EVOLVING PARAMETERS OF LOGIT MODEL USING GENETIC ALGORITHMS

Ming Zhong, University of New Brunswick
Pawan Lingras and Will Blades, Saint Mary's University
John Douglas Hunt, University of Calgary

Introduction

Disaggregate travel demand models were developed based on discrete choice analysis methods to solve the aggregation problems existing in aggregate models (Ben-Akiva and Lerman 1985). Disaggregate modeling has evolved from binary choice, multinomial choice to probit, nested logit, ordered logit, distributed parameter logit, combinatorial logit, and mixed/cross-nested logit (Train 2003). Discrete choice models have been investigated extensively by many researchers (Batsell and Louviere 1992, Ben-Akiva and Lerman 1985, Bunch et al. 1992, Cameron 1992, Dickie et al. 1987, Hunt and McMillan 1997, Ashiabor et al. 2007). These studies used different types of data from widely different areas, and developed models with various specifications and socioeconomic variables. They clearly show that, to some extent, discrete travel demand modeling is both theoretically consistent and practically successful. A close examination to the above studies, however, indicates that multinomial and nested logit models are most widely applied, even more complicated/realistic models exist. This may be due to their simplicity and convenience in modeling. The rare applications of the complicated models, such as mixed logit and generalized extreme value (GEV), may indicate that complicated formulation and estimation process associated with these models is indeed difficult for practitioners to comprehend and use. In this study, we want to show relative simplicity of GA estimation approach in these aspects.

A literature review indicates that classic nonlinear programming algorithms (i.e., Newton-Raphson method) and simulations have been used to estimate parameters of logit models. In these methods, a maximum likelihood function is constructed based on the sum of logarithm of the probability of chosen modes and the algorithms or simulations are used to search a set of 'best-fit' parameters. For most of these classic optimization approaches, only chosen behavior was considered in the maximum likelihood functions and the rest information (such as ranking) is discarded. This is largely due to the limitation of computer capability of the classic algorithms. On the other hand, GAs are known for their capability of handling vast computing tasks and therefore they are used in this study to explicitly considering ranking information during the estimation process.

In this paper, a simulated ranking synthetic datasets, analogous to those collected in travel demand surveys, is used. Then GAs are used to re-estimate these coefficients and ASCs based on not only the chosen choice data, but also the ranking information provided. GAs are used to evolve the parameters of logit models in order to reproduce the observed rankings from simulated dataset. Three fitness functions based on exact and partial matching are tested in this study.

This paper is organized as the following: A review of logit modeling and genetic algorithms is right after the Introduction; then the paper illustrates how to use GAs for estimation of parameters of logit models; testing results are then presented; and finally the conclusions of this study are given.

Review of Logit Models

Logit choice models enjoy wide-ranging use in transportation demand modeling. These models are one part of a larger family of disaggregate choice models that has evolved and expanded over the past 50 or so years to include probit, logit, nested logit, ordered logit, distributed parameter logit, combinatorial logit, and mixed/cross-nested logit (Train, 2003). The overall postulation in disaggregate behaviour modelling is that the probability of an individual choosing a given alternative is a function of the socioeconomic characteristics of the individual and the relative attractiveness of the alternative (Ben-

Akiva and Lerman 1985). The attractiveness of alternatives is represented using the concept of utility, which is a numeric measure of the attractiveness an individual associates with an alternative. In the strictest sense, utility is not a property of alternatives, rather it is derived from the bundle of attributes that describe the alternative as perceived and valued by the individual. This derivation of a utility value from the attributes of the alternative by the individual is represented using a utility function, as follows:

$$U(a,i) = F\{X(a), c(i), K\} \tag{1}$$

Where:
U(a,i) = utility for individual *i* associates with alternative *a*
X(a) = vector of numeric measures of attributes of alternative *a*
C(i) = vector of numeric measures of characteristics of individual *i*
K = vector of utility function parameters.

The individual's choice behaviour is viewed as an exercise in maximizing this utility, either consciously or unconsciously, by selecting the alternative that provides the bundle of attributes with the greatest utility.

$$U_{in} = \max\{U_{1n}, U_{2n}, \cdots, U_{jn}\} \tag{2}$$

Based on random utility theory, the probability that any element *i* in C_n is chosen by decision maker *n* is given by:

$$P_n(i) = \Pr(U_{in} > U_{jn}, \forall j \in C_n) \tag{3}$$

The utility of each alternative is usually divided into a deterministic (V_{in}) and random component (ε_{in}) as the following:

$$\begin{aligned} \boldsymbol{P_n(i)} &= \boldsymbol{\Pr(U_{in} > U_{jn}, \forall j \in C_n, j \neq i)} \\ &= \boldsymbol{\Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n, j \neq i)} \end{aligned} \tag{4}$$

By assuming that all the disturbances ε_{in} are (1) independently distributed, (2) identically distributed, and (3) Gumbel-distributed with a location parameter η and a scale parameter $u > 0$, then

$$P_n(i) = \frac{e^{uV_{in}}}{\sum_{j \in C_n} e^{uV_{jn}}} \quad (5)$$

The estimation of parameters of logit models has been traditionally based on maximum likelihood optimization. First a maximum likelihood function like the following is constructed and optimization searching algorithms, such as Newton-Raphson and Steepest Ascent, are used to search for 'best-fit' parameters.

$$£ = P[a_1,1] \cdot P[a_2,2] \cdot P[a_3,3] \cdot \ldots \cdot P[a_n,n] \cdot \ldots \cdot P[a_N,N] \quad (6)$$

Where:

£ = the likelihood the model assigns to the observed behaviour
n = index representing particular observations in the set of observations
N = the full set of observations
a_n = the alternative observed to be chosen in observation n
$P[a_n,n]$= probability assigned by the model to the choice of alternative a_n in observation n, which is the probability calculated by the model for the observed choice.

For analytical tractability this is usually transformed into logs, giving the following:

$$L = \log £ = \Sigma \log\{P[a_n,n]\} \quad (7)$$

Where:

$n \in N$
L = 'log-likelihood' of the model assign to the observed behaviour.

For the above equation the observed data are fixed and L changes in value as the parameter values change. If the model is perfect, then

$P[a_n,n] = 1.0$ for all $n \in N$. In that case L would have a value of 0. To the extent that the model is not perfect, each $P[a_n,n]$ will have a value something between 0.0 and 1.0; with the values tending to get closer to 1.0 as the model gets better. It follows that L will have a value less than 0.0, with the value getting closer to 0.0 (larger) as the model gets better.

A literature review indicates that the parameter estimation techniques used so far are mostly classic optimization algorithms, such as Newton-Raphson, Steepest Ascent, and BHHH algorithm. Hyuk-Jae (2007) examined the effectiveness of eight classic optimization algorithms and concluded the Newton-Raphson method was most effective. Behat (2003) reviewed the development of transportation discrete choice models back to its origin from McFadden (1978) and pointed out the framework of Generalized Extreme Value (GEV) models and the substantial progress in simulation methods to estimate likelihood functions have made econometric models considerably enhanced. Bastin et al. (2006) discussed issues associated with model estimation using numerical and simulation methods. These research works are all limited to developing more realistic classic statistical models and more accurate estimation techniques with various numerical and simulation methods. Little or no research has been done to use GAs to estimate parameters of logit models.

Review of Genetic Algorithms

The origin of Genetic Algorithms (GAs) is attributed to Holland's (1975) work on cellular automata. There has been significant interest in GAs over the last two decades. The range of applications of GAs includes such diverse areas as: job shop scheduling, training neural nets, image feature extraction, and image feature identification (Buckles and Petry 1994). Grefenstette (1986) showed that GAs consistently outperformed both classical gradient search techniques and various forms of random search on more difficult problems, such as optimizations involving discontinuous, noisy, high-dimensional, and multimodal objective functions. This section gives a brief overview of GAs and should not be considered to cover the whole variety of GAs.

GAs follow the principles of evolution through natural selection. The domain knowledge is represented using a candidate solution called an *individual*. In most cases an individual is a single *chromosome* represented by a vector of length n: $C = (c_i \mid 1 \le i \le n)$, where c_i is called a gene. Some GAs decode the information preserved in the chromosome to form what is known as a phenotype solution. In this manner, several different chromosomes can be decoded to give the same phenotype. For example, Gen *et al*. (1997) provided an encoding scheme that was then decoded by an $O(n^2)$ algorithm into a path in the network. The proposed encoding scheme, however, does not differentiate between coded chromosomes and decoded phenotype solutions.

A group of individuals is called a *population*. Successive populations are called *generations*. A GA starts from initial generation $G(0)$, and for each generation $G(t)$ generates a new generation $G(t + 1)$. There are at least two methods to generate the next generation: the generational replacement model and the steady state model. This paper considers the former.

The genetic algorithm is based on two fundamental evolutionary concepts (DeJong 1998):

(1) A Darwinian notion of *fitness*, which describes an individual's ability to survive; and
(2) *Genetic operators*, which determine the next generation's genetic makeup based upon the current generation.

Conventionally, genetic operations are achieved through *crossover* and *mutation* operators. The crossover operator – shown in Fig. 1 - creates new individuals called *offspring*, by recombining the genetic material of two individuals, deemed the *parents*. Individuals with higher fitness scores are selected with greater probability to be parents and "pass on" their genes to the next generation. This is known as the fitness proportional selection method. Other selection methods exist but are not discussed in this paper.

Crossovers allow exploitation of successful subspaces of the solution space. The mutation operator – shown in Fig. 2 - randomly alters one or more genes in an individual. Mutations add genetic diversity to the population. Through mutation, GAs can search

previously unexplored sections of the solution space. Mutations consequently assure that the entire search space is connected (Buckles and Petry 1994). Through crossover and mutations, GAs are able to simultaneously explore new subspaces while exploiting successful ones.

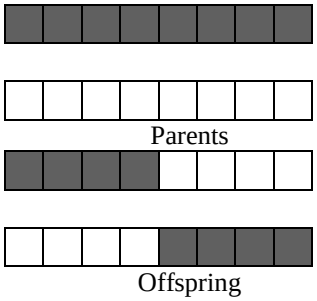

Fig. 1. Crossover operation

For the GA developed in this paper several parameters are left for the user to define: population size, number of generations, crossover ratio, and mutation ratio. Each is defined below:

1) Population size is the number of individuals in the population.
2) The number of generations is the number of times the simulation is to be run. It is used as one of our stopping criteria.
3) For a crossover ratio *p*, when two parents are selected they have probability *p* of producing two new offspring and a (1-*p*) probability of simply being copied for the next generation.
4) For a mutation ratio *q*, each new individual of the next generation has a probability *q* of undergoing a genetic mutation.

Original chromosome

Mutated chromosome

Fig. 2. Mutation operation

Design of Genetic Algorithms for Evolving Logit Model

This section describes the design of the Genetic Algorithms (GAs) for evolving the parameters of the Logit model described in Section 2. The data provides us with the values of four attributes $X_{m,k}$ for all the seven modes for each of the observations, as well as the preference ranking for each mode. The objective of our evolution is to determine the values of $\alpha_{m,k}$ and β_m such that the utility values for each mode $V_m = \sum_{k=1}^{4} \alpha_{m,k} \times X_{m,k} + \beta_m$ will provide the rankings that maximally match the ones found in the original dataset.

Since there are seven hypothetical modes and the utility function for each mode has five parameters ($a_{m,1}$, $a_{m,2}$, $a_{m,3}$, $a_{m,4}$ *and* β_m , where m = 1, 2, …, 7). A genome consisting of thirty-five genes – one for each parameter in the model is used in this study. They can be logically viewed as seven subsequences – one per mode - of five modes as shown in Fig. 3(a). Each gene subsequences consists of five genes. Fig. 3(b) shows an example of the subsequence for mode 1. The first four genes in the subsequence for mode *m* corresponds to $\alpha_{m,k}$, where *k* goes from 1 to 4 for the four attributes. The fifth gene in the subsequence for mode *m* will be β_m, which is the alternative specific constant of a given mode.

Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7

(a) Complete genome

$\alpha_{1,1}$	$\alpha_{1,2}$	$\alpha_{1,3}$	$\alpha_{1,4}$	β_1

(b) Expanded portion for Mode 1

Fig. 3 Genetic Encoding for the Logit Model

The GAs evolve the values of the thirty-five parameters such that the values of V_m provide the rankings that maximally match those observed in the simulated dataset. The fitness function, shown in Fig. 4, first calculates the values of $V_m = \sum_{k=1}^{4} \alpha_{m,k} \times X_{m,k} + \beta_m$ using the values given by the genome. The modes are then ranked using the values of V_m. The number of matches between the actual and projected ranks is the relative fitness of the genome as shown in Fig. 4. We experimented with different formulas to calculate matches. The simplest one looked for exact match, that is i^{th} ranked mode must be exactly the same for both observed and projected ranking. The first fitness function is referred to as Exact Matches Fitness Function (EMFF). The second possibility was when i^{th} actual rank matched either i^{th}, $(i+1)^{th}$, or $(i-1)^{th}$ projected rank it was considered a match. The second fitness function is referred to as Exact and Partial Matches Fitness Function (EPMFF). In the third experiment if i^{th} actual rank matched i^{th} it was considered a match with weight of 1, and if i^{th} actual rank matched either $(i+1)^{th}$ or $(i-1)^{th}$ it was considered a match with weight of 0.5. The third fitness function is referred to as Exact and Weighted Partial Matches Fitness Function (EWPMFF). The parameters of logit models are evolved with GAs and corresponding utility values of seven modes considered are computed and ranked from the highest to the lowest. And then the obtained ranks will be compared to those observed from the simulated data and the matches based on each of above three criteria are counted. The objective function of GAs is to obtain the maximum number of matches.

For all the observations:

Matches = 0

For all the modes m = 1 to 7:

Calculate $V_m = \sum_{k=1}^{4} \alpha_{m,k} \times X_{m,k} + \beta_m$

Rank the modes based on V_m

For all the ranks r = 1 to 7

Count Matches

Return Matches as the fitness

Fig. 4 Fitness Function for the Logit Model

Results and Discussions

The experiments were conducted using the synthetic data obtained from a previous study by Zhong and Hunt (2006). The simulated ranking synthetic datasets was developed by specifying the utility function coefficients and alternative specific constants (ASCs) as a priori. The data consisted of 15,000 records. Each record contained information about seven modes of transportation. The information was generated from simulations consisting of four attributes quantifying the appreciated attributes of the seven hypothetical modes being considered, their utility values and the rankings of them. For details of simulation process, please see Zhong and Hunt (2006). The three fitness functions described in the earlier section were used for three runs of the Genetic Algorithms. The population size of 100 was allowed to evolve for 300 generations. The probability of mutation was set at 0.1, and the probability of crossover was 0.7. Table 1 shows the exact matches for the three fitness functions. It can be seen that the number of exact matches went down when we allowed for partial matches and gave them equal weighting as the exact matches. This seems logical, because of the equal contributions of exact and partial matches to the fitness function, GAs can generate a set of parameters to produce a large number of partial matches by sacrificing the number of exact matches. On the other hand, it is interesting to note that, when the weighting for partial matches was halved, the number of exact and partial matches went up. It appears that, by differentiating the weights of the exact match (1.0) and the

partial match (0.5), GAs were forced to generate a better set of parameters which resulted in a higher number of both exact and partial matches. This is probably analogue to what happened in the real-world evolution: the harsher the environment, the stronger the survived generation. In this case, the 'stronger' generation means a set of better parameters.

Table 1 Number of Exact and Partial Matches for Three Fitness Functions

Fitness function	Exact matches	Partial matches	Total
EMFF	48572	NA	48572
EPMFF	41055	39247	80302
EWPMFF	49721	43807	93528

A further analysis of the exact matches by ranks is seen in Table 2. It can be seen that addition of equally weighted partial matches did not affect the top three ranks, but led to significant reduction in the matching of the bottom rank (Rank 6 and 7). The halving of the weight for the partial matches, on the other hand, resulted in much better matches for the bottom ranks. Although it is difficult to justify why the EWPMFF resulted in the highest number of exact and partial matches, it appears from the study results that an appropriate fitness function is important. Constructing an appropriate fitness function seems setting up a good environment for GAs to evolve.

Conclusions

A literature review indicates that classic optimization algorithms have been dominantly used to estimate parameters of logit model. Research using advanced optimization techniques was not found. This paper describes modeling of Logit models using genetic algorithms. The GAs were used to estimate parameters of linear Logit models in this study. In contrast to classic optimization algorithms, GAs are better in global optimization and handling vast computing tasks. Classic estimation approaches include developing a maximum likelihood function which only considers the chosen behavior. The maximum likelihood functions developed in most cases ignore the other useful information, such as travelers' ranking of all the other

modes being considered. This is largely because of the limited computing capability of classic optimization algorithms.

Table 2 Number of Exact Matches by Rank for Three Fitness Functions

Rank	EMFF	EPMFF	EWPMFF
1	13714	13714	11975
2	7514	7512	7040
3	4487	4587	4591
4	4974	5129	5198
5	8177	8331	8322
6 and 7	9706	1782	12595
Total	48572	41055	49721

GAs were used in this study to solve the problem mentioned above. Mode ranking information was explicitly considered in GA estimation process. Three fitness functions were used to evaluate how accurately GAs re-estimated parameters of logit models by comparing reproduced ranks with the observed ones. The preliminary experiments reported in this study used counted exact and partial matches resulted from using these fitness functions. It was found that using the equal weight for both exact and partial matches led to lower accuracy of predictions for low ranked modes. Reducing the weight for partial matches significantly restored and increased the accuracy of prediction of low ranked modes. It appears that setting up an appropriate fitness function is critical for successful parameter estimation with GAs.

The results reported here are preliminary. Since GAs are based on random number generation, the experiments will have to be repeated multiple times and average results should be used. These average results will be further compared with the existing techniques in a future study.

As mentioned before, a simple linear type of logit model was used in this study for illustration purpose. However, it should be noted that the approach described is generic enough that can be used

for different forms of Logit models. Future research will compare the performance of GAs with classic techniques for more complicated model types.

ACKNOWLEDGEMENTS

The authors are grateful towards the Natural Science and Engineering Research Council (NSERC), Canada and Institute of Advanced Policy Research (IAPR) of University of Calgary for their financial support.

References

Ashiabor, S. Baik, H. and Trani, A. Logit Models for Forecasting Nationwide Intercity Travel Demand in the United States, Transportation Research Record 2007, pp. 1 – 12, (2007).

Bastin, F., Cirillo, C. and Toint, P.L. Application of an Adaptive Monte Carlo Algorithm to Mixed Logit Estimation. *Transportation Research: Part B*, Vol. 40, pp. 577-593, (2006).

Batsell, R.R. and Louviere, J.J. Experimental Choice Analysis, *Marketing letter*. 2, pp. 199-214, (1992).

Bhat, C.R., Econometric models of choice: Formulation and estimation. Paper presented at *the 10th International Conference on Travel Behaviour Research*, Lucerne, August 10-15, (2003).

Ben-Akiva, M. and Lerman, *S.R. Discrete Choice Analysis*. MIT Press, Cambridge, Mass, (1985).

Buckles, B.P. and Petry, F.E. *Genetic Algorithms*. IEEE Computer Press, Los Alamitos, California (1994).

Bunch, D.S., Bradly, M., Golob, T.F., Kitamura, R. and Occhuiuzzo, G. *Demand for Clean-Fuel Vehicles in California: A Discrete Choice Stated Preference Survey*, Institute of Transportation Studies, University of California, Irvine, CA (1991).

Cameron, T.A. Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods. *Land Econometrics* 68, pp. 302-317, (1992).

DeJong, K. Learning with Genetic Algorithms: An Overview. *Machine Learning* 3, Kluwer Academic, Hingman, Mass., pp. 121-138, (1998).

Dickie, M., Fisher, A. and Gerking, S. Market Transactions and Hypothetical Demand Data: A Comparative Study, *Journal of American Statistics Association* 82, pp. 69-75, (1987).

Gen, M., Cheng, R., and Wang, D. Genetic algorithms for solving shortest path problems, in: *Proceedings of 1997 IEEE International Conference on Evolutionary Computing*, pp. 401–406, (1997).

Grefenstette, J. Optimization of Control Parameters for Genetic Algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, 16, No. 1, pp. 122-128 (1986).

Holland, J.H. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, (1975).

Hunt, J.D. and McMillan, J.D.P. Stated-Preference Examination of Attitudes Toward Carpooling to Work in Calgary, *Transportation Research Record* 1598, pp. 9-17, (1997).

Hyuk-Jae, R. *A Study on the Competitiveness of Eight Different Estimation Algorithms for Multinomial Logit Mode Choice Modelling Using Analytical Derivatives*. *M.A.Sc. thesis,* Carleton University, Ottawa, Canada, (2007).

McFadden, D. Modeling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, 75–96, (1978).

North-Holland, Amsterdam.Train, K., *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA, (2003).

Zhong, M. and Hunt, J.D. "Sensitivity Analysis of Logit Formulation and Estimation". Paper presented at *the Fifth International Conference on Traffic & Transportation Studies (ICTTS)*, Xi'an, China, August 2-4, (2006).