Artificial Neural Networks Applied to Taxi Destination Prediction

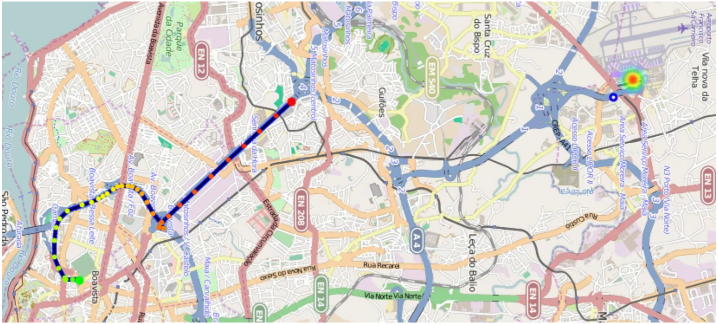
Alexandre de Brébisson Étienne Simon Alex Auvolat
Team

MILA, University of Montreal

September 7, 2015
Porto ECML/PKDD

Table of contents

Task



- City of Porto
- Evaluation score: Haversine distance
- Improve the dispatch system of the taxis

Datasets

Training:

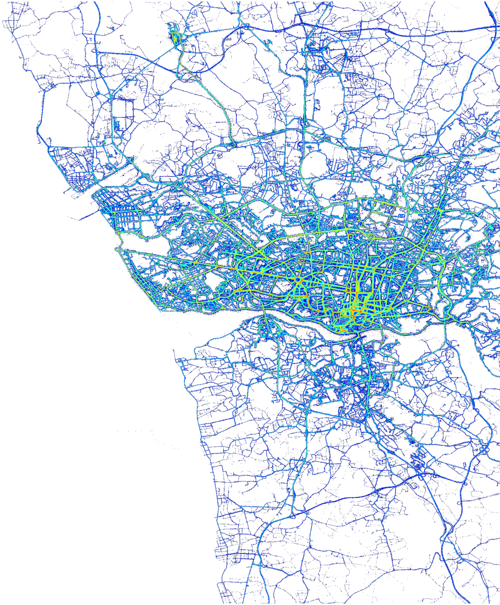
- 1.7 million training complete trajectories,
- (Mean, Std) path length: (12min, 11min),
- Longest ride: 16h 10min,
- Taxi lost in Iran.

Testing:

- **Kaggle testing:** 320 testing prefixes,
- **Custom testing dataset**: 20 000 testing prefixes.

Features

- the complete taxi ride: a sequence of GPS positions (latitude and longitude) measured every 15 seconds.
- metadata associated to the taxi ride:
  - client ID,
  - stand ID,
  - taxi ID,
  - timestamp of the beginning of the ride.



Embeddings for meta-data

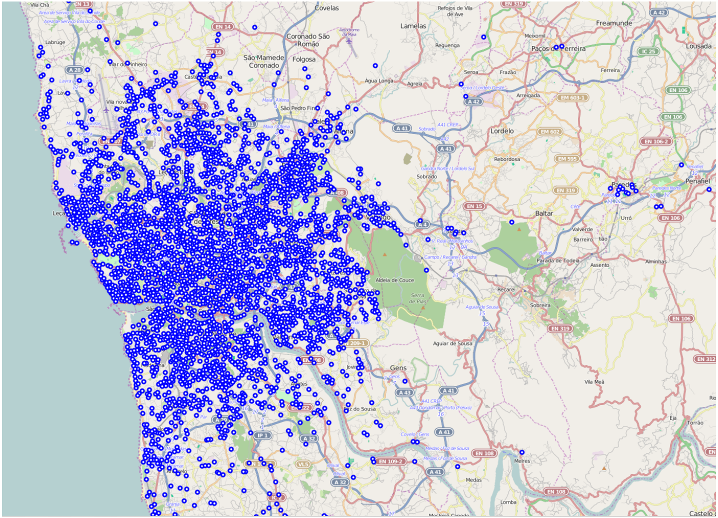
Timestamp → (Week of the year, Day of the week, quarter hour of the day)

$$
\begin{array}{ccc}
\text{Monday} & \xrightarrow[\text{embedding}]{\hspace{3cm}} & (w_{1,1}, \ldots, w_{1,10}) \\
\vdots & & \vdots \\
\text{Sunday} & \xrightarrow[\text{embedding}]{\hspace{3cm}} & (w_{7,1}, \ldots, w_{7,10})
\end{array}
$$

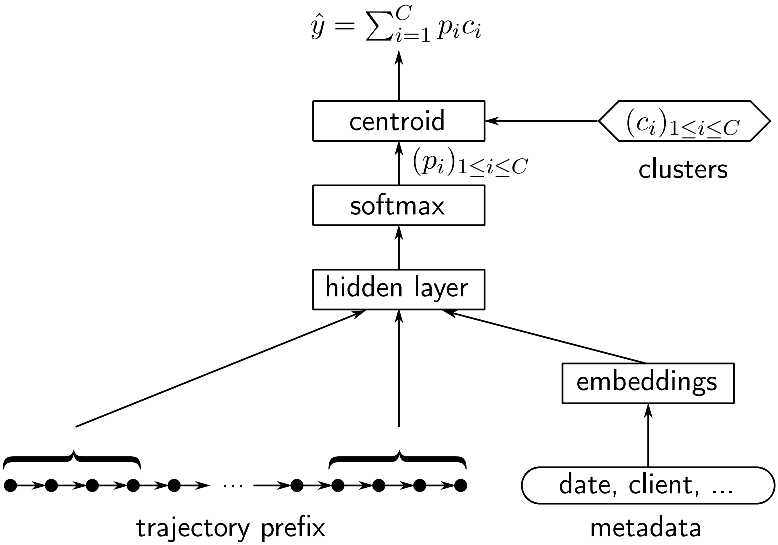
Metadata	Number of possible values	Embedding size
Client ID	57106	10
Taxi ID	448	10
Stand ID	64	10
Quarter hour of the day	96	10
Day of the week	7	10
Week of the year	52	10

Structure of the outputs

Mean-shift algorithm to find 3392 clusters.



Winning Architecture



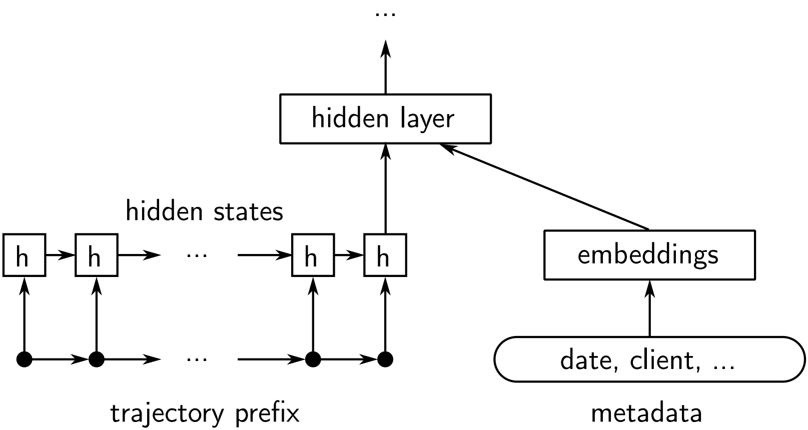
Training

- Training data is composed of complete trajectories, whereas our model processes prefixes.
- Training and Testing data distributions.
- Equirectangular approximation of the Haversine distance as the training cost:

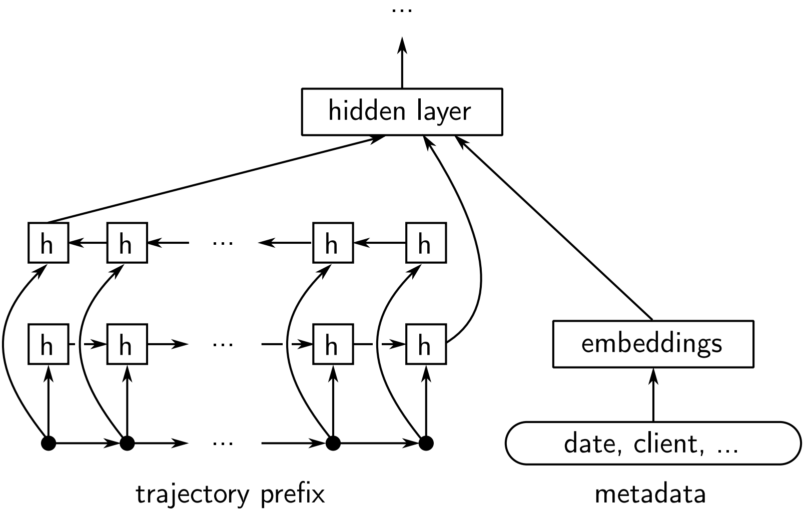
$$d(x, y) = R\sqrt{\left((\lambda_y - \lambda_x)\cos\left(\frac{\phi_y - \phi_x}{2}\right)\right)^2 + (\phi_y - \phi_x)^2}$$

- SGD with momentum.

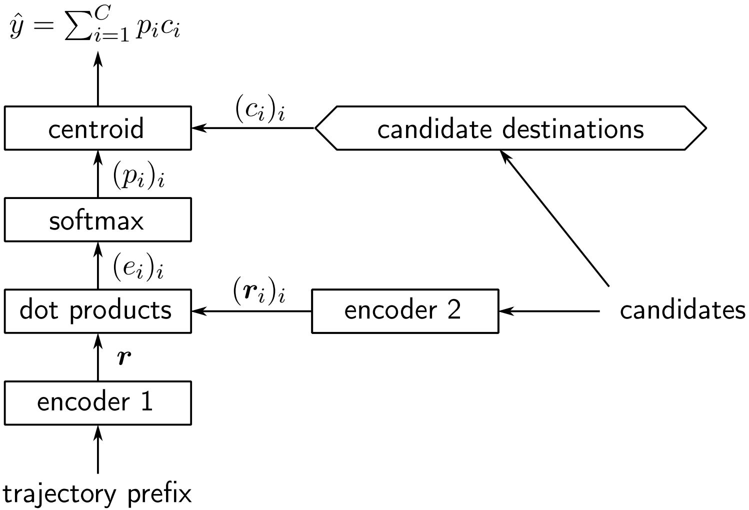
RNN


...
hidden layer
hidden states
h
h
...
h
h
embeddings
...
date, client, ...
trajectory prefix
metadata

Bidirectional RNN



Memory-Network-like Architecture



Results

	Model	Custom Test	Kaggle Test
1	MLP, clustering (winning model)	2.81	**1.87**
2	MLP, clustering, no embeddings	2.93	2.17
3	MLP, clustering, embeddings only	4.29	3.76
4	RNN	3.14	2.39
5	Bidirectional RNN	3.01	2.33
6	Bidirectional RNN with window	**2.60**	2.06
7	Memory network	2.87	2.20
	Second-place team	-	2.09
	Third-place team	-	2.11
	Average competition scores[1]	-	3.11
	Constant predictor (center of Porto)	4.71	4.19

[1] Only the scores better than the center of Porto predictor are selected. 381 teams.

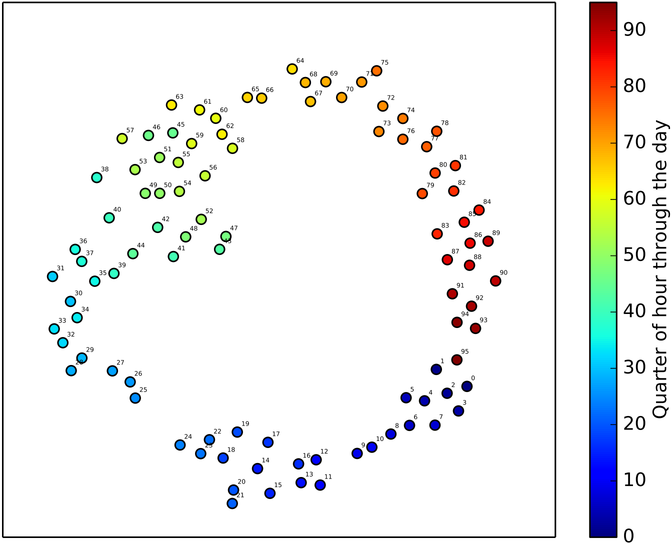
Confidence of the models

Null hypothesis: the errors of the models have the same mean.

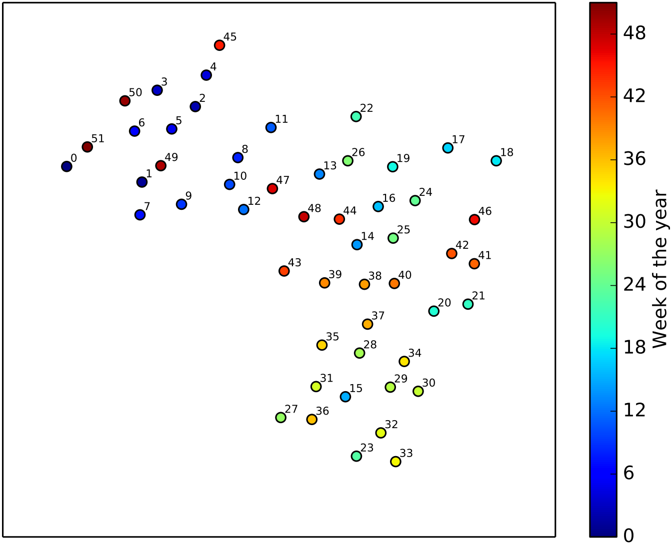
Model	1	2	4	5	6
2	$4.0 \cdot 10^{-2}$				
4	$1.4 \cdot 10^{-6}$	$1.6 \cdot 10^{-3}$			
5	$2.2 \cdot 10^{-3}$	$1.4 \cdot 10^{-1}$	$2.9 \cdot 10^{-2}$		
6	$\mathbf{3.8 \cdot 10^{-4}}$	$6.5 \cdot 10^{-8}$	$0.0 \cdot 10^{+0}$	$4.2 \cdot 10^{-11}$	
7	$1.9 \cdot 10^{-1}$	$1.6 \cdot 10^{-1}$	$1.9 \cdot 10^{-5}$	$1.6 \cdot 10^{-2}$	$2.1 \cdot 10^{-6}$

1 MLP, clustering
2 MLP, clustering, no embeddings
4 RNN
5 Bidirectional RNN
6 Bidirectional RNN with window
7 Memory network

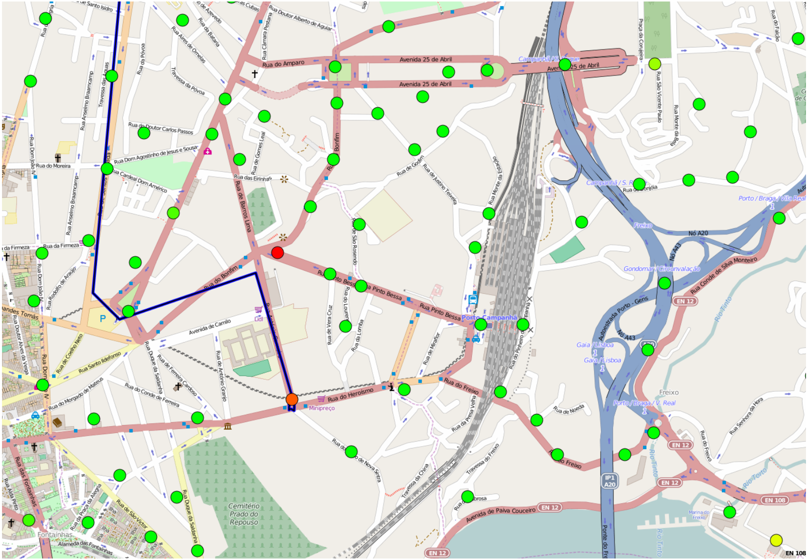
t-SNE: Quarter of Hour



t-SNE: Weeks


45
4
3
50
2
22
11
6
5
51
17
18
8
0
49
26
19
13
1
10
47
9
24
16
7
12
48
44
46
14
25
42
41
43
39
38
40
21
20
37
35
28
34
31
29
30
15
27
36
32
23
33
48
42
36
30
24
18
12
6
0
Week of the year

Softmax output



Conclusion

- Almost fully automated compared to other published solutions,
- Raw GPS points, no discretization,
- Use of meta-data,
- Little preprocessing,
- No post-processing,
- No ensembling.

Artificial Neural Networks Applied to Taxi Destination Prediction
Alexandre de Brébisson, Étienne Simon, Alex Auvolat, Pascal Vincent, Yoshua Bengio.
`https://github.com/adbrebs/taxi`