Curiosity simulation

Franziska Brändle, Charley M. Wu & Eric Schulz

March 2020

1 Problem statement

We want to explore a target function $f(x)$ on some bounded set $\mathcal{X}$, which we will take to be a subset of $\mathbb{R}^D$ by using a Gaussian Process ($\mathcal{GP}$) as a probabilistic model for $f(x)$ and then exploiting this model via an acquisition function to decide where in $\mathcal{X}$ to evaluate the function next. Our goal is to assess how different acquisition functions differ in their mean confidence ranks over their sampled input points.

2 Gaussian Process

Let $f(\boldsymbol{x})$ be a function mapping an input $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$ to an output y. A $\mathcal{GP}$ defines a distribution $p(f)$ over such functions. A $\mathcal{GP}$ is parametrized by a mean function $m(\boldsymbol{x})$ and a kernel function, $k(\boldsymbol{x}, \boldsymbol{x}')$:

$$m(\boldsymbol{x}) = \mathbb{E}\left[f(\boldsymbol{x})\right] \tag{1}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}\left[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))\right] \tag{2}$$

At time t, we have collected observations $\mathbf{y}_{1:t} = [y_1, y_2, \ldots, y_t]^\top$ at inputs $\boldsymbol{x}_{1:t} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$. For each outcome y_t, we assume $y_t = f(\boldsymbol{x}_t) + \epsilon_t$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Given a $\mathcal{GP}$ prior on the functions $f(\boldsymbol{x}) \sim \mathcal{GP}\left(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right)$, the posterior over f is a $\mathcal{GP}$ with

$$m_t(\boldsymbol{x}) = k_{1:t}(\boldsymbol{x})^\top (\boldsymbol{K}_{1:t} + \sigma^2 \boldsymbol{I}_t) \boldsymbol{y}_{1:t} \tag{3}$$

$$k_t(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{k}_{1:t}(\boldsymbol{x})^\top (\boldsymbol{K}_{1:t} + \sigma^2 \boldsymbol{I}_t)^{-1} \boldsymbol{k}_{1:t}(\boldsymbol{x}') \tag{4}$$

where $\boldsymbol{k}_{1:t}(\boldsymbol{x}) = [k(\boldsymbol{x}_1, \boldsymbol{x}), \ldots, k(\boldsymbol{x}_t, \boldsymbol{x})]^\top$, $\boldsymbol{K}_{1:t}$ is the positive definite kernel matrix $[k(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1,\ldots,t}$, and $\boldsymbol{I}_t$ is a t by t identity matrix.

3 Simulation details

We use the Radial Basis Function (RBF) kernel as a component of the GP function learning algorithm, which specifies the correlation between inputs.

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{\lambda}\right) \tag{5}$$

We sample a target function $f^* \sim \mathcal{GP}(0, k(x, x'))$ over a discretized input space from -5 to 5 with equal distance between each points. We corrupt the signal of this function with a noise of $\sigma = 0.1$, such that

$$f^*(x) + \epsilon, \ \epsilon \sim \mathcal{N}(0, \sigma) \tag{6}$$

We then use a Gaussian Process to explore the sampled function by sampling options based on different acquisition functions. For all simulations, we set the length-scale of both the generating and the sampling RBF kernel to $\lambda = 1$.

4 Acquisition functions

We compare three different acquisition functions: novelty-based sampling, complexity approximations, and upper confidence bound sampling.

4.1 Novelty-based sampling

Novelty-based (NB) sampling picks the next point that currently has the highest predictive uncertainty

$$\mathrm{NB}(\boldsymbol{x}) = \sigma(\boldsymbol{x}). \tag{7}$$

NB sampling treats uncertainty as positive and samples the option it is currently the most uncertain about. This mimics novelty-based approaches to curiosity.

4.2 Complexity approximation

Our implementation of complexity approximation (CA) sampling, tries to maximize the average expected reduction in predictive error

$$\mathrm{CA}(\boldsymbol{x}) = \int_{\mathcal{D}} \Delta\sigma(\boldsymbol{x}) \,\mathrm{d}y, \tag{8}$$

which we approximated by summing up the current posterior standard deviation over all options and then comparing it to the same sum when new observations points were Monte Carlo-sampled for the evaluated input.

4.3 Upper confidence bound sampling

Upper confidence bound (UCB) sampling picks the next point that currently has the highest upper confidence bound

$$\mathrm{UCB}(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \beta\sigma(\boldsymbol{x}). \tag{9}$$

UCB sampling is an optimistic strategy that samples based on an explicit exploration-exploitation trade-off, and has been proven to produce sub-linear regret. We set the exploration parameter to $\beta = 3$ to mimic highly curious agents.

5 Simulation details

We let each acquisition function explore 10,000 target functions over 10 trials each. Thus, there are 10,000 simulation runs for each acquisition function. On each trial, a softmax choice rule transforms each model's valuations into a probability distribution over options:

$$p(\mathbf{x}) = \frac{\exp(q(\mathbf{x})/\tau)}{\sum_{j=1}^{N} \exp(q(\mathbf{x}_j)/\tau)}, \tag{10}$$

where $q(\mathbf{x})$ is the predicted value of each option $\mathbf{x}$ for a given acquisition function (e.g., $q(\mathbf{x}) = \mathrm{UCB}(\mathbf{x})$ for UCB), and τ is the temperature parameter. We set $\tau = 0.001$ to mimic minimally noisy sampling.

For every trial per simulation, we rank all options based on their current uncertainty, measured by $\sigma(x)$, and then track the average confidence rank, which is the opposite of an options uncertainty rank. After each run, we calculate the average confidence ranks over trials, leading to a collection of 10,000 such mean ranks for each acquisition function.