, the magnitude of RMSE for calibration is higher than for testing. This is related to the nature of measurement, where the RMSE is a parametric measurement while R^2 is not. Because the size of sample and total trip numbers for calibration is bigger than that for testing, the magnitude of RMSE for calibration is bigger than for testing (see Figures 10-13).

To conclude, it can be seen that the performance of neural models is nonlinear towards the number of HL and HL nodes. Taking a moderate number of HL nodes such as ten nodes is recommended. More nodes will increase the model complexity and over fitting chance. In addition, using more or less than ten nodes for HL is not always advantageous or disadvantageous to the model performance. Then, the RMSE and R^2 for calibration and testing may be higher and lower than the expected ones. However, it should be noted that other issues in the model testing such as the output back scale and balancing have yet to be considered. It is expected that better performance of neural models can be obtained after the output is properly back scaled and balanced.

5. CONCLUSIONS

From the results and discussion above, some important findings can be drawn. These findings are important in the next modelling process and determine the performance of neural model in forecasting trip distribution. They form the basis of the model development. Those findings are summarized below.

- More HL in the neural model is also not followed by a faster convergence speed of the training.
- There is an apparent trend for the average number of stopped epoch to drop when there are more nodes in the HL.
- There is nonlinear relationship between the HL, HL nodes, and the stopped epoch numbers.
- There is a greater testing performance variation among the neural models with single and double HLs than the model without HL in estimating trip numbers.
- The performance for neural models with single HL has a greater variation than the double HL neural models in estimating tip numbers.

Finally, a moderate number of nodes, such as ten nodes, are suggested for HL as more nodes will increase the model complexity and hence lead to potential over fitting.

REFERENCES

BLACK, W. R. 1995. Spatial interaction modeling using artificial neural networks. *Journal of Transport Geography,* 3**,** 159-166.

CARVALHO, M. C. M., DOUGHERTY, M. S., FOWKES, A. S. & WARDMAN, M. R. 1998. Forecasting travel demand: a comparison of logit and artificial neural network methods. *The Journal of the Operational Research Society,* 49**,** 711-722.

DOUGHERTY, M. 1995. A review of neural networks applied to transport. *Transportation Research Part C: Emerging Technologies,* 3**,** 247-260.

MOZOLIN, M., THILL, J. C. & LYNN, U. E. 2000. Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. *Transportation Research Part B: Methodological,* 34**,** 53-73.

MURAT CELIK, H. 2004. Modeling freight distribution using artificial neural networks. *Journal of Transport Geography,* 12**,** 141-148.

SHIR-MOHAMMADLI, M., SHETAB-BUSHEHRI, S. N., POORZAHEDY, H. & HEJAZI, S. R. 2010. A comparative study of a hybrid Logit–Fratar and neural: network models for trip distribution: case of the city of Isfahan. *Journal of Advanced Transportation,* 45**,** 80-93.

TEODOROVIC, D. & VUKADINOVIC, K. 1998. *Traffic Control and Transport Planning: A Fuzzy Sets and Neural Networks Approach*, Massachusetts, USA, Kluwer Academic Publisher.

TILLEMA, F., VAN ZUILEKOM, K. M. & VAN MAARSEVEEN, M. F. A. M. 2006. Comparison of Neural Networks and Gravity Models in Trip Distribution. *Computer-Aided Civil and Infrastructure Engineering,* 21**,** 104-119.

YALDI, G., TAYLOR, M. A. P. & YUE, W. L. Year. Improving Artificial Neural Network Performance in Calibrating Doubly-Constrained Work Trip Distribution by Using a Simple Data Normalization and Linear Activation Function. *In:* Paper of The 32 Australasian Transportation Research Forum, 2009a Auckland, New Zealand. Available at www.patrec.org/atrf.aspx.

YALDI, G., TAYLOR, M. A. P. & YUE, W. L. 2009b. Using Artificial Neural Network in Passenger Trip Distribution Modelling (A Case Study in Padang, Indonesia). *Journal of Eastern Asia Society for Transportation Studies (in press).*

YALDI, G., TAYLOR, M. A. P. & YUE, W. L. 2010. Refining the Performance Neural Network Approach in Modelling Work Trip Distribution by Using Lavemberg-Marquardt Algorithm. *Journal of the Society for Transportation and Traffic Studies (JSTS).*

YALDI, G., TAYLOR, M. A. P. & YUE, W. L. 2011. Forecasting origin-destination matrices by using neural network approach: A comparison of testing performance between back propagation, variable learning rate and levenberg-marquardt algorithms. *The 34th Australasian Transportation Research Forum.* Adelaide, South Australia.

ZHANG, G., PATUWO, B. E. & HU, M. Y. 1998. Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting,* 14**,** 35-62.