**Course 7: Nature-Inspired Search**

**Mechanism of Evolution:**

The mechanism of evolution is a fundamental concept that underlies EAs. It involves the processes of selection, variation, and heredity. In nature, organisms with favorable traits have a higher chance of survival and reproduction, leading to the propagation of those traits in subsequent generations. Similarly, in EAs, the fittest individuals from a population are selected for reproduction, undergo variation through genetic operators (mutation and recombination), and pass on their genetic material to the next generation.

**Simulation of Nature:**

EAs simulate the natural evolution process to solve optimization problems. Instead of directly solving a problem analytically, EAs explore a population of potential solutions and iteratively improve them over generations. By emulating the principles of natural evolution, EAs can effectively explore the search space and converge towards optimal or near-optimal solutions.

**Evolution Algorithms:**

**a) Main Characteristics of EAs:**

EAs possess several key characteristics:

* Population-based: EAs maintain a population of individuals (also called chromosomes or solutions) instead of focusing on a single solution. This allows for exploration of multiple potential solutions simultaneously.
* Iterative improvement: EAs iteratively evolve the population through genetic operators to gradually improve the quality of solutions over generations.
* Stochastic search: EAs employ randomization to explore the search space, balancing exploration and exploitation.
* Adaptive: EAs dynamically adjust the search process based on the performance of individuals in the population.

**b) Design Elements of EAs:**

1. **Representation:**

In evolution algorithms (EAs), the representation of individuals (also known as chromosomes or solutions) is a crucial aspect. It determines how potential solutions are encoded and manipulated within the algorithm. The representation can be classified into external and internal levels and can take various forms, including linear (discrete and continuous) and tree-based representations.

* **External Level:** At the external level, the representation refers to how the solution is perceived and interacted with by the external environment. It is the form in which the solution is presented or communicated outside of the algorithm. The external level representation should be interpretable and meaningful to the problem domain.
* For example, in a traveling salesperson problem, the external representation could be a sequence of cities visited. This sequence can be easily understood and evaluated by humans.
* **Internal Level:** The internal level of representation pertains to how the solution is encoded and manipulated within the EA. It is the internal structure or data structure used to represent the solution. The internal level representation may not be directly interpretable by humans and is designed to facilitate the computational processes of the algorithm.

**a) Linear Representation:**

Linear representations involve encoding the solution as a sequence or vector of discrete or continuous values. Here are some examples:

* **Linear Discrete Binary Representation:** This representation encodes the solution using binary values (0s and 1s) within an array. Each position in the array corresponds to a specific gene or decision variable. For example, consider solving a knapsack problem where you have a set of items to choose from, and each item can either be selected (1) or not selected (0). An example binary representation could be [1, 0, 1, 1, 0], where the 1s represent the selected items.
* **Linear Discrete Non-Binary Representation:** In this representation, discrete values other than binary are used. For example, in a scheduling problem where tasks need to be assigned to different time slots, each gene in the chromosome can represent a specific time slot.
* **Linear Integer Representation:** This representation uses integer values to encode the solution. For example, in a job scheduling problem, each gene can represent the index of the machine to which a job is assigned.
* **Linear Random/Permutation Representation:** This representation involves representing the solution as a random or permuted order of elements. For example, in a permutation problem, such as the traveling salesperson problem, the chromosome can represent the sequence in which cities are visited.
* **Linear Class-Based Representation:** In some problems, the solution can be represented using classes or categories. For instance, in a classification problem, the chromosome can encode the class labels for different features.

**b) Tree-Based Representation:**

Tree-based representations employ hierarchical structures to encode the solution. These structures are particularly useful when the problem exhibits a hierarchical relationship between variables. For example, in a symbolic regression problem, where the goal is to find a mathematical expression that fits a given dataset, the chromosome can be represented as a tree structure, with each node representing an operation or variable.

1. **Population:**

The population in EAs refers to the collection of individuals (chromosomes) being explored simultaneously. The population evolves over generations through genetic operators such as selection, mutation, and recombination. Here are some key concepts related to the population:

* **Concept:** The population represents the set of potential solutions being evaluated and improved by the EA. It allows for exploration of multiple solutions concurrently.
* **Initialization:** The population is initialized at the beginning of the algorithm with a set of randomly generated or predefined individuals. The initial population should cover a diverse range of solutions to facilitate exploration of the search space.
* **Model:** The model of an EA determines how the population evolves over generations. Two common models are:
  + **Generational EA:** In this model, a new population is created in each generation. The parent population is completely replaced by the offspring population generated through selection, variation operators, and possibly elitism.
  + **Steady-state EA:** In this model, only a subset of the population is replaced in each generation. Typically, one or a few individuals are selected for replacement, while the rest of the population remains unchanged. This approach can maintain diversity in the population and facilitate the exploitation of promising solutions.

It is important to choose **an appropriate population size**, balancing the need for **exploration** and **exploitation**. A larger population size allows for more diverse exploration but increases computational costs, while a smaller population may converge prematurely or get trapped in local optima.

By **managing** the population through selection, variation operators, and population models, EAs can effectively explore the search space, discover promising solutions, and evolve towards better solutions over generations.

1. **Fitness Function:**

The fitness function is a crucial component of an evolution algorithm (EA) that evaluates the quality or fitness of an individual within a population. It represents the problem-specific criterion for measuring how well a particular solution solves the problem at hand.

* **Concept**: The fitness function assigns a numerical value, known as fitness or objective value, to each individual in the population. This value quantifies the performance or suitability of the solution based on the problem's objectives. The objective can be either to maximize or minimize, depending on the problem context.
* **Example:** Let's consider a simple example of a genetic algorithm applied to a knapsack problem. The knapsack problem involves selecting a subset of items to maximize the total value while staying within a weight constraint.
  + In this case, the fitness function would take an individual (chromosome) as input, which represents a potential solution or a set of selected items. The function would calculate the total value of the items in the individual's knapsack. If the total weight of the items exceeds the knapsack's capacity, the fitness would be penalized or set to a low value.
  + The fitness function's calculation could be as follows:
    - Iterate over each gene (item) in the individual's chromosome.
    - If the gene is set to 1 (indicating that the item is selected), add its value to a running total.
    - If the total weight exceeds the knapsack's capacity, set the fitness to a low value (e.g., 0).
    - Otherwise, set the fitness to the total value.
  + The goal of the algorithm would be to find individuals with higher fitness values, indicating a better selection of items that maximize the total value while respecting the weight constraint.

1. **Selection:**

Selection is the process of choosing individuals from the current population to serve as parents for generating offspring in the next generation. It plays a crucial role in driving the evolutionary process by favoring individuals with higher fitness values, which are more likely to contribute desirable characteristics to future generations.

* **Concept:** The concept of selection is inspired by the principle of "survival of the fittest" in nature. It aims to mimic natural selection by favoring individuals that are more adapted to their environment or have better fitness.
* **Properties(not important):** Selection operators in EAs typically exhibit the following properties:
  + Fitness Proportionate: The probability of an individual being selected as a parent is directly proportional to its fitness value. Individuals with higher fitness have a higher chance of being selected.
  + Elitism: Elitism involves preserving a certain number of the best individuals (based on their fitness) from one generation to the next without undergoing any modifications. This ensures that the best solutions discovered so far are carried forward.
  + Diversity Maintenance: Selection should also aim to maintain diversity in the population. This prevents premature convergence and facilitates exploration of different regions of the search space. Techniques like tournament selection or stochastic universal sampling can help maintain diversity.
* **Mechanism(important):** There are various selection mechanisms used in EAs, including:
  + Roulette Wheel Selection: Each individual's selection probability is proportional to its fitness value. A roulette wheel is imagined, with each individual's fitness value determining the size of its corresponding slice on the wheel. A random spin of the wheel determines the selected individuals.
  + Tournament Selection: Individuals are randomly selected in pairs (tournament size) from the population, and the one with the highest fitness value is chosen as a parent. This process is repeated until the desired number of parents is selected.
  + Rank-Based Selection: Individuals are ranked based on their fitness values, and the selection probability is assigned based on their rank. Higher-ranked individuals have a higher probability of being selected.
  + These selection mechanisms determine the individuals that contribute genetic material to the next generation, promoting the propagation of favorable characteristics and driving the evolution of the population.

1. **Recombination Selection:**

Recombination selection refers to the process of selecting individuals from the population for recombination or crossover, which combines their genetic material to create new offspring. It is a crucial step in evolving new solutions by exploiting the information encoded in the selected parents.

* **Concept:** Recombination, also known as crossover, involves taking genetic material from two or more parents and creating new offspring by exchanging or combining segments of their genetic representations. The goal is to produce offspring that inherit beneficial characteristics from their parents.
* **Mechanism:** Recombination selection is the mechanism by which individuals are chosen as parents for the recombination process. The selection process is typically biased towards individuals with higher fitness values, as they are more likely to contribute desirable traits to the offspring.
* **Implications:** Recombination selection involves selecting a subset of individuals from the population, either randomly or using specific selection mechanisms like tournament selection or rank-based selection. These selected individuals serve as parents for the recombination process.
* **Efficiency:** By choosing individuals with higher fitness values, recombination selection increases the likelihood of generating offspring with improved fitness and promoting the exploration of the search space for better solutions.

1. **Selection of recombination:**

* Selection for recombination refers to the process of selecting individuals from the population to serve as parents for generating offspring through recombination or crossover. There are several selection methods used in evolutionary algorithms, and here we will dive deeper into three common types: ranking selection, tournament selection, and survival selection.
* **Ranking selection** assigns a rank to each individual in the population based on their fitness values. The ranking can be done in ascending or descending order, depending on whether the objective is to maximize or minimize the fitness. Individuals with higher ranks are more likely to be selected as parents.
* One popular ranking selection method is the **Roulette Wheel Selection**, also known as **fitness proportionate selection**. It assigns selection probabilities to individuals based on their rank. The selection probability of an individual is determined by dividing its rank by the sum of all ranks. This way, individuals with higher ranks have a higher probability of being selected.

Example: Suppose we have a population of 10 individuals with their fitness values and corresponding ranks as follows:

Individual Fitness Rank

1 80 3

2 90 1

3 70 5

4 85 2

5 75 4

The selection probabilities for each individual would be:

Individual Fitness Rank Selection Probability

1 80 3 0.1

2 90 1 0.3

3 70 5 0.05

4 85 2 0.2

5 75 4 0.15

* **Tournament selection** involves randomly selecting a subset of individuals from the population (tournament size) and choosing the individual with the highest fitness as the parent. This process is repeated until the desired number of parents is selected.

Tournament selection provides a simple and efficient way to maintain selection pressure without requiring the entire population to be sorted or ranked. The tournament size determines the intensity of the selection pressure, where larger tournament sizes favor stronger individuals.

Example: Let's consider a population of 10 individuals and a tournament size of 3. The selection process would involve randomly selecting three individuals, comparing their fitness values, and selecting the one with the highest fitness as the parent. This process is repeated until the desired number of parents is chosen.

* **Survival selection** is the process of selecting individuals to survive and move on to the next generation after recombination and mutation. It ensures that the offspring, which inherit characteristics from their parents, replace less fit individuals in the population.
* Survival selection can be based on different strategies, such as elitism, where the best individuals from the current generation are automatically selected to survive. Another approach is age-based selection, where individuals compete for survival based on their age or how long they have been in the population.

1. **Variation operatiors:**

* Variation operators, including mutation and recombination, introduce genetic diversity into the population, allowing for exploration of new solutions beyond the scope of the existing individuals. These operators play a crucial role in driving the evolutionary process.

1. **Mutation:**

* **Aim:** Mutation is an operator that introduces small random changes or modifications to the genetic material (chromosomes) of individuals. Its aim is to promote exploration by adding new genetic information to the population.
* **How it works:** Mutation works by randomly altering the values or structure of genes within an individual's chromosome. The extent of the changes introduced by mutation is usually small, ensuring that the modified offspring are still similar to their parents.
* **Types of Mutation:**
* **Strong Mutation:** Strong mutation involves making significant changes to the genetic material. It can lead to drastic modifications in the offspring, potentially exploring entirely new regions of the search space.
* **Weak Mutation:** Weak mutation introduces minor changes to the genetic material, aiming to fine-tune the solutions rather than drastically altering them.
* **Random Resetting Mutation:** Random resetting mutation involves randomly selecting one or more genes and resetting their values to random values within the allowed range.
* **Creep Mutation:** Creep mutation involves gradually modifying the values of genes over generations by adding a small random value to them.
* **Swap Mutation:** Swap mutation selects two genes and swaps their values, altering the order or arrangement of genetic material.
* **Insertion Mutation:** Insertion mutation selects a gene and inserts it at a random position within the chromosome, shifting the neighboring genes.
* **Inversion Mutation:** Inversion mutation selects a segment of the chromosome and reverses the order of genes within that segment.
* **Scramble Mutation:** Scramble mutation selects a segment of the chromosome and randomly shuffles the order of genes within that segment.
* **K-opt Mutation:** K-opt mutation involves selecting K genes and reordering them in a different way, potentially introducing significant changes to the chromosome structure.
* **Uniform Mutation:** Uniform mutation randomly selects genes and modifies their values uniformly within the allowed range.
* **Non-Uniform Mutation:** Non-uniform mutation adjusts the magnitude of changes based on a specific function or probability distribution. It allows for more controlled exploration, gradually reducing the magnitude of changes over time.

1. **Recombination:**

* **Concept:** Recombination, also known as crossover, combines genetic material from two or more parents to create new offspring. It promotes the exchange and recombination of beneficial genetic information, leading to potentially improved solutions.
* **Types of Recombination:** Recombination can vary based on the type of representation used for the chromosomes. Here, we'll focus on binary and integer representation from two parent chromosomes:
* N-Cutting Point Crossover: In N-cutting point crossover, the chromosome is divided into N segments, and the corresponding segments from both parents are swapped to create offspring.
* Uniform Crossover: Uniform crossover involves randomly selecting bits from the parent chromosomes and exchanging them to create offspring.
* Order Crossover: Order crossover is primarily used for permutation representation. It selects a random segment from one parent and copies it to the offspring, then fills the remaining positions with the genes from the other parent in the order of appearance.
* Discrete Crossover: Discrete crossover combines genes from both parents based on certain rules or conditions, aiming to preserve specific patterns or structures present in the parents.
* Arithmetic Crossover: Arithmetic crossover is used for real-valued representation. It involves taking the weighted average of corresponding gene values from both parents to generate offspring.
* Singular Arithmetic Crossover: Singular arithmetic crossover considers only one offspring generated from the average of parent values.
* Simple Arithmetic Crossover: Simple arithmetic crossover generates two offspring by taking the average of parent values.
* Complete Arithmetic Crossover: Complete arithmetic crossover creates two offspring by taking the average of parent values, and then perturbs them using specific techniques.
* Geometric Crossover: Geometric crossover is another recombination operator used for real-valued representation. It combines genetic information from parents by taking the geometric mean of corresponding gene values.
* Blend Crossover: Blend crossover is a variation of arithmetic crossover that allows for more exploration. It involves randomly selecting a range between the parent values and creating offspring within that range.

1. **Recombination or mutation advantages and disadvantages:**

* Advantages of Recombination:
  + Promotes exploration by combining genetic material from different individuals.
  + Allows for the exchange of beneficial genetic information.
  + Maintains diversity in the population.
* Disadvantages of Recombination:
  + Alone, it may not explore the entire search space.
  + Can lead to the loss of good solutions if not properly preserved or recombined.
* Advantages of Mutation:
  + Introduces random changes, enabling exploration of new regions.
  + Maintains genetic diversity.
  + Fine-tunes existing solutions.
* Disadvantages of Mutation:
  + Random changes may not always be beneficial.
  + Slower in exploring the search space compared to recombination.
  + The choice and combination of recombination and mutation depend on the problem and search space characteristics, balancing exploration and exploitation for optimization.

1. **The stop condition:** The stop condition determines when the EA should terminate. It can be based on reaching a maximum number of generations, finding a satisfactory solution, or exceeding a certain computational limit.
2. **The evaluation of an EA's performance**:

It is typically based on various metrics, such as average of solutions, median of solutions, best solution, worst solution, standard deviation of solutions(for comparisons)