

# 복합 유전자 알고리즘에서의 국부 탐색을 위한 셀룰러 학습 전략

## (A Cellular Learning Strategy for Local Search in Hybrid Genetic Algorithms)

고 명 숙<sup>\*</sup>    길 준 민<sup>\*\*</sup>  
(Myung-Sook Ko) (Joon-Min Gil)

**요 약** 유전자 알고리즘(GA: Genetic Algorithm)은 최적화 문제를 풀기 위해 생물학적 진화(evolution) 과정을 모방한 최적화 알고리즘이다. 유전자 알고리즘은 복잡한 상태 공간에서 최적 해를 찾기 위해 전통적인 최적화 기법과는 달리 유향적 임의 탐색을 행한다. 학습에 해당하는 국부 탐색(local search)을 유전적 탐색과 결합시킨 복합 유전자 알고리즘은 exploration 탐색과 exploitation 탐색의 균형을 유지시켜 줄 수 있는 한 방법이다. 모집단 내의 각 개체가 진화 과정 중에 학습한 유전적 특질들은 그 다음 세대에서 되풀이되며 이러한 학습(learning) 과정을 유전자 알고리즘과 결합시킴으로써 탐색 속도의 향상을 기대할 수 있다.

이 논문에서는 함수 최적화를 위해 학습 속도를 개선한 셀룰러 학습을 기반으로 하는 유전자 알고리즘을 제안한다. 제안하는 셀룰러 학습 전략은 셀룰러 오토마타의 주기성과 수렴성을 기반으로 하며, 유기체가 그 개체의 생명 주기의 한 세대에서 얻게되는 지식과 경험들을 자손에게 전달한다는 이론을 바탕으로 한다. 제안한 셀룰러 학습 전략의 효율을 기존의 복합 유전자 알고리즘에서의 라마기안 진화 및 볼드윈 효과와 비교하였다. 다양한 테스트 베드 함수에 대한 실험을 통하여 셀룰러 학습에 의한 개체의 국부적 향상이 전체적인 성능 향상에 기여함을 알 수 있었고 제안한 학습 전략이 기존의 방법보다 더 빨리 전역 최적 해를 찾을 수 있음을 증명하였다.

**Abstract** Genetic Algorithms are optimization algorithm that mimics biological evolution to solve optimization problems. Genetic algorithms provide an alternative to traditional optimization techniques by using directed random searches to locate optimal solutions in complex fitness landscapes. Hybrid genetic algorithm that is combined with local search called learning can sustain the balance between exploration and exploitation. The genetic traits that each individual in the population learns through evolution are transferred back to the next generation, and when this learning is combined with genetic algorithm we can expect the improvement of the search speed.

This paper proposes a genetic algorithm based Cellular Learning with accelerated learning capability for function optimization. Proposed Cellular Learning strategy is based on periodic and convergent behaviors in cellular automata, and on the theory of transmitting to offspring the knowledge and experience that organisms acquire in their lifetime. We compared the search efficiency of Cellular Learning strategy with those of Lamarckian and Baldwin Effect in hybrid genetic algorithm. We showed that the local improvement by cellular learning could enhance the global performance higher by evaluating their performance through the experiment of various test bed functions and also showed that proposed learning strategy could find out the better global optima than conventional method.

<sup>\*</sup> 종신회원 : 부천대학 전산정보처리과 교수  
kms@hangil.bucheon.ac.kr

<sup>\*\*</sup> 학생회원 : 고려대학교 컴퓨터학과  
jmgil@disys.korea.ac.kr

논문접수 : 2000년 9월 5일

심사완료 : 2001년 7월 10일

## 1. 서론

진화 이론에 의하면 진화 과정 중 모집단에 속하는 가장 적합한 개체들만이 그 다음 세대에도 생존할 수 있으며 각 개체는 생명 주기 동안에 적응 능력을 학습

하며 그 유전적 형질들이 다음 세대에도 보존됨을 알 수 있다. 또한 생명 주기동안에 학습된 특질들은 그 다음 세대에서 다시 학습할 필요가 없다. 이러한 경험 획득 과정을 학습(learning)이라 부르며 일반적으로 유전자 알고리즘에서는 국부 탐색(local search)을 통하여 학습이 이루어진다[16].

유전자 알고리즘은 개체(individual)들로 구성된 집합인 모집단(population)과 모집단에 대해 정의된 유전 연산(genetic operator)들로 구성된다. 전통적인 유전자 알고리즘은 문제 영역을 표현하기 위해 비트 스트링(bit-string)을 사용하며 이때, 각 스트링은 후보 해를 표현한다[1,2]. 유전자 알고리즘은 후보해 집합 중 해에 도달할 가능성이 높은 개체들에 유전 연산을 가함으로써 더욱 향상된 후보 해들을 산출해 나가는 방향으로 탐색을 행한다. 또한, 유전 연산인 교차 연산과 돌연 변이 연산을 통하여 현 모집단으로부터 한 세대 진행한 후의 개체들의 집단을 만들어 낸다. 이러한 과정은 종료 조건을 만날 때까지 반복해서 수행된다[1,8,10].

복합 유전자 알고리즘(HGA: Hybrid Genetic Algorithm)은 국부 탐색을 기존의 유전자 알고리즘에 결합시킨 알고리즘이다. 국부 탐색은 유기체가 환경의 변화에 의한 영향과 그 세대에 획득한 지식, 경험 등을 다음 세대에도 물려줄 수 있도록 개체들을 진화시키는 역할을 수행한다. 즉, 그 세대에 행해지는 국부 탐색 자체를 개체가 생존 기간에 환경으로부터 얻게 되는 학습 개념으로 본 방법이다. 국부 탐색과 결합된 유전자 알고리즘은 전체 탐색 공간에 대한 유망적 임의 탐색뿐만 아니라 의미있는 영역에 대해서 보다 집중적인 탐색을 할 수 있도록 한다[11,12].

국부 탐색의 대표적인 방법으로 라마키안 진화(Lamarckian Evolution)와 볼드윈 효과(Baldwin Effect)가 있다. 라마키안 진화에 의한 학습은 개체의 코딩 자체에 학습된 결과를 적용시키는 방법이며, 볼드윈 효과는 학습된 결과를 개체 코딩에는 영향을 주지않고 그 개체의 적합도 함수 값만을 변화 시킴으로써 학습을 행하는 방법이다[3,9]. 그러나, 라마키안 진화를 결합시킨 복합 유전자 알고리즘은 상태 공간이 비교적 완만한 경우에는 최적해를 빨리 찾을 수 있지만, 굴곡이 심한 상태 공간에 대해서는 학습이 진행되어도 현 상태에서 기울기 강하에 의한 국부 탐색을 수행하므로 국부 최적점에 도달한 경우엔, 더 이상의 성능 향상을 기대할 수 없다는 단점이 있다. 볼드윈 효과에 의한 탐색은 전역 최적해로의 수렴 가능성이 높은 반면 수렴 속도가 느리다는 단점을 가지고 있다.

이 논문에서는 새로운 학습 전략인 셀룰러 학습 기법을 제안하고 이 기법을 기존의 유전자 알고리즘과 결합 시킴으로써 학습 속도의 개선과 전역 최적해로의 수렴도의 향상을 보이고자 한다. 또한, Witley의 테스트 함수와 DeJong의 테스트 함수[4]에 대한 최적화 실험을 통하여 제안하는 기법의 탐색 효율에 대한 성능을 평가 분석한다.

이 논문의 구성은 다음과 같다. 2장에서는 기존의 국부 탐색에 의한 학습 기법인 라마키안 진화와 볼드윈 효과에 대해 설명한다. 3장에서는 제안하는 셀룰러 학습 전략에 대해서 기술하고, 4장에서 기존의 학습 전략을 사용하는 복합 유전자 알고리즘과 제안하는 방법과의 비교를 위해 여러 테스트 베드 함수에 대한 최적화 실험을 통하여 그 성능을 평가한다. 마지막으로 5장에서는 실험 결과에 대한 분석 및 결론과 앞으로의 연구 방향을 제시하고자 한다.

## 2. 국부 탐색에 의한 학습

학습을 행하는 국부 탐색 기법으로는 기울기 강하(gradient descent) 방법을 이용하는 라마키안 진화 방법과 볼드윈 효과가 있다.

### 2.1 라마키안 진화

라마키안 진화(Lamarckian Evolution) 알고리즘은 기울기 상승 방법인 국부적 언덕 오르기(hill-climbing) 방법을 학습 메커니즘으로 결합시킨 복합 유전자 알고리즘이다. 이 알고리즘은 환경의 변화가 자손에게 물려주는 유기체의 유전 형질(genotype)을 변화시킬 수 있다는 라마키안 이론을 근거로 한다[10]. 이 기법은 각 유기체가 그들의 생존 기간(life time)에 습득한 지식이나 경험과 같은 학습에 의해 획득되어진 향상된 형질들이 그대로 유지가 되게 함으로써 이미 획득된 향상에 대해서는 더 이상의 반복 학습을 필요로 하지 않는다는 특징을 갖는다[13].

교차 연산과 돌연 변이와 같은 유전 연산이 적용된 후의 각 자식 개체들은 현재 그 개체의 상태에서부터 국부적인 최적화를 수행한다. 국부 탐색을 적용한 후의 개체는 탐색 전의 개체와는 다른 유전 형질을 가지게 되는데 이는 국부 탐색을 수행하는 동안 탐색 공간에서의 이동 거리 및 변화가 개체 자체에 반영됨을 의미한다. 탐색 후의 개체는 지역적으로 최적 해에 해당하는 상태 공간에 있게 되며 결과적으로 국부 최적 해에 도달할 수 있도록 학습되는 것이다. 각 개체가 학습한 내용이 다음 세대에 전달되도록 하기 위해 기존의 국부 최적화 방법으로 학습을 시킨 후에 평가와 선택을 수행한다. 국

부 탐색의 기반이 되는 언덕 오르기 알고리즘은 다음 그림 1과 같다.

```

begin
  t ← 0
  repeat
    local ← FALSE
    select a current string  $v_c$  at random
    evaluate  $v_c$ 
    repeat
      select strings in the neighborhood of  $v_c$  by flipping
        single bits of  $v_c$ 
      select the string  $v_n$  from the set of new strings
        with the largest value of fitness function  $f$ 
    if  $f(v_c) < f(v_n)$ 
      then  $v_c \leftarrow v_n$ 
      else local ← TRUE
    until local
    t ← t+1
  until t = MAX
end
    
```

그림 1 언덕 오르기 알고리즘

2.2 볼드윈 효과

볼드윈 효과(Baldwin Effect)는 학습의 또 다른 형태로서 국부 최적화를 통하여 얻은 향상이 각 개체의 유전적 형질은 변화시키지 않고 그대로 유지하면서 개체의 적합도 함수 값만을 변화시키면서 학습해 나가는 방법이다[10]. 볼드윈 효과는 학습이 진화의 방향을 결정지을 수 있다는 가정으로부터 출발하였으며, 다음과 같은 관찰을 그 근거로 한다[4,16].

- 1) 가변 환경에서 종(species)이 진화할 때, 생존 기간 동안 학습할 수 있는 능력을 갖는 개체가 더 선호되는 진화 압력(evolutionary pressure)이 생긴다.
- 2) 많은 특질들을 배울 수 있는 개체는 고정된 특질과 관련된 유전 형질을 가지게 될 확률이 낮아진다.

선택 압력(selective pressure)은 학습된 행동이 진화 방향을 결정하거나 그렇게 되도록 영향을 미친다. 이 특성은 최소화 문제에서 국부적 탐색 형태의 학습을 결합시켰을 경우, 유전자 알고리즘에 의한 전체 탐색과 국부 탐색에 의한 미세한 튜닝(fine-local tuning) 효과에 의해 전체적인 탐색 효율을 높일 수 있다. 다음 그림 2.2는 국부 탐색을 수행할 경우 탐색 횟수에 따른 적합도 함수의 변화를 보여주는 예이다.

그림 2는 n단계의 국부 탐색을 적용했을 경우와 완전한 국부 탐색을 적용한 경우에 대한 탐색 후의 변화된 함수의 모형을 보여준다. 적합도 함수의 모형은 점점 덜

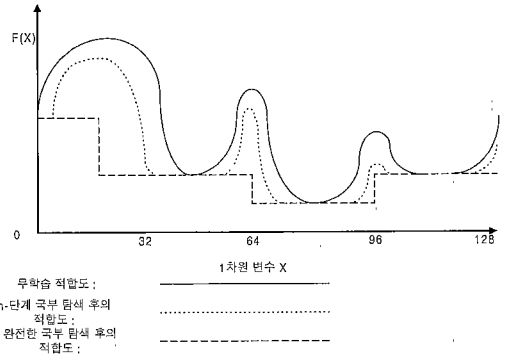


그림 2 국부 탐색 학습에 의한 적합도 함수의 변화

복잡한 최적화 문제로 수렴되고 있을 알 수 있으며, 탐색의 방향은 무 학습 단계에서의 유인소(basins of attraction)들로 향하고 있음을 볼 수 있다. 완전한 국부 탐색을 수행시킨 후의 적합도 함수의 모형을 보면 유인소(국부 최적점) 근처에 있는 모든 개체들이 동일한 적합도 함수 값을 가짐을 알 수 있다. 즉, 국부 탐색의 수렴 성질에 의해 다중 모드(multi-modal) 함수가 일정한 격에 대한 상수 함수를 조합으로 변형됨을 볼 수 있다.

2.3 기존 학습 전략의 단점

라마키안 알고리즘을 사용한 국부 탐색 방법은 한 개체에 대한 국부 탐색 결과가 항상 결정적인 특정 결과로 수렴하는 특징을 갖는다. 즉, 개체의 해밍 거리를 이용한 국부 탐색에 의해 표현되어지지 않은 개체는 결과적으로 한번도 탐색 대상이 될 수 없음을 의미한다. 따라서, 최악의 경우 표본으로 구성된 모집단 내의 개체들이 최적화 하고자 하는 함수의 상태 공간을 적절히 표현하고 있지 못할 경우에는 현재 주어진 후보 해들로부터 얻을 수 있는 최적해가 전역 최적해가 되지 못하는 결과를 낳는다. 최적화하기 위한 함수의 상태 공간이 비교적 완만하거나 또는 이웃하는 목적 함수 값의 변화가 크지 않을 경우에는 라마키안 알고리즘은 좋은 성능을 보일 수 있을 것이다. 즉 단순 유전자 알고리즘보다 목적 함수의 평가 횟수는 상대적으로 증가하지만, 수렴 속도는 상당히 빠를 것이다. 반면에 볼드윈 효과에 의한 국부 탐색은 각 개체의 유전 형질의 변화를 유발하지 않으므로 모집단의 다양성을 비교적 잘 유지할 수 있는 학습 전략이다. 비교적 최적화 하고자 하는 함수의 상태 공간이 골짜기가 많은 경우에도 전역 최적해를 잘 찾을 수 있으나 원래 함수를 근사 시키기 위해 학습 과정에 필요한 세대수가 상당히 많다는 단점이 존재한다.

라마키안 진화 방법보다 볼드윈 효과를 이용한 탐색

알고리즘이 전역 해에 잘 수렴함을 볼 수 있는데[3], 볼드윈 효과에서의 수렴 속도의 저하는 학습 과정에서 변화되는 적합도 함수 값 값들 간의 차가 작기 때문이다 (유인소 근처에서 거의 같은 값을 가지게 됨).

본 논문에서는 전역 최적해로의 수렴도와 학습 속도를 향상시키기 위한 국부 탐색 방법으로서 적합도 함수 값 변화를 이용하여 국부 탐색을 수행하는 새로운 학습 전략을 제시하고자 한다.

다음 장에서는 본 논문에서 제안한 셀룰러 학습 전략의 기본이 되는 셀룰러 오토마타에 대해 기술하고 적합도 함수 값이 어떻게 변화하면서 학습에 반영되는지 설명한다.

### 3. 셀룰러 학습

#### 3.1 셀룰러 오토마타

모델링과 시뮬레이션 도구로 많이 사용되는 셀룰러 오토마타는 완전한 피드백 장치로서, 각 셀들의 상태가 단계적으로 변하는 유한 상태 장치이다[5]. P-상태 셀룰러 오토마타는 각 셀이 0에서 p-1의 값 중 하나의 값으로 표현되는 셀들의 구성이 선형이거나 면을 이루는 1,2차원 형태로 표현된다. 오토마타는 계층의 초기 상태와 규칙들의 집합으로 이루어져 있다. 여기서 규칙은 새로운 계층의 셀의 상태를 결정하는 방법을 기술한 것으로서 현 상태의 한 셀에 이웃하는 여러 셀들의 그룹에 의해 결정된다[7].

2차원 셀룰러 오토마타에서 이웃(neighborhood)은 다음 그림 3과 같은 한 개의 셀을 중심으로 4-이웃과 8-이웃을 이용한다. 4-이웃은 8-이웃 중에서 4방향에 해당하는 셀만으로 구성되며 이를 NWSE-이웃이라 부른다.

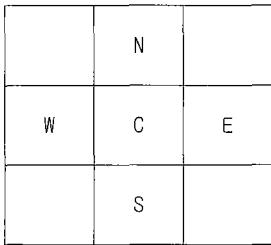


그림 3 오토마타의 4-이웃

그림 3에서 4-이웃에 해당하는 C는 중심 셀, E는 동쪽에 있는 셀, W는 서쪽에 있는 셀, S는 남쪽에 있는 셀, N는 북쪽에 있는 셀을 표시하는 레이블이다. 각 셀이 가질 수 있는 상태가 두 가지라고 가정한다면 각 셀

은 0 또는 1의 값을 갖게 되며 CSWNE 상태는 5개의 이진수의 조합에 의해 표현될 수 있다. 예를 들면 CSWNE = 11100D은 셀 C, 셀 S, 셀 W는 alive 상태이고 셀 N과 셀 E는 dead 상태임을 의미한다. 다음 단계에서 중앙 셀의 상태는 8개의 이웃 셀 또는 4개의 이웃 셀의 상태에 의해 결정되며, 각 셀들이 가능한 상태들을 조합하여 규칙을 만들 경우 4-이웃에 대해 다음 표 1과 같이 32개의 상태에 대한 규칙들을 구성할 수 있다[7].

표 1 4-이웃 상태에 대한 규칙 테이블

CSWNE	C	CSWNE	C	CSWNE	C	CSWNE	C
00000	0	01000	1	10000	1	11000	1
00001	0	01001	1	10001	1	11001	1
00010	0	01010	1	10010	1	11010	1
00011	0	01011	1	10011	1	11011	1
00100	1	01100	0	10100	1	11100	1
00101	1	01101	0	10101	1	11101	1
00110	1	01110	0	10110	1	11110	1
00111	1	01111	0	10111	1	11111	1

셀룰러 오토마타의 주기적 특성은 셀의 다음 단계에서의 상태가 자신의 상태와 이웃 셀의 상태에 의해 결정되는 생명 게임(Game of Life) 규칙에서 볼 수 있으며, 다음 상태의 셀들의 값은 다음과 같은 생명 게임 규칙에 의해 결정된다[7].

[규칙 1] 임의의 시점에서 alive인 셀은 그 셀의 8-이웃 중 2개 또는 3개의 셀이 alive인 경우 alive 상태로 살아 남는다.

[규칙 2] 임의의 시점에서 dead인 셀은 그 셀의 8-이웃 중 3개 이상의 셀이 alive인 경우만 alive 상태로 변한다.

규칙 1은 다음과 같은 경우에 셀이 dead 상태로 변화될 수 있다. 첫 번째 경우는 임의의 시점에서 alive인 셀에 대하여 그 셀의 8-이웃 중 4개 이상의 셀이 alive인 경우이며 이때 과다 밀집(overcrowdness)으로 인해 dead 상태로 변할 수 있다. 두 번째 경우로는 임의의 시점에서 alive인 셀에 대하여 그 셀의 8-이웃 중 1개 이하의 셀이 alive인 경우 외로움(loneliness)으로 인해 dead 상태로 변하게 되는 경우이다.

생명 게임 규칙이 단계적으로 진행되었을 때 셀 집단이 특정 양상을 나타냄을 볼 수 있으며, 주요 특징으로 패턴의 주기적 양상(periodic behavior)을 들 수 있다.

셀룰러 오토마타에서의 중심 셀은 자신의 상태와 이웃하는 셀들의 상태에 따라 다음 단계의 자신의 상태가 결정된다.

본 논문에서 제안하는 셀룰러 학습에서는 하나의 중심 셀이 alive 또는 dead 상태로 변하는 것이 아니라 현재 그 개체의 적합도 함수 값이 주위의 개체들의 적합도 함수 값에 의해 어떤 방향으로 변하는지에 대한 규칙을 정함에 있어서 셀룰러 오토마타에서의 생명 게임 규칙을 학습 전략을 기반으로 하였다. 즉, 생명 게임 규칙에서처럼 하나의 셀이 이웃 셀들에 의해 다음 단계(step)에서의 생존 여부가 가려지는 것이 아니라 셀에 해당하는 하나의 개체의 적합도 함수 값이 그 세대에서 어떤 값을 가지게 되는지의 결정을 이웃 크기 내에 있는 개체들(여기서는 가상의 개체들)에 의해 결정하는 것을 말한다. 유전자 알고리즘에서 다음 세대(generation)에서의 생존 여부는 선택 과정에서 결정되는 것이며 셀룰러 학습 전략은 한 개체의 이웃 크기 내에 있는 개체들의 적합도 함수 값을 기반으로 그 개체의 적합도 함수 값을 정하는 국부 탐색 알고리즘이다. 각 세대마다 이 규칙을 적용하게 되는 모집단을 구성하는 개체들이 선택 결과에 의해 달라지고, 따라서 각 개체의 적합도 함수 값에 영향을 미치는 이웃 크기 내에 있는 가상의 개체들도 달라지게 된다.

주기성을 가지게 되는 이유는 셀룰러 오토마타에서 단계가 진행됨에 따라 alive 또는 dead 상태의 셀들이 특정 패턴을 형성하듯이 셀룰러 학습이 적용될 대상이 되는 개체들의 적합도 함수 값 또한 세대가 진행됨에 따라 특정 값(여기서는 응용 문제에 적합한)으로 수렴하게 된다. 여기서 모집단을 구성하는 각 개체들의 적합도 함수값의 분포가 세대 진행에 따라서 주기적으로 어떤 일정 패턴을 형성하면서 변화해 나가는 모양을 볼 수 있는데, 이러한 현상으로부터 모집단 내의 개체들의 적합도 함수값의 분포가 주기성을 나타냄을 알 수 있다.

다수성(majority) 규칙은 세 개의 이웃하는 셀 그룹에 의해 한 셀의 새로운 상태가 셀 그룹의 투표(vote)에 의해 결정된다는 규칙을 기반으로 하는 오토마타의 수렴 특성을 잘 나타내는 규칙이다. 임의의 셀의 8-이웃 중 5개 이상의 셀의 상태가 같은 경우 그 셀은 여전히 같은 상태로 남게 되거나 이웃하는 셀의 상태로 변하게 되는 결과를 얻는다[7]. 즉, 8-이웃이 중심이 되는 셀이 이웃 셀들의 다수성을 결정하는 주된 역할을 한다고 볼 수 있다. 이 규칙의 특성은 임의의 상태 값으로부터 출발하여 일정 단계가 반복된 후에는 더 이상의 상태 변화를 보이지 않고 특정 패턴으로 수렴하게 됨을 볼 수

있다.

각 세대마다 모집단을 구성하는 개체들의 적합도 값은 셀룰러 학습에 의해 이웃 크기 내에 있는 개체들의 적합도 함수 값의 영향으로 국부 탐색을 수행하게 된다. 각 세대 별로 비교해 본다면 모집단을 구성하는 어떤 개체든지 그 개체의 적합도 함수 값이 특정 함수 값(여기서는 최대화 또는 최소화 응용에 따라서)으로 편향되어 가는 현상을 볼 수 있다. 이 현상은 일반 유전자 알고리즘에서의 수렴 현상과 비슷하지만 차이점은 다음과 같다. 먼저 일반 유전자 알고리즘은 선택 과정에서 적합도 함수 값이 높은 개체들을 그 다음 세대에 모집단 구성원으로 선택할 확률이 높음으로 인해 모집단을 구성하는 개체들이 특정 값으로 수렴하는 것이다(이 경우에는 개체들의 유전 형질도 서로 비슷하게 됨). 반면에 본 논문에서의 수렴 현상은 선택 압력에 의한 것이 아닌 각 개체가 그 세대에서 셀룰러 학습을 통하여 유전 형질은 그대로 보존하면서 적합도 함수 값이 변하는 것이므로 모집단을 구성하는 개체들에 대한 다양성이 유지될 수 있으며 또한 모집단을 구성하는 적합도 함수 값이 돌연 변이와 같은 특정 유전 인자의 변화에 의한 적합도 함수 값의 변화 이외에는 일정 세대동안의 학습과 진화를 거친 후에는 세대의 진행에도 더 이상 변화되지 않게 되는 것이다.

다수성 규칙은 다시 말하면 수렴 현상에 대한 규칙이라고 할 수 있다. 셀룰러 학습에 의한 적합도의 변화가 주변 셀(이웃 크기 내의 셀들)들의 영향에 의해 특정 값으로 수렴한다는 현상과 대응된다는 개념에서 다수성 규칙에 대응되는 특징을 셀룰러 학습에서 찾아볼 수 있다.

생명 게임 규칙에 의한 셀룰러 오토마타는 일정 단계 동안 규칙들이 반복적으로 적용되었을 때 일련의 셀 그룹이 주기적인 패턴을 나타내는 양상을 보인다. 다수성 규칙이 적용된 셀룰러 오토마타는 단계가 진행됨에 따라 특정 패턴으로 수렴함을 볼 수 있다. 본 논문에서는 오토마타의 이러한 특징들을 국부 탐색 과정에서 적합도 함수의 변이에 반영함으로써 함수의 수렴을 유도하고자 한다.

### 3.2 셀룰러 학습

셀룰러 학습은 셀룰러 오토마타에 의해 생성되는 패턴의 주기성과 수렴 성질을 현 상태의 적합도 함수 값을 변화시키는 데 적용한 학습 전략이다. 셀룰러 학습은 셀룰러 오토마타의 격자 공간에 대한 개념을 모집단을 구성하는 개체들에 대하여 각 개체를 기준으로 현 세대에서의 이웃 크기에 내 존재할 수 있는 4-이웃의 가상 개체들을 통하여 학습을 해 나가는 국부 탐색 방법이다.

중심 셀에 해당하는 개체는 모집단을 구성하는 개체들이 되고 이 개체의 적합도 함수 값에 영향을 주게되는 개체들은 이웃 크기 범위에 속하는 균등 분할 4-이웃 (함수 최적화의 경우 입력 매개 변수  $x_1, x_2$ 에 대한 이웃 크기 범위 내에서의 균등 위치 값이 됨)이 형성되고 각각에 대한 적합도 함수 값은 현재 최적화시키고자 하는 함수에 의해 계산하는 방식으로 학습을 수행한다. 여기서 4-이웃 내에는 실제 모집단 내에 존재하는 개체들이 포함되며 또한, 실제 이웃 크기 내에 존재 가능한 개체들도 포함되도록 한다. 이웃 크기는 절대 거리의 개념으로 정한 것이므로 셀룰러 학습을 적용시키는 과정에서 이웃 크기 내에 실제로 포함되는 개체들이 자기 자신을 제외하고 전혀 존재하지 않을 경우는 이웃 크기로부터 생성될 수 있는 가상 개체와 또한 그에 대응되는 적합도 함수값을 적용시킴으로써 셀룰러 학습을 행하도록 하였다.

다음은 셀룰러 학습을 적용하기 위한 전처리 단계로서 경계 초기화 기법과 부\_모집단(sub\_population) 구성에 대해 기술한다.

3.2.1 경계 초기화

탐색 공간의 전체적 특성을 반영하기 위해 적합도 함수 값의 상관 계수(correlation coefficient) 값에 대한 통계적 성질을 이용해서 탐색 공간의 굴곡 정도(ruggedness)를 측정한다. 굴곡 정도가 큰 탐색 공간은 이웃 점들과의 관련성이 적음을 나타내고, 굴곡 정도가 작은 탐색 공간(완만한 탐색 공간)은 이웃 점들과의 관련성이 많음을 의미한다. 굴곡 정도가 심한 탐색 공간에서의 원거리 이동(long jump) 또는 균일 학습 기법의 적용은 정보의 손실(적합도에 대한)을 야기할 수 있으므로, 탐색 공간의 굴곡 정도에 따라 적용적으로 학습을 해 나가는 것이 바람직하다. 굴곡 정도를 측정하는 방법으로는 정보 이론과 물리학 분야에서 많이 사용하는 상관 계수 값인  $\rho[h, \Delta f]$ 가 있다[6]. 여기서  $h$ 는 상태 공간 상의 두 점  $p_1, p_2$ 간의 해밍 거리(hamming distance)이고,  $\Delta f = |f(p_1) - f(p_2)|$ 이다. 이때,  $\rho[x, y]$ 는 다음 식 (1)과 같이 계산한다.

$$\rho[x, y] = \frac{\overline{(x-x)(y-y)}}{\sigma_x \sigma_y} \quad (1)$$

이  $\rho[x, y]$  값이 클 경우, 원거리 이동은 적합도 값의 변화를 크게하며, “근거리 이동(small jump)”은 적합도 값의 변화를 줄여준다. 탐색 공간의 전체적 특성에 대한 척도(measure)는 공간상의 임의의 한 점에서 국부 최적점으로 가는 언덕 오르기거리(hill-climbing walk)의 평균값이다. 이 값은 상태 공간의 최적값들의 수와는 반비

례하는 특성을 갖는다. 예상 거리(walk length)의 기대 값( $R$ )의 상한값(upper bound)은  $D/2$ 이다(여기서,  $D$ 는 공간의 차원이고,  $R_{max} = 30/2 = 15$ ). 참고 문헌 [6]에서 Kauffman은 완전한 임의 상태 공간상의  $R$  값이  $\log_2(D-1)=4.9$ 임을 증명하였다. 전체 탐색 공간에 대한 상관 계수 값의 계산을 위하여 모집단을 구성하는 초기해 개체들을 출발점으로 하여 그 개체와 가장 가까운 국부 최적점(local optimum)까지 힐 크라이밍(hill-climbing) 국부 탐색을 적용한다. 그 다음 국부 최적점에 도달 할때까지 변화하는 각 탐색 공간 상의 개체들에 대한 적합도 함수값들의 대한 평균값을 측정함으로써 전체 탐색 공간에 대한 상관 계수값을 얻는다.

본 논문에서는 상관 계수를 이용하여 학습이 적용되는 단위인 섬(island)의 개수를 결정하고 각 섬에 대한 초기화를 수행하는 경계 초기화 단계를 행한다. 세대가 진행됨에 따라 셀룰러 학습의 결과로 함수의 상태 공간의 굴곡 정도를 나타내는 값이 커지므로(상관 계수 값은 굴곡 정도와 반비례하므로), 섬의 크기는 커지게 되고 반대로 섬의 개수는 감소하게 된다.

경계 초기화 기법은 임의의 초기화에 의한 표본화 오류를 줄이고 셀룰러 학습이 수행될 수 있는 섬을 구성하기 위한 방법으로서 초기 모집단을 구성할 때 구간별로 임의로 초기화를 행하는 방법이다.

다음 그림 4는 2차원 함수에서 섬의 개수가 4일 경우의 정의 구역  $x$ 에 대한 구간 표시와 그에 대응되는 개체의 표현형과 경계 초기화의 예를 보여주는 그림이다.

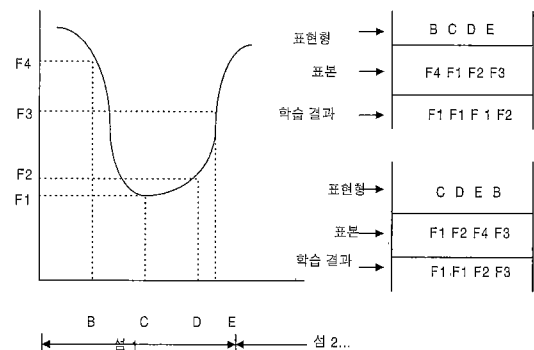


그림 4 섬의 크기가 4인 경우의 경계 초기화 예

그림 4에서 B,C,D,E를 포함하는 각 구간은 셀룰러 유전자 알고리즘의 고립 간격(isolate distance)에 의해 분리되는 섬에 대응되며, 각 섬 내에서는 이웃 크기에 따라 셀룰러 학습을 수행한다.

3.2.2 부\_모집단 구성

경계 초기화된 모집단에 대해 유전 연산이 수행된 후 전체 모집단은 다음과 같은 두 개의 부 모집단을 구성하게 된다. 부\_모집단 1(sub\_population 1)과 부\_모집단 2(sub\_population 2)의 구성은 다음 그림 5와 같다.

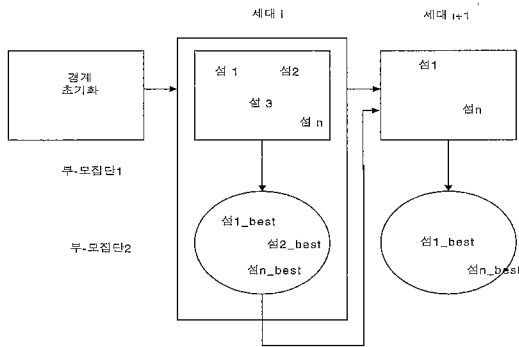


그림 5 부\_모집단의 구성

부\_모집단 1의 구성 요소는 경계 초기화에 의해 만들어진 섬들이다. 각 섬에서 가장 높은 적합도를 갖는 개체만이 부\_모집단 2로 보내질 수 있으며, 한 세대가 진행된 후 각 섬에서 최대 적합도 값을 갖는 개체가 이전 세대의 최대 적합도 값을 갖는 개체(부\_모집단 2에 포함되어 있는)보다 그 값이 높을 경우, 부\_모집단 2를 구성하는 개체들이 새로운 개체로 대체된다. 그림 3.3은 각 세대 i와 i+1에 대한 각 섬n에 대한 처리 과정을 보여주는 그림이다.

부\_모집단 내의 섬의 개수는 각 섬의 성능 향상률이 높아짐에 따라서 그 크기가 커지며, 반면에 섬의 수는 줄어들게 된다. 그 이유는 각 섬이 이웃하는 섬과의 성능 향상률 및 적합도 값의 비교를 통하여 하나의 섬으로 합쳐지기 때문이다. 각 부\_모집단의 구성과 역할은 다음과 같다.

1) 부\_모집단 1

초기의 전체 탐색 공간에 대한 상관 계수를 측정하고 이 값에 의해 섬의 크기를 결정한다. 이웃 크기를 1로 초기화하고, 초기화된 각 섬 내에서 셀룰러 학습을 수행한다. 이 중 가장 높은 적합도 함수 값을 갖는 개체를 그 섬의 대표로 결정한다. 이때, 개체 자체의 유전 형질은 변화시키지 않도록 하며 각 섬의 대표를 부\_모집단 2로 보낸다.

2) 부\_모집단 2

부\_모집단 2는 부\_모집단1로부터 만들어진 각 섬을 대표하는 개체들로 구성된다. 각 섬의 개체들은 다음 세대

로 전달될 때 전 세대에서 셀룰러 학습 전략에 의해 학습한 내용을 적합도 함수에 반영함으로써 전 세대에 개체 자체의 적합도가 낮더라도 학습된 결과에 의해 그 향상 정도가 높을 경우에는 부\_모집단 2에 포함되도록 한다. 부\_모집단 2를 구성하는 개체들은 세대가 바뀔 경우 각 섬의 이전 세대를 대표하는 개체의 적합도와 비교하여 더 높은 적합도를 갖는 개체가 계속 남아 있을 수 있도록 한다. 부\_모집단 2를 구성하는 개체들에는 학습에 의해 자체의 유전 형질은 그대로 유지하면서 적합도 값의 향상을 갖는 개체들도 포함되게 함으로써 선택 압력에 의해 발생하는 성급한 수렴(조기 수렴) 현상을 줄일 수 있다. 또한, 각 섬을 대표하는 개체들간의 유전 연산을 행함으로써 전체 상태 공간에 대한 전역 탐색 효과를 얻게된다.

3.2.3 셀룰러 학습 전략

셀룰러 학습은 개체의 유전 형질은 변화시키지 않고 향상된 정도를 적합도 함수 값에 적용시킨 학습 전략이다. 각 단계마다 규칙에 따라서 한 개체의 적합도 함수값은 이웃 크기 내에 존재하는 각 개체 중 가장 높은 적합도 함수 값을 갖는 개체의 적합도 함수 값으로 대체된다. 이 규칙은 이웃 크기 내에 속하는 여러 개체들의 적합도 함수 값에 의해 영향을 받으므로, 항상 고정된 값을 산출하지 않고 이웃 크기가 변화에 따라 적응적으로 그 값이 변화한다. 셀룰러 학습 전략은 다음 그림 6과 같다.

```

S = ∪_{i=1}^n {Island(I_i)} // S : 섬의 전체 집합
B = { } // 집합 S의 n개의 섬으로부터의
           best_solution의 집합
           // max_island_number : 최대 섬의 수
           // max_island_size : 섬의 크기
for i = 1 to max_island_number {
  for j = 1 to max_island_size {
    S'_i = optimum(s_{i-1}^j, s_i^j, s_{i+1}^j)
           // s_i^j의 이웃 크기내에서의 최적값
  }
  b_i = best_solution in s'_i
  B = B ∪ { b_i }
}
for i = 1 to max_island_number {
  I_i ← s'_i
}
reduced_number = 0
for i = 1 to max_island_number {
  if ( | b_i - b_{i+1} | < α ) {
    I_i = { I_i } ∪ { I_{i+1} }
    reduced_number = reduced_number + 1
  }
}
max_island_number ← max_island_number -
reduced_number
// α: threshold
    
```

그림 6 셀룰러 학습 전략

이웃하는 점에 대해 각 점을 대표하는 개체의 적합도 함수 값의 차가 임계값보다 작을 경우, 이웃하는 두 점은 한 개의 점으로 합쳐지게 된다. 따라서 학습이 진행되어 감에 따라 점의 개수는 점차 감소하며 또한 그 크기는 커지게 된다. 그 효과로는 하나의 점으로 합쳐진 개체들은 두 점 내의 개체들 중에서 최대값(최대화 문제의 경우) 또는 최소값(최소화 문제의 경우)으로 수렴하게 됨을 볼 수 있다.

셀룰러 학습에 의한 국부 탐색 방법의 장점은 다음과 같이 요약할 수 있다. 첫째, 개체의 형질에 영향을 주지 않으므로 학습이 진행되어도 개체의 다양성 유지가 가능하며, 선택 압력에 의한 국부 최적값(local optimum)에 빠지는 현상을 막을 수 있다. 둘째, 이웃 크기 내에서의 학습은 셀룰러 학습이 수행하는 단계 반복 개념을 이웃 크기의 증가로 대체할 수 있기 때문에 기울기 강하 기법에 의한 학습보다 학습 속도의 향상을 유도한다. 셋째, 전체 탐색 공간에 대한 균등 탐색을 가능하게 하는 심분할에 의한 학습은 이웃 크기 내에서의 학습을 통한 지역적 탐색(exploitation)과 동시에 각 점을 대표하는 개체들간의 유전 연산을 통한 전역 탐색(exploration)의 효과를 적절히 결합시킬 수 있다.

3.2.4 셀룰러 학습의 효과

셀룰러 오토마타가 지니는 주기성과 수렴성을 학습에 적용하기 위하여 먼저, 이웃 크기(n)와 적용될 규칙을 정해야 한다. n값(이웃 크기 :  $n = 2b+1$ )은 초기에 가장 작은 이웃 크기를 형성하는 b값이 1인 경우이다(이때 n값은 3이 된다). n값은 셀룰러 학습이 진행되면서 성능 향상도에 따라 증가하게 된다. 그 이유는 셀룰러 학습이 각 개체가 주위 환경(이웃 크기 내에 존재하는 개체들)을 고려하여 다음 단계에서의 자신의 함수값을 얻게 되기 때문이다. 즉, 학습이 진행됨에 따라 이웃과의 연관성으로 점점 서로 유사한 값을 나타내게 되며, 따라서 점의 크기도 커진다. 결국 학습이 수행된 후 각 점은 일정 형태로 수렴함을 볼 수 있다.

이러한 방법으로 결정된 이웃 크기에 대한 셀룰러 학습을 위해 적용되는 규칙은 이웃 크기내에 존재하는 개체들 중에서 가장 높은 적합도 함수값을 갖는 개체의 적합도 함수값이 다음 단계에서의 그 개체의 적합도 함수값에 적용되도록 하는 것이다. 따라서 이 규칙은 이웃 크기내에 속하게 되는 개체에 따라 다른 값들을 결과로 산출하게 되며, 항상 고정되어 있는 것이 아니라 이웃 크기의 변화에 따라 영향을 받게 되며 현재 이웃 크기내에 속해 있는 개체들에 적용적으로 값을 산출해 낸다. 단, 이러한 셀룰러 학습이 진행되기 위해서는 경계 초기화되어야 한다.

왜냐하면 일반 유전자 알고리즘에서 사용하는 임의 초기화의 경우는 이웃 크기를 정했을 때 셀룰러 학습이 행하는 바로 이웃하거나 또는 어떤 일정 거리(범위) 내에서의 단계 반복이라는 개념을 적용할 수 없기 때문이다.

3.3 셀룰러 학습 기반 유전자 알고리즘

본 논문에서 제안하는 유전자 알고리즘은 단일 교차 연산과 wheel of fortune 선택 메커니즘을 사용하였으며 다음과 같은 9개의 요소로 구성되어 진다. 각 요소에 대한 설명은 다음 그림 7과 같다.

$$AGA = (P^0, \lambda, l, s, \Omega, f, t, b, I)$$

$P^0 = (a_1^0, \dots, a_n^0) \in I^l$	$I = \{0,1\}^l$	: 초기 모집단
$\lambda \in N$		: 모집단 크기
$l \in N$		: 개체의 길이
$s : I^l \rightarrow I^l$		: 선택 연산자
$\Omega \subseteq \{ \omega : I \times I^l \rightarrow P \rightarrow I \}$		: 유전 연산자 (여기서, $P = \{ p : I \rightarrow [0,1] \}$ )
$f : I \rightarrow R$		: 적합도 함수
$t : I^l \rightarrow \{0, 1\}$		: 종결 조건
$i \in [ 1 .. l/2 ]$		: 점 크기 제어 변수

그림 7 유전자 알고리즘의 구성 요소

임의 초기화와 상관 계수 값에 의해 경계 초기화를 적용한 후 셀룰러 학습 및 유전 연산을 수행하는 전체 알고리즘을 단계별로 기술하면 다음과 같다.

- [단계 1] 임의 초기화  
초기 모집단을 임의로 초기화 시킨다.
- [단계 2] 상관 계수 측정  
단계 1에서 구성한 모집단에 대한 상관 계수를 구한다.
- [단계 3] 경계 초기화  
단계 2로부터 얻은 상관 계수에 의해 점의 개수를 결정하고 실제 부-모집단 1을 구성한다.
- [단계 4] 셀룰러 학습  
이웃 크기값을 1로 초기화시킨 후 부-모집단 1에 대한 셀룰러 학습을 행하고 각 개체의 실제 적합도 함수 값을 계산한다.
- [단계 5] 적합도 확률 분포에 의한 선택
  - i) 각 개체  $v_i$  ( $i = 1, \dots, pop\_size$ ) 에 대한 적합도를 계산한다.  
 $eval(v_i) = f(x), i = 1, \dots, pop\_size$
  - ii) 모집단의 전체 적합도를 구한다.  
 $F = \sum_{i=1}^{pop\_size} eval(v_i)$
  - iii) 각 개체  $v_{ii}$  ( $i = 1, \dots, pop\_size$ ) 에 대해 선택 확률  $P_i$ 를 계산한다.



$$F_i = eval(v_i)/F$$

iv) 각 개체  $v_i$  ( $i = 1, \dots, pop\_size$ ) 에 대한 누적 확률 분포값  $q_i$ 를 계산한다.

$$q_i = \sum_{j=1}^i P_j$$

v) 난수 발생기에 의해 구간  $[0 .. 1]$  사이의 값  $r$ 에 대해,  $r$ 값이  $q_i$  값보다 작으면 첫 번째 개체( $v_i$ )를 선택하고, 그렇지 않을 경우에는 다음 조건을 만족하는 개체  $v_i(2 \leq i \leq pop\_size)$ 를 선택한다.

$$q_{i-1} < r \leq q_i$$

[단계 6] 돌연 변이

개체  $v_i^t = \langle u_1, \dots, u_n \rangle$ 에 대해서 돌연 변이 확률  $P_m$ 에 의해 돌연 변이 연산을 적용한다. 난수 발생기에 의해 얻은 구간  $[0 .. 1]$  사이의 값  $r$ 에 대해  $r$ 값이 돌연 변이 확률  $P_m$ 보다 작을 경우 돌연 변이 연산이 수행된다. 연산이 적용된 후의 개체는 다음과 같이 표현되며, 여기서 개체를 구성하는 위치  $k$ 에 있는 유전 형질이  $u_k'$ 로 바뀐 결과이다.

$$v_i^t = \langle u_1, \dots, u_k', \dots, u_n \rangle$$

[단계 7]부-모집단 1과 부-모집단 2 구성

각 설을 대표할 수 있는 최대 적합도 값을 갖는 개체들만 부모집단 1로부터 추출하여 부-모집단 2를 구성한다.

[단계 8]부-모집단 2에 대한 유전 연산

부-모집단 2에 대해 단일 교차 연산을 수행한다.

[단계 9] 종료 조건

성능 향상이 거의 없거나 최대 세대에 도달한 경우에 종료하고 아니면 단계 4로 간다.

#### 4. 실험 및 성능 평가

이 장에서는 제안하는 셀룰러 학습 전략의 성능을 기존의 국부 탐색 방법인 라마키안 진화와 볼드윈 효과와 비교하여 평가하고자 한다. 최적화 실험에 많이 사용되

는 드중과 쉼퍼에 의해 제안된 테스트 베드 함수들[4]을 대상으로 실험을 행하였다. 실험에 사용된 함수들은 다음 표 2와 같다. 유전자 알고리즘에 사용된 매개 변수들의 값은 다음 표 3과 같다(함수 F3에서 V값은 8000임).

표 3 실험에 사용된 매개 변수

매개 변수	값
교차 연산률 :	0.6
돌연 변이률 :	0.05
모집단 크기	20
개체의 길이	33
최대 세대 횟수	30
종결 조건	최대 세대수 또는 성능향상률의 변화량

수렴도를 측정하는 방법 중의 하나로서 최적해로 수렴해 나가는 과정에서 모집단 내의 평균 적합도 함수 값(Favg)과 최대 적합도 함수 값(Fmax)을 이용하여 수렴도를 측정할 수 있다[2]. 각 세대에 대한 Fmax와 Favg값을 계산하면 최적 값(최대값 또는 최소값)으로 가까워질수록 그 계산 값이 작아지는 현상을 이용하여 실험 함수에 대한 수렴 곡선을 구하는 것이다[2].

평균 적합도 함수 값은 세대가 진행됨에 따라 모집단 내의 개체들이 더 높은 적합도 함수 값을 얻게 되므로 그에 비례하여 높아지며, 각 세대에 대한 가장 높은 적합도 함수 값(Fbest)을 가지는 개체는 다음 세대에도 그대로 유지되도록 하였으므로 여기서, 세대 진행에 따른 Fbest, Fmax-Favg, Favg값의 변화를 살펴보는 것도 성능을 비교할 수 있는 중요한 척도가 될 것이다. 다음 그림 8과 그림 9는 라마키안 진화와 볼드윈 효과에 의해 학습을 결합시킨 유전자 알고리즘의 수렴도를 측정하기 위한 테스트 결과로서 함수 F1에 대한 Fbest, Fmax-Favg, Favg값의 분포를 보여주는 그림이다.

표 2 테스트 함수

함수 번호	함수	변수 범위	참고 문헌
F1	$f_1(x_1, x_2) = 21.5 + x_1 \sin(4\pi x_1) + x_2(20\pi x_2)$	$-3.0 \leq x_1 \leq 12.1$ $4.1 \leq x_2 \leq 5.8$	[8]
F2	$f_2(x_i) = 200 + \sum_{i=1}^{20} x_i^2 - 10 \cos(2\pi x_i)$	$-5.12 \leq x_i \leq 5.11$	[3]
F3	$f_3(x_i) = V + \sum_{i=1}^{20} -x_i \sin(\sqrt{ x_i })$	$-5.12 \leq x_i \leq 5.11$	[3]
F4	$f_4(x_i) = \sum_{i=1}^5 integer(x_i)$	$-5.12 \leq x_i \leq 5.11$	[4]
F5	$f_5(x) = \sum_{i=1}^{30} ix_i^4 + Gauss(0, 1)$	$-1.28 \leq x_i \leq 1.28$	[4]

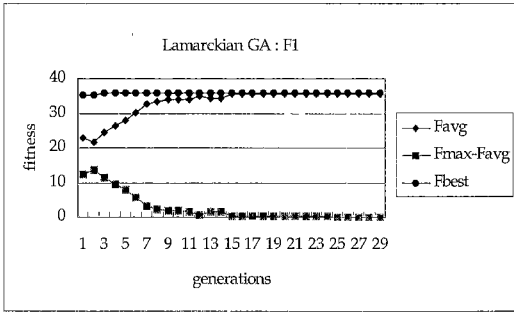


그림 8 라마키안 진화에 의한 Fmax-Favg, Favg 값의 분포 : F1

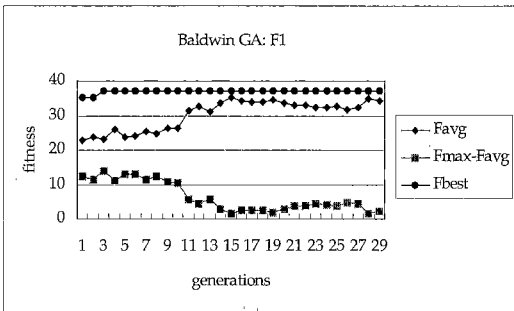


그림 9 볼드윈 효과에 의한 Fmax-Favg, Favg 값의 분포 : F1

실험 결과에서 알 수 있듯이 라마키안 진화에 의한 방법이 볼드윈 효과에 의한 방법보다 수렴 속도가 월등히 빠름을 볼 수 있다. 라마키안 알고리즘에 의한 학습 결과를 보면 15세대 이후부터는 최대 적합도 함수 값이 모집단의 평균 적합도 함수 값과 거의 유사해 짐을 볼 수 있다. 이로부터 각 세대에 대한 국부 탐색이 전체 모집단 내의 개체의 적합도 함수값을 거의 균일하게 유지되도록 함으로써 최적해로 수렴하는 속도가 매우 빠름을 알 수 있으나 반면에 전역 해에 도달할 수는 없었음을 볼 수 있다. 볼드윈 효과에 의한 방법은 수렴 속도가 다소 느리지만 모집단의 평균 적합도 함수의 값의 분포로부터 알 수 있듯이 개체들의 다양성 유지와 함께 전역해로 수렴함을 볼 수 있었다.

다음 그림 10은 셀룰러 학습에 의한 Fbest, Fmax-Favg, Favg값의 변화를 관찰한 실험 결과이다. 각 값의 분포로부터 모집단의 수렴 정도가 볼드윈 효과보다 빠름을 알 수 있고, 전역 최적 해에 도달하였음을 볼 수 있다. 또한 평균 적합도 함수 값(Favg)의 분포 곡선이 변화가 작은 완만한 증

가를 보이므로 모집단내의 개체들의 다양성이 세대가 진행됨에 따라 비교적 잘 유지되고 있음을 알 수 있다.

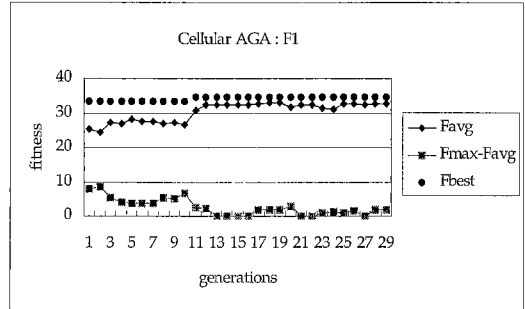


그림 10 셀룰러 학습과 AGA에 의한 Fmax-Favg, Favg 값의 분포 : F1

다음 그림 11과 그림 12는 각 알고리즘의 세대가 진행됨에 따른 각 세대에서의 Fbest값의 변화와의 변화를 보여주는 그림이다. 볼드윈 효과와 셀룰러 학습에 의한 탐색 결과에 대한 Favg값이 상대적으로 낮게 나타나는 이유는 탐색이 진행되는 동안 개체의 다양성이 잘 유지되고 있음을 의미한다.

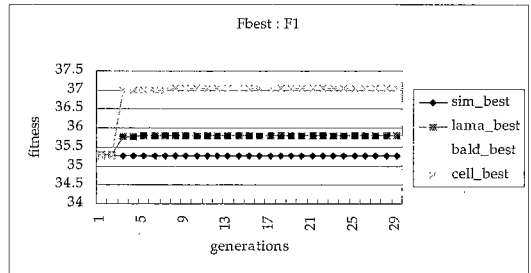


그림 11 Fbest 값의 분포

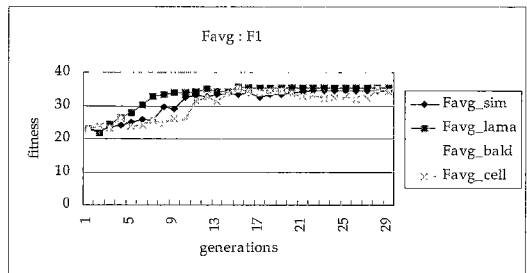


그림 12 Favg 값의 변화

다음 표 4은 함수 F2, F3에 대해서 각 알고리즘을 30 회 수행시켰을 때 전역 최적 해를 찾은 횟수를 나타내는 표이다. 함수 F2에 대해서는 비교적 잘 수렴되었지만 함수 F3에 대한 실험 결과를 보면 단순 유전자 알고리즘은 최적 해를 찾지 못하였으며 라마키안 알고리즘은, 최적해 수렴 확률이 볼드윈 효과와 셀룰러 학습에 의한 결과보다 낮음을 볼 수 있다. 함수 F2에 대해서는 실험 함수의 상태 공간이 국부 최적 점을 비교적 적게 포함하는 특성으로 인하여 라마키안 알고리즘도 좋은 성능을 보였는데, 이로부터 상태 공간에 대한 정보를 미리 알고 있으면서 빠른 학습 효과를 요구하는 문제에 대해서는 라마키안 학습도 비교적 좋은 결과를 제공하여 줄 수 있으리라는 것을 예측할 수 있다.

표 4 테스트 함수의 수렴 횟수 비교

함수 번호	SGA	LGA	BGA	CGA
F2	0	30	30	30
F3	0	9	19	20

그림 13은 세대 진행에 따른 단순유전자 알고리즘과 볼드윈, 셀룰러 학습 전략에 의한 적합도 함수 값의 변화를 보여주는 그림이다(최소화 문제 테스트 함수 F4,

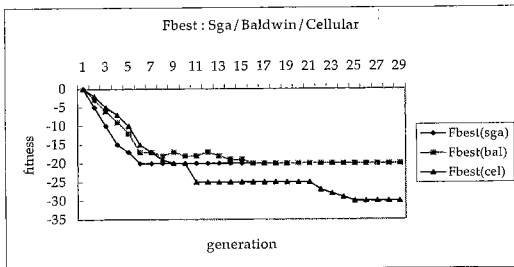


그림 13 최적해 분포(F4)

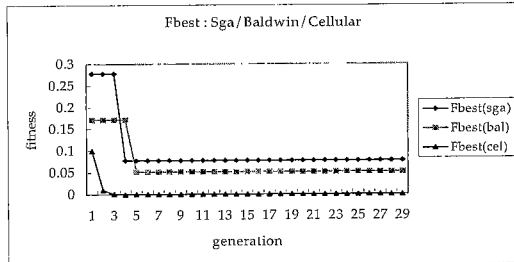


그림 14 최적해 분포(F5)

F5). 그림에서 볼 수 있듯이 셀룰러 학습 전략이 단순 유전자 알고리즘이나 볼드윈 효과를 적용한 알고리즘보다 학습 속도면에서 빠르고 최적 해에 더 빨리 도달함을 알 수 있다. 또한 그림 14은 볼드윈 알고리즘과 단순 유전자 알고리즘이 테스트 함수 F5에 대해서는 거의 수렴하는 형태를 보임에도 불구하고 전역 최적 해에 수렴하지 못함을 알 수 있다.

### 5. 결론

셀룰러 학습은 수렴 속도와 전역해로의 수렴 확률을 높이기 위해 부분 탐색을 수행하는 학습 전략이다. 이 방법은 분할된 각 섬에 대해 학습을 수행하는데, 모집단을 초기화시키는 단계에서 인의로 초기화를 행하지 않고 전체 상태 공간의 상관 계수를 측정한 후 상관 계수 값에 따라 적절하게 전체 상태 공간을 섬으로 분할 한 후 경계 초기화를 수행함으로써 표본화 오류를 줄인 학습 방법이다. 상관계수 값은 전체 상태 공간의 특징을 나타내는 값으로서 그 절대값이 클수록 모집단 내의 개체들이 밀접하게 연관되어 있음을 의미하므로 따라서 다양성을 유지할 수 있는 척도로 쓰일 수 있다. 세대 진행에 따른 모집단의 다양성 유지 정도를 알아보기 위해 상관 계수 값을 각 섬에 대해 측정해 보았으며, 그 결과로부터 셀룰러 학습에 의한 알고리즘이 라마키안과 볼드윈 방법보다 월등히 다양성 유지에 효과적임을 실험으로 증명하였다.

본 논문에서 제안하는 학습 전략은 복잡한 상태 공간을 나타내는 함수를 완만한 형태의 공간으로 근사 시켜 나가면서 최적해로 수렴할 수 있도록 해주는 학습 전략이다. 이때, 개체의 형질을 그대로 보존시키고 적합도 함수 값만을 이웃하는 개체의 값에 의해 보정해 나감으로써 모집단의 다양성을 유지하여 지역 최적 해에 수렴할 확률을 줄일 수 있었다. 또한 근사화시키는 학습 전략에 의해 전역 최적해를 찾는 수렴속도를 꺾을 수 있었다.

본 논문에서 제안하는 알고리즘은 다음과 같이 요약할 수 있다.

- 1) 유기체가 그 세대에 획득한 지식, 경험 및 환경의 변화에 대한 내용을 유전 형질은 변화시키지 않으면서 다음 세대로 물려줄 수 있는 셀룰러 학습 전략을 제안하였다.
- 2) 골목이 심한 상태 공간을 갖는 함수의 최적화 문제에 대해서 모집단의 다양성 유지와 함께 근사 상태 공간을 형성하는 방법을 통하여 전역 최적 해로의

수렴 확률을 높일 수 있다.

- 3) 상태 공간을 상관 계수 값에 의해 섬으로 구분한 후 초기화시킴으로써 모집단의 임의 초기화에서 발생할 수 있는 표본화 오류를 줄일 수 있다.
- 4) 기존 학습 전략에 비해 학습 속도의 개선을 꾀하였다.

앞으로의 연구 과제로는 셀룰러 학습 전략의 다중 최적화 문제로의 적용과 셀룰러 학습 기법의 탐색 공간상의 변이에 대한 기존 학습 전략간의 비교에 대한 연구이다.

### 참 고 문 헌

- [1] M. Srinivas and L. M. Patnaik, "Genetic Algorithms: A Survey," IEEE Trans. Computers, June, 1994, pp. 17-26.
- [2] M. F. Bramlette, "Initialization, Mutation and Selection Methods in Genetic Algorithms for Function Optimization," Proc. 4th Int. Conf. on Genetic Algorithms, Morgan Kaufmann Pub., 1991, pp. 128-135.
- [3] D. Whitley, V. S. Gordon and K. Mathias, "Lamarckian Evolution, The Baldwin Effect and Function Optimization," Parallel Problem Solving from Nature-PPSN III, Springer-Verlag Pub., 1994, pp. 6-15.
- [4] D. E. Goldberg, Genetic Algorithms in Search Optimization and Machine Learning, Addison Wesley Pub., 1989.
- [5] F. Gruou and D. Whitley, "Adding Learning to the Cellular Development of Neural Networks: Evolution and the Baldwin Effect," MIT Journal on Evolutionary Computation, Vol. 1, No. 3, 1997, pp. 213-233.
- [6] M. Lipsitch, "Adaptation on Rugged Landscapes generated by Iterated Local Interactions of Neighboring Genes," Proc. 4th Int. Conf. on Genetic Algorithms, Morgan Kaufmann Pub., 1991, pp. 128-135.
- [7] H. O. Peitgen, H. Jurgen and D. Sque, Chaos and Fractals-New Frontiers of Science, Springer-Verlag New York, Inc., 1992.
- [8] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, Springer-Verlag Pub., 1992.
- [9] D. Whitley, "Modeling Hybrid Genetic Algorithms," Genetic Algorithms in Engineering and Computer Science, John Wiley Pub., 1995, pp. 191-201.
- [10] M. Mitchell, An Introduction to Genetic Algorithms, MIT Press, 1996.
- [11] M. Syrakow and H. Szczerbicka, "Combination of Direct Global and Local Optimization Methods," IEEE Int. Conf. on Evolutionary Computation, Vol. 1, 1995, pp. 326-333.
- [12] F. Vavak, K. Jukes and T. C. Fogarty, "Learning the Local Search Range for Genetic Optimization in Non-stationary Environments," IEEE Int. Conf. on Evolutionary Computation, Vol. 1, 1995, pp. 355-360.
- [13] A. Adamatzky, Identification of Cellular Automata, Taylor & Francis Pub., 1994
- [14] B. H. Voorhees, Computational Analysis of One-dimensional Cellular Automata, World Scientific Pub., 1996.
- [15] L. Davis, Genetic Algorithms and Simulated Annealing, Pitman Pub., 1987.
- [16] T. M. Mitchell, Machine Learning, The McGraw-Hill Companies, inc., 1997.



고 명 숙

1989년 이화여자대학교 물리학과 졸업(학사). 1993년 고려대학교 대학원 컴퓨터학과 졸업(이학석사). 1998년 고려대학교 대학원 컴퓨터학과 졸업(이학박사). 1999년 ~ 2000년 고려대학교 정보통신 기술공동연구소 연구조교수. 2001년 ~ 현재 부천대학 전산정보처리과 교수. 관심분야는 퍼지 및 신경망 이론, 유전자 알고리즘, 셀룰러 오토마타, 인공 생명 등임



길 준 민

1994년 고려대학교 자연과학대학 전산학과 졸업(학사). 1996년 고려대학교 대학원 전산학과 졸업(이학석사). 2000년 고려대학교 대학원 컴퓨터학과 졸업(이학박사). 관심분야는 퍼지 시스템, 유전자 알고리즘, 셀룰러 오토마타, 이동 컴퓨팅 등임