

새로운 최적화 알고리즘 - 유전 알고리즘

정재학

영남대학교 화학공학과

Genetic algorithm as a new optimization algorithm

Jae Hak Jung,

Dept. of Chem. Eng., Yeungnam University

서론

최적화 문제는 그 문제의 특성에 따라 해를 효과적으로 찾아내기 어려운 경우를 자주 찾아볼 수 있다. 최근 과학 및 공학적 문제가 풀이하기 난해한 소위 NP-complete로 규정지워지는 경우가 자주 나타나며 이들은 주로 미분 불가능점을 다수 포함하는 noisy 형태 혹은 combinatorial 형태, 이산 시스템의 최적화 문제 및 혼합정수형, 혼합 이진수형 문제이다. computer system의 발달로 이러한 문제들을 풀이하여 최적 혹은 최적에 매우 근사한 준 최적치를 얻어내는 방법론이 최근 많이 연구 개발되어지고 있으며 이들중 백미에 해당한다고 할 수 있는 방법이 montecarlotic한 randomized technique 들이며 유전 알고리즘은 그들 중 하나이다.

유전 알고리즘은 1970년대 초 Michigan 대학에서 John Holland에 의해 최초로 연구되어 졌으며, 세계적으로 인정받고 주목받기 시작한 것은 1985년도 제1회 ICGA(International Conference on Genetic Algorithms and their Application)에서 많은 논문이 발표되면서부터라 할 수 있다. 유전 알고리즘은 자연 생태계의 진화 과정에서 부모세대가 그들의 자손세대에 의해 대체되어 나가는 과정에서 관찰, 연구된 몇가지 처리과정 중 적자생존(Survaval of fittest)의 원리와 자연계의 유전 법칙인 부모의 교배에 의한 자연 선택(natural selection)의 원리를 알고리즘으로 표현하여 학습에 의한 경험 축적(machine learning) 혹은 최적화(optimization)를 수행하는 컴퓨터 응용방법이다.

유전 알고리즘이 모방한 자연 생태계의 진화 과정은 다음으로 요약될 수 있다. 첫째, 진화는 염색체(chromosome) 즉 생물의 구조를 부호화(encoding) 하기 위한 최소 요소들에 의해 일어나며, 둘째 자연도태는 소위 환경 및 목적하는 바에 대해 우수성이 높은 염색체 집단인 유전자를 보다 많이 재생하며, 셋째, 돌연변이는 부모의 유전형질을 다른 형질로 바꾸며, 넷째, 교배(Crossover) 및 재조합(Reproduction)은 부모의 유전자들이 결합하여 각각이 가진 염색체들을 부분적으로 교환하여 다른 그러나 부분적으로 닮은 자식을 생산해 낸다. 유전 알고리즘이 기존 알고리즘과 다른 점은 코드화된 parameter(유전자)들의 set에 의해 조작되며, 한 세대(generation)에 여러 개체(population)가 있어 한 점의 탐색이 아닌 동시에 다발적인 여러 개체의 병렬 탐색(parallel search)을 수행하며, 목적함수를 수학적,기하학적 조작으로 조작하지 않는 payoff 혹은 black box 형태의 정보 탐색을 수행하며, deterministic rule에 의하지 않는 확률적(probabilistic) rule에 의한 탐색을 수행한다. 이 방법은 매우 단순하며, 어떤 종류의 문제에도 적용성이 뛰어나고, 효율성(efficiency)과 재현성(efficacy)의 균형을 잘 이루며, 매우 강건한(robust) 탐색기법으로 평가되고 있다.

유전 알고리즘의 기초이론

유전 알고리즘의 탐색 적응성 혹은 알고리즘 수행에 따른 결과의 진보를 증명하는 기본 이론이 schema의 정의로부터 출발한다. 유전 알고리즘을 수행하기 위한 개체 표현은 염색체를 상징하는 스트링 표현임을 앞서 밝힌 바 있다.

John Holland는 이들 유전 정보를 갖고 있는 염색체를 0과1로 구성된 2진 스트

링을 사용하였고 다음에 8 bit 스트링을 가진 A라는 개체의 예를 나타내었다.

$$A = a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 = 0 1 1 1 0 0 1 0$$

여기서 한 예로써 schema H=*11*0*** 를 정의한다면, 개체들 중 a₂, a₃ 은 1, a₅ 는 0 그외에는 무엇이든 관계없는 개체를 schema H 영역의 개체라 정의한다.

즉 schema H 영역의 개체는 a₂, a₃, a₅ 염색체의 발현성이 낮은 개체들인 것이다.

여기서 다음 두가지는 특성이 정의된다. 첫째 defining length(δ)로써 *외에 1혹은 0이 시작된 지점으로 부터 끝난 지점까지의 거리이다. 그렇다면 $\delta(H)$ 는 3이 된다.

둘째 order of schema(O)로써 고정된 스트링의 수를 말하며 O(H)는 3이 된다.

또 자손을 번식시키기 위해 교배의 장(mating pool)에 선택되는 부모는 적합도가 우수한 형질의 개체이다. 여기서 적합도는 최적화 문제를 위해 목적함수가 될 것이고 적합도의 판별은 개체가 가진 유전정보 스트링으로 부터 계산되어 진다.

특정 schema를 갖는 개체집단(schemata)의 다음 세대에서 재조합될 기대치는 다음과 같이 계산되어 질 수 있다.

$$m(H, t+1) = m(H, t) \frac{f(H)}{\bar{f}}$$

m(H,t)는 한세대 A(t)에서 schema H를 갖는 개체 수이며, f(H)는 평균 schema H가 재조합되는 평균 적합도이며 \bar{f} 는 총 개체들의 평균 fitness 이다.

P_s는 스트링이 교배 후 다음 세대에 살아남을 확률이라면, schema H가 살아남을 확률은 교배위치가 $\delta(H)$ 밖에 존재해야만 가능성이 있다. 그래서

$$P_s \geq 1 - P_c \frac{\delta(H)}{l-1}$$

l는 스트링의 길이이며, l-1은 defining length가 선택될 가능한 site 수이다. P_c 는 무작위 선택에 의해 교배 자체가 더 우수한 성능을 가질 확률이다.

$$\therefore m(H, t+1) = m(H, t) \frac{f(H)}{\bar{f}} \left[1 - P_c \frac{\delta(H)}{(l-1)} \right]$$

P_m을 돌연변이가 일어날 확률이라면

$$m(H, t+1) = m(H, t) \frac{f(H)}{\bar{f}} \left[1 - P_c \frac{\delta(H)}{(l-1)} - o(H)P_m \right]$$

의 식으로 돌연변이에 의한 영향을 표현할 수 있다. 위 식이 내포하는 의미는 짧고, order 가 낮고, 평균이상의 fitness를 갖는 schemata, 즉 우수한 형질의 자손이 점차 증가함을 증명하고 있으며, 이것을 schema theorem이라 한다.

단순유전알고리즘과 μ -유전알고리즘

최적화문제를 풀이하기 위한 유전알고리즘의 경우 Min or Max f(적합도 함수)의 목적함수를 사용하게 되고 각 개체는 스트링의 배열로 이루어지며, 이 스트링 배열이 주어질 경우 적합도 함수값이 얼마인지를 계산할 수 있어야 한다. 단순유전알고리즘은 기본 3가지 조작자인 재조합, 교배, 돌연변이 조작자만을 사용한다.

1. 최적화 대상함수 f(적합도 함수)를 정의하며, 개체의 스트링 수를 정한다.
2. 한 세대의 개체수(population)를 정하고 무작위로 각 개체의 스트링을 정한다
3. 각 개체를 적합도함수인자로 변환하여 적합도를 계산한다.
4. 부모 개체들중 적합도에 따라 후손을 재조합 할 수 있는 기회의 가능성을 부여한후 교배를 위해 해당 개체를 재조합 한다.
5. 두 개체간 교배와 돌연변이로 부모세대와 같은 개체수의 자손세대를 만든다.
6. 만족할만한 iteration수 만큼 3항으로 부터 6 항까지를 반복한다.

한 예로서 4개의 개체를 갖는 부모세대(A₁,A₂,A₃,A₄)가 있다고 하고 이들은 8개의 염색체 스트링을 갖는다고 하자. 각 개체의 적합도가 각각 0.32, 0.18, 0.06, 0.44 이라면 다음 Table 1 에 재조합확률, 예상되는 재조합 횟수를 나타내었다.

Table 1. Selection to fitness

개체	적합도(f)	선택확률	예상분포(1)	예상분포(2)
A ₁	0.32	32%	2	1
A ₂	0.18	18%	0	1
A ₃	0.06	6%	0	0
A ₄	0.44	44%	2	2
합	1.00	100%	4	4

즉 A₄ 가 적합도가 가장 큰 개체 이므로 재조합확률이 가장 크다. 예상분포(2)의 경우가 선택이 되었을 경우 교배는 다음과 같이 행해진다.

A₁ = 00000000, A₄ = 00110011, A₄ = 00110011, A₂ = 00001111

먼저 A₁ 과 A₄ 가 교배하고 교배점은 무작위로 정한다. 교배점이 스트링 4, 5 사이라면 결과적으로 도출된 두 자손은 A'₁ = 0000|0011, A'₄ = 0011|0000 이 된다. 같은 방법으로 A₄, A₂ 로써 다른 2개의 개체를 만들어 한 세대의 개체수 (population size)를 채운다. 교배의 장으로 보다 적합도가 우수한 개체가 재조합될 확률이 높고 적합도가 낮은 개체는 도태된다. 돌연변이는 약 1000 개체중 1개 정도의 빈도로 무작위로 선택된 재조합 개체의 무작위로 선택된 염색체 스트링 한 개를 반대의 형질로 단순히 바꾼다. 돌연변이에 의한 기대효과는 국지 최적점에서 빠져나와 보다 우수한 전체 최적치를 구할수 있도록 하고자 함에 있다.

μ -유전알고리즘은 한 세대의 개체수가 단순유전알고리즘에 비해 매우 작은 약 30~200 정도로 제한하여 계산시간을 크게 줄일 수 있도록 제안된 방법을 말한다. 개체수를 줄이는 것은 동시탐색의 병렬처리 point 수를 줄이는 것으로 계산시간은 크게 줄어으나, 성능이 크게 감소될 우려가 있다. 그래서 여기서는 다음의 몇가지 진보된 조작자(advanced operator)를 단순유전알고리즘에 추가한다.

1. 집단의 크기를 5개로 잡을경우 4개는 재조합 및 교배, 돌연변이로 정하고 나머지 1개는 부모세대중 가장 적합도가 큰 개체로 둔다. (엘리트 규칙)
2. 적합도가 떨어지는 개체의 스트링의 일부 서열을 역전시켜 적합도가 우수하게 될 경우 우수해진 개체를 채택한다. (inversion operator)
3. 교배의 방법을 단순 교배에서 PMX(partially matched crossover), OX(oder crossover), CX(cycle crossover)등을 적절히 선택한다.
4. 교배를 위해 두 개체가 배우자의 개념으로 재조합된다. 이때 배우자를 random 하게 선택 함으로써 서로다른 형질의 부모가 자손을 만들어 낼 가능성을 높인다. (segregation and translocation operator)
5. 매우 적합도가 우수한 형질은 재조합후 교배를 거치지 않고 복제하며 매우 적합도가 낮은것은 재조합 확률을 '0' 으로 주어 제거시킨다.
6. 한세대의 개체들을 두 그룹으로 나누어 한 그룹은 남성, 다른 그룹은 여성으로 정의하여 반드시 양성이 교배하도록 하고 생겨나는 자손도 양성으로 나눈다. (sexual determination operator)

유전알고리즘의 응용분야

유전알고리즘은 지난 10년간 매우 다양한 분야에서 응용 연구되어져 최근에는 의학, 생물학, 자연과학, 공학 뿐만아니라 경영학, 사회학에까지 응용이 확대되고 있으며, 구체적으로는 함수최적화, 패턴인식 - image processing, control, operations research, 구조최적화 및 neural network 에도 응용되고 있다. 그 외에 유전알고리즘과 fuzzy logic, parallel computing 혹은 parallel processing, knowledge base techniques 등과의 접목으로 보다 우수한 결과를 나타내는 연구가 활발하다.

다음 Table 2 에 1988년 까지의 유전알고리즘 연구분야의 비율과 1989년 부터 1992년 까지의 유전알고리즘의 연구분야비율을 나타내었다.

Table 2. The areas of application of genetic algorithm

	RESEARCH AREA	~1987(%)	1988~1991(%)
1	Biology	7	9
2	Eng. & O.R.	17	22
3	Genetic algorithms	35	13
4	Hybrid Techniques	11	20
5	pattern recognition	6	12
6	Parallel GA implementations	8	7
7	Social science	5	6
8	Computer science	9	10
9	physical science	2	1

결론 및 토론

유전 알고리즘의 응용분야는 매우 넓고 그 성능 역시 주목할 만 한 것으로 판명되고 있다. 적용에 성공을 거두는 문제의 특성은 난해한 비선형문제이거나 combinatorial 혹은 이산시스템의 많은 미분 불가능점을 갖는 문제들이 대부분이다. 일반적인 unimodal 혹은 심하지 않은 multi-modal 함수들의 경우 미분가능성을 이용한 수학적 최적화에 비해 그다지 우수하다고 할 수 없다. 모든 도구가 적재적소에 사용되어야만 그 효율 및 성능이 최대한 발휘됨은 두말할 나위가 없다. 최적화 문제는 계산시간과 그 성능(해의 질)의 양쪽간에 trade off가 존재하는 것이 일반적이다. 컴퓨터의 발전속도가 크게 발전하기는 하지만 NP-complete 영역의 문제의 크기가 증가하는 속도를 따라잡기는 불가능해 보인다. 그래서 컴퓨터 성능의 발달을 극대화시킬 수 있는 알고리즘으로써 유전 알고리즘은 크게 각광받게 되었다. 유전 알고리즘은 최적해를 항상 보장하지는 못하지만 비교적 신속히 우수한 해를 찾아주며 앞으로 발전될 컴퓨터 시스템의 발전방향에 매우 잘 부합되는 알고리즘이다. 가장 중요한 유전 알고리즘의 장점은 역시 강건성에 있다. 효율(efficiency)과 가용성(efficacy)의 조화와 그 두가지 성능이 모두 높기 때문에 유전 알고리즘은 최소한의 수정으로 매우 다양한 넓은 범위의 최적화문제에 잘 적용되어 매우 우수한 성능을 보여준다.

참고문헌

1. D. E. Goldberg, "Genetic algorithms in search optimization & machine learning", Addison wesley, 1989.
2. L. Davis, "Handbook of genetic algorithm", Van Nostrand Reinhold, 1991.
3. P. Wayner, "Genetic algorithms", BYTE, January, pp. 361, 1991
4. J. H. Holland, "Genetic algorithms", Scientific American, July, pp. 44, 1992.
5. E. R. Crowe and C. A. Vassiliadis, "Artificial intelligence: starting to realize its practical promise", Chem. Eng. Prog., January, pp. 22, 1995.
6. C.H.Lee, J.H.Jung and I.Lee, " A genetic algorithm for scheduling of multi-product batch processes", Comput. chem. Engng, submitted, 1995.
7. D. E. Goldberg, "Computer-aided pipeline operation using genetic algorithms and rule learning. Part 1: genetic algorithms in pipeline optimization, Engineering with Computers, vol. 3, pp. 35, 1987.
8. K. E. Kinnear, Jr. "Advances in genetic programming", The MIT Press, 1994.
9. Proceedings of the fourth international conference on Genetic Algorithm and their applications, Edited by R. Belew, L. Booker and M. Kaufmann, 1991.
10. J. H. Holland, "Adaptation in natural artificial systems, MIT Press, 1992.
11. M. J. Quinn, "Parallel computing, theory and practice", McGRAW-Hill, 1994.