EE-490(h) - Lab In Data Science

Robust Journey Planning

Firas Kanoun, Sami Ben Hassen, Zeineb Sahnoun

Contents

1. Motivation

- Multi-modal journey planning is becoming increasingly popular
- However nowadays, most used models use a deterministic approach.
- In real life, transport networks feature uncertainty.
- Journey plans should be able to handle such unexpected situations without compromising the journey duration beyond some acceptable level.

2. Assumptions & Simplifications

- We only consider stations within a range of 10kms from Zürich HB.
- A given journey will take time at most 1h, which is reasonable considering the 10km constraint.
- We assume statistical independence between the observed transportation delays.
- Walking time at different connections takes 2mins.
- Walking speed for a random user of the system is 4km/h.

3. Data Cleaning

- We only consider rows where the actual times of arrival and departure were measured (AN_PROGNOSE_STATUS is set to GESCHAETZT)
- We chose to use data from 04/2018 because actual measures times were most frequent than other more recent months.
- We only keep useful columns (namely date, trip_id, train_id, station_id, stop_name, arrival_time, actual_arrival_time, departure_time, actual_departure_time)

	date	trip_id	train_id	BPUIC	stop_name	arrival_time	actual_arrival_time	departure_time	actual_departure_time
0	20.04.2018	85:11:1255:001	1255	8503000	Zürich HB	20.04.2018 08:26	20.04.2018 08:27:41	20.04.2018 08:37	20.04.2018 08:39:43
1	20.04.2018	85:11:1258:001	1258	8503000	Zürich HB	20.04.2018 21:23	20.04.2018 21:25:19	20.04.2018 21:34	20.04.2018 21:34:51
2	20.04.2018	85:11:1507:002	1507	8503000	Zürich HB	20.04.2018 06:30	20.04.2018 06:30:14	20.04.2018 06:39	20.04.2018 06:39:52
3	20.04.2018	85:11:1507:002	1507	8503016	Zürich Flughafen	20.04.2018 06:49	20.04.2018 06:49:43	20.04.2018 06:51	20.04.2018 06:52:01
4	20.04.2018	85:11:1509:003	1509	8503000	Zürich HB	20.04.2018 07:30	20.04.2018 07:31:24	20.04.2018 07:39	20.04.2018 07:39:36

4. Graph Construction & Shortest Path

- We reconstruct the transportation system by grouping by trip_id, removing redundant itineraries and taking the longest one.
- We then create a graph where each node represents a station, and edges represent a segment of an itinerary with departure/arrival times and durations.
- We run Dijkstra's algorithm to compute the shortest path from station A to station B.

	trip_id	departure	departure_time	arrival	arrival_time	time
0	85:11:20453:001	Bassersdorf	1900-01-01 13:56:00	Zürich Flughafen	1900-01-01 14:01:00	5
1	85:11:20453:001	Zürich Flughafen	1900-01-01 14:03:00	Zürich Oerlikon	1900-01-01 14:07:00	4
2	85:11:20453:001	Zürich Oerlikon	1900-01-01 14:08:00	Zürich Wipkingen	1900-01-01 14:10:00	2
3	85:11:20453:001	Zürich Wipkingen	1900-01-01 14:10:00	Zürich HB	1900-01-01 14:16:00	6
4	85:11:20453:001	Zürich HB	1900-01-01 14:21:00	Zürich Wiedikon	1900-01-01 14:24:00	3

5. Probabilistic Model

- We use an exponential distribution to model the probability of a missed connection between two itinerary segments.
- We use the mean of previous delays of the given trip_id.

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- The resulting **Robustness** parameter will be equal to “the bottleneck” probability of successful connection in our itinerary.

6. Validation & Visualization

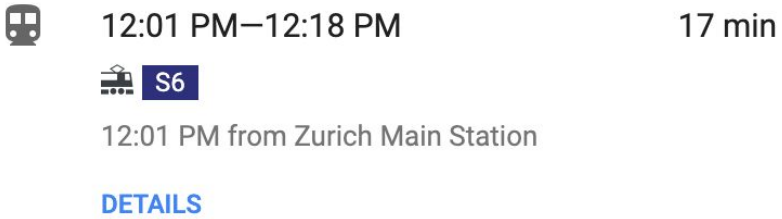
We use the itinerary from Zürich HB to Zürich Affoltern as a validation example.

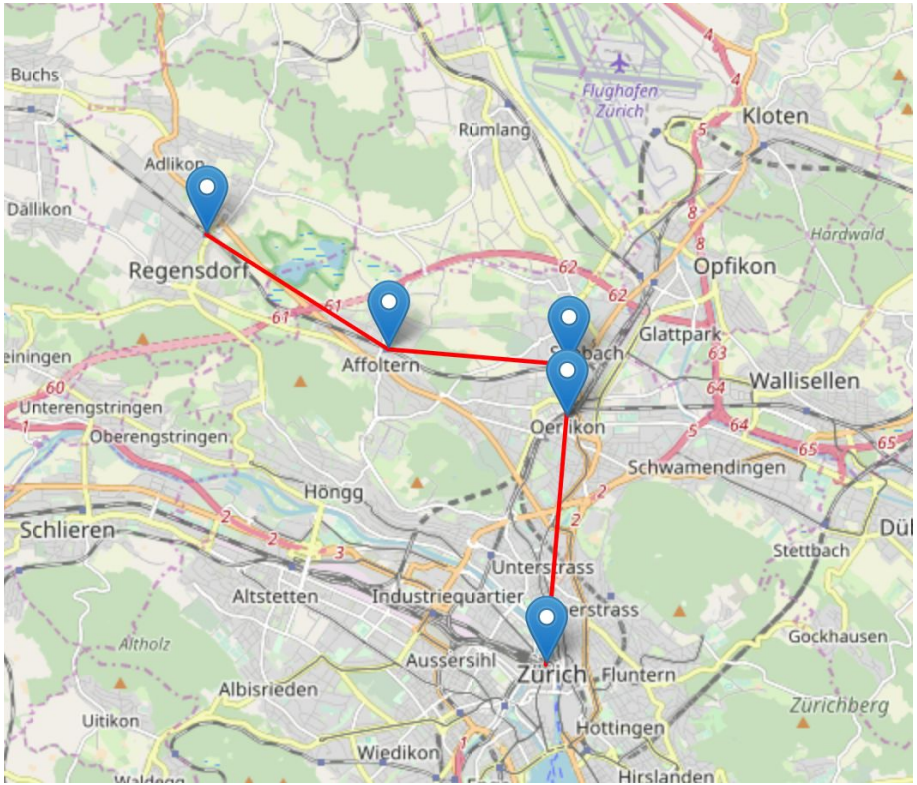
Comparing our itinerary to the one we get from Google Maps, both take 17mins but our algorithm displays a Robustness score equal to 0.83 which will help the user determine how likely he will arrive on time.

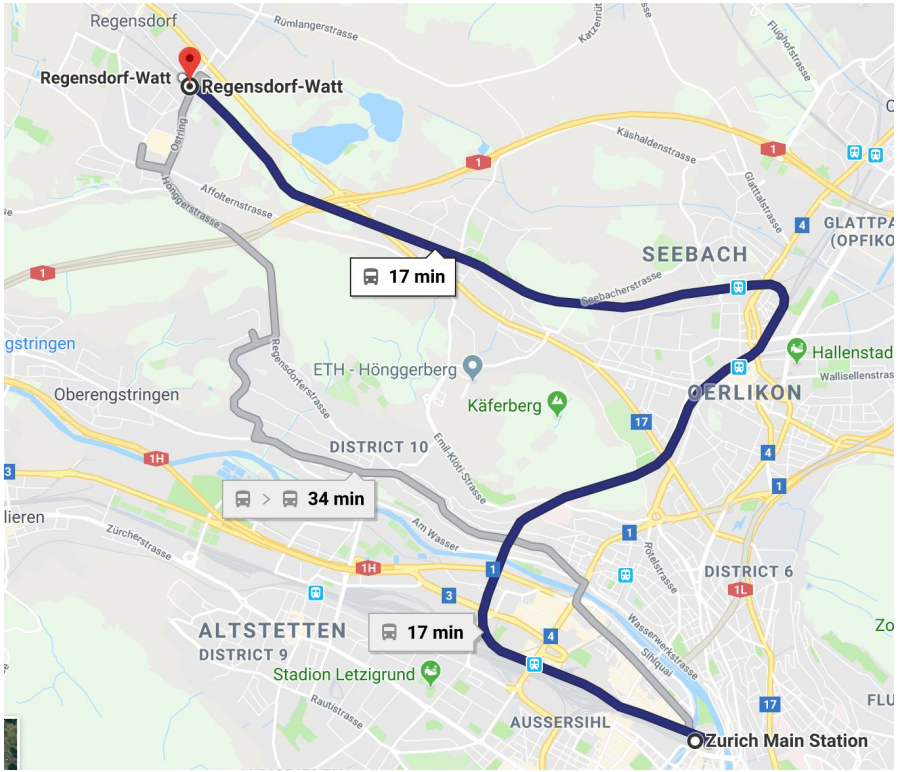
Our output:

```
start at Zürich HB at time 12:00
--------------------
train from Zürich HB at 12:01:00 arrives in Zürich Oerlikon at 12:07:00 85:11:2640:001 with probability 1
--------------------
train from Zürich Oerlikon at 12:09:00 arrives in Zürich Seebach at 12:11:00 85:11:18644:001 with probability 0.83098
66845939339
--------------------
train from Zürich Seebach at 12:11:00 arrives in Zürich Affoltern at 12:14:00 85:11:18644:001 with probability 1
--------------------
train from Zürich Affoltern at 12:14:00 arrives in Regensdorf-Watt at 12:18:00 85:11:18644:001 with probability 1
--------------------
Robustness of this journey is 0.8309866845939339
```

Google Maps output:




Buchs
Dällikon
Adlikon
Regensdorf
Affoltern
Rümlang
Flughafen Zürich
Kloten
Hardwald
Opfikon
Glattpark
Wallisellen
Oerlikon
Unterengstringen
Oberengstringen
Schwamendingen
Höngg
Schlieren
Stettbach
Unterstrass
Altstetten
Industriequartier
Gockhausen
Aussersihl
Zürich
Fluntern
Albisrieden
Zürichberg
Uitikon
Hottingen
Wiedikon
Hirslanden


Regensdorf
Regensdorf-Watt
Regensdorf-Watt
17 min
SEEBACH
GLATTPA
(OPFIKO
Hallenstad
OERLIKON
ETH - Hönggerberg
Käferberg
Oberengstringen
DISTRICT 10
34 min
DISTRICT 6
ALTSTETTEN
DISTRICT 9
17 min
Stadion Letzigrund
AUSSERSIHL
Zurich Main Station