Deep Learning Applied to Destination Prediction

Étienne Simon
MILA, University of Montreal, Canada
esimon@esimon.eu

Problem

Goal
Predicting the destinations (latitude and longitude) of taxi trips based on initial partial trajectories (which we call prefixes) and some meta-information associated to each ride.

The problem has the following inputs:
- the taxi ride: a sequence of GPS positions (latitude and longitude) measured every 15 seconds. At training time we have the full trajectories (the last point is the destination), at test time we only have access to one of its prefixes.
- metadata associated to the taxi ride:
  - if the client called the taxi by phone, then we have a client ID.
  - if the client called the taxi at a taxi stand, then we have a taxi stand ID.
  - otherwise we have no client identification.
  - the taxi ID.
  - the time of the beginning of the ride.

Input: Embeddings

In order to get a distributed representation of the metadata, we use lookup tables. For example, given the ID of the taxi, the information associated with it will be represented as a 10-dimensional real valued vector.

These matrices are parameters of the model, that is, they are updated by the training algorithm.

Output: Destination Clustering

The set of points on the right is the result of a mean-shift algorithm on the destinations of the training set.

The extracted 3992 clusters centers $c_i$ are then used in our models to encode a prior on the taxi destinations.

Our models output a probability distribution $p$ on the clusters, they are used to compute a convex combination of clusters centers, which gives the final prediction:

$$\hat{y} = \sum_i p_i c_i$$

Winning Model: Multilayer Perceptron

To obtain the probability distribution $p$, this layer is then followed by a softmax:

$$p_i = \frac{\exp(W_i y)}{\sum_j \exp(W_j y)}$$

Best Model: Recurrent Neural Network

To process a sequence, we use a recurrent neural network (RNN):

$$h_t = W_m y_t + W_h h_{t-1}$$

The first way to deal with variable length sequences: just ignore them. In this approach, the trajectory prefix is solely represented by its first and last coordinates (the first 5 and last 5 GPS points).

The prefix representation $x$ and the embeddings of the metadata $m$ are fed into a layer using ReLU:

$$p = \max(0, W^x x + W^m m)$$

Results

This was a Kaggle competition for the ECML/PKDD 2015 discovery challenge, we got first place by being able to predict the destination with an average accuracy of 1.87 km.

About this project


With the support of the following organizations for research funding and computing support: Nuance Foundation, Samsung, NSERC, Calcul Quebec, Compute Canada, the Canada Research Chairs and CIFAR.