



УДК 004.023:004.855.5:004.832.23

## МЕТОД ДИНАМИЧЕСКОГО ВЫБОРА ВСПОМОГАТЕЛЬНЫХ КРИТЕРИЕВ В МНОГОКРИТЕРИАЛЬНЫХ ЭВОЛЮЦИОННЫХ АЛГОРИТМАХ

И.А. Петрова<sup>а</sup>, А.С. Буздalова<sup>а</sup>, А.А. Шалыто<sup>а</sup><sup>а</sup> Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

Адрес для переписки: irenepetrova@yandex.com

### Информация о статье

Поступила в редакцию 02.03.16, принята к печати 13.04.16

doi: 10.17586/2226-1494-2016-16-3-460-466

Язык статьи – русский

**Ссылка для цитирования:** Петрова И.А., Буздalова А.С., Шалыто А.А. Метод динамического выбора вспомогательных критериев в многокритериальных эволюционных алгоритмах // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 3. С. 460–466. doi: 10.17586/2226-1494-2016-16-3-460-466

### Аннотация

**Предмет исследования.** Предложена модификация метода EA+RL, являющегося одним из методов повышения эффективности эволюционных алгоритмов при помощи вспомогательных критериев. Проведено сравнение предложенной модификации с существующими методами повышения эффективности эволюционных алгоритмов на примере задачи коммивояжера. **Метод.** В методе EA+RL обучение с подкреплением используется для выбора оптимизируемого критерия, целевого или одного из вспомогательных, на каждой итерации однокритериального эволюционного алгоритма. Предложенная модификация метода EA+RL позволяет использовать данный подход в многокритериальных эволюционных алгоритмах. В отличие от метода EA+RL, в предложенной модификации на каждом шаге многокритериального эволюционного алгоритма оптимизируются целевой критерий и один из вспомогательных, выбираемый при помощи обучения с подкреплением. **Основные результаты.** Проведено сравнение предложенной модификации метода EA+RL с существующими методами повышения эффективности эволюционных алгоритмов с помощью вспомогательных критериев на примере задачи коммивояжера. В методах EA+RL и предлагаемой его модификации применялись алгоритмы обучения с подкреплением в стационарной и нестационарной средах. Показаны преимущества решения задачи с использованием предлагаемой модификации метода EA+RL, применяемой совместно с алгоритмом обучения с подкреплением в нестационарной среде, по сравнению с использованием ранее известных методов выбора вспомогательных критериев в эволюционных алгоритмах. **Практическая значимость.** Предложенный в работе подход позволяет повысить эффективность эволюционных алгоритмов, которые применяются для решения NP-трудных задач дискретной оптимизации. К таким задачам относятся, в частности, поиск оптимального маршрута и составление расписаний.

### Ключевые слова

эволюционный алгоритм, обучение с подкреплением, многокритериальная оптимизация, задача коммивояжера

## ADAPTIVE SELECTION OF AUXILIARY OBJECTIVES IN MULTIOBJECTIVE EVOLUTIONARY ALGORITHMS

I.A. Petrova<sup>а</sup>, A.S. Buzdalova<sup>а</sup>, A.A. Shalyto<sup>а</sup><sup>а</sup> ITMO University, Saint Petersburg, 197101, Russian Federation

Corresponding author: irenepetrova@yandex.com

### Article info

Received 02.03.16, accepted 13.04.16

doi: 10.17586/2226-1494-2016-16-3-460-466

Article in Russian

**For citation:** Petrova I.A., Buzdalova A.S., Shalyto A.A. Adaptive selection of auxiliary objectives in multiobjective evolutionary algorithms. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 3, pp. 460–466. doi: 10.17586/2226-1494-2016-16-3-460-466

### Abstract

**Subject of Research.** We propose to modify the EA+RL method, which increases efficiency of evolutionary algorithms by means of auxiliary objectives. The proposed modification is compared to the existing objective selection methods on the example of travelling salesman problem. **Method.** In the EA+RL method a reinforcement learning algorithm is used to select an objective – the target objective or one of the auxiliary objectives – at each iteration of the single-objective evolutionary algorithm. The proposed modification of the EA+RL method adopts this approach for the usage with a multiobjective evolutionary algorithm. As opposed to the EA+RL method, in this modification one of the auxiliary objectives is selected by reinforcement learning and optimized together with the target objective at each step of the multiobjective evolutionary

algorithm. **Main Results.** The proposed modification of the EA+RL method was compared to the existing objective selection methods on the example of travelling salesman problem. In the EA+RL method and its proposed modification reinforcement learning algorithms for stationary and non-stationary environment were used. The proposed modification of the EA+RL method applied with reinforcement learning for non-stationary environment outperformed the considered objective selection algorithms on the most problem instances. **Practical Significance.** The proposed approach increases efficiency of evolutionary algorithms, which may be used for solving discrete NP-hard optimization problems. They are, in particular, combinatorial path search problems and scheduling problems.

#### Keywords

evolutionary algorithm, reinforcement learning, multiobjective optimization, travelling salesman problem

### Введение

Существуют практически значимые задачи оптимизации, в которых информация о градиенте оптимизируемой функции недоступна. К ним относятся, например, задачи комбинаторной оптимизации, такие как построение расписаний, а также задачи поисковой инженерии программного обеспечения, где требуется оптимизировать программы или входные данные для них. Точные алгоритмы решения многих из этих задач являются слишком неэффективными для их применения в практических приложениях, а для некоторых классов задач таких алгоритмов доказуемо не существует. Для получения решений приемлемого качества можно использовать метаэвристические алгоритмы оптимизации, такие как эволюционные алгоритмы.

Критерий оптимизации в эволюционном алгоритме (ЭА) представляется в виде функции приспособленности (ФП). Известно, что в некоторых случаях можно повысить эффективность ЭА, оптимизирующего некоторую целевую ФП, если дополнительно использовать вспомогательные ФП [1]. В существующих методах использования вспомогательных ФП предлагается либо оптимизировать одновременно все ФП, либо выбирать ФП на каждом этапе оптимизации случайным образом. Эти методы обладают рядом недостатков:

- первый подход работает слишком медленно на задачах с большим числом вспомогательных ФП или большим временем вычисления ФП;
- второй подход не учитывает возможных особенностей задачи, которые можно было бы использовать для более эффективного выбора ФП;
- в обоих подходах предполагается, что вспомогательные ФП специально разработаны так, чтобы повышать эффективность оптимизации целевой ФП. На практике добиться выполнения этого требования сложно, особенно в случае автоматической генерации вспомогательных ФП.

Ранее одним из авторов настоящей работы был предложен метод автоматического адаптивного выбора ФП в однокритериальном эволюционном алгоритме. В предложенном методе для выбора ФП впервые было применено обучение с подкреплением [2, 3]. Эффективность метода была подтверждена теоретически [4] и экспериментально путем применения его для решения ряда задач оптимизации [5, 6].

В данной работе излагаются результаты исследований, расширяющих область применения предложенного метода на многокритериальные эволюционные алгоритмы. Использование многокритериальных эволюционных алгоритмов для решения задач однокритериальной оптимизации позволяет производить сравнение точек поиска, равнозначных при однокритериальной постановке задачи, и тем самым предотвращать «застывание» в локальном оптимуме [7].

В следующем разделе описывается разработанный ранее метод выбора вспомогательных ФП в однокритериальных эволюционных алгоритмах. Далее предлагается модификация данного метода с целью применения к многокритериальным эволюционным алгоритмам и приводятся результаты экспериментов, позволяющие оценить эффективность предложенной модификации.

### EA+RL: метод выбора вспомогательных ФП в однокритериальных эволюционных алгоритмах

В методе EA+RL (Evolutionary Algorithm + Reinforcement Learning) [2, 3] агент обучения с подкреплением взаимодействует со средой, в качестве которой выступает ЭА. Схема метода представлена на рис. 1, где  $t$  – номер текущей итерации. Агент выбирает ФП из набора, включающего как вспомогательные ФП, так и целевую ФП. Далее выбранная ФП передается ЭА. Следующее поколение ЭА генерируется с использованием этой ФП. Агенту возвращаются численное вознаграждение и некоторое представление состояния среды. На основе полученной информации агент обновляет стратегию выбора ФП, и описанный процесс повторяется.

Целью агента обучения с подкреплением является максимизация суммарного вознаграждения. В методе EA+RL вознаграждение после каждой итерации основано на изменении значения целевой ФП. Таким образом, максимизация суммарного вознаграждения ведет к максимизации роста целевой ФП за весь период оптимизации.

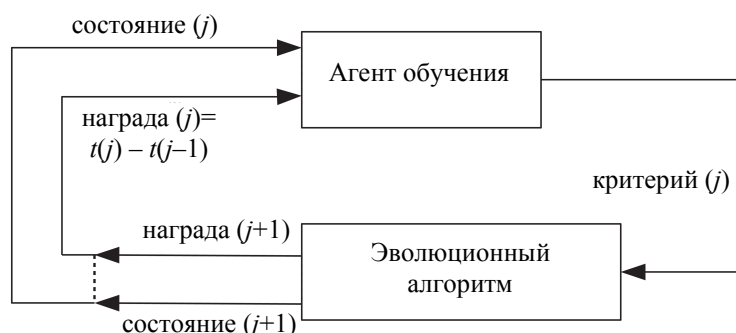


Рис. 1. Схема метода EA+RL

**МОЕА+RL: метод выбора вспомогательных критериев оптимизации в многокритериальных эволюционных алгоритмах**

В отличие от метода EA+RL, в предлагаемом методе МОЕА+RL (Multi-Objective EA + Reinforcement Learning) в качестве среды выступает многокритериальный ЭА, а действием является выбор вспомогательной ФП, оптимизируемой на текущей итерации алгоритма одновременно с целевой ФП.

Отметим, что, в отличие от метода EA+RL, в методе МОЕА+RL целевая ФП оптимизируется явно. Неявная оптимизация целевой ФП позволяет использовать однокритериальный ЭА, который обычно требует меньше вычислительных затрат, чем многокритериальный. Однако результаты данной работы показывают, что иногда неявной оптимизации целевой ФП недостаточно, и метод EA+RL в таком случае является неэффективным.

При исследовании эффективности метода МОЕА+RL в настоящей работе рассматривались алгоритмы обучения с подкреплением в стационарной и нестационарной средах. Среда является стационарной, если эффективность применения выбранного действия зависит только от состояния среды. Однако в случае, когда свойства вспомогательных ФП меняются в процессе оптимизации, эффективность применения одного и того же действия в одном и том же состоянии может быть разной.

Ранее было проведено исследование применимости существующих алгоритмов обучения с подкреплением в нестационарной среде для выбора вспомогательных ФП, меняющих свойства в процессе оптимизации. Однако полученные результаты были хуже, чем при использовании алгоритмов обучения с подкреплением в стационарной среде. В связи с этим недавно авторами был предложен прототип алгоритма обучения с подкреплением в нестационарной среде, специально разработанный для работы совместно с методом EA+RL и МОЕА+RL [8].

В настоящей работе применяется улучшенная версия предложенного ранее прототипа. Так же как в алгоритме классического Q-обучения, на каждой итерации агент применяет действие  $a$  к среде, находящейся в состоянии  $s$ . Затем значение ожидаемой награды  $Q(s, a)$  обновляется в соответствии с полученной наградой. Основной идеей алгоритма является перезапуск алгоритма обучения при выполнении двух условий, отражающих изменение свойств вспомогательных ФП и остановку ЭА в локальном оптимуме. Первое условие выполнено, если награда меньше или равна нулю в течение нескольких последовательных итераций и в заданном числе ( $k_1$ ) этих итераций награда строго меньше нуля. Второе условие выполнено, если награда меньше или равна нулю в течение нескольких последовательных итераций и в заданном числе ( $k_2$ ) этих итераций награда равна нулю.

**Решение задачи коммивояжера при помощи ЭА с использованием вспомогательных ФП**

В задаче коммивояжера рассматривается множество из  $n$  городов и матрица  $\mathbf{M}$  расстояний между ними, размером  $n \times n$ . Значение  $\mathbf{M}(c_1, c_2)$  соответствует длине пути из города  $c_1$  в город  $c_2$ . Целью задачи коммивояжера является нахождение пути  $\pi$  минимальной длины  $D(\pi)$ , проходящего через каждый город ровно один раз с возвратом в исходный город. Задача коммивояжера является NP-полной [9].

Существует несколько методов решения задачи коммивояжера с помощью ЭА с использованием вспомогательных ФП. Особью ЭА является путь, а в качестве целевой ФП выступает длина пути. Рассмотрим существующие подходы к решению задачи коммивояжера при помощи вспомогательных ФП.

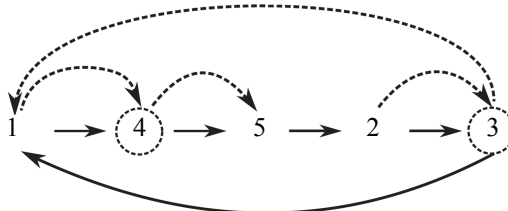
В одном из подходов, предложенном Knowles и др. [7], вспомогательные ФП получаются путем разбиения целевой ФП. Путь разбивается двумя городами  $a$  и  $b$  на два подпути. Таким образом, целевая ФП разбивается на две вспомогательные, которые соответствуют длинам подпутей. В данном методе на каждом шаге многокритериального ЭА оптимизируются две вспомогательные ФП вместо целевой.

В другом подходе к решению задачи коммивояжера при помощи вспомогательных ФП, предложенном Jensen [10], используются новые вспомогательные ФП, коррелирующие с целевой ФП. Значение вспомогательной ФП для особи, соответствующей пути  $\pi$ , вычисляется следующим образом:

$$h(\pi, p) = \sum_{i=1}^p \mathbf{M}(c_{\pi[\pi^{-1}[p[i]] \ominus 1]}, c_{\pi[i]}) + \mathbf{M}(c_{\pi[i]}, c_{\pi[\pi^{-1}[p[i]] \oplus 1]}) \tag{1}$$

где  $\mathbf{M}(c_1, c_2)$  соответствует длине пути из города  $c_1$  в город  $c_2$ ;  $p$  – некоторое подмножество городов,  $i \ominus 1 = \begin{cases} i-1, & i > 0 \\ n, & i = 0 \end{cases}$ ;  $i \oplus 1 = \begin{cases} 1, & i = n \\ i+1, & i < n \end{cases}$ . Подмножество городов  $p$  создается случайным образом, для каждого города вероятность попасть в  $p$  равна 50%.

Наиболее эффективно на каждом шаге многокритериального ЭА оптимизировать целевую ФП и одну из вспомогательных [10]. Каждая из вспомогательных ФП оптимизируется на протяжении равного числа итераций многокритериального ЭА. Порядок выбора вспомогательных ФП является случайным. На рис. 2 представлен пример вычисления вспомогательной ФП  $h(\pi, p)$  для пути  $\pi = (1, 4, 5, 2, 3)$ , где подмножество городов  $p = \{4, 3\}$ .



$$D(\pi) = \mathbf{M}(1, 4) + \mathbf{M}(4, 5) + \mathbf{M}(5, 2) + \mathbf{M}(2, 3) + \mathbf{M}(3, 1)$$

$$h(\pi, p) = [\mathbf{M}(1, 4) + \mathbf{M}(4, 5)] + [\mathbf{M}(2, 3) + \mathbf{M}(3, 1)]$$

Рис. 2. Пример вычисления вспомогательной ФП  $h(\pi, p)$  [11]

Однако при выборе вспомогательных ФП в случайном порядке может быть выбрана неэффективная вспомогательная ФП. Использование неэффективной вспомогательной ФП может привести к отдалению текущего найденного решения от оптимального значения. Исходя из этого, Jähne и др. [11] предложили модификацию данного метода, в которой не требуется выбирать оптимизируемую ФП. Создаются два подмножества городов: подмножество  $p$ , формируемое описанным выше способом, и подмножество  $p^c$ , являющееся дополняющим к подмножеству  $p$ . Затем создаются вспомогательные ФП,  $h_1(\pi, p)$  и  $h_2(\pi, p^c)$ , определяемые в соответствии с формулой (1). На каждом шаге многокритериального ЭА оптимизируются две вспомогательные ФП вместо целевой.

### Описание экспериментов

В настоящей работе результаты решения задачи коммивояжера при помощи методов EA+RL и MOEA+RL сравнивались с результатами, полученными при использовании трех описанных выше методов. Исследовалась применимость алгоритмов обучения с подкреплением в стационарной и нестационарной среде. Все алгоритмы запускались на фиксированном числе вычислений ФП. Подход Knowles и др. и метод EA+RL требуют гораздо больше вычислений ФП, чем метод MOEA+RL и подходы, предложенные Jensen и Jähne и др. В связи с этим метод EA+RL не сравнивался с подходами Jensen и Jähne и др. По той же причине метод MOEA+RL не сравнивался с подходом Knowles и др. Метод EA+RL сравнивался с традиционным ЭА, методом имитации отжига (ИО) [12] и подходом, предложенным Knowles и др. Метод MOEA+RL сравнивался с подходами Jensen и Jähne и др. Рассматривались два алгоритма обучения с подкреплением. Одним из них является стационарный алгоритм Q-обучения с  $\epsilon$ -жадной стратегией [13]. Другим является предложенный алгоритм обучения с подкреплением в нестационарной среде, описанный выше.

В ходе предварительных экспериментов было получено, что наилучшие результаты достигаются при использовании награды, используемой в работе [14]. Функция награды выглядит следующим образом:

$$r_t = \begin{cases} 1, & \text{если } g_{t+1} - g_t > 0 \\ 0, & \text{если } g_{t+1} - g_t = 0 \\ -1, & \text{если } g_{t+1} - g_t < 0 \end{cases}$$

где  $g_t$  и  $g_{t+1}$  – лучшее значение целевой ФП в поколениях  $t$  и  $t+1$  соответственно. Все рассматриваемые алгоритмы запускались по 30 раз на каждом экземпляре задачи коммивояжера, затем результаты усреднялись. Так же как и в работах Knowles и др., а также Jähne и др., вспомогательные ФП создавались один раз и использовались во всех запусках. Также в ходе предварительных экспериментов было получено, что наилучшие результаты достигаются при вероятности перезапуска, равной 0,5.

### Экспериментальное исследование метода EA+RL

В экспериментальных исследованиях метода EA+RL для решения задачи коммивояжера использовались те же параметры, что и в экспериментах Knowles и др., в том числе использовался только оператор мутации с двумя изменениями [10]. Поколение ЭА состояло из 100 особей. Число вычислений ФП взято из работы [7]. Для задач  $\text{gap20}$ ,  $\text{gap50}$ ,  $\text{euc50}$  число вычислений ФП равно  $5 \cdot 10^5$ , а для задач  $\text{euc100}$ ,


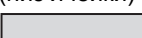
kroB100 –  $2 \cdot 10^6$ . Алгоритм  $\epsilon$ -жадного  $Q$ -обучения применялся со следующими значениями параметров:  $\alpha = 0,6$ ,  $\gamma = 0,01$ ,  $\epsilon = 0,3$ . Для предложенного алгоритма обучения с подкреплением в нестационарной среде в ходе предварительных экспериментов были выбраны следующие значения параметров:  $\alpha = 0,6$ ,  $\gamma = 0,1$ ,  $k_1 = 10$ ,  $k_2 = 500$ .

Результаты экспериментального сравнения метода EA+RL с другими методами представлены в табл. 1. Для каждого экземпляра задачи коммивояжера среднее значение представлено сверху ячейки, а лучшее полученное значение представлено внизу ячейки. В первой колонке содержится название задачи. Во второй колонке содержится наилучшее известное решение. Третья колонка (ЭА) содержит результаты применения традиционного ЭА. В следующих двух колонках содержатся результаты применения метода ИО и подхода Knowles и др. (Knowles), взятые из работы [7]. Следующие две колонки содержат результаты применения метода EA+RL с  $\epsilon$ -жадным  $Q$ -обучением (С / К) и предложенным алгоритмом обучения с подкреплением в нестационарной среде (НС / К). В обоих методах использовались вспомогательные ФП, предложенные Knowles и др. Следующие две колонки содержат результаты запусков метода EA+RL с  $\epsilon$ -жадным  $Q$ -обучением (С / J) и предложенным алгоритмом обучения с подкреплением в нестационарной среде (НС / J) с использованием двух вспомогательных ФП, предложенных Jähne и др. Первый и второй результаты выделены темно-серым и светло-серым цветами соответственно. Среднеквадратичное отклонение среднего значения ФП составило 0,8%.

Можно видеть, что использование метода EA+RL с использованием предложенного алгоритма обучения с подкреплением в нестационарной среде и вспомогательных ФП, предложенных Jähne и др., дает наилучшие результаты среди всех рассмотренных алгоритмов на большинстве задач.

Задача	Лучшее	ЭА	ИО	Knowles	С / К	НС / К	С / J	НС / J
ran20	1,91	2,03	2,55	2,66	1,93	1,92	1,96	1,92
		1,91	2,54	2,54	1,91	1,91	1,91	1,91
ran50	2,04	2,63	2,30	2,32	2,49	2,42	2,55	2,29
		2,34	2,13	2,18	2,19	2,14	2,29	2,06
euc50	5,03	5,62	5,72	5,78	5,51	5,49	5,51	5,48
		5,37	5,69	5,69	5,37	5,37	5,37	5,37
euc100	7,12	8,27	7,98	7,97	8,14	8,09	8,28	8,09
		7,96	7,85	7,79	7,91	7,95	8,01	7,90
kroB100	22141	23296	22529	22546	22952	22776	23161	22391
		22509	22217	22141	22432	22243	22611	22139

Таблица 1. Среднее (верх ячейки) и лучшее (низ ячейки) значение целевой ФП.

 – первый результат;  – второй результат

### Экспериментальное исследование метода MOEA+RL

При исследовании применимости метода MOEA+RL для решения задачи коммивояжера рассматривались экземпляры задачи коммивояжера, решаемые в работах Jensen и Jähne и др. Условия задач были взяты с сайта TSPLIB<sup>1</sup>. В рассматриваемых задачах было от 100 до 1002 городов. Число в названии задачи соответствует числу городов. Применяемые операторы мутации и кроссовера, а также эвристика 2-ort, используемая при решении задач коммивояжера, были такими же, как и в работах Jensen и Jähne и др. Поколение многокритериального ЭА состояло из 100 особей. Число вычислений ФП для каждой задачи вычислялось по формуле из работы [11]:  $E(n) = \sqrt{n^3} \times 15$ , где  $n$  – число городов. В алгоритме  $\epsilon$ -жадного  $Q$ -обучения использовались следующие параметры:  $\alpha = 0,6$ ,  $\gamma = 0,01$ ,  $\epsilon = 0,3$ . Для второй версии предложенного подхода были выбраны параметры  $\alpha = 0,6$ ,  $\gamma = 0,1$ ,  $k_1 = 10$ ,  $k_2 = 10$ . Различие в выбранных в ходе предварительных экспериментов значениях параметра  $k_2$  в методе EA+RL и MOEA+RL можно объяснить различным числом вычислений ФП, и, как следствие, различным числом итераций алгоритма.

Результаты применения метода MOEA+RL представлены в табл. 2. В первой колонке содержится название задачи. Во второй колонке содержится наилучшее известное решение. В следующих четырех колонках содержатся результаты применения методов MOEA+RL с предложенным алгоритмом обучения с подкреплением в нестационарной среде (НС MOEA+RL), MOEA+RL с  $\epsilon$ -жадным  $Q$ -обучением (С MOEA+RL), Jähne и др. (Jähne) и Jensen (Jensen–Jähne). Во всех этих методах использовались две вспомогательные ФП, предложенных Jähne и др. Последняя колонка содержит результаты метода Jensen, с использованием десяти вспомогательных ФП, предложенных Jensen, как в экспериментах работы [10].

<sup>1</sup> <http://comopt.ifl.uni-heidelberg.de/software/TSPLIB95/>

Формат представления результатов в табл. 2 аналогичен формату табл. 1. Среднеквадратичное отклонение среднего значения ФП составило 0,05%.

Можно видеть, что метод MOEA+RL с использованием предложенного алгоритма обучения с подкреплением в нестационарной среде превосходит остальные алгоритмы на большинстве рассмотренных задач. Данный метод статистически различим с остальными рассмотренными методами, что было показано при помощи многокритериального знакового теста с уровнем статистической значимости  $\alpha = 0,05$  [15].

Задача	Лучшее	HC MOEA+RL	C MOEA+RL	Jähne	Jensen–Jähne	Jensen
kroB100	22141	22144	22145	22150	22158	22155
		22139	22139	22139	22139	22139
kroD100	21294	21342	21353	21344	21349	21347
		21294	21294	21294	21294	21294
kroE100	21294	22093	22095	22169	22095	22100
		22068	22068	22068	22068	22068
eil101	629	641,39	641,84	641,50	641,59	641,95
		640	640	640	640	640
pr124	59030	59030	59030	59030	59032	59052
		59030	59030	59030	59030	59030
bier127	118282	118324	118394	118387	118408	118394
		118293	118293	118293	118293	118293
pr136	96772	96975	97000	96980	97193	97063
		96785	96835	96795	96785	96835
kroA150	26524	26540	26558	26533	26557	26558
		26524	26524	26524	26524	26554
kroB150	26130	26153	26166	26170	26166	26174
		26127	26127	26127	26127	26127
pr152	73682	73693	73702	73904	73820	73821
		73683	73683	73687	73683	73820
pr439	107217	107675	107677	107748	108035	107743
		107241	107248	107301	107258	107248
rat575	6773	6869	6872	6874	6863	6877
		6833	6824	6847	6835	6826
pr1002	259045	263158	263318	263425	263184	263189
		261444	261970	261231	262023	260971

Таблица 2. Среднее (верх ячейки) и лучшее (низ ячейки) значение целевой ФП.  
 – первый результат;  – второй результат

### Заключение

Было проведено сравнение методов EA+RL и MOEA+RL с другими методами выбора вспомогательных функций приспособленности на примере задачи коммивояжера. В качестве алгоритмов обучения с подкреплением рассматривались алгоритмы обучения в стационарной и нестационарной средах. Алгоритм обучения с подкреплением в нестационарной среде был разработан авторами специально для применения совместно с методами EA+RL и MOEA+RL. Использовались вспомогательные функции приспособленности, предложенные в работах Knowles и др., Jensen и Jähne и др.

В результате проведенных исследований можно сделать следующие выводы.

- Наиболее эффективным среди рассмотренных методов решения задачи коммивояжера с использованием вспомогательных функций приспособленности является метод MOEA+RL с использованием предложенного алгоритма обучения с подкреплением в нестационарной среде.
- Наилучшие результаты решения задачи коммивояжера были получены при использовании вспомогательных функций приспособленности, предложенных Jähne и др.

- Методу EA+RL требуется гораздо больше вычислений функций приспособленности, чем методу MOEA+RL. Отсюда следует, что иногда неявной оптимизации целевой функции приспособленности недостаточно.
- Несмотря на то, что современные исследования направлены на улучшение подхода Jähne и др., метод, в котором на каждом шаге оптимизируется целевая функция приспособленности и одна из вспомогательных, может быть более эффективным в случае использования подходящего метода выбора вспомогательной функции приспособленности (в данном случае это использование предложенного алгоритма обучения с подкреплением в нестационарной среде).

#### Литература

1. 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
2. Буздalова А.С., Буздalов 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.
3. Buzdalov A., Buzdalova A. Increasing efficiency of evolutionary algorithms by choosing between auxiliary fitness functions with reinforcement learning // Proc. 11<sup>th</sup> Int. Conf. on Machine Learning and Applications. Boca Raton, USA, 2012. V. 1. P. 150–155. doi: 10.1109/ICMLA.2012.32
4. Buzdalov M., Buzdalova A., Shalyto A. A first step towards the runtime analysis of evolutionary algorithm adjusted with reinforcement learning // Proc. 12<sup>th</sup> Int. Conf. on Machine Learning and Applications. Miami, USA, 2013. V. 1. P. 203–208. doi: 10.1109/ICMLA.2013.42
5. 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
6. Buzdalov M., Buzdalova A., Petrova I. Generation of tests for programming challenge tasks using multi-objective optimization // Proc. Genetic and Evolutionary Computation Conference Companion. Amsterdam, Netherlands, 2013. P. 1655–1658. doi: 10.1145/2464576.2482746
7. 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.
8. Petrova I., Buzdalova A., Buzdalov M. Improved selection of auxiliary objectives using reinforcement learning in non-stationary environment // Proc. 13<sup>th</sup> Int. Conf. on Machine Learning and Applications. Detroit, USA, 2014. P. 580–583. doi: 10.1109/ICMLA.2014.99
9. Applegate D.L., Bixby R.E., Chvatal V., Cook W.J. The Traveling Salesman Problem: A Computational Study. Princeton University Press, 2007. 608 p.
10. 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.
11. Jähne M., Li X., Branke J. Evolutionary algorithms and multi-objectivization for the travelling salesman problem // Proc. 11<sup>th</sup> Annual Genetic and Evolutionary Computation Conference. Montreal, Canada, 2009. P. 595–602. doi: 10.1145/1569901.1569984
12. Скобцов Ю.А., Федоров Е.Е. Метаэвристики. Донецк: Ноулидж, 2013. 426 с.
13. Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. Cambridge: MIT Press, 1998. 322 p.
14. Afanasyeva A., Buzdalov M. Optimization with auxiliary criteria using evolutionary algorithms and reinforcement learning // Proc. 18<sup>th</sup> Int. Conf. on Soft Computing MENDEL. Brno, Czech Republic, 2012. P. 58–63.
15. Derrac J., Garcia S., Molina D., Herrera F. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms // Swarm and Evolutionary Computation. 2011. V. 1. N 1. P. 3–18. doi: 10.1016/j.swevo.2011.02.002

<i>Петрова Ирина Анатольевна</i>	– аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, irenepetrova@yandex.com
<i>Буздalова Арина Сергеевна</i>	– аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, abuzdalova@gmail.com
<i>Шалыто Анатолий Абрамович</i>	– доктор технических наук, профессор, заведующий кафедрой, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, shalyto@mail.ifmo.ru
<i>Irina A. Petrova</i>	– postgraduate, ITMO University, 197101, Saint Petersburg, Russian Federation, irenepetrova@yandex.com
<i>Arina S. Buzdalova</i>	– postgraduate, ITMO University, 197101, Saint Petersburg, Russian Federation, abuzdalova@gmail.com
<i>Anatoly A. Shalyto</i>	– D.Sc., Professor, Head of Chair, ITMO University, 197101, Saint Petersburg, Russian Federation, shalyto@mail.ifmo.ru