Neural networks for analyzing service quality in public transportation

By: Concepción Garrido, Rocío de Oña & Juan de Oña

Concepción Garrido,Rocío de Oña & Juan de Oña (2015) *Neural networks for analyzing service quality in public transportation*. **Expert Systems with Applications, 41(15), 6830-6838** DOI: 10.1016/j.eswa.2014.04.045

Neural Networks for Analysing Service Quality in Public Transportation

Concepción Garrido, Rocío de Oña and Juan de Oña[(a)]

TRYSE Research Group. Department of Civil Engineering, University of Granada, ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain),

[(a)] Corresponding author. Phone: +34 958 24 99 79, Fax: +34 958 24 61 38, jdona@ugr.es

Abstract

It is essential to take into account the service quality assessment made by the passengers of a public transportation system, as well as the weight or relative importance assigned to each one of the attributes considered, in order to know its strengths and weaknesses. This paper proposes using Artificial Neural Networks (ANN) to analyze the service quality perceived by the passengers of a public transportation system. This technique is characterized by its high capability for prediction and for capturing highly non-lineal intrinsic relations between the study variables without requiring a pre-defined model. First, an ANN model was developed using the data gathered in a Customer Satisfaction Survey conducted on the Granada bus metropolitan transit system in 2007. Next, three different methods were used to determine the relative contribution of the attributes. Finally, a statistical analysis was applied to the outcomes of each method to identify groups of attributes with significant differences in their relative importance. The results show that statistical significant differences exist among several categories of attributes that have a greater or lesser impact on service quality and satisfaction. All the methods agree that Frequency is the most influential attribute in the service quality, and that other attributes such as Speed, Information and Proximity are also important.

Keywords: Service quality; bus transit; neural networks; ANN; MLP; Profile; Perturb; Connection Weights

1. Introduction

Currently, an extended use of public transport modes among citizens is one of the key aims of public administrations. Nowadays, the role of public transportation is viewed as an alternative to private cars instead of being just the support for the movement of passengers (Simoes, 2013). For individuals, car travel is generally perceived as more comfortable, flexible and faster for supporting busy lifestyles (Jakobsson Bergstad et al., 2011). However, the excessive use of these private vehicles generates environmental and social problems in cities (e.g., pollution, traffic congestion, noise, etc.), greatly exacerbating of the unsustainability of citizens' mobility.

Public transport services have to prove that they can compete with other modes, by guaranteeing effective and high quality services. The authorities are attempting to impose strong incentives on operators (Mouwen and Rietveld, 2013) by using a good definition of service quality and a good measuring method. Given that public transport services are offered directly to customers, the resultant quality of a service should be seen as an outcome of user perception (Das and Pandit, 2013) because as Bordagaray et al. (2013) stated, "Without the consumer, the market has no reason to exist". Therefore, the level of quality in a service will be high when the performance of the service fits passengers' needs and expectations.

In recent decades, practitioners, managers and researchers have focused their attention on this point of view (De Oña and De Oña, in press), striving to learn more details about how passengers evaluate a service, by considering the impact of various attributes that characterize it. Several authors have stated that service quality is a complex, fuzzy and abstract concept (Carman, 1990; Parasuraman et al., 1985), mainly because of the three properties of service: intangibility, heterogeneity and inseparability, but also because of the subjective nature of considering passengers' opinions for measuring this quality. In the literature, there are very different methods for determining this influence, although there is no consensus as to which is the best one. That is why measuring service quality is still a challenge for researchers and transport planners.

Various authors pointed to the existence of several categories of attributes that have a greater or lesser impact on service quality and satisfaction. Philip and Hazlett (1997) proposed a model with a hierarchical structure, based on three classes of attributes: pivotal, core and peripheral attributes. This model was subsequently contrasted for the rail transportation industry (Tripp and Drea, 2002). The pivotal attributes exert the greatest influence on the satisfaction levels. The UNE-EN 13186 (2003) standard classifies the service's characteristics into basic, proportional and attractive, depending on how compliance and non-compliance affects customer satisfaction. The Transit Capacity and Quality of Service Manual (TRB, 2004) groups attributes into availability factors (more important to passengers), and comfort and convenience factors (less important). Eboli and Mazzulla (2008) empirically demonstrated the existence of two categories of attributes (basic and not basic) in the preferences showed by users.

The influence of these characteristics on passengers' overall evaluation can be determined in different ways. For example, in recent years, Structural Equation Models have gained popularity among researchers, such as de Oña et al. (2013), Eboli and Mazzulla (2007; 2012), Irfan et al. (2011) or Ngatia et al. (2010). For others (e.g., Bordagaray et al. 2013; Eboli and Mazzulla, 2008;2010; Hensher, 2014; Hensher and Prioni, 2002; Hensher et al. 2003; Marcucci and Gatta, 2007) discrete choice models are a great method for deriving the importance of service quality attributes. However, most of these models have their own model assumptions and pre-defined underlying relationships between dependent and independent variables, such as normal data, linear relationships between dependent and independent variables, low multi-colinearity, and so on. According to Garver (2003), these assumptions are almost always violated in customer satisfaction research.

However, as Kikuchi (2012) says "the traditional paradigms of prediction, diagnosis, and regulation or optimization are not sufficient to deal with the extremely complex social and human systems, of which transportation engineering and planning are part". Following this idea, a novel insight has been proposed by de Oña el at. (2012; in press) applying a data mining technique for overcoming the aforementioned weaknesses and analyzing quality of service for transit operation. The methodology used was a decision tree model, which needs neither model assumptions nor pre-defined underlying relationships between the independent and the dependent variables. Following this direction, and because of the powerful results obtained with the decision tree model, the authors' interest for other data mining techniques increased.

To the authors' knowledge, the neural network approach, which is also a non-parametric model with similar advantages to the tree models has not been used before for analyzing service quality in public transportation, although it has been successfully used in other transportation engineering fields such as choice behavior (Lee et al., 2010; Xie et al. 2003). As an example, Lee et al. (2010) applied the ANN and decision trees methodology to analyze the factors affecting car drivers' alternative route choice; while Xie et al. (2003) modeled work travel mode based on three different methodologies: decision trees, ANN and multinomial logit models. Both studies concluded that ANN achieved the best fitting of the problem, with higher accuracies than the decision tree models. Thus, the aim of the present study is to use an ANN approach to investigate the influence of service characteristics on passengers' overall evaluation of a service to know the relative importance assigned to the service quality attributes. Three different methods of relative contribution (Connection Weights, Perturb and Profile) will be used. Another objective of this paper is to verify the hypotheses of Eboli and Mazzulla (2008) on the existence of different categories of attributes, and to determine if significant differences exist among groups of attributes that have higher and lower importance in overall service quality.

This paper is structured in five sections. First, the experimental context is described, specifying the survey conducted for collecting the data, and the main characteristics of the sample and perceptions rates. The following section is about the methodology and framework adopted in this context. In the fourth section the results of the analysis are shown and discussed. Finally, a brief concluding section is reported.

2. Experimental context

The survey supporting the research targeted sample of users of the metropolitan public transport service operating in Granada (Spain). The service was provided by a bus system in which 15 bus companies operated, connecting different urban agglomerations of the metropolitan area of Granada. The Transport Consortium of Granada carried face-to-face interviews on March 2007, during five days of a week. Passengers were interviewed at the main bus stops of the service, collecting a final random sample of 858 people.

The questionnaire was structured into two main sections. The aim of the first section was to collect data concerning: general information on the trip (e.g., time of the interview, bus stop, line, operator, origin, destination), socioeconomic characteristics of passengers (e.g., gender, age, private vehicle availability) and travel habits (travel reason, frequency, ticket, complementary modes from origin to bus stop and from bus stop to destination).

The majority of the respondents were female (Table 1). More than a half of the respondents were 18 to 30 years old, and only 9.5% were older than 60. Most of the people sampled (61.1%) had a private vehicle available for doing the trip. About a 29.4% of people traveled for business reasons and a similar percentage for studies purposes. The rest of the respondents traveled for other reasons, such as doctor, shopping, holidays and so on. Most of the passengers traveled with an almost daily or frequently frequency, while occasionally and sporadic passengers represented only about 10% of the sample. The most usual complementary mode used for reaching the bus stop or for reaching the destination from the bus stop was on foot. Other complementary modes had also been used, such as car, urban bus, motorbike, etc., although its representativeness was low. Finally, the consortium card and the standard ticket were the most widespread types of tickets among passengers, representing more than an 80% when combined.

The second section was more oriented to collect passenger opinions about the service. Specifically, users rated importance and perception on 12 service quality attributes that characterize the service, and also rated the overall quality of the service. A cardinal scale from 0 to 10 was used for state the ratings. The attributes used to characterize the service included information, punctuality, safety on board, driver courtesy, bus interior cleanliness, bus space, bus temperature, accessibility to/from the bus, fare, speed, frequency of service and stops proximity to/from origin/destination.

(Table 1 here)

Table 2 shows the average rates calculated from the collected data. It can be seen that there is very little variation in the importance rates stated by the passengers in the survey, considering that all the attributes are highly important. The average value of importance is concentrated in the 8.5 to 9.5 range. Therefore, this importance is uniform and practically equal in all the attributes. This is one of the serious drawbacks encountered when studying the importance of variables based on the stated opinions of passengers (Weinstein, 2000; de Oña et al., 2012).

(Table 2 here)

On the contrary, the average perception rates show higher differences among attributes. In all the cases they are lower than the mean values of the importance rates. They are concentrated in a range from 6 to 8. Nonetheless, these values are quite good because all the attributes are perceived with at least adequate quality (>6), and some of them with quite good quality (>7). The attributes characterized by the highest levels of quality were Driver courtesy, Safety on board and Bus temperature, and the one judged with the lowest level of quality was the Fare. Finally, passengers are quite satisfied with the service, with an average overall evaluation rate of 7.10.

3. Methodology

3.1. Artificial Neural Networks (ANN)

ANN are information processing systems based on the behavior of the human brain (Martín del Brío and Sanz, 2006). ANN capture the inherent information from the considered variables and learn from the existing data, even when noise is present (Kasabov, 1996), therefore no formulation or a priori model is required (Watts and Worner, 2008).

ANN´s structure is composed of elemental information processing units, called neurons. They are organized into several layers and interconnected with each other through synaptic weights. Synaptic weights represent the intensity of the interaction between every pair of neurons, and the activation functions calculate the potential of every neuron (Martín del Brío and Sanz, 2006).

The multilayer perceptron (MLP) with the back propagation learning algorithm is used in this study, since it is the most widely used type of ANN in numerous previous researches (Gedeon, 1995), and it is also a universal function approximator (Funahashi, 1989). The information always flows from the input neurons to the output neurons, and no feedback exists. The back propagation algorithm, first introduced by Werbos (1974) and further developed by Rumelhart and McClelland (1986), is the most popular learning rule (Azadeh et al., 2011).

A set of data with their target outputs is fed into the network during the supervised learning process, and an error function, represented by a hills and valleys surface, is defined. The synaptic weights values are iteratively updated until the provided output tends to be the desired, and the error function descends along the surface towards a local minimum. By defining the momentum and the learning ratio parameters, the learning convergence accelerates (Hagan et al., 1996), because the former modifies the fixed learning rhythm depending on the sign of the updated weights and the latter controls the size of the weights changes. A cross validation process has to be carried out during the learning phase to avoid overfitting. The training algorithm continues until a pre-defined number of 20,000 epochs. Thus, the database is randomly divided into three sets: training, validation and test (Bishop, 1995 and Haykin, 1999), in a 70/15/15 ratio. The test set, which contains T data, determines the global performance of the trained ANN, according to the MAPE (Delen et al., 2006) equation (see Eq. 1).

$$\boldsymbol{MAPE} = \frac{1}{T} \cdot \sum_{i=1}^{T} \boldsymbol{abs}\left(\frac{\boldsymbol{Actual\ value\ i - Set\ point\ value\ i}}{\boldsymbol{Set\ point\ value\ i}}\right) \qquad (1)$$

A three layer MLP was implemented using MATLAB software (Beale el al., 2007). The input layer was made up of twelve neurons, corresponding to the predictor variables (see Table 2 and Figure 1), while a single neuron was in the output layer, which represents the overall quality of service. Several architectures with H € [1;30] neurons in the hidden layer were trained (see Figure 1). Logarithmic sigmoidal functions were used in every layer, and 0.1 and 0.9 values were selected for the learning rate and the momentum factor parameters, respectively.

(Figure 1 here)

The data have been pretreated prior to entry into the ANN in order to accelerate the training time and improve the convergence (Masters, 1993; Martín del Brío and Sanz, 2006). Therefore, the min-max formula (Delen et al., 2006) has been used for scaling the initial variable values into the [0;1] interval, in a way that this range coincides with the limits of the activation functions (see Eq. 2).

$$\boldsymbol{v'} = \frac{\boldsymbol{v - min_A}}{\boldsymbol{max_A - min_A}}(\boldsymbol{new\_max_A} - \boldsymbol{new\_min_A}) + \boldsymbol{new\_min_A} \qquad (2)$$

The algorithm can stagnate in a local minimum because several valleys can exist in the error surface and there is no guarantee that the learning algorithm will descend towards the global minimum when the MLP is trained only once. To cover a larger number of possibilities of finding the best-behavior ANN, and in line with Paliwal and Kumar (2011), every architecture (with the same number of neurons in the hidden layer) was independently trained M times with different small initial random weights each time, since the learning algorithm can tend to reach different local minimums. Finally, the ANN architecture with the best global MAPE was selected.

3.2 Relative importance of the attributes

The following three methods will be employed in the analysis of the quality service attributes. These methods have been applied by several authors in different research fields (Gervey et al., 2003; Olden and Jackson, 2002).

3.2.1. Connection weights

This method uses the synaptic weights of the trained ANN for quantifying input attributes contribution. For every input neuron, it calculates the sum of the product of the connection input neuron – hidden neuron with the connection hidden neuron – output neuron. The higher the sums for a given input neuron, the higher the relative importance of the corresponding attribute. For more information about the method see Olden and Jackson (2002).

3.2.2. Profile

This method divides the range of values of every variable [0,1] into a number of J intervals (e.g., 10 in this study, in a way that 11 values are obtained). All variables except one are fixed successively at their first quartile, median, third quartile and maximum, while the remaining variable successively adopts each of the J+1 values of its interval for the five fixed values, so that the ANN gives five different values, of which their median is selected. Finally, the graphic representation of the median corresponding to every value of the interval is a curve of variation that indicates how the output variable (overall service quality) is affected by the changes in the input variables values. Thus, the larger variation in the ordinate axis´ values, the larger relative importance of the attribute. For more information about the method see Lek et al. (1995).

3.2.3. Perturb

Small amounts of noise are applied to each input neuron, until the 50% of the original value is perturbed, while the remaining input neurons keep unaltered. The change in the Mean Square Error (MSE) assesses the relative importance of each attribute, therefore the larger the MSE for each input perturbation, the more the relative importance of the corresponding attribute. For more information about the method see Gevrey el al. (2003).

3.3 Statistical analysis

The statistical analysis of the results was performed by means of non-parametric tests, due to the data assumption of normality is not complied. Kruskal-Wallis test is used to explore the hypothesis that significant differences exist in the importance rates of the service quality attributes, and Dunn multiple comparison method is applied to identify which are the attributes responsible of these differences. In our analyses we performed three times the non-parametric tests, one for each relative contribution method applied (connection weights, profile and perturb), on a quantitative dependent variable (importance) and qualitative independent variable (service quality attributes). Thereby, homogeneous groups of attributes can be identified with statistical differences among each other.

4. Results

4.1. Neural Networks (NN)

Every NN architectures was independently trained M=50 times, and a range of MAPE´s values were obtained for each of them (see Table 3). All the trained ANN reached a very high accuracy, above 90% in all the cases. This agrees with the results obtained by other authors who have applied ANN to evaluate the relative contribution of the service quality variables to other non-related with transportation fields, such as the education sector (Mahapatra and Khan, 2006), the service sector (Lin, 2007; Deng et al., 2008; Deng and Pei, 2009; Larasati et al., 2012), or the level of job satisfaction perceived by the employees from several gas refinery enterprises (Azadeh et al., 2011). All of them also reached a very high accuracy in their outcomes. Moreover, this reinforces the idea maintained by Garver (2002) that ANN are an adequate technique for evaluating the relative importance of the customer satisfaction variables. In addition, their accuracy is significantly higher than that obtained by de Oña el al. (2012) when using decision trees for analyzing the relative contribution of the service quality attributes, which was between 59.72% and 62.16%.

Although there are not large differences in the accuracy of the trained architectures, those with 6 neurons in the hidden layer present lower values in the MAPE range, so this architecture was the selected to carry out the following phases of this research.

(Table 3 here)

The attributes' importance obtained from each of the methods was scaled in the range [0;100]. A value of 100 was assigned to the highest importance obtained from each of the methods, while the remaining variables' importance in the same method was scaled according to this relation. This transformation is necessary for adequately comparing and analyzing the results between methods, since the values of relative importance differ among the methods by several orders of magnitude. Figure 2 shows the relative importance of the service quality attributes, expressed as a percentage, according to connection weights, profile and perturb methods.

(Figure 2 here)

The Connection Weights method commits the highest relative importance to Frequency (100.0%), followed by Speed (76.0%), Information (66.7%) and Proximity (55.5%). Safety (51.4%), Punctuality (51.4%) and Courtesy (47.8%) are also considered as of high importance, while Temperature (36.7%), Space (36.4%) and Fare (32.0%) are of middle importance, and Cleanliness (27.4%) and Accessibility (14.6%) are the least important attributes.

As with the Connection Weights method, the Profile method also assigns primary relative importance to Frequency (100.0%), and a quite high relative importance to Speed (77.7%), Information (64.2%) and Proximity (60.2%). The relative importance is also high to the Punctuality (54.4%); Safety (53.3%) and Courtesy (48.6%) attributes, and medium to Temperature (38.4%), Fare (36.4%) and Space (27.2%). Accessibility (17.3%) and Cleanliness (3.4%) are in the last positions in the ranking of importance.

In the Perturb method, Frequency (100.0%) reaches the highest relative importance, although Speed (63.6%) and Information (42.9%) are also important attributes. Unlike the results of the previous methods, Punctuality (33.0%), Safety (33.0%), Courtesy (30.3%), Proximity (23.3%) and Temperature (22.3%) are considered as of medium influence; Fare (17.9%) and Space (14.5%) as of low importance; and finally Accessibility (7.6%) and Cleanliness (7.1%) are the least important attributes.

By observing the outcomes of the application of the Connection Weights, Perturb and Profile methods, it can be said that all of them present similarities in percentage distribution terms and in the assigned position in the relative importance ranking (see Figure 3), especially between the Connection Weights and the Profile methods. The Perturb method shows more discrepancies, mainly in the Proximity attribute, placing it in the seventh position in the relative importance ranking, while the other two methods raise it to the fourth position. All the methods agree that Frequency is the globally most important attribute for evaluating the service quality in the bus public transportation; that Speed, Information and Punctuality are of great importance; and that Proximity, Safety and Punctuality are also relevant attributes. On the other hand, Accessibility and Cleanliness are the less influential attributes.

(Figure 3 here)

These results are consistent with those extracted from other previous studies (Dell´Olio et al. 2010; de Oña et al., 2012; Eboli and Mazzulla, 2010; 2011), in which different techniques (such as regression models, decision trees or relative importance indexes) have been applied for analyzing the service quality in public transportation.

4.2. Statistical analysis

A statistical analysis was conducted to confirm the statistical differences between the importance rates of some of the service quality attributes describing the Granada public transport service. This process was carried out using non-parametric techniques (Kruskal-Wallis and Dunn tests) because of the non-normality of the data. The analysis pointed out that there are significant statistical differences with a 95% confidence level between some of these attributes and across the different importance rates deduced with each method (connection weights, perturb and profile) (see Table 4).

(Table 4 here)

For the Connection Weights method, 7 homogeneous groups were identified. The variables included in each group do not present statistically significant intra-group differences, but they present statistically significant inter-groups differences. For the Perturb algorithm, 8 groups were also extracted, with the attributes that compose these homogeneous groups coinciding in almost all of them. Finally, the Profile data only allows 7 groups to be identified. These groups are marked with letters, and their average rates of importance decrease with the succession of the alphabet.

By observing the outcomes of the analysis (see Table 4), Frequency is identified in the three analysis as one independent group (Group a), that has statistical significant differences with the rest of the attributes. Frequency presents the highest influence on overall service quality, with the highest average rate in all cases. In the same direction, at the Connection weights and Perturb data, Speed, Information and Proximity also achieve high importance on service quality, making up group b and c, with significant differences compared to the rest of the groups. Regarding the results for the Profile's data, group b is also composed of Speed, Information and Proximity. This indicates that, after the attribute Frequency of the service, this group of three attributes is the one that produce the highest impact on the passengers' overall satisfaction.

On the contrary, Accessibility is the variable that presents the least average rate of importance. For the Connections Weights data, Group g is composed of this variable only and presents considerable differences with the other groups of variables. At the same time, this pattern is identified with the data of the Perturb (group h) and the Profile (group f) algorithms. Accessibility is characterized by having significant differences with the rest of groups and being one of the least important attributes (the second lowest average rate for the Perturb data and the Profile data).

For Connections Weights, Group f is composed of Bus interior cleanliness, Space on board, Bus temperature and Fare. Significant statistical differences are deduced with respect to the rest of homogeneous groups, except those attributes that belong to Group e. With regards to Group g for Perturb data, the same variables are identified, except for Cleanliness of the service. For Profile, the Group f is formed by Space, Temperature and Accessibility. There is no exact coincidence in the attributes identified in this group among the different methods. However, if we consider the attributes belonging to the last two Groups at the same time, in all of them (attributes whose importance average rates are lower to the rest of the groups), the same attributes are identified (Cleanliness, Space, Temperature, Accessibility), with exception of

Fare, which is only considered to belong to these two homogeneous groups for the Connection Weights and Perturb data.

The rest of the attributes (Punctuality, Safety, Courtesy and Fare) are of average importance. They do not produce the highest impact on the overall evaluation (such as the Frequency, Speed, Information and Proximity) nor the lowest impact (such as Accessibility, Cleanliness, Space and Temperature). Eboli and Mazzulla (2008) also distinguished among different categories of attributes according to their impact on the overall service quality of a bus transit service in Cosenza (Italy). They defined basic and non-basic attributes based on the preferences showed by the users. The Frequency and Proximity of the service (named "walking distance" in their research) were identified as having the highest impact and were defined as basic attributes. Our results agree with this point. Cleanliness and Courtesy were considered to be non-basic attributes (Eboli and Mazzulla, 2008), and these attributes were discovered to have a low and medium impact on service quality in the present research.

Thus, this statistical analysis can be used to identify homogeneous groups of attributes, according to their higher or lower impact on passengers' perception of the service quality (importance rates).

5. Conclusions

The main objective of this paper was to validate the use of Artificial Neural Networks (ANN) for modeling the service quality of public transportation systems. More specifically, we used ANN to investigate the impact that several characteristics describing a transit service have on passengers' overall evaluation its quality. The metropolitan public bus service of Granada (Spain) was considered as a study case, using the data collected on a survey conducted by the Transport Consortium of Granada in 2007.

The ANN are proposed in this research because of its numerous advantages over more traditional parametric models (such as regression models, structural equation models or logit/probit models), but also over other non-parametric models, such as decision trees. ANNs provide higher fits of the phenomenon under study.

Three different methods were applied to determine the relative contribution of the service quality attributes, and both percentages and ranking positions show that Frequency, Speed, Information and Proximity are the users' most important attributes for evaluating the service quality perceived; although Punctuality, Safety and Courtesy must also be taken into account as relevant attributes. To the contrary, Cleanliness and Accessibility are considered to be the least important attributes. The validity of this methodology is corroborated by the fact that the outcomes agree with those obtained by other authors who have used different techniques such as regression models, decision trees or importance indexes (Dell´Olio et al. 2010; de Oña et al., 2012; Eboli and Mazzulla, 2010).

This study rises a well-understanding about which groups of attributes produce a higher contribution to passengers' perceptions of service quality, and which ones play a less important role in their evaluation. Various previous works (e.g. Eboli and Mazzulla, 2008; Philip and Hazlett, 1997; TRB, 2004; UNE, 2003) have pointed to the existence of several categories of attributes that have a greater or lesser impact on SQ and satisfaction. Philip and Hazlett (1997) propose a model with a hierarchical structure, based on three classes of attributes: pivotal, core and peripheral attributes. The pivotal attributes exert the greatest influence on the satisfaction levels. Core attributes are the amalgamation of the people, processes and the service organizational structure through which consumers must interact and/or negotiate so that they can achieve/receive the pivotal attribute. And the peripheral attributes can be defined as the "incidental extras" designed to add "roundness" to the service encounter and make the whole experience for the consumer a complete delight. The UNE-EN 13186 (2003) standard classifies the service's characteristics into basic, proportional and attractive, depending on how compliance and non-compliance affects customer satisfaction. The Transit Capacity and Quality of Service Manual (TRB, 2004) groups attributes into availability factors (more important to passengers), and comfort and convenience factors (less important). Eboli and Mazzulla (2008) demonstrated the existence of two categories of attributes (basic and not basic) empirically from the preferences showed by users. Basic attributes compromise SQ when their level is low, and non-basic attributes are considered secondary service characteristics that affect SQ if they are present, but do not compromise it if they are absent.

In this study several homogeneous groups of attributes were identified, with significant statistical differences with respect to other groups regarding their influence on overall service quality. These homogeneous groups were similar across the three statistical analyses: one for the data derived from the Connection weights method, and the other two from the Perturb and Profile methods. Seven homogeneous groups were identified with the Connection weights and Profile data, and eight groups from the Perturb data. Cleanliness, Space, Temperature and Accessibility correspond to the attributes with the lowest impact on passengers' overall evaluation, while Frequency, Speed, Information and Proximity correspond to the categories with the highest influence. This is similar to research carried out in other bus public services (Eboli and Mazzulla, 2008) where Frequency and Proximity were also determined as basic attributes, having the highest impact on service quality, while Cleanliness and Courtesy were considered to be non-basic attributes.

These results provide powerful information for future policy-making due to the fact that once the transport planners know the variables that users value the most; they can define more efficient strategies for their investments. In the future, it would be also interesting to extend the interviews to non-users of the public transport service, in order to discover, not only the opinions of the current passengers about the level of quality provided, but also how non-users perceive service quality. This insight could be used to attract new users to the public transport system.

The main advantages of this approach can be summarized in the following points:

- This methodology allows to mitigate the inherent instability of ANN models, which is an important improvement in the field of "black-boxes" techniques, to which ANN belong, since until now there is no consensus about what method of relative importance must be used for determining the variables´ relative importance. This approach has demonstrated to solve this problem when using the profile, perturb and connection weight methods as methods for identifying the variables' contribution.
- The high capability of prediction and accuracy that characterize ANNs are met in the used database. And this accuracy is much higher than those obtained by some other methods as decision trees, regression models or importance indexes (Dell´Olio et al. 2010; de Oña et al., 2012; Eboli and Mazzulla, 2008; 2010) in the field of service quality in public transportation systems.

- The ranking of importance obtained for each variable is similar to those achieved by other techniques of widespread use in the field of service quality and transportation.

The main disadvantage of this approach is the duration of the calculus, since a large number of ANN must be trained and tested, and each method for determining the variables´ relative importance (profile, perturb and connection weight) must be calculated as many times as ANN are included in the selected architecture.

Since it is not easy to select the optimal ANN set, due to the intrinsic difficulty of finding the optimal or sub-optimal ANN, as future research directions, it would be interesting to know how the selection of ANN architecture affects the ranking of relative importance of the variables. In this sense, it would be also interesting to analyze how the outcomes of the approach vary when a different database is used, or what happen when the accuracy obtained from the ANN is much lower than the obtained with the current database. Therefore, one could wonder whether the approach keeps the same effectiveness.

Acknowledgements

Support from Consejería de Innovación, Ciencia y Economía of the Junta de Andalucía (Spain) (Research Project P08-TEP-03819, co-funded by FEDER) is gratefully acknowledged. The authors also acknowledge the Granada Consorcio de Transportes for making the data set available for this study.

References

Azadeh, A., Rouzbahman, M., Saberi, M. & Mohammad Fam, I., (2011). An adaptative neural network algorithm for assessment and improvement of job satisfaction with respect to HSE and ergonomics program: the case of a gas refinery. Journal of Loss Prevention in the Process Industries, 24, 361-370.

Beale, M.H., Hagan, M.T. & Demuth, H.B., (2007). Neural Network Toolbox 7. User´s Guide. MathWorks, Inc. 3 Apple Hill Drive Natic, MA 01760-2098.

Bishop, C.M., (1995). Neural networks for pattern recognition. Oxford University Press.

Bordagaray M, Dell'Olio L, Ibeas A, Cecín P. (2013). Modelling user perception of bus transit quality considering user and service heterogeneity. Transportmetrica, DOI: 10.1080/23249935.2013.823579.

Carman J. (1990). Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions. Journal of Retailing, 66, 33-55.

Das S, Pandit D. (2013). Importance of user perception in evaluating level of service for bus transit for a developing country like India: a review, Transport Reviews: A Transnational Transdisciplinary Journal, DOI:10.1080/01441647.2013.789571.

Delen, D., Sharda, R. & Bessonov, M., (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis and Prevention, 38, 434-444.

Dell´Olio, L., Ibeas, A. & Cecín, P. (2010). Modelling user perception of bus transit quality. Transport Policy, 17(6), 388-397.

Deng, W-J, Chen, W-C & Pei, W., (2008). Back-propagation neural network based importance-performance analysis for determining critical service attributes. Expert Systems with Applications, 34, 1115-1125.

Deng, W-J & Pei, W., (2009). Fuzzy neural based importance-performance analysis for determining critical service attributes. Expert Systems with Applications, 36, 3774-3784.

De Oña, J. & De Oña, R. (in press). Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches. Transportation Science.

De Oña, J., de Oña, R., & Calvo, F. J. (2012). A classification tree approach to identify key factors of transit service quality. Expert Systems with Applications, 39, 11164–11171.

De Oña J, de Oña R, Eboli L, Mazzulla G. (2013). Perceived service quality in bus transit service: A structural equation approach. Transport Policy 29: 219-226.

De Oña J, de Oña R, Eboli L, Mazzulla G. (in press). Heterogeneity in perceptions of service quality among groups of railway passengers. International Journal of Sustainable Transportation. http://dx.doi.org/10.1080/15568318.2013.849318

De Oña, J. & Garrido, C. (in press). Extracting the contribution of independent variables in neural network models: a new approach to handle instability. Neural Computing and Applications. http://dx.doi.org/10.1007/s00521-014-1573-5

Eboli L, Mazzulla G. (2007). Service quality attributes affecting customer satisfaction for bus transit. Journal of Public Transportation 10(3):21-34.

Eboli L, Mazzulla G. (2008). A Stated Preference Experiment for Measuring Service Quality in Public Transport. Transportation Planning and Technology, 31(5), 509-523.

Eboli, L., & Mazzulla, G. (2010). How to capture the passengers' point of view on a transit service through rating and choice options. Transport Reviews, 30, 435-450.

Eboli, L., & Mazzulla, G. (2012). Structural Equation Modelling for Analysing Passengers' Perceptions about Railway Services. Procedia-Social and Behavioural Science, 54, 96-106.

Funahashi, K.I., (1989). On the approximate realization of continuous mappings by neural networks. Neural Networks, 2, pp. 183-192.

Garver, M.S., (2002). Using data mining for customer satisfaction research. Marketing research, 14(1), 8-12.

Garver MS. (2003) Best practices in identifying customer-driven improvement opportunities. Industrial Marketing Management, 32(6), 455–466.

Gedeon, T.D., Wong, P.M. & Harris, D (1995). Balancing the bias and variance: network topology and pattern set reduction techniques. Proceedings of International Workshop on Artificial Neural, IWANN95, pp. 550-8. Torremolinos, España.

Gevrey, M., Dimopoulos, I. & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecological Modelling, 160, 249-264.

Hagan, M.T., Demuth, H.B. & Beale, M.H., (1996). Neural network design. Campus Publishing Service. Colorado University Bookstore. ISBN 0-9717321-0-8.

Haykin, S. (1999). Neural networks. A comprehensive Foundation. 2nd Edition. Pentince-Hall.

Hensher DA. (2014). The Relationship Between Bus Contract Costs, User Perceived Service Quality and Performance Assessment. International Journal of Sustainable Transportation 8(1): 5-27.

Hensher, D.A., & Prioni, P. (2002). A service quality index for an area-wide contract performance assessment regime. Journal of Transport Economics and Policy, 36(1), 93-113.

Hensher, D. A., Stopper, P., & Bullock, P. (2003). Service quality-developing a service quality index in the provision of commercial bus contracts. Transportation Research Part A, 37, 499-517.

Irfan Syed Muhammad, Mui Hung Kee Daisy, & Shahbaz Saman (2011). Service Quality in Rail Transport of Pakistan: A Passenger Perspective. In Proceedings of 3rd SAICON: International Conference on Management, Business Ethics and Economics (ICMBEE), December 28-29, 2011, Lahore, Pakistan.

Jakobsson Bergstad C, Gamble A, Hagman O, Polk M, Garling T. (2011) Affective-symbolic and instrumental-independence psychological motives mediating effects of socio-demographic variables on daily car use. Journalof Transport Geography 19(1),33–38.

Kasabov, N.K., (1996). Foundations of neural networks, fuzzy systems and knowledge engineering. MIT Press.

Kikuchi, S. (2012). Thoughts on the future of artificial intelligence and transportation. In Transportation Research Circular E-C168: Artificial Intelligence Applications to Critical Transportation Issues. Ed. Transportation Research Board, Washington DC, pp. 139-140.

Larasati, A., De Yong, C. & Slevitch, L., (2012). The application of neural network and logistics regression models on predicting customer satisfaction in a student-operated restaurant. Procedia Social and Behavioral Sciences, 65, 94-99.

Lek, S., Beland, A., Dimopoulos, I., Lauga, J. & Moreau, J. (1995). Improved estimation, using neural networks, of the food consumption of fish populations. Marine and Freshwater Research, 46, 1229-1236.

Lin, W-B., (2007). An empirical of service quality model from the viewpoint of management. Expert Systems with Applications, 32, 364-375.

Mahapatra, S.S. & Khan, M.S., (2006). A methodology for evaluation of service quality using neural networks. Proceedings of the International Conference on Global Manufacturing and Innovation, July 27-29.

Marcucci, E., & V. Gatta. (2007). Quality and public transport service contracts. European Transport, 36, 92-106.

Martín del Bío, B. & Sanz Molina, A., (2.006). Neural networks and fuzzy systems. Editorial RA-MA.

Masters, T (1993). Practical neural networks recipes in C++. Academic Press.

Mouwen A, Rietveld P. (2013). Does competitive tendering improve customer satisfaction with public transport? A case study for the Netherlands. Transportation Research Part A 51, 29–45.

Ngatia, G. J., Okamura, T., & Nakamura, F. (2010). The Structure of Users' Satisfaction on Urban Public Transport Service in Developing Country: the Case of Nairobi. Journal of the Eastern Asia Society for Transportation Studies, 8.

Olden, J.D. & Jackson, D.A., (2002). Illuminating the "black-box": a randomization approach for understanding variable contributions in artificial neural networks. Ecological Modelling, 154, 135-150.

Paliwal, M. & Kumar, U.A., (2011). Assessing the contribution of variables in feed forward neural network. Applied Soft Computing, 3690-3696.

Parasuraman A, Zeithaml VA, Berry LL. (1985) A conceptual model of service quality and its implications for future research. Journal of Marketing, 49, 41–50.

Philip G. & Hazlett SA. (1997) The measurement of service quality: a new P-C-P attributes model. International Journal of Quality & Reliability Management, 14(3), 260-286.

Transportation Research Board (2004) Transit Capacity and Quality of Service Manual, second ed.

Redman L, Friman M, Garling T, Hartig T. (2013) Quality attributes of public transport that attract car users: A research review. Transport Policy 25, 119–127.

Rumelhart, D.E. & McClelland, J.L., (1986). Parallel distributed processing: Explorations in the microstructure of cognition. In Foundations. Cambridge, MA: MIT Press.

Simoes A. (2013) Mobility-impaired travelers and public transport: a framework to assess quality of service. Theoretical Issues in Ergonomics Science, 14:3, 247-257.

Transportation Research Board (2004) Transit Capacity and Quality of Service Manual, second ed.

Tripp C, Drea JT. (2002) Selecting and promoting service encounter elements in passenger rail transport. The Journal of Services Marketing, 16(5), 432-442.

UNE-EN 13186 (2003) Transportation. Logistics and Services. Public Passenger Transport. Service Quality Definition, Targeting and Measurement. AENOR.

Watts, M.J., & Worner, S.P., (2008). Using artificial neural networks to determine the relative contribution of abiotic factors influencing the establishment of insect pest species. Ecological Informatics, 3, 64-74.

Weinstein A. (2000) Customer satisfaction among transit riders. How customer rank the relative importance of various service attributes. Transportation Research Record, 1735, 123–132.

Werbos, P.I., (1974). Beyond regression: New tools for prediction and analysis in the behavior sciences. Ph.D. thesis. Cambridge, MA: Harvard University.

List of Figures:

Figure 1. Artificial neural network architecture

Figure 2. Relative importance of each attribute for predicting service quality for each one of the methods (connection weights, profile and perturb).

Figure 3. Ranking of relative importance of each service quality attribute by methods (Connection Weights, Profile, Perturb).

List of Tables:

Table 1. Sample characteristics

Table 2. Importance and perceptions average rates.

Table 3. MAPE´s range for considered ANN architectures

Table 4.- Homogeneous groups with average values (n=50) significantly different among other groups

Table 1.- Sample characteristics

Characteristics	Statistics
1.Gender	Male (33.0%), Female (67.0%)
2.Age	18-30 (56.5%), 31-60 (34.1%), > 60 year-olds (9.5%)
3.Private vehicle availability	Yes (38.9%), No (61.1%)
4.Travel reason	Occupation (29.4%), Studies (22.9%), Doctor (14.2%), Shopping (4.4%), Personal activities (18.7%), Holidays (0.2%), Leisure time (8.6%), Others (1.5%)
5.Frequency	Almost daily (67.98%), Frequently (20.58%), Occasionally (8.94%), Sporadic (2.49%)
6. Complementary modes from origin to bus stop	On foot (77.6%), Car (1.9%), Urban bus (16.9%), Motorbike (0.5%), Others (3.1%)
7. Complementary modes from bus stop to destination	On foot (94.5%), Car (2.1%), Urban bus (2.3%), Motorbike (0.2%), Others (0.9%)
8.Type of ticket	Consortium card (49.6%), Standard ticket (41.2%), Senior citizen pass (4.8%), Others (4.4%)

Table 2.- Importance and perceptions average rates.

Attributes	Average Importance rates	Average Perception rates
Information	8.62	6.86
Punctuality	9.14	7.41
Safety on board	8.98	7.73
Driver courtesy	8.77	7.96
Bus interior cleanliness	8.86	7.46
Bus space	8.66	7.21
Bus temperature	8.72	7.43
Accesibility to/from the bus	8.91	6.90
Fare	8.77	6.44
Speed	8.73	7.30
Frequency of service	9.05	6.99
Proximity to/from origin/destination	8.71	7.43
Overall Service Quality		7.10

Table 3.- MAPE´s range for considered ANN architectures

H	Min	Average	Standard Deviation	Max
1	0.0400	0.0531	0.0082	0.0874
2	0.0401	0.0532	0.0090	0.0943
3	0.0438	0.0525	0.0045	0.0679
4	0.0372	0.0519	0.0067	0.0668
5	0.0424	0.0528	0.0062	0.0765
6	0.0325	0.0475	0.0064	0.0580
7	0.0417	0.0514	0.0052	0.0643
8	0.0437	0.0520	0.0044	0.0858
9	0.0436	0.0513	0.0036	0.0620
10	0.0368	0.0530	0.0074	0.0744
11	0.0399	0.0505	0.0051	0.0674
12	0.0384	0.0520	0.0086	0.0938
13	0.0396	0.0526	0.0106	0.0997
14	0.0427	0.0517	0.0043	0.0616
15	0.0354	0.0526	0.0065	0.0706
16	0.0365	0.0514	0.0087	0.0853
17	0.0379	0.0513	0.0056	0.0610
18	0.0416	0.0513	0.0070	0.0761
19	0.0406	0.0513	0.0046	0.0600
20	0.0419	0.0516	0.0049	0.0635
21	0.0395	0.0508	0.0056	0.0611
22	0.0346	0.0524	0.0078	0.0863
23	0.0424	0.0518	0.0700	0.0742
24	0.0393	0.0497	0.0047	0.0591
25	0.0385	0.0508	0.0067	0.0715
26	0.0405	0.0524	0.0063	0.0708
27	0.0417	0.0520	0.0078	0.0923
28	0.0375	0.0507	0.0056	0.0633
29	0.0427	0.0534	0.0075	0.0789
30	0.0413	0.0517	0.0091	0.0890

H: number of hidden neurons.

Table 4.- Homogeneous groups with average values (n=50) significantly different among other groups

	CONNECTION WEIGHTS	PERTURB	PROFILE
Information	2.864[b,c]	2.540[b,c]	0.787[b,c,d]
Punctuality	2.262[d]	1.950[d,e]	0.785[c,d]
Safety on board	2.098[d]	1.950[d,e]	0.785[c,d]
Driver courtesy	2.063[d,e]	1.797[e,f]	0.783[d,e]
Bus interior cleanliness	1.255[f]	0.421[h]	0.768[g]
Bus space	1.627[e,f]	0.859[g]	0.777[f]
Bus temperature	1.590[e,f]	1.319[f,g]	0.779[e,f]
Accessibility to/from the bus	0.660[g]	0.449[h]	0.776[f]
Fare	1.359[f]	1.060[f,g]	0.785[c,d]
Speed	3.284[b]	3.768[b]	0.790[b]
Frequency of service	4.315[a]	5.925[a]	0.797[a]
Proximity to/from origin/destination	2.357[c,d]	2.163[c,d]	0.787[b,c]

a,b,c,d,e,f,g,h: denotes differences statistically significant ($p<0.05$). Two or more variables with the same letter in the same column denote homogeneous subgroup.

Figure(s)

Figure 1.- Artificial neural network architecture

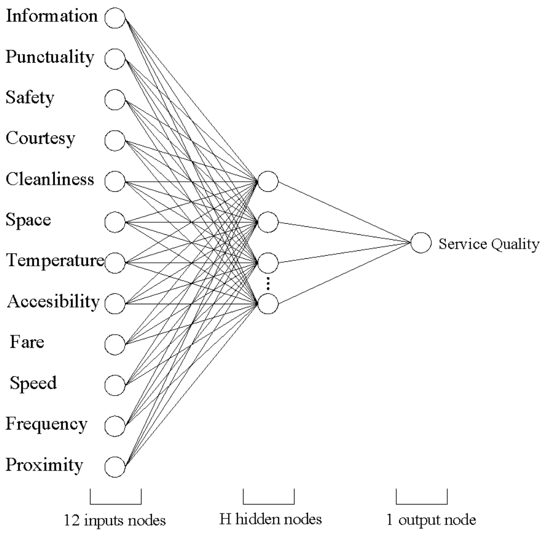

Figure 2.- Relative importance of each attribute for predicting service quality for each one of the methods (connection weights, profile and perturb).

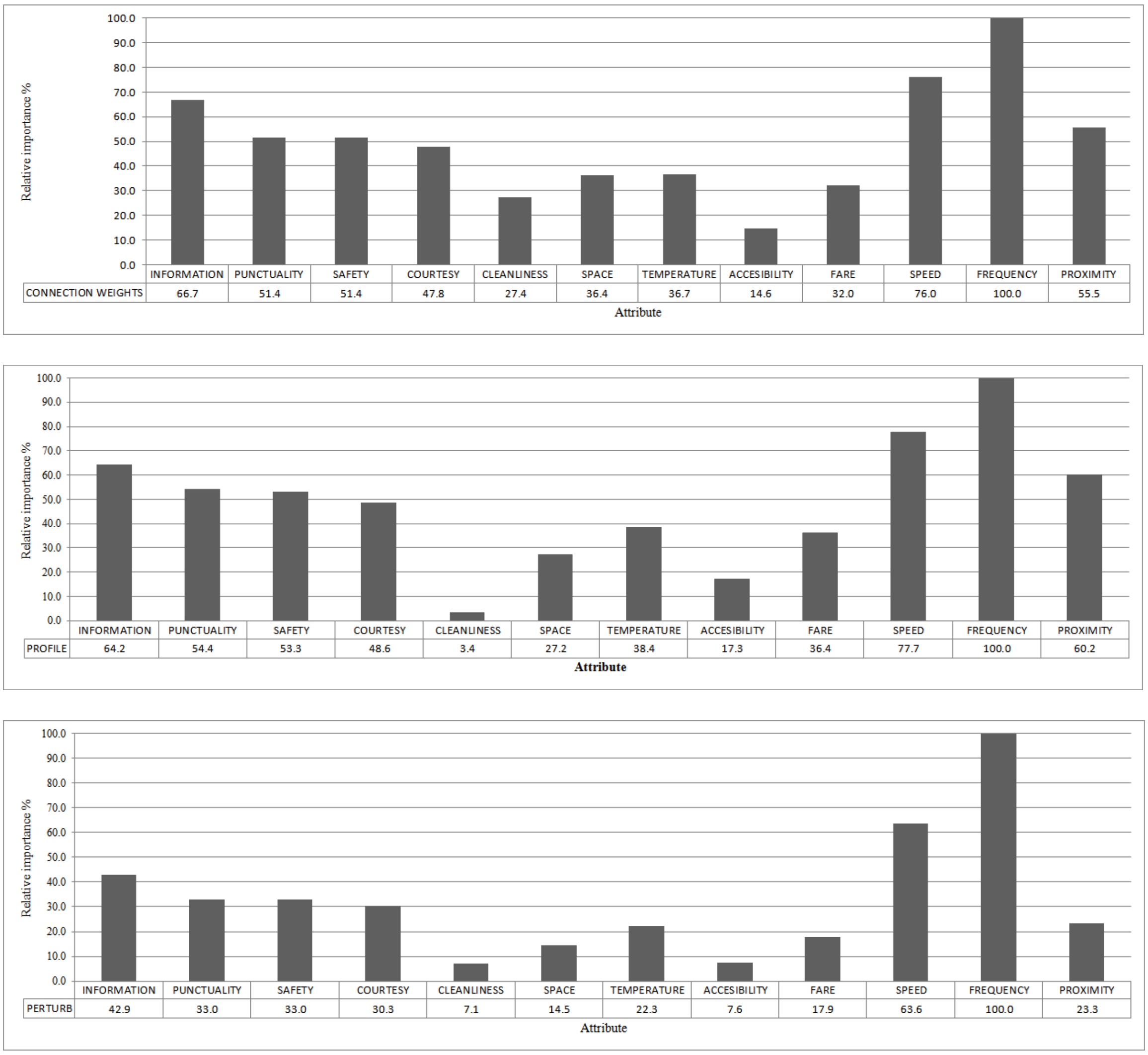

Figure 3.- Ranking of relative importance of each service quality attribute by methods (Connection Weights, Profile, Perturb).