Active Data Selection with Faults and Changepoints

Michael A. Osborne, Roman Garnett and Stephen J. Roberts
Department of Engineering Science
University of Oxford
Oxford, OX1 3PJ
United Kingdom
{rgarnett,mosb,sjrob}@robots.ox.ac.uk

1 Introduction

We describe a Bayesian formalism for the intelligent selection of observations from sources that may intermittently undergo faults or changepoints. Such active data selection is performed with the goal of taking as few observations as necessary in order to maintain a reasonable level of uncertainty about the variables of interest. The presence of faults/changepoints is not always obvious and therefore our algorithm must first detect their occurrence. Having done so, our selection of observations must be appropriately altered. Faults corrupt our observations, reducing their impact; changepoints (abrupt changes in the characteristics of data) may require the transition to an entirely different sampling schedule. Our solution is to employ a Gaussian process formalism that allows for sequential time-series prediction about variables of interest along with a decision theoretic approach to the problem of selecting observations.

2 Gaussian process prediction in the presence of changepoints

Gaussian processes (GPs) offer a powerful method to perform Bayesian inference about functions [1]. For a function $y(x)$, the GP gives a prior distribution over its values $\boldsymbol{y}$ on a subset $\boldsymbol{x}$ of its domain that are completely specified by a mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{K}$, $p(\,\boldsymbol{y} \mid I\,) \triangleq \mathrm{N}(\boldsymbol{y}; \boldsymbol{\mu}(\boldsymbol{x}), \mathbf{K}(\boldsymbol{x}, \boldsymbol{x}))$. Here I, the *context*, includes prior knowledge of both the mean and covariance functions, which generate $\boldsymbol{\mu}$ and $\mathbf{K}$ respectively. We will incorporate knowledge of relevant functional inputs, such as x, into I for notational convenience.

An example is the squared exponential covariance function, given by $K^{(SE)}(x_1, x_2; \{\lambda, \sigma\}) \triangleq \lambda^2 \exp\left(-\frac{1}{2}(\frac{|x_1 - x_2|}{\sigma})^2\right)$. The parameters λ and σ represent respectively the characteristic *output* and *input scales* of the process. They are examples of the set of hyperparameters, collectively denoted as θ, that are required to specify our covariance and mean functions. Other covariance functions can be constructed for a wide variety of problems [1]. We have developed covariance functions that allow us to model changepoints and faults of many different types [2]. Changepoint covariances are also specified by hyperparameters, such as the location and type of each changepoint.

Note that we typically do not receive observations of y directly, but rather of potentially noise corrupted versions z of y. We consider the Gaussian observation likelihood $p(\,\boldsymbol{z} \mid \boldsymbol{y}, \theta, I\,)$, and usually take simple independent, stationary noise contributions. A sensor fault implies that the relationship between y and the observed values z is temporarily altered [2]. In order to describe such faults, we include additional hyperparameters into θ specifying their time of occurrence, duration and type. Our principled probabilistic approach will allow us to extract whatever information faulty observations may contain that is pertinent to inference about the plant process.

We define the set of observations available to us as $(\boldsymbol{x}_d, \boldsymbol{z}_d)$. Conditioning on these observations, I, and θ, we are able to analytically derive our predictive equations for the vector of function values $\boldsymbol{y}_\star$ at inputs $\boldsymbol{x}_\star$ $p(\boldsymbol{y}_\star \mid \boldsymbol{z}_d,\, \theta,\, I)$ [1]. We use the sequential formulation of a GP given by [3] to

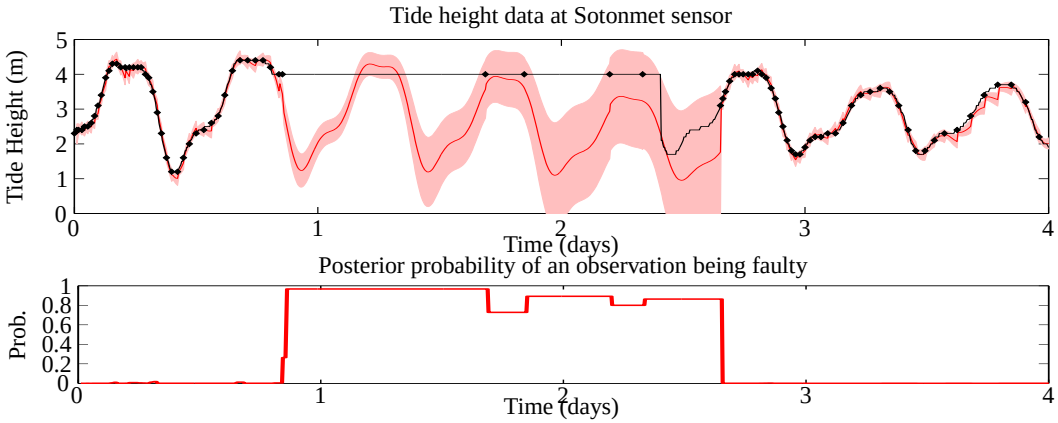

Figure 1: Active data selection over intermittently faulty tide height data from the Sotonmet sensor—above, the selected observations and consequent predictions (with lookahead $\epsilon = 5$ mins), below, the posterior probability that an observation taken at a time $t + \epsilon$ would be faulty, given all observations taken up to and including time t.

perform sequential prediction using an adaptive moving window. After each new observation, we use rank-one updates to the covariance matrix to efficiently update our predictions in light of the new information received. The computational savings made by these choices mean our algorithm can be feasibly run on-line.

Of course, we can rarely be certain about θ *a priori*. These hyperparameters must hence be assigned an appropriate prior distribution and then marginalized to give $p(\boldsymbol{y}_\star \mid \boldsymbol{z}_d, I)$. Although the required integrals are non-analytic, we can efficiently approximate them by use of Bayesian Monte Carlo [4] techniques, giving a mean $\boldsymbol{m}(\boldsymbol{y}_\star \,|\, \boldsymbol{z}_d, I)$ and covariance $\mathbf{C}(\boldsymbol{y}_\star \,|\, \boldsymbol{z}_d, I)$ for $p(\boldsymbol{y}_\star \mid \boldsymbol{z}_d, I)$. Effectively, we average over a range of models compatible with the data, giving a covariance that captures our underlying uncertainty about the correct model. This includes uncertainty about the faultiness of data. A similar approximation can also be made to evaluate the non-analytic integrals required to determine the posterior distribution for any hyperparameter [2].

3 Active data selection in the presence of changepoints

Consider the decision about sampling facing us at an arbitrary time t. Our goals in making this decision are two-fold. Firstly, we aim to minimise the uncertainty we possess about variables of interest $\boldsymbol{y}_\star \triangleq \{y_l, l = 1, \ldots, L\}$, where y_l is the value of our field at the location of sensor l for the lookahead-shifted time $t + \epsilon$. Secondly, we aim to minimise the number of costly observations. We can now define the expected loss associated with choosing an observation from sensor l_c as $\Lambda(l_c \mid \boldsymbol{z}_d, I) \triangleq \int \left(\sum_{l=1}^{L} \sqrt{\mathbf{C}(y_l \,|\, z_c, \boldsymbol{z}_d, I)} \right) p(z_c \mid \boldsymbol{z}_d, I) \, \mathrm{d}z_c + C_c$, where we have marginalised over the possible observations z_c received from sensor l_c. This integral is non-analytic, and as such will be approximated by further application of Bayesian Monte Carlo techniques. The term C_c denotes the cost of the observation. We also have the loss associated with not taking an observation at all $\Lambda(\emptyset \mid \boldsymbol{z}_d, I) \triangleq \sum_{l=1}^{L} \sqrt{\mathbf{C}(y_l \,|\, \boldsymbol{z}_d, I)}$. So, defining $l_m \triangleq \operatorname{argmin}_{l_c = 1, \ldots, L} \Lambda(l_c \mid \boldsymbol{z}_d, I)$, if $\Lambda(l_m \mid \boldsymbol{z}_d, I) < \Lambda(\emptyset \mid \boldsymbol{z}_d, I)$, we sample at l_m; otherwise, we do not sample at all.

4 Results

We have applied our methods to data drawn from two sensor networks. Selected observations, plotted as black diamonds, were used to generate predictions, plotted using a red line for the mean and pink for the ± 1 SD error bars. The full list of all available observations serves as a measure of "ground truth" (in the absence of faults), plotted as a black line.

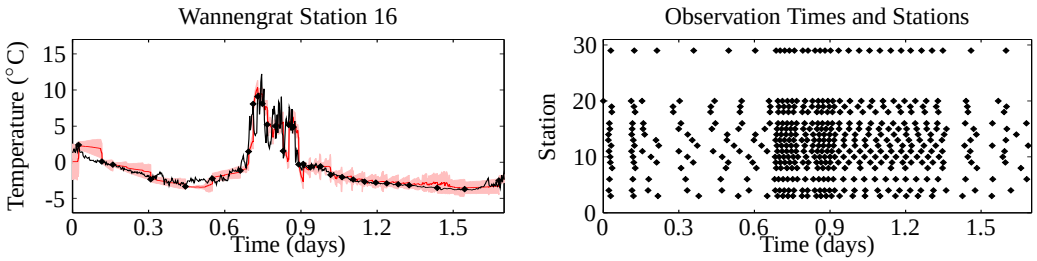

Figure 2: Active data selection of ambient temperatures at 16 Wannengrat sensor stations for $\epsilon = 0$.

4.1 Bramblemet weather sensor network

We tested our algorithm on a network of weather sensors located on the south coast of England [3]. In particular, we performed active data selection over tide height data in which readings from a sensor became stuck at an incorrect value. As such, we used a model that allowed for such changes in the observation likelihood. Results are plotted in Figure 1. Here, the run length is the number of days since the last changepoint. Our model correctly identified the beginning and end of the fault, and schedules more observations as it begins to suspect that the fault may have ended.

4.2 Wannengrat weather sensor network

Our second dataset was drawn from the Wannengrat Alpine Observatory being built above and around the town of Davos as part of Swiss Experiment (see `http://www.swiss-experiment.ch/index.php/Wannengrat:Home`). The changepoints in output scale that exist within this dataset , along with the limited battery life of the remote sensors, make it a natural fit for our methods. Figure 2 demonstrates active data selection over a 16-sensor ambient temperature dataset. Note that the algorithm becomes more reluctant to make simultaneous observations from multiple sensors subsequent to learning that their readings are strongly correlated. It can be seen that there is a dramatic increase in sampling frequency coincident with the volatile fluctuations in temperature (registered at all sensors) that begin at about $t = 0.7$ days.

5 Conclusion

We have introduced a new sequential algorithm for performing active data selection in the presence of changepoints or faults. We employ a principled, Bayesian framework throughout, performing prediction using a Gaussian process with an appropriate changepoint covariance. We then used those predictions to evaluate the optimal decision-theoretic sampling policy, given the goal of minimising the average uncertainty for a defined cost of taking an observation. Tests on real sensor networks demonstrate the efficacy of our approach.

References

[1] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[2] R. Garnett, M. A. Osborne, S. Reece, A. Rogers, and S.J. Roberts. Sequential Bayesian prediction in the presence of changepoints and faults. Technical report, University of Oxford, 2009. available at `http://www.robots.ox.ac.uk/~mosb/PARG0901.pdf`.

[3] M. A. Osborne, A. Rogers, S. Ramchurn, S. J. Roberts, and N. R. Jennings. Towards real-time information processing of sensor network data using computationally efficient multi-output Gaussian processes. In *International Conference on Information Processing in Sensor Networks 2008*, pages 109–120, April 2008.

[4] C. E. Rasmussen and Z. Ghahramani. Bayesian Monte Carlo. In S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15. MIT Press, Cambridge, MA, 2003.