
Методы глобальной ОПТИМИЗАЦИИ

Генетические алгоритмы

Генетические алгоритмы (ГА)

Адаптивные методы поиска, которые в последнее время часто используются для решения задач функциональной оптимизации.

Они основаны на генетических процессах биологических организмов: биологические популяции развиваются в течении нескольких поколений, подчиняясь законам естественного отбора и по принципу "выживает наиболее приспособленный" (survival of the fittest), открытому Чарльзом Дарвином.

Подражая этому процессу генетические алгоритмы способны "развивать" решения реальных задач, если те соответствующим образом закодированы

Эволюция: естественный отбор и генетическое наследование

- *Естественный отбор*: наиболее приспособленные особи лучше выживают и приносят больше потомства, чем менее приспособленные
 - *Основной закон наследования*: потомки похожи на родителей. В частности, потомки более приспособленных родителей будут, скорее всего, одними из наиболее приспособленных в своем поколении
-

Строение животной клетки

- Почти в каждой клетке любого животного имеется набор **хромосом**, несущих информацию об этом животном. Основная часть хромосомы - нить **ДНК** (**молекула дезоксирибонуклеиновой кислоты**), которая состоит из четырех видов специальных соединений - нуклеотидов, идущих в определенной последовательности. Нуклеотиды обозначаются буквами А, Т, С и G, и именно порядок их следования кодирует все генетические свойства данного организма. Говоря более точно, ДНК определяет, какие химические реакции будут происходить в данной клетке, как она будет развиваться и какие функции выполнять.
- **Ген** - это отрезок цепи ДНК, отвечающий за определенное свойство особи, например за цвет глаз, тип волос, цвет кожи и т. д. Вся совокупность генетических признаков человека кодируется посредством примерно 60 тыс. генов, суммарная длина которых составляет более 90 млн. нуклеотидов.
- Различают два вида клеток: половые и соматические. В каждой соматической клетке человека содержится 46 хромосом. Эти 46 хромосом - 23 пары, причем в каждой паре одна из хромосом получена от отца, а вторая - от матери. Парные хромосомы отвечают за одни и те же признаки - например, отцовская хромосома может содержать ген черного цвета глаз, а парная ей материнская - ген голубоглазости. Существуют определенные законы, управляющие участием тех или иных генов в развитии особи. В частности, в данном примере потомок будет черноглазым, так как ген голубых глаз является "слабым" (рецессивным) и подавляется геном любого другого цвета.
- В половых клетках хромосом только 23, и они непарные. При оплодотворении происходит слияние мужской и женской половых клеток и образуется клетка зародыша, содержащая как раз 46 хромосом.

Факторы, влияющие на наследственность

- **Кроссинговер** - парные хромосомы соматической клетки сближаются вплотную, затем их нити ДНК разрываются в нескольких случайных местах и хромосомы обмениваются своими частями. Этот процесс обеспечивает появление новых вариантов хромосом
- **Мутация** - изменение некоторых участков ДНК. Мутации также случайны и могут быть вызваны различными внешними факторами, такими, как радиоактивное облучение. Если мутация произошла в половой клетке, то измененный ген может передаться потомку и проявиться в виде наследственной болезни либо в других новых свойствах потомка. Считается, что именно мутации являются причиной появления новых биологических видов, а кроссинговер определяет уже изменчивость внутри вида (например, генетические различия между людьми)

Задача оптимизации

- Таким образом, эволюция - это процесс постоянной оптимизации биологических видов. Естественный отбор гарантирует, что наиболее приспособленные особи дадут достаточно большое потомство, а благодаря генетическому наследованию, что часть этого потомства не только сохранит высокую приспособленность родителей, но будет обладать и некоторыми новыми свойствами. Если эти новые свойства окажутся полезными, то с большой вероятностью они перейдут и в следующее поколение. Так происходит накопление полезных качеств и постепенное повышение приспособляемости биологического вида в целом.
 - Таким образом решается задача оптимизации видов в природе, похожий метод можно применять для решения различных реальных задач оптимизации - наиболее распространенного и важного для практики класса задач
-

Генетический алгоритм – метод глобальной оптимизации

- Генетические алгоритмы имитируют процессы наследования свойств живыми организмами и генерируют последовательности новых векторов , содержащие оптимизированные переменные $w=(w_1, w_1, \dots, w_n)$.
 - При этом выполняются операции трех видов: селекция, скрещивание и мутация.
-

Отличие ГА от традиционных методов оптимизации

Генетические алгоритмы:

- обрабатывают не значения параметров самой задачи, а их закодированную форму
- осуществляют поиск решения исходя не из единственной точки, а из некоторой популяции
- используют только целевую функцию, а не ее производные либо иную дополнительную информацию
- применяют вероятностные, а не детерминированные правила выбора

Основные понятия ГА

- **Популяция** – это конечное множество особей
- **Особи**, входящие в популяцию, в генетических алгоритмах представляются хромосомами с закодированным в них множествами параметров задачи, т.е. решений, которые иначе называются точками в пространстве поиска
- **Хромосомы** (другие названия – цепочки или кодовые последовательности) – это упорядоченные последовательности генов
- **Ген** (также называемый свойством, знаком или детектором) – это атомарный элемент генотипа, в частности, хромосомы



Основные понятия ГА

- **Генотип** или структура – это набор хромосом данной особи. Следовательно, особями популяции могут быть генотипы либо единичные хромосомы (в довольно распространенном случае, когда генотип состоит из одной хромосомы).
- **Фенотип** – это набор значений, соответствующих данному генотипу, т. е. декодированная структура или множество параметров задачи (решение, точка пространства поиска).
- Каждая хромосома может кодировать только один параметр задачи. Но так как в одном генотипе может быть несколько хромосом, то один генотип может закодировать несколько параметров. Например: генотип включает в себя две хромосомы, каждая хромосома состоит из 3 генов, значит, первые 3 гена генотипа кодируют один параметр, а вторые 3 гена – второй параметр. Проиллюстрируем пример генотипа и фенотипа:



Основные понятия ГА

- **Аллель** – это значение конкретного гена.
 - **Локус** или позиция указывает место размещения данного гена в хромосоме (цепочке).
 - **Локи** – это множество позиций генов.
-

Основные понятия ГА

- Очень важным понятием в генетических алгоритмах считается **функция приспособленности**, иначе называемая функцией оценки. Она представляет меру приспособленности данной особи в популяции. Эта функция играет важнейшую роль, поскольку позволяет оценить степень приспособленности конкретных особей в популяции и выбрать из них наиболее приспособленные (т.е. имеющие наибольшие значения функции приспособленности) в соответствии с эволюционным принципом выживания «сильнейших») (лучше всего приспособившихся). Функция приспособленности также получила свое название непосредственно из генетики. Она оказывает сильное влияние на функционирование генетических алгоритмов и должна иметь точное и корректное определение
- Очередная популяция в генетическом алгоритме называется **поколением**, а к вновь создаваемой популяции особей применяется термин «новое поколение» или «поколение потомков»

Упрощенный «Репродуктивный план Д. Холланда»

1. Сконструировать исходную популяцию. Ввести точку отсчета поколений $t=0$. Вычислить приспособленность каждой хромосомы в популяции, а затем среднюю приспособленность всей популяции.
2. Установить $t=t+1$. Произвести выбор двух родителей для реализации оператора кроссинговера. Он выполняется случайным образом пропорционально приспособляемости родителей.
3. Применить оператор кроссинговера к выбранным хромосомам. Далее с вероятностью 0,5 выбрать один из потомков и сохранить как член новой популяции $P_i(t)$. После этого к $P_i(t)$ применить оператор мутации и полученный хромосому потомка сохранить как $P_k(t)$
4. Обновить текущую популяцию заменой отобранных хромосом на потомков $P_k(t)$
5. Вычислить приспособленность каждой хромосомы в популяции, а затем среднюю приспособленность всей популяции.
6. Если $t=t_{\text{заданному}}$, то перейти к 7, иначе к 2.
7. Конец

Простой генетический алгоритм

Описан Д. Гольдбергом на основе работ Д. Холланда

- Создание начальной популяции хромосом (альтернативных решений)
- Определение (задание) функций приспособленности для особей популяции (оценивание)
- (Начало цикла)
 - Репродукция (селекция)
 - Скрещивание и\или мутация
 - Вычисление функций приспособленности для всех особей
 - Формирование нового поколения
 - Если выполняются условия останова, то (конец цикла), иначе (начало цикла).

Создание начальной популяции хромосом

Случайным образом инициализируется определенная популяция хромосом - векторов $w = (w_1, w_1, \dots, w_n)$.

Размер популяции, как правило, пропорционален количеству оптимизируемых параметров. Слишком малая популяция хромосом приводит к замыканию в неглубоких локальных минимумах. Слишком большое их количество чрезмерно удлиняет вычислительную процедуру и также может не привести к точке глобального минимума.

Кодирование вектора $w = (w_1, w_2, \dots, w_n)$

- Компоненты вектора $w = (w_1, w_2, \dots, w_n)$ могут кодироваться в двоичной системе (обычный код или код Грея) либо натуральными числами. После чего отдельные биты представляют конкретные значения параметров, подлежащих оптимизации
- Каждому вектору w сопоставляется определенное значение целевой функции. Генетические операции выполняются для подбора таких значений переменных w_i , при которых $F(w) \rightarrow \max$ (функция приспособленности)

Кодирование признаков, представленных целыми числами

- Битовое значение этого признака - используется ген определенной длины, достаточной для представления всех возможных значений такого признака

Недостатки:

- соседние числа отличаются в значениях нескольких битов, так например числа 7 и 8 в битовом представлении различаются в 4-х позициях, что затрудняет функционирование генетического алгоритма и увеличивает время, необходимое для его сходимости
 - Код Грея
-

Двоичное кодирование**Кодирование по коду Грея**

Десятичный код	Двоичное значение	Шестнадцатеричное значение	Десятичный код	Двоичное значение	Шестнадцатеричное значение
0	0000	0h	0	0000	0h
1	0001	1h	1	0001	1h
2	0010	2h	3	0011	3h
3	0011	3h	2	0010	2h
4	0100	4h	6	0110	6h
5	0101	5h	7	0111	7h
6	0110	6h	5	0101	5h
7	0111	7h	4	0100	4h
8	1000	8h	12	1100	Ch
9	1001	9h	13	1101	Dh
10	1010	Ah	15	1111	Fh
11	1011	Bh	14	1110	Eh
12	1100	Ch	10	1010	Ah
13	1101	Dh	11	1011	Bh
14	1110	Eh	9	1001	9h
15	1111	Fh	8	1000	8h

Пример

- Допустим, что у объекта имеется 5 признаков, каждый закодирован геном длиной в 4 элемента. Тогда длина хромосомы будет $5 \cdot 4 = 20$ бит

0010 1010 1001 0100 1101

Признак	Значение гена	Двоичное значение признака	Десятичное значение признака
Признак 1	0010	0011	3
Признак 2	1010	1100	12
Признак 3	1001	1110	14
Признак 4	0100	0111	7
Признак 5	1101	1001	9

Селекция

Репродукция (селекция) – заключается в выборе по рассчитанным значениям функции приспособленности тех хромосом, которые будут участвовать в создании потомков для следующей популяции, т.е. для очередного поколения. Такой выбор производится согласно принципу естественного отбора, по которому наибольшие шансы на участие в создании новых особей имеют хромосомы с наибольшими значениями функции приспособленности

В результате процесса селекции создается родительская популяция, также называемая родительским пулом с численностью, равной численности текущей популяции.

Виды операторов селекции

- Селекция на основе рулетки
 - Селекция на основе заданной шкалы
 - Элитная селекция
 - Турнирная селекция
-

Метод рулетки

- Каждой хромосоме может быть сопоставлен сектор колеса рулетки, величина которого устанавливается пропорциональной значению функции приспособленности данной хромосомы. Поэтому чем больше значение функции приспособленности, тем больше сектор на колесе рулетки. Все колесо рулетки соответствует сумме значений функции приспособленности всех хромосом рассматриваемой популяции.
- Каждой хромосоме, обозначаемой ch_i для $i=1..N$ (где N обозначает численность популяции) соответствует сектор колеса $v(ch_i)$, выраженный в процентах согласно формуле:

$$v\left(ch_i\right) = p_s\left(ch_i\right) \times 100\%$$

$$p_s\left(ch_i\right) = \frac{F\left(ch_i\right)}{\sum_{i=1}^N F\left(ch_i\right)}$$

$F\left(ch_i\right)$ - значение функции приспособленности хромосомы ch_i

$p_s\left(ch_i\right)$ - вероятность селекции хромосомы ch_i

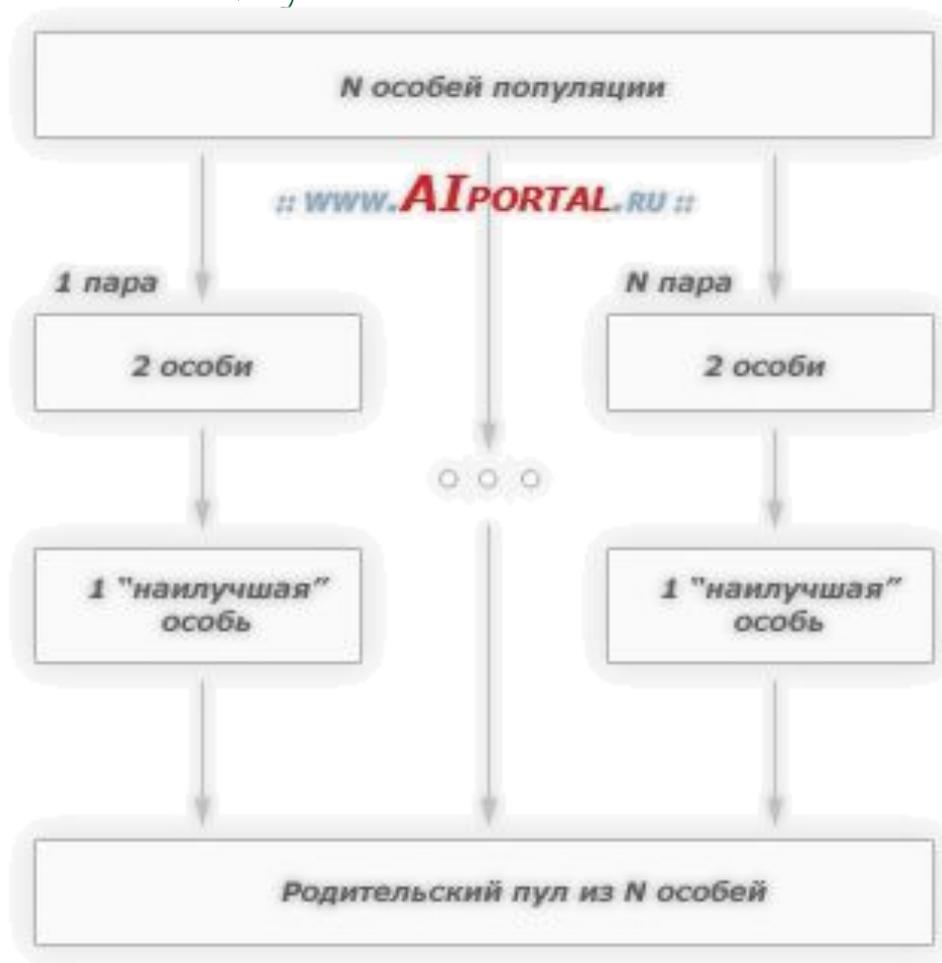
Колесо рулетки

- Селекция хромосомы может быть представлена как результат поворота колеса рулетки, поскольку «выигравшая» (т.е. выбранная) хромосома относится к выпавшему сектору этого колеса. Очевидно, что чем больше сектор, тем больше вероятность «победы» соответствующей хромосомы. Поэтому вероятность выбора данной хромосомы оказывается пропорциональной значению ее функции приспособленности.
- Если всю окружность колеса рулетки представить в виде цифрового интервала $[0, 100]$, то выбор хромосомы можно отождествить с выбором числа из интервала $[A, B]$, где A и B обозначают соответственно начало и окончание фрагмента окружности, соответствующего этому сектору колеса; очевидно, что $0 \leq A < B \leq 100$. В этом случае выбор с помощью колеса рулетки сводится к выбору числа из интервала $[0, 100]$, которое соответствует конкретной точке на окружности колеса.

Турнирная селекция

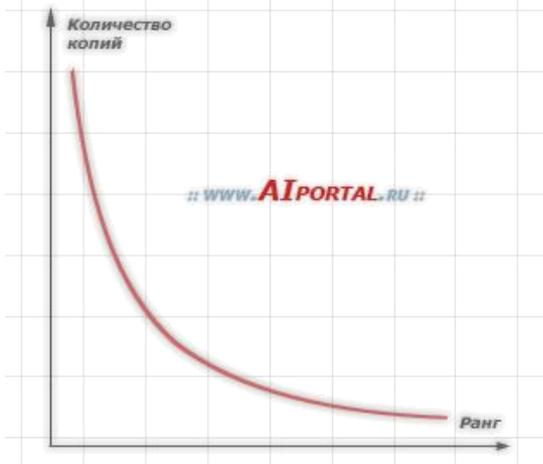
- При турнирной селекции все особи популяции разбиваются на подгруппы с последующим выбором в каждой из них особи с наилучшей приспособленностью. Различаются два способа такого выбора: детерминированный выбор и случайный выбор. Детерминированный выбор осуществляется с вероятностью, равной 1, а случайный выбор – с вероятностью, меньшей 1. Подгруппы могут иметь произвольный размер, но чаще всего популяция разделяется на подгруппы по 2-3 особи в каждой.
- Турнирный метод пригоден для решения задач как максимизации, так и минимизации функции. Помимо того, он может быть легко распространен на задачи, связанные с многокритериальной оптимизацией, т.е. на случай одновременной оптимизации нескольких функций. В турнирном методе допускается изменение размера подгрупп, на которые подразделяется популяция. Исследования подтверждают, что турнирный метод действует эффективнее, чем метод рулетки.

Метод турнирной селекции для подгрупп, состоящих из двух особей



Ранговая селекция

- При ранговой селекции особи популяции ранжируются по значениям их функции приспособленности. Это можно представить себе как отсортированный список особей, упорядоченных по направлению от наиболее приспособленных к наименее приспособленным (или наоборот), в котором каждой особи приписывается число, определяющее ее место в списке и называемое рангом. Количество копий каждой особи, введенных в родительскую популяцию, рассчитывается по априорно заданной функции в зависимости от ранга особи. Пример такой функции может быть следующий график.



- Достоинство рангового метода заключается в возможности его применения как для максимизации, так и для минимизации функции. Достоинство рангового метода заключается в возможности его применения как для максимизации, так и для минимизации функции.

Операторы кроссинговера

Языковая конструкция, позволяющая на основе преобразования хромосом родителей создавать потомков.

- Одноточечный оператор кроссинговера
 - Двухточечный оператор кроссинговера
 - Упорядоченный оператор кроссинговера
 - ...
-

Мутация

- Оператор мутации (mutation operator) необходим для "выбивания" популяции из локального экстремума и способствует защите от преждевременной сходимости. Достигаются эти чудеса за счет того, что инвертируется случайно выбранный бит в хромосоме
- Можно выбирать некоторое количество точек в хромосоме для инверсии, причем их число также может быть случайным. Также можно инвертировать сразу некоторую группу подряд идущих точек. Вероятность мутации значительно меньше вероятности кроссинговера и редко превышает 1%. Среди рекомендаций по выбору вероятности мутации нередко можно встретить варианты $1/L$ или $1/N$, где L - длина хромосомы, N - размер популяции.

Выбор родительской пары

- *Случайный* выбор родительской пары ("*панмиксия*"), когда обе особи, которые составят родительскую пару, случайным образом выбираются из всей популяции, причем любая особь может стать членом нескольких пар. Несмотря на простоту, такой подход универсален для решения различных классов задач. Однако он достаточно критичен к численности популяции, поскольку эффективность алгоритма, реализующего такой подход, снижается с ростом численности популяции.
- *Селективный*. Его суть состоит в том, что "родителями" могут стать только те особи, значение приспособленности которых не меньше среднего значения приспособленности по популяции, при равной вероятности таких кандидатов составить брачную пару. Такой подход обеспечивает более быструю сходимость алгоритма. Однако из-за быстрой сходимости селективный выбор родительской пары не подходит тогда, когда ставится задача определения нескольких экстремумов, поскольку для таких задач алгоритм, как правило, быстро сходится к одному из решений.
- Другие два способа - *инбридинг* и *аутбридинг*. Оба эти метода построены на формировании пары на основе близкого и дальнего "родства" соответственно. Под "родством" здесь понимается расстояние между членами популяции как в смысле геометрического расстояния особей в пространстве параметров (для фенотипов), так и в смысле хэммингового расстояния между хромосомными наборами особей (для генотипов). В связи с этим будем различать генотипный и фенотипный (или географический) инбридинг и аутбридинг. Под инбридингом понимается такой метод, когда первый член пары выбирается случайно, а вторым с большей вероятностью будет максимально близкая к нему особь. Аутбридинг же, наоборот, формирует брачные пары из максимально далеких особей.