Iranian Journal of Science & Technology, Transaction B, Engineering, Vol. 31, No. B5, pp 547-560
Printed in The Islamic Republic of Iran, 2007

EVALUATING THE PARAMETERS AFFECTING URBAN TRIP-GENERATION*

M. ARABANI** AND B. AMANI

Dept. of Civil Engineering, University of Guilan, Guilan, I.R. Iran
Email: M_Arbani@yahoo.com
Dept. of Civil Engineering, Faculty of Higher Education, Azad University of South Tehran, Tehran, I. R. Iran

Abstract– Various parameters affect the number of household trips, therefore it can be claimed that the first step in the modeling of trip-generation is identification of the most important of these parameters. In this research the effective parameters on trip-generation are determined and evaluated by using the rough-set approach as a modern mathematical method. The primary purpose of the present paper is to find the smallest sets of attributes having quality equal to the general quality of the defined characteristics in the information system. So, the number of parameters used in the modeling is reduced. To evaluate the results, the best algorithm of the defined attributes in the information system is identified by making use of the stepwise (stepwise linear regression) method and then compare the results with each other. The results of the research indicate that the rough-set theory has a better and stronger operational capability in identifying the effective parameters in trip-generation modeling. By taking advantage of these results, the amount of necessary data for trip-generation modeling is decreased and the speed and effectiveness of information processing are increased considerably.

Keywords– Trip-generation, rough-set theory, decision-making algorithms

1. INTRODUCTION

Updating and predicting travel-demand models, particularly trip-generation models (TGM), have been discussed in different studies as a very important problem so far. These discussions and investigations have not been only related to the conformity of the models from the view point of space and time circumstances, but also there have been many arguments even regarding the problems of model specifications and defined levels of aggregation [1]. The level of aggregation of models can be divided into three categories including: Person, Household and Zone.

Some investigators, including Ortuzar and Willumsen, have emphasized that trip-generation models with more disaggregate levels enjoy better estimation and conformity [2]. Atherton and Ben-Akiva have shown that disaggregate models can evaluate the variance and travel-behavioral attributes of the variables better than the others [3]. Dowens and Gyenes have proven the said models are more able to evaluate the future predictions than the zonal-based ones. Consequently, it is expected that more disaggregate models are set forth to better estimation and prediction [4].

Supernak *et al*. have proposed the person-category approach with respect to the nature and identity of trip and travel-making parameters [5]. They have claimed that such models can estimate and evaluate better. The sample size required to develop a person-category model is several times smaller than that required to estimate a household-category model. Despite several advantages, the person-based models

*Received by the editors September 7, 2004; final revised form February 21, 2007.
**Corresponding author

were not adopted. The major limitations of these models are the difficulty of studying family interaction effects, determining household income, and the way this is distributed among family members. However, Supernak *et al*. argue that it is not clear how vital these considerations are and how they can be effectively incorporated, even in household-based models [6]. Nevertheless, their models were set forth as an implicit fashion in the last years of the 1960s and were not used from then on. Finally, the experts concluded at the beginning of the 1970s that the most suitable study unit concerning trip-generation modeling is household [7]. (For more details regarding the said issues, refer to the issues proposed by Fleet & Robertson and Ortuzar & Willumsen) [2, 8].

Consequently, the related experts, through long years and in conjunction with existing ideas, have constructed various types of models and endeavored to evaluate the number of generated trips by newer and better models. The general trends of the constructed and corrected models are as follows:

- Regression-based models [9].
- Cross-classification models [10].
- Discrete choice models such as Probit and Logit models [11].
- Simulations methods such as Smash, Amos & Starchild System, fuzzy-logic models and artificial neural networks [12].

One of the important problems regarding the construction of trip-generation models is the identification of effective parameters on the number of trips produced by a household. Accordingly, the experts and researchers have always tried to use the parameters indicating the number of trips generated by a household in the best way. Therefore, to gain this goal, they have relied on their knowledge and experience.

Gradually, by developing methods like artificial neural network, the way was paved for the selection of the parameters for experts. Nevertheless, the aforementioned methods did not enjoy enough efficiency and accuracy to cross out the surplus parameters and determine the shortest decision-making algorithms. In this regard, the rough-set theory can be considered as an efficient method to analyze the data and information and delete the extra factors and attributes existing in the information system. One of the paramount advantages of using this method is the recognition of the shortest decision-making algorithms with very high precision. In this research, the rough-set theory has been used to determine the most important parameters affecting the number of household trips.

2. TRIP-GENERATION

Trip can be termed as a one-way movement from a point of origin to another point called destination and the trip-generation stage is the first one out of four-stages concerning the trip demand prediction models which will quantify the total number of generated trips from an origin [7].

a) Factors affecting trip-generation

The number of trips produced by a household is highly dependent on family size and household-structure. Another factor affecting the number of household trips is its residential-density. This factor has been investigated in trip-generation models, generally by parameters such as the distance to CBD, availability of alternative modes of transport and average income level. Thus if all of these factors are supposed at the same levels, generated trips in denser zones will be more than less-density areas. Nevertheless, we should not forget that the number of walking trips in denser areas has been more; therefore these trips have to be separated from vehicular ones [13].

1. Mode Availability: Perhaps one of the most significant and outstanding defects of the previous studies about trip-generation is the lack of investigation about the effect of availability of alternative modes of transport, particularly private car, on the number of trips produced by a household [9].

Experts have strived to investigate the problem by various studies. Therefore, to this end, they studied two different cases, car-owning and non-car-owning households. However, they did not get a logical result. In other stages, some revision to the problem was tried through the definition of the variables regarding trip-generation models which can explain whether different household members use a private car or not [13]. In this study, the following classification has been used to investigate the possibility of using a private car by different household members: 1) Non-car-ownership, 2) Car-ownership, when a distinct member uses it, 3) Car-ownership, when all household members can use it.

2. Accessibility: One of the matters that should be analyzed in investigating the rates of produced trips in an instance of time is the evaluation of the effect of accessibility on trip-generation. It is clear that the number of trips made by each mode of transport is under the influence of availability and level of services of that mode to the community. Nevertheless, one of the distinct problems, always set forth regarding four-stage classic models, is the ignorance of the changes happening in communication networks devices, and also the availability of alternative transport modes [13].

3. Forecasting Variables in Trip-generation Analysis: The identification of variables used to predict the trip-generation rate has always been the center of discussions and investigations for transport programmers. These variables were usually considered as family-size, number of owned vehicles and income of a household. Thus, the researchers, from the beginning of 1980, became gradually interested in improving the trip-generation models using behavioral sciences methods and theories and studying the behavioral attributes of trip-makers.

In this regard, the main idea was the fact that the social circumstances of individuals affect their opportunities and activity choices. For example, the study of this issue, that a person lives alone or in-group, can influence his/her trips accomplishment, requiring coordination with other members of the household. So, one of the conceivable solutions is the analysis of *age-structure* and *life-style* of a household [7].

Salmon views an individual's lifestyle as the result of the individual's attitude and aim during his/her personal life. Salmon has claimed that the person's orientation in life can be assessed through three major decisions, 1) The decision to form a household, 2) The decision to participate in the labor-force, 3) The decision about spending free-time preferences [14].

In other words, in order to study a person's attitude during his/her life, we can point out some parameters including preferences for residence location, degree of career orientation, preferences for specific types of automobiles and entertainment preferences. Generally there are two functional and major elements defining and clarifying a person's life-style: his/her personality and occupational status, which can considerably clarify a person's orientation in life [15].

The number of accomplished trips by individuals is a function of allocated time and money to trips for doing their activities in different locations. Then, the life-style is treated as an effective and important parameter regarding trip-generation.

One of the empirical methods whereby the aforementioned issues can be appraised is the definition of a parameter correlated with the stages of life (or *family cycle*). These stages are: 1)The time when a couple has no children, 2) The time when pre-school children appear, 3)The time when the youngest child reaches school age, 4) The time when a youth leaves home and either lives alone or with other young adults, or marries, 5) The time when all the children of the family live independently and parents are still not retired, 6) The time when the parents are retired [7].

One of the other important issues concerning the analysis of household travel-behavior is the study of role commitments carried out by family members. With respect to the existing needs, every family member has duties and the types of duties is a function of some factors consisting of social norms, individuals' experiences, personality factors and the role of each member in the family.

In addition to the above-mentioned parameters, there are other factors such as social-culture, environment and different technologies available to users, affecting the trip-generation rate, but they are not easily explainable and interpretable [15].

4. Summary: Generally, with respect to the aforementioned issues, the effective factors on trips generated by a household are categorized as follows: 1) Household-structure, 2) Family-size, 3) CBD, 4) Availability and cost of activities, 5) Personality and life-style, 6) Technology available to users, 7) Household-income, 8) Social values and norms, 9) Transportation systems characteristics and its policies, 10) Car-ownership, 11) Network of communicational routes, 12) Educational level of the household members, 13) Mode availability, 14) Age-structure, 15) Residential-density, 16) Value of land.

3. ROUGH-SET THEORY

The Rough-set theory was proposed by Pawlak as a new mathematical tool for analyzing vague, uncertain, and imprecise information [27]. The Rough-set theory has been very effective in extracting laws from decision tables, automated extraction of rules from clinical databases, learning conceptual relations, extracting rules of water demand prediction, data mining and knowledge discovery, and managing the pavement database [17-20]. Recently Kryszkiewicz presented a general rough-set framework for dealing with incomplete information systems [21, 22].

Rough-set analysis is essentially a nonparametric statistical method that is able to handle a diverse and less immediate tangible set of factors. It provides a formal tool for transforming a data set, such as a collection of past observations or a record of experiences, into structured information in a way that it can classify objects having distinctive patterns of attributes. It is not always possible to distinguish objects on the basis of available information (descriptors). The imperfect information causes the indiscernibility of objects through the values of the attributes describing them and prevents them from being unambiguously assigned to a given single set. In this case, the only sets that can be precisely characterized with regard to the values of ranges of such attributes are lower and upper approximations of the set of objects [23]. We will now set out the basic principles of this method (for more details, see also Pawlak 1991; Van den Bergh *et al.* 1997; Slowinski and Stefanowski 1994; and Greco *et al*. 1995) [24-27].

Human knowledge based on experiences (e.g. concerning decision making in a specific field) is often recorded in a structure called the information system. This information system contains information about particular cases (objects, states, observations, events, etc.) and factors and attributes effective on them (features, variables, characteristics, symptoms, etc.). The set of attributes consists of two types. The first type (called condition attributes) is concerned with the results of some tests or measurements, data from observation, anamnesis, symptoms of cases, states etc. The other (called decision attributes) is concerned with some expert's decisions, diagnoses, classification and evaluation of the results considering the object attributes [28].

With reference to a certain finite set of objects U, it is assumed possible to perceive the differences existing between them by observing some information associated with each of them. A finite set Q of attributes is identified, which serves to identify and characterize these objects. Since the rough-set theory aims to classify and distinguish data on the basis of different values their attributes assume with reference to each object, each attribute $q \in Q$ must be able to assume different values in its domain U_q.

There must be, therefore, at least two of these values for the attribute to be a significant basis for the required characterization. If an attribute is quantitative, its domain is, in practice, partitioned into a suitable number of sub-intervals, which give a good description of the phenomenon studied, so as to avoid ending up with a distribution of values with a high number of modalities, which would not be useful for the analysis intended. The difficult choice of the bounds (called norms) used to define these sub-intervals is the importance of ensuring a correct application of this approach so that too much information is not lost in the translation of the original quantitative attribute-values into qualitative coded values.

At this point, every $x \in U$ may be introduced to a vector whose components are the distinct evaluations of x with respect to every attribute of Q and called the description of x in terms of attribute-values from set Q. The table, containing the descriptions of every $x \in U$ by means of the attributes of the set Q, is known as the information table. It is also possible to obtain a description of $x \in U$ in terms of any one subset of attributes $P \subseteq Q$.

A fundamental concept of the rough-set theory is that of the binary relation of indiscernibility, denoted by I_P. Two objects $x, y \in U$ are said to be P-indiscernible by means of the set of attributes $P \subseteq Q$ if they have the same description. Thus the binary relation I_P is reflexive, symmetric, and transitive (equivalence relation); its classes, that is, the subsets of U containing all the objects with the same description in terms of the attributes from subset P, and only these, are called P-elementary sets. If all the attributes of Q are considered, the Q-elementary sets are called atoms. The P-elementary sets, $P \subseteq Q$, generate a partition of U, in which every object $x \in U$ belongs to one and only one P-elementary set.

To explain the rough-set theory, it is necessary to introduce two other key concepts. Let $P \subseteq Q$ be a subset of attributes and $X \subseteq U$ a subset of objects of U. P-lower approximation of X, denoted by P_LX is a subset of U with the elements as the objects belonging to the P-elementary sets contained in the set X, and only these. In other words, the elements of P_LX are all the elements of U belonging to all the classes generated by the indiscernibility relation I_P and contained in X.

We define the P-upper approximation of X, denoted by P_UX as the subset of U with the elements all being the objects belonging to the P-elementary sets having at least one element in common with the set X, and only these. In other words, the elements of P_UX are all the elements of U belonging to all the classes generated by the indiscernibility relation I_P that have at least one representative belonging to X, and only these.

The difference between these sets is known as the P-boundary of X, denoted by $Bn_P(X) = P_UX–P_LX$. Therefore, $P_LX \subseteq X \subseteq P_UX$ results and, consequently, if an object x belongs to P_LX, it is also an element of X; if x belongs to P_UX, it may belong to the set X; therefore $Bn_P(X)$, constitutes the "doubtful region" (with reference to its elements, nothing can be said with certainty about its belonging to the set X). The indiscernible classes generated by I_P, therefore, constitute the basic instrument of the rough-set theory for better recognition and evaluation of data. This knowledge is intended as a family of partitions of U, generated by the indiscernibility relation I_P on U, $P \subseteq Q$.

A P-rough-set is the family of all subsets of U which includes the same lower and upper P-approximations. The intention is thus to approximate a set X, $X \subseteq U$, by means of a pair of sets associated with it, called lower approximation, P_LX, and upper approximation, P_UX, of X, which can then be considered as a particular case of the interval set. Only if $P_UX = P_LX$ does X prove to be equal to the union of a certain number of P-elementary sets and is called P-definable. Clearly, in this case (and only in this case), it is possible to affirm with certainty whether x, $x \in U$, belongs to X, $X \subseteq U$, using the set of attributes P. Moreover the accuracy of the approximation of X, equal to

$$\frac{\text{card } (P_L X)}{\text{card } (P_U X)} \tag{1}$$

will be at the maximum value (i.e., equal to 1). In general, therefore, the aim of the rough-set analysis is to establish whether x is an element of X based on the lower and upper approximations of X, rather than directly by means of a specific characteristic function.

Let $Y = (Y_1, Y_2, ..., Y_n)$ be a certain classification of U. Regarding the classification of Y, P-lower approximation and P-upper approximation respectively, are the sets in Y that have as their elements the P-lower and P-upper approximations, that is $P_LY= (P_LY_1, P_LY_2 ,..., P_LY_n)$ and $P_UY= (P_UY_1, P_UY_2 ,..., P_UY_n)$. The quality of the approximation of the partition Y considering the set of attributes P, denoted by $\gamma_P(Y)$, can be gained by the ratio of the total number of P-correctly classified objects (i.e., belonging to the P-lower approximations of Y_i, i = 1, 2,..., n), to the total number of objects considered. This ratio is called the quality of the classification and will have its maximum value (equal to one) if, and only if, each class Y_i of Y proves P-definable.

Another fundamental concept in this theory is that of attribute reduction (i.e., given a classification Y of the objects of U, the goal is searching for a minimal possible set of independent attributes (R) that has the same quality of classification as the original set of attributes P). The minimal subset $R \subseteq P \subseteq Q$ that $\gamma_R(Y) = \gamma_P(Y)$ is called Y-reduct of P and denoted by $RED_Y(P)$. (Note that a single table of information may have more than one reduct). The intersection of all the Y-reducts is known as Y-core of P, that is, $CORE_Y(P) = \bigcap RED_Y(P)$. Naturally the core contains all the attributes of P, which are the most important attributes in the information table (i.e., the most relevant for a correct classification of the objects of U).

In other words, in order to analyze the information table, it is sufficient to use any one of the reduced attributes $R \subseteq Q$. So, the classification Y of the objects of U may be characterized only without eliminating any information and there is no need for any other defined attributes of Q–R. On the other hand, each of the attributes not belonging to the core may be neglected without deteriorating the quality of the classification considered, but if any one attribute belonging to the core were eliminated from the information table, it will not be possible to obtain the highest quality of approximation with the remaining attributes.

Consequently, as mentioned above, the rough-set theory is essentially a classification method devised for non-stochastic information and by making use of it, the following results are obtainable:

- Evaluation of the relevance of particular condition attributes;
- Construction of a minimal subset of variables ensuring the same quality of description as the whole set (i.e., reducts of the set of attributes);
- Intersection of those reducts giving a core of attributes that cannot be eliminated without disturbing the quality of the description of the set of attributes; and
- Identification and elimination of irrelevant attributes [23].

4. APPLICATION OF THE ROUGH-SET THEORY IN THE ANALYSIS OF INFORMATION AND RECOGNITION OF TRIP-GENERATION ALGORITHM

Firstly, to be able to make use of the rough-set theory to detect the effective parameters on urban trip-generation, the germane data and information have to be gathered and provided.

In this research, in order to provide the data, six thousand questionnaires containing impressive specifications and characteristics on household trip-generation were distributed in the city of Rasht. Therefore, in the first step, most of the questionnaires were distributed among teachers, students, instructors, managers, experts and employees in the universities and schools. In the second step, in order to complete the obtained statistics and cover the other social categories, the rest of the questionnaires were distributed among physicians, engineers, businessmen and etc. Finally 4300 questionnaires have been used in the analysis from the total 6000.

The information was then categorized in a table, the row of which reflects the specifications of a household. Each column of the table indicates one of the characteristics considered for households and the last column shows the number of daily trips produced by those households.

Secondly, in order to utilize the rough-set theory to analyze the information table provided in the previous step, the information should be classified. To this end, each of the impressive attitudes and characteristics on the household trip-generation has been classified. The results of this classification has been shown in Tables 1 and 2. Thus, if we substitute the associated code for each investigated characteristic of households, the informational table can be provided.

In this table all the samples in the form of distinct categories have been set forth, concerning the dependent attribute. In other words, the information system is an information table in which each row indicates one of the samples (a household), that has been studied, and each column shows one of the evaluated attributes of that household in a categorized form. The last column implies the decision parameter. A portion of the input information system has been given in Table 3.

a) Rough-set out-put

In order to analyze the information and apply the rough-set theory, Rose2-Lite software has been used and its input data is the above information system.

The results obtained from the application of the rough-set theory in the analysis of the information system produced by the number of household trips and their characteristics can be divided into two major parts as follows:

1. Decision-making algorithms (Reducts): The attained decision-making algorithms are composed of defined independent parameters, existing in the system, by means of which the number of household produced trips can be estimated and evaluated regardless of the other parameters. In other words, these algorithms enjoy the quality equal to the total defined attributes in the information system in question. The algorithms resulted from the rough-set analysis are set forth in Table 4. With respect to the results of the rough-set theory, there are ten different reducts which evaluate the number of daily trips produced by each household. The shortest algorithm obtained in this study comprises five attributes and predicts the number of daily trips of each household by the combination of income level, number of full-time workers, CBD, household-structure and life-cycle status parameters. On the whole, each of the ten decision-making algorithms can be utilized for estimating the number of produced trips and these algorithms are not, from the viewpoint of quality and accuracy of approximation, different from each other. However, it is noteworthy that the shortest algorithm is always superior to the others.

Another noticeable point in rough–set analysis is that there are sometimes attributes in the information systems that are present and common among all the reducts. In this study, as Table 4 indicates, these attributes are household-income and the number of full-time workers in the family, which are called Core because without them, a more exact evaluation of the dependent parameter is not conceivable. These parameters are paramount regarding the analysis of household travel-behavioral characteristics and strongly affect the number of trips produced by each household.

2. Accuracy and quality of classification: We can evaluate the accuracy of independent parameters in predicting the trip rate's level separately. In addition, the accuracy and quality of the classification of categories and parameters defined in the information system can be appraised. The results concerning the accuracy of different levels (or thresholds) of the decision attribute have been given in Table 5. As it can be seen in the table, the accuracy and quality of the classification are equal to 1. This means that based on the attributes defined in the information system, the number of trips produced by each household can be strongly estimated.

5. APPLICATION OF LINEAR REGRESSION METHOD IN THE ANALYSIS OF INFORMATION AND RECOGNITION OF TRIP-GENERATION ALGORITHM

One of the most regular methods in making trip-generation models is the linear regression method. In this case, an attempt is made to find a linear relationship between the number of trips produced or attracted by the zone, and the average socioeconomic characteristics of the households of that zone. Selecting the best form of the linear regression equation and determining its parameters require experience and abundant study about the subject in question. Applying the linear regression model is a common form of using a correlation model as follows**:**

$$Y = a_0 + a_1 x_1 + a_2 x_2 + \ldots a_n x_n \tag{2}$$

Table 1. Categorization of investigated dependent variable

Decision attribute			
Level	Number of Household Trips		
1	$- \leq 5$	3	$10 < - \leq 15$
2	$5 < - \leq 10$	4	$15 < -$

Table 2. Categorization of investigated independent variables

Explanatory variables (Condition attributes)	
1 Family Size 1 1 person 2 2 persons 3 3 or 4 persons 4 5 or more persons	8 Household structure No. employed / Other adults 1 0 / 1 2 0 / 2 or more 3 1 / 1 or less 4 1 / 2 or more 5 2 or more / 1 or less 6 2 or more / 2 or more
2 Car-ownership 1 0 2 1 3 2 or more	9 Accessibility Taxi / Bus 1 Suitable / Suitable 2 Suitable / Unsuitable 3 Unsuitable / Suitable 4 Unsuitable / Unsuitable
3 Income (10000 Rials) 1 $- \leq 60$ 2 $60 < - \leq 90$ 3 $90 < - \leq 180$ 4 $180 < - \leq 300$ 5 $300 < - \leq 450$ 6 $450 < - \leq 700$ 7 $700 < - \leq 1000$ 8 $1000 < -$	10 Family-cycle 1 Single 2 Having a small child in the family 3 When the youngest child is at school age 4 When the children are at junior high level or higher 5 When one of the children has left the family 6 When all children live separately and parents are employed (or couples with no children). 7 When all children live separately and parents are retired
4 No. of female persons 1 0 person 2 1 person 3 2 or more persons	11 Value of land 1 High 2 Medium 3 low
5 No. of full-time workers 1 0 person 2 1 person 3 2 or more persons	12 Residential density 1 Urban (high density) 2 Suburban (medium density) 3 Rural (low density)
6 No. of Persons having driving-license 1 0 person 2 1 person 3 2 or more	13 Mode availability 1 No access 2 The family car, most of the time, is used by one of the family members, e.g. the father. 3 The family car is simply at the disposal of all family members
7 Distance to CBD (kilometer) 1 $- \leq 1$ 2 $1 < - \leq 2$ 3 $2 < - \leq 5$ 4 $5 < - \leq 10$ 5 $10 < -$	

Table 3. Part of coded table for rough-set analysis

Objects	Condition attributes													Decision attrib.
	1	2	3	4	5	6	7	8	9	10	11	12	13	
1.	2	2	5	3	2	3	1	3	1	4	2	1	3	1
2.	3	3	6	3	2	3	2	4	2	4	1	1	3	2
3.	1	1	2	1	1	2	1	1	1	1	3	1	1	1
4.	3	3	6	2	2	3	3	4	2	5	1	1	3	2
5.	3	2	5	2	2	3	2	6	1	4	3	1	3	2
6.	4	2	5	3	1	3	2	2	2	4	2	1	2	3
7.	3	1	4	2	2	3	3	4	2	4	1	1	1	2
8.	1	1	1	1	2	1	5	1	3	1	3	3	1	1
9.	3	1	5	3	2	3	1	4	2	5	2	1	1	3
10.	4	2	6	3	1	2	2	6	1	4	1	1	2	4
11.	3	1	4	1	1	3	3	2	1	1	3	2	1	2
12.	3	1	5	2	1	3	3	5	2	4	3	2	1	2
13.	1	1	3	1	1	2	3	1	2	1	3	2	1	1
14.	3	1	8	2	2	3	3	5	2	4	1	1	1	3
15.	2	1	5	2	1	1	3	2	2	4	1	1	1	1
16.	1	1	1	2	2	1	5	1	3	1	3	3	1	1
17.	4	2	7	2	3	3	1	6	2	4	3	1	3	2

Table 4. Reducts and core

Reducts:

Algorithm no.1 {Income, No. of Full-Time Workers, Household Structure, Family Cycle, CBD}

Algorithm no.2 {Income, No. of Full-Time Workers, No. of Persons Having Driving License, Family Cycle, Accessibility, Family Size}

Algorithm no.3 {Income, No. of Full-Time Workers, Residential Density, Family Cycle, Accessibility, Family Size}

Algorithm no.4 {Income, No. of Full-Time Workers, No. of Persons Having Driving License, Household Structure, Accessibility, Family Size}

Algorithm no.5 {Income, No. of Full-Time Workers, No. of Persons Having Driving License, Household Structure, Mode Availability, Family Size}

Algorithm no.6 {Income, No. of Full-Time Workers, No. of Persons Having Driving License, Household Structure, Family Cycle, No. of Females Person}

Algorithm no.7 {Income, No. of Full-Time Workers, No. of Persons Having Driving License, Household Structure, Accessibility, Family Cycle}

Algorithm no.8 {Income, No. of Full-Time Workers, No. of Persons Having Driving License, Household Structure, Mode Availability, Family Cycle}

Algorithm no.9 {Income, No. of Full-Time Workers, No. of Female Persons, Household Structure, Residential Density, Family Cycle}

Algorithm no.10 {Income, No. of Full-Time Workers, Mode Availability, Household Structure, Residential Density, Family Cycle}

Core

{Income, No. of Full-Time Workers}

In which y, x_i and a_i stand for dependent variable, independent variables and model parameters respectively, and are determined before anything else in the procedure.

The dependent variable in a household-based trip-generation model is the rate of household trip production. These models were introduced in the early 1970s. Each independent parameter in the models is an indication of a particular specification and characteristic of household travel-behavior. Therefore providing a household-based linear regression model needs full studies about the characteristics of household travel-behavior. Making use of the parameters, which can define the number of household trips in their best form, is a key problem because applying all the parameters in a model is a difficult and impractical task. Accordingly, in the modeling and forming regression equations, the stepwise method is used to build the shortest and the most appropriate linear regression equations. In this case, different parameters are tested to achieve the best linear combination in providing the model with the minimum number of independent parameters in it. So the data collected in the research was analyzed and it was found that the income parameter is the first and the most important parameter of household specifications having the highest amount of correlation with the dependent parameter, then family-size and car-ownership parameters are respectively the second and the third ones in the information system. Obtained linear regression equations are shown in Table 6. It is observed that by adding and replacing other additional parameters in the outcome structure of the equation of these three independent parameters, slight changes happen in the rate of R^2 which can be accepted. This means by applying these three independent parameters with high value of correlation, one can predict the number of trips produced by household members.

6. THE COMPARISON BETWEEN ROUGH-SET AND LINEAR REGRESSION METHODS

In this research, the decision-making algorithms obtained from the rough-set analysis and the algorithm resulted from the stepwise linear regression analysis are studied. The results of these investigations, accomplished by statistical software through the regression method, are listed in Table 7. In this table the shortest and the most important algorithms obtained from the rough-set analysis are recorded. In addition, the rate of the correlation coefficient of some linear equations resulting from different algorithms used in the various and common trip-generation models are evaluated in this research. So, we can point out the algorithm that contains the parameters of car-ownership, CBD, household-structure and income, which are key parameters in different trip-generation models (such as cross-classification method).

As shown in Table 7, the difference between the values of R^2 resulting from the decision-making algorithm of the rough-set analysis to estimate the number of household trips and that of the stepwise linear regression algorithm is 1.5%, which is inconsiderable and can be neglected. In addition, the R^2 value obtained from the rough-set algorithm is slightly different from the value of R^2observed for different evaluated algorithms of attributes. For example, we can mention the income, car-ownership, CBD, and household-structure parameters, which are the basic parameters defined in different trip-generation models. Therefore, it can be concluded that the algorithm resulting from rough-set analysis enjoys a suitable correlation coefficient to estimate the number of household trips, in comparison with the other evaluated algorithms.

The amounts of accuracy and quality of approximation have been used as a criterion in the current research. These values represent how carefully the defined independent parameters can predict the dependent parameter in the algorithm. In other words, the accuracy and quality of the evaluated algorithm in analyzing characteristics and parameters of household travel-behavior are estimated by the aforementioned criteria. Therefore, the value of accuracy and the quality of approximation were studied for different decision-making algorithms and consequently it was observed that other algorithms of attributes have less values of accuracy and quality of approximation in comparison to reducts resulting from rough-set analysis. For example, it was observed that the algorithm resulting from stepwise analysis and the algorithm obtained from the most important attributes defined in different trip-generation models, have lesser values of accuracy and quality of approximation. For instance, the difference between the values of accuracy and quality of approximation in algorithms resulting from rough-set analysis and stepwise linear method one are 64% and 48% respectively, which is indicative of the great difference between the results of the two methods.

Based on the above studies, it can be concluded that the multivariable linear regression equation obtained from the rough-set algorithm has a desirable correlation coefficient in estimating the number of household daily generated trips and enjoys the highest values of accuracy and quality of approximation (in comparison with other algorithms) in the analysis of the information and predicting the number of trips. Therefore, the decision-making algorithms resulting from rough-set analysis are the most assured algorithms in the information processing and trip-generation modeling and can work better than the other algorithms in analyzing the reasons of differences among the number of daily trips produced by different households.

Table 5. Accuracy and quality of the classification regarding the number of household trips level

Level of household trips	Accuracy
1	1
2	1
3	1
4	1
Accuracy of approximation:	1
Quality of approximation:	1

Table 6. Stepwise linear regression equations

Step	Equations	R^2	Parameters
1	$Y = 2.04X_1 - 0.14$	68.5%	Income
2	$Y = 1.35X_1 + 1.94X_2 - 2.74$	75.8%	Family size
3	$Y = 0.94X_1 + 2.05X_2 + 1.87X_3 - 2.05$	79.6%	Car ownership

Table 7. Comparison of correlation coefficient, quality and accuracy of approximation of different algorithms

Algorithm	R^2	Accuracy of approx	Quality of approx
{C ، E ، H ، J ، G}*	78.1%	1	1
{A ، B ، C}	79.6%	0.36	0.52
{B ، C ، H}	77.9%	0.48	0.69
{C ، B ، G ، H}	78.3%	0.68	0.82
{A ، B ، C ، H}	79.8%	0.66	0.80
{A ، B ، C ، G}	80.7%	0.56	0.72
{A ، B ، C ، G ، H}	80.3%	0.80	0.89
{A ، B ، C ، G ، H ، I}	80.6%	0.86	0.94
{A ، B ، C ، G ، H ، I ، K ، L}	80.9%	0.86	0.94

- * Shortest decision-making algorithms resulted from Rough-set Analysis
- A = Family Size, B = Car Ownership, C = Income, D = No Of Females, E = No. Of Full-Time Workers,
  F = No. Of Persons Having Driving License, G = CBD, H = Household Structure, I = Accessibility,
  J = Family Cycle, K = Value Of Land, L = Residential Density, M = Mode Availability
- The R^2 value resulted from the linear regression of all parameters (each 13 parameters) is 85%.

7. CONCLUSION

In the research, the rough-set theory and stepwise linear regression method have been used to determine the shortest decision-making algorithms in providing and analyzing trip-generation models. To this end, thirteen important characteristics of household travel-behavior specifications have been evaluated. The most important results are as follows:

1. The main impressive characteristics on daily trip-generation of a household are household-income and number of full-time workers (these characteristics have been obtained through Rough-set theory analysis).
2. The shortest decision-making algorithm resulting from rough-set analysis suggests the combination of household-income, number of full-time workers, household-structure, family-cycle and CBD parameters to estimate the number of trips produced by a household. This algorithm can be used instead of all the other parameters, to evaluate the number of daily trips produced by a household.
3. The results obtained from the stepwise linear regression method suggest the use of a decision-making algorithm consisting of a set of income, family-size and car-ownership parameters to estimate the number of daily trips produced by a household. It is worthy to mention that the parameters resulted from this analysis are the same parameters presented in the cross-classification approach which were used as key parameters in trip-generation models in previous years.
4. In this research, the correlation coefficients of the algorithms, which result from the rough-set analysis and the stepwise linear regression method to estimate the number of household trips, have been compared with each other. It was observed that the difference between the values of R^2 in algorithms concerning rough-set analysis and the stepwise linear regression method is negligible. It was also observed that the value of R^2resulting from the algorithm of rough-set analysis is slightly different from the value of R^2 obtained from the other evaluated algorithms of attributes. For instance, we can mention the algorithm of household-income, household-structure, car-ownership and CBD, which contains the most important parameters defined in different trip-generation models.
5. The Rough-set theory makes use of the quality of approximation value as a criterion to determine the best decision-making algorithms. This value represents how carefully the independent parameters, defined in the algorithms, can predict the dependent parameter. Therefore, in order to evaluate the results obtained from the rough-set analysis, the quality of approximation values are investigated for the algorithms concerning stepwise linear regression analysis and the ones resulting from the important parameters applied in trip-generation models (for instance, we can enumerate the set of household-income, car-ownership, household-structure and CBD parameters which are mostly the main parameters defined in trip-generation models). In this research, it was observed that other evaluated algorithms of attributes enjoy fewer values of accuracy and quality of approximation, in comparison with algorithms resulting from the rough-set analysis. So, the rough-set output reducts enjoy the highest values of accuracy and quality of approximation in analyzing households' travel-behavior specifications and estimating the number of their daily trips.

According to the studies accomplished in this research, it can be claimed that the reduced algorithms resulting from rough-set analysis are the most reliable algorithms to analyze and explain the information about the household produced trips. Thus, using the rough-set output reducts in trip-generation modeling, a significant amount of information necessary for the modeling can be reduced and the speed and effectiveness of information processing are increased considerably.

REFERENCES

1. Valentin, C. A., Prashker, J. N. & Shiftan, Y. (2003). Analysis of trip-generation in Israel for the years 1984, 1996/7 and spatial & transferability of trip-generation demand models. *T.R.B. 2003* Annual Meeting.
2. Ortuzar, J. & Willumsen, L. G. (1994). *Modeling transportation*. 2nd Edition, pp. 94-126.
3. Atherton, T. & Ben-Akiva, M. (1976). Transferability and updating of disaggregate travel demand models. *Transportation Research Record* (T.R.R.), No. 610, pp. 12-18.
4. Dowens, J. D. & Gyenes, L. (1976). Temporal stability and forecasting ability of trip-generation models in reading. TRRL Laboratory Report 726.
5. Supernak, J., Talvitie, A. & DeJohn, A. (1983). Person-category trip-generation model. *Transportation Research Record* (TRR), No. 944, pp. 74-83.
6. Supernak, J. (1983). Transportation modeling: lessons from the past and tasks for the future. *Transportation* 12, 70-90.
7. Ortuzar, J. D. D. & Willumsen, L. G. (2001). *Modeling transport*. 3rd Edition, John Wiley & Sons, pp. 199-245.
8. Fleet, C. & Robertson, S. (1968). Trip-generation in the transportation planning process. *Highway Research Board Record* 240, pp. 11-31.
9. Wilmot, C. G. (1995). Evidence on transferability of trip-generation models. *Journal of Transportation Engineering*, Vol. 09, pp. 405-410.
10. Walker, W. T. & Olanipekun, O. A. (1989). Interregional stability of household trip-generation rates from the 1986 New Jersey Home Interview Survey. *Transportation Research Record* 1220 (TRR), pp. 389-402.
11. Zhao, H. (2000). Comparison of two alternatives for trip-generation. Transportation Research Board, *79th Annual Meeting*, January 9-13, Paper No. 000546, Washington D.C.
12. Shafahi, Y. & Farzaneh, F. (2002). Application of NNs and NFSs in forecasting trip demand in large cities. T.R.B. *81th Annual Meeting*, Washington D.C.
13. Lane, R., Powell, T. J. & Smith, P. P. (1971). *Analytical transport planning*. Gerald Duckworth, pp. 116-128.
14. Salomon, I. (1983). Life-styles: a boarder perspective on travel behavior. *In Recent Advances in Travel Demand Analysis*, Gower, Aldershot, Hampshire, England.
15. Papacostas, C. S. & Prevedouros, P.D. (1993). *Transportation engineering and planning*. 2nd Edition, Prentice-Hall International, pp. 345-358.
16. Attoh-Okine, N. O. (2002). Combining use of rough-set and artificial neural networks in doweled-pavement-performance modeling-a hybrid approach. *Journal of Transportation Engineering*, May/June.
17. Skowron, A. (1995). Extracting laws from decision tables. *Compute. Intelligence*, Vol. 11, No. 2, pp. 371-388.
18. Tsumoto, S. (1998). Automated extraction of medical expert systems rules from clinical databases based on rough-set theory. *Information Science*, Vol. 112, pp. 67-84.
19. An, A. J. (1996). Discovering rules for water demand prediction: an enhanced rough-set approach. *Eng. Applic. Artif. Intell*., Vol. 9, No. 6, pp. 645-653.
20. Attoh-Okine, N. O. (1997). Rough-set application to data mining principles in pavement management database. *J. Compute. Civ. Eng*., Vol. 11, No. 4, pp. 231-237.
21. Kryszkiewicz, M. (1998). Chapter 21: Properties of incomplete information systems in the framework of rough-sets, rough-set and knowledge discovery. L. Polliowski and A. Skowron, eds., pp. 422-450.
22. Kryszkiewicz, M. (1999). Rules in incomplete information systems. *Information Science*, Vol. 113, pp. 271-292.
23. Nijkamp, P. & Pepping G. (1998). Meta-analysis for explaining the variance in public transport demand elasticities in Europe. *Journal of Transportation and Statistics*, Vol. 1, No. 1.

24. Van den Bergh, J. C. J. M., Button, K., Nijkamp, P. & Pepping, G. (1997). Meta-analysis of environmental policies. Dordrecht, The Netherlands: Kluwer.
25. Slowinski, R. & Stefanowski, J. (1994). *Handing various types of uncertainty in the rough-set approach, rough-sets, fuzzy sets and knowledge discovery*. Edited by Ziarko, W. P. Berlin, Germany: Springer.
26. Greco, S., Matarazzo, B. & Slowinski, R. (1995). Rough-set approach to multi-attribute choice and ranking problems. ICS Research Report 38/95, Warsaw University of Technology, Warsaw.
27. Pawlak, Z. (1991). *Rough-sets: Theoretical aspects of reasoning about data*. Kluwer Academic, Boston.
28. Predki, B. & Wilk, Sz. (1999). Rough-set based data exploration using rose system. In: Z.W.Ras, A.Skowron, eds, *Foundations of intelligent systems*, Lecture Notes in Artificial Intelligence, Vol. 1609. Springer-Verlag, Berlin, pp. 172-180.