

Адаптивная настройка параметров эволюционного алгоритма с динамическим разбиением диапазона с помощью обучения с подкреплением

Рост А. Ю., магистр кафедры компьютерных технологий Университета
ИТМО, arkrost@gmail.com

Петрова И. А., аспирант кафедры компьютерных технологий Университета
ИТМО, irenepetrova@yandex.com

Буздalова А. С., аспирант кафедры компьютерных технологий
Университета ИТМО, abuzdalova@gmail.com

Аннотация

Недавно был предложен эффективный метод адаптивной настройки параметров эволюционного алгоритма, основанный на обучении с подкреплением. В данном методе диапазоны допустимых значений параметров разбиваются на несколько подынтервалов до запуска алгоритма. Однако изменение разбиения во время работы способствует увеличению эффективности работы алгоритма. Алгоритм EARPC является одним из наиболее эффективных методов, использующих динамическое разбиение диапазона.

В данной работе предлагаются два метода адаптивной настройки параметров эволюционного алгоритма. Один из них является улучшением существующего метода настройки параметров с помощью обучения с подкреплением за счет динамического разбиения диапазона допустимых значений параметров при помощи алгоритма EARPC. Второй метод основан на применении Q-обучения с адаптивным выделением множества действий агента. Было проведено сравнение предложенных подходов с существующими методами на нескольких задачах численной оптимизации с использованием различных конфигураций эволюционного алгоритма. Наилучшие результаты были получены при использовании метода, основанного на применении Q-обучения с адаптивным выделением множества действий агента.

Введение

Обозначим эффективность эволюционного алгоритма (ЭА) как число вычислений функции приспособленности (ФП), необходимое для нахождения оптимума. Эффективность ЭА зависит от значений его параметров, например, вероятности мутации и скрещивания. Значения параметров могут быть выбраны до запуска ЭА, однако оптимальное значение параметра может меняться в ходе процесса оптимизации. Поэтому необходим метод адаптивного выбора значений параметров ЭА.

В настоящей работе рассматривается случай выбора параметра из некоторого вещественного диапазона. Зачастую задачу выбора значений параметров дискретизируют, разделяя диапазон допустимых значений параметра на интервалы. Разбиение на интервалы может производиться до запуска алгоритма и не меняться в процессе его работы. Однако было показано, что изменение разбиения во время работы повышает эффективность работы алгоритма [1, 2]. В частности, значение параметра можно подобрать тем точнее, чем меньше шаг разбиения. Одним из наиболее эффективных методов с динамическим разбиением диапазона значений параметров является EARPC [1, 6].

Опишем алгоритм EARPC. В алгоритме EARPC каждый из параметров настраивается отдельно от остальных. Каждому выбранному значению v_i параметра p_i соответствует оценка его качества q_i . Для выбора значения параметра сохраненные пары (v_i, q_i) разбиваются на два кластера, например, алгоритмом k -средних. Затем интервал значений параметра разбивается на два подынтервала, минимизируя энтропию. На каждом из подынтервалов считается среднее качество Q_i . Один из подынтервалов выбирается с вероятностью, пропорциональной Q_i . В качестве значения параметра выбирается случайное значение из выбранного подынтервала.

Недавно в работе Karafotias et al. был предложен эффективный алгоритм настройки параметров с использованием обучения с подкреплением (далее К-алгоритм) [5, 6]. Однако в данном методе разбиение диапазона значений параметра не меняется в ходе процесса оптимизации. Также не производилось сравнение данного подхода с методом EARPC.

Опишем К-алгоритм. При использовании обучения с подкреплением (reinforcement learning, RL) [4, 9] для настройки параметров ЭА [7, 8] в качестве среды выступает ЭА, а действием является выбор подынтервала, из которого случайным образом выбирается значение параметра. Полученная агентом награда является некоторой функцией от значений ФП текущего и предыдущего поколений. В К-алгоритме значения всех параметров выбираются одновременно. Пусть есть k настраиваемых параметров. Диапазон каж-

дого из параметров v_i разбивается перед запуском алгоритма на m_i подынтервалов. Действием агента является одновременный выбор подынтервалов для каждого из параметров. Поэтому число возможных действий агента равно $\prod_{i=1}^k m_i$. Полученная награда вычисляется как $r = c \cdot \left(\frac{f_{t+1}}{f_t} - 1 \right)$, где f_t — лучшее значение функции приспособленности, полученное на t -ой итерации, а c — некоторая константа. Отметим, что награда всегда положительна в случае, если ЭА не ухудшает лучшее полученное решение.

Для определения состояний среды используется двоичное дерево решений [10]. На ребрах дерева находятся условия на наблюдаемые характеристики ЭА, например, прирост среднего значения ФП в поколении. Листьям дерева соответствуют состояния среды, для которых хранятся значения ожидаемой награды и кортежи опыта, полученные в этом состоянии. Таким образом, в К-алгоритме по наблюдаемым характеристикам ЭА определяется состояние среды s . Затем агент на основе сохраненных значений ожидаемой награды для s выбирает действие, применяет его к среде и сохраняет полученный кортеж опыта. Далее при помощи алгоритма Колмогорова-Смирнова состояние s разбивается на два новых. Если точка разбиения найдена, то сохраненные в состоянии s кортежи опыта разделяются по новым состояниям в соответствии с разбиением.

Предложенные методы

Предлагается два новых метода адаптивной настройки параметров с динамическим разбиением диапазона. Первый из них совмещает К-алгоритм и EARPC. Второй основан на обучении с подкреплением и разбиении диапазонов значений параметров при помощи критерия Колмогорова-Смирнова.

Метод E+K

Данный метод совмещает в себе К-алгоритм и EARPC. В К-алгоритме диапазоны значений параметров разбиваются на подынтервалы до запуска ЭА. В данном алгоритме производится динамическое разбиение диапазонов на подынтервалы. Однако при изменении разбиения меняется и множество действий агента. Поэтому необходимо изменить способ выбора параметров К-алгоритма. В предложенном алгоритме в листьях дерева решений хранятся только полученные кортежи опыта. Значения параметров выбираются при помощи алгоритма EARPC, применяющегося к хранящимся в соответствующем текущему состоянию листе кортежам опыта.

Метод KS+RL

Данный метод основан на применении критерия Колмогорова-Смирнова (KS) и обучении с подкреплением (RL). Предварительные эксперименты показали, что формирование состояний при помощи двоичного дерева поиска неэффективно. Поэтому в методе KS+RL используется среда с одним состоянием. В отличие от К-алгоритма, каждый параметр настраивается отдельно, поэтому для каждого настраиваемого параметра v_i используется свой агент Q-обучения $agent_i$. Изначально у $agent_i$ есть только одно действие — выбор значения параметра из интервала $[min_i, max_i]$, где min_i и max_i это минимальное и максимальное значения параметра v_i . После того как все агенты выбрали значение параметра, выбранные значения $v = (v_1 \dots v_n)$ передаются ЭА. Затем вычисляется награда r и сохраняется пара — значения параметров v и полученная за них награда r . Процесс повторяется до тех пор, пока не накапливается достаточное число сохраненных пар для того, чтобы разбить диапазон допустимых значений параметра при помощи критерия Колмогорова-Смирнова. Затем диапазон разбивается на две части, аналогично тому, как производилось разбиение состояний в К-алгоритме. Каждый из полученных подынтервалов снова разбивается на две части. Таким образом, число полученных подынтервалов равно 2^i , где i — некоторая константа. Множество действий агента обновляется в соответствии с разбиением. Агент выбирает действие до тех пор пока значения ожидаемой награды не становятся почти одинаковыми для хотя бы одной пары возможных действий агента. В этом случае агент не может выбрать какое из действий наиболее эффективно, поэтому происходит переразбиение диапазона допустимых значений.

Описание экспериментов

Целью ЭА являлось нахождение вектора, на котором достигается минимум заданной функции с точностью ϵ . В качестве функций использовались функции сферы, Розенброка, Леви и Растригина. В качестве ЭА использовалась $(\mu + \lambda)$ эволюционная стратегия. В качестве особи ЭА рассматривался набор x_1, \dots, x_n , состоящий из n вещественных чисел, $x_i \in [min_i, max_i]$. В качестве эволюционных операторов использовался только оператор мутации, трансформирующий каждое число x_i в $x_i + \sigma dx_i$, где $dx_i \sim \mathcal{N}(0, 1)$, а σ — настраиваемый параметр. Если $x_i + \sigma dx_i > max_i$, то x_i трансформировался в max_i , а если $x_i + \sigma dx_i < min_i$, то в min_i . Ожидаемым поведением методов адаптивного выбора параметров ЭА является уменьшение σ при приближении к оптимуму. Значения параметра σ лежат в интервале $[0, k]$, где k — некоторая константа. При увеличении параметра k увеличивается

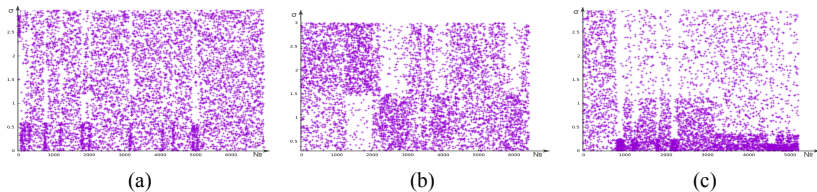


Рис. 1: Выбранные значения σ в К-алгоритме (a) и методах E+K (b) и KS+RL (c)

диапазон допустимых значений параметра σ , что усложняет задачу поиска его оптимального значения.

Результаты

В Таблице 1 представлено среднее число поколений, необходимое для достижения оптимума. Первые три колонки содержат значения параметров k , μ and λ . Следующие 20 колонок содержат результаты оптимизации четырех функций с различным ландшафтом — сферы, Растригина, Леви и Розенброка. Для каждой функции представлены результаты предложенного метода KS+RL (K+R), алгоритма Q-обучения (Q), К-алгоритма (K), алгоритма EARPC (E) и предложенного метода E+K (E+K). Последняя строка таблицы содержит общее число конфигураций ЭА, на которых соответствующий алгоритм превзошел другие алгоритмы. Серым цветом выделен лучший результат для каждой из конфигураций ЭА. Отметим, что мы не сравниваем различные конфигурации ЭА, а число вычислений ФП в поколении для каждой из конфигураций одинаково для всех алгоритмов выбора параметров. Поэтому использование числа поколений вместо числа вычислений ФП не влияет на результаты сравнения алгоритмов настройки параметров. Предложенный метод KS+RL превосходит остальные рассмотренные методы на большинстве экземпляров задач. В соответствии с множественным знаковым тестом [3], метод KS+RL различим с остальными методами с уровнем статистической значимости $\alpha = 0.05$.

На рис. 1 показаны значения настраиваемого параметра σ , выбираемые при помощи двух предложенных алгоритмов и К-алгоритма в процессе оптимизации функции Растригина. По горизонтали отложен номер итерации ЭА, а по вертикали — выбранное значение σ . Можно видеть, что предложенный алгоритм KS+RL улучшает значение настраиваемого параметра на протяжении всей работы ЭА, в отличие от других методов.

Таблица 1: Среднее число поколений, необходимое для достижения оптимума

k	μ	λ	Функция сферы					Функция Растригина					Функция Левы					Функция Розенброка				
			A	Q	K	E	E+K	A	Q	K	E	E+K	A	Q	K	E	E+K	A	Q	K	E	E+K
1	1	1	2434	8769	8048	5258	4830	3631	9653	8385	7794	8689	3496	7200	7265	7986	14092	5124	15058	13418	9003	12905
1	1	3	2207	4221	2683	3942	3070	1776	2226	2069	3148	3610	1980	3305	2903	2789	3688	2301	3914	4167	3553	2701
1	1	7	878	1085	2620	1653	2247	1226	1605	1757	1422	1422	1321	1584	1600	1923	3820	1411	1791	2330	2296	1941
1	5	1	1450	1664	2076	4472	3893	1666	1706	2281	4341	4530	1778	1898	1865	2372	2246	1859	2311	2809	5730	4453
1	5	3	569	706	824	1589	845	918	894	942	1958	2197	855	862	794	1632	2171	1053	869	748	1393	2749
1	5	7	368	390	473	642	568	502	570	675	1679	1329	356	606	444	1426	1236	401	533	665	1130	1258
1	10	1	703	959	747	1438	728	1008	1103	1105	2681	2926	884	840	804	2353	1451	1617	1497	1593	2213	3085
1	10	3	331	378	358	534	400	622	604	665	1460	1485	533	502	624	1491	1134	683	488	616	1161	1225
1	10	7	186	195	167	142	422	358	489	473	1103	1858	308	331	294	510	766	483	290	235	391	429
2	1	1	4342	25192	28523	31738	16182	4744	23663	21043	27865	19142	4947	13420	14789	25721	30653	5165	23371	27239	20005	20819
2	1	3	2333	7681	6478	5152	4233	1839	6405	6825	9748	8997	12020	6884	6601	7477	4612	3461	7295	7169	7166	13342
2	1	7	1464	3360	3739	1688	2956	1160	2388	2183	3806	4085	1205	3521	2370	2533	2967	1753	3488	2968	5874	7870
2	5	1	1891	3814	3468	4908	5173	2467	3944	4029	5961	5750	1935	3517	3369	7222	5365	1990	8076	5810	11153	11856
2	5	3	974	994	1163	1631	946	1128	1614	1780	2293	1720	1085	1231	1326	3039	2943	1396	2116	2705	3775	4542
2	5	7	616	716	756	1511	626	967	1012	1461	933	1180	770	792	1089	1153	2180	934	1001	1247	1775	2584
2	10	1	1164	2397	1833	778	876	1722	2108	2255	2411	2807	1803	1884	2197	5139	3072	2022	3415	2787	8768	4603
2	10	3	459	445	465	719	788	988	1420	1116	1486	1234	887	988	1006	1045	2381	1181	1037	1106	3719	1648
2	10	7	188	320	252	380	413	615	856	712	627	922	618	731	564	806	1064	707	811	666	1740	2558
3	1	1	8055	36698	29112	54868	30710	5159	29082	23305	24249	27327	5216	21470	21953	28900	35119	6535	28702	36597	32533	43832
3	1	3	2826	7845	9115	12446	11985	2821	7977	10726	6541	16040	2900	9029	7071	6966	10883	3391	11427	9873	13061	10068
3	1	7	1427	3907	4813	10118	9207	1517	4680	4266	4144	5408	1700	4139	4019	3375	2062	2573	5820	6462	9797	9775
3	5	1	2447	8328	5886	3348	12313	2704	6208	5419	5953	7665	3105	6966	6742	10480	9059	3148	12438	6791	8952	11581
3	5	3	1445	2790	2222	1379	1848	1544	2514	2096	2823	2304	1808	2467	2664	4593	3714	1721	2289	3673	4434	6799
3	5	7	896	919	777	1424	1105	902	1120	1629	1033	1055	1017	1929	1788	594	1439	1231	1626	1255	3951	4914
3	10	1	1996	2398	2531	3173	3745	2048	3747	3258	5095	3873	2071	2901	2943	5210	4814	2352	4473	6531	8320	10539
3	10	3	1053	1074	1206	1114	171	1296	1740	1523	1013	1028	1330	1462	1832	3140	2389	1677	2000	1650	2611	2479
3	10	7	409	391	427	410	537	955	995	1203	779	839	667	1117	799	521	845	1101	1137	1098	1742	1907
Итого			20	2	2	2	1	22	2	0	3	0	19	1	4	2	0	19	3	5	0	0

Заключение

В работе предложены два метода адаптивной настройки параметров эволюционных алгоритмов с помощью обучения с подкреплением. В данных алгоритмах множество действий агента формировалось в процессе оптимизации, за счет разбиения диапазона допустимых значений параметра в ходе работы алгоритма. Один из предложенных алгоритмов был основан на двух существующих подходах к настройке параметров эволюционного алгоритма: алгоритме *EARPC* и методе, предложенном *Karafotias et al.* Второй алгоритм основан на том, что диапазон допустимых значений настраиваемого параметра разбивается при помощи критерия Колмогорова-Смирнова.

Предложенные методы адаптивной настройки параметров эволюционных алгоритмов были протестированы на четырех модельных задачах с использованием 27 конфигураций эволюционного алгоритма. Было проведено сравнение предложенных методов с существующими алгоритмами. Наилучшие результаты были получены при использовании второго предложенного алгоритма. Было продемонстрировано, что данный метод улучшает значения параметров на протяжении всего процесса оптимизации в отличие от остальных рассмотренных методов.

Литература

- [1] A. Aleti and I. Moser. Entropy-based adaptive range parameter control for evolutionary algorithms. In *Proceedings of Genetic and Evolutionary Computation Conference*, pages 1501–1508, 2013.
- [2] A. Aleti, I. Moser, and S. Mostaghim. Adaptive range parameter control. In *Proceedings of Congress on Evolutionary Computation*, pages 1–8, 2012.
- [3] J. Derrac, S. Garcia, D. Molina, and F. Herrera. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, 1(1):3–18, 2011.
- [4] A. Gosavi. Reinforcement learning: A tutorial. survey and recent advances. *INFORMS Journal on Computing*, 21(2):178–192, 2009.
- [5] G. Karafotias, Á. E. Eiben, and M. Hoogendoorn. Generic parameter control with reinforcement learning. In *Proceedings of Genetic and Evolutionary Computation Conference*, pages 1319–1326, 2014.
- [6] G. Karafotias, M. Hoogendoorn, and A. Eiben. Parameter control in evolutionary algorithms: Trends and challenges. *Evolutionary Computation, IEEE Transactions on*, PP(99):1–1, 2014.
- [7] J. E. Pettinger and R. M. Everson. Controlling Genetic Algorithms with Reinforcement Learning. In *Proceedings of Genetic and Evolutionary Computation Conference*, page 692, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.
- [8] Y. Sakurai, K. Takada, T. Kawabe, and S. Tsuruta. A method to control parameters of evolutionary algorithms by using reinforcement learning. In *Proceedings of 2010 Sixth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)*, pages 74–79, 2010.
- [9] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, USA, 1998.
- [10] W. T. B. Uther and M. M. Veloso. Tree based discretization for continuous state space reinforcement learning. In *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence, AAAI '98/IAAI '98*, pages 769–774, Menlo Park, CA, USA, 1998.