Australasian Transport Research Forum 2011 Proceedings
28 - 30 September 2011, Adelaide, Australia
Publication website: http://www.patrec.org/atrf.aspx

Forecasting origin-destination matrices by using neural network approach: A comparison of testing performance between back propagation, variable learning rate and levenberg-marquardt algorithms

Gusri Yaldi[1], Michael A P Taylor[2], Wen Long Yue[3]

[1] PhD Candidate, Barbara Hardy Institute, Adelaide, SA, Australia
[2] Director, Barbara Hardy Institute, Adelaide, SA, Australia
[3] Senior Lecturer, Program Director, Barbara Hardy Institute, Adelaide, SA, Australia

Abstract

The previous studies suggest that the Neural Network (NN) approach is able to model the commodity, migration and work trip flows. However, its generalization performance is poor, compared to the well known doubly constrained gravity model. This paper is intended to fix the testing performance of NN by training the models with the Levenberg-Marquardt (LM) algorithm, while the previous studies used standard Back propagation (BP), Quickprop and Variable Learning Rate (VLR) algorithms. The main difference between those algorithms is the technique used in defining the optimum connection weights. Then, the trained and validated model is used to forecast trip numbers from different Trip production and Trip Attraction dataset. The testing results suggest that the Root Mean Square Error (RMSE) are 168, 152 and 125 for model trained with BP, VLR and LM respectively, while the R^2 values are 0.194 0.315, 0.505. The models trained by using BP and VLR have underestimate of forecasted total trip numbers, while the LM algorithm has slightly higher numbers. The same data set is then calibrated by using Hyman's technique for the doubly constrained gravity model. The testing RMSE and R^2 for gravity model is 127 and 0.507 respectively. It means the NN model has about the same level of error and goodness of fit as the gravity model, for testing level. Based on these results from this study, it can be concluded that the testing performance of NN approach can be refined to the same level as doubly constrained gravity model when the model is trained by LM algorithm.

1. Introduction

1.1 The origin-destination matrices

The information of the volume of people and goods travelling from a group of origin zones to a group of destination zone is a pivotal factor for transportation planning, and traffic management and operation. A matrix is usually used as a means to represent this information. It is a two-dimensional table, where each cell inside the table represents the volume of traffic moving from a specific origin zone to a specific destination zone. It can be various links or routes in an urban transportation network. This matrix is commonly called as the Origin-Destination (O-D) matrix or the trip matrix. Thus, it can be seen that there are three important knowledge in an O-D matrix table, namely (1) the magnitude of traffic volume, (2) the O-D pattern, and (3) the total trip production (P) and Attraction (A). Those are the basic information used in different aspects of transportation systems planning and operations.

The information in the O-D matrix is used for various purposes, such as new road designs and existing road improvements (widening or adding more lanes) due to increasing demand of transport services and facilities. It is also fundamental in investigating the impacts of the implementation of traffic operation scenarios to current traffic situation, social, and environmental sectors. The scenarios may involve the route change, road closure due to road work or maintenance, emergency evacuation due to natural disaster such as earthquake, tsunami, bush fire, and also major flood. The traffic operators can anticipate the situations likely to occur and hence notify the road users with enough time prior to the changes. Therefore, O-D matrices are the basic source of information for many purposes and must be prepared rigorously.

1.2 The method of O-D matrix estimation

According to Taylor et al. (2000) O-D matrices can be estimated by three different ways, namely (1) Direct observation, (2) Synthesis, and (3) Modelling. Among the three methods in constructing and updating the O-D matrices, Doblas & Benitez (2005) suggested that the first method provide the most accurate results, the second one has been widely used since the 1970s as an alternative cheap and quick technique by synthesizing the trip matrix tables from traffic count data. The third method employs techniques such as the gravity model or multinomial logit model. It requires knowledge of an observed matrix for initial calibration, and can still be relatively inaccurate.

Regardless its accurate results, the drawbacks of direct observation were reported in many studies, such as Nihan & Davis (1987), Cremer & Keller (1987), Sherali et al.(2003), Sherali et al.(1994), Doblas & Benitez (2005), and Nie et al.(2005). It includes:

1. Expensive, requiring a large budged
2. time-consuming, especially in interviewing the individual respondents
3. Unable to estimate the impact of changes in trip patterns over time, including the peak and off peak trips
4. Unable to capture the impact of land-use development or changes
5. Biased results are frequently found
6. Causing disruption to the road user, especially during the road side survey/interview, and hence cannot be undertaken often

Therefore, the second method is preferred over the other two. This is partly true. However, many projects, especially in developing countries, purposely involve a high number of human resources in order to provide more jobs due to high rate of unemployment. The labour costs are certainly much lower thus it is much cheaper than in developed countries. In addition, there will always be surveys periodically undertaken for various census or other purposes. Thus, the transport or travel survey can be integrated with those surveys.

It is also infeasible to use traffic counts at many intersection or links in developing or poor countries. There are at least two reasons, namely (1) high initial investment in the transport sector, which is not likely to be the top priority of the decision makers and government officials, and (2) security issues where it is unsafe to install the traffic counters at the roadside. Thus, it will require 24 hour manned security protection, and hence will trigger new cost components. In order word, a method which is applicable in one country could be infeasible in others. There is no single solution, and hence alternative methods are always expected.

With many considerations, the direct method is preferred. However, the trip distribution model, like the gravity model, is often used due to high cost requirement of the direct method. The gravity model is relatively cheap to develop although it is unable to generate accurate and precise estimations. Therefore, an alternative trip distribution model that can forecast the pattern of people and commodity movement will be a significant aid, especially for developing countries. Alternative techniques that can cover the disadvantages of traditional gravity model, but are able to estimate more accurately, are often investigated by many researchers. An alternative modelling technique is not always the best one, however, it may become the most effective and efficient tool for specific case.

Neural Network (NN) approach is one of the alternatives modelling that could overcome the disadvantage of other modelling procedures. This paper describes the application of NN for trip distribution (TD) modelling, especially at the testing or generalization level. The calibration level has previously been discussed in Black (1995), Mozolin et al.(2000), Yaldi et al.(2009a), and Yaldi et al. (2010). This research is part of the effort to build a modelling framework so that the NN can be properly applied in TD modelling.

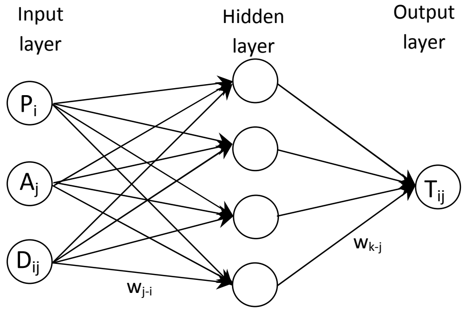
2. The neural network approach: alternative modelling tool

The early use of the Neural Network (NN) approach for the fully constrained spatial movement problem was promoted by Black (1995). It was started by examining the ability of this artificial intelligent technique in estimating trip flow for three region problem. It was found that the NN model had a good performance where the Root Mean Squared Error (RMSE) between observed and estimated trip flow is close to zero (based on non-normalized observed and estimated flow numbers). Then, the total estimated trip flow numbers produced by and attracted to origin and destination zones indicated that the NN model could satisfy both trip Production (P) and trip Attraction (A) constraints. The NN models also had better performances than the traditional doubly constrained gravity model for estimating seven sets of commodity flow. Its error levels can be only half of those for the gravity models. Thus, the NN approach can display exceptionally good pattern recognition.

As a computation technique based on iterative processes, the NN model (see Figure 1 for proposed model structure) estimates the outputs by minimizing the deviation between model outputs and the target values. This process is called training or learning. The neurons/nodes in the first layer take the data into the network. It flows along the connections and is scaled by the values of connection weights (w_{ji} and w_{kj}), which are initially randomly selected. The output is computed by transferring the scaled input through an internal transfer function, which is commonly a logistic function (Black 1995; Dougherty 1995). The estimated trip flow is obtained through the accumulation of the summation from all of nodes in hidden layer, then transformed it by the activation function in output layer node, in this case also the sigmoid function. More details can be found in the literatures, for examples Rumelhart et al. (1986), and Hagan & Menhaj (1994).

The ultimate goal of the training process is minimizing the error term until optimum values of connection weights (w_{ji} and w_{kj}) are achieved through the use of specific training or learning algorithm (TA). Therefore, TA is a fundamental aspect of the NN model performance (Dougherty 1995; Teodorovic & Vukadinovic 1998). The study by Black (1995) used Back Propagation (BP) learning algorithm.

Figure 1: The NN model architecture



The second study that used NN approach for fully constraint trip distribution model was reported by Mozolin et al. (2000). The NN models were used to estimate the work trip distribution numbers in the Atlanta Metropolitan Area. The Quickprop learning algorithm was adopted to train the models. It was reported that the NN models tended to have inconsistent performance with poor generalization ability. In addition, it was unable to satisfy both Production and Attraction constraints. These findings were opposite to the results reported by Black (1995). The third study that adopted NN approach to model work trip distribution was Yaldi et al. (2009b). Although a different learning algorithm was used in the training process, the results were similar to the study by Mozolin et al. (2000). The models were trained by using Variable Learning Rate (VLR) algorithm.

Among these three different studies, the first one suggested that NN approach has exceptionally good patter recognition ability, while the other two suggested the opposite situation. Thus, the latest study by Yaldi et al. (2010) compared three different learning algorithms, namely Back Propagation (BP), Variable Learning Rate (VLR) and Levenberg-Marquardt (LM), at the calibration level. Comparison also involved the doubly constrained gravity model. The uniqueness of this study is that the experiments were repeated 30 times for each model (previous studies had a maximum number of five repetitions, e.g. Mozolin et al. (2000)). Therefore, statistical analysis can be applied to examine the consistency of NN models performance. Further, the iteration number or epoch for each experiment is limited to 100 times, while previous studies had up to 150.000 and 100.000 epochs for Black (1995) and Mozolin et al. (2000) respectively. Training with this high number would cause the models to be over fitted.

The statistical tests suggest that only NN models trained with the LM algorithm have a significantly lower error than the doubly constrained gravity model. It also has a higher goodness of fit (correlation coefficient/r). The P and A constraints are able to be satisfied, also only when the model is trained by LM algorithm. Thus, it can be concluded that neither BP nor VLR algorithms are suitable for training for the fully constrained spatial movement problem. Black (1995) used BP algorithm in the first study that uses NN approach for people and freight movements. Although the NN model for three-region problem suggests that both P and A constraints could be satisfied, however, it is uncertain whether it also could be satisfied for the commodity and migration data. Given these promising results, the Yaldi et al. (2010) is extended to the testing level.

Thus, in this research the NN models are trained by using three TAs, namely the standard BP, the VLR, and the LM. The selection of those three TAs is due to:

1. BP was known as the most famous TA, widely used, and also used in the early spatial movement interaction study by using NN approach as reported by Black (1995) with relatively better performance than well-known doubly constrained-gravity model. However, it was used for calibration level and the epoch number is up to 150000 iterations.
2. As one of technique to improve the drawback of BP by ad-hoc technique, the VLR was found can fix the BP problem (Jacobs 1988; Vogl et al. 1988) and considered as the best method to increase the convergence rate of standard BP (Popescu et al. 2009)
3. There are attempts to search TA with better performance than VLR by keeping the standard neural network and criterion function, however, the optimization technique is modified (Barnard 1992). It can be obtained by utilizing the second order approach, and LM has been proven can be a fast and efficient TA (Wilamowski et al. 2001). It was also found that the LM algorithm can converge in various cases where the VLR and also other forms of second order approach like CG, are failed to converge (Hagan & Menhaj 1994).
4. Spatial movement interaction study is a unique case. It poses extra constraints in the model outputs, and the second order approach based TA like the LM can satisfy those constraints better than VLR and BP (Yaldi, Taylor & Yue 2010)

3. Model development and methodology

3.1 Neural net architecture

The models are developed to forecast the work trip. Its structure is illustrated by Figure 1. It has three input nodes representing the Trip Production (P_i), Trip Attraction (A_j) and Distance (D_{ij}). There is one node in the output layer, the estimated trip number (T_{ij}). Each node is connected to the hidden layer nodes by connection weights w_{ji} and w_{kj}.

The work trip data is based on the 2005 home interview survey conducted in Padang City, West Sumatra, Indonesia. This study area includes 36 zones. Simple data normalization method is used in this study. Simple normalization will convert the input data to the range [0, 1]. Matlab 7.0.1 is used to develop the network, where the initial connection weights are randomly defined by the modelling tool.

The model is trained by feeding the input pattern forward and calculates the cumulative deviation between the target value and the model input. The training is undertaken in batch mode. Then, the error is back propagated to the model in order to adjust each connection weight within the model. That is why the model is called feed forward back propagation neural nets. Then, the model is trained again with the updated connection weight values. This recursive process is undertaken until a combination of optimum connection weight is achieved, indicated by the error of the model less than a specific error threshold. However, the three algorithms have different techniques in the calculation of weight change magnitude. The details of training procedures for BP, VLR, and LM and be found in literatures such as Rumelhart et al. (1986), Volg et al. (1988), and Hagan & Menhaj (1994). Those literatures explain and discuss the BP, VLR, and LM algorithm respectively.

3.2 Training and test sets

There are usually two kinds of input data sets in neural networks, namely training and testing data sets. The training data set is used in estimating the model parameters/variables while the testing data set is for evaluating the forecasting ability of the model. Zhang et al. (1998) suggest another data set, a validation sample. This is used to avoid over fitting and to determine the iteration stop point.

Guidance on how to divide the data sets into developing (training and validation sets) and evaluation sample does not exist so far. However, the problem characteristics, data type and the size of the available data are the factors considered in the data subdivision. In general, the data is split into three blocks randomly. However, spatial interaction modelling poses unique characteristics, especially the doubly constrained model. The data comes in the form of a two-dimensional matrix (the trip or O-D matrix). Generally, it contains the magnitude of traffic, movement pattern, and also the Trip Production (P_i) and Trip Attraction (A_j) (which are the row and column sums of the table). This characteristic will determine the method in splitting the original data into several blocks, namely the training, validation, and testing blocks.

The previous study by Black (1995) did not report how the data was divided for each block as it only undertook the training stage equivalent to the calibration process. Subsequently Mozolin et al. (2000) and Yaldi et al.(2009b) reported that the vector data was randomly split for training, validation and testing in the same way as commonly undertaken by other kinds of studies. The disadvantage of this split method is that it is difficult to measure whether the model outputs are able to satisfy the constraints, for example is the P_i and A_j constraints for the doubly constrained model. Both studies by Mozolin et al. (2000) and Yaldi et al. (2009b) reported that the NN ability to satisfy both P and A constraints was inadequate.

It is unlikely the model output alone can ensure the P and A constraint satisfaction assessment. Thus, it will negatively influence the performance of the NN models. Therefore, an alternative data split method is proposed, that is suitable to the characteristics of the spatial interaction data and model. The data split is undertaken by preserving the matrix form of the data. It can be achieved by randomly selecting the zone number for three blocks, and then forming the data vectors for each block. The results can be seen in Table 1. The advantage of this technique is the performance of the NN model toward P_i and A_j constraints can be assessed and should contribute positively to its performance in forecasting the trip numbers for unseen data set (generalization ability).

Table 1: Data blocks

Block number	Zone numbers	Total (Zones)	Remarks
1st Block, Training zones	20,26,33,6,3,9,32,16,22,1,12,8,19,35	14	40% of total zone numbers
2nd Block, Validation zones	29,25,23,36,10,4,14,15,27,30,34	11	30% of total zone numbers
3rd Block, Testing zones	21,11,31,17,18,2,13,28,24,5,7	11	30% of total zone numbers
Total		36	100%

The procedures to split and to prepare the data by this technique are explained below.

1. Determine the number of total zones in the study area. It is 36 zones for this model.
2. Determine the percentage for each block (training, validation, and testing)
   Literature review suggests that there are several configurations of data sets, for examples are 90% vs. 10%, 80% vs. 20%, and 70% vs. 30%. Dia (2001) used a composition 60%, 10% and 30% for training, testing and validating samples respectively. Carvalho et al. (1998) used 70% vs. 30% as training and testing data set for mode choice modelling by NN approach. Thus, this study used 40, 30, 30 per cent configuration for each block respectively.
3. Compute the number of zones for each block
4. Randomly select the zone number for each block, manually or by using any computing tools
5. Form the O-D matrices for each block
6. Arrange the vector data for each block based on each O-D matrices, so that it forms the column matrix. Each vector data set consists of P_i, A_j, D_{ij} as the Input pattern vector

For example, a three region problem given by Black (1995) is illustrated below (Tables 2 and 3). Black (1995) illustrated how to prepare the vector data for both input pattern and target. The readers are suggested to read the paper entitled “Spatial Interaction Modelling Using Artificial Neural Network”. It can be considered as one of the earliest studies, if not the earliest one, of NN application in the spatial interaction modelling.

Table 2: Flows and marginal totals

	A	B	C	Total/P_i
A	15	4	1	20
B	18	21	1	40
C	17	5	18	40
Total/A_j	50	30	20	100

Table 3: Distance between regions

A	B	C
2	3	4
3	2	5
4	5	2

Table 4: Input pattern

Productions	Attractions	Distance
20	50	2
20	30	3
20	20	4
40	50	3
40	30	2
40	20	5
40	50	4
40	30	5
40	20	2

Table 5: Target values

Flow	O-D pattern	Pattern number
15	1-1	1
4	1-2	2
1	1-3	3
18	2-1	4
21	2-2	5
1	2-3	6
17	3-1	7
5	3-2	8
18	3-3	9

7. Arrange the vector data for each block, which consists of t_{ij} as the target vector (Tables 4 and 5)
8. Normalize the input pattern for each block

3.3 Data normalization

In order to avoid computational problems, to meet the algorithm requirement and to enhance the learning process, the input and output data of neural network must be normalized. Zhang et al. (1998) provides some common normalization method. Those are:

1. Linear transformation to [0,1]
$$x_n = (x_0 - x_{min})/(x_{max} - x_{min}) \quad (1)$$
2. Linear transformation to [a, b]
$$x_n = \frac{(b-a)(x_0 - x_{min})}{x_{max} - x_{min}} + a \quad (2)$$
3. Statistical normalization
$$x_n = (x_0 - \bar{x})/s \quad (3)$$
4. Simple normalization
$$x_n = x_0/x_{max} \quad (4)$$

Another normalization method that can be adopted is by dividing the entire input pattern vector, except the distance, by the total number of trip for each block. The distance is normalized by its maximum value. The advantage of this normalization is that there is only one unique factor used to normalized the P, A, and t_{ij} data, and hence it becomes simpler than dividing them with their maximum values. This method was used by Black (1995).

Besides, normalizing the data to much lower values can avoid the model failing to recognize the unseen relationship between the input and target patterns. This is due to larger numerical values of the data will produce larger numerical summation in each node (except for the input layer nodes). Consequently, the outputs will be transformed to its maximum value for numbers above certain limit. Thus, the model output will be possibly far from the desired one.

The re-normalization computation of the model outputs will be also simpler. It can be multiplied with the cumulative trip production, or trip attraction, dependent upon the constraint condition. Yaldi et al.(2009a) demonstrated that the NN model with simple normalization performs better than the statistical and linear transformation for training or calibration. Thus, this study will use simple normalization; however, the total trip number is used in normalizing the data instead of the maximum value for P_i, A_j, and t_{ij}. Meanwhile, the distance is normalized by dividing it with its maximum value.

4. Model output analyses

4.1 Descriptive statistic discussion

The discussion on the model outputs for three TAs (BP, VLR, and LM) is started by the comparison of the performance of the model trained by each TA. The performance of the model here is measure by looking at the Root Mean Square Error (RMSE) between the observed and the estimated trip numbers. As stated in the title of this paper, the performance comparison is for the testing level only. The testing level can be described as the ability of the NN model to estimate the trip number, where the input trip matrix has not been seen by the NN model before. In order word, the NN model is used to forecast the trip number distribution with new set of input matrix consisting of the new P_i, A_j and D_{ij} data. This data is not used in the training and validation process before. The new data set is supplied to NN model which has been trained and validated before, and produces the testing trip numbers from a set of origin zone i to a set of destination zone j (T_{ij}^t). Thus, the RMSE for testing ($RMSE^t$) becomes:

$$RMSE^t = \sqrt{\frac{\sum(t_{ij}^t - T_{ij}^t)^2}{z}} \qquad (5)$$

Where :

t_{ij}^t = the observed trip number from origin zone i to destination zone j for testing data

T_{ij}^t = the estimated trip number from origin zone i to destination zone j for testing data

z = number of ij pairs
= $i \times j$

The results for 30 trials are reported in the Table 6. The advantage of undertaking the multi experiments is that the impact of different states of initial connection weights to the model performance can be noticed. Table 6 shows the $RMSE^t$ and also the the R^2 coefficient for each experiment and TA. Firstly, the $RMSE^t$ for BP has a wider range than the other two. It indicates that the variation among the BP performance is greater than VLR and LM, while the LM is the lowest. If we compare the number of maximum stopped epoch, BP is trained to a maximum number of 100000 (average 61409) epochs. It is 295 (average 190) and 27 (average 14) epochs for VLR and LM respectively. The stopped epoch means the number of epoch when the training is stopped due to the increase of validation error for more than five times since the last time it decreased.

Allowing the learning rate to shift according to the state of the total error has significantly reduced the number of average training epoch from 61409 to 190 times. Furthermore, using the second order approach like LM has further considerably decreased the average epoch to 14 iterations. This is a moderate epoch number, since lower number will probably diminish the LM testing performance. Using the second partial derivative of the error function as in the LM algorithm has significantly increased the convergence rate. This is true as indicated by Rumelhart et al. (1986) that training the NN model by using second order partial derivative based algorithm will converge faster than BP, which is a gradient descent. The VLR is basically a gradient descent; however, an ad hoc modification is made by allowing the learning rate to vary according to the state of the total error. This ad hoc modification has also increased the convergence rate; however, LM is still much quicker than VLR.

Table 6: The testing performance of NN trained with different TAs

Experiment #	BP		VLR		LM	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
1	178	0.016	161	0.251	136	0.404
2	187	0.000	158	0.243	145	0.330
3	195	0.014	195	0.015	129	0.505
4	201	0.052	162	0.215	132	0.542
5	208	0.056	156	0.249	158	0.246
6	202	0.050	166	0.197	126	0.502
7	172	0.052	130	0.507	128	0.468
8	161	0.249	161	0.249	131	0.445
9	165	0.176	165	0.176	134	0.499
10	174	0.080	174	0.088	134	0.464
11	141	0.376	141	0.376	130	0.480
12	186	0.098	189	0.150	132	0.446
13	142	0.350	140	0.361	133	0.452
14	169	0.107	154	0.239	138	0.403
15	166	0.104	166	0.103	132	0.444
16	192	0.042	159	0.268	145	0.344
17	172	0.029	173	0.027	144	0.426
18	182	0.009	167	0.116	127	0.474
19	181	0.000	151	0.330	136	0.425
20	175	0.027	175	0.026	134	0.425
21	179	0.163	179	0.163	125	0.501
22	146	0.358	155	0.343	132	0.445
23	200	0.031	146	0.318	143	0.483
24	200	0.051	156	0.235	141	0.369
25	183	0.021	152	0.293	138	0.392
26	213	0.023	174	0.022	125	0.507
27	173	0.023	147	0.310	136	0.493
28	212	0.025	156	0.267	169	0.143
29	195	0.080	175	0.017	127	0.506
30	171	0.084	191	0.097	134	0.437
Average1	181	0.092	162	0.208	136	0.433
Average2	168	0.194	152	0.315	125	0.505

4.2 Method to compute the $\overline{RMSE}^t$

There are two methods to compute the average performance (average $\overline{RMSE}^t$) of the model testing. They are as follow.

- **First method**
  1. Compute the un-normalized T_{ij}^{tk}
     $uT_{ij}^{tk} = T_{ij}^{tk} \times nf$, for k= 1, 2, ...i×j
     (6)
     Where “nf“ is the normalization factor

  2. Compute the deviation (Δ^k) between observed trip number and un-normalized T_{ij}^{tk}
     $\Delta^k = t_{ij}^{tk} - uT_{ij}^{tk}$, for k= 1, 2, ...i×j (7)

  3. Compute the cumulative of squared deviation ($(\Delta^k)^2$)
     $\sum(\Delta^k)^2 = (t_{ij}^{tk} - uT_{ij}^{tk})^2$, for k=1, 2, ...,i×j (8)

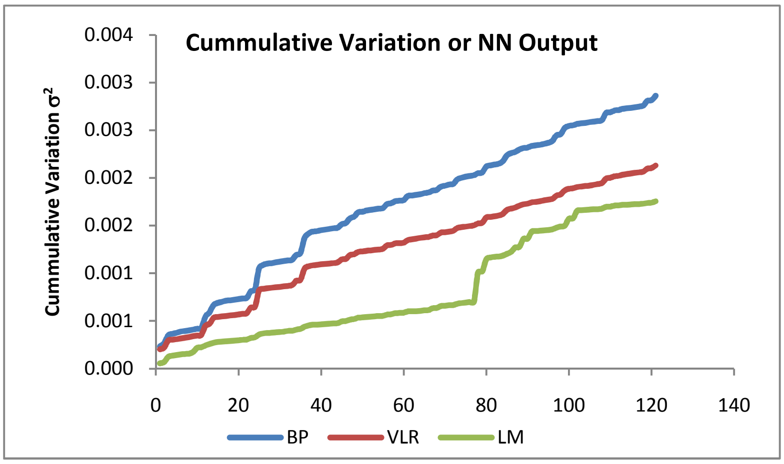
  4. Compute the $RMSE^t$
     $RMSE^t = \sqrt{\frac{\sum(\Delta^k)^2}{z}}$,

5. Compute the average $RMSE^t$
   $\overline{RMSE}^t = \sum \frac{RMSE^t}{n}$
   where n = number of experiment
   = 30 experiments

- **Second method**
  1. Compute the average T_{ij}^{tk}
     $\bar{T}_{ij}^{tk} = \frac{\sum_n T_{ij}^{tk}}{n}$, for k=1, 2, ...,i×j (9)
  2. Compute the un-normalized $\bar{T}_{ij}^{tk}$
     $u\bar{T}_{ij}^{tk} = \bar{T}_{ij}^{tk} \times nf$ (10)
  3. Compute the deviation (Δ^k) between observed trip number and un-normalized $\bar{T}_{ij}^{tk}$
     $\Delta^k = t_{ij}^{tk} - u\bar{T}_{ij}^{tk}$, for k= 1, 2, ...i×j (11)
  4. Compute the cumulative of squared deviation ($(\Delta^k)^2$)
     $\sum(\Delta^k)^2 = (t_{ij}^{tk} - u\bar{T}_{ij}^{tk})^2$, for k=1, 2, ...,i×j (12)
  6. Compute the $\overline{RMSE}^t$
     $\overline{RMSE}^t = \sqrt{\frac{\sum(\Delta^k)^2}{z}}$ (13)

The results can be seen in Table 6. The second method has the average RMSE lower than the first one, by about 6-8%. It is related to the variation among the T_{ij}^{tk} of different experiments. The gap between the second and the first method will get greater when the variation among the T_{ij}^{tk} of different experiments is bigger. The results in Table 6 have proven and Figure 2 graphically supports that. The BP has the highest variation and LM is the lowest. The second method has also lifted up the R^2 coefficients for all TAs.

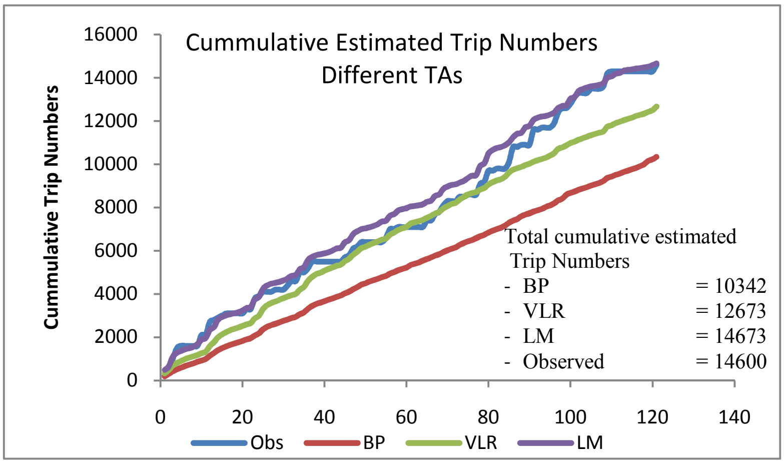
Figure 2: The cumulative variation of NN outputs



4.3 Trip distribution and cumulative trip numbers

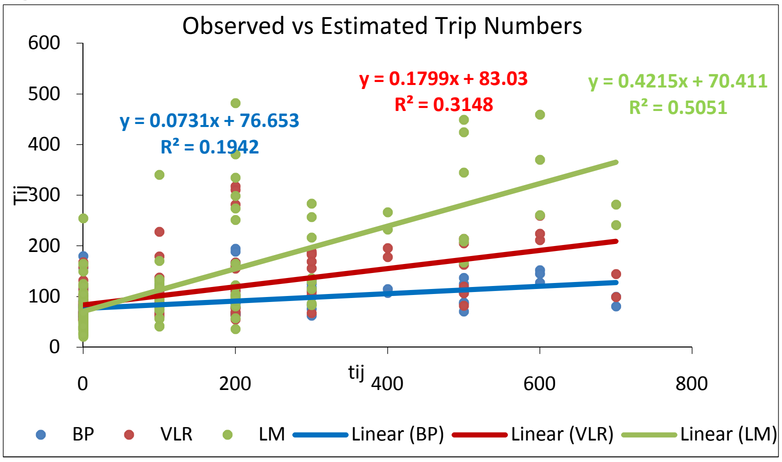
Since the second method shows better results than the first, the next section will also be based on that method. Figure 3 illustrates the cumulative trip numbers for testing estimated by NN models trained with different TAs. Training NN model with BP algorithm generates the trip number much lower than the real one. The gap is getting bigger to the right. It occurs for both BP and VLR. However, it can be seen that VLR has a smaller gap, and it produce good estimation in the middle of the graph. It is the opposite of the LM. The estimation for the first and the last of one third of the graph has lower gap than in the middle part. The total cumulative trip numbers for each TA can be seen in the right bottom part of the graphs.

Figure 3: Cumulative estimated trip numbers



Both BP and VLR generate underestimated results. A linear relationship is displayed in Figure 4. It shows the slopes and the R^2 for each TA is much lower than the LM. On the other hand, the constant for LM is lower than the other two. The estimated trip numbers represented by the vertical axis are located closer to zero points for BP and VLR than the LM. In addition, the estimated trip numbers for BP and VLR are positioned much closer to each other than the LM. Thus, it brings the slopes for BP and VLR down and increases their constants.

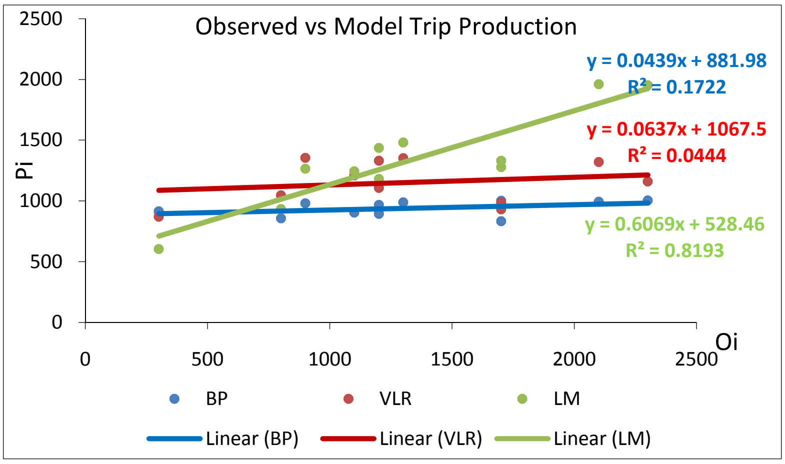
Figure 4: Linear relationship between observed and estimated trip numbers



4.4 Trip production (P_i) and trip attraction (A_j) satisfaction

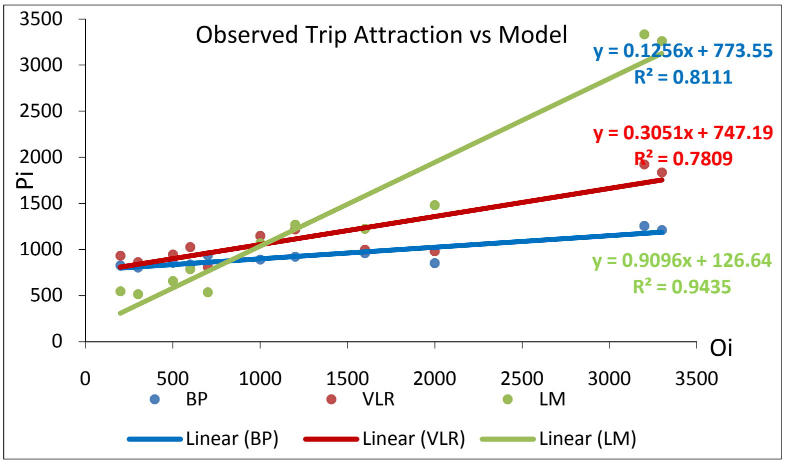
The modelling procedures of forecasting for future or other O-D trip matrices require the independent variables like the Trip Production (P_i), Trip Attraction (A_j), and Trip Length (D_{ij}). The availability of P_i and A_j data determines the model constraint(s). For example there are unconstrained, singly constrained (P_i), singly constrained (A_j), and doubly constrained (D_{ij}) gravity models. Among them, the doubly constrained is the one mostly used to model the journey to work. Therefore, the evaluation of the estimated trip numbers generated by NN models should cover its ability to distribute the trips proportionally and satisfy the constraint(s). In this study, the model is categorized as doubly constrained NN model. The input data consists of P_i, A_j, and D_{ij}. Thus, the NN model to proportionally distribute the trip numbers must satisfy both P_i and A_j constraints. Figures 5 and 6 show the results. In order to simplify the comparison, the intercept is set to zero.

Figure 5: The observed and model trip production



The figures suggest that in general, the NN model has a better ability to satisfy the A_j than P_i constraint. It may be related to the arrangement of input pattern consisted of P_i, A_j and D_{ij}. However, the NN model trained with LM for different configurations of input pattern such as "A_j, D_{ij}, P_i" and "D_{ij}, A_j, P_i", has distributed the trip number with insignificant difference in the outputs.

Figure 6: The observed and model trip attraction



Then, the R^2 coefficient for A_j is much higher than for P_i for three of them. The ability of NN model trained with BP and VLR to satisfy the P_i for testing level is far below the LM. The slope and intercept for BP and VLR is much lower and higher than LM. Thus, it can be seen that the high value of R^2 for A_j of BP and VLR, represents the percentage of data that is close to the fitting line. It does not represent the goodness-of-fit between the observed and model outputs. A majority of the points are near the fitting line and tend to be flat. This suggests that there is a weak relationship between the observed and estimated P_i and A_j, although majority of the points are very close to the fitting line. In addition, the results tend to be uniform.

5. Conclusions

The NN models have trained with three different TAs, namely the BP, VLR and LM. Then, the trained and validated network is used to forecast O-D matrices for new data set, which has not been used either in the training or validation process. The results suggest that the testing performance of NN model trained with LM is better than the other two. In addition, it is the only one which has the same level of testing performance as the doubly-constrained gravity model, calibrated by Hyman's algorithm. There are also other factors which can contribute to the NN performance, namely the data split and normalization methods, and also methods in measuring the RMSE and R^2 for the testing outputs.

Although NN models trained with LM perform better than others, it requires more memory in the computation of Jacobian matrices. In addition, training the model in batch mode will also increase the memory usage. It will rise with the increasing number of input pattern size and also the desired output vectors. However, the rapid developments in computer technology should be able to solve that issue. In addition, there is research devoted to refining the LM to lower its memory usage requirements. Finally, it is impossible to undertake the P_i and A_j satisfaction assessment, if the data is split according to the random vector pattern rather than based on random zone number. The method described in this paper overcomes that problem. Finally, the neural network model can use more information for the training and forecasts the trip distribution once the proposed framework for neural network application in trip distribution has been successfully developed. Thus, it becomes the further work of this research.

References

Barnard, E 1992, 'Optimization for training neural nets', *Neural Networks, IEEE Transactions on,* vol. 3, no. 2, pp. 232-240.

Black, WR 1995, 'Spatial interaction modeling using artificial neural networks', *Journal of Transport Geography,* vol. 3, no. 3, pp. 159-166.

Carvalho, MCM, Dougherty, MS, Fowkes, AS & Wardman, MR 1998, 'Forecasting travel demand: a comparison of logit and artificial neural network methods', *The Journal of the Operational Research Society,* vol. 49, no. 7, pp. 711-722.

Cremer, M & Keller, H 1987, 'A new class of dynamic methods for the identification of origin-destination flows', *Transportation Research Part B: Methodological,* vol. 21, no. 2, pp. 117-132.

Dia, H 2001, 'An object-oriented neural network approach to short-term traffic forecasting', *European Journal of Operational Research,* vol. 131, no. 2, pp. 253-261.

Doblas, J & Benitez, FG 2005, 'An approach to estimating and updating origin-destination matrices based upon traffic counts preserving the prior structure of a survey matrix', *Transportation Research Part B: Methodological,* vol. 39, no. 7, pp. 565-591.

Dougherty, M 1995, 'A review of neural networks applied to transport', *Transportation Research Part C: Emerging Technologies,* vol. 3, no. 4, pp. 247-260.

Hagan, MT & Menhaj, MB 1994, 'Training feedforward networks with the Marquardt algorithm', *IEEE Transactions on Neural Networks* vol. 5, no. 6, pp. 989-993.

Jacobs, RA 1988, 'Increased rates of convergence through learning rate adaptation', *Neural Networks,* vol. 1, no. 4, pp. 295-307.

Mozolin, M, Thill, JC & Lynn, UE 2000, 'Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation', *Transportation Research Part B: Methodological,* vol. 34, no. 1, pp. 53-73.

Nie, Y, Zhang, HM & Recker, WW 2005, 'Inferring origin-destination trip matrices with a decoupled GLS path flow estimator', *Transportation Research Part B: Methodological,* vol. 39, no. 6, pp. 497-518.

Nihan, NL & Davis, GA 1987, 'Recursive estimation of origin-destination matrices from input/output counts', *Transportation Research Part B: Methodological,* vol. 21, no. 2, pp. 149-163.

Popescu, M-C, Balas, V, Olaru, O & Mastorakis, N 2009, 'The Backpropagation Algorithm Functions for the Multilayer Perceptron', paper presented at the 11th WSEAS International Conference on Sustainability in Science Engineering, Timisoara, Romania.

Rumelhart, DE, Hinton, GE & Williams, RJ 1986, 'Learning representations by back-propagating error', *Nature,* vol. 323, pp. 533-536.

Sherali, HD, Narayanan, A & Sivanandan, R 2003, 'Estimation of origin-destination trip-tables based on a partial set of traffic link volumes', *Transportation Research Part B: Methodological,* vol. 37, no. 9, pp. 815-836.

Sherali, HD, Sivanandan, R & Hobeika, AG 1994, 'A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes', *Transportation Research Part B: Methodological,* vol. 28, no. 3, pp. 213-233.

Taylor, MA, Bonsall, PW & Young, W 2000, *Understanding Traffic Systems: Data, Analysis and Presentation*, Second edn, Ashgate Publishing Ltd, Hants, England.

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

Vogl, T, Mangis, J, Rigler, A, Zink, W & Alkon, D 1988, 'Accelerating the convergence of the back-propagation method', *Biological Cybernetics,* vol. 59, no. 4, pp. 257-263.

Wilamowski, BM, Iplikci, S, Kaynak, O & Efe, MÖ 2001, 'An Algorithm for Fast Convergence in Training Neural Networks', *IEEE,* vol. 3, pp. 1778-1782.

Yaldi, G, Taylor, MAP & Yue, WL 2009a, 'Improving Artificial Neural Network Performance in Calibrating Doubly-Constrained Work Trip Distribution by Using a Simple Data Normalization and Linear Activation Function', paper presented at the Paper of The 32 Australasian

Transportation Research Forum, Auckland, New Zealand. Available at www.patrec.org/atrf.aspx.

Yaldi, G, Taylor, MAP & Yue, WL 2010, 'Refining the Performance Neural Network Approach in Modelling Work Trip Distribution by Using Lavenberg-Marquardt Algorithm', *Journal of the Society for Transportation and Traffic Studies (JSTS)*.

Yaldi, G, Taylor, MAP & Yue, WL 2009b, 'Using Artificial Neural Network in Passenger Trip Distribution Modelling (A Case Study in Padang, Indonesia)', *Journal of Eastern Asia Society for Transportation Studies,* vol. 8, pp. 682-693.

Zhang, G, Patuwo, BE & Hu, MY 1998, 'Forecasting with artificial neural networks:: The state of the art', *International Journal of Forecasting,* vol. 14, no. 1, pp. 35-62.