

УДК 681.3.001.63

О.А. Мелихова

ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ ДЛЯ ПОСТРОЕНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА*

Рассматривается возможность использования методов направленного случайно-детерминированного поиска для решения дискретных задач оптимизации, точнее, для решения NP-полных комбинаторных задач. Эффективность этих методов проявляется при решении задач с большой размерностью. В работе используется понятие схема, облегчающее описание функционирования генетического алгоритма. Кроме того, проводится анализ основных факторов, влияющих на преобразования схем в генетическом алгоритме. Приводится основная теорема о генетических алгоритмах. Рассматриваются определения порядка схемы, охвата схемы, эффекты совместного воздействия селекции, скрещивания и мутации. Приводится теорема о схемах при использовании генетических алгоритмов для решения оптимизационных задач, в которых целевая функция может принимать отрицательные значения. Также приводится кодирование параметров задачи в генетическом алгоритме. Рассматривается использование эволюционных вычислений для решения оптимизационных задач.

Эволюционные вычисления; хромосомное представление; кодирование; длина схемы; вероятность выживания схемы.

О.А. Melikhova

GENETIC ALGORITHMS APPLICATION FOR THE ARTIFICIAL INTELLECT SYSTEMS CONSTRUCTION

The possibility of the directed stochastic-determined search methods use for the discrete optimization problems solution, more exactly, for the solution of the NP-complete combinatorial problems is considered. The effectiveness of these methods is shown by the solution of the problems with the big dimension. The concept scheme, facilitating the genetic algorithm functioning description is used in the paper. Besides, the analysis of the main factors, influencing on the scheme transformations in the genetic algorithm is developed. The main theorem about the genetic algorithms is reduced. The definitions of the scheme order, scheme envelopment, the effects of the combined actions of selection, crossing and mutation are considered. The scheme theorem under using of genetic algorithms for the decision of the optimization problems in which the goal function may accept the negative values is reduced. The parameters coding of the problem in the genetic algorithm is reduces also. The use of evolutionary computations for the optimization problems decision is considered.

Evolutionary computations; chromosomes representation; coding; scheme length; scheme survival probability.

Введение. Использование эволюционных вычислений для решения оптимизационных задач ориентировано на задачи комбинаторной оптимизации, успешное решение которых позволяет повысить эффективность принятия решений в интеллектуальных системах [1]. Решение NP-полных комбинаторных задач в настоящее время представлено множеством различных методов и алгоритмов. Точные методы не эффективны при большом числе исходных данных, а эвристические алгоритмы обладают большими погрешностями. Использование методов эволюционных вычислений для решения дискретных задач оптимизации методом направленного случайно-детерминированного поиска позволяет устранить названные недостатки. Ос-

* Работа выполнена при частичной поддержке РФФИ (проект № 11-01-00975).

новное преимущество эволюционных вычислений заключается в возможности решения многомодальных задач с большой размерностью за счет сочетания элементов случайности и детерминированности, как это происходит в природе. Детерминированность этих методов заключается в моделировании природных процессов отбора, размножения и наследования, происходящих по определенным правилам. Основным при этом является закон эволюции: «выживает сильнейший», обеспечивающий улучшение результата. Другим важным фактором эффективности эволюционных вычислений является моделирование размножения и наследования. Рассматриваемые варианты решений могут по определенному правилу порождать новые решения, наследующие лучшие черты своих «предков». В качестве случайного элемента можно использовать моделирование процесса мутации. Случайно измененные характеристики решения приводят к новому направлению в процессе эволюции и могут ускорить процесс выработки лучшего решения.

Кодирование параметров задачи в генетическом алгоритме. Выбор исходной популяции связан с представлением параметров задачи в форме хромосом, с так называемым хромосомным представлением, определяющим способ кодирования. В классическом генетическом алгоритме применяется двоичное кодирование, то есть аллели всех генов в хромосоме равны 0 и 1. Длина хромосом зависит от условий задачи, то есть от количества точек в пространстве поиска. Генетические алгоритмы находят применение в основном в задачах оптимизации [1, 2].

Допустим, что ищется максимум функции $f(x_1, x_2, \dots, x_n) > 0$ для $x_i \in [a_i, b_i] \subset R; i=1,2,\dots,n$ и требуется найти решение с точностью до q для каждой переменной x_i . В такой ситуации необходимо разбить интервал $[a_i, b_i]$ на $(b_i - a_i) \cdot 10^q$ одинаковых подинтервалов. Это означает применение дискретизации с шагом $r = 10^{-q}$. Наименьшее натуральное число m_i , удовлетворяющее неравенству

$$(b_i - a_i) \cdot 10^q \leq 2^{m_i} - 1, \quad (1)$$

определяет необходимую и достаточную длину двоичной последовательности, требуемой для кодирования числа из интервала $[a_i, b_i]$ с шагом r . Каждой такой двоичной последовательности соответствует десятичное значение числа, представляемого данным кодом (с учетом правил перевода десятичных чисел в двоичную форму). Пусть y_i обозначает десятичное значение двоичной последовательности, кодирующей число x_i . Значение x_i можно представить выражением

$$x_i = a_i + y_i \frac{b_i - a_i}{2^{m_i} - 1}. \quad (2)$$

Таким способом задаются фенотипы, соответствующие кодовым последовательностям с длиной m_i [1–3].

Основная теорема о генетических алгоритмах. Для того чтобы лучше понять функционирование генетического алгоритма, используем понятие *схема* и сформулируем основную теорему, относящуюся к генетическим алгоритмам и называемую теоремой о схемах [1, 3, 4]. Понятие *схема* было введено для определения множества хромосом, обладающих некоторыми общими свойствами, то есть подобных друг другу. Если аллели принимают значения 0 и 1 (рассматриваются хромосомы с двоичным алфавитом), то схема представляет собой множество хромосом, содержащих нули и единицы на некоторых заранее определенных позициях. При рассмотрении схем удобно использовать расширенный алфавит $\{0, 1, *\}$, в который помимо 0 и 1 введен дополнительный символ *, обозначающий любое допустимое значение, то есть 0 или 1. Например,

$$\begin{aligned} 10*1 &= \{1001, 1011\} \\ *01*10 &= \{001010, 001110, 101010, 101110\} \end{aligned}$$

Считается, что хромосома принадлежит данной схеме, если для каждой j -й позиции (локуса), $j = 1, 2, \dots, L$, где L – длина хромосомы; символ, занимающий j -ю позицию хромосомы, соответствует символу, занимающему j -ю позицию схемы, причем символу * соответствуют как 0, так и 1. То же означает утверждение: хромосома соответствует схеме и хромосома представляет схему. Отметим, что если в схеме присутствует m символов *, то эта схема содержит 2^m хромосом. Кроме того, каждая хромосома длиной L принадлежит к 2^L схемам. Цепочки длиной 2 соответствуют четырем различным схемам, а цепочки длиной 3 – восьми схемам.

Генетический алгоритм базируется на принципе трансформации наиболее приспособленных особей. Пусть $P(0)$ означает исходную популяцию особей, а $P(k)$ – текущую популяцию (на k -й итерации алгоритма). Из каждой популяции $P(k)$, $k = 0, 1, \dots$ методом селекции выбираются хромосомы с наибольшей приспособленностью, которые включаются в так называемый родительский пул $M(k)$. Далее, в результате объединения особей из популяции $M(k)$ в родительские пары и выполнения операции скрещивания с вероятностью p_c , а так же операции мутации с вероятностью p_m формируется новая популяция $P(k+1)$, в которую входят потомки особей из популяции $M(k)$.

Следовательно, для любой схемы, представляющей хорошее решение, желательно, чтобы количество хромосом, соответствующих этой схеме, возрастало с увеличением количества итераций k .

На соответствующее преобразование схем в генетическом алгоритме оказывают влияние три фактора: селекция хромосом, скрещивание и мутация. Проанализируем воздействие каждого из них с точки зрения увеличения ожидаемого количества представителей отдельно взятой схемы.

Обозначим схему символом S , а количество популяций $P(k)$, соответствующих этой схеме – $c(S, k)$. Следовательно, $c(S, k)$ можно считать количеством элементов множества $P(k) \cap S$.

Исследуем влияние селекции. При выполнении этой операции хромосомы из популяции $P(k)$ копируются в родительский пул $M(k)$ с вероятностью, определяемой выражением (3). Пусть, $F(S, k)$ – среднее значение функции приспособленности хромосом из популяции $P(k)$, которые соответствуют схеме S . Если

$$P(k) \cap S = \{ch_1, \dots, ch_{c(S, k)}\}, \quad (3)$$

то

$$F(S, k) = \frac{\sum_{i=1}^{c(S, k)} F(ch_i)}{c(S, k)}. \quad (4)$$

Величина $F(S, k)$ также называется приспособленностью схемы S на k -й итерации [1, 3, 5].

Пусть $\mathfrak{Z}(k)$ обозначает сумму значений функции приспособленности хромосом из популяции $P(k)$ мощностью N , то есть

$$\mathfrak{Z}(k) = \sum_{i=1}^N F(ch_i^{(k)}). \quad (5)$$

Обозначим через $\bar{F}(k)$ среднее значение функции приспособленности хромосом этой популяции, то есть

$$\bar{F}(k) = \frac{1}{N} \mathfrak{Z}(k). \quad (6)$$

Пусть $ch_r^{(k)}$ обозначает элемент родительского пула $M(k)$. Для каждого $ch_r^{(k)} \in M(k)$ и для каждого $i = 1, \dots, c(S, k)$ вероятность того, что $ch_r^{(k)} = ch_i$, определяется отношением $F(ch_i) / F(k)$. Поэтому ожидаемое количество хромосом в популяции $M(k)$, которые равны ch_i , составит

$$N \frac{F(ch_i)}{F(k)} = \frac{F(ch_i)}{F(k)}$$

Таким образом, ожидаемое количество хромосом из множества $P(k) \cap S$, отобранных для включения в родительский пул $M(k)$, будет равно

$$\sum_{i=1}^{c(S,k)} \frac{F(ch_i)}{F(k)} = c(S, k) \frac{F(S, k)}{F(k)},$$

что следует из выражения (4).

Поскольку каждая хромосома из родительского пула $M(k)$ одновременно принадлежит популяции $P(k)$, то хромосомы, составляющие множество $M(k) \cap S$ – это те же самые особи, которые были отобраны из множества $P(k) \cap S$ для включения в популяцию $M(k)$. Если количество хромосом родительского пула $M(k)$, соответствующих схеме S , обозначить $b(S, k)$, то из приведенных рассуждений можно сделать следующий вывод [1–3].

Ожидаемое значение $b(S, k)$, то есть ожидаемое значение количества хромосом родительского пула $M(k)$, соответствующих схеме S , определяется выражением

$$E[b(S, k)] = c(S, k) \frac{F(S, k)}{F(k)}. \quad (7)$$

Если схема S содержит хромосомы со значением функции приспособленности, превышающим среднее значение, то ожидаемое количество хромосом из родительского пула $M(k)$, соответствующих схеме S , будет больше количества хромосом из популяции $P(k)$, соответствующих схеме S . Поэтому можно утверждать, что селекция вызывает распространение схем с приспособленностью «лучше средней» и исчезновение схем с «худшей» приспособленностью.

Порядок (order) схемы S , иначе называемый счетностью схемы и обозначаемый $o(S)$ – это количество постоянных позиций в схеме, т.е. нулей и единиц в случае алфавита $\{0, 1, *\}$. Например,

$$o(10*1) = 3 \quad o(*01*10) = 4 \quad o(**0*1*) = 2 \quad o(*101**) = 3$$

Порядок схемы $o(S)$ равен длине L за вычетом количества символов $*$, что легко проверить на представленных примерах (для $L = 4$ с одним символом $*$ и для $L = 6$ с двумя, четырьмя и тремя символами $*$). Легко заметить, что порядок схемы, состоящей исключительно из символов $*$, равен нулю, то есть $o(****) = 0$, а порядок схемы без единого символа $*$ равен L ; например, $o(10011010) = 8$. Порядок схемы $o(S)$ – это всегда целое число из интервала $[0, L]$.

Охват (defining length) схемы S , называемый также длиной схемы и обозначаемый $d(S)$ – это расстояние между первым и последним постоянным символом (то есть разность между правой и левой крайними позициями, содержащими постоянные символы). Например,

$$\begin{aligned} d(10*1) &= 4 - 1 = 3 & d(*01*10) &= 6 - 2 = 4 \\ d(**0*1*) &= 5 - 3 = 2 & d(*101**) &= 4 - 2 = 2 \end{aligned}$$

Охват схемы $d(S)$ – это целое число из интервала $[0, L - 1]$. Отметим, что охват схемы с постоянными символами на первой и последней позиции равен $L - 1$ (как в первом из приведенных примеров). Охват схемы с единственной постоянной позицией равен нулю, в частности, $d(**1*) = 0$. Охват характеризует содержательность информации, заключенной в схеме [1, 4, 5].

Перейдем к влиянию операции скрещивания на обработку схем в генетическом алгоритме. Отметим, что одни схемы оказываются более подверженными уничтожению в процессе скрещивания, чем другие. Например, рассмотрим схемы $S_1 = 1****0*$ и $S_2 = **01***$, а также хромосому $ch = [1001101]$, соответствующую обеим схемам. Видно, что схема S_2 имеет больше шансов «пережить» операцию скрещивания, чем схема S_1 , которая больше подвержена «расщеплению» в точках скрещивания 1,2,3,4 и 5. Схему S_2 можно разделить только при выборе точки скрещивания, равной 3. Обратим внимание на охват обеих схем, который – это очевидно – оказывается существенным в процессе скрещивания.

В ходе анализа влияния операции скрещивания на родительский пул $M(k)$ рассмотрим хромосому из множества $M(k) \cap S$, то есть хромосому из родительского пула, соответствующую схеме S . Вероятность того, что эта хромосома будет отобрана для скрещивания, равна p_c . Если ни один из потомков этой хромосомы не будет принадлежать к схеме S , значит точка скрещивания должна находиться между первым и последним постоянным символом данной схемы. Вероятность этого равна $d(S)/(L-1)$. Из этого можно сделать следующие выводы: для некоторой хромосомы из $M(k) \cap S$ вероятность того, что она не будет отобрана для скрещивания либо, что хотя бы один из ее потомков после скрещивания будет принадлежать к схеме S , ограничена снизу величиной:

$$1 - p_c \frac{d(S)}{L-1}$$

Эта величина называется вероятностью выживания схемы S . Легко показать, что если данная хромосома принадлежит к схеме S и отбирается для скрещивания, а вторая родительская хромосома также принадлежит схеме S , то оба их потомка тоже будут принадлежать схеме S .

Рассмотрим влияние мутации на родительский пул $M(k)$. Оператор мутации с вероятностью p_m случайным образом изменяет значение в конкретной позиции с 0 на 1 и обратно. Очевидно, что схема переживет мутацию только в том случае, когда все ее постоянные позиции останутся после выполнения этой операции неизменными.

Хромосома из родительского пула, принадлежащая к схеме S (то есть хромосома из множества $M(k) \cap S$), останется в этой схеме тогда и только тогда, когда ни один символ этой хромосомы, соответствующий постоянным символам схемы S , не изменится в процессе мутации. Вероятность такого события равна $(1 - p_m)^{o(S)}$. Этот результат можно представить в форме следующего вывода: вероятность того, что некоторая хромосома из $M(k) \cap S$ будет принадлежать схеме S после операции мутации, определяется выражением

$$(1 - p_m)^{o(S)}.$$

Эта величина называется вероятностью выживания схемы S после мутации.

Эффект совместного воздействия селекции, скрещивания и мутации с учетом факта, что если хромосома из множества $M(k) \cap S$ дает потомка, соответствующего схеме S , то он будет принадлежать к $P(k+1) \cap S$, ведет к построению следующей схемы репродукции (8):

$$E[c(S, k+1)] \geq c(S, k) \frac{F(S, k)}{F(k)} \left(1 - p_c \frac{d(S)}{L-1} \right) (1 - p_m)^{o(S)}. \quad (8)$$

Зависимость (8) показывает, как изменяется от популяции к популяции количество хромом, соответствующих данной схеме [1, 4]. Чем больше значение каждого из этих факторов, тем большим оказывается ожидаемое количество соответствий схеме S в следующей популяции.

Конечный результат из выражения (8), сформулируем в форме теоремы. Это основная теорема генетических алгоритмов, иначе называемая теоремой о схемах [1, 2, 5]. *Теорема:* схемы малого порядка, с малым охватом и с приспособленностью

выше средней формируют показательно возрастающее количество своих представителей в последующих поколениях генетического алгоритма. В соответствии с теоремой важным становится кодирование, которое должно обеспечивать построение схем малого порядка, с малым охватом и с приспособленностью выше средней.

Заключение. В настоящее время эволюционные вычисления представляют одну из развивающихся областей науки, исследования в которой ведут к постоянному повышению эффективности их использования и к появлению новых подходов к построению конкретных алгоритмов. Они, все чаще, применяются при интерактивном управлении технологическими процессами на металлургических и химических предприятиях. Развитие теории решения комбинаторных задач оптимизации происходит непосредственно под влиянием запросов практики, в частности, необходимости производить оптимальный выбор решений на множестве вариантов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Рутковская Д., Пилинский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И.Д. Рудинского. – М.: Горячая линия – Телеком, 2006. – 452 с.
2. Мелихова О.А., Мелихова З.А. Использование генетических алгоритмов при управлении сложными технологическими процессами // Труды Конгресса по ИС и ИТ “IS&IT’11”. – М.: Физматлит, 2011. – Т. 2. – С. 36-41.
3. Мелихова О.А., Мелихова З.А. Использование генетических алгоритмов при построении систем искусственного интеллекта // Труды Конгресса по ИС и ИТ “AIS-IT’10”. – М.: Физматлит, 2010. – Т. 1. – С.47-54.
4. Курейчик В.М. Биоинспирированный поиск с использованием сценарного подхода // Известия ЮФУ. Технические науки. – 2010. – № 7 (108). – С. 7-13.
5. Курейчик В.М., Кажаров А.А. Использование роевого интеллекта в решении NP-трудных задач // Известия ЮФУ. Технические науки. – 2011. – № 7 (120). – С. 30-36.

Статью рекомендовала к опубликованию д.т.н., профессор Я.Е. Ромм.

Мелихова Оксана Аскольдовна – Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Южный федеральный университет»; e-mail: ao@tsure.ru; 347928, г. Таганрог, ул. Энгельса, 1; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Melikhova Oksana Askoldovna – Federal State-Owned Autonomy Educational Establishment of Higher Vocational Education “Southern Federal University”; e-mail: ao@tsure.ru; 1, Engelse street, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; associate professor.

УДК 681.31

А.Ю. Полуян, И.В. Марейченко

ПОСТРОЕНИЕ БИОНИЧЕСКОГО ПОИСКА ДЛЯ ЗАДАЧ ОБ ЭКСТРЕМАЛЬНОМ ПУТИ НА ОСНОВЕ СТРАТЕГИИ АДАПТАЦИИ

Рассматривается проблема разработки алгоритма бионического поиска для задач об экстремальном пути на графе. В настоящее время разработка эффективных методов и алгоритмов для задач данного типа осуществляется много лет, являясь по-прежнему актуальной проблемой. Перспективной является разработка бионических алгоритмов (БА) на основе эволюционных стратегий, особенно при решении трудоемких задач оптимизации. К преимуществам можно отнести: возможность выполнения эволюционного и генетического поиска, а также то, что БА состоит в параллельной генерации наборов квази-оптимальных альтернативных решений с возможной «миграцией» решений между этими