ВЫБОР ФУНКЦИИ ПРИСПОСОБЛЕННОСТИ ОСОБЕЙ ЭВОЛЮЦИОННОГО АЛГОРИТМА С ПОМОЩЬЮ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

А. С. Афанасьева

студентка кафедры компьютерных технологий;
afanasyevarina@gmail.com

Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики

Аннотация: В данной работе предлагается метод, позволяющий динамически выбирать вспомогательную функцию приспособленности, наиболее выгодную для использования в эволюционном алгоритме. Метод основан на применении обучения с подкреплением. Приведены экспериментально полученные результаты его использования для решения модельной задачи, а также результаты, позволяющие сравнить предлагаемый метод с алгоритмами многокритериальной оптимизации.

Введение

В теории оптимизации известны задачи скалярной и многокритериальной оптимизации [1]. На практике возникает возможность решать модифицированную скалярную задачу оптимизации, в которой присутствуют дополнительные критерии [2]. Целью ее решения является максимизация единственной целевой функции, а дополнительные критерии имеют лишь вспомогательное значение. Целевая функция может некоторым образом зависеть от дополнительных критериев, поэтому в некоторых случаях вместо максимизации целевой функции оказывается выгодным оптимизировать дополнительные критерии.

Важным классом алгоритмов оптимизации являются эволюционные алгоритмы (ЭА), где в качестве критерия выступает целевая функция приспособленности (ФП). При наличии нескольких вспомогательных функций приспособленности выбор наиболее выгодной из них приходится производить вручную [2]. Подобный подход не вполне эффективен, так как предполагает многократный перезапуск ЭА.

Целью описываемых исследований является разработка метода, позволяющего автоматически выбирать из заранее подготовленного набора такие вспомогательные ФП, применение которых способствует ускорению «выращивания» особей с высокими значениями целевой ФП. В литературе встречаются разработки по автоматической настройке значений параметров ГА, таких как вероятность применения генетических операторов или число особей в поколении, а также по настройке некоторой фиксированной ФП [5].

Предлагаемый метод отличается от существующих подходов к настройке ГА тем, что предполагает выбор между качественно разными ФП.

Выбор ФП предлагается осуществлять с помощью обучения с подкреплением [3, 4]. Отметим, что применение обучения с подкреплением к оптимизации работы эволюционных алгоритмов мало исследовано на данный момент [5]. В предлагаемой работе подобный подход впервые применяется для контроля ФП.

Существуют также подходы к улучшению производительности алгоритмов однокритериальной оптимизации, основанные на введении многокритериальности [6]. Их применение предполагает разработку дополнительных ФП, таких что получающаяся задача многокритериальной оптимизации решается проще, чем исходная задача. Отличие предлагаемого метода от многокритериального подхода состоит в том, что вспомогательные ФП задаются заранее и могут не коррелировать с целевой функцией. Ниже будут описаны результаты эксперимента, позволяющего сравнить производительность предлагаемого метода и алгоритмов многокритериальной оптимизации.

Описание предлагаемого подхода

Для выбора вспомогательной ФП используется обучение с подкреплением [3, 4]. В алгоритмах этого типа обучение происходит одновременно с применением накопленного опыта, что позволяет выбрать и применить оптимальные ФП в течение одного запуска ЭА.

В алгоритмах обучения с подкреплением агент применяет действия к среде, которая отвечает на каждое действие наградой. В разработанном методе действие агента состоит в выборе ФП для каждого вновь сформированного поколения ЭА. Вознаграждение тем выше, чем существеннее рост значений целевой ФП. На рис. 1 представлена схема предлагаемого метода.

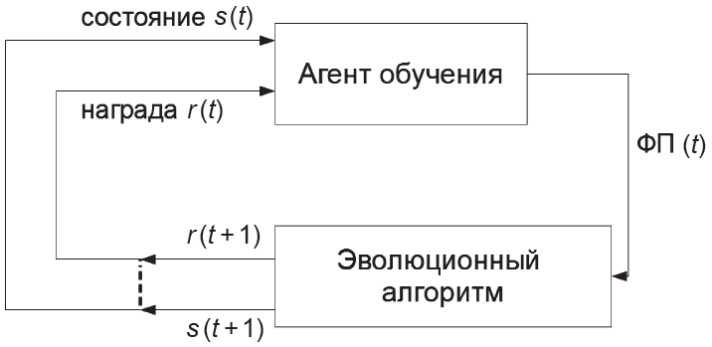

Рис. 1. Схема предлагаемого метода,
t — номер поколения ЭА

Алгоритмы обучения с подкреплением нацелены на формирование оптимальной стратегии поведения агента, следование которой приводит к максимизации суммарного вознаграждения, а следовательно, к ускорению роста целевой ФП.

Результаты экспериментов

Эффективность предлагаемого метода была проверена экспериментальным способом. Было реализовано несколько различных эволюционных алгоритмов и контролирующих их алгоритмов обучения с подкреплением, а именно Q-learning [3], Delayed Q-learning [7], R-learning [8] и Dyna [3]. В ходе первого эксперимента предлагаемый метод использовался для решения модельной задачи, в которой на различных этапах оптимизации выгодны различные вспомогательные ФП. В ходе второго эксперимента было проведено сравнение с методами многокритериальной оптимизации на примере задачи H-IFF [6], применяющейся для тестирования генетических алгоритмов.

Результаты решения модельной задачи

Рассмотрим постановку следующей модельной задачи. Особи представлены битовыми строками. Пусть x бит равны единице. Определим функции приспособленности. Целевая ФП задается формулой $g(x) = [x/k]$. Вспомогательные ФП имеют следующий вид:

$$h_1(x) = \begin{cases} x, x < p \\ p, x \geq p \end{cases}, \quad h_2(x) = \begin{cases} p, x < p \\ x, x \geq p \end{cases}.$$

Будем называть p точкой переключения.

Можно видеть, что для особей с числом единиц меньшим значения точки переключения выгоднее использовать функцию h_1 в качестве текущей ФП. Для особей, число единиц в представлении которых превышает значение точки переключения, выгодно использовать h_2.

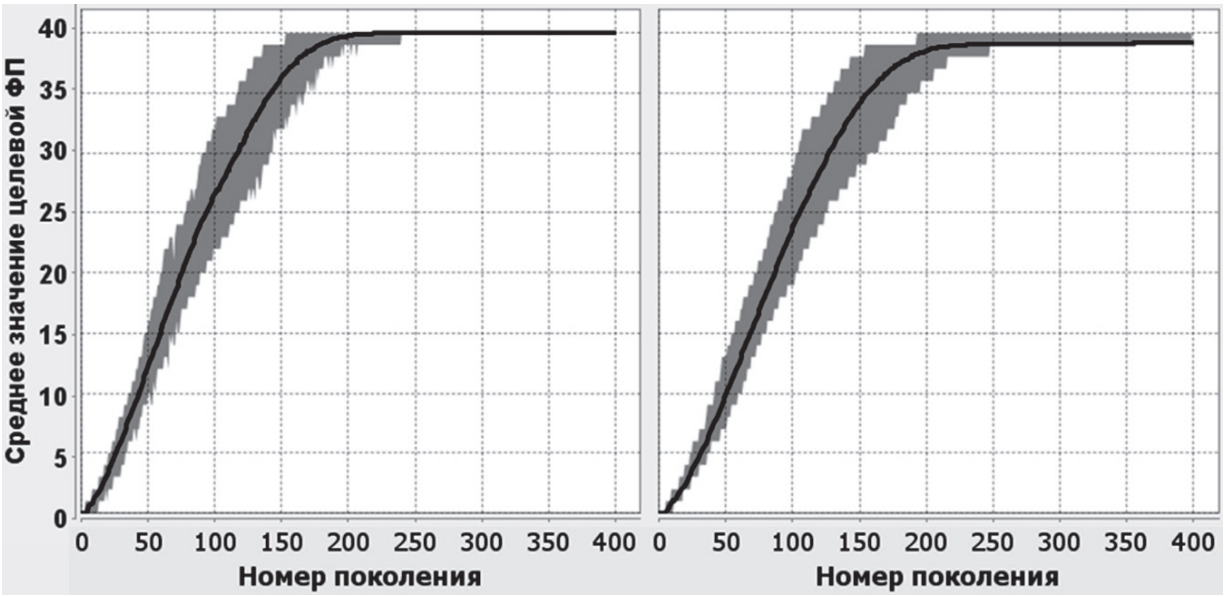

Рис. 2. Сравнительные графики роста ФП лучшей особи во время работы алгоритма, использующего Delayed Q-learning [7] (слева) и обычного ГА (справа)

Поставлен эксперимент, заключающийся в многократных запусках ГА с обучением и обычного ГА. На рис. 2 представлены графики зависимости усредненной целевой ФП лучшей выращенной в текущем поколении особи от номера поколения. Результаты эксперимента подтверждают, что применение обучения ускоряет получение особей с более высокими значениями целевой ФП.

Результаты оптимизации функции H-IFF

Определим задачу скалярной оптимизации функции H-IFF (Hierarchical-if-and-only-if function). Пространство поиска состоит из битовых строк $B = b_1, b_2, \ldots, b_l$ фиксированной длины l. Требуется максимизировать функцию H-IFF [6]:

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

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 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 [6]. Вместо исходной функции f вводятся функции f_0 и f_1:

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

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

В ходе эксперимента было реализовано новое решение задачи оптимизации H-IFF с использованием предлагаемого метода. В качестве целевой ФП выступала функция f. В качестве вспомогательных ФП были взяты функции f_0 и f_1, применяемые при оптимизации MH-IFF. 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 статье [6], 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 (M)H-IFF с помощью различных алгоритмов. Результаты отсортированы по среднему значению целевой ФП лучших особей, полученных в результате 30 запусков соответствующих алгоритмов. Вычисления запускались на фиксированное число поколений. 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 [8] позволяет преодолеть проблему остановки в локальном оптимуме столь же эффективно, как и метод PESA, и более эффективно, чем метод PAES.

Таблица 1

Результаты оптимизации (M)H-IFF.
Алгоритмы 1, 2, 4, 5, 7 реализованы с применением предлагаемого метода

№	Алгоритм	Лучшее значение	Среднее значение	σ	% одного оптимума	% двух оптимумов
1	(1+10)-ЭС+ R-learning	448	448,00	0,00	100	40
2	ГА + R-learning	448	448,00	0,00	100	10
3	PESA	448	448,00	0,00	100	100
4	ГА + Q-learning	448	435,61	32,94	87	3
5	ГА + Dyna	448	433,07	38,07	80	0
6	PAES	448	418,13	50,68	74	43
7	ГА + Delayed QL	448	397,18	49,16	53	0
8	ГА + Random	384	354,67	29,24	0	0
9	DCGA	448	323,93	26,54	3	0
10	ГА	384	304,53	27,55	0	0
11	SHC	336	267,47	29,46	0	0
12	(1+10)-ЭС	228	189,87	17,21	0	0

В табл. 2 отдельно рассмотрена оптимизация H-IFF с применением ЭС. Применяемая ЭС устроена таким образом, что решает задачу весьма неэффективно. Однако применение предлагаемого метода позволяет увеличить среднее значение целевой ФП примерно в два раза.

Таблица 2

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

№	Алгоритм	Лучшее значение	Среднее значение	σ	% одного оптимума	% двух оптимумов
1	(1+10)-ЭС+R-learning	448	448,00	0,00	100	40
2	(1+10)-ЭС	228	189,87	17,21	0	0
3	(1+5)-ЭС+R-learning	448	448,00	0,00	100	37
4	(1+5)-ЭС	216	179,07	16,99	0	0
5	(1+1)-ЭС+R-learning	448	403,49	59,48	73	10
6	(1+1)-ЭС	188	167,07	11,98	0	0

Заметим, что целью предлагаемого метода является ускорение получения особей с высокими значениями целевой ФП. Приведенные результаты свидетельствуют о том, что метод успешно справляется с этой задачей. Невысокий процент нахождения обоих возможных оптимумов объясняется тем, что предлагаемый метод не проводит многокритериальную оптимизацию. Ожидается, что в случаях, когда среди вспомогательных ФП есть ФП, не коррелирующие с целевой, предлагаемый метод будет эффективнее методов многокритериальной оптимизации.

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 контроля работы эволюционных алгоритмов, настройка выбора функций приспособленности с помощью обучения проведена впервые. В ходе экспериментов подтверждена эффективность метода: метод позволяет динамически выбирать наиболее выгодную функцию приспособленности и избегать остановки в локальном оптимуме. Метод, примененный к $(1+m)$-эволюционным стратегиям, позволяет в два раза увеличивать значение целевой функции приспособленности особей, выращиваемых за фиксированное число поколений.

Литература

1. *Лотов А. В., Поспелова И. И.* Многокритериальные задачи принятия решений: Учебное пособие. М.: МАКС Пресс. 2008.
2. *Буздалов М. В.* Генерация тестов для олимпиадных задач по теории графов с использованием эволюционных алгоритмов. Магистерская диссертация. СПбГУ ИТМО. http://is.ifmo.ru/papers/2011-master-buzdalov/ [дата просмотра: 10.05.2012]

3. *Sutton R. S., Barto A. G.* Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, 1998.
4. *Gosavi A.* Reinforcement Learning: A Tutorial Survey and Recent Advances // INFORMS Journal. Vol. 21. No. 2. 2009. P. 178–192.
5. *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.
6. *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.
7. *Strehl A. L., Li L., Wiewora E., Langford J., Littman M. L.* PAC model-free reinforcement learning // Proceedings of the 23rd international conference on Machine learning. ICML'06. 2006. P. 881–888.
8. *Mahadevan S.* Average reward reinforcement learning: foundations, algorithms, and empirical results. Machine Learning 22, 1–3. 1996. P. 159–195.