A GA-BASED TRANSPORT MODAL CHOICE MODEL*

by Katia ANDRADE**, Kenetsu UCHIDA***, Seiichi KAGAYA***,
Andre DANTAS**** and Alan NICHOLSON*****

1. Introduction

Application of soft computing techniques to transportation research has become very popular since the 70s, which formed a new category of travel behavior models sometimes called hybrid models. These models include neural networks, fuzzy logic and Genetic Algorithms (GA), and statistical methods. Hybrid models have been suggested for solving transportation problems in an attempt to explore the advantages of both methods, and here promising results have been found [1) 2)]. These techniques have been applied to various modeling problems such as: bicycle route choice behavior [3)]; modal choice model's parameters and linguistic values estimation[4)]; estimating household members' transport modal choices[5)]. However, these studies require complex decision structures and large computational processing for simulation. This limits the extent to which hybrid models can be applied in transport studies where there is a lack of time and resources.

This paper presents a hybrid transport modal choice model in which the genetic algorithm is applied for estimating parameters of a Multinomial Logit Model. The model has a simple decision structure which requires relatively modest computational capabilities and time to be estimated. Linear Utility Functions are defined for the transport modes which are included in the model structure as the fitness function. The objective function seeks to reduce the errors in the modal choice process. Moreover, the hybrid model incorporates distance as a spatial attribute variable for explaining commuters' modal choices among automobile, bus, bicycle and walk modes. After this introduction section, a brief review on GA's is presented. Section 3 introduces the Hybrid Transport Modal Choice Model structure, which is followed by the modeling results. Finally, Section 5 discusses this work's findings and makes recommendations to further research.

2. Genetic Algorithms

Genetic Algorithms were initially suggested during the 70s[6)] as search technique used for approximating solutions in optimization problems. The GA's are based on the evolution theory which considers that, in competitive scenarios, individuals with favorable traits are more likely to survive and reproduce as a result of the Natural Selection process[7)]. These algorithms are defined by three basic rules, in which the first is the Crossover, used for combining information from two or more individuals; the second is the Mutation, which generates new solutions from the current population, therefore allowing the exploration of new regions of the Search Space; and the third is the *Selection* of the fittest individuals, a process that ensures the continuing improvement of the candidate solutions [8)]. Presently, GA's are widely used in solving problems on decision making, classification, and complex numerical optimization[4)], among other applications. In this context, it seems to be an adequate tool for use in transportation problems.

Amongst genetic algorithms, Real-Coded Genetic Algorithm (RCGA) is specifically applied to parameter estimation. The RCGA adopts a real coding of the optimization parameters. It includes a process called *Real Biased Crossover*, which produces two new individuals from their ancestors, where one of them is more similar to its "best" ancestor, i.e. the one presenting better value for the objective function[8)]. This process ensures that there will be better individuals in the new generation, which does not happen for the so-called Simple Genetic Algorithm (SGA) in its basic form[7)]. Table 1 summarizes the characteristics of the RCGA and its procedures.

3. Hybrid Transport Modal Choice Model Structure

The model employs the GA technique to conduct the modeling of the transport modal choice. In this model, GA's main objective is to iteratively improve an initially random set of individuals[9)], which is formed by the choice among the automobile, bus, bicycle and walk modes. The procedures adopted for this search are the crossover among the selected good modal choices, and their further random mutation. In a constant-size population, only those individuals with high fitness are selected, whereas incorrectly predicted choices are dropped from the initial population. GA is able to find nearoptimal solutions for the investigated problem after a sufficient number of iterative applications of these two instruments [10)].

---

* Keywords: transport modeling, hybrid travel behavior model, genetic algorithm

** Student Member of JSCE, M. Sc.., Graduate School of Engineering, Hokkaido University
(Kita-13, Nishi-8, Kita-ku, Sapporo 060-8628, Japan, TEL. +81-11-706-6822, kaandrade@gmail.com)
*** Member of JSCE, Ph.D., Graduate School of Engineering, Hokkaido University
(Kita-13, Nishi-8, Kita-ku, Sapporo 060-8628, Japan, TEL. +81-11-706-6214, FAX11-706-6214)
**** Member of JSCE, PhD, Dept. of Civil Eng., University of Canterbury
(Private Bag 4800, Christchurch, New Zealand, TEL. +64-3-364-22-38, andre.dantas@canterbury.ac.nz)
***** Dr. Eng., Dept. of Civil Eng., University of Canterbury
(Private Bag 4800, Christchurch, New Zealand, TEL. +64-3-364-22-33, alan.nicholson@canterbury.ac.nz)

Table 1: RCGA characteristics and procedures

Characteristics	Procedures
• Each parameter is described by a real number. • A valid interval is defined for parameter values. • *N* random vectors are generated in the first stage of the algorithm. • The algorithm ends either by an imposed condition or by reaching the determined number of generation.	• Crossover: the population is divided in two parts, and the possibility of Biased crossover is verified (Probability of crossover equal to 0.8, with 50% of the crossings being biased ones). • Mutation: the possibility of mutation for each individual is verified (Probability of mutation equal to 0.05). • Evaluation: each individual is evaluated according to the objective function. • Adjust Function or Fitness Function: the objective function is introduced in the adjusted function. • Selection: *N* individuals are selected from the original population. • Enrichment: if the best individual was not selected for the new population, then it takes place of any element randomly.

(1) Model Input

The model input is composed by characteristics of the transport network and characteristics of a set of investigated commuters. The data used in this study was obtained from a data survey carried out in Christchurch, New Zealand. In this survey students and staff of the Canterbury University were targeted. Initially, several attribute variables obtained from the survey were analyzed so as to identify those which could be the most appropriate to model adequately the modal choices. This analysis was based on the extensively published literature about travel behavior models. Finally, travel time (*TT*), distance from household to the university (*D*), availability of parking place inside the university campus (*P*), availability of bus service (*BS*), and bicycle ownership (*BO*) were defined as the input attribute variables to the model. The distance between household and university was included in the model as a spatial variable, in an attempt of explaining the choices between motorized modes (automobile and bus) and non-motorized modes (bicycle and walk).

(2) Utility Functions

The fitness *U* is assumed as the linear modal Utility Functions. The fitness function quantifies the optimality of a solution or chromosome in a GA[10].In this study, the fitness functions were formulated as follows:

$$\begin{gathered} U_{in} = \beta'_i + \beta_{TTi}TT_i + \beta_{Di}D_i + \beta_{Pi}P_i + \beta_{BSi}BS_i + \beta_{BOi}BO_i, \\ \forall\, i \in \{a: auto, b: bus, bc: bicycle, w: walk\} \\ \text{where}: \beta_{Pi} = 0\ \forall\, i \neq a, \beta_{BSi} = 0\ \forall\, i \neq b, \beta_{BOi} = 0\ \forall\, i \neq bc \end{gathered} \quad (1)$$

where U_{an}, U_{bn}, U_{bcn} and U_{wn} are the utilities of decision maker *n* for transport modes of automobile, bus, bicycle and walk, respectively; β'_a, β'_b, β'_{bc} and β'_w are the parameters for alternative-specific constants for modes; β_{TTa}, β_{TTb}, β_{TTbc}, β_{TTw}, β_{Da}, β_{Db},β_{Dbc}, β_{Dw}, β_{Pa}, β_{BSb} and β_{BObc} are parameters for the measurable attribute variables of travel time, travel distance, parking at university, availability of bus service and bicycle ownership, specific to transport mode.

(3) Objective Function

The objective function to be minimized by the GA, with respect to β_{TTa}, β_{TTb}, β_{TTbc}, β_{TTw}, β_{Da}, β_{Db},β_{Dbc}, β_{Dw}, β_{Pa}, β_{BSb} and β_{BObc}, is the error *E* function, that is, the percent of wrong answers in the model estimation. The function *E* was formulated as:

$$E = \frac{\sum_{n=1}^{N} e_n}{N}, \quad (2)$$

where *N* is the number of investigated samples (200 input-output data pairs were used during the training process); and e_n is the individual error ($e_n = 0$, if model output is equal to observed choice; and $e_n = 1$, otherwise).

(4) Characteristics of the RCGA

The conditions imposed to the real-coded genetic algorithm are summarized in Table 2.

Table 2: Calculation Conditions for the RCGA

Condition	Value
Population size	50
Number of function calls	10000
Mutation rate	0.05
Crossover rate	0.80

(5) Model Estimation Probabilities

Multinomial Logit structure was applied for estimating choice probabilities. By following this structure, the probability of individual n selecting mode i from the choice set C_n can be written by Equation 3.

$$P_{in} = \frac{1}{1+\sum_{j \in C_n} e^{\Delta U_{ij}{}^{n}}} (\forall j \neq i) \tag{3}$$

$$\text{where: } \Delta U_{ij}{}^{n} = (U_{in} - U_{jn})$$

4. Results

The performance of the RCGA was initially measured by analyzing the average error across modal choices as shown in Figure 1. The performance analysis shows good behavior of the algorithm. Modal choices Error is equal to 27.5 percent in the initial iteration and reaches the minimum value of 16 percent after 4500 iterations, remaining constant after this point. The RCGA estimation parameters are summarized in Table 3, where all parameters have expected signs. The attribute variable distance is positive for motorized modes, whereas it is negative for non-motorized modes. This is in accordance with the survey results.

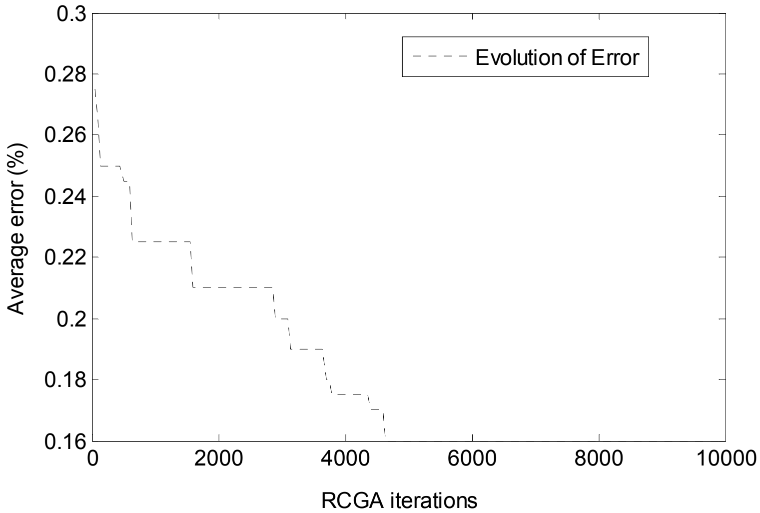

Figure 1: Evolution of Error by Iterations

Table 3: Parameter Estimation by the RCGA

Mode	β'_i	β_{TTi}	β_{Di}	β_P	β_{BS}	β_{BO}
Auto	-0.4702	-0.7132	0.0501	1.2665		
Bus	-5.9262	-0.5900	0.0747		2.2185	
Bicycle	-1.1951	-0.8560	-0.0593			2.3146
Walk	-0.8575	-0.5100	-0.1457			

A sensitivity analysis was performed so as to verify the influence of household distance to the working place in the modal choices (Figure 2). We focused on travelers' behavior regarding motorized modes and non-motorized modes. As expected, the probability of selecting both automobile and bus increase with the distance. Additionally, it was observed that there is almost no variation on the probability of selecting bus mode up to a 20 Km distance. Figure 2 shows the probability of walking to the working place is very much influenced by the household distance. On the other hand, slight variation was observed in the probabilities of selecting bicycle mode, which might reflect that cycling does not compete directly with the other three modes. These sensitivity analysis results seem to reflect correctly the investigated scenario.

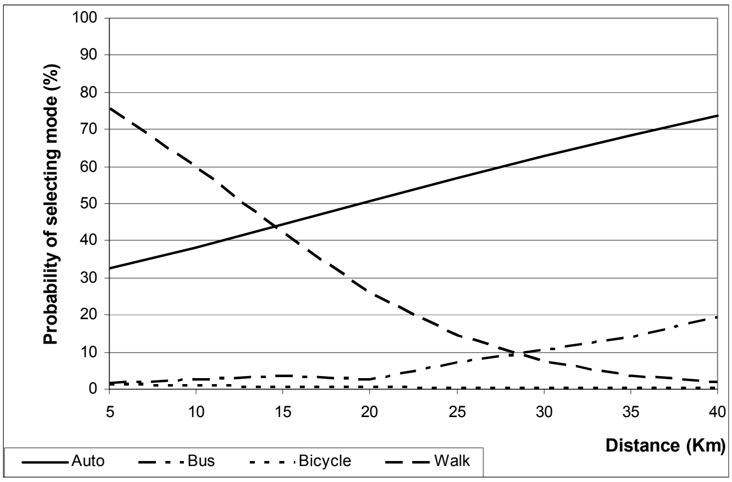

Figure 2: Travelers' Sensitivity to the Household Distance

5. Discussion and Further Work

The Real-Coded Genetic Algorithm was used for estimating parameters to be used in a Multinomial Logit modeling framework. The fitness function was composed by linear utility functions of transport modes. The objective function was formulated so as to minimize the error across choices. The model displayed good performance by presenting a reasonable evolution of the error function across iterations and a final sixteen percent error in the simulation of modal choices in the validation data set. Additionally, the parameters presented appropriate signs. For example, the positive sign of the attribute variable availability of bus service (*BS*), suggests that improvements on the bus level of service would increase its use among commuters. Moreover, the sensitivity analysis suggests that the model is able to capture correctly variations in the characteristics of investigated travelers.

The modest data set and computational capability required for developing the model seem to be advantageous characteristics of the hybrid model. However, the simplicity of the model might result in losing accuracy. In this context, further investigation should be conducted. Testing other variables in the model could increase model reliability. Moreover, Maximum Likelihood could be used for parameter estimation. Additionally, Multinomial logit model with GA-base parameters could be compared with other simulation techniques in order to verify its accuracy. Finally, the model could be tested for different discrete choice problems.

Acknowledgements

The research reported in this paper is supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan. The authors are grateful to Felipe Campelo (Research Group of Informatics for System Synthesis, Graduate School of Information Science and Technology, Hokkaido University, Japan) and Frederico G. Guimarães (Department of Electrical Engineering, Federal University of Minas Gerais, Brazil) for their significant support in the use of the Real-Coded Genetic Algorithm.

References

1) Andrade, K., K. Uchida, and S. Kagaya. Development of Transport Mode Choice Model by using Adaptive Neuro-Fuzzy Inference System. TRB, Washington DC, USA, 2006.
2) Akiyama, T., T. Takaba, and K. Mizutani. Soft Computing Approaches in Activity Based Analysis. Proceedings of the International Conference of Modeling and Management in Transportation. CD-ROM. Nagoya, Japan, 1999.
3) Hyodo, T., N. Suzuki, and Y. Takahashi. Modeling Bicycle Route Choice Behavior on Describing Bicycle Road Network in Urban Area. TRB, Washington DC, USA, 1998.
4) Mizutani, K., and T. Akiyama. Construction of Modal Choice Model with Descriptive Utility Function using Fuzzy Reasoning. Proceedings of the 9th IFSA Congress and 20th NAFIPS International Congress. CD-ROM. 2001, No 149.
5) Roorda, M, E. J. Miller, and N. Kruchten. Incorporating Within-Household Interactions into a Mode Choice Model using a Genetic Algorithm for Parameter Estimation. TRB, Washington DC, USA, 2006.
6) Holland, J. H. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. University of Michigan Press, USA, 1975.
7) Soares, G. L. Genetic Algorithms: Study, New Techniques and Applications. Master Dissertation, Federal University of Minas Gerais, Engineering School, 1997. (in Portuguese)
8) Takahashi, R. H. C. Scalar and Vector Optimization: Class Notes. Federal University of Minas Gerais, Department of Mathematics, 2004. (in Portuguese)
9) Goldberg, D.E., Genetic Algorithms in Search, Optimization Machine Learning. Addison Wesley, 1989.
10) Meister, K., M. Frick, and K. W. Axhausen. A GA-Based Household Scheduler. TRB, Washington DC, USA, 2005.