Transportation Research Part D 30 (2014) 32–37

Contents lists available at ScienceDirect

Transportation Research Part D

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

Bus line classification using neural networks

Felipe Jiménez *, Francisco Serradilla, Alfonso Román, José Eugenio Naranjo

Technical University of Madrid (UPM), University Institute for Automobile Research (INSIA), Campus Sur UPM, Carretera de Valencia, km. 7, 28031 Madrid, Spain

ARTICLE INFO

Keywords:
Cluster
Urban buses
Neural network
Bootstrap method

ABSTRACT

Grouping urban bus routes is necessary when there are evidences of significant differences among them. In Jiménez et al. (2013), a reduced sample of routes was grouped into clusters utilizing kinematic measured data. As a further step, in this paper, the remaining urban bus routes of a city, for which no kinematic measurements are available, are classified. For such purpose we use macroscopic geographical and functional variables to describe each route, while the clustering process is performed by means of a neural network. Limitations caused by reduced training samples are solved using the bootstrap method.

Introduction

Representing the complete vehicle operation of an urban bus transit system using a single drive cycle may be an excessively simplistic approach and many details concerning differences among lines may remain unobserved (Jiménez et al., 2013). Therefore, several methodologies to group routes into clusters of lines with homogeneous characteristics are proposed. This classification could provide relevant information to the bus fleet company in order to organize its vehicles in the most appropriate way, for example, to save fuel or reduce emissions considering each line's characteristics.

Clustering (Jain et al., 1999) or classification (Beniwal and Arora, 2012) algorithms differ because grouping into clusters is a non-supervised process; while a classification procedure requires that groups are predetermined and their properties must be previously identified.

These techniques are frequently used in diverse areas related with the transport industry. In addition to studies on peer transportation companies (Giuliano, 1981; Fielding et al., 1985), or road classifications according to traffic conditions, traffic characteristics and fleet compositions (Chen et al., 2008); clustering is also used for developing driving cycles or during the categorization of routes. For instance, Fotouhi and Montazeri (2013) use the centroid (k-means) method to group a sample of microcycles that represent traffic patterns in Tehran. André and Villanova (2004) study the bus network in Paris using two methods to evaluate and classify the lines from an environmental perspective: the first approach uses statistics related with route characteristics, travel time, commercial speed, annual statistical data, the irregularity of travel and information on the problems encountered; while the alternative procedure considers aspects related with the socioeconomic peculiarities of each route's setting. In André et al. (2005), the bus routes are characterized by linking analyses of bus operating conditions and urban characteristics collected and managed using a Geographic Information System (GIS).

Due to the difficulty, the high costs required, and the time consumed in the acquisition of detailed data (for example, speed profiles) for the complete operation, most studies perform the route classification process without pre-established groups and generally based on macroscopic variables. Subsequently, this classification is utilized as a guideline in order to select a limited sample of routes in which to measure the kinematic variables.

* Corresponding author. Tel.: +34 91 336 53 17; fax: +34 913365302.
E-mail address: felipe.jimenez@upm.es (F. Jiménez).

http://dx.doi.org/10.1016/j.trd.2014.05.008
1361-9209/

This article proposes a different approach by assigning a characteristic operational cycle to each line, considering that speed cycle information is a more reliable representation of the route's operation and, therefore, contains additional information that may be used during the clustering process. In this case, the preliminary cluster organization based on the kinematic data contained in a limited sample of routes has already been structured (Jiménez et al., 2013), and consequently, this paper describes the methodology utilized to allocate the remaining lines to these clusters using macroscopic variables that are available for the complete set of lines, since kinematic variables have only been gathered for the sampled routes. For this purpose, a neural network is used in order to process the macroscopic variables as input data and provide output values with the appropriate cluster recommendation for each line.

Methodology

Jiménez et al. (2013) includes a proposal of algorithms that may be utilized to classify a set of urban bus routes using the kinematic data related with the operation. However, collecting this information for a large amount of lines is usually a costly and time-consuming process, and therefore, a more practical approach consists in identifying macroscopic variables that influence the operation of the vehicles. Consequently, a method based on the analysis of macroscopic variables is required in order to assign lines to the set of clusters previously obtained from kinematic data.

Macroscopic variables

Table 1 shows the list of macroscopic variables used in this study. These variables include features related with the operation of the vehicles, the characteristics of the routes, the infrastructure on which the lines are set and the areas of the city being serviced. The metropolitan surface is segmented according to two different criteria: a broader perspective which entails zones that are delimited by the main beltways of the city, and a more detailed partition based on district and neighborhood limits. It should be noted that some of the macroscopic variables can be adapted depending on the city's geography (for instance, defining more areas in variables 15 and 16), but this circumstance does not influence the remaining application of the methodology.

Neural network

Artificial Neural Networks are recently developed computational models that are very useful in classification and clustering processes (Bishop, 1995; Hagan et al., 1996).

Table 1
Definition of the macroscopic variables.

N°	Variable	Units
Bus operation		
MV 1	Average speed	km/h
MV 2	Bus frequency (time between 2 consecutive bus arrivals)	minutes
MV 3	Number of passengers index	dimensionless
Route characteristics		
MV 4	Number of bus stops per km	dimensionless
MV 5	Number of streets per km	dimensionless
	Coincidences with other means of transport	
MV 6	Coincidences with Metro stations per km	dimensionless
MV 7	Coincidences with Metro lines per km	dimensionless
Bus route infrastructure		
Bus lane		
MV 8	% of the route distance driven along conventional bus lane	%
MV 9	% of the route distance driven along bus lane with barrier	%
MV 10	% of the route distance driven along independent bus lane	%
MV 11	% of the route distance driven along bus lane	%
City areas		
MV 12	Streets fiscal index (index between 1 – highest index and 9 – lowest index)	dimensionless
Distances to city center		
MV 13	Distance of the nearest route terminal to the city center	m
MV 14	Distance of the furthest route terminal to the city center	m
MV 15	Proximity index to the city center (index between 1 and 3 depending on the city area)	dimensionless
MV 16	Distance index to the city center (index between 1 and 3 depending on the city area)	dimensionless
Circulation along city areas limit streets		
MV 17	% of the route distance driven along district limit streets	%
MV 18	% of the route distance driven along neighborhood limit streets	%
MV 19	% of the route distance driven along district or neighborhood limit streets	%