Timeseries prediction and model comparison for fleet resource management

Supervisor: Colm Connaughton (Complexity and Mathematics)

Co-supervisor(s): Roozbeh Pazuki

External partner: L&A Consultants Ltd.

Scientific background

L&A Consultants Ltd, is a London-based data analytics company specialising in providing telemetry-based fleet management software to large organisations, particularly the emergency services. Their flagship product, IR3, is a software platform which combines mapping, event processing and data visualisation tools with streaming telemetry data to allow fleet operators to monitor the state of their fleet in real time. This system produces a lot of timeseries data. If these timeseries could be used for forecasting particular quantities of interest this would be a valuable addition to the system's capability. For example, a fleet manager might wish to forecast whether the number of vehicles in the workshop for repair is going to be bigger or smaller than average over the coming days or to obtain estimates of the number of vehicles which will be available to operate from a particular depot.

Research challenge

In this project a variety of probabilistic timeseries forecasting methods will be investigated and applied to L&A's fleet data. The aim will be to keep the underlying methodology general so that the findings can be applied to different types of data within the application domain and so that more than one type of forecast model can be applied to any given task. The kind of tasks the student might expect to deal with include the following:

- Investigate and implement different timeseries forecasting methods which could be applied to fleet data. Linear Gaussian models for continuous variables are well described in [1] which would be a good place to start. We will also need models for discrete processes driven by Poisson processes which are less well known. An able student might extend the class of models to include nonlinear models based on machine-learning techniques such as neural networks.
- Define a forecast skill measure to allow different models of the same process to be compared quantitatively. The question of forecast skill has been extensively discussed in the context of meteorology and climate science. A good place to start to get a feel for the issues is [2].
- Given that different models might perform better or worse under different conditions, L&A see value in maintaining an ensemble of models. This project should design a principled method of managing an ensemble of models and updating their weights dynamically. A good place to start to get a feel for these issues is [3].

Pre-requisites

The primary skills required for this project are a good knowledge of stochastic processes and statistical analysis and strong programming skills in order to implement these methods. Ideally the project will be done in Python since this is the language of choice for data analytics in the real world. The data management aspects should be reasonably straightforward and will not require any particular specialist data analytics skills. This would be an ideal project for a student who wants a taster of the data analytics world.

Additional considerations

If the project goes well there will be a follow-on PhD project available and the possibility of an internship with L&A as part of an on-going collaboration between L&A, the Centre for Complexity Science and the Warwick Impact Fund. Since forecasting is a current business priority for the company there is every possibility for the project to have immediate real-world impact, even at the MSc level.

References

[1] J. Durbin and S. J. Koopman. *Time series analysis by state space methods*. Oxford University Press, 2012.

[2] J. Bröcker and L. A. Smith. Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, 22(2):382–388, 2007.

[3] C. R Shalizi, A. Z. Jacobs, K. L. Klinkner, and A. Clauset. Adapting to non-stationarity with growing expert ensembles. *arXiv preprint arXiv:1103.0949*, 2011.