유전자 알고리즘을 이용한 수도쿠 퍼즐 난이도 측정

조영조*, 김병욱**
*전북대학교 IT응용시스템공학과
**동국대학교 컴퓨터공학과
e-mail:jyj529@jbnu.ac.kr, bwkim@dongguk.ac.kr

Sudoku Puzzle Difficulty Measurement using Genetic Algorithm

Yeongjo Cho*, Byoungwook Kim**
*Dept of IT Applied System Engineering, Chonbuk National University
**Dept of Computer Engineering, Dongguk University

요 약

스도쿠는 다양한 연령대의 사람들에게 인기있는 논리 기반 조합 퍼즐 게임입니다. 이러한 인기로 인해 다양한 수준의 난이도로 스도쿠 퍼즐을 생성하고 해결하기 위해 컴퓨터 소프트웨어가 개발되고 있습니다. 본 논문은 유전자 알고리즘 (Genetic Algorithm)을 이용하여 스도쿠 퍼즐의 해결 및 스도쿠 퍼즐의 난이도를 평가하는 문제를 연구한다. 유전자 알고리즘이 수행되는 동안 적합도 함수가 수렴되는 시간이 길수록 퍼즐의 난이도가 높을 것이라는 가정한다. 본 논문에서는 유전자 알고리즘을 이용하여 스도쿠 퍼즐을 해결하는 알고리즘을 개발하고, 스도쿠 퍼즐 제작자들에 의해 난이도가 정해진 실제 스도쿠 퍼즐을 개발된 알고리즘으로 해결하는데 시간을 측정한다. 측정된 시간과 퍼즐 문제의 상관 관계를 분석하여 가정을 검증한다. 실험결과에서 알고리즘의 실행시간과 퍼즐의 난이도에는 유의미한 상관 관계가 있음을 보였다.

1. 서론

스도쿠는 전 세계적으로 인기를 얻고 있는 유명한 숫자 퍼즐 게임이다. 스도쿠 퍼즐은 전통적인 마방진의 형태에 스도쿠만의 특별한 규칙을 적용해 만들어진 배열의 임의의 칸을 여러 개 비움으로써 난이도를 조절하며, 처음에 주어진 숫자들과 해당 게임의 규칙에 따라 빈칸을 모두 채우게 되면 그 퍼즐을 푼 것이 된다 [1]. 일반적으로 가로 9칸, 세로 9칸 총 81칸의 정방형 배열로 구성되어 있으며, 이는 다시 가로 3칸, 세로 3칸 총 9칸의 작은 배열 9개로 다시 구분된다 (그림 1). 스도쿠 퍼즐을 푸는 근본적인 규칙은 일정한 구역에 대해 1부터 9까지 9개의 숫자가 중복되지 않고 나열되어야 한다는 것이다 [2]. 스도쿠의 규칙은 비교적 간단하지만 조건을 만족하는 답을 찾는 것은 쉬운 일은 아니다.

8								
		3	6					
	7			9		2		
	5				7			
				4	5	7		
			1				3	
		1					6	8
		8	5				1	
	9					4		

8	1	2	7	5	3	6	4	9
9	4	3	6	8	2	1	7	5
6	7	5	4	9	1	2	8	3
1	5	4	2	3	7	8	9	6
3	6	9	8	4	5	7	2	1
2	8	7	1	6	9	5	3	4
5	2	1	9	7	4	3	6	8
4	3	8	5	2	6	9	1	7
7	9	6	3	1	8	4	5	2

(그림 1) 스도쿠 퍼즐 문제와 해답

스도쿠 퍼즐 문제는 컴퓨터 과학의 대표적인 난제인 NP-Complete 문제로 알려져 있다 [3]. 스도쿠 퍼즐의 최적해를 찾는 알고리즘들이 존재한다 할지라도 계산 시간이 너무 길 수밖에 없다. 따라서 스도쿠 문제의 효율적인 해결을 위해 다양한 휴리스틱과 탐색 알고리즘들이 개발되어 왔다. 그 중에서 유전자 알고리즘을 이용하여 스도쿠 퍼즐을 해결하는 연구가 많이 진행되었다 [4, 5, 6]. 유전 알고리즘은 자연 진화의 원리를 최적해를 탐색하는 문제에 이용하는 진화 연산의 대표적인 휴리스틱 알고리즘이다. 대부분의 연구는 유전자 알고리즘을 이용하여 스도쿠 문제를 생성하거나 주어진 문제를 빠르게 해결하려는 연구가 이루어져 왔다. 그러나 아직까지 제작된 스도쿠 문제의 난이도를 설정하는 연구를 이루어지지 않았다. 스도쿠를 이용하여 사용자들의 지능을 측정할 경우 제작된 스도쿠 문제의 정확한 난이도를 알아야 올바른 지능 측정이 될 수 있다. 따라서 스도쿠 문제의 난이도를 객관적으로 측정하는 것은 중요한 문제일 수 있다. 이에 본 연구에서는 유전자 알고리즘의 수렴 속도를 이용하여 스도쿠 문제의 난이도를 객관적으로 측정하는 방법의 유효성을 검증하고자 한다.

2. 유전자 알고리즘

유전 알고리즘은 생물학의 진화 이론을 응용한 연산

방법으로, 임의적으로 생성된 최초의 해집합을 하나의 인구 집단으로 간주하고 적합 선택, 교배, 돌연변이 등 실제 생태계에서 일어나는 현상과 유사한 적응 단계를 거쳐 최종적으로 환경에 가장 잘 적응한, 즉 가장 적합한 해를 구하는 방법이다. 단순히 무작위적인 입력을 제공하는 표준 난수 함수와는 달리 적합도(Fitness)라는 개념을 적용하여 계산이 거듭될수록 효율적으로 유용한 해에 접근할 수 있다는 것이 가장 큰 장점이다.

이를 이용하여 스도쿠의 기본 규칙 뿐만 아니라 원하는 변형 규칙을 따르는 고급 스도쿠도 생성할 수 있다. 난수를 발생시키는 특성을 가지고 있으므로 스도쿠 퍼즐에 있어서는 생성뿐만 아니라 실제 빈칸과 초기 힌트가 주어진 하나의 문제를 해결하는 데에도 유용한 도구가 될 수 있다. 특히 적합도의 판단이라는 강력한 특성을 이용하면 손으로 풀기 매우 어려운 난이도의 퍼즐이라도 유전 알고리즘을 이용하여 빠르게 풀 수 있다.

3. 스도쿠 퍼즐 해결을 위한 유전자 알고리즘

스도쿠 퍼즐을 나타내는 염색체는 81바이트 값으로 구성된다. 이들은 3x3 서브 그리드를 나타내기 위해 9바이트 값을 갖는 9개의 그리드로 분할된다. (그림 2)는 염색체 내의 2개 하위 그리드의 예를 보여준다.

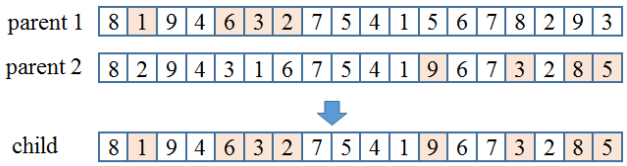
8	1	9	4	1	5
4	6	3	6	7	8
2	7	5	2	9	3

8	1	9	4	6	3	2	7	5	4	1	5	6	7	8	2	9	3

(그림 3) 염색체 내의 2개 하위 그리드의 예.

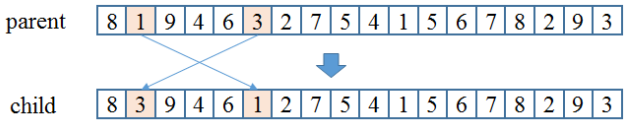
(1) 초기화: 각 염색체에 문제로부터 주어진 값을 저장한 다음 각 행이 1에서 9까지의 값의 유효한 순열이되도록 임의로 값을 생성한다.

(2) 크로스 오버: 유전자 알고리즘의 크로스 오버 연산은 모집단에 포함된 두 개의 염색체의 속성을 결합하여 모집단의 새로운 염색체를 작성한다. 일반적으로 두 개의 부모 염색체로부터 한두 개의 자식 염색체만 생성된다. 가장 기본적인 크로스 오버 방법 중 하나는 균일 크로스 오버이다. 균일 크로스 오버에서는 두 부모의 동등한 부분에서 자식 염색체가 만들어진다. (그림 4)는 크로스 오버 연산에 대해 설명한 그림이다.



(그림 4) 크로스 오버 예

(3) 뮤테이션: 뮤테이션 연산자는 새로운 세대 구성원이 생성될 때 현재 세대를 구성하고 있는 염색체에서 임의의 값을 변경하는 것이다. 뮤테이션은 서브 블록 내부에만 적용되도록 하였다. 원래, 조합 최적화에 일반적으로 사용되는 스왑 뮤테이션 전략을 사용하였다. 스왑 뮤테이션에서 두 위치의 값이 교환된다. 서브 블록 내에서 뮤테이션이 시도 될 때마다 문제로 주어진 배열이 확인되고, 임의로 선택된 위치를 변경하는 것이 불필요한 경우에는 뮤테이션 연산이 생략된다.



(그림 5) 뮤테이션 예

(4) 선택은 엘리트 전략과 토너먼트 선택의 조합을 사용하여 수행된다. 엘리트 전략에 기반하여 부모 집단에서 가장 적합한 염색체는 자식 집단에 직접 추가된다. 그런 다음 토너먼트 선택을 사용하여 두 개의 염색체를 부모 집단에서 선택한다. 교차 및 돌연변이를 사용하여 새로운 염색체가 만들어지고 자식 집단에 추가된다.

(5) 적합도 함수: 유전자 알고리즘을 이용하여 최적의 해를 찾는 적합도 함수를 설계하는 것은 조합 문제에서 어려운 작업이다. 이 경우, 우리는 여러 개의 다른 구속조건 위반에 대해 각각 에러 값을 높이는 다소 복잡한 적합도 함수를 사용한다. 첫 번째 테스트에서 각 행과 열 합계는 45와 같아야하고 각 행과 열 곱은 9!와 같아야 한다. 세 번째 요구 사항은 집합 이론에서 도출된다. 각 행 x_i 및 열 x_j는 1에서 9까지의 정수를 포함하는 집합 A이어야 한다. 그렇지 않을 경우 적합도 함수의 에러값이 높아진다.

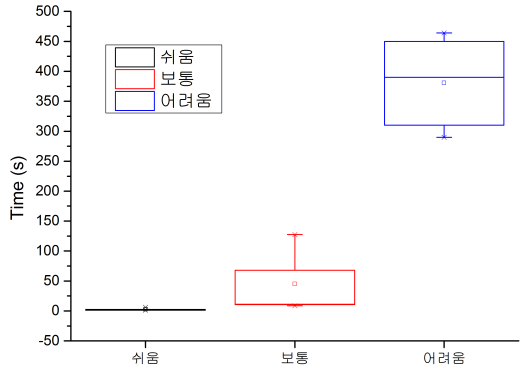
4. 실험 및 결과

알고리즘을 테스트하기 위해 온라인 스도쿠 생성기(https://sudoku.com/ko/swium/)를 사용하여 "쉬움", "보통", "어려움"의 난이도를 갖는 스도쿠 퍼즐을 난이도 별로 각각 5문항씩 생성하였다. <표 1>의 설정을 사용하여 각 퍼즐에 대해 알고리즘을 10번 수행하여 평균 수행결과를 제시하였다.

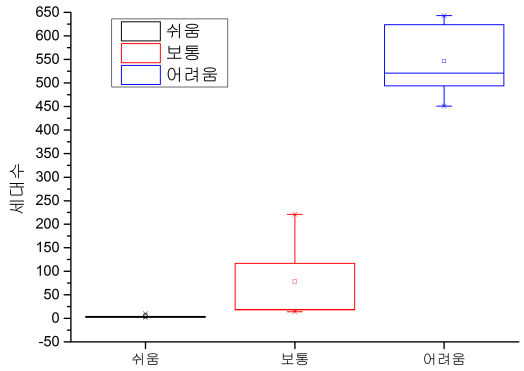
<표 1> 알고리즘 테스트에 사용되는 설정

유전자 집단 크기	500
토너먼트 크기	3
뮤테이션 비율	10%
크로스오버 비율	50%
엘리트 전략 비율	80%
최대 생성 세대수	5,000

(그림 5)는 난이도에 따른 수렴 시간을 나타낸다. (그림 5)에서 보는 바와 같이 스도쿠 문제의 난이도가 높아질수록 수렴 속도가 증가하는 것을 볼 수 있다. 난이도가 쉬울 경우 보통 난이도와 크게 차이가 나지 않았지만, 난이도가 어려운 경우 보통 난이도와 큰 차이가 나는 것을 알 수 있다. 난이도와 수렴 시간의 상관 계수는 0.876로 양의 상관관계가 높은 것으로 나타났다.



(그림 6) 난이도에 따른 수렴 시간



(그림 7) 난이도에 따른 세대수

(그림 6)은 난이도에 따른 세대 수를 나타낸다. (그림 6)에서 보는 바와 같이 스도쿠 문제의 난이도가 높아질수록 생성되는 세대 수도 증가하는 것을 볼 수 있다. 생성되는 세대수 또한 난이도가 쉬울 경우 보통 난이도와 크게 차이가 나지 않았지만, 난이도가 어려운 경우 보통 난이도와 큰 차이가 나는 것을 알 수 있다. 난이도와 생성되는 세대수의 상관 계수는 0.711로 양의 상관관계가 높은 것으로 나타났다.

5. 결론

본 논문에서는 유전자 알고리즘 (Genetic Algorithm)을 이용하여 스도쿠 퍼즐의 해결하고, 스도쿠 퍼즐의 난이도와 수렴 속도의 상관 관계에 대해 연구한다. 실험 결과는 스도쿠 퍼즐의 난이도가 증가함에 수렴 속도가 증가하였으며, 세대수 또한 증가하였다. 난이도와 수렴 속도에는 상관관계가 높은 것으로 나타났다.

본 연구를 수행하는 과정에서 난이도가 높은 스도쿠 문제의 경우, 즉 초기에 위치가 정해진 숫자의 개수가 적은 경우에 수렴이 되지 않는 경우도 발생하는 것을 확인할 수 있었다. 유전자 알고리즘의 경우 초기 유전자 집단을 랜덤으로 설정하면서 로컬 미니멈(local minimum)에 빠지면서 수렴 속도가 느려지는 것을 알 수 있었다.

유전자 알고리즘은 실험에 다양한 변수(유전자 집단 크기, 선택 전략, 뮤테이션 비율, 크로스오버 비율)들이 사용된다. 우리는 추후 본 연구에서 사용된 설정 이외의 수렴 속도를 높일 수 있는 다양한 변수에 대해 연구를 지속할 예정이다.

참고문헌

[1] Semeniuk, I.: Stuck on you. In NewScientist 24/31 December (2005) 45-47

[2] Wikipedia: Sudoku. Available at http://en.wikipedia.org/wiki/Sudoku

[3] T. Yato and T. Seta. Complexity and Completeness of Finding Anohter Solution and Its Application to Puzzles. In Proceedings of the National Meeting of the Information Processing Society of Japan (IPSJ) 2002.

[4] S. K. Lee, D. I. Seo, and B. R. Moon. A Hybrid Genetic Algorithm for Optimal Hexagonal Tortoise Problem. In Genetic and Evolutionalry Computation Conference. page 689, 2002.

[5] Xiu Qin Deng, Yong Da Li, A novel hybrid genetic algorithm for solving Sudoku puzzles, Optim Lett (2013) 7:241 - 257.

[6] Moraglio, A., Togelius, J., Lucas, S.: Product geometric crossover for the sudoku puzzle. In 2006 IEEE Congress on Evolutionary Computation (CEC2006), Vancouver, BC, Canada, July 16-21 (2006) 470-476.