A Consistent Neural Network Model for Doubly Constrained Spatial Movement Estimation

Gusri YALDI
PhD Student
ISST-Transport Systems
University of South Australia
Adelaide, South Australia, SA 5001
Fax: +61 8 8302 1880
E-mail: Gusri.Yaldi@postgrads.unisa.edu.au

Michael A P TAYLOR
Professor
Director of ISST
University of South Australia
Adelaide, South Australia, SA 5001
Fax: +61 8 8302 1880
E-mail: Michael.Taylor@unisa.edu.au

Wen Long YUE
Senior Lecturer, Program Director
ISST-Transport Systems
University of South Australia
Adelaide, South Australia, SA 5001
Fax: +61 8 8302 1880
E-mail: Wen.Yue@unisa.edu.au

Abstract: Despite its successful applications in mode choice modelling, the Neural Network (NN) Approach is often seen as a poor technique for trip distribution estimation. However, empirical results from this study show that the NN can also be used as a good trip distribution modelling tool, by using appropriate Training Algorithm (TA), which is the most critical property of NN. This study will describe the results of NN approach in work trip distribution estimation with three different TAs, namely Back Propagation (BP), Variable Learning Rate (VLR), and Levenberg-Marquardt (LM). The experiments were conducted 30 times for each TA. The results suggest that both BP and VLR models tend to have overestimated total trip numbers, while LM model generates the closest total trip number to the observed one, with less than five per cent difference. Statistical analysis also suggests that LM model is able to generate a consistent performance, while other models have a great fluctuation. Then, the findings from this study are expected can be used as an important consideration for using NN approach as an alternative simple and robust travel demand modelling tool.

Key Word: Neural Network, Training Algorithm, Work trip Distribution

1. INTRODUCTION

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 model. 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 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 ($w_{ij}$ and $w_{jk}$), which are 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 transform it by the activation function in output layer node, in this case also the sigmoid function. More details can be found in the literature, for example Black (1995) and Yaldi et al. (2009b).

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

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 pattern 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). 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 and Mozolin et al. respectively. Training with this high number would cause the models to be over fitted. Special discussion on the effect of excessive iteration number on the NN model performance is presented later in this paper.

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. More details are available in Yaldi et al. (2009a, 2010).

This paper will extend the analysis and discussion of latest previous study by Yaldi et al. (2010), focusing more in the consistency of the performance of NN models trained by using BP, VLR, and LM algorithms. The experiment number is increased to 30 experiments for each model. The criterions of performance evaluation are the total trips estimated by the model, which has not been done before. It is important to analyse the consistency of the NN model performance to make sure that the NN model has statistically the same output for each experiment, although trained by using different values of random connection weights. The dependency of the model toward specific values of connection weights can be avoided. The estimation trend, whether over estimate or underestimate, also can be investigated. Therefore, a robust NN model, where its performance does not rely upon specific connection weight values, can be promoted. The previous studies reported the results usually based on single experiment only; it is unsure whether the NN models will generate the same results statistically when it is trained again with different values of connection weights.

2. MODEL DEVELOPMENT

2.1 Model structure and training process

Figure 1 illustrates the model structure. The model is constructed by three layers, namely input, output and hidden layers. Three nodes are located in input layer representing the independent variables of Total trip Production (\(P_i\)), Total trip Attraction (\(A_j\)) in origin and destination zones, and the trip length between zones (\(D_{ij}\)). The output layer has only one node, representing the trip flow from origin \(i\) to destination \(j\) (\(T_{ij}\)) estimated by the model. Meanwhile, a hidden layer is located between input and output layers. Its node number is arbitrary. This study used a constant number of nodes in the hidden layer, as the former study by Yaldi et al. (2009b) indicated that high node numbers in hidden layer is a insignificant factor in NN model performance. It even could increase the error level as found in a study by Mozolin et al. (2000) and Carvalho et al. (1998). However, NN model with a hidden layer performs better than without one.
As described in former section, the ultimate goal of the NN model is to obtain the optimum values of connection weights by minimizing the error term. It is attained by using a specific learning algorithm. The error or difference between estimated and actual trips is defined as:

\[
diff = t_i - \sigma^i_k(u_k)
\]  

(1)

Then, the difference is squared and accumulated for all of input patterns. This cumulative squared difference is then compared with the defined error threshold/training goal. The training will continue with the next iteration when the difference is higher than the error threshold and is used in adjusting the connection weights, or it will stop when it is below or equal to threshold value. This type of training process is called batch mode.

The training algorithm plays a crucial role in the performance quality of the NN models. The history of NN development is strongly related to the training process. After experiencing a great development up to 1969, the NN was considered to be in a downturn period triggered by the monumental work by Minsky and Papert (1969). This period lasted for about two decades. Although the number of research and the amount of fund allocated for NN study decreased, the exploration of NN were continuing. Then, the resurgence of NN research was marked by the introduction of back propagation algorithm by Werbos (1974) which was popularized by Rumelhart and McClelland (1986) as described by Skias (2006), Jain et al. (1996), and Lippmann (1987).

Back propagation (also written as backpropagation), a generalization of Least Mean Squares (LMS) algorithm, was considered the most famous learning rule. It was also used in the spatial interaction modelling by Black (1995), one of the earliest study in transport area that used NN technology. Although back propagation considered as a landmark in the revival of NN, it is considered too slow in the training process, which led to the development of various learning algorithms in order to improve its convergence speeds (Jacobs, 1988, Barnard, 1992), and one of the is the variable learning rate algorithm as described by Vogl et al. (1988). Improving the training speed by allowing the learning rate to be modified during the training is one of the earliest techniques. It is also considered as the best method in increasing the standard back propagation learning rate (Popescu et al., 2009).

The improvement method described by Vogl et al. (1988) is categorized as ad hoc technique, where the standard back propagation algorithm is modified by adding a constant (momentum), allowing the learning rate to vary, or by rescaling the variables so that the learning converges more quickly (Hagan and Menhaj, 1994, Barnard, 1992). Barnard (1992) also suggested that the empirical improvement procedures proposed by Fahlman (1988), who proposed the Quickprop algorithm used by Mozolin et al. (Mozolin et al., 2000) in their study entitled “Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation”, as a heuristic approach which might negatively affect the NN performance. According to Wilamowski et al. (2001), there is improvements contributed by the heuristic approach, however, it is considered minor. Then, it is suggested that the second order approaches such as the Newton’s method, conjugate gradient or the Levenberg-Marquardt (LM) optimization technique can improve the NN performance significantly. The LM algorithm proposed by Hagan and Menhaj (1994), is the training algorithm from the second order approach which is reported as the most efficient and accepted widely (Wilamowski et al., 2001).
Thus, three different training algorithms are used to train the model in this study, namely standard Back Propagation (BP), Variable Learning Rate (VLR) and Levenberg-Marquardt (LM). To simplify the discussion, the models are termed as BP, VLR and LM models respectively. The BP represents the earliest and the most famous TA and was used in the early period of NN application in the spatial interaction movement. The second algorithm was considered as the best method in improving the convergence speed of NN model, and in the ad hoc improvement category together with the Quickprop algorithm used by Mozolin et al. (2000), where this algorithm might negatively affect the NN performance. The LM, the third algorithm used in this study, was reported widely accepted, can improve the NN performance significantly, and is the most efficient TA. More detail mathematical formulation of each algorithm can be found in Rumelhart and McClelland (1986), Vogl et al. (1988), and Hagan & Menhaj (1994).

2.2 Data
The NN models in this study are developed to forecast work trip distribution (i.e. doubly constrained origin-destination matrix). The data is obtained from the Household Survey 2005, in Padang City, West Sumatra, Indonesia. There are 36 zones. Therefore, there are 36 x 36 input sample patterns, consisting of 1296 inputs. The total number of trips is 218700.

2.3 Transfer function
The selection of activation function is dependent upon the following factors:

- The activation functions in the hidden layer must be able to capture the nonlinearity between input and output
- Different activation functions can be used in the hidden and output layers
- Given that all the data are positive, the activation in the hidden layer must not allow the summation outputs to be negative values
- The activation function in the output layer must ensure it does not generate negative outputs (estimated trips)

There are four common activation functions according to Teodorovic and Vukadinovic (1998), two of which are the Sigmoid and Linear functions. Sigmoid functions have often been used in different transportation studies, such as the studies by Mozolin et al. (2000), Carvalho et al. (1998), and Black (1995). The mathematical formulations of these functions are given below.

- Tansig
  \[ x = \frac{2}{1+\exp(-2x)} - 1 \] (2)
- Logsig
  \[ x = \frac{1}{1+\exp(-x)} \] (3)
- Linear
  \[ x = x \] (4)

![Figure 2 Common activation functions](image-url)
Thus, Logsig activation function is selected for both of hidden and output layer nodes, as it is the only one that can satisfy all of the factors mentioned above. This is the same function as used by Black (1995), however, it is different from what was used by Mozolin et al (2000), which is the double logistic function. More details regarding the performance of the neural network models for different activation functions is reported in Yaldi et al. (2009a).

2.4 Maximum epoch number
The Epoch number can be defined as the number of iterations in the training process. To define the maximum epoch for the NN models in this study, the original data is divided into three data sets, where 40 per cent is used for training or developing the model. Then, 30 per cent is used to validate the model, and the last 30 per cent is used for testing. The training is stopped when the validation error increases. Thus, the number of epoch when the training is stopped is defined as the maximum epoch. There are three maximum epoch numbers, for the NN models trained by BP, VLR and LM training algorithms. Those are 252, 51 and 19 epochs for BP, VLR and LM respectively. The epoch numbers applied in this study are significantly lower than the number of iteration from former studies by Mozolin et al. (2000) and Black (1995), where the epoch number was up to 100000 and 150000 respectively. Therefore, over-fitting can be avoided by limiting the epoch number. In addition, a significantly lower training time can be expected in this study. Special discussion on over-fitting and its impact is discussed later in this paper.

2.5. Modelling tools and performance measurements
The model was developed using the Neural Network Tool in MATLAB software version 7.0.1. The initial weights for all layers were randomly selected by the software. Each experiment has different value of connection weights each other. The weights were updated after all of the data are used in the training (batch mode). Model performance is analysed based on the average estimated total trip, standard deviation (SD), and correlation coefficient (CV). Paired t-test is also conducted to examine the impact of epoch number toward the model accuracy.

3. DISCUSSIONS AND MODEL OUTPUTS

3.1 Total estimated trip
Firstly, all models are trained with 100 epochs. Then, each model is trained again with different epoch numbers as discussed in section 3.2. The discussion in this section is based on 100 epoch results.

The total trips estimated by each NN model is illustrated in Figures 3, 4, and 5. It can be seen that the results vary for each experiment. From the 30 experiments, majority of the BP model results are above the total observed trips (obs), while the VLR and LM models have results below the total observed trips (see Figures 3, 4 and 5).

In average, the BP models estimate the total trips as double as the total actual trips, while the VLR models average total trip is about a half of the total actual trips. However, the average total trips estimated by LM model is about 2 per cent lower than the actual value. More details on the results are shown in Table 1.
Table 1 Total estimated trips and RMSE for all models (trained 100 epochs)

<table>
<thead>
<tr>
<th>Trial</th>
<th>BP</th>
<th>VLR</th>
<th>LM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ΣTij</td>
<td>RMSE</td>
<td>ΣTij</td>
</tr>
<tr>
<td>1</td>
<td>424325</td>
<td>384</td>
<td>160598</td>
</tr>
<tr>
<td>2</td>
<td>67335</td>
<td>318</td>
<td>82617</td>
</tr>
<tr>
<td>3</td>
<td>368076</td>
<td>414</td>
<td>140572</td>
</tr>
<tr>
<td>4</td>
<td>310362</td>
<td>395</td>
<td>98653</td>
</tr>
<tr>
<td>5</td>
<td>284605</td>
<td>384</td>
<td>106551</td>
</tr>
<tr>
<td>6</td>
<td>261589</td>
<td>299</td>
<td>179988</td>
</tr>
<tr>
<td>7</td>
<td>509771</td>
<td>408</td>
<td>170749</td>
</tr>
<tr>
<td>8</td>
<td>439634</td>
<td>396</td>
<td>123281</td>
</tr>
<tr>
<td>9</td>
<td>435163</td>
<td>370</td>
<td>121516</td>
</tr>
<tr>
<td>10</td>
<td>472290</td>
<td>493</td>
<td>98202</td>
</tr>
<tr>
<td>11</td>
<td>314415</td>
<td>324</td>
<td>138742</td>
</tr>
<tr>
<td>12</td>
<td>3423442</td>
<td>2549</td>
<td>1596</td>
</tr>
<tr>
<td>13</td>
<td>563419</td>
<td>610</td>
<td>73454</td>
</tr>
<tr>
<td>14</td>
<td>529769</td>
<td>397</td>
<td>131747</td>
</tr>
<tr>
<td>15</td>
<td>453959</td>
<td>405</td>
<td>156920</td>
</tr>
<tr>
<td>16</td>
<td>388565</td>
<td>419</td>
<td>85931</td>
</tr>
<tr>
<td>17</td>
<td>170328</td>
<td>276</td>
<td>223739</td>
</tr>
<tr>
<td>18</td>
<td>498631</td>
<td>336</td>
<td>227637</td>
</tr>
<tr>
<td>19</td>
<td>74085</td>
<td>305</td>
<td>75875</td>
</tr>
<tr>
<td>20</td>
<td>327068</td>
<td>286</td>
<td>188929</td>
</tr>
<tr>
<td>21</td>
<td>207674</td>
<td>290</td>
<td>111532</td>
</tr>
<tr>
<td>22</td>
<td>824533</td>
<td>673</td>
<td>58123</td>
</tr>
<tr>
<td>23</td>
<td>514980</td>
<td>595</td>
<td>62522</td>
</tr>
<tr>
<td>24</td>
<td>92949</td>
<td>339</td>
<td>67191</td>
</tr>
<tr>
<td>25</td>
<td>156317</td>
<td>236</td>
<td>202086</td>
</tr>
<tr>
<td>26</td>
<td>337014</td>
<td>530</td>
<td>32932</td>
</tr>
<tr>
<td>27</td>
<td>470709</td>
<td>436</td>
<td>119339</td>
</tr>
<tr>
<td>28</td>
<td>261649</td>
<td>412</td>
<td>46836</td>
</tr>
<tr>
<td>29</td>
<td>415213</td>
<td>558</td>
<td>41023</td>
</tr>
<tr>
<td>30</td>
<td>484210</td>
<td>504</td>
<td>47369</td>
</tr>
<tr>
<td>Average</td>
<td>462736</td>
<td>478</td>
<td>112542</td>
</tr>
<tr>
<td>SD</td>
<td>582253</td>
<td>405.37</td>
<td>58286</td>
</tr>
<tr>
<td>CV</td>
<td>1.258</td>
<td>0.848</td>
<td>0.518</td>
</tr>
</tbody>
</table>

The results in Table 1 suggest that the BP models have the highest standard deviation between observed and estimated trips, followed by VLR and LM models. Consequently, the coefficient of variation (CV) for the BP model becomes the largest. The VLR model is in the second place, and LM model is the third. Thus, the LM models predict the total trips closer to the actual value than other models. With a CV of 0.006, the LM models generate the total trips which is very much more consistent than other models. There is higher certainty that the LM model will produce total trips more or less the same as actual trips, although the value of estimation parameter/the connection weights between each layer are different and randomly selected for each experiment. Figure 6 shows that the LM models have almost insignificant variation compared to others.
The LM models also have a consistent performance in terms of RMSE. The lowest average RMSE belongs to LM models, which is 157, followed by VLR and BP with 283 and 478 trips respectively (see Table 1). LM models predict the total trips more accurately than BP and VLR models for 30 experiments, with a smaller variation in the results than other models. Their CV for RMSE is lower than the other models. It can be concluded that the LM models estimate the total trip more consistently and more accurately than other models.

A comparison with the doubly constrained gravity model also suggests that LM models have a statistically lower error (Yaldi et al., 2009a). Meanwhile, BP models tend to have overestimate results, while VLR models tend to underestimate. However, the results are based on the models trained for 100 epochs. Although it is significantly lower than the epoch
numbers used by Mozolin et al. (2000) and Black (1995), over fitting may still occur. Therefore, the maximum number of epoch must be investigated, as discussed in the next section.

![Figure 6 Total trips estimated by all models](image)

**3.2 Over fitting**

The over fitting can be defined as the NN model where it has a lower error when more iterations are applied in the training stage. However, its testing performance becomes worse. In other word, it produces better results when the same data for training is used to test the model, yet, it produces a higher error when other data is used to test the same model, as illustrated by Figure 7.

The LM model is used to illustrate the over fitting as it has the best performance and also the lowest iteration number. It is also suggested as the best model for passenger trip distribution compared to BP and VLR models (Yaldi et al., 2010).

Figure 7 depicts the performance lines of LM model for work trip estimation. The blue, red and green lines represent the performance of training, validation and testing in term of normalized mean squared error (MSE). The maximum epochs for training as displayed in Figures 7a, 7b and 7c are 19, 1000 and 10000 times respectively. The training is stopped at the 19th epoch as the validation error starts to increase (Figure 7a). The training error is 0.00279, while the testing error is 0.0064484.

Then, the same model is trained again with 1000 (Figure 7b) and 10000 (Figure 7c) epochs. It generates MSE for training and testing 0.002305 and 0.009726 for 1000 iterations, while it is 0.002302 and 0.009743 for 10000 iterations. It can be concluded that higher numbers of iterations will generate a lower error for the training, but a higher error for the testing. Therefore, the training must be stopped to avoid this over fitting. Therefore, the same BP, VLR and LM models as discussed in previous section are trained again with different epoch numbers. The LM model is trained for a maximum epoch of 19. Meanwhile, the BP and VLR model are trained for 252 and 51 epochs respectively. Those epoch numbers are obtained through the same procedures as in obtaining the LM maximum epoch number. The results are discussed in the next section.
3.3 Results for modified BP, VLR and LM models

To avoid over fitting, all models are trained again. The BP, VLR and LM models are trained for 252, 51 and 19 epochs respectively based on the procedures explained in the former section. In general, the results tend to fluctuate the same as before modification. BP model remains to have an overestimate trend. VLR models has now also an overestimate trend. BP models have more experiments where estimated total trips is lower than the actual trips, and VLR model has more experiments where estimated total trips is higher than the observed value. The same trend is also happened for LM model, yet, it is still close to the real total trip number (see Figures 8, 9 and 10). Table 2 presents more details on the results after modification.
The impact of modification toward the maximum training epoch is examined by using a paired t-test. It is aimed at measuring the significance of change to model performance, before and after modification. The results in Table 2 show that the average total trips estimated by BP and LM models experience a statistically significant changes. The average estimated total trips for BP and VLR model is much closer to the actual trips compared to before modification, however, the LM model is still the closest one.
The BP models generate an average total trips which is much improved. Although it is still an overestimate, its standard deviation is much lower than before modification. The reason is it
is allowed to be trained up to 252 epochs. Meanwhile, the LM models also experience a statistically significant change in the average estimated total trip. It has a higher standard deviation and higher CV than before modification. This is occurred because the models are trained with a lower number of iterations. Therefore, the epoch number can significantly increase or decrease the performance of the models, and becomes an important properties of the NN models. Although the LM model has a higher standard deviation than before modification, it is still the lowest of all the models, as is its CV value. This means that the LM model is still able to consistently estimate total trip number with higher accuracy than other models. Figure 11 illustrates that the LM model has the most consistent results for 30 experiments. Finally, a doubly constrained gravity model was also developed. It was calibrated by using Maximum likelihood technique proposed by Hyman (1969). The RMSE of the gravity model is 168, or about eight per cent higher than the LM model.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1. Conclusions
The analysis on the model performance in this study suggests that only the LM model has a consistent results for 30 experiments, although its epoch number is reduced to 19 epochs. It is indicated by the coefficient of variation (CV), as well as the standard deviation for both total estimated trip and the RMSE for each experiment. The BP and VLR models have a great fluctuations and tend to overestimate. Their standard deviations between estimated and observed total trips is significantly higher than the LM model. This result supports the findings of Yaldi et al. (2010) and (2009a) studies. The NN models trained by using the LM algorithm estimate total trip number with a statistically lower RMSE than those trained by using other TAs, and are also lower compared to the doubly constrained gravity model. The LM model has a higher goodness of fit. It also has a consistent performance as suggested by the analysis in this paper. Therefore, it can be finally concluded that the NN approach can be a robust tool for pattern recognition, when the models are trained by using the LM algorithm as found in this study.

4.2. Recommendations
Further study is recommended to develop the NN models with LM training algorithm for commodity flows. It is expected that the application of NN approach as a robust modelling tool for fully constraint spatial movement can be extended to freight movements. A special analysis on the training speed is also recommended, in order to investigate the convergence rate of different training algorithms.

REFERENCES


