EPSRC CDT in Distributed Algorithms

PhD Project: Scalable Gaussian Processes

University of Liverpool

PhD Student: Adam Lee
Project Partner: GCHQ

Supervisors:
Dr. Peter Green, University of Liverpool
Professor Vassil Alexandrov
Robert Allison

Project Description

Gaussian Processes (GPs) are highly flexible, non-parametric Bayesian models which attempt to infer a latent function from input data by placing a prior distribution on all possible functions which agree with the data. The flexibility and uncertainty quantification of GPs make them desirable utilities within the machine learning tool box and can be used for both regression and, less commonly, classification problems. The GP prior consists of a mean function, and a covariance function which encapsulate our prior beliefs about the latent function. The model is then trained by the data by conditioning our prior leading to a posterior distribution which is also a Gaussian Process and is fully tractable. However, full Gaussian Processes regression requires the solution of a very large linear system of equations, a task which scales cubically with the number of data we use to train the model. This limits Gaussian Processes to data-sets of size less than 10,000 points.

This project is focussed on adapting Gaussian Processes for the age of Big Data, not only in terms of scalability of the method but also in limiting the need for non-autonomous interaction. The need to specify both mean and covariance functions, which then need to be tuned such that they best represent the training data, limits a Gaussian Process' ability to learn from the data in an unsupervised manner. Recently, there has been a resurgence in research into scalable approaches for Gaussian Process regression from the wider community, both in improvements to existing algorithms and in exploitation of modern computer architecture. The scope for improving the scalability of GPs further while also integrating the possible ideas of an autonomous statistician gives way for this project.

Go to the EPSRC CDT In Distributed Algorithms website.