Mining passenger’s regional intermodal mobility from smartcard data

Xiaoyan Xiea,*, Fabien Leurenta, and Yazi Zhub

a LVMT, UMR-T 9403, École des Ponts, UGE (ex IFSTTAR, ex UPEM), UPE, 77455, Champs-sur-Marne, France
b INP ENSIACET-Toulouse, 4 Allée Emile Monso, 31030, Toulouse, France

Abstract

The Park and Ride (P&R) facilities were important emerging mobility facilities for reducing auto-dependency in cities. However, the estimation of the usages of the P&R facilities was still difficult due to the lack of related data in many cities. To this end, the objective of this paper is for mining passenger intermodal (P&R usage) information from Automated Fare Collection (AFC) data. This is accomplished by an approach of multiple data sources fusion based on AFC data, P&R facility information, Transport Survey data (EGT), and GTFS data in Greater Paris. Three main steps are considered: (1) P&R related rail transit EGT O-D trip generation, (2) P&R related rail AFC O-D trip generation, and (3) mining P&R related intermodality by a supervised learning model. We recover the intermodal O-D trips from the full population of AFC data of the 3 rail transit modes (metro, train-RER, and tram), and classify the P&R users. The usage of P&R facilities are obtained by an unsupervised learning model, and the P&R clients are classified in three groups: two groups during morning peak hours (one with short distance, and another with long distance) and one group during evening off peaks.

Peer-review under responsibility of the scientific committee of the 23rd EURO Working Group on Transportation Meeting.

Keywords: Intermodality, Activity based trip-chain model, Machine learning; AFC data; Park and Ride

1. Introduction

The Park and Ride (P&R) facilities for passenger intermodal mobility between the Public Transports (PTs) and the PVs were considered as important emerging mobility facilities for reducing auto-dependency in cities (Kimpton et al., 2020). There were 46 P&R facilities in Île-de-France (Paris area) in 2018, and will be 80 P&R facilities in 2021 (IdFM, 2020). However, the estimation of the usages of the P&R facilities was still impossible because of lack of related data. It was the same case in many cities around the world (Kimpton et al., 2020).

In the era of big data, over the decade, the Automated Fare Collection (AFC) system was widely integrated into PT system. Origin-Destination (O-D) trips on the PT network were generated from AFC data, and trip purposes were inferred for O-D trips (Alsger et al., 2018; Zhou et al., 2019). Theoretically, the passenger intermodal trips using the P&R facilities could be included in. Unfortunately, the authors assumed that the destinations of inferred O-D trips were the final destinations. The intermodal trips between PTs and Private Vehicles (PVs) were ignored.

To address the above-mentioned issue, the objective of this paper is to recover passenger intermodal (P&R usage) information from AFC data by extending our previous methodology in (Zhou et al., 2019). To accomplish that, this study will combine the AFC data with the P&R facility information, the P&R facility information, the Transport Survey data (EGT), and the GTFS data in Paris area, France. Three main steps will be considered: (1) EGT O-D trip generation, (2) AFC O-D trip generation, and (3) intermodal trip mining.

2. Models

This study extends our previous three-step method proposed for a closed PT network in (Zhou et al., 2019) to a network mixed closed and semi-closed PT networks. The aim of this part is to design an approach for passenger regional intermodal analysis in Île-de-France, France. We are interested in mining and analyzing passenger intermodal mobility related to passenger O-D trips on the PT network. A trip is a complete journey between one pair of O-D stations for an activity, including short time intermediate activity in / near an intermediate station and longtime activity. Thus, a trip is composed of one or several trip-legs, where a trip-leg is a journey between one pair of O-D stations along a single line. Thus, the PV trip in intermodality is either before the origin station or after the destination station.

* Corresponding author. Tel.: +33 1 81 66 88 77; fax: +33 1 64 15 51 40.
E-mail address: xiaoyan.xie@enpc.fr

Peer-review under responsibility of the scientific committee of the 23rd EURO Working Group on Transportation Meeting.

E 23rd EURO Working Group on Transportation Meeting, EWGT 2020, 16-18 September 2020, Paphos, Cyprus
**Studied site and multiple data sources.** This research revolves around analyzing passenger intermodality linked to the regional PT network in Île-de-France, including three PT sub-modes, train / RER, metro and tram. Among the data provided by this region, the applied datasets are: (1) passenger related data: the AFC data and the Transport Survey data; (2) the transit network information (network and its geo-location information); (3) the vehicle timetable data: the General Transit Feed Specification (GTFS) data; and (3) the P&R geo-location data in 2018 in Île-de-France. The AFC data are the anonymous Navigo data on the 11th February 2019, which are about 13 million records. The Transport Survey data are the EGT2010 (Enquête Globale Transport in 2010) data (OMNIL, 2012).

**EGT O-D trip generation.** EGT O-D trip generation is achieved directly by trip mapping on the PT network. EGT O-D trips are characterized by different trip purposes at the destination stations, including intermodality and final destination. When a destination is the final destination, there were seven O-D trip purposes around the destination stations in Île-de-France (Zhou et al., 2019), including work, professional affaire, school, leisure, shopping, personal affaire, and home.

**AFC O-D trip generation.** AFC O-D trip generation is accomplished by an Activity based Trip-chain Model (ATM). This is accomplished by an approach of data fusion based on AFC data and GTFS data. In this stage, we extended our ATM for closed RER network in (Zhou et al., 2019) to the semi-closed network for metro and tram in Île-de-France. Two sub-steps are considered for O-D trip generation: (1) the inference of an O-D pair, and (2) the inference of a transfer and the O-D trip.

**Intermodal trip mining.** Since AFC O-D trips do not record passenger intermodal information, the AFC O-D trips are confronted with EGT O-D trips for mining the intermodal trips from AFC O-D trips, using an Intermodality Inference Model (IIM). The IIM integrates a Supervised Learning (SL) algorithm, and is achieved by three main sub-steps. In the first sub-step, we match the geo-location information of P&R to the PT network model to find out the P&R related stations. The O-D trips linked to P&R stations are extracted for both AFC and EGT databases. The second sub-step is to summarize the common temporal and spatial attributes of intermodal trips of the two databases as the labels of the SL algorithm. The third sub-step is the training and validation of the SL algorithm. Three SL algorithms are applied and compared: Support Vector Machine (SVM), Decision Tree and TensorFlow.

**Implementation.** This approach is implemented in Python with big-data management and Machine Learning (ML) packages. This study completes our prototype of the Mobility Analytics Platform for Île-de-France (MAPI).

3. Results and Discussions

This study proposes an approach to mining passenger intermodal trips from AFC data. The statistical analyses and illustrations are proposed at first for the AFC records, EGT data, AFC O-D trips, EGT O-D trips, and O-D trips linked to P&R stations of the two databases. Secondly, an IIM integrates an SL algorithm based on trip spatial and temporal common attributes of the two databases to infer the intermodal trips from AFC O-D trips. The results of the three SL algorithms are cross-compared. Three model evolution metrics including the precision, the recall, and the F1-score are verified for model choice. The results show that the Decision Tree has the best performance for this study. The results of Decision Tree is further used to characterize the AFC intermodal trips and illustrated on the map. That gives the information of P&R usage by PT passengers in Île-de-France, and the clients are classified in different groups.

For further application, the information of P&R usage by PT passengers in Île-de-France can be used for supporting the future development and management of P&R in Île-de-France. Secondly, since the inconsistent dates of the applied datasets of EGT and AFC may make the results of the two datasets different, this study should be validated by using EGT and P&R datasets in 2019. Thirdly, by integrating with other data sources, more features can be recovered to describe the intermodal motility patterns. Finally, more sophisticated ML methods can also be explored to improve the process of IIM.

**Acknowledgements**

This research was supported by the Research and Education Chair ENPC-IdFM operated by ENPC in partnership with the IdFM (ex STIF), to whom the authors are grateful.

**References**


