Sampling efficiency

The efficiency of acquisition functions is influenced by many factors, e.g., the type of optimized function, the distribution of training data, the fitness of Kriging model and data noises, computational budgets etc. Therefore, the efficiency of different acquisition functions is comparable only for the case of a given the optimized function. There are four ways to compare the efficiency among different acquisition functions by offsetting the effect of the distribution of initial training data, which are illustrated in Exampl.

Example

Take three methods as an example for illustrating, viz, Expected Improvement (EI) method, Expected improvement with "plugin" (EI_plugin) method and Reinterpolation Expected Improvement (REI) method. Given a noise contained function, e.g., the one-dimensional polynomial function $y = 0.013x^4 - 0.25x^3 + 1.61x^2 - 4.1x + 8 + \mathrm{N}(0,0.1)$ in the range of $x \in [0,11]$ with known global minimum $\mathrm{y}(x = 7.7) = 3.4528$, global maximum $\mathrm{y}(x = 11.0) = 15.2930$ and defined domain of $x \in [0,11]$. The kriging model is selected as the default model with 0.1 noise level in Bgolearn Package in iterations.

(1) Trail path

N trails of optimization are done by the three different acquisition functions with randomly sampled initial training data. At each trail N_i, we randomly select 8 discrete data at the range of $x \in [0,11]$ with step of 0.1, and added new data continuously under the guide of acquisition functions. The iteration goes on until the optimal $x = 7.7$ is picked out, and the total number of iterations at trail N_i is recorded as n_i. After $N = 100$ trails, we derived 100 individual n_i and two statistical variables, viz., mean and variance of trails are calculated as $\bar{n} = \frac{1}{N}\sum_{i=1}^{N} n_i$ and $s^2 = \frac{\sum_{i=1}^{N}(n_i - \bar{n})^2}{N-1}$. The acquisition with the minimum $\bar{n}$ and s^2 is considered the most efficient. The results of the three methods are listed in Table E1.

Table E1 $\bar{n}$ and s^2 of 100 trails

function	$\bar{n}$	s^2
EI	4.71	3.91

EI_plugin	4.54	3.75
REI	4.28	2.49

By comparing trail paths, REI shows the best attribute among the three.

(2) Opportunity Cost

The opportunity cost (OC) value is defined as the difference between the global optimal and current optimal of the function values (Balachandran et al., 2016), viz., OC = $y_{min} - \mathrm{y}(x = 7.7)$, where y_{min} is the minimum function value of training dataset and $\mathrm{y}(x = 7.7) = 3.4528$ is the global optimal of the optimized function. If the optimal value is derived, there has OC = 0. The normalized opportunity cost is used usually as $\widehat{OC} = \frac{y_{min} - \mathrm{y}(x=7.7)}{\mathrm{y}(x=11.0) - \mathrm{y}(x=7.7)}$, where $\mathrm{y}(x = 11.0) = 15.2930$ is the global maximum. If the search space is too large to be explored, $\widehat{OC}$ will coverage to a certain value but seldom be zero.

N trails of optimization are done by the three different acquisition functions with randomly sampled initial training data. At each trail N_i, we randomly select 8 discrete data at the range of $x \in [0,11]$ with step of 0.1, and added new data continuously under the guide of acquisition functions. The iteration goes on until $\widehat{OC} = 0$, and the variation path of $\widehat{OC}$ at trail N_i is recorded as $\widehat{OC}_i$. After $N = 10$ trails, each path $\widehat{OC}_i$ with iteration steps at trail i are plotted as shown in Figure E1. The coverage area of each path $\widehat{OC}_i$ is denoted as S_i, the smaller the average coverage area $\bar{\mathrm{S}} = \frac{1}{N}\sum_{i=1}^{N} S_i$ is, the better effective of acquisition function has.

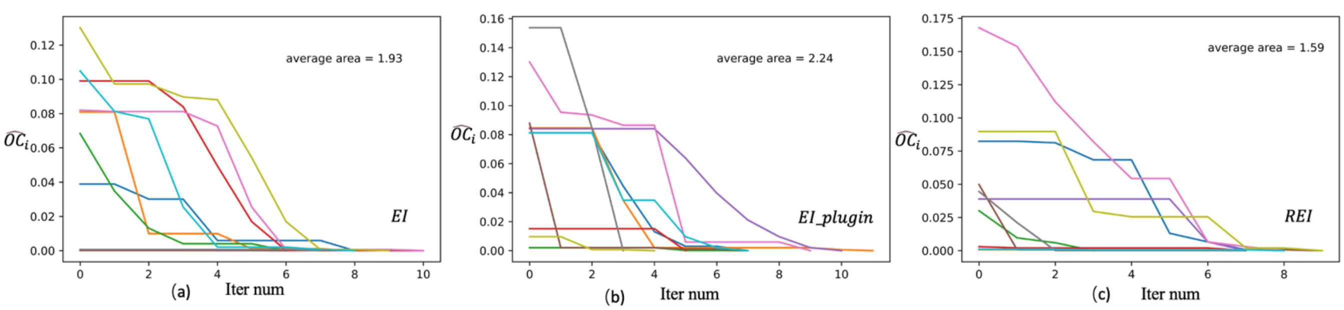

Figure E1 $\widehat{OC}_i$ with iteration steps. (a) by Expected Improvement (EI) method, (b) by Expected improvement with "plugin" (EI_plugin) method (c) by Reinterpolation Expected Improvement (REI) method.

By comparing the average coverage area, REI shows the best attribute among the three.

(3) Probability density function

For each specific trial N_i, the difference of iteration number between a compared acquisition function and a reference acquisition function is calculated as $\Delta_i = n_i^{ref} - n_i$, where n_i^{ref} and n_i are the number of iterations at trial i guide by the reference and comparation acquisition function respectively. When Δ_i=0, the acquisition function has same efficient with the reference one, $\Delta_i < 0$ means that the compared acquisition function needs more iteration steps and has less efficiency than the reference one, vice versa.

In this example, we choose EI as the reference acquisition and $N = 100$ trails of optimization are done by the three different acquisitions functions with randomly sampled initial training data. At each trail N_i, we randomly select 8 discrete data at the range of $x \in [0,11]$ with step of 0.1, and added new data continuously under the guide of acquisition functions. The iteration goes on until the optimal $x = 7.7$ is picked out, and the number of iterations at trail N_i is recorded as n_i^{ref}, $n_i^{EI\_p}$ and n_i^{REI}, respectively. Hereafter, the Δ_i value are counted, which are yielded in $\Delta_i^{EI_p} = n_i^{ref} - n_i^{EI_p}$ and $\Delta_i^{REI} = n_i^{ref} - n_i^{REI}$ for i =1,…100, respectively. After N trails, we derived 100 samples of $\Delta_i^{EI\_p}$ and Δ_i^{REI}. The distribution of $\Delta_i^{EI\_p}$ and Δ_i^{REI} are estimated through bin method and drawn in Figure E2.

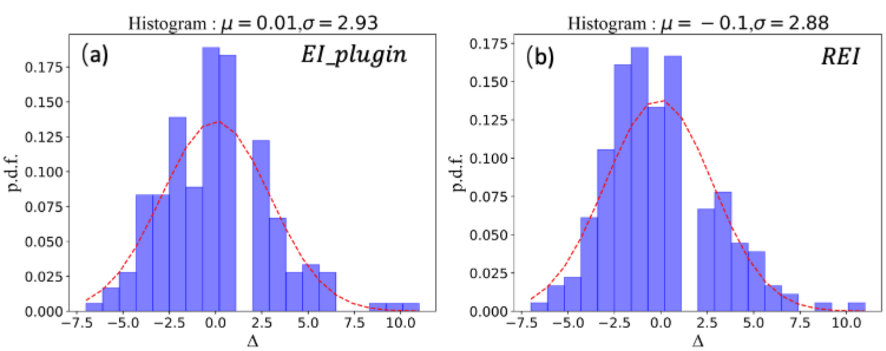

Figure E2 Probability density distribution of (a) $\Delta_i^{EI\_p}$ and (b) Δ_i^{REI}, estimated by bins=20.

The probabilities of $P\left(\Delta_i^{\mathrm{EI\_P}} > 0\right) = 0.501$ and $P(\Delta_i^{REI} > 0) = 0.486$ reflect how much better of EI_p and REI than reference EI method on efficiency. By comparing the probabilities, EI_plugin shows the best attribute among the three.

(4) Count Strategy

The efficient of acquisition functions are compared by the qualified frequency at a finite optimization step. At each trail N_i, we randomly select 8 discrete data at the range of $x \in [0,11]$ with step of 0.1, and added new data continuously under the guide of acquisition functions. The iteration goes on until attained the maximum iteration threshold of T=4 or pick out the optimal $x = 7.7$. If trail N_i find out the optimal $x = 7.7$ within four iterations, we count trial N_i as a success trial, vise as a failed trial. After $N = 100$ trails, the number of success trials of 100 trails are listed in Table E2.

Table E2 The number of success trials of 100 trials

function	success	failed
EI	50	50
EI_plugin	58	42
REI	61	39

By comparing the number of success trials, REI shows the best attribute among the three.

Reference:

Balachandran, P. V., Xue, D., Theiler, J., Hogden, J., and Lookman, T. (2016). Adaptive strategies for materials design using uncertainties. Sci. Rep. 6, 19660. doi:10.1038/srep19660.