

# PfSGA를 이용한 MLP 분류기의 구조 학습

愼 晟孝, 金 商雲

明知大學校 컴퓨터工學科

## A Structural Learning of MLP Classifiers Using PfSGA

Seong-Hyo Shin, Sang-Woon Kim

Dept. of Computer Engineering, Myongji University

shshin@ce.myongji.ac.kr

### Abstract

We propose a structural learning method of MLP classifiers for a given application using PfSGA (parameter-free species genetic algorithm), which is a combining of species genetic algorithm(SGA) and parameter-free genetic algorithm(PfGA). Experimental results show that PfSGA can reduce the learning time of SGA and has no influence of parameter values on structural learning. And we also convince that PfSGA is more efficient than the other methods in the aspect of misclassification ratio, learning rate, and complexity of MLP structure.

### 1. 서론

다층신경망(multi-layered perceptron; MLP) 분류기의 학습에 큰 영향을 미치는 위상 구조 중 입력층, 은닉층, 출력층 뉴런의 수(각각  $I, H, O$ )에서  $H$ 는 설계자의 경험에 의하여 결정하고 있다.

유전 알고리즘(genetic algorithm; GA)을 이용하여 적합한  $H$ 를 결정하는 방법에는 MLP의 가중치만을 구하는 방법[1], 위상 문자열과 가중치 문자열을 동시에 구하는 방법[2], 꼭 필요한 링크들만을 갖는 선택적 연결의 MLP를 구하는 방법[3] 등이 있다. 그러나, 이 방법들은 가중치뿐만 아니라 은닉층 뉴런의 수 및 링크의 필요여부까지 학습해야 하기 때문에 많은 학습 시간을 필요로 한다.

이에 GA에 종족(species) 개념[4]을 도입하여 최적 구조를 도출할 수 있는 종족 유전 알고리즘(species genetic algorithm; SGA) 방법이 제안되었다[5]. SGA는 전체 탐색 공간을 은닉층 뉴런의 수에 따라 종족으로 분할한 다음, 불필요하다고 판단되는 종족에 대해서는 탐색을 중단함으로써 전체적인 학습 시간을 줄이는 방법이다. 그러나, 이 방법은 돌연변이 확률, 교배 확률 등의 설정과 공간 분할 방법에 따라 학습 결과가 달라질 수 있다[5]. SGA의 문제점을 해결하는 시도로 주성분분석(principle component analysis)으로 입력 특징 공간을 압축한 후 SGA를 적용하는 PSGA 방법[6]이 있다. 이 논문에서는 또 다른 접근법으로 SGA 종

족 내의 학습 방법을 달리하여 SGA의 학습 시간을 단축시키고자 한다.

GA를 변형한 학습 방법 중, Sawai 등[7]은 파라메타 설정에 무관하며, 학습 속도를 개선시킨 학습 방법인 PfGA(parameter-free genetic algorithm)를 제안하였다. 이 방법은 탐색체의 수가 소량이며 가변적인 특징을 갖고 있으나, 탐색 공간이 큰 문제에서는 초기 학습 속도가 느리며 지역적 최소치(local minima)에 빠지는 경우가 발생한다.

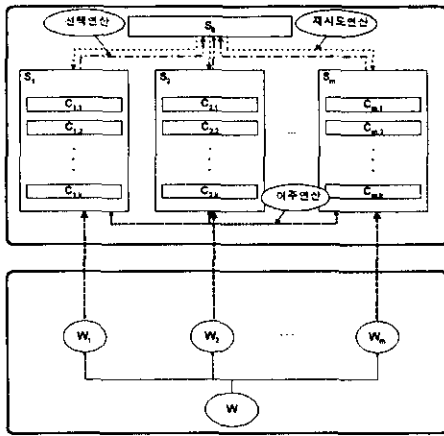
이 논문에서는 SGA와 PfGA를 결합하여 학습하는 방법(parameter-free species genetic algorithm; PfSGA)을 제안한다. PfSGA는 SGA에 의한 탐색 공간 분할 및 불필요한 공간 제거의 학습 방법과 PfGA의 빠른 학습 속도 및 파라메타의 선택이 불필요하다는 장점을 이용하였다. PfSGA는 공간 분할이 세분화되어 있어, 불필요한 공간의 제거가 SGA에 비해 신중하게 이루어지며 학습 속도 또한 향상시킬 수 있다.

### 2. PfSGA 학습

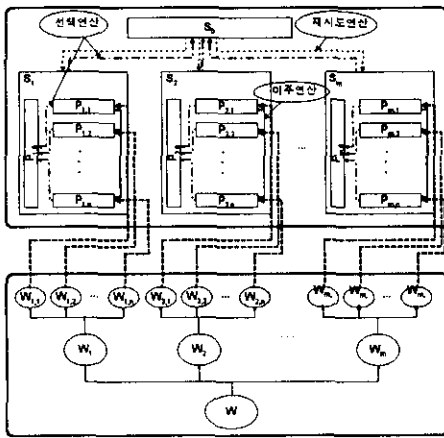
#### 2.1 PfSGA의 구조

PfSGA는 SGA의 공간 분할 및 제거와 PfGA의 빠른 학습 속도 및 파라메타 설정의 불필요성을 접합한 방법이다. SGA의 종족 분할을 수행한 후, 종족 내에 GA 학습 알고리즘을 이용하는 하나의 탐색체 풀을 위치시키는 대신, PfGA 학습 알고리즘을 이용하는 여러 개의 탐색체 풀을 위치시킨다. 즉, 종족 내의 학습 알고리즘을 GA에서 PfGA로 변환한 알고리즘이라 할 수 있다. 그러나, PfSGA의 영역 분할 및 제거 방법은 SGA에 비교할 때 차이가 있다. 그림 1은 PfSGA의 학습 구조와 각 종족에 할당된 서브공간의 관계를 나타낸 것이다.

그림 1(a)에서  $W$ 는 전체 탐색 공간이며,  $W_i$ 는  $H$ 에 의해 분할된  $m$ 개의 서브 공간이다. SGA는 전체 학습 공간을  $H$ 에 의해  $m$ 개로 분할한 후 종족별로 학습하나, PfSGA는  $W_i$ 를 다시  $n$ 개의 서브공간으로 분할하며  $W_{ij}(1 \leq i \leq m, 1 \leq j \leq n)$ 로 표기한다. 여기서,  $n$ 은 전체 링크의 수인  $(I + O) \times H$ 로 하면 된다.  $W_{ij}$ 로의 분할 기준은 연결되어 있는 링크의 수로 한다. 즉,  $W_{ij}$ 는  $W_i$  중 위상문자열에서 그 값이



(a) SGA



(b) PfSGA

그림 1. 학습 구조와 영역 분할의 관계

'1'인 비트의 수가  $j$ 개인 서브공간을 의미한다.

그림 1 (b)에서 PfSGA의 종족  $S_i$ 는  $n$ 개의 염색체 풀을 가지고 있으며  $P_{i,j}$ 로 표기한다. 이  $P_{i,j}$ 는  $W_{i,j}$  영역을 학습하게 된다.  $P_{i,j}$ 은 PfGA에서의 염색체 풀이며,  $C_{i,j,k}$ 인  $k$ 개의 염색체들을 가지고 있다. 따라서,  $C_{i,j,k}$ 는 은닉층 뉴런의 수가  $i$ 개, 연결되어 있는 링크의 수가  $j$ 개인  $k$ 번째 염색체를 뜻한다. 그리고,  $S_0$ 는 종족들에 대한 제거 연산을 제어하며, 각 종족마다  $P_{i,0}$ 를 두어  $P_{i,j}$ 들의 제거 및 생성을 제어한다.

### 2.2 PfGA의 학습 알고리즘

PfSGA의  $P_{i,j}$ 내에서는 PfGA 학습 알고리즘이 이용되며, 학습 구조는 그림 2와 같다. PfGA[7]는 기존 GA의 돌연변이 확률, 교배 확률, 초기 염색체의 수 등의 파라미터를 설계자가 지정해야 한다는 단점을 해결하고, 또한 학습하는 동안 염색체의 수가 평균 1~3개로 소량이며, 가변적인 특징을 갖고 있어 학습 시간을 단축할 수 있다. 학습은 그림 내의 번호 순서로 이루어지며 각 수행 내용은 다음과 같다.

①, ②는  $W$ 로부터 임의로 하나의 염색체를 선정하여  $P$

로, ③은  $W$ 로부터 임의로 두 개의 염색체를 선정하여  $C1, C2$ 로 하여 교배연산, ④는 교배로 생성된 자식염색체 중 한 염색체에 대하여 확률 1/2로 돌연변이를 수행, ⑤는 교배로 생성된 2염색체와  $C1, C2$ 중에서 선택-도태연산으로 선택된 1~3개의 염색체를  $P$ 로 되돌려준다. ⑥은  $P$ 의 염색체 수가 2보다 크거나 같으면 ③으로 가고, 1이면 ②로 간다.

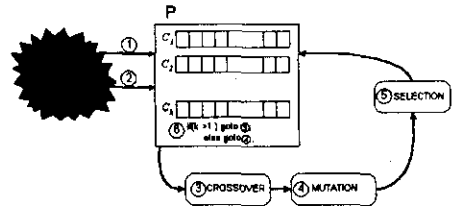


그림 2. PfGA의 학습 구조

### 2.3 PfSGA의 학습 알고리즘

그림 1(b)의 구조를 이용하는 PfSGA의 학습 알고리즘은 다음과 같다.

- [단계 1] 임의의 종족의 수  $m$ 을 결정한다.
- [단계 2] 임의의 풀의 수  $n$ 을 결정한 후, 학습 세대  $t$ 를 0으로 한다.
- [단계 3] PfGA의 ①, ②단계에 해당한다.  $P_{i,j}$ 는 서브공간  $W_{i,j}$ 에서만 염색체를 추출하며, 만약  $W_{i,j}$ 가 불필요한 공간이라고 판단되어  $P_{i,j}$ 가 제거될 경우  $W_{i,j}$ 에서의 염색체 추출은 더 이상 이루어지지 않는다.
- [단계 4] PfGA의 주 학습 단계인 ③~⑤단계로 각 종족에 대하여 적용한다.
- [단계 5] 돌연변이, 교배 연산에 의해 변형된 염색체를 적합한 풀로 이주시키는 연산이다.
- [단계 6]  $P_{i,j}$ 내 염색체  $C_{i,j,k}$ 들 중에서 적합도 값이 가장 좋은 염색체를 추출하는 단계이며 이 염색체를  $C_{i,j}(t)$ , 적합도 값을  $F_{i,j}(t)$ 라고 한다.  $P_{i,j}$ 의  $C_{i,j}(t)$ 들 중 가장 적합도 값이 큰 염색체를  $P_{i,0}$ 에  $C_i(t)$ 로 저장하며 이 때의 적합도 값을  $F_i(t)$ 로 한다.
- [단계 7] 전체 염색체들 중에서 적합도 값이 가장 좋은 염색체를 추출하는 단계로써,  $F_i(t)$ 들을 상호 비교하여 가장 높은 값을 갖는 염색체를 추출하여  $S_0$ 에 저장한다. 이 염색체와 적합도 값을 각각  $C(t)$ 와  $F(t)$ 로 한다.
- [단계 8]  $t$ 세대에서  $P_{i,j}$ 의 학습을  $L_{i,j}(t)$ 를 계산하는 단계이며, 식 (1)과같이 계산한다.
 
$$L_{i,j}(t) = \max(F_{i,j}(t')) - \min(F_{i,j}(t)), \quad \forall t, 0 \leq t \leq t$$
 식 (1)
- [단계 9] PfSGA 학습에서 풀을 없애는 제거연산이다. 제거 연산은  $F_i(t)$ 의 변화에 따라 필요에 의해 종족별로 이루어지며, 이 세대를  $r_i$ 라고 하며 다음의 조건에 의한다.
 
$$\text{if } (F_i(t) > \alpha \times F_i(r_i - 1)) \text{ then remove\_}P_{i,j}(t);$$
 여기서,  $\alpha$ 는 1.2 ~ 2.0사이의 상수이다. remove\_  $P_{i,j}(t)$ 는 학습율이 제일 부진한 풀을 제거하는 함수이다.
- [단계 10] SGA의 종족 제거 작업으로써 단계 9와 같이 불

필요한 영역의 축소를 수행하며, 이 세대를  $r_2$ 라고 한다.

if  $(F(t) > \beta \times F(r_2 - 1))$  then remove  $P_i(i)$ ;

여기서,  $\beta$ 는 2.0 ~ 4.0 사이의 상수이다.  $remove\_P_i(i)$ 는 학습율이 제일 부진한 종족  $S_i$ 가 제거된다. 즉, [단계 10]은 넓은 단위의 서브 공간, [단계 9]는 작은 서브 공간을 제거하는 연산으로써, SGA에 비해 불필요한 공간의 제거가 신중하게 이루어진다.

[단계 11] [단계 3]에서 [단계 10]을 반복한다.

### 3. 실험 및 결과 고찰

#### 3.1 실험 방법

제안 방법의 성능 평가를 위하여 IRIS, E13B, LETTER의 세 가지 실험 데이터를 사용하였다[5]. GA, SGA, PfGA, PfSGA의 네 가지 방법에 대하여 실험한 후 그 성능을 비교 평가하였다. Sun sparc/20에서 C언어를 이용하였으며, 실험 결과는 세 가지 데이터에 대해 20~30번씩 수행한 후 평균치를 표현하였다. 실험에 사용된 파라메타 값들은 다음과 같다. 이 파라메타들은 경험치에 의해 설정되었으며, 부적절하게 설정될 경우에는 학습 결과에 영향을 미친다.

- GA
  - 전체 염색체 수 = 150,
  - 학습세대 = 1,000(IRIS), 2,000(E13B, LETTER),
  - 돌연변이 확률 = 0.01, 교배 확률 = 0.7.
- SGA
  - 한 종족의 초기 염색체 수 = 20,
  - 한 종족의 최대 염색체 수 = 100,
  - 학습세대 = 1,000(IRIS), 2,000(E13B, LETTER),
  - 돌연변이 확률 = 0.01, 교배 확률 = 0.7,
  - 초기 종족의 수 = 10.
- PfGA
  - 학습세대 = 100,000(IRIS), 200,000(E13B, LETTER).
- PfSGA
  - 학습세대 = 5,000(IRIS), 10,000(E13B, LETTER),
  - 초기 종족의 수 = 10.
  - 한 종족( $S_i$ )의 초기 풀의 수( $n$ ) =  $i \times (I + O)$
  - $\alpha = 1.3$ ,  $\beta = 2.0$ .

#### 3.2 학습 속도 고찰

우선 각 방법의 학습 속도를 평가하기 위하여 네 방법을 같은 시간동안 학습한 후의 MSE를 표 1과 같이 측정하였다. 학습이 빨리 이루어지는 방법일수록 많은 오차를 줄일 수 있기 때문에 MSE가 작게 된다. 4가지 방법 중 PfGA가 모든 데이터에 대해서 가장 큰 MSE를 보임으로써 학습 속도가 가장 느리며, PfSGA가 가장 빠른 것으로 나타났다. 그러나, PfGA를 SGA와 결합(PfSGA)하게 되면, 그 성능은 탁월하게 향상되는 것을 알 수 있다. 이는 SGA를 이용하여 탐색 공간을 분할함으로써 PfGA 염색체 풀이 적은 공간에서 탐색을 수행하기 때문이다.

표 1. MSE 비교

데이터 방법	IRIS	E13B	LETTER
GA	0.026 ± 0.023	0.054 ± 0.094	0.104 ± 0.143
SGA	0.009 ± 0.012	0.019 ± 0.041	0.063 ± 0.076
PfGA	0.029 ± 0.024	0.073 ± 0.098	0.132 ± 0.151
PfSGA	0.009 ± 0.011	0.018 ± 0.032	0.059 ± 0.047

표 2는 IRIS 데이터를 학습하는 과정에서 염색체 수에 대한 변화를 나타낸 것이다. GA 방법은 초기 염색체의 수 150개가 학습이 끝날 때까지 유지된다. 그러나, PfGA는 염색체의 수가 가변적이라는 특징을 갖고 있으며 한 세대 당 평균 2.13개의 염색체를 가지고 학습한다. SGA와 PfSGA는 종족 및 풀의 제거로 인하여 염색체의 수가 변하며, SGA는 평균 157.31개로 GA와 비슷하다. PfSGA는 평균 42.17개로 PfGA에 비해 약 20배정도 염색체의 수가 많다. 따라서, 3.1절의 실험 환경에서 전체 학습 세대를 방법에 따라 달리하였으며, 모든 방법이 학습에 소요되는 시간이 동일하게 하였다.

표 2. 염색체의 수 비교

세대 방법	1	100	200	300	400	500	600	700	800	900	1000	평균
GA	150	150	150	150	150	150	150	150	150	150	150	150.00
SGA	200	200	200	200	200	200	100	100	100	100	100	157.31
PfGA	2	2	2	3	2	3	2	2	2	2	2	2.13
PfSGA	200	137	91	41	17	2	2	2	2	2	2	42.17

#### 3.3 MLP 분류기 구조 고찰

실험 데이터에 대하여 각각의 방법으로 학습한 MLP 분류기의 구조는 표 3과 같다. 각 방법에 있어서 은닉층 뉴런의 수는 별 차이가 없으나, 전체 링크의 수에 있어서는 차이를 보인다. 표 1에서 보인 MSE와 표 3의 링크의 수와는 밀접한 관계를 갖고 있다. MSE를 빨리 줄이는 방법일수록 링크의 수 또한 많이 줄일 수 있다. 특히 입력 차원 및 클래스의 수가 많아질수록 이 차이는 현저하게 나타난다. E13B와 LETTER의 경우 PfGA에서의 링크의 수는 PfSGA와 100개 이상 차이를 보인다.

MSE와 구조학습과의 관계를 나타내기 위하여 E13B 데이터에 대한 MSE와 엔트로피 값의 감소 과정

표 3. MLP분류기의 구조 비교

구조 방법	IRIS		E13B		LETTER	
	H	L	H	L	H	L
GA	2.56	12.44	7.01	278.42	12.31	254.21
SGA	2.63	12.18	6.58	208.91	13.01	209.57
PfGA	3.15	12.70	9.05	301.45	14.05	294.05
PfSGA	2.55	12.15	7.10	179.70	13.90	179.10

(H: 은닉층 뉴런의 수, L: 전체 링크의 수)

을 그림 6에 나타내었다. SGA와 PfSGA는 학습 초기부터 MSE를 줄이는 작업이 활발하게 이루어지는 데 비하여, GA와 PfGA는 이 방법들보다 더디게 진행되는 것을 그림 3의 (a)를 보면 알 수 있다. 그림 3의 (b)에서 GA와 PfGA는 엔트로피 학습이 1,400~1,600

회사에서 시작되는 것에 비해 SGA와 PfSGA는 600~800회 정도부터 이루어지며, 1,500회 정도에 거의 수렴하여 학습이 빠르게 진행된다. PfSGA와 SGA를 비교해 보면, 그림 3의 (a)의 MSE 변화량은 유사하지만 (b)의 엔트로피 학습에서는 차이를 보인다. 800~1,000회 사이에 PfSGA는 SGA보다 엔트로피 학습이 빠르게 이루어진다. 따라서, PfSGA에서 사용된 종족에 대한 분할 및 각 영역에 대한 제거 작업이 종족 내에서의 불필요한 탐색을 줄여 SGA에 비해 학습 효율이 우수함을 확인할 수 있다.

3.4. 오인식율 고찰

표 4에 나타난 각 방법에 대한 오인식율 또한 위의 표들과 같은 결과를 나타낸다. IRIS의 경우 각 방법의 학습 패턴에 대한 MSE는 차이가 있으나 오인식율은 거의 유사하다. 그러나, E13B와 LETTER의 경우에는 오인식율에서 차이를 보인다. 이는 표준 편차를 보면 알 수 있듯이 2,000회 반복의 학습 횟수동안 GA와 PfGA는 완전하게 학습하지 못하는 경우가 발생하게 되어 평균적으로 오인식율이 높다. 표 1에서 학습이 빠른 시간에 잘 이루어지는 방법일수록 표 4에서 오인식율 또한 낮게 나타났다.

표 4. 오인식율 비교(%)

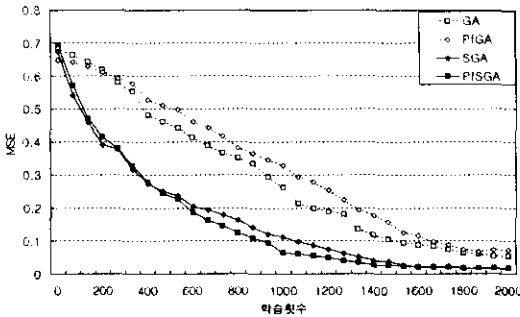
데이터 방법	IRIS	E13B	LETTER
GA	4.5 ± 0.1	6.82 ± 1.45	27.35 ± 5.34
SGA	4.3 ± 0.1	4.84 ± 0.34	23.15 ± 3.12
PfGA	4.6 ± 0.1	9.28 ± 1.99	30.02 ± 5.65
PfSGA	4.3 ± 0.1	4.78 ± 0.32	21.72 ± 2.92

4. 결론

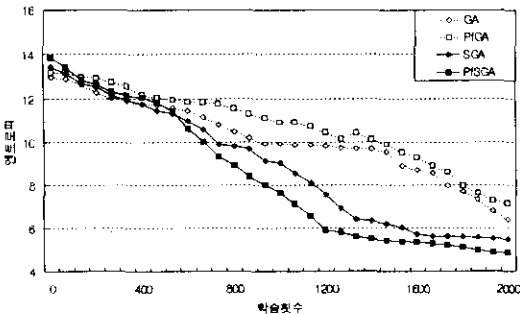
이 연구에서는 종족 유전 알고리즘 SGA와 파라메타의 설정에 무관한 유전 알고리즘 PfGA를 결합하여 MLP분류기의 위상구조를 학습하는 방법인 PfSGA를 제안하였다. 세 종류의 벤치마크 데이터를 대상으로 학습성능을 실험한 결과, SGA와 PfSGA가 GA와 PfGA에 비해 빠른 시간 내에 최적의 위상구조를 선택할 수 있음을 확인하였다. 그리고, PfGA를 단독으로 수행하는 것보다 SGA와 결합하여 학습하는 방법인 PfSGA로 학습 효율을 높일 수 있었다. 특히, 특정 차원 및 클래스의 수가 많을수록 기존의 방법에 비해 성능이 우수하다는 것을 확인하였다.

참고문헌

- [1] H. Kitano, "Empirical studies on the speed of convergence of neural network training using genetic algorithms", Proceedings of AAAI, pp. 789-795, 1990.
- [2] D. Dasgupta, and D. R. McGregor, "Designing application-specific neural networks using the structured genetic algorithm", Proceedings of COGANN, pp. 87-96, 1992.
- [3] T. Nagao, T. Agui and H. Nagahashi, "Structural evolution of neural networks having arbitrary connections by a genetic method," IEICE Trans. on Information and Systems, vol. E76-D, no. 6, pp. 689 - 697, 1993.
- [4] J. C. Potts, T. D. Gidden and S. B. Yadav, "The development and evaluation of an improved genetic algorithm based on migration and artificial selection", IEEE Trans. on SMC, Vol. 24, No. 1, pp. 73 - 86, Jan. 1994.
- [5] 신성효, 김상운, "종족 유전 알고리즘을 이용한 MLP분류기의 구조학습", 대한전자공학회논문지-C, Vol. 35, No. 2, pp. 112 - 119, 1998. 2.
- [6] S. -W. Kim, S. -H. Shin, Y. Aoki, "A structural learning of neural-network classifiers using PCA networks and species genetic algorithms", IEICE Trans. on Fundamentals., Vol. E81-A, No. 6, pp. 1183 - 1186, Jun. 1998.
- [7] 澤井 季文, 木津左千夫, 遠藤 哲郎, "パラメータフリー遺傳的アルゴリズムと定常状態遺傳的アルゴリズムとの性能比較", 電子情報通信學會論文集, Vol. J81-D-II, No. 6, pp. 1455-1459, 1998.



(a) Reduction of MSE



(b) Reduction of entropy

그림 3. E13B에 대한 MSE와 엔트로피 감소 과정