Summary of “**A Predictive Model for the Passenger Demand on a Taxi Network**”

(by Luis Moreira-Matias, Joao Gama, Michel Ferreira, Luis Damas)

This paper presents a methodology for creating short-term predictions for taxi demand in Porto, Portugal. By using time series forecasting, they were able to design a model that had 76% accuracy when tested on 4 months of real data.

It is mentioned that a balance is needed between passenger demand and taxi drivers’ search for customers, and when this balance does not exist there can either be an excess in vacant taxis and competition or long passenger waiting times and taxi unreliability. The paper focuses solely on the first scenario.

An assumption is made - that after dropping off a passenger, taxi drivers must return to one of the local taxi stands. Therefore, the focus is on choosing the best taxi stand that a driver should go to after a passenger drop-off, based on time and location. A few factors have been taken into account, but the paper is mostly concerned with the passenger demand for each stand over time.

The model that they came up with attempts to predict how many services will be demanded during a specific time period at each taxi stand.

Models used:

- Time varying Poisson model
- Weighted time varying Poisson model (for seasonality issues, where some taxi stands could be busier than usual due to an event, weather changes etc.)
- ARIMA: Autoregressive Integrated Moving Average Model (used for exploring other types of periodicity)
- Sliding window ensemble framework

Data preprocesssing: they developed a time series of taxi demand services aggregated for a period of P minutes.

They also observed that while the average waiting time for a taxi in a taxi stand is 42 minutes, the average duration of a service is 11 minutes and although the data collected shows regularity in the service, there is a high mobility intelligence discrepancy between taxi drivers.

Experiments:

An aggregation period of 30 minutes and a radius of 100 metres were chosen. The training set consisted of 13 weeks of data, while the test set had 3 weeks.

First, the ARIMA model was used in order to learn the underlying model. The time varying Poisson models (both weighted and non-weighted) were updated every 24 hours. For the sliding window ensemble framework, they used a window of 4 hours. The accuracy of each model was measured using a metric called sMAPE.

Results: each model showed at least 74% accuracy.