Continuous Destination Prediction Micro Service

Students: Muhammad Danish Farooq, Tyler Ho, Sicong Huang

Faculty Mentor: Robert B. Darling

Industry Sponsor: Telenav

Industry Mentor: Kumar Maddali, Changzheng Jiang, Shubham Agrawal

ELECTRICAL & COMPUTER ENGINEERING

Introduction

Project aims to design a machine learning model to predict a user's destination at the beginning of a trip using a continuous destination prediction approach. Model trains on daily routine data of user, so past behavior can be used to predict future intent.

Implemented model leverages multiple types of trip features such as locations, time, and coordinates, etc. Moreover, we also extract relations among destinations by clustering. Our model can continuously predict a driver's destination during a trip.

Another goal was to make meaningful clusters of destinations using the given trip features. By clustering destinations, it provides more information for the model. Such as if a location is a restaurant, home, workplace, etc. By doing so, we can feed this information to our model and increase prediction accuracies.

Experiment Data

- One-year's trip data of 5 different users consisting of
  - Approximately 1200 trips per user
  - Around 215 different destinations per user
- Each trip data consists of
  - Trip ID
  - Start & End location latitude and longitude (Lat-Lon)
  - Start & End location time zone
  - Start & End location name
  - Start & End time
  - Path trajectory in (Lat-Lon)

[1]: Brébisson, A.D., Simon, É., Auvolat, A., Vincent, P., & Bengio, Y. (2015). Artificial Neural Networks Applied to Taxi Destination Prediction. *CoRR, abs/1508.00021.*

Theoretical Model

Model combines research ideas from papers[1] and competitions. Provides a foundation to approach the problem through viable methods. The same model approach was followed in different steps for end results.

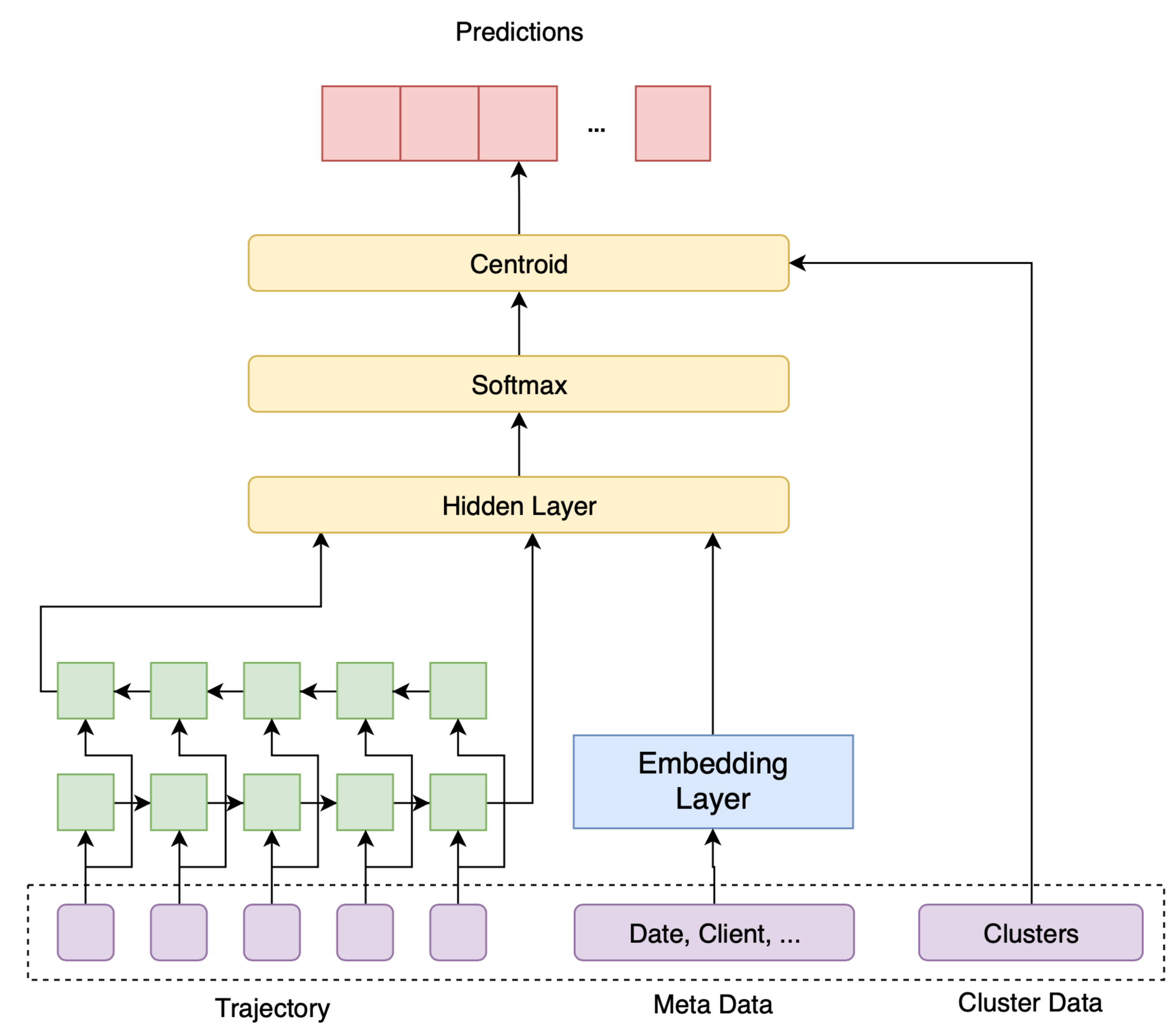

Figure 1: Graph of theoretical model for destination prediction

Base Model: Location Embedding

The base model takes start locations as input and embeds the location as a dense vector. It then predicts a probability distribution over all possible output locations

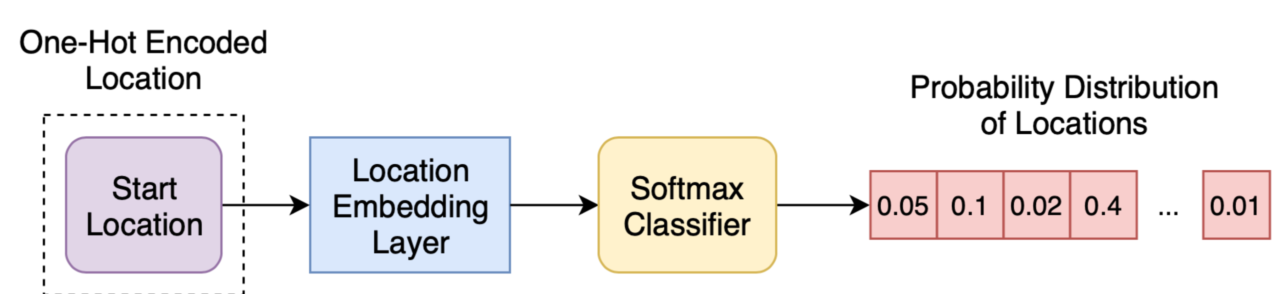

Figure 2: An illustration of the base model

Feature-Rich Model: Multi-Embedding

Multi-embedding model builds on top of base model, by including more time features as inputs. It separately embeds each time feature and makes a prediction based on the concatenation of these embeddings.

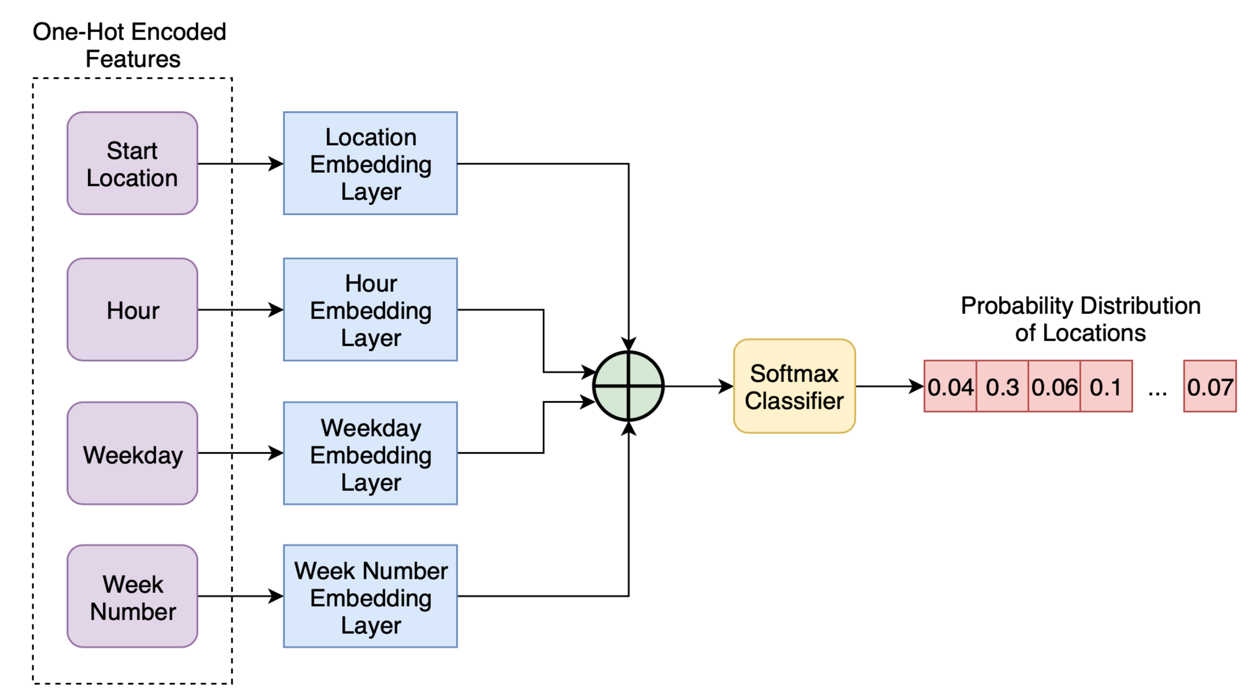

Figure 3: An illustration of multi-embedding model

Continuous Prediction: Recurrent Neural Network Model

Recurrent neural network (RNN) models the trajectory as a sequence of Lat-Lons. It takes location and time embeddings as initial state, then takes in a Lat-Lon at each time step. RNN predicts a distribution of locations for each time step. This way, as the user's current coordinate is updating, the model can continuously predict the user's destination.

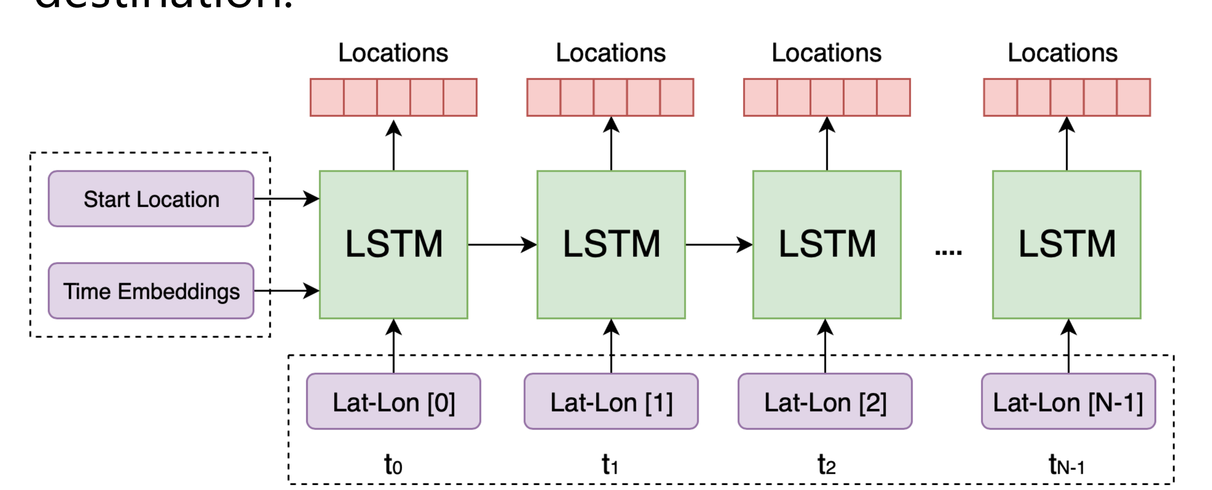

Figure 4: An illustration of recurrent neural network model

Results

%age	User 1	User 2	User 3	User 4	User 5	Average
Base Model	41.1	30.3	35.9	25.8	31.3	32.9
Double Embedding (Location & Time)	45.3	37.3	48.2	35.2	36.4	40.5
Multi Embedding	48.3	41.6	50.0	38.2	41.5	**43.9**
RNN	46.7	44.1	43.5	37.5	40.0	42.4

Table 1: Experiment results in % accuracy

Additional time features are useful as they provide 11% accuracy gain over the base model. RNN produces almost the same accuracy as same set of features are being used as in multi-embedding model.

Destination Clustering

Clustered destinations using multiple features such as frequency, end time, and idle time of trips. Used Jenk's Natural Breaks to first see if groups were formed naturally within each feature. The feature vectors were then clustered using K-Means

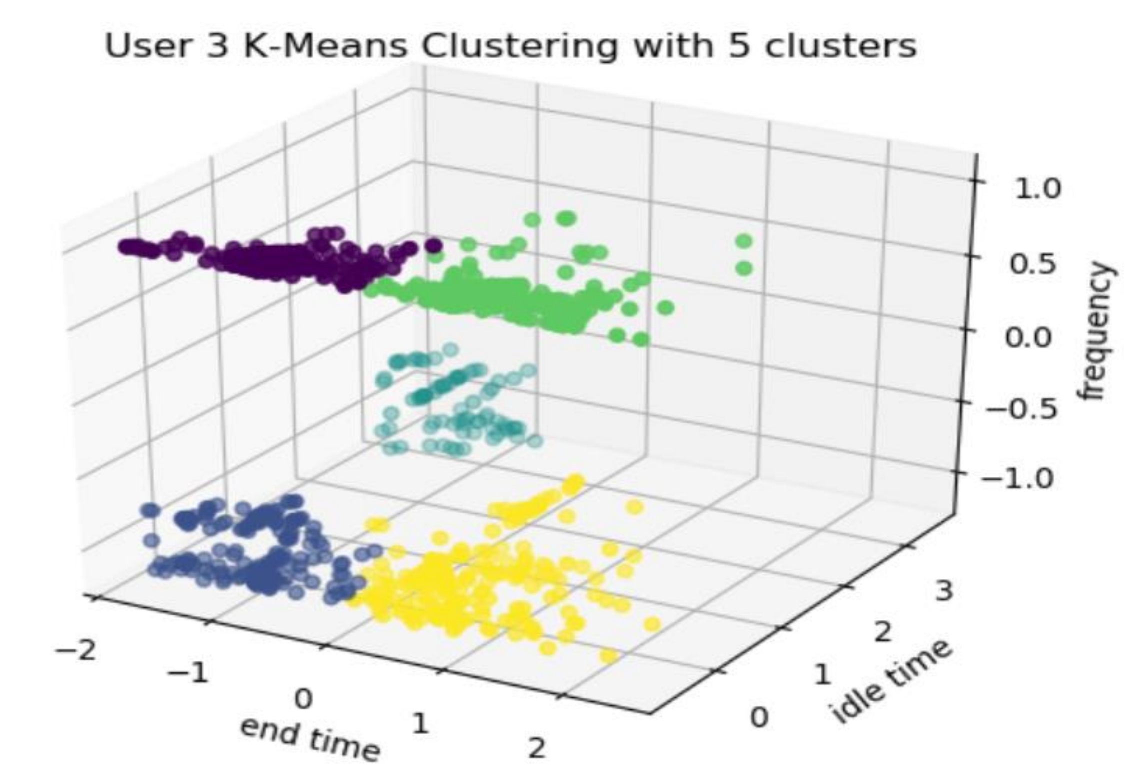

Figure 5: An example of k-means clustering with k=5

Future Improvements

- Combine destination clustering with rest of the model to get higher prediction confidence
- Experiment current model to handle time transition
- Tune model for domain transition