

УДК 004.023:004.855.5:004.832.23

ТЕОРЕТИЧЕСКИЙ АНАЛИЗ МЕТОДА ВЫБОРА ПЕРЕКЛЮЧАЮЩИХСЯ ВСПОМОГАТЕЛЬНЫХ КРИТЕРИЕВ НА ЗАДАЧЕ XdivK

И.А. Петрова^a, А.С. Буздалова^a, А.А. Шальто^a

^a Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

Адрес для переписки: irenepetrova@yandex.com

Информация о статье

Поступила в редакцию 02.03.17, принята к печати 30.03.17

doi: 10.17586/2226-1494-2017-17-3-409-416

Язык статьи – русский

Ссылка для цитирования: Петрова И.А., Буздалова А.С., Шальто А.А. Теоретический анализ метода выбора переключающихся вспомогательных критериев на задаче XdivK // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 3. С. 409–416. doi: 10.17586/2226-1494-2017-17-3-409-416

Аннотация

Предмет исследования. Проведен анализ причин неэффективности метода EA+RL на задаче оптимизации XdivK с переключающимися критериями. Предложена модификация метода EA+RL. Метод EA+RL предназначен для повышения эффективности однокритериальных эволюционных алгоритмов путем введения вспомогательных критериев. Задача XdivK характеризуется большим числом локальных оптимумов. Переключающиеся критерии оказывают помощь на одних этапах оптимизации и позволяют избежать остановки процесса оптимизации в локальных оптимумах, но мешают на других этапах. **Метод.** Для проведения теоретического анализа метода EA+RL и предложенной его модификации построены марковские цепи, моделирующие процесс оптимизации XdivK. На основе анализа вероятностей переходов в марковских цепях произведена оценка числа вычислений функции приспособленности, необходимого для нахождения оптимума XdivK. **Основные результаты.** Произведен теоретический анализ метода EA+RL и предложенной его модификации на задаче XdivK с критериями, эффективность которых меняется в зависимости от этапа оптимизации. Приведено доказательство того, что предложенная модификация, в отличие от метода EA+RL, позволяет игнорировать критерий, являющийся мешающим на данном этапе оптимизации. Получена оценка времени работы предложенной модификации. **Практическая значимость.** Предложенная в работе модификация позволяет повысить эффективность метода EA+RL, успешно применяемого для решения NP-трудных практических задач, в том числе задачи генерации тестов.

Ключевые слова

эволюционный алгоритм, обучение с подкреплением, марковские цепи, анализ времени работы

Благодарности

Исследование И.А. Петровой и А.С. Буздаловой выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-31-00380 мол. а.

THEORETICAL ANALYSIS OF DYNAMIC SELECTION OF SWITCHING AUXILIARY OBJECTIVES ON XdivK PROBLEM

I.A. Petrova^a, A.S. Buzdalova^a, A.A. Shalyto^a

^a ITMO University, Saint Petersburg, 197101, Russian Federation

Corresponding author: irenepetrova@yandex.com

Article info

Received 02.03.17, accepted 30.03.17

doi: 10.17586/2226-1494-2017-17-3-409-416

Article in Russian

For citation: Petrova I.A., Buzdalova A.S., Shalyto A.A. Theoretical analysis of dynamic selection of switching auxiliary objectives on XdivK problem. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2017, vol. 17, no. 3, pp. 409–416 (in Russian). doi: 10.17586/2226-1494-2017-17-3-409-416

Abstract

Subject of Research. The paper deals with analysis of the EA+RL method inefficiency reasons on XdivK optimization problem with switching auxiliary objectives. It is proposed to modify the EA+RL method. The EA+RL method increases efficiency of an evolutionary algorithm by introducing auxiliary objectives. The XdivK problem is characterized by a large number of local optima. Switching objectives help to escape from local optima on some stages of optimization while being obstructive on the other stages. **Method.** To perform theoretical analysis of the EA+RL method and its proposed

modification, the corresponding optimization process was modeled by Markov chains. The number of fitness function evaluations needed to reach the optimum was estimated based on the analysis of transition probabilities. **Main Results.** The EA+RL method and its proposed modification were theoretically analyzed on the XdivK problem with switching auxiliary objectives. It was proved that the proposed modification ignores obstructive objectives contrary to the EA+RL method. The lower and upper bounds on the running time of the proposed modification were obtained. **Practical Relevance.** The proposed modification increases the efficiency of EA+RL method, successfully used to solve NP-hard optimization problems, such as the test case generation problem.

Keywords

evolutionary algorithm, reinforcement learning, Markov chains, runtime analysis

Acknowledgements

The work of I.A. Petrova and A.S. Buzdalova was financially supported by RFBR research project No. 16-31-00380 mol_a.

Введение

Существуют задачи оптимизации, для которых точный алгоритм решения является неэффективным или его не существует. Примером таких задач служат задача составления расписаний, задача коммивояжера, задача о рюкзаке. Одним из методов решения данных задач является применение эволюционных алгоритмов [1]. Эволюционные алгоритмы (ЭА) основаны на принципах природной эволюции. Кандидаты на оптимальное решение задачи представляются в виде особей эволюционного алгоритма. На каждой итерации алгоритма существует набор особей, называемый поколением. То, насколько особь близка к оптимальному решению, определяется функцией приспособленности (ФП). Для получения следующего поколения к особям применяются операторы скрещивания, мутации и отбора, использующие значения ФП особей. Существуют однокритериальные и многокритериальные ЭА, предназначенные для решения задач однокритериальной и многокритериальной оптимизации соответственно.

Рассматривается задача оптимизации целевого критерия с помощью однокритериального ЭА [1]. Существует два наиболее распространенных способа оценки эффективности ЭА. В одном из них эффективность определяется как максимальное значение целевого критерия, полученное за фиксированное число вычислений ФП. Во втором, применяемом в данной работе, эффективность оценивается как число вычислений функции приспособленности, необходимое для достижения оптимума.

В некоторых случаях эффективность однокритериальной оптимизации может быть повышена путем введения вспомогательных критериев оптимизации [2–5]. Существует несколько способов применения вспомогательных критериев. Одним из методов является одновременная оптимизация всех критериев вместо целевого с помощью многокритериального ЭА [4]. В данном методе вспомогательные критерии должны быть независимы друг от друга и разработаны таким образом, чтобы оптимум целевого критерия лежал на Парето-фронте, получаемом при оптимизации вспомогательных критериев. Создание таких критериев не всегда возможно, поэтому были разработаны методы, динамически выбирающие критерий из целевого и вспомогательных на каждом шаге оптимизации [6, 7]. Критерии могут выбираться случайным образом [6] или с помощью некоторой эвристики [7]. Первый подход является общим, но он не учитывает специфику задачи. Второй подход был разработан специально для решения задачи составления расписаний, и его применение для решения других задач затруднительно. Исходя из этого, был предложен метод EA+RL (Evolutionary Algorithm + Reinforcement Learning) [8, 9] выбора критериев при помощи обучения с подкреплением [10, 11]. На каждом шаге метода EA+RL выбранный критерий – целевой или один из вспомогательных – оптимизируется с помощью однокритериального ЭА.

При решении практических задач вспомогательные критерии зачастую создаются автоматически и могут менять свои свойства в процессе оптимизации [7, 12]. Такие критерии могут быть помогающими на одних этапах оптимизации и мешающими, т.е. приводящими к ухудшению значения целевого критерия, на других. В настоящей работе такие критерии называются *переключающимися*. Эффективность метода EA+RL была доказана теоретически на ряде модельных задач [13–15]. Однако эффективность метода EA+RL в случае использования переключающихся вспомогательных критериев не была исследована. Таким образом, возникает задача проведения теоретического анализа метода EA+RL с переключающимися критериями. Теоретический анализ позволяет лучше понять механизмы работы рассматриваемого алгоритма и предложить его улучшение. Получаемая в ходе анализа асимптотическая оценка времени работы алгоритма является объективным критерием его эффективности.

В данной работе проведен теоретический анализ метода EA+RL в случае использования переключающихся вспомогательных критериев. Рассмотрена модельная задача XdivK [13] с двумя переключающимися вспомогательными критериями, причем на каждом этапе оптимизации один из них более эффективен, чем целевой, а другой является мешающим. Приведено теоретическое объяснение и экспериментальное подтверждение того, что алгоритм EA+RL не способен игнорировать критерий, являющийся мешающим на данном этапе оптимизации. Предложена модификация метода EA+RL, сохраняющая лучшую особь. Теоретически доказано, что, в отличие от метода EA+RL, предложенная модификация игнорирует критерий, являющийся мешающим на данном этапе оптимизации.

Метод EA+RL выбора вспомогательных критериев с помощью обучения с подкреплением

Метод EA+RL основан на обучении с подкреплением [10, 11]. На каждом шаге алгоритма агент обучения выбирает оптимизируемый критерий – целевой или один из вспомогательных, и применяет его к среде, в качестве которой выступает ЭА. ЭА использует выбранный критерий для создания следующего поколения и возвращает агенту так называемую *награду*. Целью агента является максимизация суммарной награды. Награда зависит от разницы максимальных значений целевого критерия в текущем и предыдущем поколениях. Чем выше значение целевого критерия в новом поколении, тем выше награда. Таким образом, максимизация суммарного вознаграждения ведет к максимизации роста целевого критерия за весь период оптимизации.

В настоящей работе рассмотрена следующая реализация метода EA+RL. В качестве ЭА выступает метод спуска со случайными мутациями (Random Local Search, RLS), особями являются битовые строки. В качестве алгоритма обучения с подкреплением использовался алгоритм Q-learning. Таким образом, использованная реализация метода EA+RL обозначается RLS+Q-learning. В методе спуска со случайными мутациями поколение состоит из одной особи. В качестве эволюционного оператора использовался только оператор мутации, инвертирующий случайный бит особи. Состояние среды определяется как значение целевого критерия текущей особи. Награда равна разности значений целевого критерия, вычисленных на особях текущего и предыдущего поколения. В листинге ниже представлен псевдокод алгоритма RLS+Q-learning.

1. Особь $y \leftarrow$ случайная битовая строка
2. Определить множество H из вспомогательных критериев и целевого критерия t
3. $Q(s, h) \leftarrow 0$ для каждого состояния s и критерия $h \in H$
4. ПОКА (оптимум целевого критерия t не найден)
5. Текущее состояние $s \leftarrow t(y)$
6. Особь $y' \leftarrow$ результат применения оператора мутации к y
7. Критерий $h: Q(s, h) = \max_{h' \in H} Q(s, h')$ (Если Q -значения одинаковы, критерии выбираются равновероятно)
8. ЕСЛИ $(h(y') \geq h(y))$ ТОГДА $y \leftarrow y'$
9. Новое состояние $s' \leftarrow t(y)$
10. Награда $r \leftarrow s' - s$
11. $Q(s, h) \leftarrow Q(s, h) + \alpha (r + \max_{h' \in H} Q(s', h') - Q(s, h))$
12. КОНЕЦ ЦИКЛА

Задача XdivK с переключающимися критериями

В задаче XdivK особью является битовая строка длины n . Пусть число единичных битов равно x .

Значение целевого критерия XdivK вычисляется как $t(x) = \lfloor \frac{x}{k} \rfloor$, где k – некоторая константа, $k < n$ и n нацело делится на k . Вводятся две функции OneMax и ZeroMax: значение функции OneMax равно x , а значение функции ZeroMax равно $n - x$. Вспомогательные критерии h_1 и h_2 вычисляются по формуле (1) и могут быть равны как OneMax, так и ZeroMax в зависимости от этапа оптимизации.

$$h_1(x) = \begin{cases} \text{ZeroMax}, & p < x \\ \text{OneMax}, & p \geq x \end{cases} \quad h_2(x) = \begin{cases} \text{OneMax}, & p < x \\ \text{ZeroMax}, & p \geq x \end{cases} \quad (1)$$

Параметр p называется точкой переключения. Критерий h_2 является мешающим при $x \in [0; p]$, а критерий h_1 является мешающим при $x \in (p; n]$.

Использование критерия, равного OneMax, позволяет различить особи с одинаковым значением целевого критерия и выбрать из них ту особь, которая содержит большее число единиц. Такая особь с большей вероятностью породит особь с более высоким значением целевого критерия. В этой связи критерий, равный OneMax, в текущем интервале значений x является эффективным, в то время как критерий, равный ZeroMax, является мешающим.

Оценка времени работы алгоритма RLS+Q-learning на задаче XdivK с переключающимися вспомогательными критериями

Для оценки времени работы алгоритма RLS+Q-learning строится марковская цепь, представляющая соответствующий процесс оптимизации (рис. 1) [15]. Состояние алгоритма обучения с подкреплением равно значению целевого критерия XdivK. Марковское состояние определяется числом единичных бит в особи. Таким образом, в каждом состоянии обучения с подкреплением содержится k марковских состояний. Математическое ожидание времени работы предложенной модификации равно числу вычислений ФП, необходимому для достижения марковского состояния n из нулевого состояния. Каждый переход в марковской цепи соответствует вычислению ФП новой особи, полученной в результате мутации. Поэтому математическое ожидание времени работы алгоритма равно числу сделанных алгоритмом переходов в марковской цепи. Метки на переходах имеют вид F, M , где F соответствует критерию, при выборе

которого может быть осуществлен переход, а M соответствует мутации, приводящей к данному переходу. Далее описано, почему переходы имеют соответствующие метки.

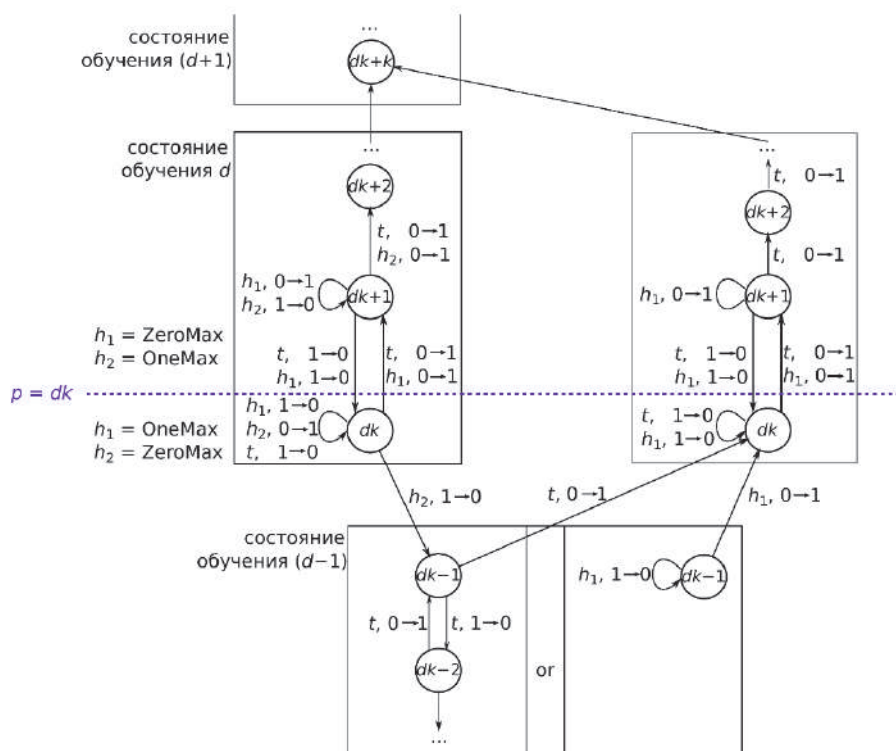


Рис. 1. Марковская цепь для анализа времени работы RLS+Q-learning на задаче XdivK с двумя переключающимися вспомогательными критериями

Пусть $x = dk$, где d – некоторое значение целевого критерия XdivK, и у агента нет опыта в состоянии обучения d . Если агент выбирает критерий, равный ZeroMax, и оператор мутации инвертирует единичный бит, то новая особь содержит $dk-1$ единичный бит и она лучше текущей особи по критерию ZeroMax. В связи с этим для следующего поколения будет выбрана новая особь. Однако значение целевого критерия новой особи меньше, чем у текущей особи. Поэтому агент получает отрицательную награду в состоянии d и переходит в состояние $d-1$. Обозначим критерий, равный OneMax при $x = dk-1$, как g . В состоянии $d-1$ агент может выбрать как критерий t , так и критерий g . Из двух критериев выбирается тот, при использовании которого агент ранее перешел в состояние d из состояния $d-1$.

После перехода из состояния $d-1$ в состояние d агент выбирает один из двух критериев – целевой t или один из вспомогательных, равный OneMax при $x = dk$. Успех агента зависит от позиции точки переключения. Если точка переключения p лежит в интервале $[dk; (d+1)k-1]$, то число единиц в особи в случае, когда x находится в интервале $[p; (d+1)k-1]$, может увеличиться только при выборе агентом целевого критерия и мутации нулевого бита. Таким образом, худший случай достигается при $p = dk$, когда число переходов, в которых должен быть выбран целевой критерий, максимально.

Из вида марковской цепи следует, что в случае, когда агент «падает» из состояния d в состояние $d-1$, ему требуется много шагов для достижения состояния $d+1$. Далее приводятся результаты экспериментов, подтверждающие данное наблюдение. Вероятность «падения» из состояния d в состояние $d-1$ равна $\frac{dk}{3n}$ и растет с ростом значения целевого критерия. Таким образом, если точка переключения находится в состоянии обучения, равном $n-k+1$, время работы алгоритма максимально.

В таблице представлено число вычислений ФП, необходимое для достижения оптимума при решении задачи XdivK с помощью алгоритма RLS+Q-learning без вспомогательных критериев, с использованием одного вспомогательного критерия OneMax и с использованием двух переключающихся критериев. Первые две колонки содержат значения n и k . Следующие две колонки содержат вычисленное теоретически значение математического ожидания числа вычислений ФП, необходимое для решения задачи XdivK без вспомогательных критериев и с одним вспомогательным критерием OneMax соответственно. Данные результаты взяты из работы [13]. Следующая колонка содержит число вычислений ФП, необходимое для решения задачи XdivK с использованием двух переключающихся критериев, усредненное по 1000 независимых запусков. В качестве значений параметров скорости обучения α и дисконтного фактора

γ в алгоритме Q-learning были выбраны $\alpha = 0,5$ и $\gamma = 0,5$. Рассматривался худший случай, в котором точка переключения равна $n-k+1$.

Из представленных в таблице результатов можно видеть, что алгоритму RLS+Q-learning с использованием двух переключающихся критериев требуется гораздо больше вычислений ФП, чем при решении задачи XdivK с использованием одного вспомогательного критерия OneMax. Несмотря на то, что на каждом шаге оптимизации один из переключающихся критериев более эффективен, чем целевой, алгоритм RLS без вспомогательных критериев справляется с нахождением оптимального значения за меньшее число итераций. Из этого следует, что EA+RL не справляется с выбором эффективного вспомогательного критерия. В следующем разделе предложена модификация, решающая данную проблему.

n	k	XdivK	XdivK+OneMax	XdivK+два переключающихся критерия
40	2	$1,19 \cdot 10^3$	$6,81 \cdot 10^2$	$3,72 \cdot 10^3$
48	2	$1,70 \cdot 10^3$	$9,56 \cdot 10^2$	$5,23 \cdot 10^3$
56	2	$2,30 \cdot 10^3$	$1,28 \cdot 10^3$	$6,81 \cdot 10^3$
64	2	$2,98 \cdot 10^3$	$1,64 \cdot 10^3$	$8,91 \cdot 10^3$
72	2	$3,76 \cdot 10^3$	$2,05 \cdot 10^3$	$1,11 \cdot 10^4$
80	2	$4,62 \cdot 10^3$	$2,51 \cdot 10^3$	$1,41 \cdot 10^4$
60	3	$3,94 \cdot 10^4$	$1,08 \cdot 10^4$	$2,95 \cdot 10^5$
72	3	$6,79 \cdot 10^4$	$1,83 \cdot 10^4$	$4,98 \cdot 10^5$
84	3	$1,08 \cdot 10^5$	$2,86 \cdot 10^4$	$7,81 \cdot 10^5$
96	3	$1,60 \cdot 10^5$	$4,23 \cdot 10^4$	$1,05 \cdot 10^6$
108	3	$2,28 \cdot 10^5$	$5,97 \cdot 10^4$	$1,63 \cdot 10^6$
120	3	$3,12 \cdot 10^5$	$8,14 \cdot 10^4$	$2,37 \cdot 10^6$
80	4	$1,72 \cdot 10^6$	$2,30 \cdot 10^5$	$2,64 \cdot 10^7$
96	4	$3,57 \cdot 10^6$	$4,72 \cdot 10^5$	$5,67 \cdot 10^7$
112	4	$6,61 \cdot 10^6$	$8,68 \cdot 10^5$	$1,01 \cdot 10^8$
128	4	$1,13 \cdot 10^7$	$1,47 \cdot 10^6$	$1,72 \cdot 10^8$
144	4	$1,81 \cdot 10^7$	$2,35 \cdot 10^6$	$2,83 \cdot 10^8$
160	4	$2,76 \cdot 10^7$	$3,57 \cdot 10^6$	$3,96 \cdot 10^8$

Таблица. Число вычислений ФП, необходимое для нахождения оптимума XdivK

Модификация метода EA+RL с сохранением лучшей особи

В методе EA+RL в случае, когда новая особь лучше текущей по выбранному критерию, в качестве особи для следующего поколения выбирается новая особь. Однако если выбран мешающий критерий, то новая особь может быть хуже текущей по целевому критерию. В таком случае ЭА теряет особь с лучшим значением целевого критерия, а для того чтобы снова найти хорошее решение, требуется много шагов ЭА.

В предложенной модификации метода EA+RL в случае, когда новая особь лучше текущей по выбранному критерию, но хуже по целевому критерию, в следующее поколение переходит текущая особь. Как и при исследовании метода EA+RL, RLS используется в качестве ЭА, а алгоритмом обучения с подкреплением является Q-learning.

Оценка времени работы предложенной модификации RLS+Q-learning на задаче XdivK с переключающимися вспомогательными критериями

Для анализа времени работы предложенной модификации метода EA+RL строилась марковская цепь, описывающая соответствующий процесс оптимизации. Данная цепь представлена на рис. 2. Обозначения состояний и переходов данной марковской цепи аналогичны обозначениям цепи, представленной на рис. 1. Далее приведено объяснение, почему переходы имеют соответствующие метки.

В случае, когда число единиц равно dk , где d – константа, агент находится в состоянии обучения, равном d , и марковское состояние равно dk . У агента нет опыта в состоянии d , поэтому критерии выбираются равновероятно. В случае, когда выбран целевой критерий или критерий, который на данном этапе оптимизации совпадает с OneMax, и оператор мутации инвертировал единичный бит, новая особь содержит $dk-1$ единичный бит. Таким образом, новая особь хуже текущей по выбранному критерию, и она не будет выбрана ЭА в качестве следующего поколения. Та же ситуация возникает, когда выбранный критерий равен ZeroMax и инвертирован нулевой бит. В случае выбора критерия, равного ZeroMax, и инверсии единичного бита, новая особь лучше текущей по выбранному критерию. Однако значение целевого критерия, вычисленное на новой особи, меньше, чем у текущей особи. Таким образом, в качестве особи для следующего поколения выбирается текущая особь. В случае выбора целевого критерия или критерия, равного OneMax, и инверсии нулевого бита новая особь переходит в следующее поколение.

Переходы в марковских состояниях dk и $dk+1$ отличаются в случае выбора целевого критерия или критерия, равного ZeroMax, и инверсии единичного бита. В этом случае значение выбранного критерия новой особи лучше или совпадает со значением выбранного критерия текущей особи. Поэтому новая особь переходит в следующее поколение. Однако новая особь содержит меньшее число единичных бит, чем текущая, поэтому алгоритм переходит в марковское состояние dk . Переходы в состояниях $dk+2, \dots, dk+k-1$ определяются аналогично переходам в состоянии $dk+1$.

Для анализа времени работы предложенной модификации также необходимо было построить марковскую цепь для RLS без вспомогательных критериев (рис. 3).

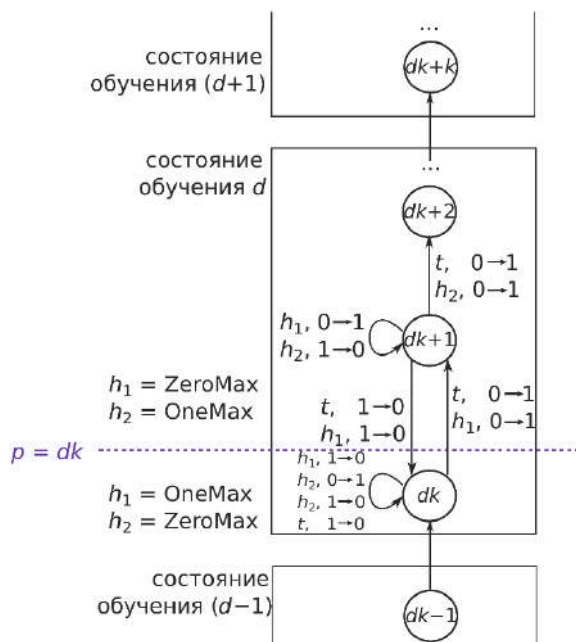


Рис. 2. Марковская цепь для предложенной модификации на задаче XdivK

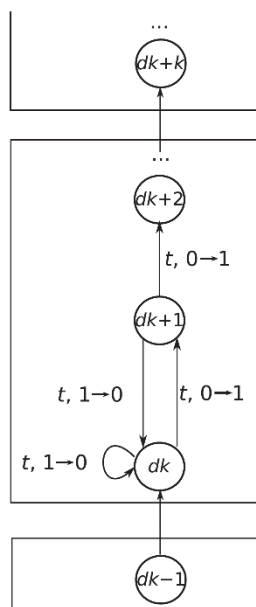


Рис. 3. Марковская цепь для RLS на задаче XdivK

Математическое ожидание времени работы предложенной модификации равно числу вычислений ФП, необходимому для достижения марковского состояния n из состояния 0. Каждый переход соответствует вычислению ФП, полученной в результате мутации особи. Поэтому математическое ожидание времени работы алгоритма, обозначаемое как $T(n)$, равно числу сделанных алгоритмом переходов в цепи:

$$T(n) = \sum_{i=0}^{n-1} E(i \rightarrow i+1), \quad (2)$$

где $E(i \rightarrow i + 1)$ – математическое ожидание числа переходов, необходимого для достижения состояния $i + 1$ из состояния i .

Значение $E(i \rightarrow i + 1)$ зависит от номера состояния i . Имеются два случая. В первом случае $i = dk$, где d – константа. Математическое ожидание числа переходов, необходимого для достижения состояния $dk + 1$ из состояния dk , вычисляется как

$$z_{dk} = E(dk \rightarrow dk + 1) = \frac{2}{3} \cdot \frac{n-dk}{n} \cdot 1 + \left(\frac{2}{3} \cdot \frac{dk}{n} + \frac{1}{3} \right) \cdot (1 + z_{dk}). \quad (3)$$

$$\text{Из (3) получается, что } z_{dk} = \frac{3n}{2(n-dk)}. \quad (4)$$

Во втором случае $i = dk + t$, где $1 \leq t \leq k - 1$. Математическое ожидание числа переходов, необходимого для достижения состояния $dk + t + 1$ из состояния $dk + t$, вычисляется как $z_{dk+t} = E(dk + t \rightarrow dk + t + 1)$:

$$z_{dk+t} = \frac{2(n-dk-t)}{3n} + \frac{2(dk+t)}{3n} \cdot (1 + z_{dk+t-1} + z_{dk+t}) + \left(\frac{dk+t}{3n} + \frac{n-dk-t}{3n} \right) \cdot (1 + z_{dk+t}). \quad (5)$$

$$\text{Из (5) получается, что } z_{dk+t} = z_{dk+t-1} \cdot \frac{dk+t}{n-dk-t} + \frac{3n}{2(n-dk-t)}. \quad (6)$$

Для оценки времени работы предложенной модификации необходимо вычислить математическое ожидание времени работы RLS без вспомогательных критериев. Методика вычисления аналогична использованной выше для предложенной модификации. Общее время выполнения также вычисляется по формуле (2). Имеются два случая: $i = dk$ и $i = dk + t$. Математическое ожидание числа переходов, необходимого для достижения состояния $dk + 1$ из состояния dk , вычисляется как

$$a_{dk} = E(dk \rightarrow dk + 1) = \frac{n-dk}{n} \cdot 1 + \frac{dk}{n} \cdot (1 + a_{dk}). \quad (7)$$

$$\text{Из (7) получается, что } a_{dk} = \frac{n}{n-dk}. \quad (8)$$

Математическое ожидание числа переходов, необходимого для перехода из состояния $dk + t$ в состояние $dk + t + 1$, равно $a_{dk+t} = E(dk + t \rightarrow dk + t + 1)$ и вычисляется следующим образом:

$$a_{dk+t} = \frac{n-dk-t}{n} \cdot 1 + \frac{dk+t}{n} \cdot (1 + a_{dk+t-1} + a_{dk+t}). \quad (9)$$

$$\text{Из (9) получается, что } a_{dk+t} = a_{dk+t-1} \cdot \frac{dk+t}{n-dk-t} + \frac{n}{n-dk-t}. \quad (10)$$

Из (4) и (8) получается, что $z_{dk} = 1,5a_{dk}$. Из выражений (2), (6), (10) при помощи математической индукции получается, что время работы предложенной модификации метода EA+RL на задаче XdivK с переключающимися критериями в 1,5 раза больше времени работы RLS. Таким образом, верхняя и нижняя асимптотические оценки времени работы предложенной модификации на задаче XdivK совпадают с оценками времени работы RLS, которые равны $O(n^{k+1})$ и $\Omega(n^k)$ [13]. Данный результат показывает, что, в отличие от метода EA+RL, предложенная модификация, несмотря на наличие мешающих критериев, имеет ту же асимптотику времени работы, что и RLS.

Заключение

Проведен анализ метода EA+RL выбора двух переключающихся вспомогательных критериев для решения задачи XdivK. В качестве эволюционного алгоритма использовался метод спуска со случайными мутациями RLS. В качестве алгоритма обучения с подкреплением использовался алгоритм Q-learning. На каждом шаге эволюционного алгоритма один из вспомогательных критериев более эффективен, чем другой является мешающим.

Показано, что в случае выбора вспомогательного критерия, являющегося мешающим на данном этапе оптимизации, лучшее найденное решение может быть утеряно, и алгоритму требуется много шагов, чтобы снова найти хорошее решение. Экспериментально подтверждено, что время работы алгоритма RLS+Q-learning на задаче XdivK во много раз превосходит время работы RLS без использования вспомогательных критериев.

Предложена модификация метода EA+RL, сохраняющая лучшую особь. Проведен теоретический анализ данной модификации. Доказано, что асимптотические оценки времени работы предложенной модификации и RLS совпадают. Таким образом, предложенная модификация справляется с игнорированием мешающего критерия, в отличие от метода EA+RL.

Литература

1. Скобцов Ю.А., Федоров Е.Е. *Метаэвристики*. Донецк: Ноулидж, 2013. 426 с.
2. Neumann F., Wegener I. Can single objective optimization profit from multiobjective optimization? / In: *Multiobjective Problem Solving from Nature: From Concepts to Applications*. 2008. P. 115–130. doi: 10.1007/978-3-540-72964-8_6
3. Neumann F., Wegener I. Minimum spanning trees made easier via multi-objective optimization // *Natural Computing*. 2006. V. 5. N 3. P. 305–319. doi: 10.1007/s11047-006-9004-x

References

1. Skobtsov Yu.A., Fedorov E.E. *Metaheuristics*. Donetsk, Noulidzh Publ., 2013, 426 p. (In Russian)
2. Neumann F., Wegener I. Can single objective optimization profit from multiobjective optimization? In: *Multiobjective Problem Solving from Nature: From Concepts to Applications*, 2008, pp. 115–130. doi: 10.1007/978-3-540-72964-8_6
3. Neumann F., Wegener I. Minimum spanning trees made easier via multi-objective optimization. *Natural Computing*, 2006, vol. 5, no. 3, pp. 305–319. doi: 10.1007/s11047-006-9004-x

4. Knowles J.D., Watson R.A., Corne D. Reducing local optima in single-objective problems by multi-objectivization // *Lecture Notes in Computer Science*. 2001. V. 1993. P. 269–283.
5. Segura C., Coello C.A.C., Miranda G., Leon C. Using multi-objective evolutionary algorithms for single-objective optimization // *4OR*. 2013. V. 11. N 3. P. 201–228. doi: 10.1007/s10288-013-0248-x
6. Jensen M.T. Helper-objectives: using multi-objective evolutionary algorithms for single-objective optimisation: evolutionary computation combinatorial optimization // *Journal of Mathematical Modelling and Algorithms*. 2004. V. 3. N 4. P. 323–347.
7. Lochtefeld D.F., Ciarallo F.W. Helper objective optimization strategies for the job-shop scheduling problem // *Applied Soft Computing*. 2011. V. 11. N 6. P. 4161–4174. doi: 10.1016/j.asoc.2011.03.007
8. Buzdalova A., Buzdalov M. Increasing efficiency of evolutionary algorithms by choosing between auxiliary fitness functions with reinforcement learning // *Proc. 11th Int. Conf. on Machine Learning and Applications*. Boca Raton, USA, 2012. V. 1. P. 150–155. doi: 10.1109/ICMLA.2012.32
9. Бuzдалова А.С., Бuzдалов М.В. Метод повышения эффективности эволюционных алгоритмов с помощью обучения с подкреплением // *Научно-технический вестник информационных технологий, механики и оптики*. 2012. № 5 (81). С. 115–119.
10. Sutton R.S., Barto A.G. *Reinforcement Learning: An Introduction*. Cambridge: MIT Press, 1998. 322 p.
11. Gosavi A. Reinforcement learning: a tutorial survey and recent advances // *INFORMS Journal on Computing*. 2009. V. 21. N 2. P. 178–192. doi: 10.1287/ijoc.1080.0305
12. Buzdalov M., Buzdalova A. Adaptive selection of helper-objectives for test case generation // *Proc. IEEE Congress on Evolutionary Computation*. Cancun, Mexico, 2013. V. 1. P. 2245–2250. doi: 10.1109/CEC.2013.6557836
13. Buzdalov M., Buzdalova A. OneMax helps optimizing XdivK: theoretical runtime analysis for RLS and EA+RL // *Proc. 16th Genetic and Evolutionary Computation Conference*. Vancouver, Canada, 2014. P. 201–202.
14. Buzdalov M., Buzdalova A. Can OneMax Help Optimizing LeadingOnes using the EA+RL Method? // *Proc. IEEE Congress on Evolutionary Computation*. Sendai, Japan, 2015. P. 1762–1768. doi: 10.1109/CEC.2015.7257100
15. Buzdalov M., Buzdalova A., Shalyto A. A first step towards the runtime analysis of evolutionary algorithm adjusted with reinforcement learning // *Proc. 12th Int. Conf. on Machine Learning and Applications*. Miami, USA, 2013. V. 1. P. 203–208. doi: 10.1109/ICMLA.2013.42
4. Knowles J.D., Watson R.A., Corne D. Reducing local optima in single-objective problems by multi-objectivization. *Lecture Notes in Computer Science*, 2001, vol. 1993, pp. 269–283.
5. Segura C., Coello C.A.C., Miranda G., Leon C. Using multi-objective evolutionary algorithms for single-objective optimization. *4OR*, 2013, vol. 11, no. 3, pp. 201–228. doi: 10.1007/s10288-013-0248-x
6. Jensen M.T. Helper-objectives: using multi-objective evolutionary algorithms for single-objective optimisation: evolutionary computation combinatorial optimization. *Journal of Mathematical Modelling and Algorithms*, 2004, vol. 3, no. 4, pp. 323–347.
7. Lochtefeld D.F., Ciarallo F.W. Helper objective optimization strategies for the job-shop scheduling problem. *Applied Soft Computing*, 2011, vol. 11, no. 6, pp. 4161–4174. doi: 10.1016/j.asoc.2011.03.007
8. Buzdalova A., Buzdalov M. Increasing efficiency of evolutionary algorithms by choosing between auxiliary fitness functions with reinforcement learning. *Proc. 11th Int. Conf. on Machine Learning and Applications*. Boca Raton, USA, 2012, vol. 1, pp. 150–155. doi: 10.1109/ICMLA.2012.32
9. Buzdalova A.S., Buzdalov M.V. Efficiency increasing method of the evolutionary algorithms by reinforcement learning. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2012, no. 5, pp. 115–119. (In Russian)
10. Sutton R.S., Barto A.G. *Reinforcement Learning: An Introduction*. Cambridge, MIT Press, 1998, 322 p.
11. Gosavi A. Reinforcement learning: a tutorial survey and recent advances. *INFORMS Journal on Computing*, 2009, vol. 21, no. 2, pp. 178–192. doi: 10.1287/ijoc.1080.0305
12. Buzdalov M., Buzdalova A. Adaptive selection of helper-objectives for test case generation. *Proc. IEEE Congress on Evolutionary Computation*. Cancun, Mexico, 2013, vol. 1, pp. 2245–2250. doi: 10.1109/CEC.2013.6557836
13. Buzdalov M., Buzdalova A. OneMax helps optimizing XdivK: theoretical runtime analysis for RLS and EA+RL. *Proc. 16th Genetic and Evolutionary Computation Conference*. Vancouver, Canada, 2014, pp. 201–202.
14. Buzdalov M., Buzdalova A. Can OneMax Help Optimizing LeadingOnes using the EA+RL Method? *Proc. IEEE Congress on Evolutionary Computation*. Sendai, Japan, 2015, pp. 1762–1768. doi: 10.1109/CEC.2015.7257100
15. Buzdalov M., Buzdalova A., Shalyto A. A first step towards the runtime analysis of evolutionary algorithm adjusted with reinforcement learning. *Proc. 12th Int. Conf. on Machine Learning and Applications*. Miami, USA, 2013, vol. 1, pp. 203–208. doi: 10.1109/ICMLA.2013.42

Авторы

Петрова Ирина Анатольевна – аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, irenepetrova@yandex.com

Бuzдалова Арина Сергеевна – аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, abuzdalova@gmail.com

Шалыто Анатолий Абрамович – доктор технических наук, профессор, заведующий кафедрой, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, shalyto@mail.ifmo.ru

Authors

Irina A. Petrova – postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, irenepetrova@yandex.com

Arina S. Buzdalova – postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, abuzdalova@gmail.com

Anatoly A. Shalyto – D.Sc., Professor, Head of Chair, ITMO University, Saint Petersburg, 197101, Russian Federation, shalyto@mail.ifmo.ru