

개체간 해밍 거리 기반의 변이연산을 적용한 유전알고리즘을 이용한 다차원 배낭 문제 탐색

정재훈^o, 이종현, 안창욱*
성균관대학교 컴퓨터공학과

e-mail : a12gjang@skku.edu, ljh08375@skku.edu, cwan@skku.edu

Genetic Algorithm Applying Modified Mutation Operator Based on Hamming Distance for Solving Multi-dimensional Knapsack Problem

Jae-Hun Jeong^o, Jong-Hyun Lee, Chang-Wook Ahn*
Dept. of Computer Science, Sungkyunkwan University

요 약

본 논문에서는 부모 개체의 해밍 거리에 기반하여 선택적 변이연산을 적용한 유전알고리즘을 제안한다. 유전자 형이 매우 유사한 개체들 간의 유전연산은 알고리즘의 탐색성능을 저하시키고 조기 수렴의 가능성을 증가시킨다. 본 논문에서는 이러한 현상을 극복하기 위하여, 교차연산 시 선택된 두 부모 개체간의 해밍 거리에 따라 그 값이 낮으면 교차연산 후 생성된 두 자식 개체 중 한쪽에게 높은 변이확률을 적용하고 다른 한쪽 자식은 부모와 비슷한 유전자 형으로 탐색을 계속하게 하여 조기 수렴을 방지하면서 해집단의 다양성 유지 기능을 향상 시켰다. 제안한 유전 알고리즘을 다차원 배낭 문제에 적용한 결과, 같은 조건에서 단순 유전 알고리즘(SGA) 보다 향상된 탐색 성능을 보여주었다.

1. 서론

본 논문에서는 유전알고리즘의 탐색성능의 향상을 위하여 교차연산 시 선택된 두 부모 개체 간의 해밍 거리에 기반한 새로운 변이연산을 제안한다. 그리고 제안된 유전 알고리즘을 다차원 배낭 문제 탐색에 적용하여 성능 평가를 수행한다.

유전 알고리즘은 자연선택, 교차, 변이의 자연현상에 기반한 경험적 무작위(Heuristic Random) 탐색 기법으로써, 최적화 문제 해결에 폭 넓게 적용되며 탁월한 탐색 성능을 나타낸다 [1]. 일반적으로 유전 알고리즘의 성능 향상을 위해서는 해집단의 다양성이 보장되어야 하며, 만약 다양성 유지에 실패하게 되면 조기수렴에 빠져 최적해 탐색에 실패할 가능성이 높아진다 [2].

유전자형이 매우 유사한 두 개체간의 교차연산은 부모의 유전자와 매우 유사한 유전자 형이 자식세대에도 반복되어 나타나기 때문에 최적해 탐색에 방해요소 작용하고, 조기 수렴 현상을 가속화 하여 세대 충분한 최적해 탐색을 하지 못하고 조기 수렴하여 고착되는 상황을 초래한다 [3].

본 논문에서는 위 문제를 극복하기 위하여 부모 개체간의 해밍 거리에 기반한 변이연산을 제안한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 유전 알고리즘과 다차원 배낭문제 대하여 설명하며, 3장에서는 제안하는 해밍 거리에 기반한 새로운 변이연산에 대해 설명한다. 4장에서는 본 논문에서 제안한 새로운 변이연산을 적용한 유전 알고리즘으로 다차원 배낭 문제를 탐색하여 성능을 평가한다. 5장에서는 결론 및 향후 연구 과제에 대해 설명한다.

2.1 유전 알고리즘(Genetic Algorithm)

유전알고리즘은 자연계의 진화 현상을 모방한 최적해 탐색 알고리즘으로써, 제한된 시간 안에서 효율적으로 복잡한 문제의 해결을 위해 사용된다[1].

유전 알고리즘의 주요 연산에는 선택과 교차, 변이연산이 있으며, 문제의 해를 유전자로 표현하는 유전형 표현방법과 각 유전자의 적합도를 평가하기 위한 함수가 요구된다.

유전 알고리즘의 큰 틀은 초기 유전자 생성을 시작으로 하여 선택과 교차, 변이연산, 적합도 평가를 반복하여 가장 적절한 유전자를 찾아가는 과정을 통해 문제를 해결해 나가는 것이다 [1]. 유전 알고리즘에서 선택연산의 역할은 낮은 적합도의 해를 걸러내고, 좋은 품질의 해가 다음 세대로 전달되도록 하는 것으로서 적자생존의 법칙이 적용되며 [4], 교차연산은 선택된 두 부모의 유전자를 부분적으로 교환하여 자손에게 물려주어 더 좋은 해를 향해 탐색하는 연산이다 [4]. 변이연산은 유전자를 변이하여 새로운 해를

^o 발표자(Presenter), * 책임저자(Corresponding Author)

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 지원 사업의 연구결과로 수행되었음

생성함으로써 다양성을 유지시키면서, 부모 유전자가 가지고 있지 않은 부분에 대한 탐색을 가능하게 한다 [4].

2.2 다차원 배낭 문제

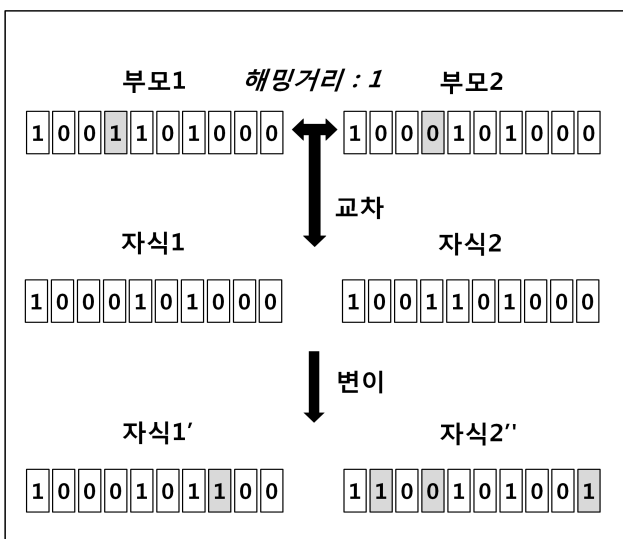
본 논문에서 다루는 다차원 배낭 문제는 잘 알려져 있고 많은 연구가 진행되고 있는 조합 최적화 문제이다. 실제 생활에서 차고 적재, 자원 할당, 금융 연산, 네트워크 스케줄, 예산 문제 등의 많은 산업과 경제적 문제들이 배낭 문제로 표현될 수 있다. 다차원 배낭 문제를 위한 기존의 해결방법으로는 동적알고리즘과 백트래킹 등의 정확한 알고리즘과 탐욕알고리즘, 타부 탐색 알고리즘, 유전 알고리즘 등의 근사적인 알고리즘이 있다 [5].

정확한 알고리즘은 변수가 작은 문제에서 항상 전역 최적해를 보장해 주지만 변수가 커지면 연산의 규모가 비현실적으로 증가하게 되는 문제점이 있다. 반면에 근사적인 알고리즘은 전역 최적해를 보장해 주지는 못 하지만, 변수가 많은 문제의 복잡성을 크게 감소시켜 빠르고 효율적으로 근사적인 답을 찾아 낼 수 있다 [5].

본 논문에서는 근사적 알고리즘의 하나인 유전 알고리즘을 다차원 배낭 문제에 적용하여 전역 최적해를 보다 근접하게 찾아 낼 수 있도록 유전 알고리즘의 변이연산의 개선을 제안하였다.

3. 두 개체간 해밍 거리에 기반한 변이연산

유전자형의 유사도가 매우 높은 두 개체 간의 교차연산은 부모 유전자와 유사한 형태의 유전자가 자식 세대에도 반복하여 생성 될 가능성이 높다[2]. 그리고 이러한 현상은 수렴이 진행됨에 따라 빈번하게 일어나게 되며, 이것은 해집단의 다양성 감소 현상을 가속화하고 충분한 해 공간 탐색을 방해함으로써 최종 해의 질을 떨어뜨리게 된다[2].



(그림 1) 제안된 새로운 변이연산

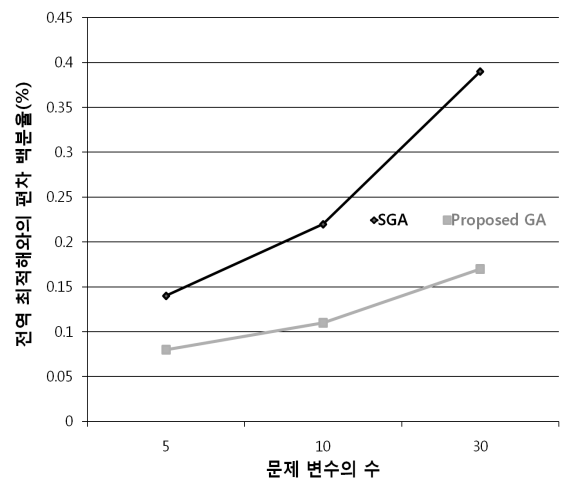
교차연산 시에 두 개체간 해밍 거리를 측정하여 두 개체의 유전자가 매우 유사하여 해밍거리가 기준보다 낮게 나온다면 (경험적으로 찾은 값인 유전자 길이의 97%) 생성된 두 자식 개체 중 한쪽 개체에는 높은 변이 확률(5%~10%)을 적용하고, 다른 하나의 개체에는 표준변이 확률(유전자 당 1개 혹은 2개 변이) 을 적용한다 [4]. 생성된 두 자식 중 한쪽 자식 유전자에만 높은 변이 확률을 적용하고, 다른 자식은 기존의 변이 확률을 그대로 적용함으로써, 확률적으로 한쪽 개체는 부모와 유사한 유전자 형태로서 해 공간탐색을 계속하여 해의 수렴을 방해하지 않으면서, 다른 한쪽 개체는 좀 더 높은 확률의 변이연산을 하여 다양성을 유지하는 탐색전략을 사용하였다.

4. 실험

본 논문에서는 Beasley가 제시한 다차원 배낭 문제 크기와 탐색공간의 형태에 따라 27개의 문제를 임의로 선택하여, 제안한 알고리즘의 성능을 단순 유전 알고리즘 (SGA)과 비교 평가 하였다 [6]. 문제의 변수(m)는 5개, 10, 25개 이고, 변수 별로 아이템의 수(n)는 각 100개, 250개, 500개 이다. 각 문제당 0.25, 0.5, 0.75의 비율로 전역 최적해의 분포형태(α)에 따라 3가지로 나누어 저었다.

한 문제당 10번의 테스트를 하여 발견한 해의 최대값과 평균값, 주어진 문제의 전역 최적해와 찾은 해의 평균값의 편차 백분율을 비교하였다.

본 논문의 모든 알고리즘은 C++로 구현되었으며, 실험에 사용된 컴퓨터는 인텔 Core i3 2.27GHz CPU, 4.0GB RAM 이다.



(그림 2) 문제 변수의 수에 따른 단순 유전 알고리즘과 해밍 거리 기반의 변이연산을 적용한 유전알고리즘의 전역 최적해와의 편차 백분율 비교

<표 3> 단순 유전 알고리즘과 부모 개체 간의 해밍거리 기반의 변이연산을 적용한 유전 알고리즘을 Beasley가 제시한 OR-library의 다차원 배낭 문제[6] 에 적용 결과 성능 비교

m	n	a	Benchmark	단순 유전 알고리즘(SGA)			해밍거리 기반의 변이확률을 적용한 GA		
				Best	Average	편차 백분율(%)	Best	Average	편차 백분율(%)
5	100	0.25	24381	24381	24320.3	0.2489	24381	24349	0.1312
		0.50	42757	42705	42639.4	0.2750	42757	42725.8	0.0729
		0.75	59822	59822	59781.3	0.0680	59822	59819.6	0.0040
	250	0.25	59312	59223	59165.9	0.2463	59312	59188.2	0.2087
		0.50	109109	109063	108965.5	0.1315	109079	109003.7	0.0965
		0.75	149659	149605	149569.9	0.0595	149646	149593.3	0.0438
	500	0.25	120130	119948	119851.6	0.1515	120042	119954.9	0.0732
		0.50	218422	218300	218222.2	0.0558	218342	218246.4	0.0366
		0.75	295828	295643	295602.9	0.0625	295747	295662.1	0.0273
Average					0.14			0.08	
10	100	0.25	23064	23044	22954.2	0.4760	23057	23053.7	0.0446
		0.50	41395	41395	41295.5	0.2403	41395	41354.7	0.0973
		0.75	57375	57375	57266.6	0.1889	57375	57360.7	0.0249
	250	0.25	59187	58998	58947.4	0.4048	59133	59032.6	0.2608
		0.50	110863	110753	110653.3	0.1891	110772	110701.8	0.1454
		0.75	151790	151705	151647.9	0.0936	151779	151683.9	0.0698
	500	0.25	117726	117490	117307.4	0.2004	117540	117347.8	0.1579
		0.50	217318	217012	216914.5	0.1408	217058	216947.6	0.1196
		0.75	304344	304106	304012.9	0.0782	304223	304061.8	0.0397
Average					0.22			0.11	
30	100	0.25	21946	21946	21735.5	0.9591	21946	21946	0.0000
		0.50	40767	40637	40463.5	0.7444	40767	40660.4	0.2614
		0.75	57494	57494	57374.2	0.2083	57494	57477.7	0.0283
	250	0.25	56693	56662	56439.2	0.4476	56677	56540.7	0.2686
		0.50	107689	107539	107440.1	0.2311	107647	107503.3	0.1724
		0.75	150083	149940	149825.3	0.1717	150009	149895.1	0.1251
	500	0.25	115868	115410	115113.2	0.3952	115519	115276.8	0.3012
		0.50	217995	217551	217427.2	0.2036	217667	217515.5	0.1504
		0.75	301627	301208	301007.7	0.1389	301291	301191	0.1113
Average					0.39			0.16	
Average					0.25			0.11	

<표1> 유전 알고리즘의 변수 설정

환경변수	내용
인구 수	2500
세대 반복 수	1000
교차 확률	90%
기본 변이 확률	0.5~1%
높은 변이 확률	5~10%
해밍 거리 경계	유전자 길이의 97%

실험 결과 선택된 모든 유형의 다차원 배낭 문제에서, 해밍 거리 기반의 변이연산을 적용한 유전 알고리즘의 최종해의 평균이 전역 최적해에 더욱 가까워 졌으며, 27개의 문제 중 26개의 문제에서 최종해의 최대값이 상승하였다.

문제의 크기가 커지면서 단순 유전 알고리즘은 편차 백분율의 크기가 가파르게 상승한 반면에 해밍 거리 기반의 변이연산을 적용한 유전 알고리즘은 완만한 상승 폭을 나타내었다.

5. 결론 및 향후 과제

본 논문에서는 부모 개체의 해밍 거리에 기반한 변이연산을 제안 하였으며, 이를 다차원 배낭 문제에 적용하였다.

제안한 방식은 두 개의 자식 해를 생산하여 부모의 해밍 거리에 따라, 하나의 자식에만 높은 변이확률을 적용하여 해집단의 다양성을 증가시키면서, 다른 하나의 자식은 부모의 해와 유사한 형태로 탐색을 계속할 수 있게 하여 해들의 수렴을 크게 방해하지 않는다.

실험결과, 단순 유전 알고리즘과 비교하여 제안된 변이연산을 적용하였을 때 모든 결과에서 다차원 배낭 문제에서의 탐색성능이 향상되었다.

제안한 방법은 구현이 매우 간단하고 명료하며, 해집단의 다양성 유지에 도움을 주어 실험결과에서 처럼 유전 알고리즘의 탐색 성능 향상에 도움을 준다.

그리고 나아가서 적합도 기반의 적응적 유전 알고리즘을 통해 더욱 정교한 변이 확률 적용한다면 탐색성능을 더욱 개선할 수 있을 것이다 [5].

또한 향후 한 세대의 전반적인 다양성 평가를 위해 해밍 거리를 사용하여 다양한 교차연산을 적용 하는 복합 유전 알고리즘과 본 논문에서 제시한 변이연산과의 융합으로 성능 개선에 대한 연구를 진행 할 것이다 [7].

참고문헌

- [1] D.E Goldberg, Genetic Algorithms in Search, Optimization & Machine Learning, Addison Wesley Publishing company, 1989
- [2] Abu Bakar Md. Sultan, Ramlan Mahmod, Md. Nasir Sulaiman, Mohd. Rizam Abu Bakar, Maintaining Diversity for Genetic Algorithm : A Case of Timetabling Problem, Jurnal Teknologi, pp. 123 - 130, Jun 2006
- [3] Taeho C. Jo, SGA Variants with Crossover based on Hamming Distance, Proceedings of the International Conference on Artificial Intelligence, vol. 1, pp. 26-31, 1999.
- [4] P.C. Chu and J.E. Beasley, A Genetic Algorithm for the Multidimensional Knapsack Problem, Journal of Heuristics, vol. 4, pp. 63-86, 1998.
- [5] Yanqin Ma and Jianchen Wan, Improved Hybrid Adaptive Genetic Algorithm for Solving Knapsack Problem, The 2nd International Conference on Intelligent Control and Information Processing, vol. 2, pp. 644-647, 2011.
- [6] Beasley, J.E. "OR-Library: Distributing Test Problems by Electronic Mail," Journal of the Operational Research Society 41, 1069 - 1072. 1990.
- [7] Mohammad Jalali Varnamkhasti, Lai Soon Lee, Mohd Rizam Abu Bakar, and Wah June Leong, A Genetic Algorithm with Fuzzy Crossover Operator and Probability, Advances in Operations Research, vol. 2012, Article ID 956498, 2012.