Research Show Window: EASTS-Japan Working Paper Series

No.08-4

A Non-Linear Analysis of Discrete Choice Behavior by the Logit Model

Shoichiro Nakayama
Dept. of Civil Engineering
Kanazawa University

Jun-ichi Takayama
Dept. of Civil Engineering
Kanazawa University

Yuichiro Yamashita
Dept. of Civil Engineering
Kanazawa University

Phone: +81-(0)76-234-4614
Fax: +81-(0)76-234-4632
email:
snakayama@t.kanazawa-u.ac.jp

Abstract

The logit model based on random utility theory has often been used for discrete choice behavior analysis. In the conventional logit model, it is assumed that variables are independent of each other and that their relationship is linear. In general, the relationship or behavior of the variables is non-linear, and the assumptions of the logit model are not always proper. Therefore, incorporating neural networks, which are suitable to make an analysis non-linearly, to the logit model is very useful. In this study, we propose the logit model using non-linear utility functions with neural network. Then, we analyze a discrete choice behavior in traffic phenomenon by the model, and examine the validity of the model.

Keywords

Non-Linear Analysis, Logit Model, Neural Network, Giveway Behavior

Note: This paper was originally presented as Nakayama, S., J. Takayama and Y. Yamashita: A Non-Linear Analysis of Discrete Choice Behavior by the Logit Model, presented at the 10th World Conference on Transport Research, Istanbul, 2004.7.

December 2008

A Non-Linear Analysis of Discrete Choice Behavior by the Logit Model

Shoichiro Nakayama*, Jun-ichi Takayama, and Yuichiro Yamashita
Department of civil Engineering, Kanazawa University
*: Corresponding author

Address: 2-40-20 Kodatsuno, Kanazawa 920-8667, Japan
Tel: +81-(0)76-234-4614
Fax: +81-(0)76-234-4632
E-mail: snakayama@t.kanazawa-u.ac.jp

Abstract

The logit model based on random utility theory has often been used for discrete choice behavior analysis. In the conventional logit model, it is assumed that variables are independent of each other and that their relationship is linear. In general, the relationship or behavior of the variables is non-linear, and the assumptions of the logit model are not always proper. Therefore, incorporating neural networks, which are suitable to make an analysis non-linearly, to the logit model is very useful. In this study, we propose the logit model using non-linear utility functions with neural network. Then, we analyze a discrete choice behavior in traffic phenomenon by the model, and examine the validity of the model.

Keywords: Non-Linear Analysis, Logit Model, Neural Network, Giveway Behavior

1. INTRODUCTION

Most of driver's behavior seems to be discrete choices. The logit model [1, 2] based on random utility theory has often been used for analysis of discrete choice behavior for decades. The logit model determines the probability of choosing a choice based on the utility. In the past, the utility has been formulated as a linear function in most studies. This represents that the (exploratory) variables work independently and that the relationship of them is not non-linear. However, in general, the variables are not necessarily independent of each other. The logit model with linear utility functions could miss some non-linear effects or phenomena of discrete choice behavior.

In this study, neural networks, which are suitable for non-linear analysis, are incorporated into the utility function, and we propose the logit model with such non-linear utility functions. Then, we can analyze a discrete choice behavior in traffic phenomena, and examine the validity of the model. As a non-linear function, polynomial expression as well as neural network is also available. However, stability of estimating parameters becomes low when including a high order term. Neural network can express various shapes as a non-linear function, can estimate stable parameters, and has already been applied prevailingly in many fields.

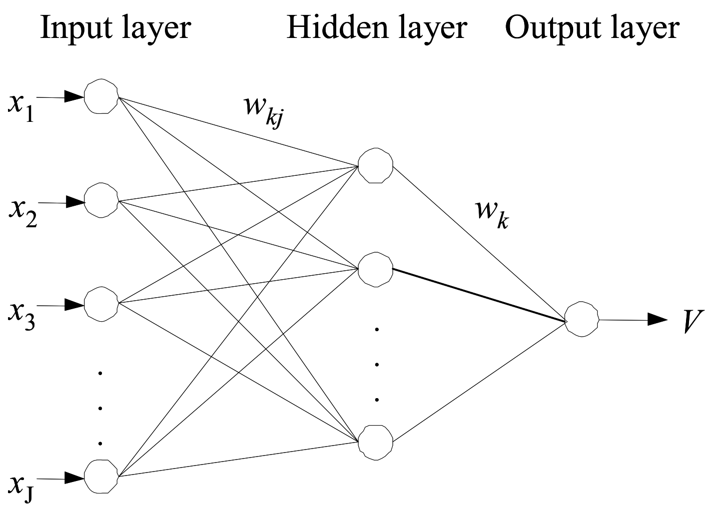

Figure 1. A 3-layer neural network

2. Neural Networks and Utility Functions

2.1. Neural Networks

A neural network model is a non-linear learning model that has been formulated based on neural networks in brains. It can deal with data that has complex structure and can analyze non-linear relationships among variables easily. Neural networks are applied as methods of data mining and elements of artificial intelligence in many studies in transportation engineering [3, 4]. Nevertheless, there are only a few applications to the analysis of discrete choice behavior in transportation engineering including travel behavior analysis [5, 6, 7]. As discussed in the previous section, a neural network model is suitable when the variables work non-linearly.

A 3-layer feed-forward neural network as shown in Figure 1 is the most frequently used formulation in recent applications of neural networks. These layers are called an input layer, a hidden layer, and an output layer, respectively. At each node in the input layer, input values entered from outside are emitted to nodes in the next layer without change. At each node in the hidden layer and the output layer, the output value from a node in the preceding layer is transformed by an activation function, and it's a transformed value is emitted to nodes in the next layer. A sigmoid function is frequently used as an activation function. In this case, an output from the hidden layer is defined as follows:

$$y_k = \frac{1}{1 + \exp\left\{-\left(\sum_j w_{kj} \cdot x_j + \theta_k\right)\right\}} \tag{1}$$

where

$y_k =$ the output from the kth node in the hidden layer (and is also the input of nodes in the output layer),

$x_j =$ the output from the jth node in the input layer (and is equal to the input of jth node in the input layer),

$w_{kj} =$ the weight between the jth node in the input layer and the kth node in the hidden layer, and

$\theta_k =$ the threshold in the hidden layer (in Figure 1, this is omitted for small space).

When a node in the output layer is single, an output from the output layer is:

$$V = \frac{1}{1 + \left\{ -\left(\sum_k w_k \cdot y_k + \theta \right) \right\}} \tag{2}$$

where

$V =$ the output from the node in the output layer,

$y_k =$ the output from the kth node in the hidden layer,

$w_k =$ the weight between the kth node in the hidden and the node in the output layer,

$\theta =$ the threshold in the output layer.

A back propagation algorithm [8] is frequently used when updating connection weights between nodes. In the back propagation algorithm, the weights are updated so as to minimize the square errors between the network output and the desired output (or teacher signal).

2.2. Utility Functions with Neural Networks

The logit model is one of discrete choice models and is used most frequently. The model calculates the probability of choosing a choice based on the utility. In the conventional logit model, a utility function is written as:

$$V_i = \sum_{j=1}^{J} w_j x_{ij} + \theta \tag{3}$$

where

$V_i =$ (the systematic component of) the utility of the ith choice,

$x_{ij} =$ the jth variable of the ith choice, and

w_j ,$\theta =$ parameters

In random utility theory, the utility of each choice is random, and the choice which has the maximum utility is taken. The utility, U, is almost always defined as $V + \zeta$, where ζ is the random term, whose mean is 0. In the logit model, ζ is Gumbel distributed, and the probability of the ith choice, P_i, is:

$$P_i = \frac{\exp(V_i)}{\sum_{i'} \exp(V_{i'})} \tag{4}$$

In this study, two kinds of non-linear utility functions with the neural network are adopted.

1) NN Utility Function 1: The conventional logit model has the linear utility function written in Equation (3). We formulate the non-linear utility function as a 3-layer feed-forward neural network with one output node. In the neural network model which has sigmoid functions as activation functions, the output is in the range from 0 to 1. However, the value of the utility is not necessarily within that range. As an activation function in the output layer, a linear function can be adopted. The activation function in the hidden layer is a sigmoid function as usual. In this case, the utility function is expressed as:

$$V_i = \sum_k w_k \cdot \left[\frac{1}{1 + \exp\left\{ -\left(\sum_j w_{kj} \cdot x_{ij} + \theta_k \right) \right\}} \right] + \theta \tag{5}$$

We shall call the logit model with this utility function NN logit I and call the utility function NN utility function I.

2) NN Utility Function 2: In NN utility function 2, as an activation in the output layer, the revised sigmoid function is used. A new parameter, γ, is incorporated into the numerator of the sigmoid function. The activation function in the hidden layer is a (normal) sigmoid function. NN utility function 2 is written as:

$$V_i = \frac{\gamma}{1 + \exp\left[-\sum_k w_k \cdot \left\{ \frac{1}{1 + \exp\left(-\sum_j w_{kj} \cdot x_{ij} - \theta_k \right)} \right\} - \theta \right]} \tag{6}$$

We shall call the logit model with this utility function NN logit II and call the utility function NN utility function II.

2.3. Estimation of Parameters

A back propagation method [8] is frequently used when estimating parameters (connection weights) in neural network models. However, the back propagation method cannot be applied to the logit model with neural network utility functions because there does not directly exist the desired data for the output of the neural network, *V*. We use the maximum likelihood estimation. The maximum likelihood estimation is the optimization of the following likelihood function:

$$L = \prod_{n=1}^{N} \prod_{i=1}^{I} (P_i^n)^{\delta_{ni}} \tag{7}$$

$$\delta_{ni} = \begin{cases} 1 & \textit{if the choice is actually taken} \\ 0 & \textit{otherwise.} \end{cases}$$

where

$L =$ the likelihood, and

P_i^n =the probability that the nth individual chooses the ith choice.

A solution of the maximization of the likelihood function satisfies the usual first-oder conditions. When the utility function is linear, the likelihood function of the sample conditioned on the parameters is convex and the ordinary optimization method such as the Newton-Raphson algorithm can be used. However, when the utility function is not linear, the likelihood function is not necessarily convex. Therefore, we estimate the parameters by the gradient method using simulated annealing. The parameters are updated by the following equation:

$$w^{(m+1)} = w^{(m)} + \eta \cdot \frac{\partial L}{\partial w} + \varepsilon^{(m)} \tag{8}$$

where

$w^{(m)}$ = the parameter (connection weight) in the mth step,

η = a small positive parameter, and

$\varepsilon^{(m)}$ = the normal random number in the mth step.

The mean of the normal random number is 0 and the variance decreases as the step goes by. In this study, $\varepsilon^{(m)}$ is $T^0 / \ln m$, where T^0 is the initial variance. The solution is obtained after sufficient iteration.

Even if the parameters fall into a local maximum, they escape from the local solution by the error term, ε. Thus, the above method can estimate the parameters properly even when the utility function is not linear.

3. Analysis

3.1. Data

We apply the model to giveway behavior at the merging point on the road. Figure 2 illustrates giveway behavior. At the merging point, some vehicles on the main lane change the lane to avoid conflicts with the merging vehicles. In this study, the behavior is assumed to be binary choice of whether or not the vehicle makes a giveway as shown in Figure 2. The survey was conducted at the merging point on Route 1 in Narano-cho, Kyoto, Japan. The sample size is 537. The explanatory

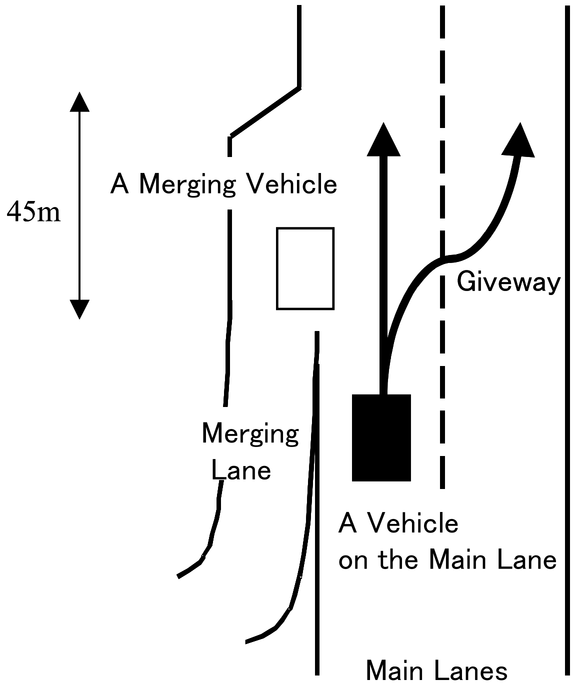

Figure 2. Giveway behavior

variables of giveway behavior are time headway between the vehicle on the main lane (the objective vehicle) and the merging vehicle and the accelerated velocity of the objective vehicle. Headway time is the elapsed time between the time that the merging vehicle passes a fixed line and the instant that the objective vehicle passes that line. The vehicle makes a giveway in order to avoid conflict with the merging vehicle, whose velocity is lower than the objective vehicle, and the objective vehicle locates on the upstream than the merging vehicle. Otherwise, the vehicle do not need to make giveway.

The difference of the utilities of both choices, V (= $V_1 - V_2$), is the function of relative velocity and headway time, where V_1 is the utility of making giveway and V_2 is the utility of not making giveway and going straight on the same lane. In this study, we do not treat V_1 and V_2 explicitly and use the utility difference, V. According to Equation (4), the probability of choosing the first gap, P_1, is calculated by:

$$P_1 = \frac{1}{1 + \exp(-V)} \tag{9}$$

3.2. Results

Table 1 shows the results of NN logit I, NN logit II and the linear logit model (conventional logit model). In light of the log-likelihood ratio, ρ^2, AIC, and hitting ratio, NN logit I models are better than the linear logit. NN logit II models except NN logit II with one node are better than the linear logit model. This implies that NN logit models are better than the linear logit model, but the NN logit models are not always better than the linear logit model. NN logit I model are better than NN logit

Table 1. The results of NN logit models

	Node	L_0^*	L^*	ρ^2	AIC	Hitting ratio
Linear Logit	-	-415.9	-287.1	0.229	580.2	76.35%
NN logit I	1	-415.9	-283.2	0.239	576.4	76.35%
	2	-415.9	-249.4	0.330	**516.8**	78.21%
	3	-415.9	-246.8	**0.337**	519.7	**78.40%**
NN logit II	1	-415.9	-284.7	0.235	581.5	76.35%
	2	-415.9	-249.2	0.330	518.4	76.35%
	3	-415.9	-247.1	0.336	522.2	76.35%

L^*: the log-likelihood, L_0^*: the initial log-likelihood

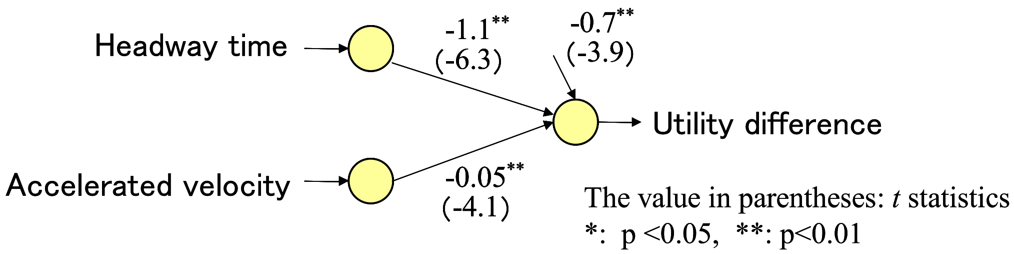

Figure 3. The parameters in the linear logit

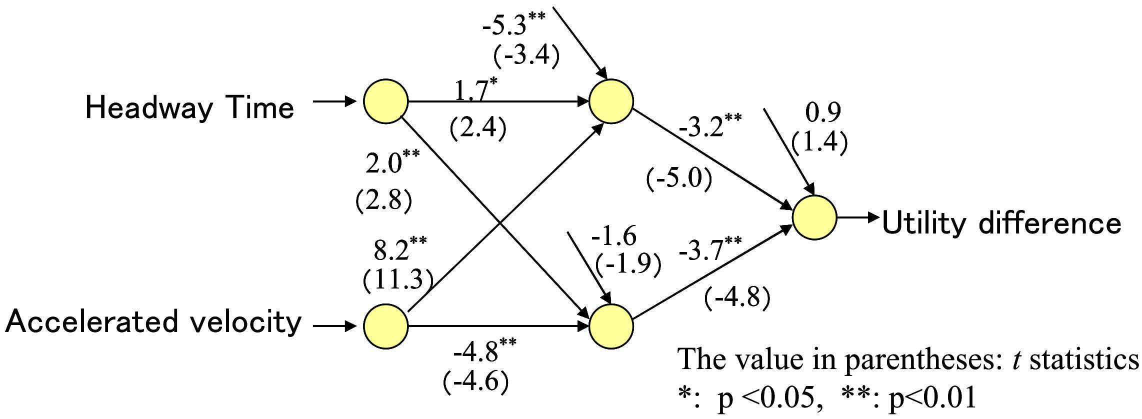

Figure 4. The parameters in NN logit I with two nodes in the hidden layer

II. The difference of NN logit I and NN logit II is the activation function in the neural network. It is important to determine which activation function is used. The log-likelihood ratio index in the NN logit I and NN logit II increases with the number of nodes in the hidden layer, and accuracy of estimation seems to improve with the number of nodes in the hidden layer. This is very natural because the number of the estimated parameters increases. AIC of NN logit I with the two nodes in the hidden layer is the highest, and it could be the best of all models. We can say that the NN logit model can analyze discrete choice behavior more elaborate than the logit model with linear utility functions if the neural network is properly set.

Figure 3 and Figure 4 present the parameters (connection weights and thresholds that are omitted in Figure 1) in the linear logit and the NN logit I model with two nodes in the hidden layer, respectively. Figure 5 and Figure 6 illustrate the relationships of the exploratory variables in the NN logit model with two nodes in the hidden layer and the linear logit model, respectively. In the figures, the values of the variables are normalized between –1 and +1. These figures are quite different each other. The probability of making giveway in the linear logit model is a flat surface in Figure 5 while one in the NN logit model is a curved surface like chevron in Figure 6. Note that Equation (9) shows that the probability of making giveway is larger with the utility difference, V, which is $-\ 1.1\ x_1 - 0.05\ x_2$, where x_1 is headway time and x_2 is accelerated velocity. In Figure 5, the headway time affects giveway behavior at a constant rate despite of accelerated velocity, and the two variables works independently, unlike Figure 6. Figure 6 illustrates that the probability of making giveway is smaller with the headway time on the whole. However, the effect of accelerated velocity on the giveway probability is not so simple as the headway time. As the accelerated velocity is close to its average, the giveway probability is higher. Thus, the effect of the accelerated velocity is not linear.

Figure 7 represents the histogram of giveway behavior. The horizontal axis is the accelerated velocity and the vertical axis is the probability. Figure 7 shows that the probability is chevron-wise. Figure 7 implies that Figure 6 in the NN logit with

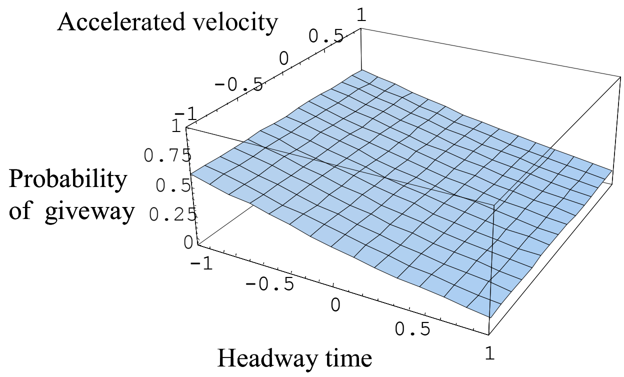

Figure 5. The utility difference in the linear logit

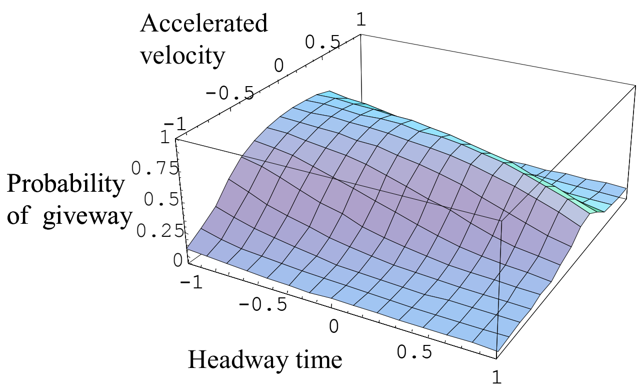

Figure 6. The utility difference in the NN logit

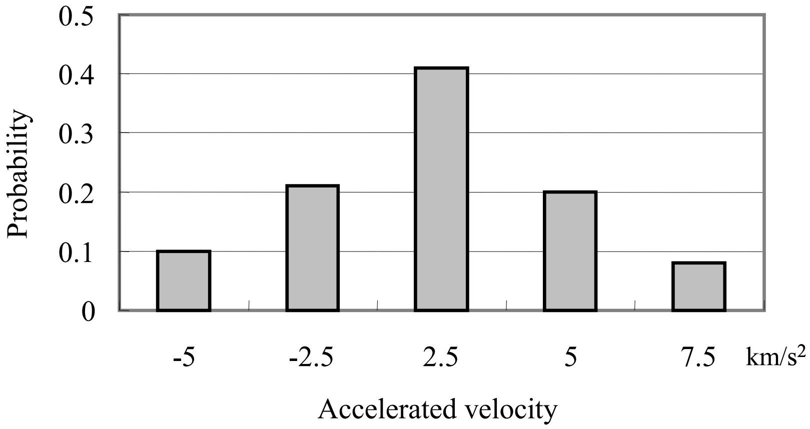

Figure 7. The histogram of giveway probability with accelerated velocity

two nodes in the hidden layer is more accurate than Figure 5 in the linear logit. In conclusion, the logit model using non-linear utility functions can be useful for discrete choice behavior analysis. Also, it is found that the NN logit model can analyze the effect or phenomena that the conventional logit model cannot catch.

4. CONCLUSIONS

The logit model based on random utility theory has often been used for discrete choice behavior analysis. In the conventional logit model, it is assumed that variables are independent and that their relationship is linear. In general, the relationship and behavior of the variables are non-linear, and the assumption of the logit model is not always proper. In this study, we proposed the logit model using non-linear utility function with neural network. Then, we can analyzed giveway behavior at the merging point on the highway, and examined the validity of the model. AICs and log-likelihood ratios of most logit models with neural network utility function (NN logit models) are higher than the logit model with linear utility function. This illustrates that the logit model using non-linear utility functions can be useful for discrete choice behavior analysis. Furthermore, it is found that the NN logit model can analyze the effect or phenomena that the conventional logit model cannot catch. Performance of the models depends on activation functions in the neural network, and the future work is to examine which activation function should be adopted more.

REFERENCES

[1] D. McFadden, 1974. Conditional Logit Analysis of Quantitative Choice Behavior, In P. Zarembka (Ed.) *Frontiers in Econometrics*, Academic Press, New York, pp. 105-142.

[2] Ben-Akiva, M, Lerman, S.R., 1985. *Discrete Choice Analysis*, MIT Press, Cambridge, MA.

[3] Dougherty, M., 1995. A Review of Neural Networks Applied to Transport, *Transportation Research* 3C 247-260.
[4] Fagri, A., Hua, J., 1992. Evaluation of Artificial Neural Network Applications in Transportation Engineering, *Transportation Research Record* 1358 71-80.
[5] Yang, H., Kitamura, R., Jovanis, P.P., Vaughn, K.M., Abdel-Aty, M.A., 1993. Exploration of Route Choice Behavior with Advanced Traveler Information Using Neural Network Concepts, *Transportation* 20 199-223.
[6] Shmueli, D., Salomon, I., Shefer, D., 1996. Neural Network Analysis of Travel Behavior: Evaluating Tools for Prediction, *Transportation Research* 4C 151-166.
[7] Hensher, D.A., Ton, T.T., 2000. A Comparison of the Predictive Potential of Artificial Neural Networks and Nested Logit Models for Commuter Mode Choice, *Transportation Research* 36E 155-172.
[8] Rumelhart, D., Hinton, G., Williams, R., 1986. Learning Internal Representations by Error Propagation, In Rumelhart, D., McClelland, J. (Eds.) *Parallel Distributed Processing* 1, MIT Press, Cambridge, MA, pp. 318-362.