

Тема 5.

Генетические алгоритмы

По материалам:

- ♦ Рутковская Д., Пилиньский М., Рутковский Л. **Нейронные сети, генетические алгоритмы и нечеткие системы**: Пер. с польск. И.Д.Рудинского. - М.: Горячая линия — Телеком, 2006. - 452 с.: ил.

Основы теории генетических алгоритмов (ГА)

Нейронные сети были созданы в результате наблюдения за естественными процессами, происходящими в **нервной системе** живых существ, и попыток воспроизведения этих процессов.

Аналогично, **генетические алгоритмы** возникли в результате наблюдения и попыток копирования естественных **процессов, происходящих в мире живых организмов**, в частности, **эволюции и связанной с ней селекции** (естественного отбора) популяций живых существ.

Идею генетических алгоритмов (genetic algorithm - GA) высказал **Джон Холланд** в конце 60-х годов XX века. Холланд был уверен в возможности составить и реализовать в виде компьютерной программы алгоритм, который будет **решать сложные задачи** так, как это делает природа — путём **эволюции**.

Алгоритмы Холланда оперировали последовательностями двоичных цифр, состоящих из нулей и единиц, которые назывались **хромосомами**.

Эти алгоритмы имитировали эволюционные процессы в **поколениях таких хромосом**. В них были реализованы механизмы **селекции** и **репродукции**, аналогичные существующим в природе.

Для отражения **приспособленности хромосомы** требовалась некоторая оценка.

Механизм селекции заключается в выборе хромосом с **наивысшей оценкой** (наиболее приспособленных), которые репродуцируют чаще, чем особи с более низкой оценкой (хуже приспособленные).

Репродукция означает создание новых хромосом в результате рекомбинации генов родительских хромосом.

Рекомбинация — это процесс, в результате которого возникают новые комбинации генов. Для этого используются две операции:

- **скрещивание**, позволяющее создать две совершенно новые хромосомы потомков путём комбинирования генетического материала пары родителей,
- **мутация**, которая может вызывать изменения в отдельных хромосомах.

В ГА применяется ряд терминов, заимствованных из генетики, прежде всего: гены и хромосомы, а также популяция, особь, аллель, генотип, фенотип.

ГА относятся к типу алгоритмов **эвристического поиска**.

Они применяются при разработке программного обеспечения, в системах ИИ, оптимизации, ИНС и в других отраслях знаний.

Следует отметить, что с их помощью решаются задачи, для которых ранее использовались только НС.

ГА часто используются совместно с НС, они могут поддерживать НС или наоборот, либо оба метода взаимодействуют в рамках гибридной системы. ГА также применяются совместно нечёткими системами.

ГА представляет собой метод, отражающий естественную эволюцию методов решения проблем, и в первую очередь, задач оптимизации.

От традиционных методов оптимизации они отличаются несколькими базовыми элементами:

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

Основные понятия

Популяция — конечное множество особей

Особи, входящие в популяцию, представляются хромосомами с закодированным с них множествами параметров задачи, т.е. решений, которые иначе называются **точками в пространстве поиска**.

В некоторых работах особи называются **организмами**.

Хромосомы (цепочки или кодовые последовательности) — это упорядоченные **последовательности генов**.

Ген (свойство, знак или детектор) — атомарный элемент **генотипа**, в частности, хромосомы.

Основные понятия

Популяция — конечное множество особей

Особи, входящие в популяцию, представляются хромосомами с закодированным с них множествами параметров задачи, т.е. решений, которые иначе называются **точками в пространстве поиска**.

В некоторых работах особи называются **организмами**.

Хромосомы (цепочки или кодовые последовательности) — это упорядоченные **последовательности генов**.

Ген (свойство, знак или детектор) — атомарный элемент **генотипа**, в частности, хромосомы.

Генотип или **структура** — это **набор хромосом** данной особи.

Следовательно, **особями популяции** могут быть **генотипы** либо **единичные хромосомы** (распространённый случай: генотип состоит из одной хромосомы).

Фенотип — это набор значений, соответствующих данному генотипу, т.е. декодированная структура или множество параметров задачи (решение, точка пространства поиска).

Аллель — значение конкретного гена, также определяемое как значение свойства или вариант свойства.

Локус или **позиция** — место размещения данного гена в хромосоме (цепочке).
Множество позиций генов — это локи.

Очень важным понятием в теории ГА является **функция приспособленности** (fitness function), иначе называемой **функцией оценки**.

Она представляет меру приспособленности данной особи в популяции. В задачах оптимизации подобная функция называется целевой функцией и максимизируется или минимизируется в зависимости от условий задачи.

При помощи функции приспособленности оценивается приспособленность каждой особи данной популяции на каждой итерации ГА, на основе чего создаётся следующая популяция особей, составляющих множество потенциальных решений проблемы, например, задачи оптимизации.

Очередная популяция в ГА называется **поколением**, а к вновь создаваемой популяции особей применяется термин «новое поколение» или «поколение потомков».

Пример представления популяций:

Рассмотрим функцию $f(x) = 2x^2 + 1$ и допустим, что x принимает значения из интервала от 0 до 15.

Задача оптимизации этой функции заключается в перемещении по пространству, состоящему из 16 точек со значениями 0, 1, ..., 15 для обнаружения той точки, в которой функция принимает максимальное или минимальное значение.

В качестве параметра задачи выступает переменная x . Множество $\{0, 1, \dots, 15\}$ составляет **пространство поиска** и одновременно — множество потенциальных решений задачи.

Каждое из 16 чисел из этого множества называется **точкой пространства поиска**, **решением**, **значением параметра**, **фенотипом**. Решение, оптимизирующее функцию, называется **наилучшим** или **оптимальным** решением.

Значения параметра x можно закодировать следующим образом:

0000	0001	0010	0011	0101	0110	0111
1000	1001	1011	1100	1101	1110	1111

Эти последовательности также называются **цепями** или **хромосомами**. В данном примере они выступают в роли **генотипов**.

Каждая из хромосом состоит из 4 генов. Значение гена в конкретной позиции называется **аллелью**, принимающей в данном случае значения 0 или 1.

Популяция состоит из особей, выбираемых среди указанных 16 хромосом. Примером популяции с численностью, равной 6 может быть, например, множество хромосом [0010, 0101, 0111, 1001, 1100, 1110], представляющих собой закодированную форму следующих **фенотипов**: [2, 5, 7, 9, 12, 14]

Функция приспособленности в этом примере задаётся выражением:

$$f(x) = 2x^2 + 1$$

Приспособленность отдельных хромосом в популяции определяется значением этой функции для значений x , соответствующих этим хромосомам, т.е. для **фенотипов**, соответствующих определённым **генотипам**.

Классический ГА

Основной ГА состоит из следующих шагов:

- инициализация, или выбор исходной популяции хромосом
- оценка приспособленности хромосом в популяции
- проверка условия остановки алгоритма
- селекция хромосом
- применение генетических операторов
- формирование новой популяции
- выбор «наилучшей» хромосомы

Селекция хромосом заключается в выборе тех хромосом, которые будут участвовать в создании потомков для следующей популяции. Такой выбор производится согласно принципу, по которому наибольшие шансы на участие в создании новых особей имеют хромосомы с **наибольшими значениями функции приспособленности**.

Существуют различные **методы селекции**:

Наиболее популярным считается т.н. «**метод рулетки**» (roulette wheel selection), когда колесо разбивается на количество секторов, равное количеству особей, а размер сектора определяется величиной значения функции приспособленности, т.е. хромосоме с наибольшим значением соответствует наибольший сектор.

Селекция хромосомы представляется как поворот колеса рулетки. Очевидно, что чем больше сектор, тем больше шансов его выпадения, т.е. в контексте селекции: больше шансов на выбор соответствующей хромосомы.

При практической реализации колесо представляется в виде цифрового интервала $[0, 100]$, а секторам соответствуют интервалы определённых длин внутри него. Тогда выбор хромосомы заключается в выборе случайного числа в диапазоне $[0, 100]$.

В результате селекции создаётся **родительская популяция** или **родительский пул** (mating pool) с численностью N .

Применение генетических операторов

Применение **генетических операторов** к хромосомам, отобранным с помощью селекции, приводит к формированию **новой популяции** от созданной на предыдущем шаге родительской популяции.

В классическом ГА применяются **два основных оператора**:

- оператор скрещивания (crossover)
- оператор мутации (mutation)

Однако стоит отметить, что оператор мутации играет второстепенную роль по сравнению с оператором скрещивания.

Применение генетических операторов

Применение **генетических операторов** к хромосомам, отобранным с помощью селекции, приводит к формированию **новой популяции** от созданной на предыдущем шаге родительской популяции.

В классическом ГА применяются **два основных оператора**:

- оператор скрещивания (crossover)
- оператор мутации (mutation)

Однако стоит отметить, что оператор мутации играет второстепенную роль по сравнению с оператором скрещивания.

Вероятность скрещивания достаточно велика: обычно $0 < p_c < 1$, тогда как вероятность мутации устанавливается весьма малой: обычно $0 < p_m < 0.1$. Это следует из аналогии с миром живых организмов. Мутация может выполняться как на популяции родителей перед скрещиванием, так и на популяции потомков.

Оператор скрещивания

- Выбираются пары хромосом из родительской популяции. Это выполняется случайным образом в соответствии с вероятностью скрещивания p_c
- Далее для каждой пары отобранных неким образом родителей разыгрывается **позиция гена** (локус) в хромосоме, определяющая так называемую **точку скрещивания**. Если длина хромосомы L , то фиксация точки скрещивания сводится к выбору числа l_k из интервала $[1, L-1]$
- В результате скрещивания пары родительских хромосом получается следующая пара потомков:
 - Потомок, хромосома которого на позициях от 1 до l_k состоит из генов первого родителя, а на позициях от l_k до L - из генов второго родителя.
 - Потомок, хромосома которого на позициях от 1 до l_k состоит из генов второго родителя, а на позициях от l_k до L - из генов первого родителя.

Оператор мутации

с вероятностью p_m изменяет состояние гена в хромосоме на противоположное (с 0 на 1 и обратно).

Например, если в хромосоме [100110101010] мутации подвергнется ген на позиции 7, то его значение, равное 1 изменится на 0, что приведёт к образованию хромосомы [100110001010].

Формирование новой популяции

заключается во включении в состав **новой популяции** хромосом, полученных в результате применения **генетических операторов**.

На каждой очередной итерации рассчитываются значения **функции приспособленности** для всех хромосом текущей популяции и проверяется условие остановки алгоритма.

В **классическом ГА** вся предшествующая популяция хромосом замещается новой популяцией, имеющей ту же численность.