An Online Recommendation System for the Taxi Stand choice Problems

Luis Moreira-Matias
luis.matias@fe.up.pt

R. Fernandes, J. Gama, M. Ferreira, J. Mendes-Moreira, L. Damas
University of Porto, LIAAD - INESC TEC and IT - Telecommunications Institute (Portugal)

1.Problem Description

The question is: "Which is the best taxi stand to go to after a passenger drop-off?". We aim use the vehicular network communicational framework to improve their reliability by combining all drivers' experience. A clever distribution of vehicles throughout the stands will decrease the average waiting time to pick-up a passenger while the distance percussed will be more profitable. Passengers will also experience a lower waiting time to get a taxi (automatically dispatched or directly picked at a stand). This tool can improve the informed driving experience by transmitting to the driver which is the stand where 1) he will wait less time to get a passenger in; or where 2) he will get the service with the greatest revenue.

2.Goals

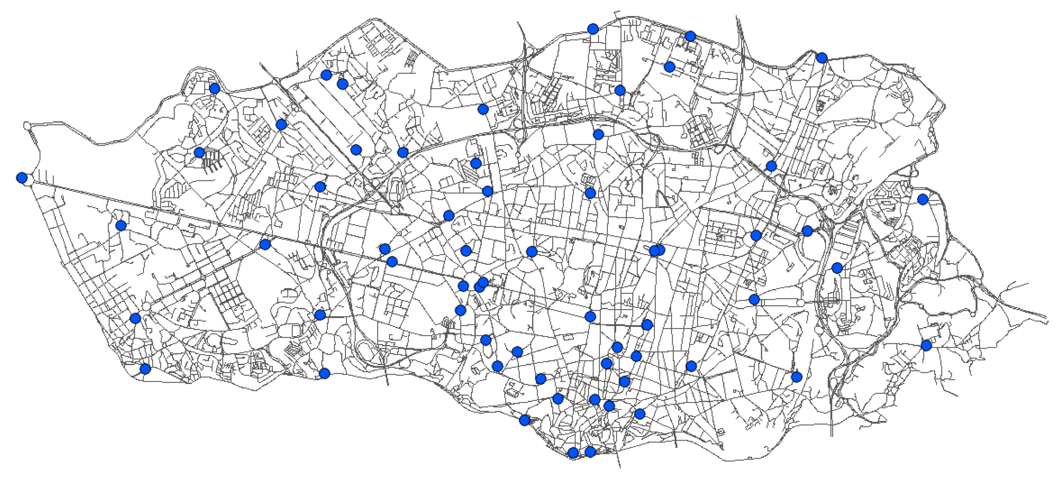
Our goal is to build an **Online Recommendation System** to improve the driver' mobility intelligence about where he can find his next passenger. This problem relies on four key variables:

- the expected price for a service over time;
- the distance/cost relation with each stand;
- how many taxis are already waiting at each stand;
- the passenger demand for each stand over time;

Based on them, we claim two contributions:

1. While the first three variables are mainly calculated, the last one relies on an **Online Prediction Model** about the spatiotemporal distribution of the passenger demand.
2. A **Recommendation Model based on a methodology to rank** the **taxi stands** according with the network status in each drop-off moment.

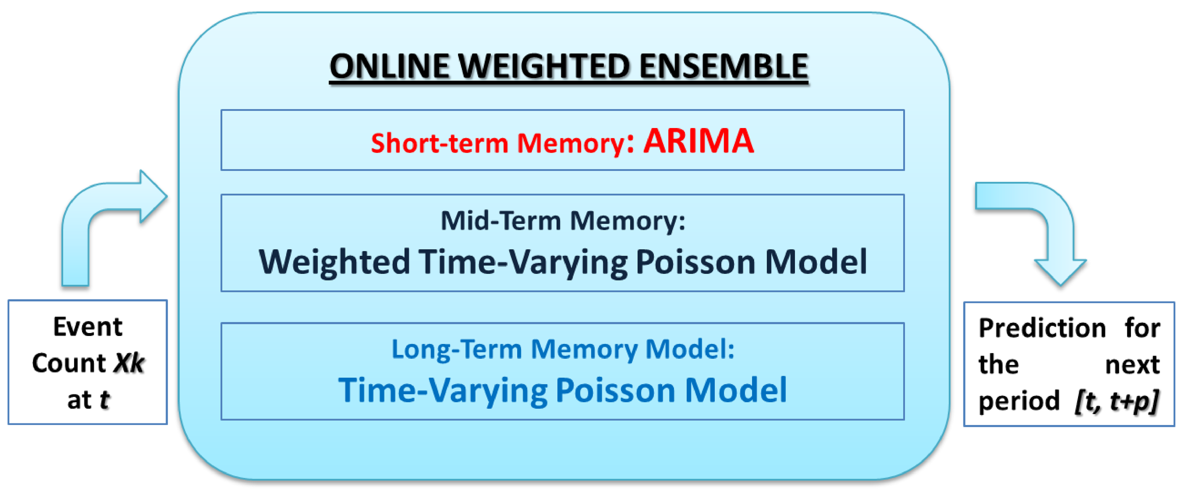
3.Case Study

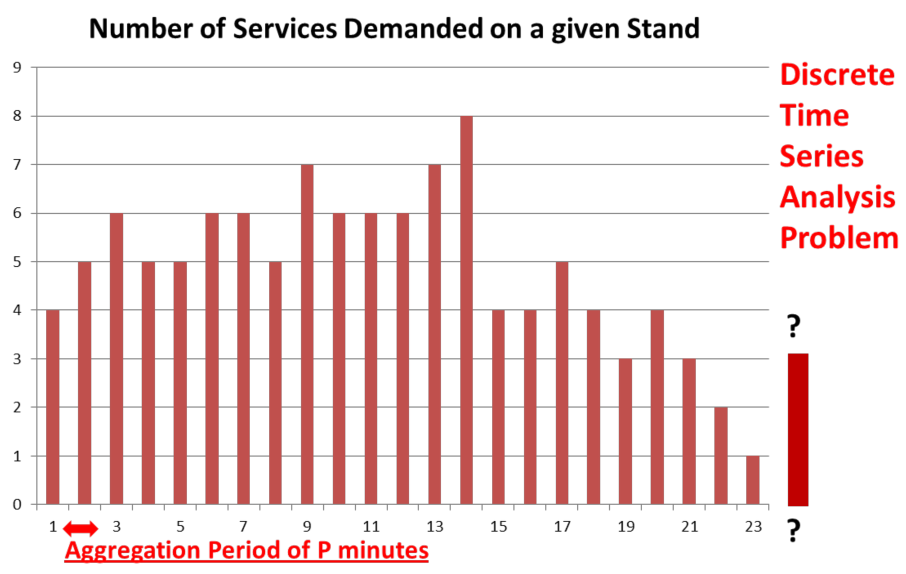


- City of Porto, Portugal. Pop.: 1.3M inhabitants
- Passenger Demand « Nr. of Vacant Taxis
- 700 taxis operating in two competitive companies(fleet A w/ 441 vehicles, fleet B w/ 250)
- **Important Regulation**: the drivers cannot run randomly searching for passengers. They are forced to head to a specific taxi stand out of the 63 existing ones in the city to wait for the next service immediately after the last passenger drop-off!
- 9 months of live data analysis w/ more than 1.3 million of taxi services;

4.Methodology

Framework Illustration





- **Time Varying Poisson Model**: We assume that the time series follows the following distribution: $P(n;\lambda) = \frac{e^{-\lambda}\lambda^n}{n!}$ where we assume that the rate λ is time-variant: $\lambda(t) = \lambda_0 \delta_{d(t)} \eta_{d(t),h(t)}$
- **Weighted Time Varying Poisson Model**: this model values more the most recent samples (i.e. the events which happen in the last Tuesday are more important than the ones that happened 3 or 4 weeks ago). The weights are calculated using the **Exponential Smoothing**: $\omega = \alpha * \{1, (1-\alpha), (1-\alpha)^2, ..., (1-\alpha)^{\epsilon-1}\}$
- **AutoRegressive Integrated Moving Average**: The **ARIMA** assumes the future value of a variable is assumed to be a linear function of several past observations and random errors. The model is automatically updated each 24h based on the auto-correlation and partial auto-correlation profiles extracted from the time series while the weights are fastly re-calculated for each prediction.
$R_{k,t} = \omega_0 + \phi_1 X_{k,t-1} + \phi_1 X_{k,t-1} + ... + \phi_p X_{k,t-p} + \varepsilon_{k,t} - \omega_1 X_{k,t-1} - \omega_1 X_{k,t-2} - ... - \omega_q X_{k,t-1}$
- **Sliding Window Ensemble**: We compute all the models independently and we use a normalized version of the error measured on a recent past fixed-sized time window to calculate a weighted mean of their output. **The Ensemble will basically choose the models by their performance!**
- **Recomendation Score**: Let w be the minutes elapsed after the last demand prediction on the instant t for the taxi-stand k, $X_{k,t+1}$ the last prediction made, $C_{k,t+w}$ the number of taxis currently parked in the stand, $L_{k,w}$ the number of services already demanded after the last prediction, $\rho_H = 1 - error$ of our model in the last H periods, v_k the distance to the stand and ε the maximum allowed distance to a taxi stand. The Recommendation Score RS_k can be obtained as following:

$$RS_k = (1 - \frac{v_k}{\varepsilon}) * ((X_{k,t+1} - C_{k,t+w} - L_{k,w}) * \rho_H) \quad (1)$$

5.Experimental results

Firstly, we evaluated our online prediction model on the live streaming data broadcasted along the vehicular network. The results are displayed on the table below.

MODEL	PERIODS			
	00->08	08->16	16->00	24h
Nr. Of Services Dem.	*110972*	*227993*	*167908*	*506873*
Poisson Mean	15,09%	19,20%	17,51%	16,84%
W. Poisson Mean	17,32%	20,66%	19,88%	18,47%
ARIMA	16,81%	18,59%	17,85%	18,51%
Ensemble	**14,37%**	**18,18%**	**17,19%**	**15,89%**

Secondly, we used the DIVERT traffic simulator to simulate the competitive scenario of our case study. We divided our services log for two fictional fleets according with the real-life proportion (A/B 70/30). While the fleet B used a common approach to the passenger-finding problem, the fleet A used our Recommendation Model. The results are displayed on the table below where WT=Waiting Time and VRD=Vacant Running Distance(kms).

Performance Metrics	A1 (RS)	B1 (common)
Average WT	38,98	40,84
Median WT	26,29	27,92
Std. Dev. WT	33,79	35,22
Average VRD	3,27	1,06
Median VRD	2,80	0,98
Std. Dev. VRD	2,53	0,54
No Service %	11,08%	19,26%

6.Conclusions

Our framework performance was promising: it predicted in real-time more than 500k taxi-services with an error of just 18%. The simulation also demonstrated that our recommendation model can be a competitive advantage in many competitive scenarios by reducing the avg. waiting time in more than 5% while compared with other fleets using the trivial passenger-finding strategies.

Funding

This work was supported by the projects DRIVE-IN: "Distributed Routing and Infotainment through Vehicular Internet-working", MISC: "Massive Information Scavenging with Intelligent Transportation Systems" and VTL: "Virtual Traffic Lights" under the Grants CMU-PT/NGN/0052/2008, MITPT/ITS-ITS/0059/2008, PTDC/EIA-CCO/118114/2010.