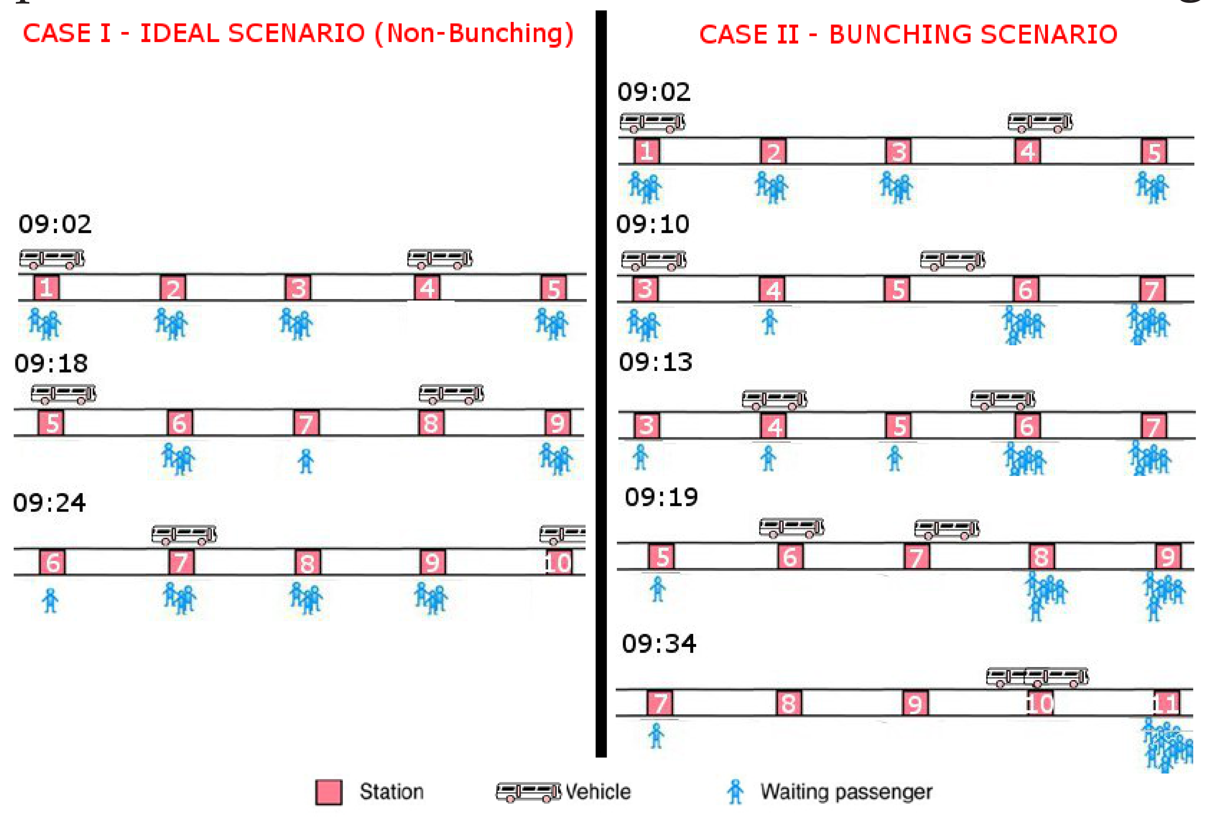
An Incremental Probabilistic Model to Predict Bus Bunching in Real-Time

Luis Moreira-Matias J. Gama, J. Mendes-Moreira, J. Freire de Sousa

luis.matias@fe.up.pt **University of Porto** and LIAAD - INESC TEC (Portugal)

1.Problem Description

In highly frequent bus routes it is more important the headway regularity than the fulfilment of the arrival time at the bus stops. In fact, a small delay of a bus provokes the raising of the number of passengers in the next stop. This number increases the dwell time and obviously, it also increases the bus' delay. On the other hand, the next bus will have fewer passengers, shorter dwell times without delays. This will continue as a *snow ball effect* and, at a further point of that route, the two buses will meet at a bus stop, forming a platoon. This phenomenon is denominated as **Bus Bunching(BB)**.



2.Goals

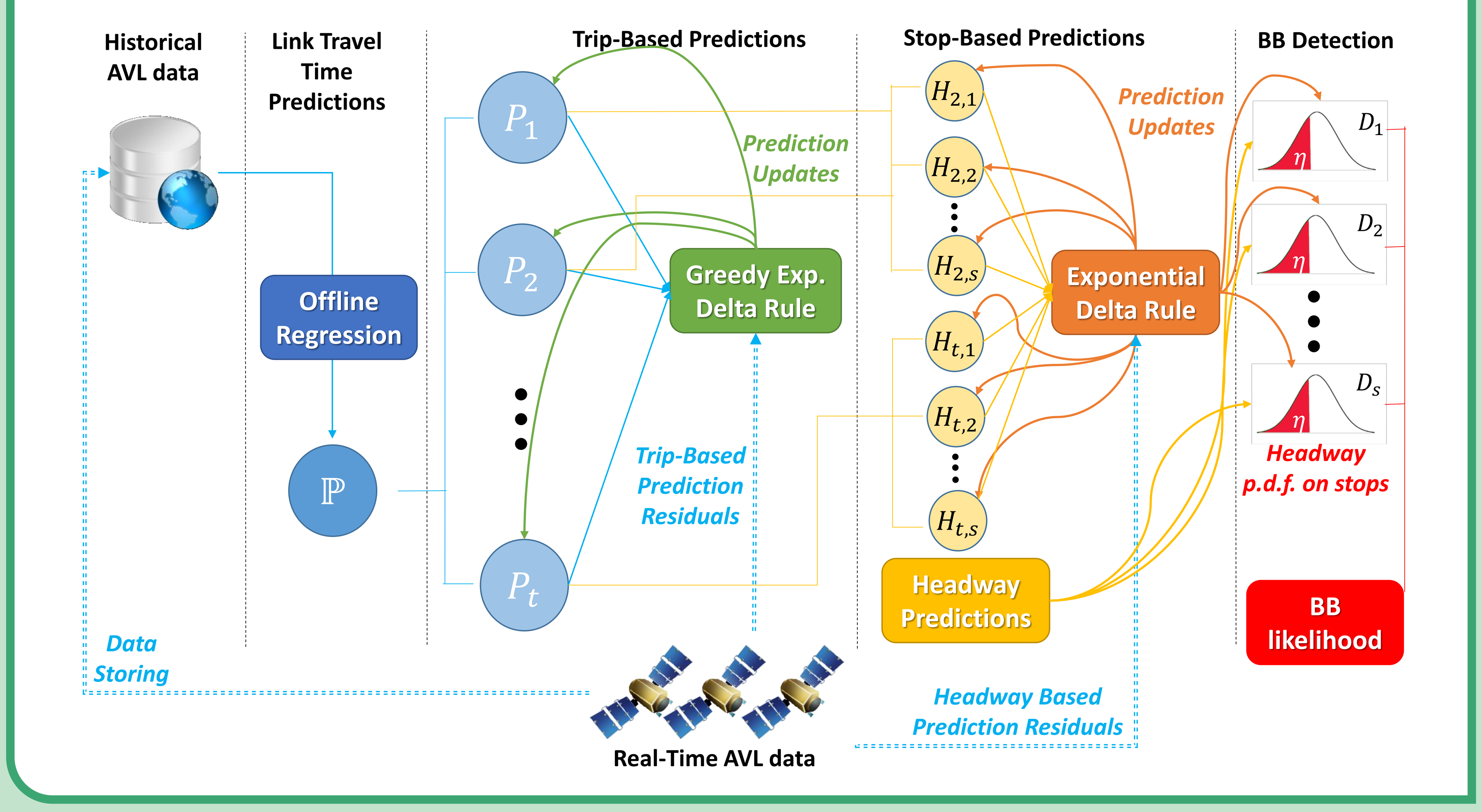
- To explore both **historical** and **real-time** Automatic Vehicle Location data by combining **Offline** and **Online Supervised Learning** techniques;
- To **predict** the *certain* short-term occurrence of such BB events in **real-time**;

3.Case Study

- City of Porto, Portugal with 1.3 million inhabitants;
- The biggest public transport operator in town (STCP) with 51 running lines;
- One year of data (2010) describing the trips of 3 lines (6 routes);
- Each data record contains a 1) route, 2) a time stamp, 3) the daytype and 4) the weekday, the respective 5) bus stop and 6) the Link Travel Time to arrive to it;

	A1	A2	B1	B2	C1	C2
Number of Trips	20598	20750	20054	19361	26739	26007
Nr. of Stops	26	26	32	32	45	45
Min. Daily Trips	44	45	56	57	65	71
Max. Daily Trips	76	76	85	84	100	101
Min. Freq.	10	11	12	13	10	10
Max. Freq.	112	100	103	120	60	60
Nr. of Trips w/ BB	682	553	437	634	1917	1702
BB Avg. Route Pos.	63.22%	74.86%	58.31%	68.54%	49.71%	53.63%

4.Methodology Illustration



5.Framework Description

Stepwise Framework:

1. To Perform **Link Travel Time (LTT) prediction** for all trips on a daily basis by employing the **Random Forests for regression** over historical trip-based data about the last 7 days;
2. These predictions are **incrementally refined** throughout the day using the averaged LTT residuals of the last trip to update all the LTT values for the next one (i.e. Greedy **Exp. Delta Rule**[1]);
3. These predictions are used to infer the future **headways** - then, they are also incrementally refined using the stop-based residuals to slightly update the predictions (i.e. **Exp. Delta Rule**[1]);
4. The headway predictions are used together with the most recent headway-based residuals to infer the **headway's *p.d.f.*** on each stop by assuming that they follow a **Gaussian** curve;
5. **A BB likelihood is calculated for each stop** based on such *p.d.f.* and combined to form a global BB *score* for the downstream stops;
6. A **BB alarm** may be triggered if this global BB likelihood goes over a certain **frequency-based** threshold;

[1] Based on the **Perceptron's Delta Rule** to update the weights on each neuron, these rule is used directly over the outputs with a dynamic α (i.e. the learning rate) - which reflects the current need of **react** the drifts on the underlying concepts (e.g. traffic jams provoking large delays on a given link).

6.Experimental Results

	A1	A2	B1	B2	C1	C2	ALL
MAE offline regression	1356.96	643.99	1475.22	1871.01	473.61	2776.57	1432.88
MAE inter-trip update	148.85	92.91	124.99	148.85	40.65	123.77	113.34
MAE incremental update	13.21	26.35	22.67	13.21	31.79	27.47	22.45
Accuracy	97.99%	96.34%	97.08%	97.83%	96.63%	93.83%	96.62%
Weighted Accuracy	93.97%	93.57%	94.57%	95.52%	95.73%	91.51%	94.14%
Precision	65.88%	40.85%	41.53%	45.70%	69.44%	51.67%	52.51%
Recall	81.81%	83.18%	83.07%	83.24%	94.48%	87.95%	85.62%
Avg. Nr. of Stops Ahead	11.85	14.78	13.88	15.01	12.96	14.52	13.83
Correct BB Predictions	558	460	363	303	1811	1497	4992
Real BB Events	682	553	437	364	1917	1702	5655

6.Discussion and Future work

By analysing the results, it is possible to conclude than the First-Order update rules are generally **improving the convergence** of our predictions to its true values. On other hand, the low Precision values shows that **our model triggers more alarms** than necessary - which it is still an acceptable behavior in the preventive context that we are facing. This work can be extended on three distinct axis:

1. on the dataset, by including a **larger dataset** containing a set of lines more representative of the entire network;
2. on the parameter setting, by conducting a **large-scale sensitivity analysis** on their values;
3. on the **corrective actions**, by proposing a method to choose **where** and **when** a action should be took to avoid BB, as well as one to choose **which** is the best one to took in each case.

Funding

This work was supported by the VTL: "Virtual Traffic Lights" (PTDC/EIA-CCO/118114/2010), by MAESTRA (ICT-2013-612944), by I-CITY - "ICT for Future Mobility" (NORTE-07-0124-FEDER-000064) and by ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and also by the Portuguese Funds through the FCT (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-037281.