Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 27 (2017) 1050–1057

Transportation Research Procedia

www.elsevier.com/locate/procedia

20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-6 September 2017, Budapest, Hungary

Land-use zone estimation in public transport planning with data mining

Viktor Nagy[a]*, Balázs Horváth[b], Richárd Horváth[c]

[a,b,c] *Széchenyi István University, Egyetem tér 1, Győr, 9026, Hungary*

Abstract

Nowadays, data sets are spreading continually, generated by different devices and systems. The modern GPS based tracking systems and the electronic tickets are producing lots of data, and we could use them, for improving the service level. These data are processable with the modern devices and methods, and we can use them for obtaining information. Thanks to the spread of data mining, these tools are not appearing only in marketing research, but also in the most various kind of scientific areas and they are advertising a new scientific revolution. Although the importance of these data sources is essential it is not widespread in transport planning except in some specific areas.

The smart card systems store the number of boarding passengers and in some cases also the alighting values. From the passengers' boarding and alighting information in a stop point we can create a time series, which shows the behavior type of the given stop points presented on graphic curves. With the help of different clustering and classification processes, these curves can be turned into groups and we can observe these groups of stop points which are defining separated zones. This is the basic step in transport modelling and the zones were determined by manual methods usually.

In this paper we examine clustering and classification methods compared to each other and check the usability of different distance measurement techniques. This paper shows the usage of these methods in public transportation and presents the background of this kind of zone distribution technic.

Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.

Keywords: data mining; zone estimation; land-usage

* Viktor Nagy. Tel.: +3696-613-561; fax: +3696-613-561.
E-mail address: nagy.viktor@sze.hu

2352-1465
Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.
10.1016/j.trpro.2017.12.145

1. Introduction

The modern GPS based tracking systems and the electronic tickets are producing lots of data, and we could use them. The importance of these data sources is essential it is not widespread in transport planning but there are researchers who work in this field in some specific areas. For example extracting origin information from AFC data (Ma et al., 2012), estimating the impact of a fare change (Wang et al. 2015), examining the network flows and mobility patterns (Zhong et al., 2015) or creating OD matrices based on smart card data (Munizaga and Palma, 2012). The cognition of the traveling behavior is an essential tool of the maintenance and correction of the service level of the public transport. The forecast of travel demand is essential in transport planning (Horváth, 2012).

The demands are continuously changing. We can talk about daily, weekly and seasonal fluctuation. The traffic demands evolve because of the different functions of areas, so if we know the characteristics of an area, we can deduce the demands too. This could be true vice versa. If we know the place and the time of the demand, we would know the type of land-use of the given part of the city. There is a correlation between the temporal patterns of passengers and land-usage (Zhong et al., 2015). The main conclusion of this paper confirms this statement. In other words, how can we create land-use zones with up to date tools using the passengers' check-in check out data as the base? In our case the passengers' data were known from a passenger counting, which was executed in the city of Győr in 2012. The database contains all of the stop points in the city, the boarding and alighting information and we can also extract the time of these. The main goal of the research is to explore the area's behavior with data mining techniques and compare to each other the results of clustering and classification.

The method assigns boarding and alighting data per hour to every single stop point. From the passengers' boarding and alighting information in a stop point, we can create time series, which show the behavior type of the given stop point, presented on a graphic curve. Based on these time series we are able to deduce the characteristics of the stop point's environment, since the different land usage yields dissimilar stop usage with well-defined peak hours.

In the case of clustering, the method compares the stop points to each other and adds a dissimilarity value based on the boarding/alighting data. With the distance measurement of time series it is possible to define that how similar are two selected stop points and their environment to each other. For such kind of distance measurement several methods are known. With the help of clustering process and the usage of R data miner software (R Development Core Team, 2008), these distance data can be turned into groups and we can observe these groups of stop points. In the case of classification the method compares the stop points to three predefined classes.

In a previous paper we took into consideration only the boarding numbers (Horváth and Nagy, 2016). In that research, we wanted to create the city's public transport zones based on the number of boarding passengers. The results were promising, but also contained lots of inaccuracy. Because of the complexity we figured out that we have to split the task into more steps for the reliable zone estimation, where the first step is to estimate the land-use of the different parts of the city. A previous study (Nagy and Horváth, 2017) examined the usage of boarding and alighting data with classification techniques. The results were not as good as in the case of using only alighting data and clustering (Nagy, 2016). Now we examine that which method is better for this kind of task, the clustering or the classification or both methods are proper and the inaccuracy is because of the boarding data. Because of that we treat the boarding numbers as irrelevant again and use only the alighting numbers.

The stop point groups define separated zones, what is a basic step in transport planning and they usually were produced with manual methods so far. So basically the method could help in the creation of an origin-destination matrix in an unused way, with creating zones automatically or revising the manual work.

2. Data description

The given data were counted in an urban area. The city has a main role in the region with well-developed industry, the population is about 130.000 and its territory is 174 km^2. The city was served with 451 stop points. The modal split in the city was about 30% for public transport, but the majority of citizen traveled by car. The database contains the lines, the directions, the vehicle types and the capacity, the schedule based and the real departure and arrival times. It contains furthermore the distance of the stop points, the arrival time and also the boarding and alighting numbers of passengers.