

УДК: 004.021

Неоднородные клеточные генетические алгоритмы

Н. М. Ершов

Факультет вычислительной математики и кибернетики,
Московский государственный университет им. М. В. Ломоносова,
Россия, 119991, г. Москва, ГСП-1, Ленинские горы, 1-52

E-mail: ershovnm@gmail.com

Получено 17 сентября 2014 г.

В работе вводится в рассмотрение понятие неоднородного клеточного генетического алгоритма, в котором ряд параметров, влияющих на работу генетических операторов, оказывается зависимым от местоположения клеток заданного клеточного пространства. Приводятся результаты численного сравнения неоднородных клеточных генетических алгоритмов со стандартными вариантами генетических алгоритмов, показывающие преимущества предложенного подхода при минимизации мультимодальных функций с большим числом локальных экстремумов. Рассматривается крупноблочная параллельная реализация неоднородных клеточных алгоритмов с использованием технологии MPI.

Ключевые слова: эволюционные алгоритмы, генетические алгоритмы, параллельные вычисления

Non-uniform cellular genetic algorithms

N. M. Ershov

*Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University,
1-2 Leninskiye Gory, GSP-1, Moscow, 119991, Russia*

In this paper, we introduce the concept of non-uniform cellular genetic algorithm, in which a number of parameters that affect the operation of genetic operators is dependent on the location of the cells of a given cellular space. The results of numerical comparison of non-uniform cellular genetic algorithms with the standard genetic algorithms, showing the advantages of the proposed approach while minimizing multimodal functions with a large number of local extrema, are presented. The coarse-grained parallel implementation of the non-uniform algorithms using the technology of MPI is considered.

Keywords: evolutionary algorithms, genetic algorithms, parallel computing

Работа выполнена при финансовой поддержке РФФИ (грант №14-07-00628 А).

Citation: *Computer Research and Modeling*, 2015, vol. 7, no. 3, pp. 775–780 (Russian).

1. Введение

Клеточные генетические алгоритмы [Alba and Dorronsoro, 2008] обладают рядом преимуществ по сравнению с обычными генетическими алгоритмами. Во-первых, за счет локальности взаимодействия между особями популяции удается более долгое время поддерживать разнообразие в популяции, что потенциально ведет к получению более качественного решения. Во-вторых, благодаря регулярности расположения особей в клеточном пространстве и отсутствию глобальных операций клеточные генетические алгоритмы хорошо и масштабируемо распараллеливаются [Whitley, 1994]. Однако, как и в обычных генетических алгоритмах, в клеточном варианте остается актуальной проблема попадания алгоритма в локальные экстремумы. В настоящей работе предлагается подход к решению этой проблемы, основанный на введении зависимости работы операторов генетического алгоритма (прежде всего мутации) от положения особи в клеточном пространстве.

2. Оптимизационная задача

Принципы работы неоднородных клеточных генетических алгоритмов рассмотрим на примере решения задачи многомерной *непрерывной* оптимизации функции $F(x)$ в области $x_k \in [-100, 100]$, $k \in \{1 \dots n\}$. Рассматривались две функции — бимодальная и мультимодальная. Первая функция

$$F(\theta, x, y) = G(1, x + 5, y + 5) + 2 G(\theta, x - 5, y - 5),$$

является двумерной и представляет собой сумму (с отрицательным знаком) двух гауссианов:

$$G(\theta, x, y) = 1 - e^{-\frac{x^2 + y^2}{2\theta}},$$

первый из которых имеет фиксированную ширину, а ширина второго определяется параметром θ . Функция имеет два минимума. Минимум в точке $(-5, -5)$ является локальным и имеет фиксированную ширину. Минимум в точке $(5, 5)$ является глобальным, значение функции в нем равно примерно 1. Параметр θ в численных экспериментах был меньше 1, поэтому область «притяжения» глобального минимума является более узкой по сравнению с локальным минимумом.

Вторая функция, с которой проводилось исследование, — это функция Растригина [Ke Tang et al., 2010]:

$$R(x) = \frac{1}{100} \sum_{k=1}^n \left[(x_k - 20)^2 + A \left(\frac{\cos(\pi x_k)}{5} \right) \right].$$

Данная функция является мультимодальной с единственным глобальным минимумом в точке $x_k = 20$, $k \in \{1 \dots n\}$. В двумерном случае в заданной области функция Растригина имеет 400 локальных минимумов, а 10-мерная функция — уже порядка 10^{13} минимумов. В численных экспериментах использовались функции с параметром $A = 2 \cdot 10^5$. Коэффициенты функции Растригина подобраны так, что локальные минимумы располагаются в точках вида $x_k = 10l$, $l \in \mathbb{Z}$, а значения функции в этих точках является суммой n квадратов целых чисел.

3. Однородные генетические алгоритмы

Сравнение эффективности неоднородного клеточного генетического алгоритма (NCGA) производилось с генетическим алгоритмом (GA) и клеточным генетическим алгоритмом (CGA). Во всех рассматриваемых вариациях алгоритмов использовались одинаковые генетиче-

ские операторы (отбор, скрещивание, мутация и миграция). Для отбора применялась *турнирная* схема, в которой пара выбранных особей решает вопрос о выживании одной из них. Особь с лучшим значением целевой функции выживает с вероятностью p_{win} . Победитель замещает своей копией место побежденного. Для скрещивания использовалась равномерная схема, когда с заданной вероятностью p_{swap} меняются местами два соответствующих гена двух заданных особей. В силу специфики решаемой задачи (непрерывная оптимизация) оператор мутации выполняет случайное изменение каждого гена заданной особи с достаточно большой вероятностью p_{mut} :

$$x_k \leftarrow x_k + \rho d_{\text{mut}},$$

где ρ — случайное действительное число из диапазона $[-1, 1]$, d_{mut} — величина мутации. Чтобы обеспечить разнообразие популяции на начальном этапе работы алгоритма и сохранить его сходимость, применялась технология *имитации отжига* по параметру d_{mut} . Оператор миграции использовался только в клеточных вариантах генетического алгоритма. Этим оператором две выбранные особи меняются местами.

Для однородного генетического алгоритма использовалась следующая схема работы. Выполнялось заданное число итераций. На каждой итерации выполнялись последовательно три оператора: отбор, скрещивание и мутация. После каждого оператора производилось полное перемешивание популяции. Для отбора и скрещивания брались пары соседних (по номерам) особей с номерами $2i$ и $2i + 1$, $i \in 0 \dots n/2$.

Однородный клеточный генетический алгоритм работал по следующей схеме. Все особи размещались по одной в клетках прямоугольной области. На каждой итерации алгоритма сначала выполнялись операторы отбора, скрещивания и миграции. Для этого клетки случайным образом делились на пары, так чтобы две клетки в одной паре всегда были соседними (т. е. имели бы общую сторону). Последним шагом на итерации выполнялась мутация всех особей.

4. Неоднородные клеточные генетические алгоритмы

Как уже упоминалось выше, проблемой однородных (клеточных или нет) генетических алгоритмов является то, что в итоге они сходятся к однородной (гомогенной) популяции. В такой популяции практически все особи имеют одинаковые геномы, поэтому наиболее мощный генетический оператор скрещивания перестает работать. Если алгоритм попал в локальный экстремум, то выйти из него можно только за счет мутации. Если мутация локальна, а более хорошее решение значительно удалено от найденного, то вероятность выхода из данного локального минимума оказывается очень низкой. Таким образом, проблема заключается в потере генетического разнообразия.

Используя клеточные генетические алгоритмы, можно поддерживать разнообразие популяции сколь угодно долгое время, не теряя при этом сходимости. Суть идеи заключается в том, что некоторые параметры генетического алгоритма делаются зависимыми от положения особи в клеточном пространстве. В простейшем варианте такой модификации подвергается величина мутации d_{mut} . В одних областях мы делаем значение этого параметра высоким, в других — низким. Поэтому первые области будут *все время* генерировать достаточно случайные решения, поддерживая таким образом необходимое разнообразие всей популяции. Области второго типа будут использоваться по своему основному назначению — селекции и скрещиванию лучших решений.

Для экспериментов было выбрано три варианта зависимости параметра d_{mut} от координат i и j клеток. В первом варианте (NCGAs, рис. 1а) параметр d_{mut} зависит только от горизонтальной координаты, так что по краям поля значение этого параметра является высоким, в середине — низким. Во втором варианте мутация в центре поля является высокой, а по его краям — низкой (NCGAm, рис. 1б). Наконец, в третьем случае, области с высокой величиной мутации, располагались в четырех углах поля (NCGAc, рис. 1в).

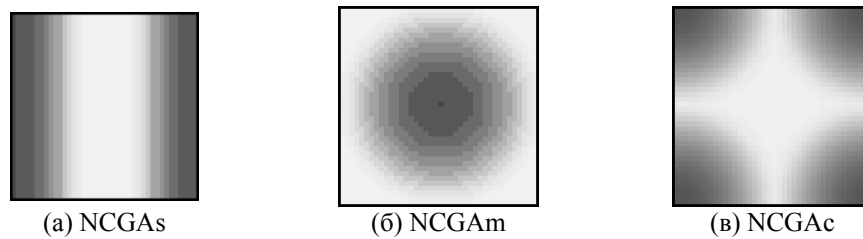


Рис. 1. Градиент величины мутации

5. Минимизация бимодальной двумерной функции

В задаче минимизации бимодальной целевой функции $F(\theta, x, y)$ измерялось частота (в процентах) обнаружения генетическим алгоритмом глобального минимума функции. Для этого выполнялось 100 запусков каждой версии генетического алгоритма. Полученные результаты показаны в таблице 1.

Таблица 1

θ	GA	CGA	NCGAs	NCGAm	NCGAc
0.001	0	0	4	4	4
0.002	0	0	5	5	5
0.005	0	0	6	8	9
0.010	0	0	10	15	14
0.020	0	0	28	24	25
0.050	3	15	61	47	54
0.100	30	76	87	72	80
0.200	74	100	100	95	97
0.500	100	100	100	100	100

Все алгоритмы в данном случае выполняли одинаковое число (500) итераций. Видно, что частота обнаружения глобального минимума неоднородными алгоритмами существенно выше по сравнению с однородными версиями. При этом заметим, что показатели неоднородного алгоритма (последние три столбца таблицы) могут быть улучшены за счет увеличения числа итераций, в то время как показатели однородных алгоритмов таким образом уже не улучшаются.

6. Минимизация мультимодальной двумерной функции

Минимизация двумерной функции Растригина $R(x, y)$ выполнялась с помощью однородных генетических алгоритмов (GA и CGA) и неоднородного алгоритма (NCGAs). Вычислялась частота обнаружения алгоритмом того или иного локального минимума R . Результаты численного эксперимента приведены в таблице 2. В первом столбце таблицы показаны первые пять локальных минимумов (0 — глобальный минимум), заметим, что в данной задаче нет локального минимума со значением 3. В остальных столбцах таблицы показаны частоты попадания указанных алгоритмов в заданные локальные минимумы. Для неоднородного алгоритма показано, как изменяются частоты ответов в зависимости от числа итераций. Таким образом, видно, что уже при 3000 итераций неоднородный алгоритм практически гарантированно находит глобальный экстремум (из имеющихся 400 локальных экстремумов).

7. Минимизация мультимодальной многомерной функции

В последнем эксперименте исследовалась частота попадания генетического алгоритма в локальные минимумы 10-мерной функции Растригина $R(x)$. В этом случае имеется порядка

Таблица 2

R	GA 500	CGA 500	NCGA 500	NCGA 1000	NCGA 2000	NCGA 3000
0	51	44	65	82	90	98
1	41	46	29	18	10	2
2	8	6	5	0	0	0
4	0	3	1	0	0	0
5	0	1	0	0	0	0

10^{13} локальных минимумов, из которых всего один является глобальным. Все минимумы имеют целые неотрицательные значения. Первые расчеты показали, что даже неоднородные алгоритмы не способны обнаружить глобальный минимум (а также близкие к нему локальные минимумы). Поэтому в алгоритм NCGAs была внесена еще одна неоднородность: вероятность выполнения оператора отбора для двух особей из соседних клеток была сделана зависящей от вертикальной координаты i :

$$p_{\text{sel}} = 1 - |i / w - 1|.$$

Таким образом, сверху и снизу клеточного поля отбор практически не работает, что дает образующимся за счет высокой мутации в угловых областях особям больше времени на улучшение целевой функции.

Численные результаты показаны в таблице 3. В первом столбце таблицы перечислены первые 8 минимумов функции, последняя строка соответствует минимумам с большими значениями функции. В однородных алгоритмах выполнялось по 500 итераций, в неоднородном – 1000, 3000 и 5000 итераций. Видно, что с увеличением числа итераций неоднородный алгоритм все чаще обнаруживает глобальный экстремум, в то время как однородные алгоритмы глобальный минимум практически не находят.

Таблица 3

R	GA 500	CGA 500	NCGA 1000	NCGA 3000	NCGA 5000
0	0	2	0	17	64
1	5	7	0	36	31
2	15	16	3	29	4
3	26	19	7	12	1
4	9	20	3	2	0
5	16	13	13	4	0
6	10	5	13	0	0
7	3	10	11	0	0
>7	16	8	50	0	0

8. Параллельная реализация

Была выполнена MPI-реализация предложенного варианта клеточных генетических алгоритмов, в которой клеточное поле делилось на горизонтальные полосы (рис. 2).

Такая схема подходит для систем с небольшим числом процессоров. Например, на рис. 3 показана зависимость ускорения от числа используемых процессоров. Если число доступных процессоров является большим, то имеет смысл рассмотреть аналогичную, но чуть более сложную реализацию, в которой поле делится на квадратные блоки.

9. Заключение

В результате выполненной работы были получены следующие результаты: введено понятие неоднородного клеточного генетического алгоритма; проведено численное сравнение рабо-

ты однородного генетического алгоритма, однородного клеточного генетического алгоритма и неоднородного генетического алгоритма на трех задачах непрерывной многомерной оптимизации; показаны преимущества предложенного подхода; предложена схема параллельной MPI-реализации клеточных генетических алгоритмов.

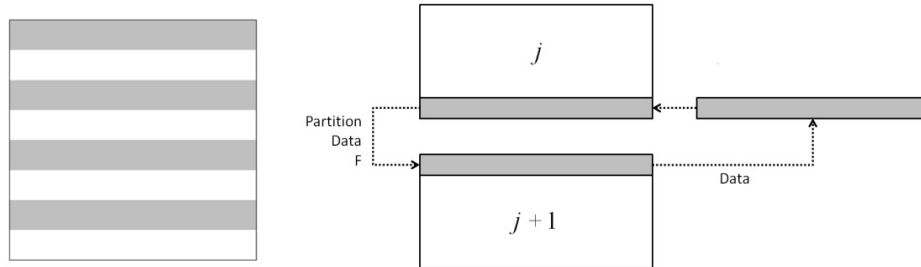


Рис. 2. Схема распараллеливания клеточного генетического алгоритма

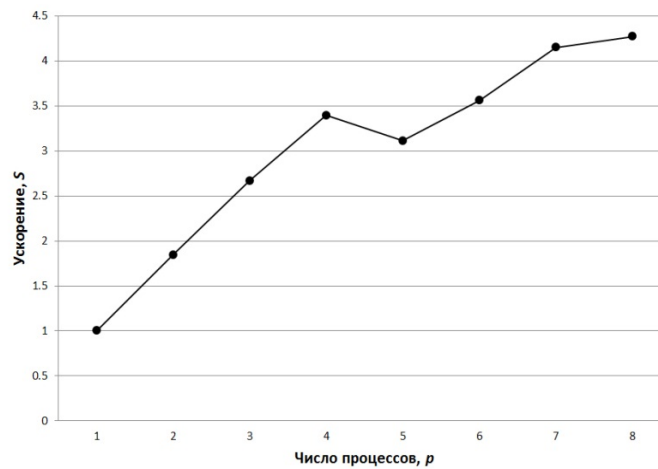


Рис. 3. Зависимость ускорения S от числа процессоров p при параллельной реализации клеточных генетических алгоритмов

Список литературы

- Alba E. and Dorronsoro B.* Cellular Genetic Algorithms. — Springer, 2008.
- Ke Tang, Xiaodong Li, Suganthan P.N., Zhenyu Yang, Weise T.* Benchmark Functions for the CEC'2010 Special Session and Competition on Large-Scale Global Optimization, Technical report, University of Science and Technology of China (USTC). — 2010.
- Whitley D.* A genetic algorithm tutorial // Statistics and Computing. — 1994. — Vol. 4, No. 2. — P. 65–85.