

Раздел I. Эволюционное моделирование, генетические и бионические алгоритмы

В.М. Курейчик

ГИБРИДНЫЕ ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ*

Введение. Создание интеллектуальных систем автоматизированного проектирования (ИСАПР) является одной из важнейших проблем науки и техники сегодня. Важной тенденцией развития и создания ИСАПР является активное развитие и использование в них бионических технологий, «биоинспирированных» алгоритмов, а также методов и средств эволюционного моделирования (ЭМ) и проектирования (ЭП) [1-5]. Под ЭП будем понимать комплекс средств автоматизированного проектирования, состоящего из семи видов обеспечений. Основным из них является математическое обеспечение, которое состоит из моделей эволюции, генетических и эволюционных алгоритмов, использующихся на всех стадиях разработки, создания и эксплуатации ИСАПР. Эволюционное проектирование на основе ИСАПР включает методы эволюционного моделирования, программирования и адаптацию к внешней среде.

В работе предлагается новая структура гибридных генетических алгоритмов для создания математического обеспечения, позволяющая повысить качество технических решений. Генетические алгоритмы (ГА) – это новая технология решения задач оптимизации, искусственного интеллекта, проектирования и т.п. [1-9]. Автор рассматривает формальное задание генетических алгоритмов, новые модифицированные генетические операторы, ориентированные на использование знаний о решаемых задачах, и описывает модифицированную структуру генетического алгоритма, позволяющую уменьшить на порядок размер популяции, что сокращает время работы ГА.

1. Генетические алгоритмы. Генетические алгоритмы (ГА) отличаются от других оптимизационных и поисковых методов и алгоритмов [1-3,6,7,9] тем, что:

- ◆ анализируют и преобразуют закодированное множество исходных параметров;
- ◆ осуществляют поиск из части популяции, популяции или множества популяций (множества альтернативных решений), а не из одного решения;
- ◆ используют целевую функцию (ЦФ), а не ее различные приращения для оценки качества альтернативных решений;
- ◆ используют детерминированные, вероятностные и комбинированные правила анализа оптимизационных задач.

В ГА предварительно анализируется множество входных параметров оптимизационной задачи и находится некоторое множество альтернативных решений, которое называется популяцией. Каждое решение кодируется как последовательность конечной длины в некотором алфавите. ГА работает до тех пор, пока не бу-

* Работа выполнена при поддержке РФФИ (гранты № 05-08-18115, № 06-01-00272) и программ развития научного потенциала высшей школы 2006-2008 гг. (РНП.2.1.2.3193, РНП 2.1.2.2238).

дет получено решение заданного качества или будет выполнено заданное количество генераций, или на некоторой генерации возникает преждевременная сходимость, когда найден локальный оптимум. Процесс эволюции основан на анализе начальной популяции и ГА начинает свою работу с создания исходного множества альтернативных решений. Затем эти «родительские» решения создают «потомков» путем случайных, направленных или комбинированных преобразований. После этого оценивается эффективность каждого альтернативного решения, которые подвергаются селекции. Во всех моделях эволюции используется принцип «выживания сильнейших», т.е. наименее приспособленные решения устраняются, а лучшие решения переходят в следующую генерацию. Затем процесс повторяется вновь [1–3, 4–6].

Генетические алгоритмы манипулируют популяцией хромосом на основе механизма натуральной эволюции. Приведем формальное определение ГА [7].

$$GA = (P_i^o, N, P_{i,k}^T, T, L_j, A, (CF, OGP, GU), GO, t),$$

где P_i^o – исходная популяция хромосом (альтернативных решений), $P_i^o = (P_{i1}^o, P_{i2}^o, \dots, P_{in}^o)$, $P_{i1}^o \in P_i^o$ – хромосома (альтернативное решение), принадлежащее i -ой исходной популяции; N – мощность популяции, т.е. число входящих в нее хромосом, $N = |P_i^T|$; $P_{ik}^T \in P_i^T$ – k -я хромосома, принадлежащая i -ой популяции, находящейся в T поколении эволюции; $T = 0, 1, 2, \dots$ – номер поколения, проходящего популяцией во время эволюции (иногда число поколений связывают с числом генераций генетического алгоритма, обозначаемых буквой G); L_j – длина i -ой хромосомы (альтернативного решения), т.е. число генов (элементов, входящих в закодированное решение, представленное в заданном алфавите), например, $|P_i^T| = L_j$; A – произвольный абстрактный алфавит, в котором кодируются хромосомы, например, $A_1 = \{0, 1\}$, $A_2 = \{0, 1, 2, \dots, 10\}$, $A_3 = \{0, 1, 2, *\}$, $A_4 = \{A, B, C, D\}$, здесь * – метка, означающая любой символ в алфавите A_2 ; (CF, OGP, GU) – целевая функция, ограничения и граничные условия, которые определяются на основе заданной модели исходной решаемой задачи; GO – генетические операторы; t – критерий окончания работы ГА.

2. Простой генетический алгоритм. Эволюционный процесс представляется как способность «лучших» хромосом оказывать большее влияние на состав новой популяции на основе длительного выживания из более многочисленного потомства. Основные этапы эволюционного поиска следующие:

1. Конструируется начальная популяция. Вводится точка отсчета поколений $t = 0$. Вычисляется приспособленность каждой хромосомы в популяции, а затем определяется средняя приспособленность всей популяции.

2. Устанавливается $t = t + 1$. Производится выбор двух родителей (хромосом) для реализации оператора кроссинговера. Он выполняется случайным образом пропорционально приспособляемости родителей.

3. Формируется генотип потомков. Для этого с заданной вероятностью производится оператор кроссинговера над генотипами выбранных хромосом. Далее с вероятностью 0,5 выбирается один из потомков $P_i(t)$ и сохраняется как член новой популяции. После этого к $P_i(t)$ последовательно применяется оператор инверсии, а затем мутации с заданными вероятностями. Полученный генотип потомка сохраняется как $P_k(t)$.

4. Определяется количество хромосом для исключения их из популяции, чтобы ее размер оставался постоянным. Текущая популяция обновляется заменой отобранных хромосом на потомков $P_k(t)$.

5. Производится определение приспособленности (целевой функции) и пересчет средней приспособленности всей полученной популяции.

6. Если $t = t_{\text{заданному}}$, то переход к 7, если нет, то переход к 2.

7. Конец работы.

Данный алгоритм известен как упрощенный «репродуктивный план Д. Холланда» [1,8,10]. Заметим, что в практических задачах вместо понятия «приспособленность» используют понятие «целевая функция».

Простой генетический алгоритм (ПГА) был впервые описан Д. Гольдбергом на основе работ Д. Холланда [2,7]. Его механизм несложен. Предварительно ПГА случайно генерирует популяцию последовательностей – хромосом (альтернативных упорядоченных и неупорядоченных решений). Затем производится копирование последовательности хромосом и перестановка их частей. Далее ПГА реализует множество простых операций к начальной популяции и генерирует новые решения. ПГА состоит из трех операторов: репродукция, кроссинговер, мутация.

Репродукция – процесс, в котором хромосомы копируются пропорционально значению их ЦФ. Копирование хромосом с «лучшим» значением ЦФ имеет большую вероятность для попадания в следующую генерацию. Рассматривая эволюцию Ч. Дарвина, можно отметить, что оператор репродукции (ОР) является искусственной версией натуральной селекции «выживания сильнейших». Он представляется в алгоритмической форме различными способами. Самый простой – создать модель «колеса рулетки», в которой каждая хромосома имеет поле, пропорциональное значению ЦФ. В генетических алгоритмах можно выделять два основных механизма воспроизводства хромосом: потомки являются точными копиями родителей (неполовое воспроизводство без мутации); потомки имеют «большие» отличия от родителей. В генетических алгоритмах в основном используют комбинации этих механизмов.

3. Модифицированные генетические операторы. При реализации генетических операторов (ГО) одним из важнейших вопросов является определение места и количества точек разрыва хромосом (альтернативных решений). При этом большое количество точек разрыва может привести к полной потере лучших решений. Маленькое количество точек разрыва часто приводит к попаданию решения в локальный оптимум, далекий от глобального. Поэтому необходим поиск разумного компромисса в этом вопросе. Автор предлагает использовать ГО на основе чисел Фибоначчи, золотого сечения, множества Кантора, дихотомического деления и др. Отметим, что необходимо проводить эксперименты и подбирать параметры, управляющие ГО на основе адаптационных процессов.

Рассмотрим построение модифицированных ГО на основе простых чисел и чисел Каталана [11]. Как известно, простые числа легко определяются на основе решета Эратосфена [11]. Приведем, например, ряд простых чисел (ПЧ) не превышающих 100: 1, 3, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97 . . .

Очевидно, что такой ряд можно продолжить. Автор предлагает длину ряда выбирать такую, чтобы последнее число ряда ПЧ было меньше, чем $L-1$. Здесь L – длина хромосомы (альтернативного решения).

Рассмотрим, например, реализацию различных ГО на основе ПЧ.

Оператор кроссинговера (ОК).

Даны две родительских хромосомы:

P_1 1 | 2 3 | 4 5 6 7 | 8 9 10;

P_2 a | b c | d e f g | h I k |.

Можно выбрать любое число из ряда длины $L-1=10-1=9$, т.е. 1,3 или 7. Известно, что можно выполнить одно, двух или многоточечный ОК. Выберем одновременно три точки разрыва. Здесь можно предложить большое число вариантов получения потомков.

Чередующийся ОК.

Тогда хромосомы-потомки получаются путем чередования строительных блоков из P_1 и P_2 :

$$P_1 = 1 bc45 67 h i k;$$

$$P_2 = a23def g89 10.$$

На рис.1 показан блок оператора кроссинговера на основе простых чисел.

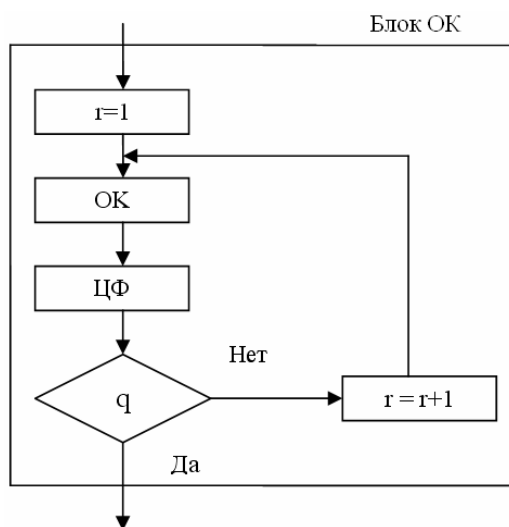


Рис.1. Блок оператора кроссинговера

Здесь r – число точек разрыва; L – длина хромосомы; $r \leq L-1$; q – проверка окончания работы блока (она может выполняться: по значению ЦФ, по времени, по числу генераций и т.п.). Если в результате ОК получаются нереальные решения, то по определенным правилам происходит замена повторяющегося гена на отсутствующий из хромосомы потомка.

ОК с разным числом точек разрыва.

Рассмотрим реализацию ГО на основе чисел Каталана. По аналогии с числами Фибоначчи существует последовательность чисел Каталана [10]. Она задается формулой:

$$k_{am}(n) = \frac{(2n)!}{(n+1)!(n!)} \tag{1}$$

Если последовательность начинается с элемента, равного 1, т.е. $n=1$, то первые 9 элементов определяются так:

$$k = 1, 2, 5, 14, 42, 429, 1430, 4862.$$

Например, при $n=4$ получим $k(4) = \frac{(2 \cdot 4)!}{(4+1)!(4)!} = \frac{12345}{12345} \cdot \frac{678}{1234} = 14.$

Очевидно, что

$$k(0) = 1, \text{ т.к. } 0! = 1 \text{ и при } n=2 \quad k = \frac{(2n)!}{(n+1)!(n!)} = \frac{(2 \cdot 2)!}{3!2!} = 2,$$

а при $n = 1 \quad k(1) = \frac{2!}{2!1!} = 1,$

$k(2) = 2, k(3) = 5, k(4) = 14, k(5) = 42, k(6) = 429, k(7) = 1430, k(8) = 4862.$

Выполнение ГА на основе чисел Каталана эффективно при наличии хромосом большой длины при малом числе испытаний. Механизм реализации ГА на основе ПЧ и чисел Каталана идентичны.

Рассмотрим **оператор мутации (ОМ)**. Пусть задана родительская хромосома:

$$P_1 = 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12.$$

Здесь также $r = L-1$. Тогда из последовательности чисел Каталана (ЧК) получим, что можно выполнить разрывы после гена 1, гена 2 и(или) гена 5.

Выполним трехточечный ОМ:

$$P_1 = 1 | 2 | 3 \ 4 \ 5 | 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12.$$

Здесь возможны следующие модификации ОМ:

а) $P_1 = 2 \ 3 \ 1 \ 4 \ 6 \ 5 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12.$

Здесь ген 1 меняется местами с геном 2. Затем ген 1 меняется местами с геном 3 и ген 5 меняется местами с геном 6.

В настоящее время перспективными считаются нечеткие ГО [12]. В них нечеткими величинами являются гены хромосомы и строительные блоки с функциями принадлежности, описываемыми кривыми Гаусса. Нечеткие множества предлагается использовать для выбора эффективных точек разрыва в ГО.

В начале 70-х годов Н. Эддридже и С. Гулд предложили новую модель эволюции – прерывистого равновесия [13]. Согласно этой новой модели эволюция происходит редкими и быстрыми толчками. Следует отметить, что эти идеи высказывали еще Ч. Дарвин, Г. де Фриз [6-10] и др. В этой модели отмечается различие причин, от которых зависят темпы микро- и макроэволюции. В настоящее время эту модель эволюции называют моделью прерывистого равновесия Гулда-Эддриджа [13]. На рис.2 и рис.3 приведены условные структуры микро- и макроэволюции согласно модели Гулда-Эддриджа.

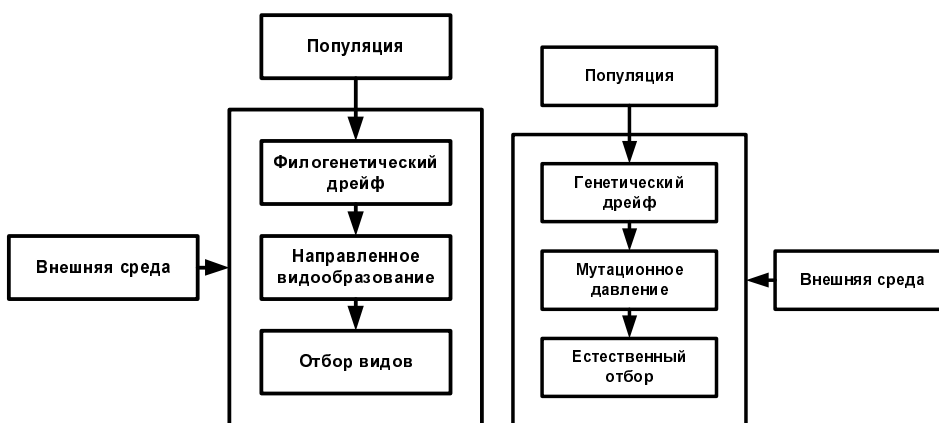


Рис.2. Структура макроэволюции

Рис.3. Структура микроэволюции

Макроэволюция основана на анализе и преобразовании популяций альтернативных решений. В процессе микроэволюции исследуется только одна хромосома.

Филогенетический дрейф – это возникновение разнонаправленных видов в силу случайных эволюционных флуктуаций [13]. Он аналогичен генетическому дрейфу особей (хромосом), но в популяциях.

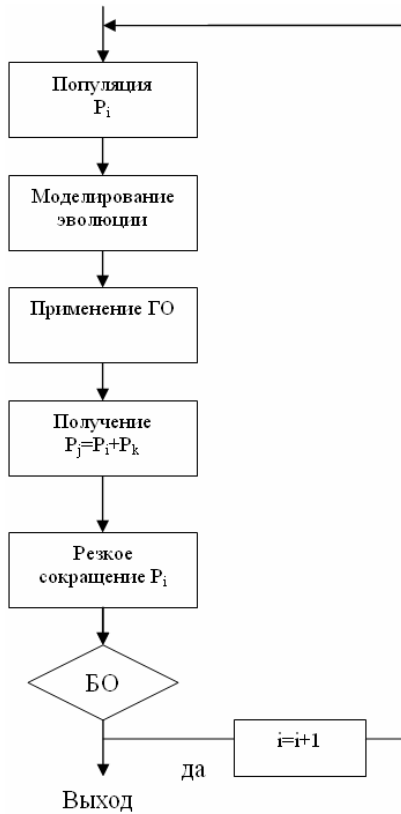


Рис.4. Схема ГА на основе эффекта «бутылочного горлышка»

Ограничения: $\sum a_i v_i \leq V_r$, $a_i \in [0, 1]$.

Стандартная целевая функция: $F = \sum c_i$.

Цель: $F \rightarrow \max$.

Был сформирован набор ЦФ, включающих традиционную ЦФ:

$$f(x) = \sum_k c_k, \tag{2}$$

$$f(x) = \sum_k c_k^2 \tag{3}$$

и несколько модифицированных ЦФ:

$$f(x) = \sum_k c_k^3. \tag{4}$$

ПГА решения задачи о рюкзаке был протестирован со всеми выбранными целевыми функциями. Случайным образом был сгенерирован набор тестовых примеров, для каждого из которых построен ландшафт ЦФ. Для каждого тестового при-

Эффект «бутылочного горлышка» [13] – это возникновение популяции с генетическим строением, определяемым случайностью сохранения немногих особей, с тем или иным генотипом, уцелевших после резкого сокращения численности родительской популяции. Приведем условную схему ГА на основе эффекта бутылочного горлышка (рис. 4).

Нет

Вариант практического использования покажем на примере задачи о рюкзаке [14]. Цель – показать, каким образом можно использовать практически разработанный комплекс алгоритмов для выбора эффективной ЦФ.

Постановка задачи следующая. Дан набор элементов и рюкзак. Каждый элемент имеет объём v и стоимость c . Необходимо разместить элементы в рюкзаке таким образом, чтобы их суммарная стоимость была максимальной. Проблема заключается в том, что не все элементы могут поместиться в рюкзаке.

Имеется n вещей и рюкзак.

- ◆ $V = \{v_1, v_2, \dots, v_n\}$ – вектор объёмов вещей;
- ◆ $C = \{c_1, c_2, \dots, c_n\}$ – вектор стоимостей вещей;
- ◆ V_r – объём рюкзака.

мера и каждой ЦФ ПГА запускался по 1000 раз. Оказалось, что выбранная ЦФ (3) является лучшей для решения задачи о рюкзаке при помощи простого генетического алгоритма.

Пример трехуровневой гибридной схемы с набором различных ГА показан на рис.5.

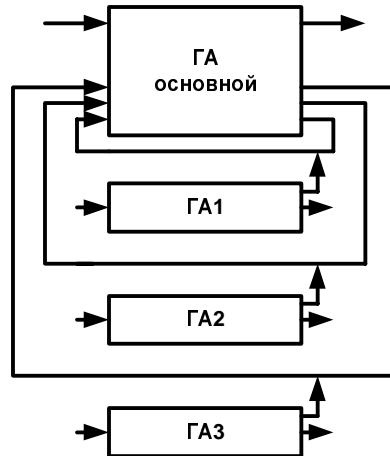


Рис.5. Пример трехуровневой схемы

Приведем основную идею новой структуры ГА. Она позволяет сократить время поиска и получить локально-оптимальные решения.

1. Конструирование популяции:
 - 1.1. На основе «создателя».
 - 1.2. Дихотомии.
 - 1.3. Чисел Фибоначчи.
 - 1.4. Чисел Каталана.
 - 1.5. Фокусировки.
 - 1.6. «Одеяла».
 - 1.7. «Дробовика».
 - 1.8. Конструирование строительных блоков на основе знаний о решаемой задаче
 2. Селекция:
 - 2.1. Элитная:
 - 2.2.1. Один со многими.
 - 2.3. Многие с одним.
 - 2.4. Лучшие с худшими.
 - 2.5. На основе ветвей и границ.
 3. Кроссинговер (ОК).
 4. Мутация.
 5. Транспозиция и транслокация с формулами выживания.
 6. Смешанный интегрированный оператор.
 7. Инверсия.
 8. Редукция популяции.
 9. Проверка окончания работы или переход на новый цикл алгоритма.
- Здесь используется альтернативное решение изменяющейся длины. Кроме того, вводится концепция нечеткости для определения ЦФ и правил реализации ГО.

В заключении отметим, что модифицированные генетические операторы должны выполняться в блоке эволюционной адаптации совместно с экспертной советующей подсистемой. Это позволит эффективно подбирать такие параметры как число точек разреза и вероятность применения генетического оператора.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Holland John H.* Adaptation and Artificial Systems: An Introductory Analysis with Application to Biology, Control, and Artificial Intelligence. USA: University of Michigan, 1975.
2. *Goldberg David E.* Genetic Algorithms in Search, Optimization and Machine Learning. USA: Addison-Wesley Publishing Company, Inc., 1989.
3. Handbook of Genetic Algorithms. Edited by Lawrence Davis. USA: Van Nostrand Reinhold, New York, 1991.
4. Practical Handbook of Genetic Algorithms. Editor I. Chambers. V.3, Washington, USA, CRC Press, 1999.
5. Handbook of Bioinspired Algorithms and Applications. Edited by S. Olarin, A. Y. Zomaya, 2006, Chapman & Hall CRC, USA.
6. *Редько В.Г.* Эволюция, нейронные сети, интеллект: модели и концепции эволюционной кибернетики. – М.: Комкнига, 2005.
7. *Гладков Л.А., Курейчик В.В., Курейчик В.М.* Генетические алгоритмы: Учебное пособие / Под редакцией В.М. Курейчика. – Ростов-на-Дону: Ростиздат, 2004.
8. *Емельянов В.В., Курейчик В.М., Курейчик В.В.* Теория и практика эволюционного моделирования. – М.: Физматлит, 2003.
9. Интеллектуальные системы: Монография / Под редакцией В.М. Курейчика. – М.: Физматлит, 2005.
10. *Гладков Л.А., Курейчик В.В., Курейчик В.М.* Генетические алгоритмы: Учебное пособие / Под редакцией В.М. Курейчика. – М.: ФИЗМАТЛИТ, 2006.
11. *Андерсон Д.* Дискретная математика и комбинаторика. – М.: Вильямс, 2003.
12. *Голубин А.В., Тарасов В.Б.* Нечеткие генетические алгоритмы / Труды международных НТК «AIS-05», «CAD-2005». – М.: Физматлит, 2005, Т.1. – С. 39-45.
13. *Назаров В.И.* Эволюция не по Дарвину: смена эволюционной модели / Учебное пособие. – М.: Комкнига, 2005.
14. *Каляда А.В., Курейчик В.М.* Анализ ландшафтов целевых функций при генетическом поиске. Труды 10 НК по искусственному интеллекту, том 2, – М.: Физматлит, 2006. – С. 392-399.

Б.К. Лебедев, В.Б. Лебедев

БИОНИЧЕСКИЕ ПРОЦЕДУРЫ НАХОЖДЕНИЯ МИНИМАЛЬНОГО ПОКРЫТИЯ*

Введение. Исходными данными задачи о покрытии множествами являются конечное множество X , а также семейство его подмножеств $G = \{X_i / i=1,2,\dots,n\}$ таких, что $X_i \subseteq X$ и $\bigcup_i X_i = X$.

Задача о минимальном покрытии множествами заключается в отыскании набора $P \subseteq G$ с минимальным числом подмножеств $X_i \in P$ и $X_i \in F$ такого, чтобы

* Работа выполнена при поддержке РФФИ (грант № 06-01-81018, № 05-08-18115) и программ развития научного потенциала высшей школы 2006-2008 гг. (РНП.2.1.2.3193, РНП 2.1.2.2238).