Topic Area:
D1

Paper Number:
4137

Authors:
Andre Dantas, Ricardo Antunes, Yaeko Yamashita, Marcus Lamar

Title:
Marketing in Public Transportation: Neural Networks approach

Abstract

In Brazil, Public Bus Transportation (PBT) responds for the most part of displacements. However, due to low level of services provided by bus operators and the increase of informal/non-regulated transportation, there has been a sharp decrease on demand. In this context, planning agencies need now to focus on efforts towards comprehension and development of strategies to satisfy various necessities of PBT's users. As a preliminary step in this direction, this work introduces the application of Neural Networks (NN) associated to Geographical Information System (GIS) in order to process the modeling of user's preferences in PBT. A case study was conducted in Taguatinga City, Federal District, Brazil and results showed the potential of the proposed modeling.

Key words: Public Transportation, GIS, NN, and Marketing.

Method of Presentation:

(1) OHP	(X)
(2) Slide Projector	( )
(3) LCD Projector	( )

Topic Area Code: D1-4137

Marketing in Public Transportation: Neural Networks approach

Andre DANTAS
Master in Transportation, Ph.D. Student
Dept. of Civil Engineering, Nagoya Institute of Technology
Nagoya, Gokiso, Showa, 466-8555, Japan
Tel/FAX: 81-52-735-5496
Email: andre@keik1.ace.nitech.ac.jp

Ricardo ANTUNES
Master in Urban Transportation

Yaeko YAMASHITA
Professor, Ph.D.
Master in Transportation, University of Brasilia
Campus Universitario Darcy Ribeiro, Predio SG-12, Asa Norte
Brasilia, Distrito Federal, Brazil

Marcus LAMAR
Professor, Ph.D.
Dept. of Electrical Engineering, Federal University of Parana, Brazil.

1. INTRODUCTION

Due to the flexibility and low cost of implementation and maintenance, PTB is the main transportation mode especially in developing countries, such as Brazil. Daily necessity of the most urban population is performed by displacements in order to supply several needs such as working, education and leisure. In developing countries, where the PTB is the only transportation alternative, bus service still lacks of a special treatment (Vasconcellos, 1996). Traditionally, planning and management of PTB have been tacked considering users such as captive users. Generally, these activities have been done in order to reach economical equilibrium of PTB system without any concern on user's needs and their economical and social conditions (EBTU, 1988) (Soares, 1997).

It is clearly noticed that PTB presents a standardized service level for all the demand supplied through the application of specific and restricted models of attendance to the users. Consequently, the inefficient service level resulted in deep changes in the conditions of the transportation market. A new alternative of displacements has been observed, which is an informal and non-regulated service (vans, minibuses and motorcycle), that is growing in several cities leading to a new situation of competition on transportation market. In this sense, recently some initiatives are focusing on market-oriented actions.

In a new market situation, due to increasing transference of captive demand to alternative service, it is fundamental to change the central focus of the management and planning in PTB towards the consideration of user's needs. In this context, marketing techniques can be useful and efficient instrument for better understanding of demand-supply relation. However, the amount of data collection activities and the consideration of spatial characteristics that affect PTB and its users simultaneously restrict the incorporation of data on user's needs. Therefore, it is essential to develop methods bearing on these limitation factors.

As a preliminary step in this direction to model the preferences of PTB's users, we propose here the application of Neural Networks (NN) associated to Geographical Information System (GIS). This model provides the forecasting of user's preferences expressed in terms of price, safety, frequency, comfort and speed based upon socio-economic and spatial data, which are acquired from GIS operations and then processed using NN. The GIS-NN integration intends to take advantage of new computational techniques based upon georeferenced data and a non-linear classifier function, which provides a flexible and self-adaptive modelling. These characteristics can be decisive when solving geographical-related problems, such as Openshaw (1997) has showed that due to their complexity it would be quite impossible to solve them with traditional statistical regression analysis.

This paper is divided into five sections. Following this introduction, a brief explanation of the main characteristics of PTB system in Brazilian cities is reported. In the sequence, the proposed model is described on section three. In the fourth section, a case study in Taguatinga City, Federal District, Brazil is described. Finally, we discuss the obtained results and general perspectives for future improvements of this research.

2. PUBLIC BUS TRANSPORTATION IN BRAZIL

Bus transportation service in the most cities of developing country has an essential role in the urban development. Along the years in Brazil, it has been consolidated as the main transportation mode, according to NTU (2000), 60 percent of daily trips in Brazilian cities, with more than 100.000 inhabitants, are performed by bus. Additionally, it is noticed that the most part of users has PTB as the only one option on their displacements.

This scenario is rapidly changing as a consequence of low service quality provided by PTB system. This process resulted in a sharp decrease on PTB's demand in the most important Brazilian metropolitan areas, increase on traffic congestion and pollution are now detected as a consequence of informal transportation with non-regulated service. On the other hand, informal transportation, which is service supplied by anyone with a motorized vehicle, is increasing its participation on modal split. This kind of service is totally nonregular in the sense that there is no safety, insurance either taxes are not obtained from its activities. Moreover, informal transportation is not concerned with frequency and traffic regulations, but only devoted to collect passengers.

Towards the development of strategies focusing on user's needs, they must be faced as "clients". Users have to be satisfied according to their characteristics, wishes and profiles. In this context, Marketing theories are instruments to be explored in the sense that PTB's market is analyzed under new paradigms that concentrate on clients' demands. In this direction, the transformation has to start by identifying potential users and clarifying what they want from the system (Kotler, 1998). Based upon this initial knowledge, it will be possible to formulate and keep services for this new reality of PBT.

Few efforts in this direction have been observed to analyze user's preferences under different perspectives and verify their influence in PBT planning. Just recently, Martins (1998) proposed a methodological framework towards the segmentation of PBT market, but it lacks of a general mathematical formulation to model the preferences and it did not incorporate spatial reality as a factor affecting PBT's users and their decision making process.

In addition to this scenario of incipient theoretical background, obtainment of information on user's preferences is decisive to correctly express PBT's market. Due to narrow relations between users and urban environment (land use, traffic system, etc), analytical evaluations of these preferences must be performed considering spatial dependencies, as pointed out by Fisher (1999) and Longley *et al*. (1999). Therefore, spatial information associated to socio-economic and behavioral data is fundamental to express the complex nature of urban problems such as the modeling of user's preferences. However, costs involved on data collection are mostly prohibitive for planning agencies of developing countries, so models have to be conceived facing such a limitation.

3. NEURAL-SPATIAL MODELLING OF USER'S PREFERENCES

In this section, we describe a neural-spatial model to forecast attributes related to daily trips as part of the effort to obtain knowledge on user's preferences. This model concentrates on the representation and incorporation of user's characteristics, urban environment and its properties and relationship between the urban space and spatial location. These are main factors that have to be carefully incorporated in order to forecast

user's preferences of PBT due to their influence on user's behavior. Traditionally, user's characteristics such as socio-economic data are obtained and treated through statistical regression, while urban related aspects are hardly analyzed and considered into transportation studies (Vemgopal and Baets, 1994). In this model, we intend to combine both and reach a comprehensive representation of PBT user's preferences.

We integrate theoretical fundamentals of Marketing, NN and GIS. From Marketing, we assimilated the necessity to treat users as clients and in the selection of socio-economic data to represent the preferences. Additionally, GIS is used to obtain, manipulate, analyze and create spatial information into the modeling process. Finally, NN is responsible for the achievement of a non-linear modeling function (Dougherty, 1995). Both GIS and NN have been extensively and successfully applied to solve transportation problems as described by Stokes and Marucci (1995) and Himanen *et al.* (1998). However, these applications have separately explored without taking fully advantage of their interrelated capabilities, but as suggested by Fisher (1999) they have to work together.

The modeling of preferences starts by defining the general assumptions to express user's behavior. So, a person (or a PBT's user) commuting from his/her residential location, is described by two main groups of characteristics: individual; and spatial. Through the computation of these characteristics, we assume that it is possible to forecast user's preference related to attributes such as price, safety, frequency, comfort and speed. Individual characteristics represent the primary level of information. Generally, they can be divided into four groups: socio-economic, demographic, user's habits and preferences related to trip's attributes. On the other hand, spatial characteristics describe the urban environment (traffic system, land use) and the relationship between potential users, as well as the influence of location on their preferences (location and level of accessibility). They can be divided into two classes: urban environment; and spatial-locational. In this modeling, these two classes of spatial characteristics are processed and treated using GIS. Throughout spatial analysis, urban environment is supposed to represent the region surrounding potential users (or clients), while spatial-locational class provides a measurement of the position on the urban space. Figure 1 summarizes this conception and the process to reach a neural-spatial modeling.

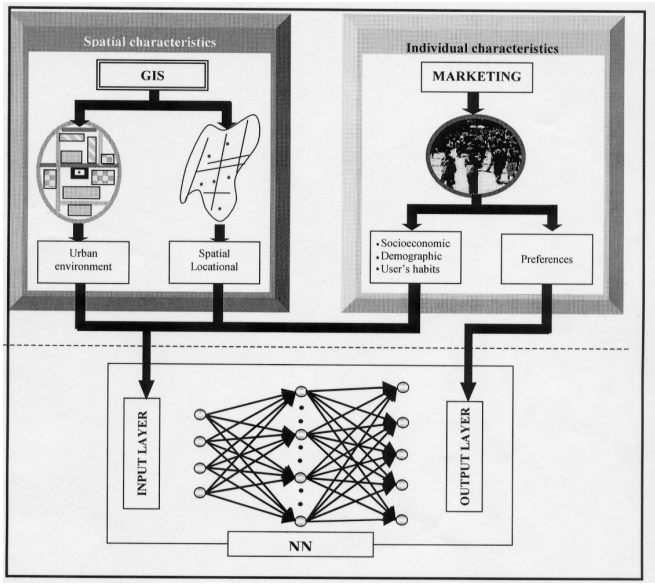


Figure 1. Framework of treatment and processing for neural-spatial modelling

Based on these general assumptions, we concentrate on the mathematical formulation of the NN, which is defined by Haykin (1999) as a massively parallel distributed processor made up of simple processing units dedicated to storage and use of experimental knowledge. In this modelling, we make use of the most applied NN structure that is a feedfoward Multilayer Perceptron (MLP). In this NN structure, we have to establish the composition of the input and the output vectors needed to training and testing activities of this modelling (Wasserman, 1989). Initially, we define the vector *X* as

$$\vec{X}_m = (\vec{SE}_m, \vec{DM}_m, \vec{HV}_m, \vec{SV}_m, \vec{US}_m, \vec{LZ}_m, \vec{NA}_m) \quad (1)$$

where:

$\vec{X}_m$ is the input vector for a sample *m*; and

$\vec{SE}_m$ $\vec{DM}_m$, $\vec{HV}_m$, $\vec{SV}_m$, $\vec{US}_m$, $\vec{LZ}_m$, $\vec{NA}_m$ are, respectively, socio-economic, demographic, user's habit, traffic system, land use, location and level of accessibility vectors for a sample *m*.

On the other hand, the output vector $\vec{Y}$ is a mathematical codification of a vector $\vec{AV}$, as presented by equation 2. This vector contains priorities av^m of user's preference for *p* attributes.

$$\vec{AV}_m = (av_1^m, av_2^m, \ldots, av_p^m) \quad (2)$$

Once $\vec{AV}$ is obtained, equation 3 is applied as described next:

$$y_k^m = \begin{cases} 1, & av_k^m = av_{\max}^m =, \text{ where } k \in (1, 2, \ldots, p) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where:

y_k^m is the coded value for attribute av_k^m;

av_k^m is the attribute value *k* for sample *m*; and

$av_{\max}^m$ is the attribute considered most important for sample *m*.

Finally, vector $\vec{Y}$ is reached by applying the following equation:

$$\vec{Y}_m = (y_1^m, y_2^m, \ldots\ldots, y_p^m) \quad (4)$$

Based on the definition of equations 1 and 4, Figure 2 presents a general NN structure.

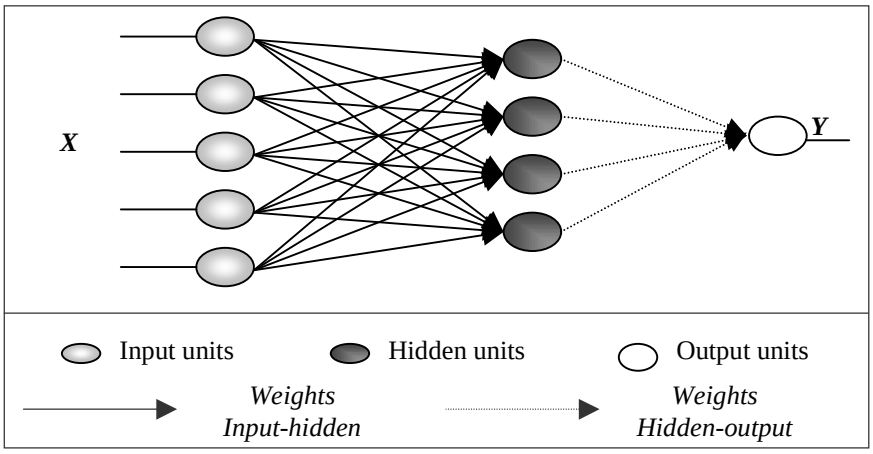


Figure 2. General NN structure

4. CASE STUDY

A case study was conducted in Taguatinga City, Federal District, Brazil, which comprehends 230 thousand people and occupies a 121,34 Km^2 area. Located 25 Km from Brasilia, Taguatinga was planned to be a satellite city, but it has developed a large variety of independent activities from the main core city, which includes industrial and commercial areas and decisively contributes to the economy of this region.

The description of this case study is conducted in three phases: GIS database construction; NN simulations; and analysis of the results.

4.1 GIS database

Initially, it was conducted the diagnosis of PBT in Taguatinga city. In this sense, we obtained data previously collected by Martins (1998). This data set contains 276 samples related to demographic, socio-economic, user's habits and preferences related to trip's attributes (price, safety, frequency, comfort and speed). Next, in order to characterize spatial characteristics, we obtained a digital map database and a set of aerial photographs (scale 1:2.000, black&white) provided by the local development governmental agency (CODEPLAN), which was incorporated into a GIS database as shown in Figure 3.

Figure 3. Digital database in the GIS software

Based upon this GIS database, firstly we digitized the traffic system (arterial, avenues, circulation axis, secondary streets and local streets) as shown in Figure 4. In order to identify land use patterns (commercial, education, heath, services, leisure, sportive clubs and religion), we applied the *United States Geological Service* (USGS) classification system (Avery and Berlin, 1990) and Taco's methodology (2000) to reach land use patterns as shown in Figure 5. The former provided a methodological structure divided in level of land use patterns, while Taco *et. al.* (2000) contributed with practical experience to perform such analysis.

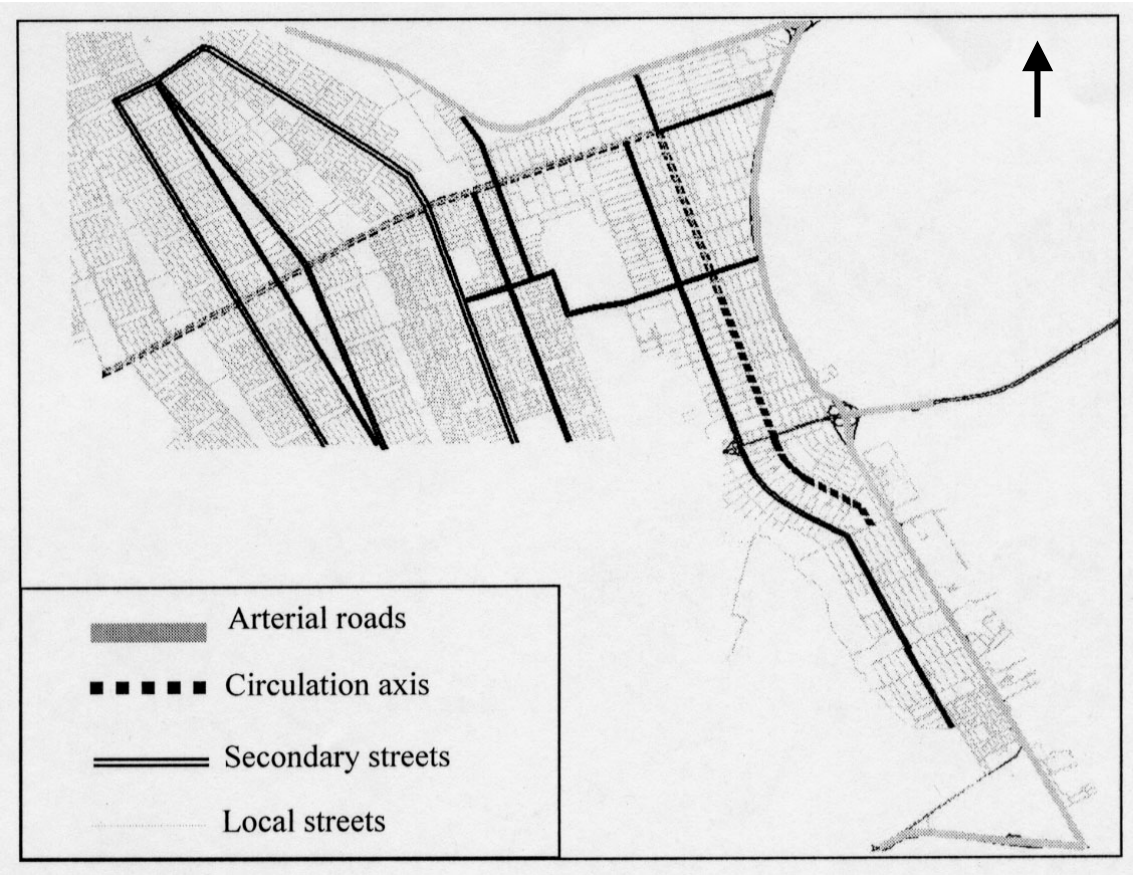


Figure 4. Traffic System

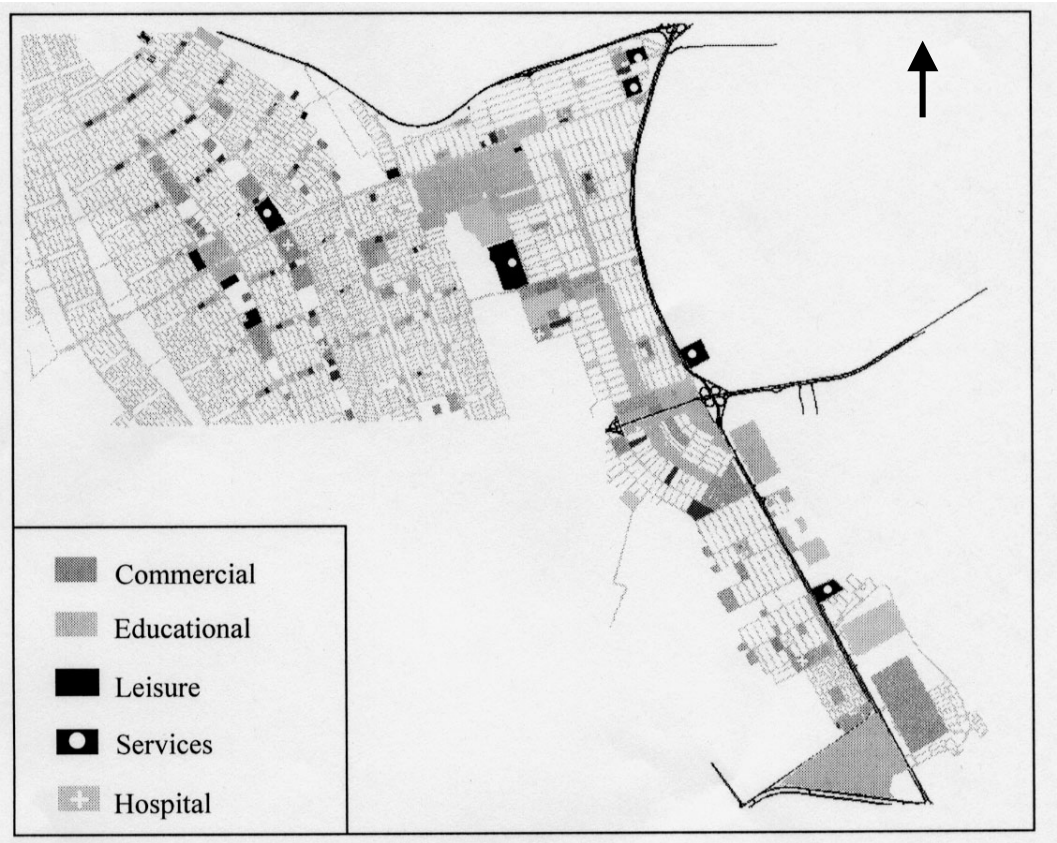
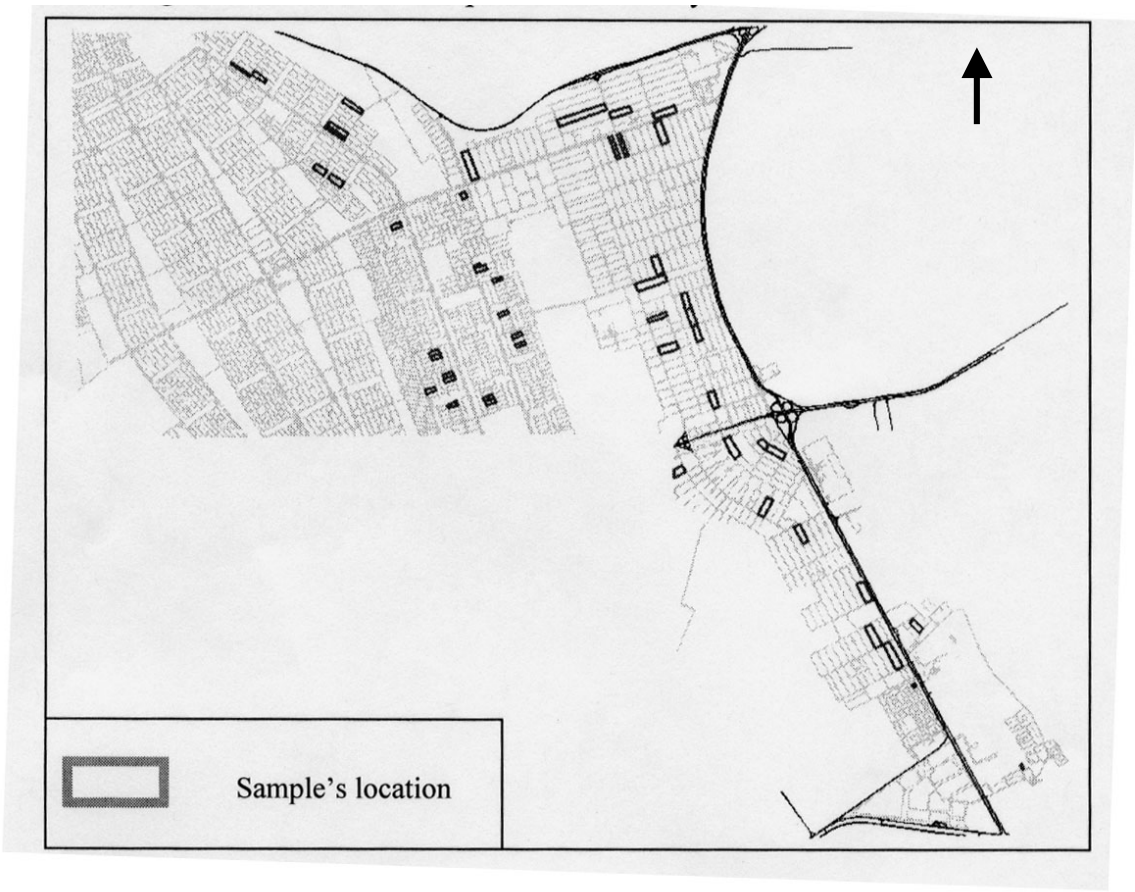


Figure 5. Land Use patterns

Data on individual characteristics were also incorporated into GIS database. In this sense, we used dada previously collected by Martins (1998). This data set contains 276 samples related to demographic, socio-economic, user's habits and preferences related to trip's attributes (price, safety, frequency, comfort, and speed). Figure 6 shows the location of these samples.



Next, we processed spatial queries on the GIS database to obtain the data needed to process NN simulations. 276 vectors containing individual and spatial characteristics were composed.

4.2 NN simulations

We defined four types of simulations to be conducted in order to evaluate different composition and treatment on the input vector. Specifically, our intention was to examine NN's efficiency using spatial data (traffic system, land use, location and level of accessibility) as well as analyze codification and normalization pre-processing procedures. Table 1 shows the considered characteristics in each type of simulation.

Table 1. Recognition rate for levels

	Variables							Pre-Processing	
	$\vec{SE}$	$\vec{DM}$	$\vec{HV}$	$\vec{SV}$	$\vec{US}$	$\vec{LZ}$	$\vec{NA}$	Codification	Normalization
Type 1	✓	✓	✓	✓	✓	✓	✓	✓	✓
Type 2	✓	✓	✓					✓	✓
Type 3	✓	✓	✓	✓	✓	✓	✓		✓
Type 4	✓	✓	✓						✓

The codification procedure is related to socio-economic, demographic and trip's habits variables. In this procedure, individual characteristics are such as gender, age and income are transformed into a binary vector $\vec{xn}_m^a$ for variable *a*. This vector is composed by $xc_{m,z}^a$ values obtained through equation 5 defined as:

$$xc_{m,z}^a = \begin{cases} 1, & \text{if } x_m^a = z \text{ where } a \in (1, 2, ..., w) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where:

x_m^a is the original value obtained from survey for variable *a* and sample *m*;

$xc_{m,z}^a$ is the coded value related to field *z*;

z is the total number of variables to be coded;

The normalization procedure is conducted to fit original values such as areas, extensions and coordinates into a limited interval ([0.1; 0.9]). Applying equation 6, we obtain a normalized vector as described next.

(6)

Where:

xn_m^a is the normalized value of variable *a* related to sample *m*;

x_m^a is the original value of variable *a* related to sample *m*;

x_{max}^a is the maximum value for variable *a*; and

x_{min}^a is the minimum value for variable *a*.

Next, we processed the data related to vector $\vec{}$ by using equation 3 and then equation 4 to obtain $\vec{Y}$ vector.

In the sequence, training (75%) and test (25%) independent data sets were generated by a random and proportional selection. Next, three-layered structures with sigmoid activation functions in the neuron outputs were defined. Using these sets, we conducted four types of simulations, as previously defined on Table 1, by applying a backpropagation algorithm

and considering a learning rate of 0.1. The networks were trained and Table 2 presents the best results reached for each type of simulation and attributes.

Table 2. Recognition rates (%) by attributes and types of simulation

	Attributes					Total
	Comfort	Frequency	Price	safety	Speed	
Type 1	33.33	55.00	0.00	80.00	38.46	53.73
Type 2	0.00	66.67	0.00	96.00	0.00	53.73
Type 3	0.00	61.11	0.00	80.00	30.77	52.24
Type 4	16.67	55.56	0.00	76.00	30.77	50.75

4.3 Analysis of the results

We notice that simulation Type 1 provided the best results among all (53,73%). This conclusion could not be reached only analyzing the total recognition rate since it is the same of simulation Type 2. So, we have to evaluate the results for each attribute and then it is clear that simulation Type 1 has a better performance than the others do. Despite the fact that the best recognition (96%) was reached for simulation Type 2 and attribute safety, NN's capability of generalization was only reached for Type 1. This means that NN-Type 1 has the capability to better forecast all attributes and it can be used for all the rest of Taguatinga City.

The combination of individual and spatial characteristics can be pointed out as the main reason for this result. Simulation Type 1 combines socio-economic, demographic, user's habits and spatial characteristics, which are normalized and coded. Especially spatial characteristics provide additional information for NN training in the sense that user's characteristics are similar all over the case study area. Then, conditions of the urban environment are decisive to describe user's preferences related to trip's attributes.

Additionally to role of spatial characteristic, we also verify a small but important contribution reached from codification and normalization procedures. The application of these two pre-processing techniques has generated a considerable improvement in the NN. This can be explained by the fact that some variables are more suitable to codification than to normalization, as used in simulation Type 3 and 4.

Another point of interest on Table 2 is the high variation of recognition rates among the attributes. It is easily noticed that comfort, price and speed attributes are poorly represented, since the former reached just 33% (Type 1) of recognition while price was completely unpredicted. Meanwhile, safety and frequency attributes presented very high rates, indicating that the training data set was able to correctly express their nature.

There are some facts that can help on the understanding of these results. Firstly, the composition of the training data set is examined. From 209 vectors in the training data set, comfort, frequency, price, safety and speed attributes are expressed by 18 (9%), 56 (27%), 17 (8%), 76 (36%) and 42 (20%) vectors, respectively. This shows a large concentration of vectors for safety attribute, while price attribute has very few samples, characterizing an imbalance training set problem which is well known when NN mainly models the dominant class.

In addition to the analysis of data set composition, we also have to consider the nature and origin of the collected data by Martins (1998) as a crucial point. Socio-economic, demographic and trip's habit data were surveyed intending to apply traditional statistical analysis without regarding the spatial nature of the problem. Consequently, we observe that there was a limited concern about sample's distribution on the space. For instance, the 276 samples can be grouped into 51 dwellings, i.e., the most part of the interviews were conducted in the same place. There were cases where 10 interviews were conducted in the same house. Obviously and unfortunately, this underestimates the real condition and needs of PBT's users. Nevertheless, the NN was able to correctly forecast some of the attributes, expressing its potential for such a problem.

5. CONCLUSION

We initially presented PBT in Brazil. In this situation, planning agencies require now instruments to develop new strategies for urban transportation services. These strategies need to be focused towards potential users, in opposition to previous notion that considered all population as homogeneous group without special requirements.

In an initial, but expected to be an efficient effort, a new conception and model to forecast user's preferences were described. They intend to be innovative by associating technological tools (GIS and NN) to a modern conception (Marketing) to obtain fundamental information for PBT's planning. It is expected that such modelling will contribute to reach user's preferences through the simultaneously incorporation of individual and spatial characteristics inside an urban environment.

In this paper, we concentrated on the mathematical description of the proposed model. We defined a general framework intending to allow future improvements according to specific needs and situations. Therefore, our modelling description has much more to be explored from now on.

The case study showed, above all, the potential of the proposed model for a real application. Through out the simulations, it was noticed the influence of the incorporation of spatial characteristics in order to provide a better "learning" and "generalization" using NN. We also verified that pre-processing procedures (codification and normalization) have an impact on the results. Finally, we observe a strong indicator that the modelling performance would be better if the collection of data had considered more properly the spatial reality. It is well known that traditional statistical models have induced to the conception of restricted representation of urban space and dynamic due to limited attention to data collection activities.

Four main perspectives for improvements of this model can be highlighted. First of all, it is necessary to develop a specific survey devoted to a neural-spatial modelling in order to avoid the concentration of data in some attributes. The second perspective is related to the creation of methodology to incorporate the influence of the urban spatial structure and its diversity into the decision making process of PBT's potential users. Next, it is fundamental to conceive a temporal-series analysis to evaluate changes along the time and their relationship with urban dynamic. Finally, new structures of NN must be simulated in order to reach a dedicated processing, which has to be specifically devoted to forecast user's preferences.

ACKNOWLEDGMENTS

The authors wish to thank the Brazilian Scientific and Technologic Development Agency (CNPq) and the Japanese Ministry of Education (Monbushou) for the scholarships that supported the development of this research.

REFERENCES

Avery, T.E. and Berlin, G. E. (1990) **Fundamental of Remote Sensing and Airphoto Intrepretation**, Maxwell Macmillan International, New York, USA.

Dougherty, M. (1995) A Review of Neural Networks Applied to Transport, **Transportation Research-C** 3(4) 247-260.

EBTU – Empresa Brasileira de Transportes Urbanos; Gerencia do Transporte Publico de Passageiros (1988) STTP, *Modulos de Treinamento, Planejamento da Operacao*. Brasilia, Brazil.

Fischer, M. M. (1999) Spatial Analysis: retrospect and prospect*;* in Goodchild, M. F.; Maguirre, D.; Rhind, D. W.; **Geographical Information Systems: Principles and Technical Issues**; Second Edition, v. 1; pp. 283-292; John Willey&Sons, USA.

Fisher. P. (1995) **Innovations in GIS 2**; Taylor&Francis, UK.

Haykin S. (1994) **Neural Networks: A Comprehensive Foundation**. Macmillan College Publishing Company, New York, USA.

Himanen, V., Nijkamp, P., Reggiani, A. (1998) ***Neural Networks in Transport Applications****;* pp. 364. Ashgate Publishing Ltd.

Kotler, P. (1998) *Administração de Marketing*. Ed. Atlas, São Paulo, Brazil.

Longley, P. A.; Goodchild, M. F.; Maguirre, D.; Rhind, D. W. (1999) **Geographical Information Systems: Principles and Technical Issues**; Second Edition, v. 1; pp. 580; John Willey&Sons, USA.

Martins, E.C.R. (1998) **Proposta Metodológica para Segmentação do Mercado de Transporte Urbano de Passageiros: Um Estudo de Caso em Taguatinga**. Dissertation in Urban Transportation, University of Brasília, Brazil.

NTU – Associacao Nacional das Empresas de Transportes Urbanos (2000)*;* **Anuário NTU 1999-2000**, Brasília, Brazil.

Openshaw, S. Openshaw, C. (1997) **Artificial intelligence in Geography***;* pp. 329, John Wiley & Sons Ltd., UK.

Soares, S. (1997) O Papel do Estado no Planejamento e na Execução dos Investimentos. in **Revista dos Transportes**, vol. 19, ANTP, n° 76.

Stokes, R. W., Marucci, G. (1995) GIS for Transportation: Current Practices, Problems and Prospects*;* **ITE Journal**; pp. 28-37; March.

Taco, P.W.G.; Yamashita, Y.; Souza, N.M. & Dantas, A.S. (2000) Trip Generation Model: A New Conception Using Remote Sensing and Geographic Information Systems. **Photogrammetrie Fernerkundung Geoinformation**, Germany, vol. 2, pp- 119-131.

Vasconcelos, E. A. (1996) **Transporte Urbano nos Países em Desenvolvimento**, Editoras Unidas São Paulo.

Vemgopal, V.; Baets, W. (1994) Neural Networks & their Applications in Marketing Management. **Journal of Systems Management**, 45(9) 16-24.

Wasserman, P. D.(1989) **Neural Computing: theory and practice**, Van Nostrand Reinhold, New York, USA.