Mining regional passenger mobility activity using multiple data sources
A case study in Paris area

Yao Zhou, Xiaoyan Xie, and Fabien Leurent
École des Ponts ParisTech (ENPC), France
Plan

- Introduction
- Applied case
- Methodology
- Results
- Discussions
Objective

Recovering passenger mobility information for AFC data

Methods

- **Activity based trip-chain model** with original trip generation algorithm
- **Machine Learning approaches** e.g. Supervised learning model
- **Multiple data sources**
  - Modern data (AFC data & GTFS data)
  - Traditional survey data (EGT data)
Closed network: **RER network** in region Ile-de-France (Paris area) including 5 regional closed transit lines, RER line A, B, C, D, E

**AFC (Automated Fare Collection) data:** Navigo data
- 14/03/2017, Ile-de-France
- Quantity: about 13 million records
- Ile de France Mobilités

**General Transit Feed Specification (GTFS) data**
- 19/07/2018, Ile-de-France
- Transilien lines timetable
- SNCF OpenData

**Transport Survey (EGT, Enquêtes Global Transport) data**
- A day in 2008, Ile-de-France
- Quantity: about 0.12 million journeys
- Ile de France Mobilités
Activity based trip-chain model + Machine Learning approach (Supervised learning models)

Modeling schema

Data Source
- AFC
- GTFS
- EGT

Activity based trip-chain model
Trip generation
- Tap-in
- Tap-out
- Trip leg
- Transfer
- Trip
- Activity
- Trip

Supervised learning models
Trip purpose prediction
- Decision Tree
- SVM
- Naive bayes

$p(C_k|x) = \frac{p(x|C_k)p(C_k)}{p(x)}$
AFC data: RER network
EGT data: RER related trips
AFC data: Trip purpose – SVM

Diameter around a station: 
R = 400m

On regional network RER

The most frequent:
• Work
• Home

Differences: AFC – EGT
• The biggest: EGT – Home, but AFC
• Work
AFC in 2017, but EGT in 2008
Discussions

Recovering passenger mobility information for AFC data: mobility purpose

- Model combining Traditional method (Activity based trip-chain model) with New approaches (Machine Learning approaches) for multiple data fusion on a regional large scale Public Transport network
- Activity based trip-chain model: original trip generation algorithm
- Multiple data sources: Modern data (AFC data & GTFS data) and Traditional survey data (EGT data)
- Side product: a tool for both management of big data and application of multiple data sources related to passenger regional mobility analyses for Île-de-France
- Limits: Survey data sample was old and small, and Lack of land use data for results validation

Perspectives

- On-going work: Results validation by land us dataset in Île-de-France
- In the future
  - Updating the survey (EGT) data 2020 for the most recent mobility characterization in Île-de-France
  - Extending mono-mode study to multi-mode study to recovering more mobility information
Thanks!
Questions?

Contact: xiaoyan.xie@enpc.fr