

유전자 알고리즘 (Genetic Algorithm)

존 홀랜드(John Holland)가 1975년에 저서 "Adaptation on Natural and Artificial Systems" 에서 처음 소개된 최적화 기법

Meta Heuristic (다양한 문제에 대해 경험적으로 솔루션을 제시하는 알고리즘)

소개

주어진 문제에 대해서 최적의 솔루션을 찾는 것.

뚜렷한 답을 찾기가 어려운데, 그 과정이나 해답에 대한 설명이 크게 필요하지 않으면서, 어느 정도 좋은 결과가 필요할 때

알고리즘 특성상 왜 도중에 이러한 값이 도출되었는지 설명하기가 굉장히 어렵다. 또한 대체적으로 좋은 결과를 내는 이유는 이 설명이 어려운 이유, 무작위성이 존재하기 때문이다. 이러한 무작위성은 알고리즘의 설명을 어렵게 하는 동시에 성능을 개선한다.

이런 유전 알고리즘의 장점과 단점은 모두 무작위성이다. 장점은 무작위성, 돌연 변이 등과 같은 요소로 대체로 좋은 솔루션을 제공하지만, 이 값이 왜 나왔는지에 대한 설명을 하기가 어렵고(대부분의 머신 러닝 알고리즘이 그렇지만!) 속도가 느리다.

적자생존

1. 환경에 적합한 개체가 살아남는다.
2. 살아남은 개체는 번식을 한다.
3. 살아남은 개체는 각각 부모가 되어서 자손을 생성하는데, 자손은 부모의 유전자를 받고, 그리고 때때로 돌연변이 유전자를 받아 다음 세대를 살아간다.

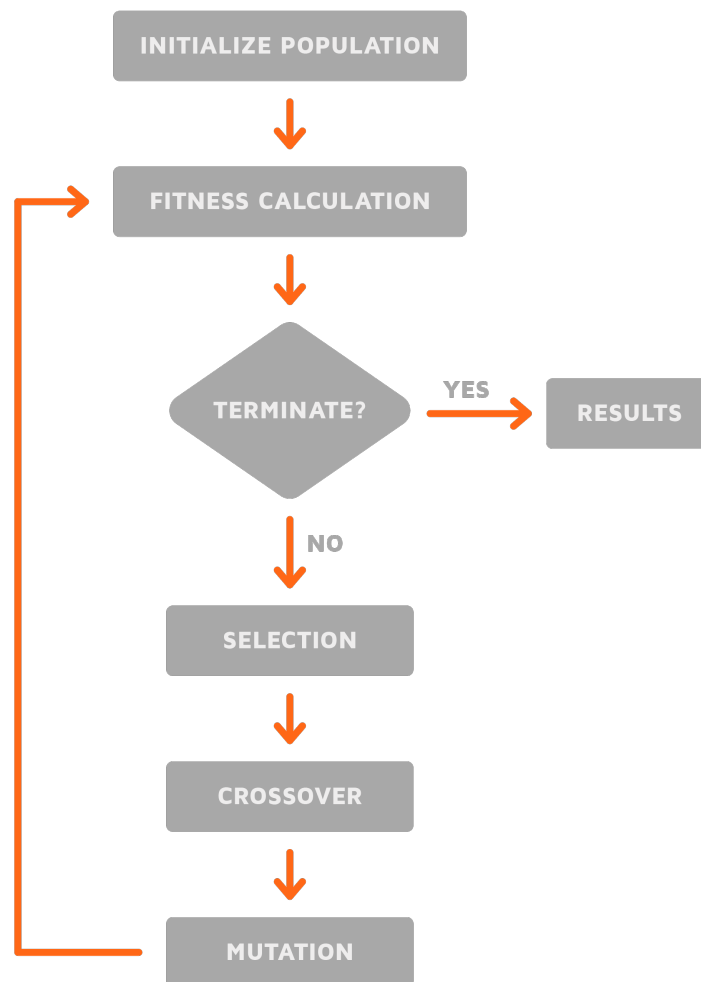
이 위의 세 가지를 반복적으로 수행하면 환경에 적합한, 즉 좋은 적합도 값을 내는 개체들이 살아남고 그 개체들이 좋은 솔루션을 제시한다.

용어

- 염색체(Chromosome): 생물학적으로 유전 물질을 담고 있는 하나의 집합. 염색체는 유전자를 포함
- 유전자(Gene): 염색체(chromosome)을 구성하는 하나의 유전 정보
- 자손(Offspring): 특정 시간에 존재한 부모 염색체의 성질을 이어받은 다음세대 염색체들을 부모 염색체의 자손 염색체 (자손 염색체는 부모 염색체와 비슷한 성질을 가짐)
- 적합도(Fitness): 각각의 염색체가 가지는 고유 값으로, 해당 문제의 해(목표값)에 얼마나 적합한지를 나타냄.
- 교차율 또는 교배 (Crossover): 2개의 염색체를 조합하여 새로운 염색체를 생성한다. (보통 교차율은 0.7 (70%)로 정의)

- 돌연변이(Mutation): 생물학적 돌연변이와 같은 것으로 특정 확률로 염색체 또는 유전자를 변형. 변이 없이 알고리즘을 진행하다보면 여러 세대가 지나고 같은 데이터를 가진 염색체만 남는 경우가 있다. 변이를 통해 새로운 데이터에 대한 가능성을 열어둔다. (보통 변이율은 0.001 (0.1%)로 정의)

절차



1. Fitness Function

환경에 적합하다는 것은 주어진 문제에 보다 좋은 솔루션을 제시. 이를 위해서는 적절한 평가를 위한 함수를 정의할 필요 (적합도 함수, fitness function)

2. Selection

- tournament: 전체 individual의 풀에서 샘플링을 해서 그중에서 적합도가 높은 individual을 부, 모로 선택하고, 필요한 부, 모의 개수를 선택할 만큼 반복해서 이를 시행한다.
- truncation : 특정 상위 몇 프로를 선택하고, 나머지를 제외
- fitness proportional selection : 함수의 적합도 값을 sampling될 확률 값으로서 변환

하여 적합도가 높으면 쉽게 선택되고 그렇지 않으면 선택되지 않도록 한다.

- elitism : 3단계인 번식 과정을 거치지 않고(자손 생성) 우수한 individual이 그냥 그다음 단계로 진출한다.

3. Crossover & Mutation

- **Crossover:** 부와 모로부터 chromosome을 섞어서 만듦
 - 반반씩 섞는 방법,
 - 보다 다채롭게 섞는 방법,
 - 완전히 랜덤으로 부와 모로부터 chromosome을 선택하는 방법 등
- **Mutation:** 일정 확률로 돌연 변이를 만들어주는데, 이 돌연 변이들이 유전 알고리즘의 정체성을 나타내면서 알고리즘의 성능을 강화. 돌연변이가 필요한 이유는 지역 최적값에 골몰될 수도 있는 가능성을 넘어서, 우연히 더 나은 솔루션을 찾을 수 있는 기회를 제공해주기 때문