

УДК 519.6

Адаптация функции приобретения в байесовской оптимизации проектных решений

Агасиев Талех Азер оглы

agtaleh@mail.ru

МГТУ им. Н.Э. Баумана, Москва, Россия

При решении прикладных задач параметрической оптимизации проектных решений широко применяют алгоритмы байесовской оптимизации. Использование суррогатных моделей в этих алгоритмах позволяет уменьшить число обращений к вычислительно затратной целевой функции за счет поиска перспективных значений ее аргументов. Перспективные значения аргументов выбирают на каждой итерации байесовской оптимизации по результатам решения задачи глобальной оптимизации некоторой функции приобретения, вычисляемой по прогнозируемым суррогатной моделью значениям целевой функции. Выбор типа функции приобретения и значений ее параметров может значительно влиять на эффективность решения оптимизационных задач. Актуальной является проблема автоматизации выбора и настройки функции приобретения алгоритма байесовской оптимизации в процессе решения прикладных задач. В работе предложен новый критерий соответствия функции приобретения ландшафту целевой функции. Приведены результаты вычислительных экспериментов, указывающие на повышенную эффективность байесовской оптимизации с использованием предложенного критерия.

Ключевые слова: оптимизация проектных решений, байесовская оптимизация, суррогатная модель, гауссовский процесс, функция приобретения

Введение. При решении прикладных задач параметрической оптимизации проектных решений широко применяют алгоритмы байесовской оптимизации, использующие суррогатные модели целевой функции [1]. Время решения прикладных задач оптимизации в значительной степени зависит от числа обращений к вычислительно сложной целевой функции. Использование суррогатных моделей позволяет уменьшить это число за счет приближенной оценки значений целевой функции в процессе поиска перспективных значений ее аргументов. Суррогатные модели на основе гауссовского процесса каждому вектору значений аргументов целевой функции ставят в соответствие нормальное распределение приближенных значений целевой функции. Перспективные значения аргументов целевой функции выбирают по результатам решения задачи глобальной оптимизации некоторой *функции приобретения* [2]. Функция приобретения формализует степень перспективности вектора значений аргументов целевой функции на основе параметров нормального распределения ее прогнозируемых значений.

Функции приобретения в байесовской оптимизации. Рассматриваем задачу глобальной непрерывной параметрической оптимизации в следующей постановке:

$$\min_{X \in D_X} f(X) = f(X^*) = f^*,$$

где $X = (x_1, \dots, x_{|X|})$ — $|X|$ -мерный вектор варьируемых параметров (размерность задачи); $D_X \subset \mathbb{R}^{|X|}$ — область поиска; $f(X) \in \mathbb{R}^1$ — целевая функция; X^* , f^* — искомые оптимальный вектор и значение целевой функции.

При использовании суррогатных моделей $\hat{f}(X)$ на основе гауссовских процессов каждому вектору X ставят в соответствие нормальное распределение предсказанных значений целевой функции $N(\mu_{\hat{f}}(X), \sigma_{\hat{f}}^2(X))$. На каждой итерации r алгоритма байесовской оптимизации выбирают перспективный вектор X_{r+1} следующим образом:

$$X_{r+1} = \arg \max_{X \in D_X} A(p, \hat{f}_r(X)),$$

где $A(p, \hat{f}_r(X))$ — функция приобретения; p — свободный параметр функции приобретения; $\hat{f}_r(X)$ — суррогатная модель, построенная на итерации r по обучающей выборке $\{X_k, f_k\}_{k=1}^r$.

Широко известны следующие функции приобретения [2].

Функция Lower Confidence Bound (LCB)

$$A_{LCB}(p, \hat{f}_r(X)) = -\mu_{\hat{f}}(X) + p \cdot \sigma_{\hat{f}}(X),$$

где $p \geq 0$ — параметр функции приобретения, определяющий вклад оценки достоверности прогноза на оценку перспективности вектора X .

Функция Probability of Improvement (PI)

$$A_{PI}(p, \hat{f}_r(X)) = \Phi\left(\frac{f_r^* - \mu_{\hat{f}}(X) - p}{\sigma_{\hat{f}}(X)}\right),$$

где $\Phi(\cdot)$ — кумулятивная функции распределения; f_r^* — лучшее вычисленное значение целевой функции на предыдущих итерациях $k \in [1:r]$; $p \geq 0$ — параметр функции приобретения, позволяющий повысить приоритет векторов X , для которых высока вероятность улучшить текущее значение f_r^* хотя бы на величину p .

Функция Expected Improvement (EI)

$$z_{\hat{f}}(p, X) = f_r^* - \mu_{\hat{f}}(X) - p,$$

$$A_{EI}(p, \hat{f}_r(X)) = z_{\hat{f}}(p, X) \cdot \Phi\left(\frac{z_{\hat{f}}(p, X)}{\sigma_{\hat{f}}(X)}\right) + \sigma_{\hat{f}}(X) \cdot \phi\left(\frac{z_{\hat{f}}(p, X)}{\sigma_{\hat{f}}(X)}\right),$$

где $\Phi(\cdot)$ — кумулятивная функция распределения, $\phi(\cdot)$ — функция плотности вероятности; f_r^* — лучшее вычисленное значение целевой функции на предыдущих итерациях $k \in [1:r]$; $p \geq 0$ — параметр функции приобретения, аналогичный параметру функции $A_{PI}(p, \hat{f}_r(X))$.

Для перечисленных функций уменьшение значения параметра p приводит к локализации поиска. В литературе также предложены другие функции приобретения [3, 4]. Сравнительный анализ эффективности различных функций приобретения выполняют по результатам вычислительных экспериментов с решением тестовых задач оптимизации. Однако нет оснований полагать, что выбранная функция приобретения окажется эффективной на задачах отличных от тех, которые использовались в вычислительных экспериментах. Авторы исследований также рекомендуют менять значения параметров выбранной функции приобретения в процессе решения задачи. В связи с этим актуальной является разработка методики адаптивного выбора типа функции приобретения и значений ее параметров в процессе решения прикладных задач оптимизации.

Предлагаемый метод адаптации функции приобретения. На итерации r алгоритма байесовской оптимизации определяем наилучшую функцию приобретения A_r^* путем максимизации предлагаемого критерия *соответствия приобретения* — Improvement Preservation (IP):

$$\max_{A_r \in D_A} IP(A_r) = IP(A_r^*),$$

$$IP(A_r) = \frac{\sum_{i=1}^N \sum_{j=i+1}^N \max(I_r^i, I_r^j) \cdot s_r^{i,j}}{\sum_{i=1}^N \sum_{j=i+1}^N \max(I_r^i, I_r^j)},$$

$$s_r^{i,j} = \begin{cases} 1, & \text{если } A_r^i < A_r^j \text{ и } I_r^i < I_r^j, \\ 1, & \text{если } A_r^i > A_r^j \text{ и } I_r^i > I_r^j, \\ 1, & \text{если } A_r^i = A_r^j \text{ и } I_r^i = I_r^j, \\ 0, & \text{иначе,} \end{cases}$$

где N — число точек в тестовой выборке $\{X_i, f_i\}_{i=1}^N$, A_r^i — значение функции приобретения A_r в точке X_i , вычисленное на основе прогноза модели $\hat{f}_r(X)$, которая построена по обучающей выборке $\{X_k, f_k\}_{k=1}^r$; I_r^i — значение «истинной» функции приобретения в точке X_i , которое оцениваем как прирост значения целевой функции относительно лучшего известного значения f_r^* на данной итерации r :

$$I_r^i = I_r(X_i) = \begin{cases} \frac{f_r^* - f_r(X_i)}{\max(|f_r^*|, 1)}, & \text{если } f_r(X_i) < f_r^*, \\ 0, & \text{если } f_r(X_i) \geq f_r^*. \end{cases}$$

Значение функции приобретения $I_r(X_i)$ положительное, если значение целевой функции $f(X_i)$ в данной точке лучше значения f_r^* , иначе значение $I_r(X_i)$ равно нулю. Критерий IP вычисляем как среднее арифметическое взвешенное значений $s_r^{i,j}$, полученных по результатам попарных сравнений значений функций A_r и I_r в точках тестовой выборки. Вес $\max(I_r^i, I_r^j)$ каждо-

го значения $s_r^{i,j}$ определяет степень значимости соответствия результатов сравнения по значениям функций A_r и I_r в точках i, j . Например, если сравнивают точки i, j с нулевым приростом значения целевой функции $I_r^i = 0$, $I_r^j = 0$, то результат их сравнения по функции A_r не учитывается. Чем выше значения функции приобретения I_r в точках i, j , тем важнее при использовании функции A_r сохранить результат сравнения точек по значениям функции I_r . Критерий IP оценивает степень соответствия рассматриваемой функции приобретения A_r и «истинной» функции приобретения I_r .

Примеры графиков различных функций приобретения для $|X| = 1$ на итерации $r = 6$ приведены на рис. 1. Красной линией на графике 1а обозначено наилучшее известное значение f_r^* . Красными точками на графиках 1б-1д обозначены перспективные точки X_{r+1} , в которых достигается максимум соответствующей функции приобретения.

Вычислительный эксперимент. Выполнен сравнительный анализ эффективности байесовской оптимизации с предлагаемым методом адаптации функции приобретения на основе критерия IP. Для проведения вычислительного эксперимента используем набор ВВОВ [5], состоящий из 24 тестовых оптимизационных задач, каждую из которых решаем с числом варьируемых параметров $|X| = 2, 4, 8$ и максимальным числом итераций $N = 30, 60, 120$ соответственно. Каждую из задач решаем 10 раз со случайным начальным приближением. Для решения задач используем открытую реализацию алгоритма байесовской оптимизации в библиотеке *ByesOpt* [6] с функциями приобретения LCB, PI и EI. На базе библиотеки *ByesOpt* также реализована адаптивная функция приобретения *Adaptive*, с которой на каждой итерации алгоритма на основе критерия IP выбирается наилучшая из функций приобретения LCB, PI и EI. По результатам решения задачи вычисляем нормированное наилучшее значение целевой функции $\bar{f}^* \in [0, 1]$. Значение $\bar{f}^* = 0$ соответствует известному минимуму тестовой функции, а значение $\bar{f}^* = 1$ — максимальному наилучшему значению целевой функции, найденному по результатам всех запусков алгоритма на этой задаче.

Результаты вычислительных экспериментов представлены на рис. 2, где по оси абсцисс отложены нормированные наилучшие найденные значения $\bar{f}^*$, а по оси ординат — общее число запусков алгоритма оптимизации, для которых найденное значение целевой функции лучше данного. На рисунке видно, что эффективность алгоритма оптимизации хуже при использовании функции приобретения LCB. Вероятно, это связано с умолчательным значением параметра $p = 2.576$ в библиотеке *BayesOpt*, которое излишне глобализует поиск при относительно малом вычислительном бюджете. В среднем наибольшая эффективность алгоритма достигается с функцией приобретения *Adaptive*, которая реализует предложенный метод адаптации функции приобретения. Усредненные по результатам всех запусков значения $\bar{f}^*$ приведены в таблице. При использовании функции приобретения *Adaptive* наблюдаем прирост эффективности более 11% по лучшему найденному значению $\bar{f}^*$.

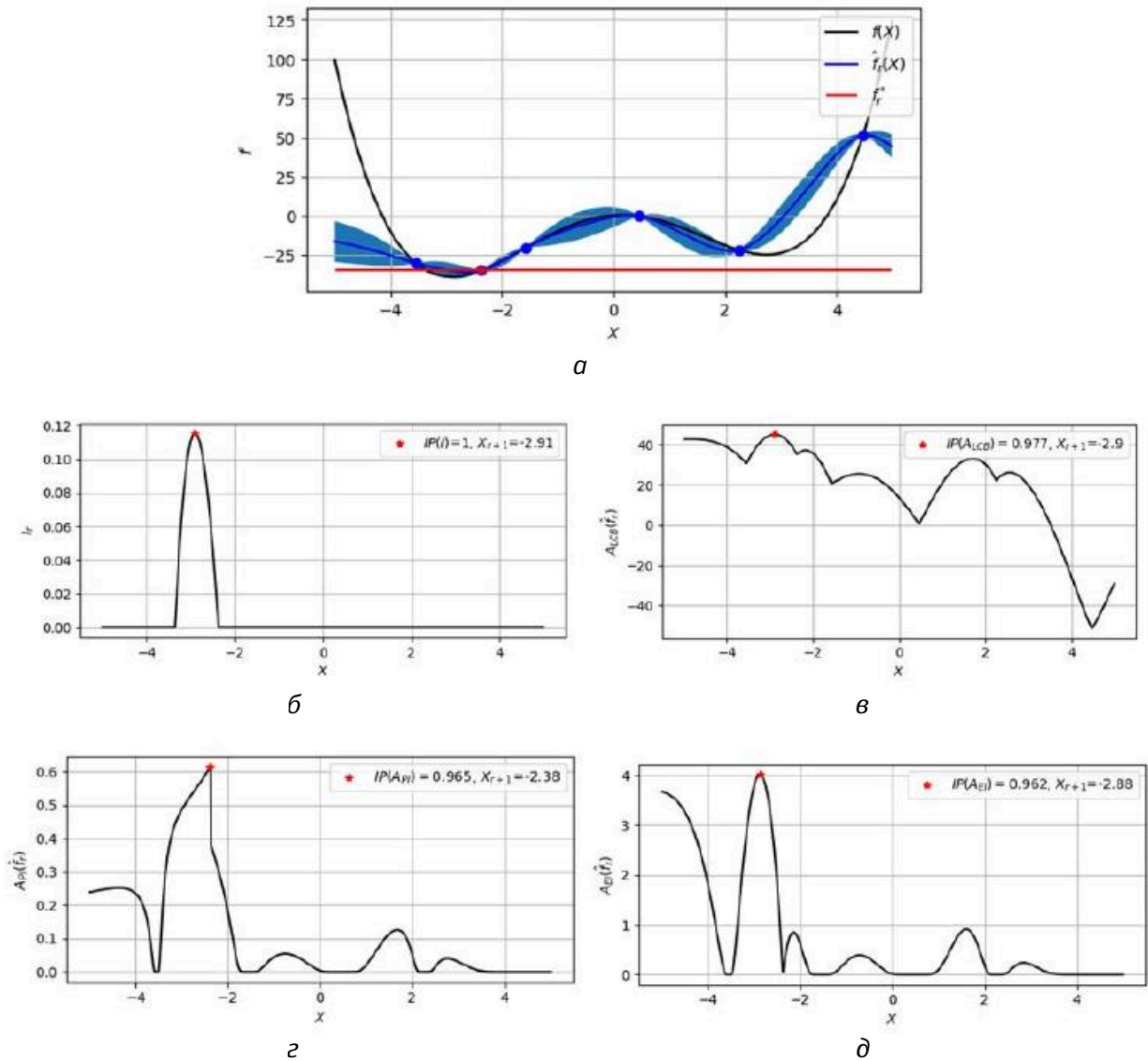


Рис. 1. Суррогатная модель $\hat{f}_r(X)$ целевой функции $f(X)$ на итерации $r=6$, $|X| = 1$ (а); графики функций приобретения IP (б), LCB (в), PI (г) и EI (д)

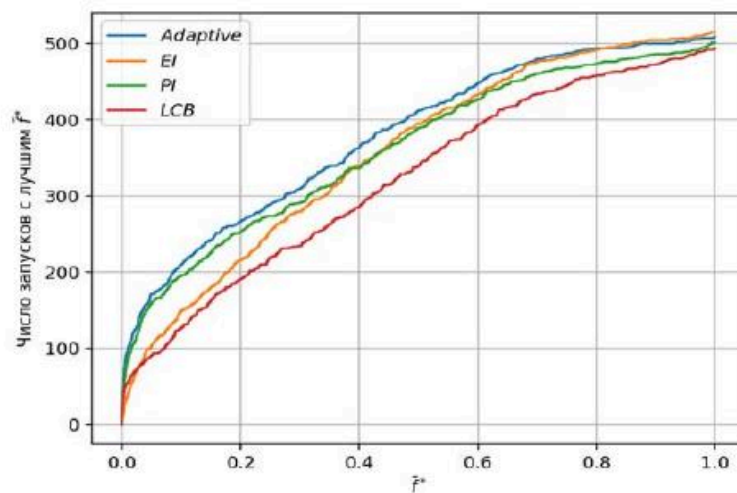


Рис. 2. Профили производительности алгоритма оптимизации при использовании различных функций приобретения

Усредненные значения $\bar{f}^*$ при использовании различных функций приобретения

Функция приобретения	Усредненное значение $\bar{f}^*$
LCB	0,381
PI	0,305
EI	0,315
Adaptive	0,271

Следует заметить, что в условиях решения прикладных задач оптимизации функция $I_r(X)$ неизвестна, тестовая выборка отсутствует ввиду высокой вычислительной сложности функции $f(X)$ и значение критерия IP не может быть вычислено напрямую. В процессе решения прикладных задач оптимизации значение критерия IP может быть вычислено приближенно с использованием методов машинного обучения. Эффективность такого подхода будет исследована в последующих работах.

Заключение. При выборе функции приобретения в байесовской оптимизации необходимо принимать во внимание особенности целевой функции, суррогатной модели и вычислительный бюджет задачи. Актуальна проблема адаптивного выбора типа функции приобретения и значений ее параметров в процессе решения прикладных задач оптимизации. Одним из перспективных направлений является разработка методики оценки эффективности функции приобретения в процессе решения прикладных задач. В работе предложен критерий *соответствия приобретения* (IP), который использован для выбора наилучшей функции приобретения в процессе решения тестовых оптимизационных задач. Результаты вычислительных экспериментов указывают на прирост эффективности алгоритма оптимизации при использовании предложенного подхода.

Список источников

- [1] Frazier P.I., Wang J. Bayesian optimization for materials design. Information science for materials discovery and design, 2016, pp. 45–75. URL: <https://arxiv.org/pdf/1506.01349> (дата обращения 01.06.2024).
- [2] Gan W., Ji Z. and Liang Y. Acquisition functions in Bayesian optimization. 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), IEEE, 2021, pp. 129–135. <https://doi.org/10.1109/ICBASE53849.2021.00032>
- [3] Wang H., van Stein B., Emmerich M., Back T. A new acquisition function for Bayesian optimization based on the moment-generating function. IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017, pp. 507–512. <https://doi.org/10.1109/SMC.2017.8122656>
- [4] Kanazawa T. One-Parameter Family of New Acquisition Functions for Efficient Global Optimization. International Joint Conference on Neural Networks (IJCNN), IEEE, 2022, pp. 1–8. <https://doi.org/10.1109/IJCNN55064.2022.9892219>
- [5] Varelas K., El Hara O.A., Brockhoff D. et al. Benchmarking large-scale continuous optimizers: The bbob-largescale testbed, a COCO software guide and beyond. Appl. Soft Comput., 2020, vol. 97, art. no. 106737. <https://doi.org/10.1016/j.asoc.2020.106737>
- [6] Nogueira F. Bayesian Optimization: Open Source Constrained Global Optimization Tool for Python. URL: <https://github.com/bayesian-optimization/BayesianOptimization> (accessed December 6, 2024).

Adaptation of acquisition function parameters during Bayesian optimization of design solutions

Agasiev Taleh

agtaleh@mail.ru

BMSTU, Moscow, Russia

Bayesian optimization algorithms are widely used for solving applied parametric design optimization problems. These algorithms employ surrogate models to reduce the number of evaluations of the computationally expensive objective function by searching for promising candidate vectors of its arguments. At each iteration of Bayesian optimization, promising candidate vectors are identified by globally optimizing an acquisition function, whose values are evaluated based on the surrogate model's predictions. The choice of the acquisition function type and parameter values can significantly impact the efficiency of solving optimization problems. A critical challenge lies in automating the selection and tuning of the acquisition function parameters while solving real-world optimization problems. The article proposes a new criterion for matching the acquisition function to the landscape of the objective function. The results of computational experiments are provided, demonstrating the increased efficiency of Bayesian optimization algorithm when using the proposed criterion.

Keywords: design optimization, Bayesian optimization, surrogate model, Gaussian process, acquisition function