A Bayesian Approach to Detect Pedestrian Destination-Sequences from WiFi Signatures: Data description

Antonin Danalet *

11th March 2014

Report TRANSP-OR yymmdd (draft)
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
`transp-or.epfl.ch`

Abstract

This technical report describes the data used in (Danalet et al.; 2013): WiFi traces, pedestrian Semantically-Enriched Routing Graph (SERG), and Potential Attractivity measure (PAM). It is a complement to the data available in the folder `Data`.

Key words

Network traces, pedestrians, semantically-enriched routing graph (SERG), potential attractivity measure

*ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory (TRANSP-OR), antonin.danalet@epfl.ch

Conditions of use

Contents

1 WiFi Traces

Network traces are defined in (Danalet et al.; 2013), Section 3.1. The specific data used in the case study on EPFL campus are described in Section 5.1. They can be found in `Data/WiFiTraces/1405data.csv`.

The `csv` file contains 6 columns.

The first column and second columns contain the timestamp $\hat{t}$; The third and the fourth ones contain the x-y coordinates of the measurement in Swiss coordinates (coordinate reference system: CH1903 / LV03, EPSG: 21781); The fifth column contains the floor; And finally, the sixth column contains the cF factor as defined in Figure 6 in Danalet et al. (2013).

This data collection has been performed by Richard Timsit from the Network services team (DIT-TI) of EPFL's Information Technology Domain (DIT). The first step consisted in integrating a supplementary module in the network infrastructure. The tool is Cisco Mobility Services Engine (MSE). It was lent by Cisco for this research, thanks to François Pujol. Then, the data were extracted and consolidated with the localization of the network infrastructure (localization of the access points) and with the data from the Radius server. Since it consisted here of the personal data of one of the authors, they were not anonymized. An automatic system for collecting regularly these data was implemented and some corrections had to be performed during the data collection campaign (e.g., some access points were changed or modified).

2 Pedestrian semantically-enriched routing graph (SERG)

Pedestrian semantically-enriched routing graph (SERG) is defined in Danalet et al. (2013), Section 3.2. The specific data used in the case study on EPFL campus are described in Section 5.2. They can be found in `Data/SERG/` .

The folder contains 7 `sql` files.

`destinations.sql`, `locaux.sql` and `routing.sql` build 4 tables : `edges` (with the edges of the graph), `vertices` (with the vertices), `destinations` (with points of interest) and `locaux` (all other rooms that are not points of interest). `vertex_id` is the main link between all these tables.

2.1 Edges

The table `edge` contains 15 columns:

edge_id is an ID for each edge.

vertex_id1 and vertex_id2 are the IDs of the two extremities of the edge, defined in table `vertices`.

level1 and level2 represent the floor of each extremity of the edge. It is used to define if the edge is going up or down in the shortest path algorithm.

type is the type of infrastructure to change floor using this edge. It takes values 9.1, 9.2 or 9.3. 9.1 represents a ramp. 9.2 represents stairs. 9.3 represents a lift. It is used in the shortest path algorithm.

door has value S when a door on this edge is locked. This is used in the shortest path algorithm.

network takes values 0, 10, 20 or 30. It represents the hierarchical status of the edge, as defined in Section 5.2 in Danalet et al. (2013). It is used in the shortest path algorithm.

length is the length of the edge. It is used in the shortest path algorithm.

the_geom is the PostGIS geometry column.

2.2 Vertices

The table `vertices` contains 3 columns:

vertex_id is an ID for each vertex. It is used in `edges`, `destinations` and `locaux`.

level is the floor of the vertex, between -2 and 7.

the_geom is the PostGIS geometry column.

2.3 Points of interest

The table `destinations` contains 6 columns:

id is an ID for each point of interest.

etage is the floor, between -2 and 7.

theme is the category of the point of interest, corresponding to the categories presented in Table 3, Appendix B in Danalet et al. (2013).

nom is the name of the point of interest.

vertex_id corresponds to the vertex in the pedestrian graph, defined in the table `vertices`.

the_geom is the PostGIS geometry column.

2.4 Rooms

The table `locaux` contains 6 columns:

gid is an ID for each room.

etage is the floor, between -2 and 7.

room_abr is the abbreviation for the name of the room, used to describe each room on EPFL campus.

room_din is the code for the type of the room, defined in `Data/SERG/DIN-CUS_277.pdf` (in French). Of particular interest on campus are categories 2, 3 and 5. Category 2 is office work, with offices (2.1) and conference rooms (2.3). Category 3 is laboratory work. Category 5 is about teaching.

vertex_id corresponds to the vertex in the pedestrian graph, defined in the table `vertices`.

the_geom is the PostGIS geometry colum.

The `create` files are just to make it faster to install it.

The tables must be PostGIS-compatbile. The commands look like:

```
psql -U myUsername -d myDatabase -f /usr/share/postgresql/8.4/contrib/postgis.sql
psql -U myUsername -d myDatabase -f /usr/share/postgresql/8.4/contrib/spatial_ref_sys.sql
```

Finally, `test_routing.sql` is the weighted shortest path described in Appendix C.

This SERG is owned by EPFL. The data have been generated by Florent Deseneux and his team in the Real Estate and Infrastructure Department (DII) from EPFL. They drew the lines with meta-data with AutoCAD. It has been converted to a network in a database by Yves Bolognini from Camptocamp. It is used by `http://map.epfl.ch`, developed by Camptocamp, with a shortest path algorithm, presented in Danalet et al. (2013) and available in the `test_routing.sql` file. In order to ask for the permission to use it, contact Florent Deseneux from EPFL.

3 Potential attractivity measure

Potential attractivity measure is defined in Danalet et al. (2013), Section 4.1.2. The specific data used in the case study on EPFL campus are described in Section 5.3. They can be found in `Data/PAM/` .

The folder contains 1 `sql` backup file and raw data. The `sql` backup file contains 7 columns.

week_day contains the week day. 0 is Monday, 1 is Tuesday, etc.

node_id contains the `vertex_id` from SERG.

quarter_hour contains the time interval as the number of quarters of an hour from the beginning of the day. 0 is the time interval 0:00 to 0:15, 1 is 0:15 to 0:30, etc.

attractivity contains the potential attractivity measure, as defined in Danalet et al. (2013).

category contains the category of the attractivity. It corresponds to its definition in (Danalet et al.; 2013).

The category of the potential attractivity measure in the table is either empty, disagregate or aggregate.

Empty means that these potential attractivity measures don't vary for different people. It includes

- **Office occupation from SAP** `Data/PAM/RawData/EPFL_Capacity_offices.csv`, all day long
- **Capacity of conference rooms** 3 (chosen by the modeler), all day long
- **Occupation of language classes** `Data/PAM/RawData/Cdl_EPFL_rooms` for the schedule, around 13 people[1]
- **Capacity of the multimedia center for personalized learning** 30 people[2], schedules in `Data/PAM/RawData/Cdl_EPFL_EMA`,
- **Points of interest** are originally defined in `http://map.epfl.ch`. They are first filtered as described in Appendix B in Danalet et al. (2013). Then,

[1] Communication with the head of the Language Center EPFL
[2] Communication with the head of the Language Center EPFL

Shop (Négoce) has capacity 30 (defined by the modeler) and is open from 6am to 6pm from Monday to Friday[3]

Library has capacity 860 and is open every day of the week from 7am to midnight[4]

Restaurants have a capacity defined in `Data/PAM/RawData/Seats_restaurants_EPFL.xlsx`, both indoor and outdoor. For restaurants with no pre-defined capacities, the modeler defined it (30 for "Roulotte Esplanade", 15 for "Roulotte Diagonale", 30 for "Cafeteria BM", 5 for "Cafeteria PH"). The schedules are extracted from an EPFL webpage[5]. In one case ("Cafeteria PH"), no schedules are defined, and thus we assume it is open 24/24.

Post office has a capacity of 13 when it is open and of 3 when it is closed (ATM is still available). This has been defined by the modeler. The schedules are available on EPFL webpage[6].

Bank has been defined having a capacity of 3 (ATMs), open 24/24, 7/7 ("Crédit Suisse Esplanade")

Student union ("AGEPoly") has a capacity of 3 (modeler's decision) and schedule are available on their webpage[7].

For all other points of interest, a default value of 1 has been defined, unless there is a higher occupation from SAP data.

Disaggregate represents the disaggregate prior as defined in Danalet et al. (2013). It contains the class schedules available on the class webpage[8] for one of the authors. The number of registered students was 30 at the beginning of the semester, 27 at the end, and 26 in the registration system on January 9, 2013. 26 is used in the `sql` dump. For other classrooms, capacity has been fixed to 1. It represents the fact that even if no class is scheduled, classrooms are usually open and accessible (e.g., to work).

Aggregate represents the aggregate prior as defined in Danalet et al. (2013). It contains the number of registered students and the class schedules as defined in

`Data/PAM/RawData/class_schedules.xls`.

[3] Schedule available on `http://restauration.epfl.ch/negoce`
[4] `http://rolexlearningcenter.epfl.ch/page-45189-en.html`
[5] `http://restauration.epfl.ch/horaires`
[6] `http://information.epfl.ch/commerces`
[7] `http://agepoly.epfl.ch/agepoly/contact`
[8] `http://transp-or.epfl.ch/courses/OptSim2012/schedule.php`

All data have been collected on webpages (see `Data/PAM/RawData/Webpages`) or from different services at EPFL: Brigitte Forster Vosicki, administrative manager in Centre de langues, provided `Cdl_EPFL_EMA.doc` and `Cdl_EPFL_rooms.doc`, Cédric Junillon, deputy of head of service in registrar's office, provided `class_schedule.xls`, Yves Junod, computer scientist in the Knowledge and Information Services and Philippe Pichon, project manager finance & controlling from the real estate and infrastructure department management, provided `EPFL_Capacity_offices.csv`, Roland Deléchat, project manager in the Vice Presidency for Planning and Logistics, provided `Seats_restaurants_EPFL.xlsx`

References

Danalet, A., Farooq, B. and Bierlaire, M. (2013). A Bayesian Approach to Detect Pedestrian Destination-Sequences from WiFi Signatures, *Technical report*, Transport and Mobility Laboratory, ENAC, Ecole Polytechnique Fédérale de Lausanne, Lausanne.
URL: *http://infoscience.epfl.ch/record/189759*