Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 109C (2017) 506–512

Procedia Computer Science

www.elsevier.com/locate/procedia

The 8th International Conference on Ambient Systems, Networks and Technologies (ANT 2017)

Hidden content of passenger data in public transport

Viktor Nagy[a], Balázs Horváth[b]

[a]nagyvi@sze.hu, Széchenyi István University, Egyetem tér 1., Győr 9024, Hungary
[b]hbalazs@sze.hu, Széchenyi István University, Egyetem tér 1., Győr 9024, Hungary

Abstract

Nowadays, data sets are spreading continually, generated by different devices and systems. In most cases, these data are stored, and the service providers don't use the information they contain, what even more they delete these data to save space. However, these data are processable with the modern devices and methods, and we can use them for obtaining information. This paper presents a possible application of the digital raw materials, taking the public transport passengers boarding and alighting information as a base. Based on these, 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. With the help of distance measurement and classification techniques it is possible to define how similar are the stop points to predefined patterns and we can create stop point groups which define separated zones. The paper shows a usage method of distance measurement methods and classification in public transportation and presents the background of this kind of land-use zone distribution technic.

Keywords: classification; land-use; time series, transport planning

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 examining dwell times[1], extracting origin information from AFC data[2], estimating the impact of a fare change[3], examining the network flows and mobility patterns[4] or creating OD matrices based on smart card data[5]. The cognition of the traveling behaviour is an essential

*Viktor Nagy. Tel.: +36 96 613 561; fax: +36 96 613 561.
E-mail adress: nagy.viktor@sze.hu

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.05.320

tool of the maintenance and correction of the service level of the public transport. The forecast of travel demand is essential in transport planning[6].

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-useage[4]. This paper tries to examine and confirm this. In other words, how can we create land-use zones with modern 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 by Universitas-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.

In the first step, the method assigns boarding and alighting data per hour to every single stop point and creates time series. 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 second step, the method compares the stop points to predefined classes and adds a dissimilarity value based on the boarding and alighting data. With the distance measurement of time series it is possible to define that what kind of land-use pattern predominate in the selected stop point and its environment. The next step is the group creation, where we create three kinds of groups, what we split in the last step to five groups. These 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.

In previous research we used clustering techniques for that problem. In a paper we took into consideration only the boarding numbers[7]. 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 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. In an other study[8] we treated the boarding numbers as irrelevant and used only the alighting numbers as base. The method created 3 types of zones (residential, service and industrial areas) with more than 80% accuracy. In this paper we wanted to specify our results and create 5 land-use types (family house area, housing estate area, shopping area, city center and industrial area) and use classification instead of clustering.

2. Data description

The given data were counted in an urban area, Győr, Hungary in 2012. The city has a main role in the region with well-developed industry, the population is about 130.000 and its territory is 174 km2. The city was served by 42 bus lines with 451 stop points (no metro or tram is available in the region). 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. The examined data was cleaned; the total number of boarding passengers in the observed day was 75284. In bigger cities the proportion of public transport users is also bigger, so this value in a city with 2 million inhabitants could be easily 3 million boardings per day.

We did not need a precision in minutes. We assign the alighting information to a stop point, we summarize the alighting numbers per hours, so we get one row (a time series) for each stop point. Thus the cells of the received table contain the sum of one hour boarding and alighting passengers of a given stop point. The total value of passengers using one stop point is obviously different. If we represent them in a diagram, we can see that similar kind of stop points may seem different and we cannot compare them. For the proper comparison we used normalization, so we did not use the exact numbers, but the proportion of the given hour at the stop point, so the values become comparable, as it can be seen in Figure 1.