

## Лекция 3

### Тема 3.2 Генетические алгоритмы

Содержание:

1. Основные понятия генетических алгоритмов.
2. Генетические операторы.

Эволюционные изменения признаков и свойств организмов обусловлены изменением генотипов, поэтому понимание основных генетических процессов, происходящих в популяции, необходимо для всей современной эволюционной теории. Успехи генетики позволили классифицировать и изучить основные формы наследственной изменчивости, а также выяснить значение изменчивости разных типов для протекания эволюционного процесса. Изучение роли разнообразных проявлений изменчивости в эволюции позволило сформулировать представление о генетических основах эволюции. В основе изменчивости как всеобщего явления живой природы лежит ковариантная редупликация. Именно в процессе матричного копирования ДНК и РНК происходят ошибки последовательности расположения нуклеотидов, связанные с заменой одного нуклеотида другим или сдвигом рамок считывания. Эти процессы лежат в основе изменчивости на молекулярном уровне. Общие причины генетической изменчивости чрезвычайно разнообразны. Генетические алгоритмы возникли в результате наблюдения и попыток копирования естественных процессов, происходящих в мире живых организмов, в частности, эволюции, связанной с ней селекцией (естественного отбора) популяции живых существ. Идея генетических алгоритмов была высказана в конце шестидесятых – начале семидесятых годов XX века. Она была основана на желании составить и реализовать в виде компьютерной программы алгоритм, который будет решать сложные задачи так, как это делает природа – путем эволюции.

Современная библиография по генетическим алгоритмам давно перевалила за 9000 наименований и продолжает непрерывно увеличиваться. Однако, несмотря на такое обилие литературы, довольно трудно точно сформулировать, чем именно они являются – квинтэссенцией эволюционных перестроек в природных популяциях организмов, универсальным средством описания адаптаций в популяциях искусственных объектов, или мощной поисковой процедурой с претензиями на решение задач глобальной оптимизации.

Целью данной работы является изучение генетических алгоритмов как способа оптимизации, их эффективности и трудоемкости. В качестве решаемой задачи была выбрана задача коммивояжера, поскольку она очень хорошо изучена, имеет эффективные способы решения, для того, чтобы сравнить с полученными результатами. Также одной из целей данной работы является изучение распространения генетических алгоритмов на модель с несколькими взаимодействующими популяциями.

Согласованность и эффективность работы элементов биологических организмов наводит на мысль о возможности использования принципов биологической эволюции с целью оптимизации практически важных для человека систем.

В 1975 г. вышла основополагающая книга Дж. Холланда (Holland) —Адаптация в естественных и искусственных системах, в которой был предложен генетический алгоритм – алгоритм, основанный на принципах естественного отбора Ч. Дарвина. Генетические алгоритмы относят к области мягких вычислений. Понятие «мягких вычислений» (Лофти Заде, 1994 г.) подразумевает под собой совокупность неточных, приближенных методов решения задач, зачастую не имеющих решение за полиномиальное время. Такие задачи возникают в области биологии, медицины, гуманитарных наук, менеджменте. Методы мягких вычислений хорошо дополняют друг друга, и часто используются совместно. В область мягких вычислений входят такие методы как:

- нечеткая логика;
- нейронные сети;
- вероятностные рассуждения;
- байесовские сети доверия<sup>1</sup>;
- эволюционные алгоритмы.

Генетический алгоритм представляет собой метод, отражающей естественную эволюцию методов решения проблем, и в первую очередь задач оптимизации. Генетические алгоритмы – это процедуры поиска, основанные на механизмах естественного отбора и наследования. В них используется эволюционный принцип выживания наиболее приспособленных особей. Они отличаются от традиционных методов оптимизации несколькими базовыми элементами. В частности, генетические алгоритмы обладают рядом отличительных свойств:

1. кодирование параметров – генетические алгоритмы обрабатывают не значения параметров самой задачи, а их закодированную форму;
2. операции на популяции – генетические алгоритмы осуществляют поиск решения исходя не из единственной точки (начальное приближение), а из некоторой популяции;
3. использование минимума информации о функции – генетические алгоритмы используют только целевую функцию, а не производные либо иную дополнительную информацию;
4. рандомизация операций – генетические алгоритмы применяют вероятностные, а не детерминированные правила выбора.

Перечисленные свойства приводят в результате к устойчивости генетических алгоритмов.

Сфера применения генетических алгоритмов – это в основном оптимизация многопараметрических функций. Прикладное же применение

генетических алгоритмов весьма обширно. Они применяются при разработке программного обеспечения в системах искусственного интеллекта, оптимизации, искусственных нейронных сетях и в других отраслях знаний. Следует отметить, что с их помощью решаются задачи, для которых ранее использовались нейронные сети. В этом случае генетические алгоритмы выступают просто в роли независимого от нейронных сетей метода, предназначенного для решения той же самой задачи. Примером может служить задача коммивояжера, изначально решавшаяся при помощи сети Хопфилда. Генетические алгоритмы часто используются совместно с нейронными сетями. Они могут поддерживать нейронные сети или наоборот, либо оба метода взаимодействуют в рамках одной гибридной системы, предназначенной для решения конкретной задачи. Генетические алгоритмы так же применяются совместно с нечеткими системами.

Но не стоит расценивать генетические алгоритмы как своеобразную панацею для задач оптимизации. С большой вероятностью генетические алгоритмы покажут, как минимум не лучшие результаты по сравнению со специально разработанными методами для решения специализированных задач. Большой плюс эволюционных вычислений в предоставляемом ими унифицированном подходе к решению самых различных проблем.

Генетические алгоритмы показывают блестящие результаты при решении сложных переборных задач (большинство из которых NP-полные, т.е. не решаются полным перебором за полиномиальное время), таких, например, как задача коммивояжера и поиск булевых термов.

При описании генетических алгоритмов используются определения, заимствованные из генетики в упрощенном виде и основные понятия линейной алгебры.

Вектор – упорядоченный набор чисел, называемых компонентами вектора.

Булев вектор – вектор, компоненты которого принимают значения из двух элементного (булева) множества значений, как правило  $\{0, 1\}$ .

Популяция – конечное множество особей.

Особь (индивидуум, организм) – набор хромосом с закодированными в них множествами параметров задачи, то есть решений, которые иначе называются точками в пространстве поиска (search points).

Хромосома (цепочка или кодовая последовательность) – это вектор генов. Хромосома может быть представлена в виде булева вектора, полученного с помощью двоичного кодирования либо кода Грея (см. приложение А). Хромосома обозначается, как правило  $A$ .

Ген (свойство, знак или детектор) – атомарный элемент генотипа, в частности, хромосомы. Несет в себе наследственную информацию. Обозначается  $x$ .

Связь хромосомы и гена изображена на рисунке 1.

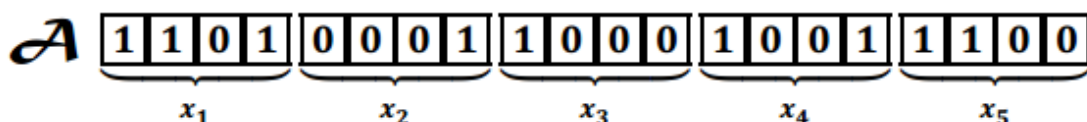


Рисунок 1 – Распределение наследственной информации по длине хромосомы.

*Генотип* (структура) – набор хромосом данной особи. Следовательно, особями популяции могут быть генотипы либо единичные хромосомы (в самом простом случае, генотип состоит из одной хромосомы).

*Фенотип* – набор значений, соответствующих данному генотипу, то есть это декодированная структура или множество параметров задачи (решение, точка пространства поиска).

*Аллель* – значение конкретного гена.

*Локус* – позиция, указывающая место размещения данного гена в хромосоме.

Множество позиций генов – *локи*.

Очень важным понятием в генетических алгоритмах считается *функция приспособленности* (fitness function), иначе называемая *функция оценки*. Эта функция играет важнейшую роль, поскольку позволяет определить степень приспособленности конкретных особей в популяции и выбрать из них наиболее приспособленные (то есть имеющие наибольшие значения функции приспособленности) в соответствии с эволюционным принципом выживания сильнейших. Функция приспособленности так же получила свое название из генетики.

В задачах оптимизации функция приспособленности, как правило, оптимизируется, (точнее сказать, максимизируется) и называется целевой функцией. В задачах максимизации целевая функция преобразуется, и проблема сводится к максимизации. В теории управления функция приспособленности может принимать вид функции погрешности, а в теории игр – стоимостной функции. На каждой итерации генетического алгоритма приспособленность каждой особи данной популяции оценивается при помощи функции приспособленности, и на этой основе создается следующая популяция особей, составляющих множество потенциальных решений проблемы.

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

## 2. Генетические операторы

Генетические операторы необходимы для того, чтобы применить принципы наследственности и изменчивости к популяции. Помимо отличительных особенностей, о которых будет рассказано ниже, для всех операторов определено такое свойство как *вероятность*. То есть описываемые операторы не обязательно применяются ко всем скрещиваемым особям, что вносит дополнительный элемент неопределенности в процесс поиска решения. В данном случае, неопределенность не подразумевает негативный

фактор, а является своеобразной "степенью свободы" работы генетического алгоритма.

Оператор кроссинговера (кроссовера, скрещивания, рекомбинации) – оператор, при котором хромосомы обмениваются своими частями. Моделирует процесс скрещивания особей.

Пусть имеются две родительские особи с хромосомами **A** и **B**.  
Случайным

образом определяется точка внутри хромосомы, в которой обе хромосомы делятся на две части и обмениваются ими. Назовем эту точку *точкой кроссинговера (crossover point)*, иногда она так же называется точкой разрыва. Описанный процесс изображен на рис. 2.

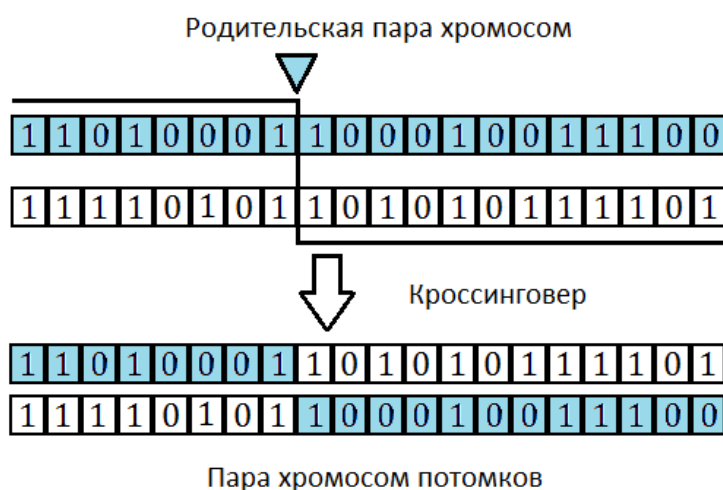


Рисунок 2 – Одноточечный кроссинговер.

Данный тип кроссинговера называется *одноточечным (single-point crossover)*, т.к. при нем родительские хромосомы «разрезаются» только в одной случайной точке.

В *двухточечном (и многоточечном кроссинговере (multi-point crossover) вообще) кроссинговере* хромосомы рассматриваются как циклы, которые формируются соединением концов линейной хромосомы. Для замены

сегмента одного цикла сегментом другого цикла требуется выбор двух точек разрыва. В таком представлении, одноточечный кроссинговер может быть рассмотрен как двухточечный, но с одной точкой разреза, зафиксированной в начале строки. Следовательно, двухточечный кроссинговер решает ту же самую задачу, но более полно. В настоящий момент исследователи соглашаются, что двухточечный кроссинговер лучше, чем одноточечный. Кроме описанных типов кроссинговера так же существует *однородный кроссинговер (uniform crossover)*. Его особенность заключается в том, что значение каждого бита в хромосоме потомка определяется случайным образом из соответствующих битов родителей. Для этого вводится некоторая величина  $0 < p < 1$ , и если случайное число больше  $p$ , то на  $i$ -ую позицию первого потомка попадает  $i$ -ый бит первого родителя, а на  $i$ -ую позицию второго –  $i$ -ый бит второго родителя. В противном случае к первому потомку попадает бит второго родителя, а ко второму – первого. Такая операция проводится для всех битов хромосомы. При использовании однородного кроссинговера родителей может быть больше двух. В случае если используется три особи, то такой кроссинговер называется триадным (triadic crossover).

Вероятность кроссинговера самая высокая среди генетических операторов и равна обычно 60% и больше.

Немаловажным при реализации генетических алгоритмов является стратегия выбора родительской пары:

- Панмиксия (panmixia) - когда обе особи, которые составят родительскую пару, случайным образом выбираются из всей популяции, причем любая особь может стать членом нескольких пар.
- Инбридинг (inbreeding) - первый член пары выбирается равномерно случайно, а вторым с большей вероятностью будет максимально близкая к нему особь. Под

- «родством» здесь понимается расстояние между членами популяции в смысле геометрического расстояния особей в пространстве параметров (фенотипов), например, для векторов можно использовать евклидово расстояние.
- Аутбридинг (outbreeding) или кроссбридинг - формирует родительские пары из максимально далеких особей.