# On Predicting the Taxi-Passenger Demand:

# A Real-Time Approach

Knowledge about where the taxi request will actually emerge can be an advantage for the driver - especially when there is no economic viability of adopting random cruising strategies to find their

next passenger. The GPS historical data is one of the main variables of this topic because it can reveal underlying running mobility patterns.

This kind of data represents a new opportunity to learn/predict relevant patterns while the network is operating (i.e. in real-time).

*Can we guarantee that the taxi spatial distribution over time will always meet the demand? Even when*

*the number of running taxis already does?*

**They are focused into predicting the short-term spatiotemporal distribution of the taxi-passenger demand by using machine learning algorithms capable of learning and predicting in a real-time environment.**

**How**: using learning concepts originally proposed to a well-known online algorithm for supervised learning of binary classifiers: (functions that can decide whether an input, represented by a vector of numbers, belongs to some specific class or not) the **perceptron**.

It is a type of linear classifier, i.e. a classification algorithm that makes its predictions based on a linear predictor function combining a set of weights with the feature vector. The algorithm allows for online learning, in that it processes elements in the training set one at a time.

They propose a discrete time series framework to predict the event count (i.e. number of services) for the next P-minutes with a periodicity τ of 30 minutes. This framework handles three distinct types of memory:

1. short term (ARIMA - AutoRegressive Integrated Moving Average)
2. mid term
3. 3) long term one (both based in time-varying poisson models ).

This model presented three main contributions facing the existing literature :

1. It builds accurate predictions on a stream environment (i.e. using a real-time

test bed);

2. Part of the model is able to forget some past data by summarizing it into sufficient statistics;

3. It is able to update itself on a short amount of time 1 reusing the last real event count to learn about the novelty thereby introduced;

However, such approach presents two relevant limitations:

1. it just produces predictions each 30 minutes while the decision process is made in real-time (i.e. can we guarantee that a prediction made at 8:00am is still informative at 8:20am?);
2. the ARIMA weights are fitted (i.e. re-calculated using an offline learning process) and not updated before each prediction by reusing the entire time series of recent event counts plus the most recent one.

A complex learning model was used to build predictions about the taxi passenger demand in a real-time environment. They extended the typical definition of an ARIMA model to an incremental one using the delta rule. (a rule firstly introduced in the perceptron algorithm which is able to update its weights step by step)

This approach was tested using 2 case studies:

1. In Porto, where there’s more taxis than the demand
2. In Odivelas and Loures, which have less taxis than the demand

Their model was able to produce predictions about the spatiotemporal distribution of the demand during the next 30 minutes with a periodicity of 5 minutes.

The results demonstrated the relevance of their contribution: they maintained the aggregated error ratio lower than 24% in case A and 22% in case B. On the other hand, they were able to reduce the typical computational time used to build each prediction by 40%.