Delay propagation on a suburban railway network

Guillaume Dalle[1], Yohann De Castro[2], Axel Parmentier[1]

[1] CERMICS, École des Ponts ParisTech
{guillaume.dalle,axel.parmentier}@enpc.fr
[2] Institut Camille Jordan, École Centrale de Lyon
yohann.de-castro@ec-lyon.fr

Keywords : *railway operations, train delay prediction, dynamic bayesian network, variational inference*

1 Introduction

Understanding how train delays appear and evolve is fundamental to all phases of railway operations. During planning, a delay model is necessary to build a resilient timetable, where small-scale perturbations can be absorbed instead of amplified. In the operational phase, such a model can supplement expert knowledge and provide guidance for real-time traffic regulation and passenger information.

A major challenge for delay prediction is the amount of interactions between trains: since they must share the same finite resources, one train's delay can affect many others by blocking the track or monopolizing a driver. These interactions are taken into account intuitively by railway operatives, but as the complexity of the system grows, predictive algorithms become increasingly useful to assist them. Designing a real-time delay prediction method is the purpose of the present work, which benefited from a collaboration with the French railway company SNCF.

2 The propagation model

Our approach is based on the dataset we were given, namely a record of event times (arrivals and departures) generated by a suburban railway system. This is a commonly available type of data, but it lacks the additional ingredient on which many propagation models rely: precedence relations between events. These are determined by resource conflicts, see e.g. [2]. To compensate for our ignorance of precedence relations, we identify infrastructure constraints as the main source of interactions: the basic principle is illustrated on Figure 1.

We represent these infrastructure constraints by thinking of each train's journey as a path through a directed *milestones graph*, which has one vertex for each possible event and location. This leads to the definition of the *network jam*, a hidden variable expressing the amount of congestion on the edges of the milestones graph. Its probabilistic evolution is structured as a

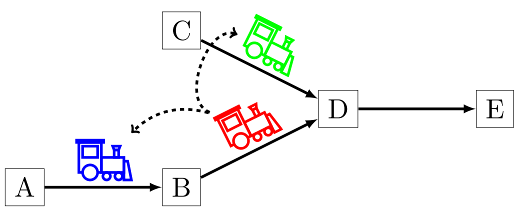

FIG. 1: Delay propagation due to infrastructure conflicts (inspired by [3])

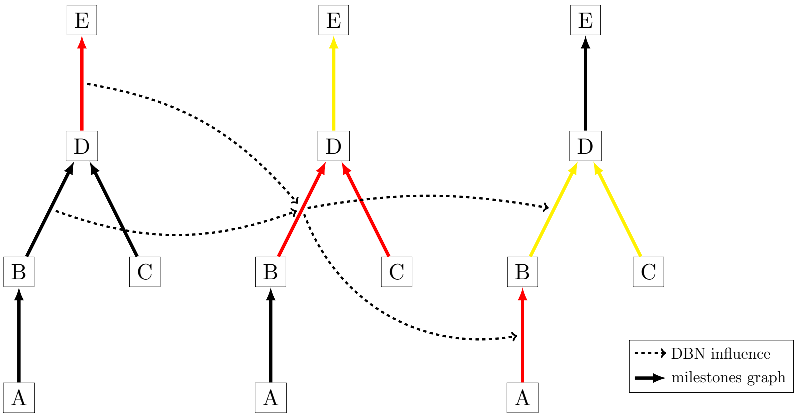

FIG. 2: The network jam and its temporal evolution

time-heterogeneous Dynamic Bayesian Network, which enables spatial propagation (as seen on Figure 2). The successive event times for each train follow a random walk influenced both by the network jam and by an individual noise term.

We first perform a statistical analysis on a simplified version of our model: a Vector AutoRegressive process of order 1 with partial and noisy observations, whose weights matrix we seek to estimate. In this setting, we exploit Fano's method and obtain a lower bound on the minimax estimation risk, i.e. the worst-case error of the best possible algorithm. This lower bound quantifies the error as a function of a few size parameters (number of edges in the network, number of days of observations, number of trains per day, etc.) and the signal-to-noise ratio.

3 Implementation and evaluation

We then propose a generic implementation that separates the expression of the model from the learning and prediction process. This is achieved by encoding the model as a probabilistic program, to which we apply Stochastic Variational Inference using the library Pyro [1].

We finally present numerical tests, both on a simulated dataset and on actual event logs. While these tests underline the computational limitations of our approach, they also demonstrate its feasibility and generality. The predictive power of the model on real data can still be improved, but preliminary results suggest that the largest benefits may be observed when significant perturbations occur, a situation which is indeed the most valuable use case of the method.

References

[1] Eli Bingham, Jonathan P. Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D. Goodman. Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*, 20(28):1–6, 2019.

[2] Francesco Corman and Pavle Kecman. Stochastic prediction of train delays in real-time using Bayesian networks. *Transportation Research Part C: Emerging Technologies*, 95:599–615, October 2018.

[3] Bernardo Monechi, Pietro Gravino, Riccardo Di Clemente, and Vito D. P. Servedio. Complex delay dynamics on railway networks from universal laws to realistic modelling. *EPJ Data Science*, 7(1):35, September 2018.