

МЕТОД ПОИСКА ОПТИМАЛЬНЫХ ЗНАЧЕНИЙ ПАРАМЕТРОВ ГЕНЕТИЧЕСКОГО АЛГОРИТМА

Бурдин Александр Михайлович

начальник лаборатории АО «НПЦАП», РФ, г. Москва

THE GENETICS' ALGORITHM OPTIMAL PARAMETERS SEARCH METHOD

Alexandr Burdin

Head of laboratory, NPCAP, Moscow

Аннотация. В докладе доказываем необходимость поиска оптимальных значений параметров адаптированного генетического алгоритма оптимизации модульных испытательных систем. За основу разработанного метода легли принципы машинного обучения. Результаты работы позволили увеличить точность работы генетического алгоритма при оптимизации модульных испытательных систем.

Abstract. In the report proved the necessary of searching optimal values of adopted genetics' algorithm parameters. The method based on machine learning principles. The results of the work made it possible to increase the accuracy of the genetic algorithm when optimizing modular test systems.

Ключевые слова: оптимизация; генетический алгоритм; машинное обучение.

Keywords: optimization; genetic algorithm; machine learning.

Современные тенденции развития испытательного оборудования для цифровой аппаратуры показывают перспективность использования модульного принципа при разработке. Этот подход позволяет создать систему, соответствующую таким критериям измерений, которые труднее реализовывать с традиционным набором аппаратуры. Постоянный реинжиниринг испытываемых приборов заставляет разработчиков аппаратуры не отставать в совершенствовании своих изделий.

Модули испытательной системы необходимо подбирать оптимально, при этом не снижая функциональных возможностей испытательного комплекса. В [1] предложена методика оптимизации модульных блочных систем на основе адаптированного, проблемно-ориентированного генетического алгоритма. В данной методике при реализации генетического алгоритма в качестве значений основных параметров таких, как размер популяции, количество особей в поколении, вероятность мутации и кроссовера, брались не оптимальные значения. Такой подход может вести к неточности работы алгоритма, и учитывая его природу (за счёт количества поколений) к неоправданно долгой работе.

Появилась потребность в методике нахождения наилучших параметров генетического алгоритма в контексте задачи оптимизации модульных испытательных систем.

В качестве меры точности работы генетического алгоритма берётся разность значения целевой функции, полученной методом полного перебора (как эталонный и дающий наиболее точный результат) и генетическим алгоритмом.

Были проведены шесть экспериментов на аналогичных друг другу задачах, отличавшихся разной размерностью пространства поиска (количеством возможных комбинаций). В качестве параметров генетического алгоритма выступали одни и те же числа таблица 1.

Таблица 1

Параметры генетического алгоритма

Параметр	Значение
Размер популяции	50
Количество поколений	50
Вероятность кроссовера	0,80
Вероятность мутации	0,05

На рисунке 1 представлен график зависимости разности значений целевой функции, полученных методом полного перебора и значением целевой функции найденным генетическим алгоритмом от размерности пространства поиска. Из графика видно, что разность растёт с увеличением размерности пространства поиска, что доказывает необходимость подстраивать параметры генетического алгоритма в зависимости от размерности пространства поиска.

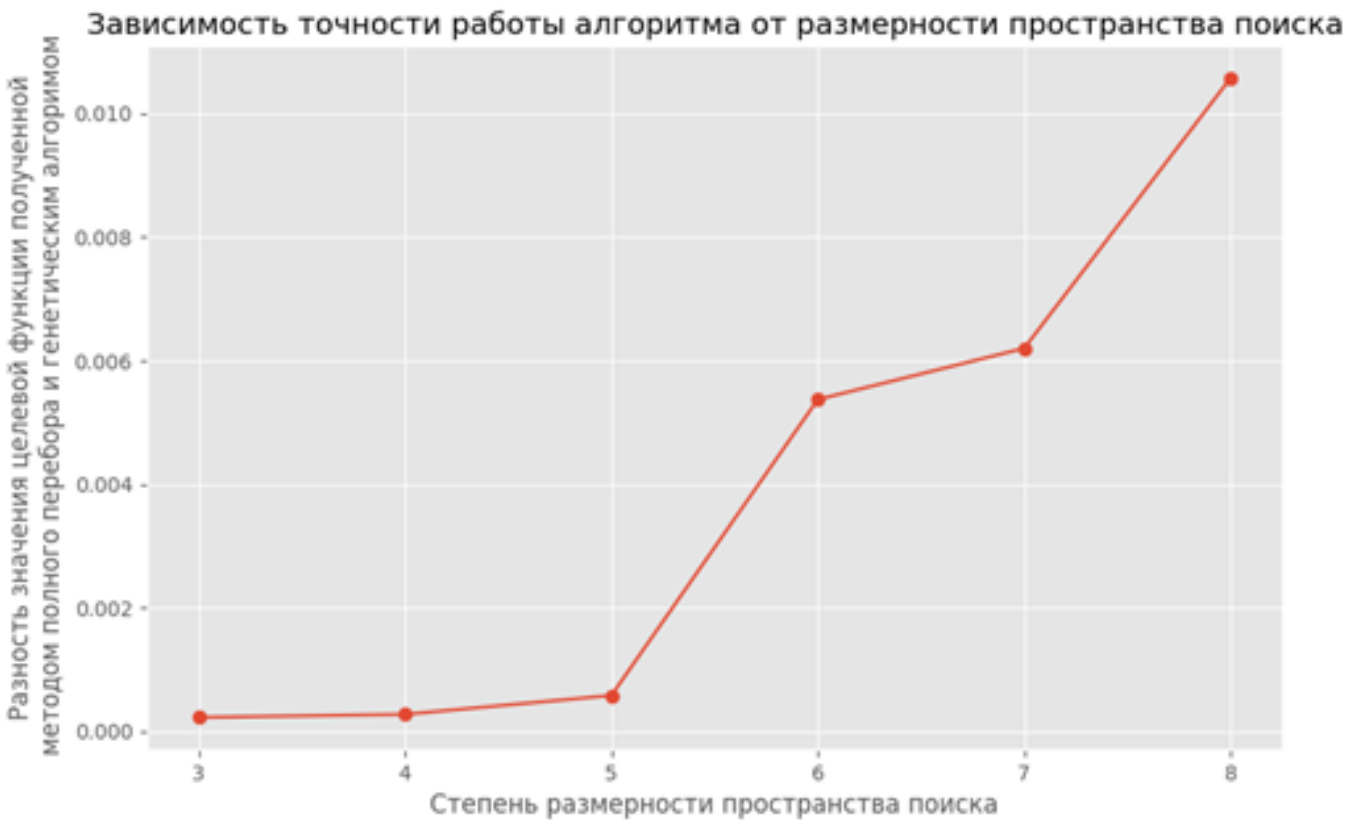


Рисунок 1. График зависимости точности работы генетического алгоритма от размерности пространства поиска

Для определения оптимальных параметров генетического алгоритма необходимо построить математическую модель ошибки, исследовать её и найти точку оптимума, где ошибка алгоритма будет минимальна.

Построить математическую модель можно при помощи классического метода полного факторного эксперимента (ПФЭ). ПФЭ даёт линейную зависимость точности алгоритма от изучаемых параметров, что может дать не корректную картину (если существует не линейная зависимость точности от параметров).

Наиболее полную модель ошибки генетического алгоритма можно построить при помощи методов машинного обучения, где модель получая на вход некоторый объём данных «обучается» (подстраивает коэффициенты). Точность полученной модели возможно оценить, используя тестовую выборку.

Данные для построения модели взяты из серии экспериментов, где входные параметры такие, как степень размерности пространства поиска, размер популяции, количество поколений в популяции, вероятность кроссовера и шанс мутации выбирались случайным образом из диапазона, представленного в таблице 2.

Таблица 2.

Значения параметров генетического алгоритма

№	Фактор	Минимальное значение	Максимальное значение
F ₁	Степень размерности пространства поиска	3	8
F ₂	Размер популяции, особей	10	200
F ₃	Количество поколений, шт	10	200
F ₄	Вероятность кроссовера, %	5	100
F ₅	Вероятность мутации, %	5	100

Количество необходимых экспериментов, получено опытным путём согласно выводам, изложенным в [2], то есть количество экспериментов, при котором точность модели перестала расти (рисунок 2).

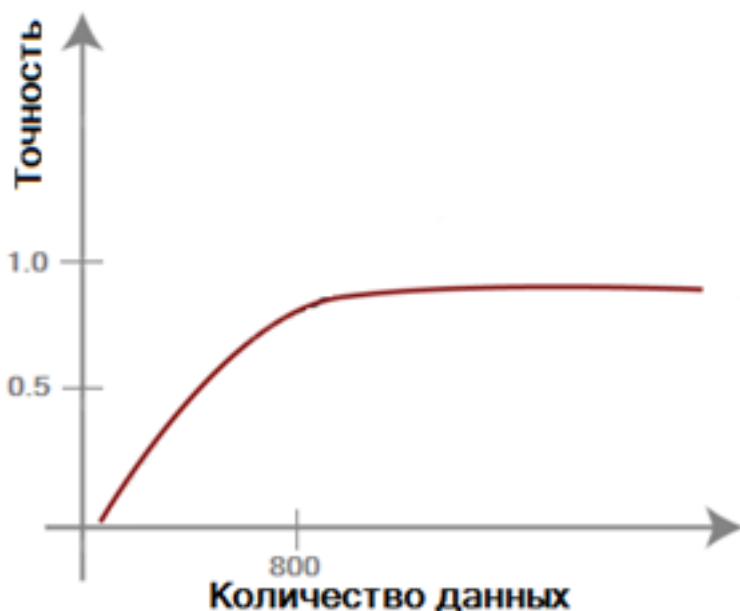


Рисунок 2. Зависимость точности модели ошибки генетического алгоритма от количества данных

Из рисунка 2, видно, что точность модели ошибки перестала расти примерно при 800 опытах.

Данные случайным образом разделены на обучающую и тестовую выборки в соотношении 0.8 и 0.2 соответственно. Построены регрессионные модели с разными степенями полиномизации и оценены их точности по коэффициенту R^2 (таблица 3).

Таблица 3

Точность регрессионной модели от степени полиномизации

Степень полиномизации	Коэффициент R^2
1	0.08
2	0.53
3	0.57
4	0.92
5	0.92
6	0.90

Наиболее полно описывает точность адаптированного генетического алгоритма регрессионные модели со степенями полиномизации больше четырёх.

Далее оценены значимости параметров методом Backward Elimination со стандартным значением $\alpha = 0.05$, после которого наименее статистически значимым оказался параметр вероятность кроссовера (рисунок 3).

	P> t
F1	0.000
F2	0.000
F3	0.015
F4	0.001
F5	0.226



	P> t
F1	0.000
F2	0.000
F3	0.007
F4	0.001

Рисунок 3. Результат применения метода *Backward Elimination*

Таким образом наиболее значимыми параметрами модели точности работы генетического алгоритма являются: размерность пространства поиска, размер популяции, количество поколений в популяции и шанс мутации.

В результате описанных выше действий получена математическая модель ошибки адаптированного генетического алгоритма для оптимизации модульных испытательных систем:

$$\begin{aligned}
 E = & 1.48 - 1.07F_1 + 6.38e^{-4}F_2 - 1.64e^{-4}F_3 - 1.28e^{-2}F_4 + 2.79e^{-1}F_1^2 - \\
 & 2.09e^{-4}F_1F_2 + 5.97e^{-5}F_1F_3 - 7.77e^{-3}F_1F_4 + 3.97e^{-5}F_2F_4 + 1.65e^{-4}F_3F_4 + \\
 & 5.95e^{-2}F_4^2 - 3.13e^{-2}F_1^3 - 3.59e^{-6}F_1^2F_2 - 6.23e^{-6}F_1^2F_3 - 3.59e^{-6}F_1^2F_2 - \\
 & 6.23e^{-6}F_1^2F_3 + 2.02e^{-3}F_1^2F_4 + 3.37e^{-5}F_1F_2F_4 - 5.61e^{-5}F_1F_3F_4 - 5.59e^{-3}F_1F_4^2 - \\
 & 1.16e^{-4}F_2F_4^2 - 1.66e^{-4}F_3 - 3.34e^{-2}F_4^3 + 1.28e^{-3}F_1^4 + 1.75e^{-6}F_1^3F_2 - \\
 & 1.35e^{-4}F_1^3F_4 - 3.10e^{-6}F_1^2F_2F_4 + 2.53e^{-6}F_1^2F_3F_4 + 5.60e^{-4}F_1^2F_4^2 - \\
 & 2.02e^{-6}F_1F_2F_4^2 + 2.10e^{-5}F_1F_3F_4^2 - 2.54e^{-3}F_1F_4^3 + 9.75e^{-5}F_2F_4^3 + \\
 & 6.05e^{-5}F_3F_4^3 + 1.55e^{-2}F_4^4
 \end{aligned}$$

$$F_1 > 0, F_2 > 0, F_3 > 0, F_4 \in [0; 100],$$

где

E - ошибка алгоритма;

F_1 - степень размерности пространства поиска;

F_2 - размер популяции;

F_3 - количество поколений;

F_4 - вероятность кроссовера.

Критическая значимость коэффициентов: $t_{кр.} \cdot S_{коэф.} \approx 1,48e-06$.

Необходимо найти такие значения размера популяции F_2 , количества поколений F_3 ,

вероятности кроссовера F_4 для каждой размерности пространства поиска F_1 , при которых ошибка E будет стремиться к нулю.

Усечённым методом Ньютона определены оптимальные значения параметров генетического алгоритма (таблица 4).

Таблица 4

Оптимальные значения параметров генетического алгоритма при разных размерностях пространства поиска

Степень размерности пространства поиска	Размер популяции, шт.	Количество поколений, шт.	Вероятность мутации в поколении, %	
3	40	12	75	
4	65	16	80	
5	39	63	65	
6	47	53	72	
7	51	71	83	
8	55	82	91	
9	59	89	62	

В результате проделанной работы разработана методика нахождения оптимальных параметров адаптированного генетического алгоритма для оптимизации модульных испытательных систем на основе методов машинного обучения. Данная методика позволила построить точную (относительно других методов) модель ошибки генетического алгоритма. Благодаря предложенной методике повысилось качество работы адаптированного генетического алгоритма для оптимизации модульных испытательных систем, что дало возможность разрабатывать более оптимальные испытательные системы.

Список литературы:

1. Бурдин А.М. Васин В.Н. Автоматизированная система многокритериальной оптимизации блочного-модульного испытательного комплекса устройств обмена. Труды ФГУП «НПЦАП». Системы и приборы управления, №2, 2016, С. 45-54.

2. Andrew Ng. Machine Learning Yearning-Draft. deeplearning.ai. 2018. P. 105.