Kalpa Publications in Computing

Volume 14, 2023, Pages 7–9

Proceedings of V XoveTIC Conference. XoveTIC 2022

kalpa Computing

SIGTRANS: Geographical Information System for the analysis and management of public TRANSport*

Pablo Gutiérrez-Asorey[1], Nieves R. Brisaboa[1], and Tirso V. Rodeiro[1]

Universidade da Coruña, Centro de Investigación CITIC, Database Lab, A Coruña, Spain
{pablo.gutierrez,nieves.brisaboa,
tirso.varela}@udc.es

Abstract

Our project is aimed at the creation of SIGTRANS, a tool focused on addressing the need of efficiently storing and analyzing the vast amount of data related to the use of public transport networks. This is a highly relevant research topic given the changes urban mobility is experiencing, including but not limited to those motivated by climate change. We will provide transport authorities and operators with a system, combining the use of GIS technologies, compact data structures and advanced algorithms, to facilitate the exploration and exploitation of the available data. This data refers to both the *offer* in terms of infrastructure and mobility services available for the citizens, and the *demand* (that is, the use citizens are expecting out of these services). The analysis of this data will then serve as the foundation for further improvements to public transport services.

1 Introduction

Our objective is to create SIGTRANS, a novel tool for the analysis and management of data referring to the use of public transport network in urban and/or metropolitan areas. Nowadays, the widespread integration of traveler cards means that a lot of data referring to individual boardings into any means of public transport are being generated daily. This enables public transport authorities to analyze the mobility of users within transport networks. However the vast amount of data that results from the use of these traveler cards means that its storage and exploitation should be addressed as a Big Data problem. Therefore, SIGTRANS will not only facilitate the visualization and exploitation of mobility data on public transport networks, but also offer an efficient, both space and access time wise, solution for its storage.

Our proposed solution will integrate: 1) a representation of the data tailored for the analysis of the movements of users within the transport network on an individualized level, 2) the efficient storage and exploitation of the data thanks to a Compact Data Structure (CDS) and 3) user interfaces based on GIS technologies to visualize and query the data stored in the CDS.

*This work was supported by the CITIC research center funded by Xunta de Galicia, FEDER Galicia 2014-2020 80%, SXU 20% [ED431G 2019/01 (CSI)]; MCIN/ AEI/10.13039/501100011033 [PID2020-114635RB-I00], [PDC2021-120917-C21]; by GAIN/Xunta de Galicia [ED431C 2021/53] GRC and by Xunta de Galicia [ED481A/2021-183]

For this project, the Regional Consortium of Transportation of Madrid, Spain[1], provided us with data referring to all traveler card transactions registered in the city of Madrid, in the years 2019 and 2021.[2], for subway, suburban train, trolley car, urban bus, and interurban bus. The resulting dataset amounted to roughly 140 GiB per year. For reference, the month of January of 2019 includes 151,094,796 records.

2 Representation of trips and alighting estimation

For the analysis of user movements on the public transport network, it is necessary to organize the data of the traveler cards as trips. A trip is the complete journey of an user from one point to another within the network. Any trip can be think off as a sequence of one or more coherent trip-stages (both in terms of space and time continuity) either on the same or different means of transport, and each with its corresponding boarding and alighting stops. Note that, as most stops, regardless of the means of transport, do not validate traveler cards on alighting, the majority of the data available to us refers only to boardings.

While this makes it difficult to organize the data as trips, thanks to the spatial continuity expected of consecutive trip-stages on the same trip, as well as the symmetry on the movements of travellers (the last trip of any given day for most people ends on the same area their first trip started from), it is possible to design algorithms to infer alighting stops based on the boarding data, including the last alighting stop of every trip.

Works such as [1] show the potential of exploiting these continuity and symmetry principles, with [2] further improving its results by also integrating the use of machine learning techniques. We have adapted these ideas to develop our own algorithm, its general flow shown on Figure 1. We can also validate our results thanks to a subset of trip-stages for which we have information about their alighting stop (around 9% of our records correspond to an alighting).

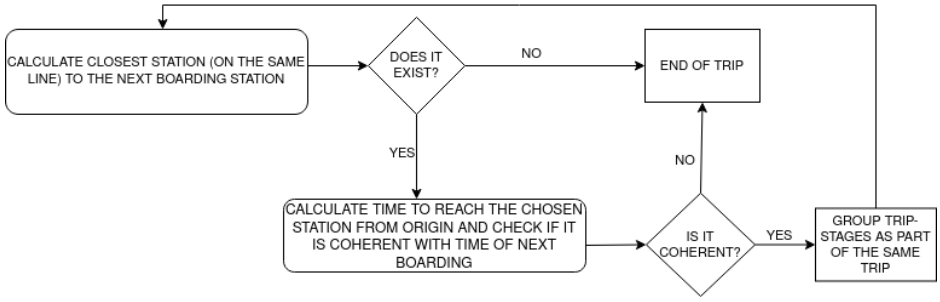

Figure 1: Flow of the algorithm for the prediction of alighting stop

3 Compact data structures for trip storage

In order to efficiently store and exploit the trips data, we are developing a new solution based on Compress Data Structures (CDS) [4]. A CDS is a data structure with self-indexing properties,

[1] https://www.crtm.es/

[2] 2020 was excluded from the analysis due to the anomalous use of the network caused by the COVID pandemic

capable of storing data in compressed form while still allowing for its processing without the need of decompression. The compressed data can then reside in main memory all the time, eliminating the need for disk accesses and thus yielding faster query times [5].

The development of CDS solutions for Big Data scenarios is a very active research field, and one we are very familiar with. We have already achieved promising results working with data referring to the use of public transport networks [3] by using a CDS supported by a modified version of a Compressed Suffix-Array (CSA) [6].

Furthermore, on the same work we also proposed T-Matrices (Trip Matrices) based on the idea of applying image rendering techniques to studying public transport loads. These could potentially simplify the approach to store aggregated data while speeding up cumulative queries on public transport information.

4 User interfaces for trip analysis

In order to visualize and exploit the wide range of data referring to user mobility on transport networks, we designed four main user interfaces, each devoted to different aspects relevant for transport management:

- **Detailed network usage**: for analyzing how many passengers have used a single line or boarded at a given stop for a given time frame and how this data evolves over time.
- **Origin-destination demand**: for analyzing overall demand by users for specific origin-destination trips from one area of the city to another. This kind of analysis was typically performed using origin-destination surveys.
- **Anomaly analysis**: for showing information on potential origin-destination pairs with a high demand but no convenient connections.
- **Accessibility analysis**: for identifying the connectivity issues between stops.

All in all, we will provide transport authorities with a detailed query interface for each of their individual needs.

References

[1] Azalden Alsger, Behrang Assemi, Mahmoud Mesbah, and Luis Ferreira. Validating and improving public transport origin–destination estimation algorithm using smart card fare data. *Transportation Research Part C: Emerging Technologies*, 68:490–506, 2016.

[2] Behrang Assemi, Azalden Alsger, Mahboobeh Moghaddam, Mark D. Hickman, and Mahmoud Mesbah. Improving alighting stop inference accuracy in the trip chaining method using neural networks. *Public Transport*, 12(1):89–121, 2020.

[3] Nieves Brisaboa, Antonio Fariña, Daniil Galaktionov, Tirso V Rodeiro, and Andrea Rodriguez. Improved structures to solve aggregated queries for trips over public transportation networks. *Information Sciences*, 584, 11 2021.

[4] Gonzalo Navarro. *Compact Data Structures: A Practical Approach.* Cambridge University Press, USA, 2016.

[5] Hasso Plattner and Alexander Zeier. *In-memory data management: technology and applications.* Springer Science & Business Media, 2012.

[6] Kunihiko Sadakane. New text indexing functionalities of the compressed suffix arrays. *Journal of Algorithms*, 48:294–313, 09 2003.