EPSRC CDT in Distributed Algorithms

PhD Project: Using Next Generation Machine Learning and AI Techniques to Aggregate Information Pertinent to Automotive Glazing

University of Liverpool

PhD Student: Kieron McCallan
Project Partner: NSG Pilkington

Supervisors:
Leszek Gasieniec, University of Liverpool
Peter Green, University of Liverpool/Engineering Data Analytics
Graham Siddons, NSG

Project Description

Simulation techniques exist to assess how feasible it is for existing production processes to be used to make a new glazing part for a car. While instances of these simulations have been run extensively in the past, it is unlikely that the new glazing part will be identical to one that has been assessed previously. At present, optimising the parameters of the simulation to provide such assessments of new parts is labour-intensive and iteratively re-running the simulation is time-consuming. It would be highly advantageous if it were possible to identify similar historic glazing parts, identify the associated inputs to the simulations, interpolate between the associated pre-existing simulation outputs and so predict the feasibility for a new part. It would also be desirable to recommend the inputs to the simulations that, when run, would be most likely to reduce the uncertainty associated with any such interpolation. In fact, there are a variety of simulation techniques each of which has been historically applied to a different set of glazing parts and real-world data that have been used to validate these simulations. These historically simulated and real-world outputs agree in general, but it is very important to understand the disparities, eg. when the simulations predict something is feasible and, in reality, it is not. It is therefore important for any analysis to consider the similarities between these outputs as a function of the inputs.

Multi-Output Gaussian Processes can represent the uncertainty associated with interpolation and extrapolation across the historic simulation and real-world outputs. Given the articulation of uncertainty, Bayesian Optimisation can be used to tackle this problem of recommending which simulation to run next given a generic objective of understanding the historic data or a specific objective of assessing feasibility of a new glazing part. However, Gaussian Processes require a mathematical model that defines similarity, ie a way to look at the shape of two glazing parts and assess whether the parameters pertinent to assessing feasibility are similar. Estimating the parameters of a mathematical model for similarity is an example of hyper-parameter estimation. Hyper-parameter estimation can be solved using existing numerical Bayesian techniques typified by Markov Chain Monte Carlo (MCMC). MCMC is inherently designed for a single processor and is therefore likely to be slow in this context. Thankfully, there is an alternative technique, the Sequential Monte Carlo (SMC) sampler, that makes it possible to exploit parallel processing to make the same inferences as MCMC in a fraction of the time. This ensures that the scheduling of any simulations can exploit the same High-Performance Computing as the simulations themselves.

Go to the EPSRC CDT In Distributed Algorithms website.