IB Case Study Vocabulary

Genetic algorithms (2022)

Websites

Genetic Algorithms Tutorial

Machine Learning and Simulated Annealing

Novelty Search

Population Initialization in Genetic Algorithms

Vocabulary

A typical **Genetic Algorithm** implementation would require

a) a genetic description of the solution space,

b) and a fitness function to evaluate it.

Brute force approach - a brute force algorithm **solves a problem through exhaustion**: it goes through all possible choices until a solution is found. The time complexity of a brute force algorithm is often proportional to the input size. Brute force algorithms are simple and consistent, but very slow.

Combinatorial optimization - seeks to improve an algorithm by using mathematical methods either to reduce the size of the set of possible solutions or to make the search itself faster.

Computational intractability – are problems for which there exist **no** efficient algorithms to solve them.

Convergence - refers to the **limit of a process** and can be a useful analytical tool when evaluating the expected performance of an optimization algorithm. An iterative algorithm is said to converge when, as the iterations proceed, the output gets closer and closer to some specific value.

Crossover / crossover operator - crossover is a genetic operator used to vary the programming of a chromosome or chromosomes from one generation to the next. In genetic algorithms and evolutionary computation, crossover, also called recombination, is a genetic operator used to combine the genetic information of two parents to generate new offspring.

Elitis – A strategy in evolutionary algorithms where the best one or more solutions, called the elites, in each generation, are inserted into the next, without undergoing any change. This strategy usually speeds up **the convergence of the algorithm**.

Exploration vs exploitation – exploration is the process of finding a new region searching for the best solution whereas exploitation is the process of updating solutions based on the best solution aiming to enhance existing ones.

Fitness / fitness function / fitness landscape – The phenotypes and genotypes favored by natural selection aren't necessarily just the ones that survive best. Instead, they're the ones with the highest overall fitness. **Fitness** is a measure of how well organisms survive and reproduce, with emphasis on "reproduce." Officially, fitness is defined as the number of offspring that organisms with a particular genotype or phenotype leave behind, on average, as compared to others in the population.

The ***fitness function*** is what is applied on the members of a population, or on the solution space to rank how close the member is to solving the original problem. Members are *selected* based on their fitness functions (more fit, more probability of getting selected), and their parameters or genes are made to *crossover* to generate an offspring.

Heuristic – Heuristics in computer science and artificial intelligence are “rules of thumb” used in algorithms to assist in finding approximate solutions to complex problems. Often, there’s simply too much data to sift through in order to come to a solution in a timely manner, so a heuristic algorithm is used to trade exactness for speed. However, because heuristics are based on individual rules unique to the problem they are solving, the specifics of the heuristics vary from problem to problem.

Heuristics aim to produce solutions in a *reasonable time frame* that are *good enough* for solving the problem at hand. The solution produced using a heuristic may not be the perfect or exact solution, but it’s valuable as an approximate or best-guess solution. Some problems would require hundreds of thousands of years for an exact answer, but we can product an approximate solution almost instantly.

Hill Climbing – Hill Climbing is a form of heuristic search algorithm which is used in solving optimization related problems in Artificial Intelligence domain. The algorithm starts with a non-optimal state and iteratively improves its state until some predefined condition is met. The condition to be met is based on the heuristic function. The aim of the algorithm is to reach an optimal state which is better than its current state. The starting point which is the non-optimal state is referred to as the base of the hill and it tries to constantly iterate (climb) until it reaches the peak value, that is why it is called Hill Climbing Algorithm.

Hill Climbing Algorithm is a memory-efficient way of solving large computational problems. It considers the current state and immediate neighboring state. The Hill Climbing Problem is particularly useful when we want to maximize or minimize any function based on the input which it is taking. The most commonly used Hill Climbing Algorithm is “Travelling Salesman” Problem” where we have to minimize the distance travelled by the salesman. Hill Climbing Algorithm may not find the global optimal (best possible) solution but it is good for finding local minima/maxima efficiently.

Initial parameters - The selection of Genetic Algorithm (GA) parameters (selection mechanism, crossover and mutation rate) are problem dependent. The objective of parameter selection is to use initial values for them that will optimize the code for the specific problems, so they produce the best results.

Selection of genetic algorithm parameters can be done by **performing a sensitivity study on the algorithm**. Perform the optimization study varying one parameter at a time keeping others constant. When you find the optimum value for that parameter you repeat the process for the rest of the parameters.

Local extrema – A local extremum of a function is the point at which a **maximum** or **minimum** value of the function in some open interval containing the point is obtained.

Mating pool - is a concept used in evolutionary computation, which refers to a family of algorithms used to solve optimization and search problems. The mating pool is formed by candidate solutions that the selection operators deem to have the highest fitness in the current population.

Mutation / mutation rate – A **mutation** is a change in a DNA sequence. In genetics, the **mutation rate** is the frequency of new mutations in a single gene or organism over time. Mutation rates are not constant and are not limited to a single type of mutation, therefore there are many different types of mutations. Mutation rates are given for specific classes of mutations.

Novelty search – novelty: the quality of being new, original, or unusual.

One of the biggest problems in evolutionary computing is the preservation of diversity: as far as evolutionary algorithms perform the optimization steps, individuals in the population tend to get closer to each other, in terms of genotype, thus preventing any further improvement. For instance, **Novelty Search** does so by measuring the "novelty" of each new individual, according to some predefined metric, and comparing it to an archive of solutions that were highly novel when they were first discovered.

Offspring – children or descendants that inherit traits from their parents.

Optimization - is the process of modifying a system to make some features of it work more efficiently or use fewer resources.

Population - A **population** is a group of organisms of the same species that are found in the same area and can interbreed. A population is the smallest unit that can evolve—in other words, an individual can't evolve.

Premature convergence - an evolutionary algorithm means convergence of algorithm before the global optimum solution is reached. Normally, when any evolutionary algorithm is trapped into local minima it is termed as premature convergence. When this occurs, the parental solutions are not able to generate offspring that are superior to their parents, through the aid of genetic operations.

Problem space or solution space - the *solution space* is the initial population. The *genetic description* involves representing each member of the population as a set of parameters with binary string encoding, called a **gene**.

Ranking - sorting a population according to fitness value.

Roulette wheel selection - Roulette wheel is the simplest selection approach. In this method all the chromosomes (individuals) in the population are placed on the roulette wheel according to their fitness value. Each individual is assigned a segment of roulette wheel. The size of each segment in the roulette wheel is proportional to the value of the fitness of the individual - the bigger the value is, the larger the segment is. Then, the virtual roulette wheel is spun. The individual corresponding to the segment on which roulette wheel stops are then selected. The process is repeated until the desired number of individuals is selected. Individuals with higher fitness have more probability of selection. This may lead to biased selection towards high fitness individuals. It can also possibly miss the best individuals of a population. There is no guarantee that good individuals will find their way into next generation. Roulette wheel selection uses exploitation technique in its approach.

Selection strategy – an approach used in the selection of chromosomes for reproduction.

Simulated annealing - Simulated annealing is a technique that is used to find the best solution for either a global minimum or maximum, without having to check every single possible solution that exists. This is extremely helpful when addressing massive optimization problems. Simulated annealing has its origins in metallurgy. In metallurgy, annealing refers to the process of heating metal to a high temperature and then slowly cooling it in a controlled environment.

Stochastic universal sampling – SUS is like Roulette wheel selection, but instead of having just one fixed point, it uses multiple fixed points. Therefore, all the parents are chosen in just one spin of the wheel. This setup encourages the highly fit individuals to be chosen at least once.

Termination condition - The termination condition of a Genetic Algorithm is important in determining when a GA run will end. Typically, you want a termination condition so that the solution is close to the optimal, at the end of the run.

Tour/Tour-based models **arrange travel into units called tours**. Tours are travel events that start at one location and return to that same location. For example, when a person travels to work and returns home, this a home-based work tour. To optimize a tour-based problem you must find the shortest route (path) from available routes. One well known tour-based problem is the Traveling Salesman Problem.

Tournament selection – a selection model where individuals are selected from the population at random and of these individuals the best one becomes a parent. The same process is repeated for selecting the next parent. Tournament selection is popular in literature as it can even work with negative fitness values.

Truncation selection - Truncation selection is a selection method used in genetic algorithms to select potential candidate solutions for recombination. In truncation selection the candidate solutions are ordered by fitness, and some proportion, p, (e.g. p = 1/2, 1/3, etc.), of the fittest individuals are selected and reproduced 1/p times. Truncation selection is less sophisticated than many other selection methods and is not often used in practice.

One exception is in animal and plant breeding, truncation selection is a standard method in **selective breeding** in selecting animals to be bred for the next generation. Animals are ranked by their phenotypic value on some trait such as milk production, and the top percentage is reproduced.