

**1. Обучение, обучающие выборки**

**Обучение** (в психологии) – усвоение новых знаний.

Новые умения и навыки приобретаются путем тренировки и упражнения.

**Обучение** (в работах по ИИ): «любое изменение в системе, приводящее к улучшению решения задачи при ее повторном предъявлении или к решению другой задачи на основе тех же данных» - Н. Саймон.

**Принцип полноты базовых знаний.** Возможность/невозможность «обучения с нуля».

**Проблемы полноты и репрезентативности обучающей выборки** (при пополнении базы знаний).

**Некоторые важные термины**

**Генеральная совокупность** – вся изучаемая выборочным методом статистическая совокупность объектов и/или явлений, имеющих общие качественные признаки или количественные переменные

**Выборка (выборочная совокупность)** – часть объектов из генеральной совокупности, отобранных для изучения, с тем чтобы сделать заключение о всей генеральной совокупности. Для того, чтобы заключение, полученное путем изучения выборки, можно было распространить на всю генеральную совокупность выборка должна обладать свойством репрезентативности.

**Репрезентативность (представительность)** – свойство выборки отражать характеристики изучаемой генеральной совокупности.

**Репрезентативная выборка** – выборка, имеющая такое же распределение относительных характеристик, что и генеральная совокупность.

**Ошибки выборки** – отклонение статистической структуры выборки от структуры соответствующей генеральной совокупности.

**Произвольная выборка** – эмпирическая выборка, не имеющая вероятностного обоснования, складывающаяся на основе случая, причем выбор каждого случая не влияет на любой другой случай.

Источник – [www.glossary.ru](http://www.glossary.ru).

**2. Символьное обучение (обучение в ПРОСТРАНСТВЕ ПОНЯТИЙ)**

Основные операции в пространстве понятий: *обобщение, специализация.*

**Индуктивное обучение как поиск в пространстве понятий**

**Пример:** Пусть в пространство понятий входит некоторое абстрактное понятие:

**Object (Sizes, Colors, Shapes)**

и известно, что его признаки принимают такие значения:

**Sizes = {large, small}; Colors = {red, blue, white, green}; Shapes = {round, polygon}**

Индуктивное обучение в этом пространстве – поиск понятия, удовлетворяющего всем примерам обучающей выборки.

Пусть, далее, у нас есть единственный обучающий пример – **Object1 (small, red, round).**

Результатом обучения может стать пополнение пространства понятий таким новым частным случаем («маленький красный шар»/«маленький красный мяч» и т.п.). Это – *специализация.*

Пусть появляется второй обучающий пример – **Object2 (large, red, round).**

Результатом обучения может стать пополнение пространства понятий этим новым частным случаем («большой красный шар»/«большой красный мяч» и т.п.). Это тоже – *специализация.*

Можно выполнить и некоторые операции обобщения, построив и добавив в общее пространство новые понятия:

**Object3 (X, red, round)** – («красный шар»/«красный мяч» и т.п.)

**Object4 (X, Y, round)** – («шар»/«мяч» и т.п.)

**Основные операции обобщения**

1. Замена конкретного значения понятия на переменную:

**Colors (X, red) & Shapes (X, cube) → Colors (X, Y) & Shapes (X, cube)**

(«красный куб») → («куб <любого цвета>»)

2. Исключение конъюнкта:

**Sizes (X, small) & Colors (X, red) & Shapes (X, cube) → Colors (X, red) & Shapes (X, cube)**

(«красный куб малого размера») → («красный куб»)

3. Добавление дизъюнкта:

**Colors (X, red) & Shapes (X, cube) → Colors (X, red) & ((Shapes (X, cube) V Shapes (X, pyramid))**

(«красный куб») → («красный куб или красная пирамида»)

4. Замена конкретного объекта или частного понятия общим понятием (на основе иерархии классов):

**Colors (X, red) → Colors (X, rainbow-color)** («красный» → «цвета радуги»)

**Shapes (X, polyhedron) → Shapes (X, solid)** («многогранник» → «геометрическое тело»)

Роль *отрицательных примеров* в предотвращении излишнего обобщения.

В последние годы определенную популярность в работах по ИИ получил подход к моделированию процессов обучения/развития на основе так называемых *генетических алгоритмов*.

### 3. Понятие о генетических алгоритмах

(Использован материал бывшего доступным в 2002 году сайта:

<http://www.ai.tsi.lv/ru/index.htm>, Борисов А.Н. Курс: Генетические алгоритмы, 2002)

**Генетические алгоритмы** (ГА) - это стохастические, эвристические оптимизационные методы, впервые предложенные Джоном Генри Холландом в книге «Адаптация в естественных и искусственных системах» (1975). Они основываются на идее эволюции с помощью естественного отбора, выдвинутой Дарвином.

ГА работают с совокупностью "**особей**" - популяцией, каждая из которых представляет возможное решение данной проблемы. Каждая особь оценивается мерой ее "**приспособленности**" согласно тому, насколько "хорошо" соответствующее ей решение задачи. В природе это эквивалентно оценке того, насколько эффективен организм при конкуренции за ресурсы. Наиболее приспособленные особи получают возможность "воспроизводить" потомство с помощью "перекрестного скрещивания" с другими особями популяции. Это приводит к появлению новых особей, которые сочетают в себе некоторые характеристики, наследуемые ими от родителей. Наименее приспособленные особи с меньшей вероятностью смогут воспроизвести потомков, так что те свойства, которыми они обладали, будут постепенно исчезать из популяции в процессе эволюции. Иногда происходят мутации, или спонтанные изменения в генах.

Таким образом, из поколения в поколение, хорошие характеристики распространяются по всей популяции. Скрещивание наиболее приспособленных особей приводит к тому, что исследуются наиболее перспективные участки пространства поиска. В конечном итоге популяция будет сходиться к оптимальному решению задачи. Преимущество ГА состоит в том, что он находит приблизительные оптимальные решения за относительно короткое время.

ГА состоит из следующих компонентов: 1) **Хромосома** (Решение рассматриваемой проблемы. Состоит из генов); 2) **Начальная популяция** хромосом; 3) **Набор операторов** для генерации новых решений из предыдущей популяции; 4) **Целевая функция** для оценки приспособленности (fitness) решений.

Чтобы применять ГА к задаче, сначала выбирается метод кодирования решений в виде строки. Фиксированная длина ( $l$ -бит) двоичной кодировки означает, что любая из  $2^l$  возможных бинарных строк представляет возможное решение задачи.

Стандартные операторы для всех типов генетических алгоритмов это: *селекция, скрещивание и мутация*.

#### Селекция

Оператор селекции (reproduction, selection) осуществляет отбор хромосом в соответствии со значениями их функции приспособленности. Существуют как минимум два популярных типа оператора селекции: рулетка и турнир.

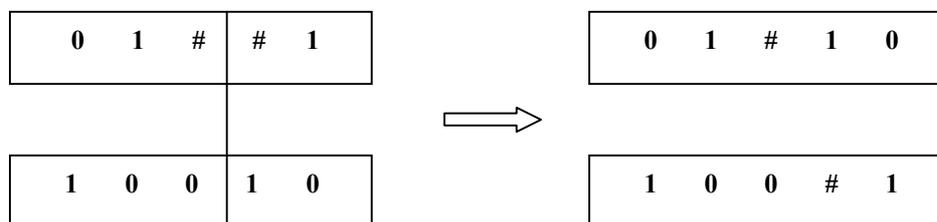
- **Метод рулетки** (roulette-wheel selection) - отбирает особей с помощью  $n$  "запусков" рулетки. Колесо рулетки содержит по одному сектору для каждого члена популяции. Размер  $i$ -ого сектора пропорционален некоторой величине вычисляемой по формуле.

При таком отборе члены популяции с более высокой приспособленностью с большей вероятностью будут чаще выбираться, чем особи с низкой приспособленностью.

- **Турнирный отбор** (tournament selection) реализует  $n$  турниров, чтобы выбрать  $n$  особей. Каждый турнир построен на выборке  $k$  элементов из популяции, и выбора лучшей особи среди них. Наиболее распространен турнирный отбор с  $k=2$ .

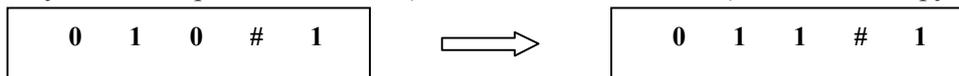
#### Скрещивание

Оператор скрещивания (crossover) осуществляет обмен частями хромосом между двумя (может быть и больше) хромосомами в популяции. Может быть одноточечным или многоточечным. Одноточечный кроссовер работает следующим образом. Сначала, случайным образом выбирается одна из  $l-1$  точек разрыва. Точка разрыва - участок между соседними битами в строке. Обе родительские структуры разрываются на два сегмента по этой точке. Затем, соответствующие сегменты различных родителей склеиваются и получают два генотипа потомков.



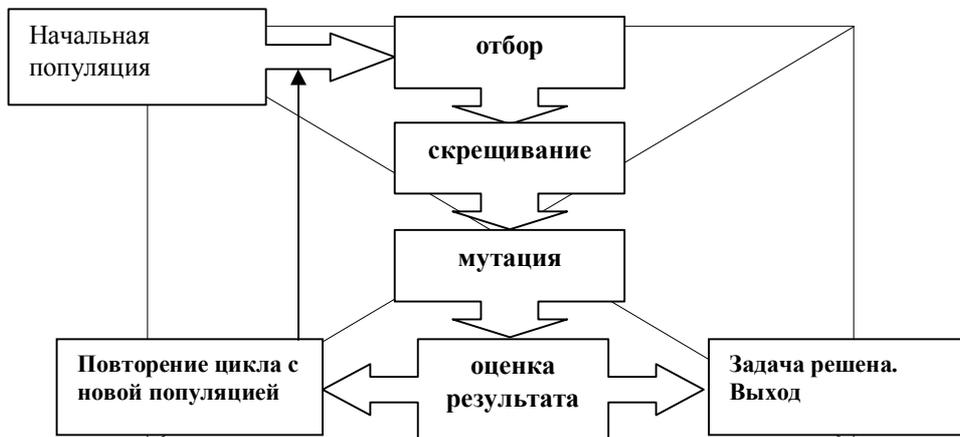
**Мутация**

Мутация (mutation) - стохастическое изменение части хромосом. Каждый ген строки, которая подвергается мутации, с вероятностью  $P_{mut}$  (обычно очень маленькой) меняется на другой ген.

**Схема работы ГА**

Работа ГА представляет собой итерационный процесс, который продолжается до тех пор, пока не выполнятся заданное число поколений или какой-либо иной критерий останова. На каждом поколении ГА реализуется отбор пропорционально приспособленности, кроссовер и мутация.

Схема работы простого ГА выглядит следующим образом:



Критерии останова алгоритма:

- нахождение глобального, либо субоптимального решения;
- исчерпание числа поколений, отпущенных на эволюцию;
- исчерпание времени, отпущенного на эволюцию.

Генетические алгоритмы служат, главным образом, для поиска решений в очень больших, сложных пространствах поиска. Примеры реальных областей применения:

- оптимизация запросов в базах данных;
- задачи на графах (задача коммивояжера, раскраска);
- задачи компоновки;
- составление расписаний.

**Замечание:** Генетические алгоритмы и «метод проб и ошибок».