Expert Systems with Applications 39 (2012) 11164–11171

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A classification tree approach to identify key factors of transit service quality

Juan de Oña *, Rocio de Oña, Francisco J. Calvo

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

ARTICLE INFO

Keywords:
Service quality
Bus transit
Data mining
Classification and regression trees (CART)
Non-parametric techniques

ABSTRACT

A key aspect to take into consideration when developing indices to evaluate transit service quality is to determine how much weight passengers give to each attribute when making a global assessment of service quality (SQ). The simplest method of a direct question in customer satisfaction survey (CSS) poses a number of problems, and therefore statistical regression methods have been developed to infer attribute importance on the basis of CSS or stated preference surveys. However, most regression models have their own model assumptions and pre-defined underlying relationships between dependant and independent variables. If these assumptions are violated, the model could lead to erroneous estimations. This paper proposes using a classification and regression tree (CART) that does not require any pre-defined underlying relationship between dependent and independents variables, to identify the key factors affecting bus transit quality of service. The paper uses the data gathered in a CSS conducted on the Granada metropolitan transit system in 2007, which was a non-research oriented survey. Two CART models were developed to compare the key attributes identified before and after making passengers reflect on the main aspects of the system. The outcomes show that, in a preliminary evaluation, passenger perception of SQ is basically influenced by frequency. After being asked to evaluate all the attributes, however, other attributes (e.g. proximity, speed and safety) become more important than frequency.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Promoting more sustainable modes of transport to alleviate the problems resulting from excessive use of the private car in most metropolitan areas (congestion, pollution, noise, etc.) is one of the main concerns of transport planners. Therefore, public transport (PT) service managers seek to diminish the use of private cars by promoting a consumer-based PT service and on-going quality enhancement that will lead to higher customer satisfaction. According to the Handbook for Measuring Customer Satisfaction and Service Quality (HMSCCQ) (Transportation Research Board., 1999), an increase in customer satisfaction translates into retained markets, increased use of the system, newly attracted customers, and a more positive public image.

SQ is related to a series of attributes that describe the PT service. To a large degree, it depends on the decisions system managers adopt regarding the scope of the service (in terms of territory and schedules), the type of service provided, and so on. Many authors consider that SQ should be measured from the customer's perspective since, as Berry, Zeithaml, and Parasuraman (1990) point out, "customers are the sole judges of service quality". Therefore, SQ can be measured by capturing passengers' perception of the attributes that describe the service.

Several approaches have been used to estimate the relative importance of each attribute with regards to the SQ perceived by each customer. The methods can be classified as stated importance methods (asking customers to rate each attribute on an importance scale) or derived importance methods (deriving a measure of attribute importance by statistically testing the strength of the relationship of individual attributes with overall satisfaction).

Stated importance methods are the most intuitive and simplest of the two methods. However, this approach has several drawbacks (Weinstein, 2000): (a) it requires a significant increase in the length of the survey, and can depress the overall response rate and accuracy of the survey; (b) it can sometimes yield insufficient differentiation among mean importance ratings, with customers evaluating nearly all of the measures near the top of the scale; or (c) attributes may be rated as important even though they in fact have little influence on overall satisfaction.

That is why recent years have seen the development of derived importance methods based on CSS (Cavana, Corbett, & Lo, 2007; Dell'Olio et al., 2010; Eboli & Mazzulla, 2007; Jen, Tu, & Lu, 2011; Joewono & Kubota, 2007; Nurul-Habib, Kattan, & Islaam, 2011; Pakdil & Aydin, 2007; Weinstein, 2000) and from stated preference surveys (Cirillo et al., 2011; Eboli & Mazzulla, 2008a, 2008b, 2010; Hensher & Prioni, 2002; Hensher, Stopher, & Bullock, 2003; Dell'Olio et al., 2011).

Many authors (Cirillo, Eboli, & Mazzula, 2011; Dell'Olio et al., 2010; Eboli & Mazzulla, 2008b; Eboli & Mazzulla, 2011) point to

* Corresponding author. Tel.: +34 958 24 99 79; fax: +34 958 24 61 38.
E-mail address: jdona@ugr.es (J. de Oña).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.03.037

the heterogeneity of passengers' perception of different aspects of the service. The heterogeneity is due to the qualitative nature of certain aspects that characterise the services, the different attitudes passengers have towards the use of PT, the different ways of viewing aspects of the service, and the social and economic characteristics of passengers and their preferences (Eboli & Mazzulla, 2011). It has even been shown that the same person may change his or her evaluation if they are made to reflect on certain important aspects of the service (Dell'Olio et al., 2010).

This heterogeneity represents a problem for many techniques that intend to measure SQ. Some authors (Dell'Olio et al., 2010) propose specific models after conducting stratified sampling based on the social and demographic characteristics of the passengers (i.e. models for women, for the elderly, according to income level, etc.). This poses two limitations, however: (a) if the samples are small, stratifying is a problem because a data set may be under-represented, and (b), it may be possible to obtain the weight of the variables entered in the model (service characteristics and perceptions), but the weight of the socio-economic variables in the model are impossible to know. Other authors (Eboli & Mazzulla, 2008b; Cirillo et al., 2011) have proposed mixed logit models to introduce such heterogeneity in the models.

However, most of these models have their own model assumptions and pre-defined underlying relationships between dependent and independent variables. If these assumptions are violated, the model could lead to erroneous estimations of the likelihood of quality of service.

The classification and regression tree (CART), a non-parametric model with no pre-defined underlying relationship between the target (dependent) variable and the predictors (independent variables), has been widely employed in business administration, agriculture, industry, and engineering. With the ability to automatically search for the best predictors and the best threshold values for all predictors to classify the target variable, CART has been shown to be a powerful tool, particularly for dealing with prediction and classification problems.

The application of CART in road safety analysis has been advocated by many authors (Abdel-Aty, Keller, & Brady, 2005; Chang & Chen, 2005; Chang & Wang, 2006; Chen & Jovanis, 2000; Council & Stewart, 1996; Kuhnert, Do, & McClure, 2000; Magazzù, Comelli, & Marinoni, 2006; Pande, Abdel-Aty, & Das, 2010; Qin & Han, 2008; Sohn & Shin, 2001; Yan & Radwan, 2006; Yan, Richards, & Su, 2010). The method has also been used for analysing other aspects of traffic engineering: Washington and Wolf (1997) used decision trees to forecast trip generation; Washington, Wolf, and Guensler (1997) considered using trees to determine modal correction factors for motor vehicle emissions; and Hallmark, Guensler, and Fomunung (2002) used trees to identify geometric and operational roadway characteristics that influenced vehicle activity. Although there are some recent decision tree applications for analysing quality of service in others industries (Huang & Hsueh, 2010; Wong & Chung, 2007), the authors have not found any application of CART to analyse quality of service for bus transit operation.

Therefore, the main purpose of this study is to examine whether or not the CART model can effectively identify the key factors affecting bus transit SQ. Another aim of this paper is to verify the hypothesis of Dell'olio et al. (2010) regarding the different evaluation passengers make of service attributes before and after making them reflect on those attributes. This paper uses the data set collected by a CSS conducted in the Granada metropolitan transit system in 2007, which was a non-research oriented survey. However, this paper shows how a non-research oriented survey can be used in researching critical elements and provides an approach to increasing the collaboration between researchers and the industry.

Section 2 presents the methodological approach and a description of the available data. Sections 3 and 4 follow with the model results and discussion. The paper concludes with a summary and directions for future research.

2. Materials and methods

2.1. Methodology

Decision trees are an important data mining technique for the classification and prediction of a class variable. Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts to use the model to predict the classes of objects (Han & Kamber, 2006). When the value of the target variable is discrete, a classification tree is developed, whereas a regression tree is developed for the continuous target variable. The CART method can be used to develop either type of tree. Because this study aims to model the SQ perceived by an individual, a classification tree is developed.

The development of a CART model generally consists of three steps. The first step is tree growing. The principle behind tree growing is to recursively partition the target variable to maximize "purity" in the two child nodes. Then, the data in each child node are more homogeneous than those in the upper parent node. To achieve this, a set of candidate split rules is created, which consists of all possible splits for all variables included in the analysis. These splits are then evaluated and ranked based on the criteria named Gini reduction, to choose amongst the available splits at every non-terminal node. The Gini reduction criteria measure the "worth" of each split in terms of its contribution toward maximizing the homogeneity through the resulting split. If a split results in splitting of one parent node into B branches, the "worth" of that split may be measured as follows:

$$Worth = Impurity(Parent\ node) - \sum_{b=1}^{B} P(b) * Impurity(b) \tag{1}$$

where Impurity (Parent node) denotes the Gini measure for the impurity (i.e., non-homogeneity) of the parent node, and P(b) denotes the proportion of observations in the node assigned to branch b. The impurity measure, Impurity (node), may be defined as follows:

$$\text{Impurity(node)} = 1 - \sum_{i=1}^{I} \left(\frac{\text{number of class i casses}}{\text{all cases in the node}} \right)^2 \tag{2}$$

If a node is 'pure' (i.e., all the cases are of the same class) then this measure (Eq. (2)) will have minimum value, and its value will be higher for less homogeneous nodes. If one considers the definition of "worth" according to Eq. (1), a split resulting in more homogeneous branches (Child nodes) will have more "worth".

While developing a CART, this criterion is applied recursively to the descendents, to achieve Child nodes having maximum worth, which in turn become the parents to successive splits, and so on. The splitting process is continued until there is no (or less than a pre-specified minimum) reduction in impurity and/or the limit for minimum number of observation in a leaf is reached. Following this process a saturated tree is obtained. The saturated tree provides the best fit for the data set which it is constructed from, but overfits the information contained within the data set because this overfitting does not help in accurately classifying another data set. Therefore, to develop a CART model, the data is usually divided into two subsets, one for learning (or training) and the other for testing (or validation). The learning sample is used to split nodes, while the testing sample is used to compare the misclassification. The saturated tree is then constructed from the learning data.

Overly large trees could result in higher misclassification when applied to classify new data sets. Now, to lessen the complexity of the saturated tree that overfits the learning data and create simpler

trees, the tree is "pruned" in the second step. This pruning is performed according to the cost-complexity algorithm, which is based on removing the branches that add little to the predictive value of the tree. After pruning a branch, if the increase in the misclassification cost is sufficiently lower than the decrease in the complexity cost, that branch will be pruned, and a new tree is created. As more and more nodes are pruned away, simpler and simpler trees are the result. At the end of the tree pruning process, the relationship between the misclassification costs and tree complexity in terms of the number of terminal nodes (given in Fig. 1) is obtained.

The last step is to select an optimal tree from the pruned trees. The principle behind selecting the optimal tree is to find a tree with respect to a measure of misclassification cost on the testing dataset (or an independent dataset), so that the information in the learning dataset will not overfit. As shown in Fig. 1, when the tree grows larger and larger, the misclassification cost for the learning data decreases monotonically, indicating that the saturated tree always gives the best fit to the learning data. On the other hand, in the misclassification cost for the testing data, first there is a decrease, and then an increased is observed, after reaching a minimum. Then, the optimal tree is the one that has the least misclassification cost for the test data. More detailed description of CART analysis and its applications can be found in Breiman, Friedman, Olshen, and Stone (1998).

One of the most valuable outcomes provided by CART analysis is the value of the standardized importance of independent variables, which reflects the impact of such predictor variables on the model. The information is obtained for all the independent variables, making it easy to find which ones are the most important.

2.2. Data

The data used in this study come from a CSS conducted in the Metropolitan PT service in the city of Granada (Spain) in 2007. Granada is a medium-sized city in southern Spain with a population of 517,923 in the metropolitan area. A Granada Area Transport Consortium was created in 2003 to coordinate transit bus service management in the Metropolitan Area. The PT service in the metropolitan area is provided by a bus system in which 15 bus companies operate in 18 transport corridors. In 2007, the metropolitan PT system carried more than 10.7 million passengers.

The Transport Consortium conducted the first CSS to evaluate SQ in the Granada Metropolitan PT system in 2007. 858 surveys were conducted at bus stops on the lines in the metropolitan area network. The CSS gathered information on 35 items (see Table 1), which can be classified into five groups: demographic profile, travel behaviour, importance of service attributes, perceived SQ attributes, and global evaluation of SQ. The last group was asked twice during the survey: once at the beginning of the survey (Pre-Evaluation) and again at the end of the questionnaire (Post-Evaluation), when the passengers had been made to reflect on the service's characteristics. Table 2 gives a description of the categories used for the items in the demographic profile and passengers' travel behaviour. A three-point semantic scale (poor, fair and good) was used for measuring the performance perception of SQ attributes and the overall evaluation of SQ (Pre- and Post-Evaluation), and the importance of each SQ attribute was measured with a numeric ten-point scale.

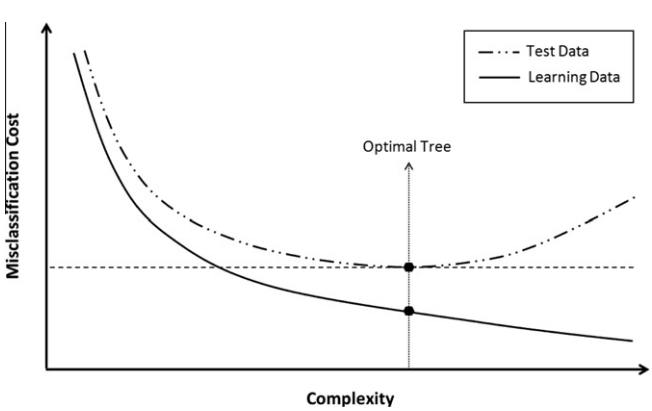

Fig. 1. Relationship between tree complexity and misclassification costs.

Two different models were built to classify the dependent variable (Pre- and Post-Evaluation) and identify the attributes that play a key role in the classification of this variable. To that end, CART decision trees were developed and implemented using 10-fold cross-validation of the sample. The independent variables (21) used were the variables for the customers' demographic profile (3), their travel behaviour (6) and perception of SQ attributes (12). The dependent variable for Model 1 is overall Pre-Evaluation SQ and the dependent variable for Model 2 is overall Post-Evaluation SQ.

The stated importance for each attribute indicated by the respondents was compared with the derived importance obtained from the CART algorithm in the two models built (Pre- and Post-Evaluation).

3. Results

3.1. Decision trees

Fig. 2 shows the CART for the dependent overall Pre-Evaluation SQ variable. The interpretation of the tree is given below. A root node (Node 0) is divided into two child nodes (Node 1 and Node 2). It is used as a splitter the variable that obtains the maximum 'purity' of the two child nodes. In this case, the splitter is Frequency. Node 1 shows the data related to passengers who have a Good or Fair perception of service Frequency. In turn, Node 1 is divided into two terminal nodes or child nodes (Node 3 and Node 4) on the basis of the Punctuality variable. Terminal Node 3 shows that if Punctuality and Frequency are rated as Good or Fair, the overall evaluation of SQ (Pre-Evaluation) is likely to be perceived as GOOD (67.8%). Terminal Node 4 shows that if Frequency of service is rated as Good or Fair and Punctuality is stated as Poor, there is a 45.8% likelihood that the occupant will consider that the global SQ is FAIR.

The passengers who have a Poor perception of service Frequency are on the right branch of the tree. In this case, Node 2 is divided into two terminal nodes (Node 5 and Node 6). Terminal Node 5 indicates that a passenger who travels for a reason other than Occupation, Studies or Doctor, and who rates service Frequency as Poor, will rate SQ as FAIR in 50.8% of cases. If the reason for travelling is Occupation, Studies or Doctor (i.e. compulsory mobility) and Frequency has been rated as Poor (Terminal Node 6), the evaluation of SQ will be POOR in 49.4% of cases.

This first decision tree produced two levels (depth below the root node), 7 nodes and 4 terminal nodes or leaves. A 10-fold cross-validation of the sample was used to give us a precision ratio of the categorisation of the variable class of 59.72%, indicating that the model's precision was acceptable (Wong & Chung, 2007).

The CART built for the overall Post-Evaluation SQ variable produced 5 levels, 23 nodes and 12 terminal nodes (see Fig. 3). In this case, the root node divides into two child nodes after the Punctuality variable. The data of the passengers who have a Good perception of service Punctuality are on the left branch of the tree, giving 6 terminal nodes (8, 10, 15, 16, 17 and 18). All these terminal nodes

Table 1
Customer satisfaction survey's items.

Type	Variable	Symbol
Demographics	Sex	Sex
	Age	Age
	Occupation	Occupation
Travel behaviour	Travel reason	Travelreason
	Use frequency	Usefrequency
	Type of ticket	Ticket
	Private vehicle available	Privatevehicle
	Complementary modes from origin to bus stop	Modesto
	Complementary modes from bus stop to destination	Modesfrom
Importance of service quality attributes	Information	Information
	Punctuality	Punctuality
	Safety on-board	Safety
	Driver courtesy	Courtesy
	Bus interior cleanliness	Cleanliness
	Bus space	Space
	Bus temperature	Temperature
	Accessibility to/from the bus	Accessibility
	Fare	Fare
	Speed	Speed
	Frequency of service	Frequency
	Stops proximity to/from origin/destination	Proximity
Perception of service quality attributes	Information	Information
	Punctuality	Punctuality
	Safety on-board	Safety
	Driver courtesy	Courtesy
	Bus interior cleanliness	Cleanliness
	Bus space	Space
	Bus temperature	Temperature
	Accessibility to/from the bus	Accessibility
	Fare	Fare
	Speed	Speed
	Frequency of service	Frequency
	Stops proximity to/from origin/destination	Proximity
Overall evaluation of quality of service	Overall pre-evaluation of service quality	Pre-evaluation
	Overall post-evaluation of service quality	Post-evaluation

Table 2
Categories of demographic variables, users travel behaviour and quality attributes.

Variable	Categories
Sex	1. Men; 2. Women
Age	1. {18–30}; 2. {31–60}; 3. {>61}
Occupation	1. Worker; 2. Student; 3. Housewife; 4. Other
Travel reason	1. Occupation; 2. Studies; 3. Doctor; 4. Other
Use frequency	1. Sporadic; 2. Occasionally; 3. Very often; 4. Almost everyday
Type of ticket	1. Standard ticket; 2. Consortium Pass; 3. Senior Citizen Pass; 4. Other
Private vehicle available	1. Yes; 2. No
Complementary modes from origin to bus stop	1. Walking; 2. Vehicle
Complementary modes from bus stop to destination	1. Walking; 2. Vehicle

predict that passengers will rate SQ as GOOD or FAIR. This implies that a passenger who rates service Punctuality as Good will give the service a (Post-Evaluation) global evaluation of FAIR or higher. If the same passengers who rated Punctuality as Good also rate Proximity and Safety as Good and Fare is not rated as Poor (Terminal Node 15), it is very likely that the overall evaluation of the service will be GOOD (76.2%). On the other hand, if Punctuality, Speed and Fare are rated as Good, the overall evaluation of the service is likely to be GOOD (Terminal Node 17, 67.9%), even if Proximity is considered Poor or Fair.

The passengers who have a Poor or Fair impression of the Punctuality variable are on the right branch of the tree, where six terminal nodes are obtained (12, 13, 14, 20, 21 and 22). On the basis of the Frequency variable, Node 2 is divided into two child nodes: Node 5 for the passengers who rate Frequency as Good or Fair and Node 6 for all other passengers.

If Punctuality is not Good and Frequency is rated as Poor, passengers will probably not give the service a GOOD overall evaluation. After Node 6, Space is used as a splitter and two terminal nodes are obtained (Nodes 13 and 14). Terminal Node 13 shows that if Punctuality is not Good and Frequency and Space are Poor, the overall evaluation of the service will probably be POOR (62.5%). In the event that Space is not Poor, the global service evaluation will probably be FAIR (68.4%, Terminal Node 14).

After Node 5 (passengers who do not rate Punctuality as Good but do not consider Frequency to be Poor) four terminal nodes are obtained (12, 20, 21 y 22), in which the overall evaluation of the service is FAIR or higher. This implies that if the service provided gives passengers the impression that Frequency is Fair or Good, their global evaluation of the service will not be POOR, even if Punctuality is not Good. Moreover, if the passengers perceive service Fare and Speed as Good and the Reason to travel is Occupation or Other, the overall evaluation of the service will probably be GOOD (69.0%, Terminal Node 21).

The Post-Evaluation CART shows that the model has a global precision value of 62.16%, which indicates that the model is stable and its precision is acceptable. The precision of this model is somewhat higher than the precision of the previous model, which implies that overall Post-Evaluation SQ can be predicted more accurately than overall Pre-Evaluation SQ.

3.2. Decision rules

One of the main advantages of decision trees as opposed to other modelling methods is that they provide effective "If-then" rules that make the model very practical and easy to interpret from the perspective of management by PT operators and managers.

Each decision tree gives as many rules as the existing number of terminal nodes. Table 3 shows the four rules from Decision Tree One (Pre-Evaluation CART), which uses the variables Frequency, Travel reason and Punctuality. One of the rules identifies the conditions that must be given for the overall evaluation of service to have a high likelihood of being considered GOOD (Node 6). In this model, two rules for an overall evaluation of FAIR and one rule for an evaluation of POOR were identified.

Table 4 shows the 12 rules of Decision Tree Two (Post-Evaluation CART) that use the attributes Punctuality, Frequency, Proximity, Space, Fare, Speed, Safety and Travel reason to identify rules that are useful to service managers. It bears mentioning that only one rule was found to imply a high probability that the overall evaluation of service will be POOR (Node 7): If Frequency and Space are perceived as Poor and Punctuality is not Good, the overall evaluation of service is likely to be POOR (62.5%). On the contrary, three rules for GOOD evaluations and eight rules for FAIR evaluations were identified. Finally, it can be seen that the precision values of the rules taken from the Post-Evaluation CART are higher than the ones taken from the Pre-Evaluation CART, with a minimum value of 58.8%.

3.3. Importance of the variables

The CART modelling process has a crucial phase in which the variables that are of key importance in the prediction of the

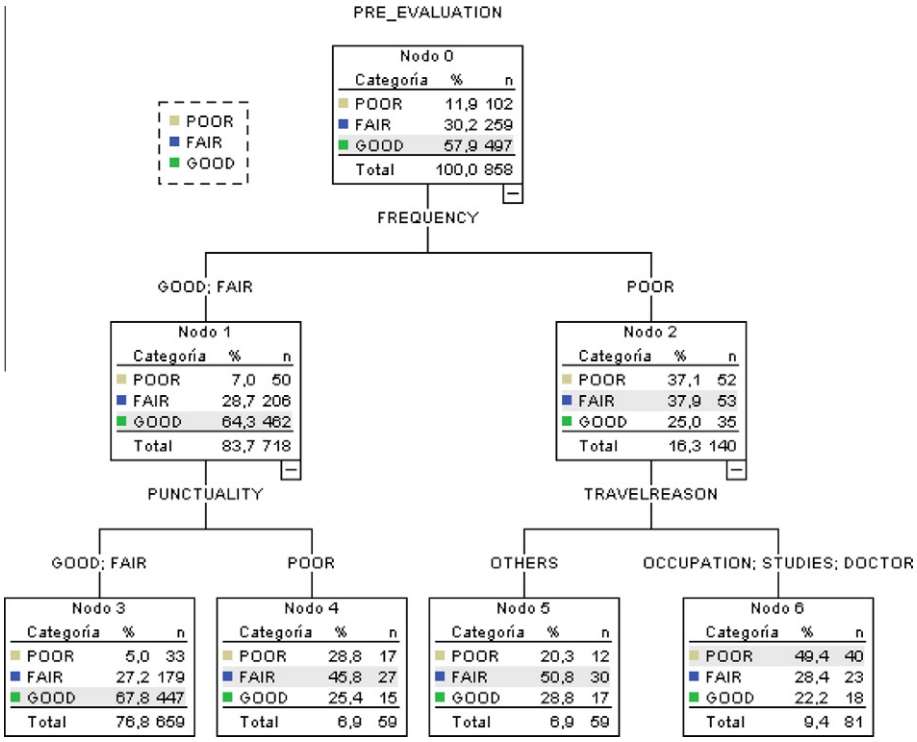

Fig. 2. Pre-Evaluation CART tree.

dependent variable are identified. This is achieved by using the importance index (Kashani & Mohaymany, 2011), of which a standardized form has been used in this paper to reflect the importance of each independent variable on the model.

Table 5 shows the standardized importance of the attributes deduced from each of the two models (Pre- and Post-Evaluation CARTs), and the importance stated by passengers in the CSS. It can be seen that there is very little variation in the evaluations stated by the passengers in the surveys, considering that all the attributes are highly important. The average value of attribute importance in the CSS is concentrated in the 8.5 to 9.5 range (on a scale of 1 to 10). Therefore, their standardized 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).

Analysing the importances derived from the Pre-Evaluation CART, Frequency is the attribute with the most weight, at quite a distance from the other attributes. A priori, this would imply that passengers rate SQ on Frequency alone. Eboli and Mazzulla (2008b), Eboli and Mazzulla (2010) also identified service Frequency as the attribute that had the greatest impact on SQ. Dell'Olio et al. (2010), Dell'Olio et al. (2011) identified Frequency as one of the most important attributes. The Pre-Evaluation CART reveals that Frequency is a key attribute of major impact when passengers have a preliminary idea of how the service operates. Moreover, Frequency serves as the tree's root variable, splitting the passengers that evaluate SQ as GOOD onto the left branch and the passengers that evaluate it as POOR onto the right branch, while the passengers who give a FAIR evaluation are split in either direction. Speed and Punctuality are also attributes that carry considerable weight on SQ in the Pre-Evaluation CART, although at quite a distance from Frequency. This matches the results of other recent studies (Dell'Olio et al., 2010; Eboli & Mazzulla, 2010) in which Punctuality and service Reliability have been identified as one of the most important attributes for passengers.

After the passengers have reflected on the service attributes, however (because they have been asked about them), a higher number of attributes gain weight in the overall perception of quality, whereas the weight of Frequency on the overall evaluation decreases. Table 5 shows that, apart from the attributes considered to be important in the Pre-Evaluation (Frequency, Speed and Punctuality), other attributes such as Proximity, Safety and Fare are also identified as important in the Post-Evaluation CART when the passengers have been made to reflect on them. They can attain standardized importance values that exceed 60.3%.

Therefore, after making the passengers reflect on the variables that can have an impact on their perception of the evaluation of PT, the importance of Frequency diminishes and the role of other service attributes in the overall perception of SQ increases, such as Proximity, Safety and Fare. These outcomes match those obtained by Dell'Olio et al. (2010). They compared an overall evaluation of SQ before and after making passengers reflect on the importance of certain fundamental system variables which they may not have previously considered. Ordered probit models were used. In their first model (pre-evaluation), Dell'Olio et al identified Reliability of Service (RS) and Waiting Time (WT) (which could be considered equivalent to Punctuality and Frequency in this paper) as the two variables that had the greatest impact on passenger's overall evaluation of SQ. Likewise, in their second model (post-evaluation), the importance of Frequency diminished as the importance of other attributes increased.

4. Discussion

The CART model provides good overall predictions for both decision trees (Pre- and Post-Evaluation) with a model precision value of 59.72% and 62.16%, respectively. This demonstrates that CART analysis is an appropriate methodology for analysing PT quality of service.

This approach has a number of benefits compared to other widely used parametric models. The main advantage of the CART

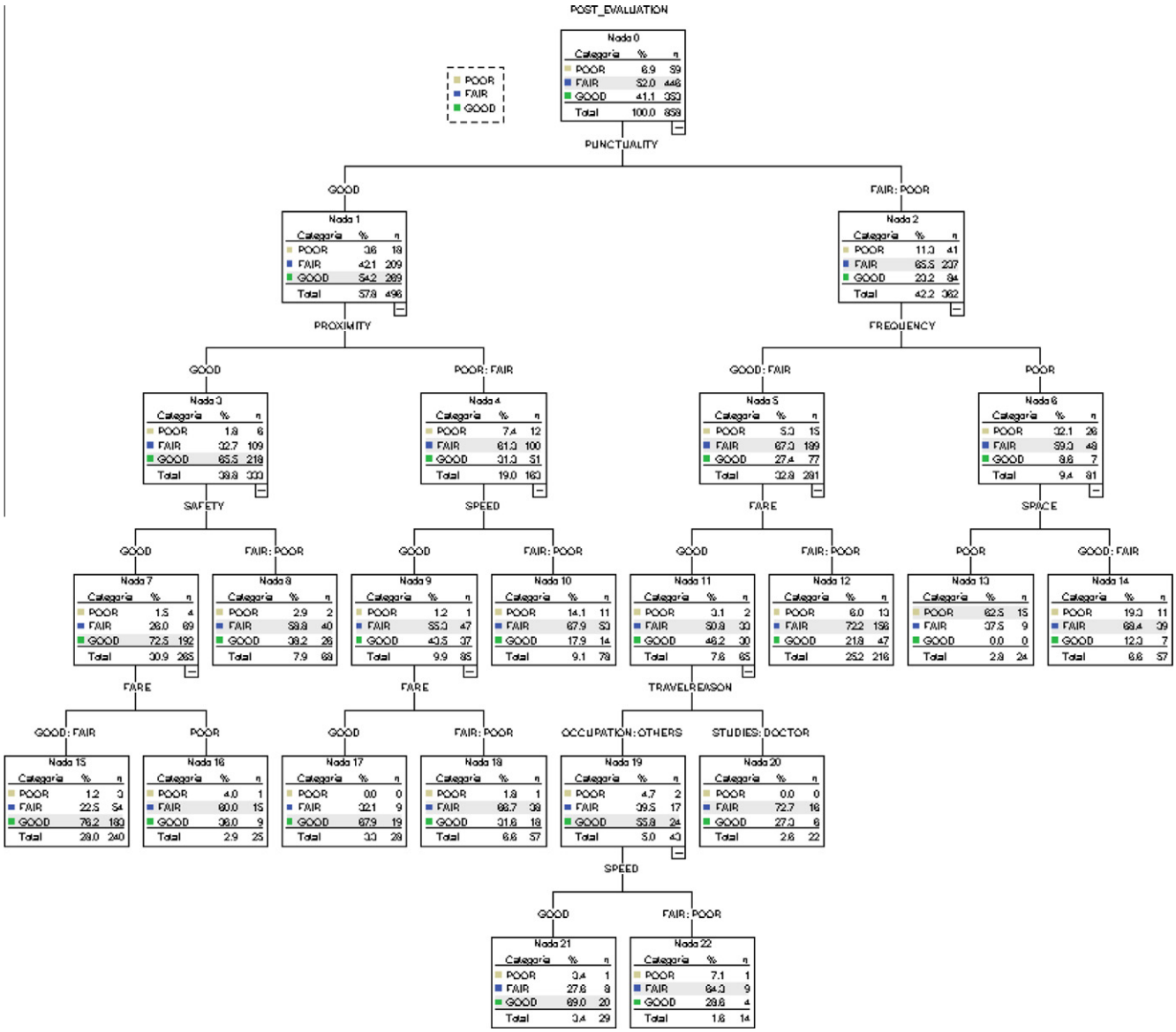

Fig. 3. Post-Evaluation CART tree.

Table 3
Rules for overall Pre-Evaluation of service quality.

Node	Rule		Accuracy rate (%)
	If	Then	
3	"Frequency" and "Punctuality" are rated as Good or Fair	Service is rated as "Good"	67.8
4	"Frequency" is rated as Good or Fair and the perception of "Punctuality" is Poor	Service is rated as "Fair"	45.8
5	"Frequency" is rated as Poor, and the "Travel Reason" is Others	Service is rated as "Fair"	50.8
6	"Frequency" is rated as Poor, and the "Travel Reason" is different to Others	Service is rated as "Poor"	49.4

model is that the outcomes of the analysis are easy to understand. This is due to the graphic representation afforded by the results and the creation of a framework of "If-then" rules. Moreover, the CART analysis allows a great many explanatory variables to be processed and the most important variables are easy to find. In this case, variables of different natures were used, some of which pertained to passengers' social and demographic characteristics and travel behaviour, and others to service features. This afforded knowledge of the weight of all the variables in the model. Another advantage of the CART analysis is that it does not need to specify a functional form. In regression analysis, if the model is misspecified, the estimated relationship between dependent variable and independent variables as well as model predictions will be erroneous. Another advantage is that CART analysis can effectively handle multi-collinearity problems, which is one of the major drawbacks in regression models. SQ measured from the passengers' perspective tends to show this type of relationships owing to the fact that passengers group items together even when satisfaction surveys

Table 4
Rules for overall Post-Evaluation of service quality.

Node	Rule		Accuracy rate (%)
	If	Then	
15	"Punctuality" is rated as Good, "Proximity" and "Safety" are rated as Good, and there is a perception of "Fare" as Good or Fair	Service is rated as "Good"	76.2
17	"Punctuality" is rated as Good, "Proximity" is rated as Poor or Fair, and there is a perception of "Speed" and "Fare" as Good	Service is rated as "Good"	67.9
21	"Punctuality" is rated as Poor or Fair, "Frequency" is not rated as Poor, the perception of "Fare" is Good, the "Travel Reason" is Occupation or Others, and "Speed" is rated as Good.	Service is rated as "Good"	69.0
8	"Punctuality" and "Proximity" are rated as Good but the perception of "Safety" is other than Good.	Service is rated as "Fair"	58.8
10	"Punctuality" is rated as Good, and "Proximity" and "Speed" are rated as Poor of Fair.	Service is rated as "Fair"	67.9
12	"Punctuality" is rated as Poor or Fair, "Frequency" is not rated as Poor, and the perception of "Fare" is Poor or Fair	Service is rated as "Fair"	72.2
14	"Punctuality" is rated as Poor or Fair, "Frequency" is rated as Poor, and the perception of "Space" is Good or Fair	Service is rated as "Fair"	68.4
16	"Punctuality", "Proximity" and "Safety" are rated as Good, and there is a perception of "Fare" as Poor	Service is rated as "Fair"	60.0
18	"Punctuality" is rated as Good, "Proximity" is rated as Poor or Fair, there is a perception of "Speed" as Good, and "Fare" is rated as Poor or Fair	Service is rated as "Fair"	66.7
20	"Punctuality" is rated as Poor or Fair, "Frequency" is not rated as Poor, the perception of "Fare" is Good, and the "Travel Reason" is Studies or Doctor	Service is rated as "Fair"	72.7
22	"Punctuality" is rated as Poor or Fair, "Frequency" is rated as other than Poor, the perception of "Fare" is Good, the "Travel Reason" is Occupation or Others, and "Speed" is rated as Poor or Fair	Service is rated as "Fair"	64.3
13	"Punctuality" is rated as Poor or Fair, "Frequency" is rated as Poor, and the perception of "Space" is Poor	Service is rated as "Poor"	62.5

Table 5
Stated and derived attributes' importance.

Stated importance		Derived importance using cart			
		Pre-Evaluation		Post-Evaluation	
Punctuality	100%	Frequency	100%	Proximity	100%
Frequency	98.9%	Speed	55.2%	Speed	78.8%
Safety	98.3%	Punctuality	49.3%	Safety	72.3%
Cleanliness	96.9%	Travelreason	15.9%	Frequency	68.4%
Accessibility	96.8%	Usefrequency	11.6%	Fare	64.1%
Courtesy	95.7%	Ticket	6.4%	Punctuality	60.3%
Fare	95.5%	Proximity	5.4%	Space	46.7%
Speed	95.5%	Age	3.5%	Courtesy	42.0%
Temperature	95.4%	Occupation	0.2%	Temperature	38.5%
Proximity	95.0%	Safety	0	Information	28.3%
Space	94.7%	Cleanliness	0	Travelreason	20.4%
Information	94.1%	Accessibility	0	Accessibility	17.6%
		Courtesy	0	Cleanliness	11.9%
		Fare	0	Sex	5.4%
		Temperature	0	Occupation	4.6%
		Space	0	Usefrecuency	2.2%
		Information	0	Modesfrom	1.6%
		Sex	0	Age	0.9%
		Modesfrom	0	Ticket	0.8%
		Privatevehicle	0	Privatevehicle	0.7%

are asking about completely different aspects (such as passengers' evaluation of service punctuality with the evaluation of speed, for instance). When regression analysis is applied and serious correlations between independent variables exist, not only will the variability of estimated coefficients be inflated, but the interpretation of relationship between independent variables and dependent variable will also become difficult. In contrast, when CART analysis is applied, correlation problems between independent variables would not be a great concern.

Not all are benefits, however. The classification tree models are generally "unstable" because building the trees is based on their seed number, which is random, and therefore different trees could be obtained and the results might vary. This is the reason why tree models are often used only to identify important variables and other modelling techniques are used to develop final models.

5. Conclusions

This paper presents a new approach to evaluating SQ in PT systems. Two non-parametric tree-based (CART) models were proposed to study which variables played a more important role in the perception of SQ by PT passengers in the Metropolitan Area of Granada. Model 1 (Pre-Evaluation CART) permitted the identification of the variables which, a priori, are more important in passengers' perception of SQ. Model 2 (Post-Evaluation CART) shows that when the passengers reflected on the services characteristics, other variables gained importance and even surpassed the former ones.

With the analysis of the two models it is found that passengers' perceived quality of service is practically limited to Frequency, Speed and Punctuality in the preliminary evaluation. Once they are made to reflect on other aspects that define the service, however, of other quality-related attributes gain in importance, such as Proximity to the bus stop, Safety on board and the Fare, while the impact of the previous three attributes diminishes (Frequency, Speed and Punctuality).

The data used in this paper come from a CSS that was not intended for this type of research. A rather simple statistical frequency analysis was the main target, in order to evaluate transit SQ. However, application of an advanced modelling technique proved that this kind of data, which was collected for a different purpose, can be used to reveal very interesting details for managers, and government and PT operators, since it gives them valuable information on which to base effective decision-making to promote the use of PT.

These measures can be used to enhance service attributes to which passengers are unconsciously more sensitive in their preliminary evaluation, or to enhance the service attributes that are more important in the second evaluation. This last approach, however, would not make sense unless it is accompanied at the same time by a publicity campaign highlighting the service functions that passengers did not pay attention to at first.

This study demonstrates that CART is a good alternative for analysing PT quality of service. In terms of future work, it would be interesting to compare the results between CART model and

statistical models. As discussed previously, statistical models based on CSS or based on stated preference surveys have been widely employed to analyse the attributes that most influence PT quality of service. By comparing the identified attributes between the CART and statistical models, it can provide valuable insights into the underlying relationship between attributes and the quality of service. However, comparisons between parametric models and non-parametric tree-based models should be made with caution, because the tree-based models are generally "unstable". In addition, if the comparisons of prediction accuracy between parametric models and tree-based models are intended, bagging procedures might be helpful to get a reliable prediction from tree-based models.

Acknowledgements

This study is sponsored by the Consejería de Innovación, Ciencia y Economía of the Junta de Andalucía (Spain) through the Excellence Research Project denominated "Q-METROBUS-Quality of service indicator for METROpolitan public BUS transport services". The authors also acknowledge the Granada Consorcio de Transportes for making the data set available for this study.

References

Abdel-Aty, M., Keller, J., & Brady, P. A. (2005). Analysis of types of crashes at signalized intersections by using complete crash data and tree-based regression. *Transportation Research Record, 1908*, 37–45.

Berry, L.L., Zeithaml, V.A., & Parasuraman, A. (1990). Five imperatives for improving service quality. Sloan Management Review Summer 9-38.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1998). *Classification and regression trees*. Chapman & Hall/CRC.

Cavana, R. Y., Corbett, L. M., & Lo, Y. L. (2007). Developing zones of tolerance for managing passenger rail service quality. *International Journal of Quality & Reliability Management, 24*(1), 7–31.

Chang, L., & Chen, W.-H. (2005). Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research, 36*(4), 365–375.

Chang, L., & Wang, H.-. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis and Prevention, 38*(5), 1019–1027.

Chen, W.-, & Jovanis, P. P. (2000). Method for identifying factors contributing to driver-injury severity in traffic crashes. *Transportation Research Record, 1717*, 1–9.

Cirillo, C., Eboli, L., & Mazzulla, G. (2011). On the asymmetric user perception of transit service quality. *International Journal of Sustainable Transportation, 5*, 216–232.

Council, F. M., & Stewart, J. R. (1996). Severity indexes for roadside objects. *Transportation Research Record, 1528*, 87–96.

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

Dell'Olio, L., Ibeas, A., & Cecín, P. (2011). The quality of service desired by public transport users. *Transport Policy, 18*(1), 217–227.

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. (2008a). A stated preference experiment for measuring service quality in public transport. *Transportation Planning and Technology, 31*(5), 509–523.

Eboli, L., & Mazzulla, G. (2008b). Willingness-to-pay of public transport users for improvement in service quality. *European Transport, 38*, 107–118.

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

Eboli, L., & Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy, 18*, 172–181.

Hallmark, S., Guensler, R., & Fomunung, I. (2002). Characterizing on-road variables that affect passenger vehicle modal operation. *Transportation Research Part D, 7*(2), 81–98.

Han, J., & Kamber, M. (2006). Data mining: Concepts and techniques. *Morgan Kaufmann*.

Hensher, D. A., & Prioni, P. (2002). A service quality index for area-wide contract performance assessment. *Journal of Transport Economics and Policy, 36*(1), 9–113.

Hensher, D. A., Stopher, 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.

Huang, C., & Hsueh, S. (2010). Customer behavior and decision making in the refurbishment industry – A data mining approach. *Journal of Civil Engineering and Management, 16*(1), 75–84.

Jen, W., Tu, R., & Lu, T. (2011). Managing passenger behavioral intention: An integrated framework for service quality, satisfaction, perceived value, and switching barriers. *Transportation, 38*, 321–342.

Joewono, T. B., & Kubota, H. (2007). User perception of private paratransit operation in Indonesia. *Journal of Public Transportation, 10*(4), 99–118.

Kashani, A. T., & Mohaymany, A. S. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Safety Science, 49*, 1314–1320.

Kuhnert, P. M., Do, K.-., & McClure, R. (2000). Combining non-parametric models with logistic regression: An application to motor vehicle injury data. *Computational Statistics and Data Analysis, 34*(3), 371–386.

Magazzù, D., Comelli, M., & Marinoni, A. (2006). Are car drivers holding a motorcycle licence less responsible for motorcycle–Car crash occurrence?: A non-parametric approach. *Accident Analysis and Prevention, 38*(2), 365–370.

Nurul-Habib, K. M., Kattan, L., & Islaam, T. (2011). Model of personal attitudes towards transit service quality. *Journal of Advanced Transportation, 45*, 271–285.

Pakdil, F., & Aydin, Ö. (2007). Expectations and perceptions in airline services: An analysis using weighted SERVQUAL scores. *Journal of Air Transport Management, 13*, 229–237.

Pande, A., Abdel-Aty, M., & Das, A. (2010). A classification tree based modeling approach for segment related crashes on multilane highways. *Journal of Safety Research, 41*(5), 391–397.

Qin, X., & Han, J. (2008). Variable selection issues in tree-based regression models. *Transportation Research Record, 2061*, 30–38.

Sohn, S. Y., & Shin, H. (2001). Pattern recognition for road traffic accident severity in Korea. *Ergonomics, 44*(1), 107–117.

Transportation Research Board. (1999). A Handbook for Measuring Customer Satisfaction and Service Quality, TCRP Report 47.

Washington, S., & Wolf, J. (1997). Hierarchical tree-based versus ordinary least squares linear regression models: Theory and example applied to trip generation. *Transportation Research Record, 1581*, 82–88.

Washington, S., Wolf, J., & Guensler, R. (1997). Binary recursive partitioning method for modeling hot-stabilized emissions from motor vehicles. *Transportation Research Record, 1587*, 96–105.

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

Wong, J., & Chung, P. (2007). Managing valuable Taiwanese airline passengers using knowledge discovery in database techniques. *Journal of Air Transport Management, 13*(6), 362–370.

Yan, X., & Radwan, E. (2006). Analyses of rear-end crashes based on classification tree models. *Traffic Injury Prevention, 7*(3), 276–282.

Yan, X., Richards, S., & Su, X. (2010). Using hierarchical tree-based regression model to predict train–vehicle crashes at passive highway-rail grade crossings. *Accident Analysis and Prevention, 42*(1), 64–74.