УДК 004.85

МЕТОД ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ЭВОЛЮЦИОННЫХ АЛГОРИТМОВ С ПОМОЩЬЮ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

А.С. Буздалова, М.В. Буздалов

Предлагается метод скалярной оптимизации, основанный на применении эволюционных алгоритмов, контролируемых с помощью обучения с подкреплением. Обучение применяется для динамического выбора наиболее эффективной функции приспособленности для каждого вновь генерируемого поколения эволюционного алгоритма. Представлены результаты эксперимента по решению модельной задачи H-IFF с помощью предлагаемого метода. Проведено сравнение разработанного метода с методами многокритериальной оптимизации. Эксперименты показывают, что предлагаемый метод позволяет повысить эффективность работы эволюционных алгоритмов.

Ключевые слова: скалярная оптимизация, многокритериальная оптимизация, обучение с подкреплением, эволюционные алгоритмы, H-IFF.

Введение

Существуют различные способы повышения эффективности скалярной оптимизации. Некоторые из них основаны на использовании вспомогательных критериев. Например, задача скалярной оптимизации может быть преобразована в задачу многокритериальной оптимизации путем разработки дополнительных критериев, обладающих определенными заранее заданными свойствами, что позволяет избежать остановки поиска решения в локальном оптимуме [1]. Также в качестве источника вспомогательных критериев может выступать предметная область [2]. В этом случае свойства критериев чаще всего заранее не известны, причем они могут меняться в зависимости от того, на каком этапе находится процесс оптимизации. Важно отметить, что в настоящей работе задача оптимизации самих вспомогательных критериев не ставится. В то же время в традиционной теории многокритериальной оптимизации одинаково важны все критерии [3]. В поставленной задаче должен быть оптимизирован только один целевой критерий.

В данной работе предлагается метод повышения эффективности скалярной оптимизации с использованием вспомогательных критериев. Предполагается, что набор критериев задан заранее и об их свойствах ничего не известно. Таким образом, возникает задача скалярной оптимизации со вспомогательными критериями. Задача решается с применением эволюционного алгоритма (ЭА) (evolutionary algorithm, EA) [4], настраиваемого во время выполнения с помощью обучения с подкреплением (reinforcement learning, RL) [5, 6]. В дальнейшем предлагаемый метод будет называться EA+RL.

Метод EA+RL позволяет выбирать из заранее подготовленного набора наиболее эффективную функцию приспособленности (ФП), соответствующую критерию оптимизации, для генерации каждого последующего поколения эволюционного алгоритма. В других существующих методах настройки эволюционных алгоритмов обычно настраиваются вещественные параметры фиксированной ФП, причем настройка ФП освещена в литературе в меньшей степени, чем настройка иных параметров ЭА [7, 8].

Новизна предлагаемого подхода заключается в применении обучения с подкреплением для настройки ЭА. Обучение с подкреплением является современной развивающейся технологией, применимость которой в различных областях человеческой деятельности находится в процессе исследования [6]. Насколько известно авторам, существуют лишь две работы, в которых исследуется возможность использования обучения с подкреплением для настройки ЭА [8, 9]. В обеих работах рассматривается настройка вещественных параметров, таких как, например, вероятность мутации или размер поколения. Данная работа вносит вклад в исследование применимости обучения с подкреплением к настройке ФП.

В предыдущей работе авторов настоящей статьи [10] предложен прототип разработанного метода, предназначенный для решения конкретной модельной задачи с помощью настройки генетического алгоритма (ГА). Результаты, представленные в той работе, подтверждают, что разрабатываемый метод позволяет динамически выбирать наиболее эффективную ФП. В настоящей работе формулируется задача скалярной оптимизации со вспомогательными критериями, что позволяет дать обобщенное описание предлагаемого метода EA+RL, применимое для решения любой задачи, сводящейся к сформулированной. Эффективность предлагаемого метода протестирована на задаче оптимизации функции H-IFF, применяющейся для тестирования ГА, а также для иллюстрации методов повышения эффективности скалярной оптимизации путем сведения ее к многокритериальной оптимизации. Проведено сравнение EA+RL с упомянутыми методами.

Задача оптимизации со вспомогательными критериями

Рассмотрим формализацию задачи скалярной оптимизации со вспомогательными критериями. Обозначим как W дискретное пространство, в котором осуществляется поиск решений. Пусть $X \subset W$ – множество допустимых решений, содержащихся в пространстве поиска. Определим целевой критерий $g : W \to R.$ Определим множество H, состоящее из k вспомогательных критериев: $H = \{h_i\}_{i=1}^{k}, h_i : W \to R.$ Целью описываемой задачи является максимизация целевого критерия g с использованием вспомога-

тельных критериев H для ускорения процесса оптимизации: $g(x) \to \max_{x \in X}$. Решением задачи является $x^* \in X : g(x^*) \geq g(x), \forall x \in X$.

В общем случае характер корреляции между целевым и вспомогательными критериями неизвестен. Однако на практике часто возникает ситуация, при которой некоторые вспомогательные критерии коррелируют с целевым, по крайней мере, на некоторых этапах процесса оптимизации [2]. Предположение о том, что некоторые вспомогательные критерии обладают подобными полезными свойствами, позволяет использовать их для ускорения процесса оптимизации.

Задача обучения с подкреплением

Опишем задачу повышения эффективности ЭА, решающего задачу скалярной оптимизации со вспомогательными критериями, как задачу обучения с подкреплением [5]. Для этого достаточно задать множество действий агента A, способ определения состояний среды $s \in S$ и функцию вознаграждения $K : S \times A \to X \subseteq R$.

Будем обозначать особи, выращиваемые ЭА, как x. Пусть G_i – i-ое поколение. Множество действий A соответствует множеству функций приспособленности, состоящему из g – целевой ФП и элементов множества H – вспомогательных ФП. Применение действия реализуется как выбор некоторой ФП $f_i \in A$ в качестве функции, используемой для оценки приспособленности особей ЭА, и формирования поколения G_i: $A = H \cup g$.

Введем обозначение для лучшей особи поколения G_i, обладающей максимальным значением выбранной для этого поколения ФП f_i: $z_i = \arg\max_{x \in G_i} f_i(x)$. Также введем обозначение для нормированной разности значений некоторой ФП, вычисленной на лучших особях двух последовательных поколений: $\Delta(f,i) = \dfrac{f(z_i) - f(z_{i-1})}{f(z_i)}, f \in A$.

Каждому поколению ЭА поставим в соответствие состояние среды. Состояние s_i, соответствующее поколению G_i, представляет собой вектор ФП $f \in A$, упорядоченный по убыванию значений нормированных разностей $\Delta(f,i)$: $s_i = \langle f_1, f_2 \ldots f_{k+1} \rangle, \Delta(f_1,i) \geq \Delta(f_2,i) \geq \ldots \geq \Delta(f_{k+1},i)$. В том случае, если для некоторых f_a, f_b значение $\Delta(f_a,i)$ совпадает со значением $\Delta(f_b,i)$, функции f_a, f_b располагаются в заранее установленном порядке. Например, пусть число вспомогательных ФП k=2 и в некотором поколении G_i выполняется неравенство $\Delta(h_2,i) = \Delta(g,i) > \Delta(h_1,i)$. Тогда соответствующее состояние среды может иметь вид $s_i = \langle h_2, g, h_1 \rangle$ или $s_i = \langle g, h_2, h_1 \rangle$ в зависимости от начальной договоренности.

В заключение определим функцию вознаграждения $K : S \times A \to R$, которая вычисляется после выбора действия f_i в состоянии s_{i-1} и генерации поколения G_i: $K(s_{i-1}, f_i) = g(z_i) - g(z_{i-1})$. Таким образом, вознаграждение зависит от разности значений целевой ФП, посчитанной на лучших особях двух последовательных поколений. Значение вознаграждения наиболее высоко, когда целевая ФП растет. Заметим, что в обучении с подкреплением целью агента является максимизация суммарной награды, причем для ряда алгоритмов обучения с подкреплением доказана их сходимость к оптимальной стратегии поведения [11]. Следовательно, задача обучения с подкреплением определена таким образом, что оптимальные действия агента будут приводить к максимизации прироста целевой ФП.

Описание алгоритма EA+RL

Предлагаемый метод позволяет управлять ходом выполнения эволюционного алгоритма путем назначения текущей ФП для каждого вновь сгенерированного поколения. Можно выделить две независимые сущности, составляющие основу метода: модуль обучения и эволюционный алгоритм. Будем называть эволюционный алгоритм средой обучения. Модулю обучения могут быть переданы награда и состояние среды. Он способен сообщать действие, которое необходимо применить к среде. В листинге представлен псевдокод разработанного алгоритма.

```
1. Установить номер текущего поколения: i ← 0
2. Сгенерировать начальное поколение G0
3. ПОКА (условие останова ЭА не выполнено)
4.     Вычислить состояние si и передать его модулю обучения
5.     Получить ФП для следующего поколения fi+1 из модуля обучения
6.     Сгенерировать следующее поколение Gi+1
```

Модуль обучения может быть реализован на основе произвольного алгоритма обучения с подкреплением и взаимодействовать с произвольным эволюционным алгоритмом. В ходе выполнения работы было реализовано четыре различных алгоритма обучения: Q-learning [6], Delayed Q-learning [11], Dyna [5] и R-learning [5]. Для обозначения различных реализаций предлагаемого метода EA+RL будем заменять в названии метода «EA» на название конкретного эволюционного алгоритма, «RL» – на название алгоритма обучения с подкреплением. Например, если с помощью предлагаемого метода реализуется контроль над ГА с помощью алгоритма обучения Q-learning, то соответствующая реализация метода будет называться ГА+Q-learning.

Модельная задача H-IFF

Определим задачу скалярной оптимизации функции H-IFF (Hierarchical-if-and-only-if function) [1]. Пространство поиска состоит из битовых строк $B = b_1 b_2 \ldots b_l$ фиксированной длины l. Требуется максимизировать функцию H-IFF:

$$f(B) = \begin{cases} 1, |B| = 1; \\ |B| + f(B_L) + f(B_R), |B| > 1 \wedge (\forall i\{b_i = 0\} \vee \forall i\{b_i = 1\}); \\ f(B_L) + f(B_R), \textit{иначе.} \end{cases}$$

Функция задана таким образом, что существует два оптимальных решения: строка, полностью состоящая из единиц, и строка, полностью состоящая из нулей. Особенностью задачи является то, что поиск ее оптимального решения с помощью эволюционных алгоритмов часто останавливается в локальном оптимуме. Существует подход к решению этой проблемы, при котором скалярная задача оптимизации H-IFF заменяется многокритериальной задачей оптимизации функции MH-IFF [1]. Вместо исходной функции f вводятся функции f_0 и f_1:

$$f_n(B) = \begin{cases} 0, |B| = 1 \wedge b_1 \neq n; \\ 1, |B| = 1 \wedge b_1 = n; \\ |B| + f_n(B_L) + f_n(B_R), |B| > 1 \wedge \forall i\{b_i = n\}; \\ f_n(B_L) + f_n(B_R), \textit{иначе.} \end{cases}$$

Затем проводится максимизация предложенных функций с помощью алгоритмов многокритериальной оптимизации. Этот подход позволяет найти решения с более высокими значениями исходной функции, чем подход, основанный на скалярной оптимизации.

Задача максимизации функции H-IFF может быть представлена как задача скалярной оптимизации целевой функции $g = f$ со вспомогательными критериями $H = \{f_0, f_1\}$. Подобное представление задачи позволяет использовать предлагаемый метод для повышения эффективности эволюционных алгоритмов, применяемых для ее решения.

Описание и результаты эксперимента

В ходе эксперимента было реализовано решение задачи оптимизации H-IFF предлагаемым методом. Использовалось два различных эволюционных алгоритма: генетический алгоритм (ГА) и $(1 + m)$-эволюционная стратегия (ЭС). В ГА с вероятностью 70% применялся оператор одноточечного кроссовера и оператор мутации, инвертирующий каждый бит каждой особи с вероятностью $2 / l$. В ЭС оператор мутации инвертировал один бит каждой особи, выбранный случайным образом.

Параметры эксперимента соответствовали параметрам, примененным в работе [1], что позволяет сравнить новые результаты с результатами, полученными ее авторами. Длина особи составляла 64 бита. Соответствующее максимально возможное значение H-IFF равно 448. В табл. 1 представлены результаты оптимизации функций H-IFF и MH-IFF с помощью алгоритмов скалярной и многокритериальной оптимизации соответственно. Результаты отсортированы по среднему значению целевой ФП лучших особей, полученных в результате 30 запусков соответствующих алгоритмов. Вычисления запускались на фиксированное число поколений, равное 500000. Успешными считаются запуски, в которых была выращена особь с максимальной приспособленностью. Алгоритмы 1, 2, 4, 5, 7 реализованы с помощью предлагаемого метода с использованием различных алгоритмов обучения. Результаты 3, 6, 9, 11 получены авторами статьи, причем алгоритмы 3 и 6 (PESA и PAES) являются алгоритмами многокритериальной оптимизации. Можно видеть, что предлагаемый метод в случае использования алгоритма обучения R-learning [5] позволяет преодолеть проблему остановки в локальном оптимуме столь же эффективно, как и метод PESA, и более эффективно, чем метод PAES.

Также был проведен эксперимент, показавший, что если среди вспомогательных ФП есть мешающая ФП, оптимизация по которой ведет к убыванию целевой ФП, предлагаемый метод по-прежнему эффективен, с его помощью удается вырастить оптимальную особь в 92% запусков. Однако алгоритмы

многокритериальной оптимизации в этом случае не позволяют выращивать особи с максимальным значением целевой ФП, так как они оптимизируют все предложенные критерии, в том числе мешающий.

№	Алгоритм	Лучшее значение	Среднее значение	σ	% успешных запусков
1	(1+10)-ЭС + R-learning	448	448,00	0,00	100
2	ГА + R-learning	448	448,00	0,00	100
3	PESA	448	448,00	0,00	100
4	ГА + Q-learning	448	435,61	32,94	87
5	ГА + Dyna	448	433,07	38,07	80
6	PAES	448	418,13	50,68	74
7	ГА + Delayed QL	448	397,18	49,16	53
8	ГА + Random	384	354,67	29,24	0
9	DCGA	448	323,93	26,54	3
10	ГА	384	304,53	27,55	0
11	SHC	336	267,47	29,46	0
12	(1+10) -ЭС	228	189,87	17,21	0

Таблица 1. Результаты оптимизации H-IFF и MH-IFF

В табл. 2 отдельно рассмотрена оптимизация H-IFF с использованием ЭС. Применяемая ЭС устроена таким образом, что решает задачу весьма неэффективно: ни в одном из запусков не удается получить особь с максимальной приспособленностью. Однако применение предлагаемого метода позволяет добиться выращивания оптимальной особи в 73% запусков в случае использования наименее эффективной (1+1)-ЭС и в 100% запусков в остальных рассмотренных случаях.

№	Алгоритм	Лучшее значение	Среднее значение	σ	% успешных запусков
1	(1+10)-ЭС + R-learning	448	448,00	0,00	100
2	(1+10)-ЭС	228	189,87	17,21	0
3	(1 + 5)-ЭС + R-learning	448	448,00	0,00	100
4	(1 + 5)-ЭС	216	179,07	16,99	0
5	(1 + 1)-ЭС + R-learning	448	403,49	59,48	73
6	(1 + 1)-ЭС	188	167,07	11,98	0

Таблица 2. Результаты оптимизации H-IFF с помощью эволюционных стратегий. Алгоритмы 1, 3, 5 реализованы с применением предлагаемого метода

Заключение

Предложен метод, повышающий эффективность скалярной оптимизации со вспомогательными критериями. Метод основан на выборе функции приспособленности эволюционного алгоритма с помощью обучения с подкреплением. Работа вносит вклад в исследование применимости обучения с подкреплением для настройки эволюционных алгоритмов. В ходе эксперимента подтверждена эффективность метода, а также проведено его сравнение с методами многокритериальной оптимизации. Предлагаемый метод, примененный к (1+m) эволюционным стратегиям для решения задачи оптимизации функции H-IFF, позволяет получать особи с максимальной возможной приспособленностью в 73–100% запусков, в то время как с помощью эволюционных стратегий без обучения не удается вырастить оптимальную особь.

Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы.

Литература

1. Knowles J.D., Watson R.A., Corne D. Reducing Local Optima in Single-Objective Problems by Multi-objectivization // Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization EMO '01. – London, UK: Springer -Verlag. – 2001. – P. 269–283.

2. Буздалов М.В. Генерация тестов для олимпиадных задач по теории графов с использованием эволюционных алгоритмов. Магистерская диссертация. СПбГУ ИТМО, 2011 [Электронный ресурс]. – Режим доступа: http://is.ifmo.ru/papers/2011-master-buzdalov/, свободный. Яз. рус. (дата обращения 21.06.2012).
3. Лотов А.В., Поспелова И.И. Многокритериальные задачи принятия решений: Учебное пособие. – М.: МАКС Пресс, 2008. – 197 с.
4. Luke S. Essentials of Metaheuristics [Электронный ресурс]. – Режим доступа: http://cs.gmu.edu/~sean/book/metaheuristics/, свободный. Яз. англ. (дата обращения 21.06.2012).
5. Kaelbling L.P., Littman M.L., Moore A.W. Reinforcement Learning: A Survey // Journal of Artificial Intelligence Research. – 1996. – V. 4. – P. 237–285.
6. Gosavi A. Reinforcement Learning: A Tutorial Survey and Recent Advances // INFORMS Journal on Computing. – 2009. – V. 21. – № 2. – P. 178–192.
7. Eiben A.E., Michalewicz Z., Schoenauer M., Smith J.E. Parameter Control in Evolutionary Algorithms // In Parameter Setting in Evolutionary Algorithms. – 2007. – P. 19–46.
8. Müller S., Schraudolph N.N., Koumoutsakos P.D. Step Size Adaptation in Evolution Strategies using Reinforcement Learning // Proceedings of the Congress on Evolutionary Computation, IEEE. – 2002. – P. 151–156.
9. Eiben A.E., Horvath M., Kowalczyk W., Schut M.C. Reinforcement Learning For Online Control Of Evolutionary Algorithms // Proceedings of the 4th International Conference On Engineering Self-Organising Systems ESOA'06. – Springer -Verlag, Berlin, Heidelberg, 2006. – P. 151–160.
10. Афанасьева А.С., Буздалов М.В. Выбор функции приспособленности особей генетического алгоритма с помощью обучения с подкреплением // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 1 (77). – С. 77–81.
11. Strehl A.L., Li L., Wiewora E., Langford J., Littman M.L. PAC Model-Free Reinforcement Learning // ICML'06: Proceedings of the 23rd International Conference On Machine Learning. – 2006. – P. 881–888.

Буздалова Арина Сергеевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, aduzdalova@gmail.com

Буздалов Максим Викторович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, mbuzdalov@gmail.com