Predictive Multimodal Trip Planner: A New Generation of Urban Routing Services

Ahmed Amrani, Kevin Pasini, Mostepha Khouadjia

IVA Project, IRT SystemX, Paris-Saclay, France

TransitData
Paris
09/07/2019
Engagement of the transport operators to enhance their communication towards their passengers:

- Develop an interoperable multimodal information system
- Provide relevant and useful travel information
- Enhance the passenger travel experience with this information

Project Partners

(July 2017, July 2021)
- Increasing interest for route planning in public transport due to the crowded and jammed traffic.

- Availability of digital footprints provide a relevant opportunity to develop innovative decision-making tools for urban routing of passengers.

- Possibility to translate these footprints in network quantitative indicators (attendance, load on board, ...), and qualitatives (comfort, accessibility, painfulness, ...).
Motivations and Interests

- Aim to develop *new generation of trip planner for public transport*

- Introduce these forecasting indicators in the *passenger routes*:
  - Exploitation of the collected data for *user centric needs*.
  - Help the passenger to *better planning* his journey by considering the forthcoming evolution of the traffic.
  - **Support** the passenger during his travel and particularly when incidents arise on the network.
  - Enhance the *travel experience* of the passengers.
Integrate into passengers trips qualitative forecasted indicators given by Machine Learning models for:

- Attendance at the station
- Load on board the trains

These indicators are updated for each route segment and for each station

Proof-Of-Concept playground is the Transilien line H operated by SNCF.
Data sources

- **Attendance at the station**
  - Validation logs of smart cards obtained from Automatic Fare Collection System (AFCS)

- **Load on board for each train**
  - Automatic Vehicle Location (AVL) with counting given by boarding and alighting sensors
  - Reconstructed train load at each train stop
Predictive Trip Planner: The proposed architecture

**DATA LAKE**
- ETL
- Batch

**MODELS FACTORY**
- Feature Engineering
- Machine Learning
- Forecasting Models
  - Load on board train
  - Attendance in station

**END-USER APPLICATION**

**DATA SOURCES**
- Batch
- API
- Real-Time

**PREDICTIVE PLANNING**
- Trip Planner
  - Request (orig, dest, date)
  - Trips
- Trip-indicators Combiner
  - Request (orig, dest, date)
  - Trips, loads, attendance
  - Response
- navitia.io

**Forecasting Models**
- Trip-request (date, time, station, mission)
- Response (load on board, attendance)

**Real-Time**
- Load on board train
- Attendance in station
Train load

- The discretization is subject to the time of train schedule at the station
- Different station profiles (frequency, #mission, transport demand)
- Train load impacted by several factors (calendar, service)
- Missing data represent 15% of raw dataset (trains without sensors)

Attendance at the station

- Time series are structured by step of 15 minutes
- Ticketing data aggregated for each step
- Missing data could have different senses (failed AFCS, incidents,...)
### Forecasting Load on Board: Machine Learning Workflow

**Load raw data**
- Arrival and departure time of the train at the station
- Aggregated passenger load at the train level

**Training Dataset**
- Calendar information (months, days, week, day-off, holidays,....)
- Theoretical time passage at the station
- Train service information such as its mission and route. (short-term features)
- Recorded delay at the station
- Lag of eight last recorded loads at the station (long-term features)

**Algorithm**
- XGradient Boosting (XGB)

**Evaluation**
- Cross-validation (k=3)
- Grid search

**Testing Dataset**

**Model**

**Prediction**

**Production data (request)**
Forecasting attendance: Machine Learning Workflow

- Ticketing volume at the station
- Aggregated at station level for each quarter of an hour.

Ticketing raw data

Testing Dataset

Training Dataset

Algorithm

- Random Forests (RF)

Evaluation

- Cross-validation
- Grid search

Model

Prediction

Production data (request)

- Calendar information (months, days, week, day-off, holidays,...) (long-term feature)
- Lag of six last recorded volumes at the station (short-term feature)
XGB and RF are better than the baseline model CA

Ability to capture the last dynamic by using the lag of last observations

Short term features (ST) improved slightly performance for train load forecasting, but significantly for station attendance forecasting

<table>
<thead>
<tr>
<th>Models</th>
<th>Suburb</th>
<th>Inner-city</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WAPE</td>
<td>RMSE</td>
</tr>
<tr>
<td>Train score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>13.9</td>
<td>29.2</td>
</tr>
<tr>
<td>XGB LT</td>
<td>8.4</td>
<td>17.2</td>
</tr>
<tr>
<td>XGB LT + ST</td>
<td>7.3</td>
<td>15.6</td>
</tr>
<tr>
<td>Test score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>19.1</td>
<td>39.8</td>
</tr>
<tr>
<td>XGB LT</td>
<td>18.8</td>
<td>38.9</td>
</tr>
<tr>
<td>XGB LT + ST</td>
<td>16</td>
<td>33.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>61.48</td>
<td>18.18</td>
</tr>
<tr>
<td>RF LT</td>
<td>56.42</td>
<td>16.77</td>
</tr>
<tr>
<td>RF LT + ST</td>
<td>11.16</td>
<td>4.44</td>
</tr>
<tr>
<td>Test score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>62.06</td>
<td>18.84</td>
</tr>
<tr>
<td>RF LT</td>
<td>61.41</td>
<td>18.53</td>
</tr>
<tr>
<td>RF LT + ST</td>
<td>31.17</td>
<td>12.9</td>
</tr>
</tbody>
</table>
Predictive Trip Planner: Display

- **Graphical User Interface (GUI)**

  Introduce the request by giving the origin and destination and time of departure.

  Selection of the first obtained journey and display its steps with attendance and load forecasting indicators.
The interest of the integration between a trip planner and forecasting models to provide a new added value for:

- Enrich the passenger information in multimodal transport system
- Improve the passenger travel experience

Forecasting models demonstrated the effectiveness of machine learning methods for such prediction tasks

Future research:

- Introduce new indicators concerning abnormal situations (incidents, disturbance) such as their severity
- Plan the trips based on these indicators as criteria for passenger travel
Thanks for your attention

https://www.irt-systemx.fr/project/iva/

mostepha.khouadjia@irt-systemx.fr