

유전알고리즘을 이용한 신경망 최적화 기법

한승수, 송경빈, 홍덕현, 최준림  
 대구효성가톨릭대학교 자동차공학부, 경북대학교 전자전기공학부

Optimizing Neural Network Using Genetic Algorithms

Seung-Soo Han, Kyung-Bin Song, Dug Hun Hong, and Jun-Rim Choi  
 Catholic University of Taegu-Hyosung, Kyungpook National University

**Abstract** - 신경망은 선형 시스템 뿐만 아니라 비선형 시스템에 있어서도 탁월한 모델링 및 예측 성능을 갖고 있다. 하지만 좋은 성능을 갖는 신경망을 구현하기 위해서는 최적화 해야할 파라미터들이 있다. 은닉층의 뉴런의 수, 학습율, 모멘텀, 학습오차 등이 그것인데 이러한 파라미터들은 경험에 의해서, 또는 문헌들에서 제시하는 값들을 선택하여 사용하는 것이 일반적인 경향이다. 하지만 신경망의 전체적인 성능은 이러한 파라미터들의 값에 의해서 결정되기 때문에 이 값들의 선택은 보다 체계적인 방법을 사용하여 구하여야 한다. 본 논문은 유전 알고리즘을 이용하여 이러한 신경망 파라미터들의 최적 값을 찾는데 목적이 있다. 유전 알고리즘을 이용하여 찾은 파라미터들을 사용하여 학습된 신경망의 학습오차와 예측오차를 심플렉스 알고리즘을 이용하여 찾은 파라미터들을 사용하여 학습된 신경망의 오차들과 비교하여 본 결과 유전 알고리즘을 이용하여 찾은 파라미터들을 이용했을 때의 신경망의 성능이 더욱 우수함을 알 수 있다.

1. 서 론

프로세스를 개발하는 사람들에게는 프로세스 제어나 고장진단, 성능향상 등을 이루기 위해서 정확하고 효율적인 모델이 필수적이다. 최근에는 신경망을 사용하여 이러한 모델링 방법을 많이 개발하여 사용하고 있다. 특히 반도체 공정과 같은 복잡한 시스템을 모델링 하기 위해서는 일반적으로 많이 사용하는 통계적 방법에 의존하는 것보다는 신경망을 이용하게 되면 모델의 오차나 예측 오차에 있어서 더욱 우수한 결과를 얻을 수 있다는 연구 결과가 많이 나오고 있다 [1-3]. 신경망을 이용하여 모델링을 하기 위해서는 신경망을 구성하는 여러 가지 파라미터들을 조정해야할 필요가 있다.

다층 신경망 회로는 하나의 입력층, 하나의 출력층, 그리고 다수의 은닉층으로 구성되어 있으나 하나의 은닉층만 사용해도 연속함수를 추정할 수 있다고 알려져 있다 [4]. 하지만 은닉층의 뉴런의 수를 최적화 하는 문제는 매우 어려운 일이다. 이러한 신경망의 구조적인 문제 이외에도 학습율과 모멘텀, 그리고 학습허용오차 등과 같은 학습 파라미터들의 값에 의해서 신경망의 성능에 많은 변화가 생긴다. 이러한 구조적 파라미터와 학습 파라미터를 최적화 하는 문제는 많은 연구자들에 의해서 연구되어 왔다. Kim과 May는 구조적 파라미터가 학습 개략화를 신경망으로 수행했을 때의 성능에 미치는 영향을 연구하였다 [5]. 그들은 실험계획법에 의해 구조적 파라미터와 학습 파라미터를 변화시켜가면서 Nelder와 Mead의 심플렉스 알고리즘을 사용해서 최적 파라미터를 결정하는 방법을 제안하였다. 하지만 심플렉스 알고리즘은 시작점에 의한 영향이 매우 크며, 또한 국부 최적점에 안착하기 쉽다.

1970년대에 Holland는 기존의 "hill - climbing" 방법을 대체할 유전알고리즘을 제안하였다 [6]. 유전알고리즘은 진화연산의 일종이며, 매우 크고 복잡한 문제를

해결하는데 매우 적합하다. 본 연구에서는 유전알고리즘을 신경망의 구조적 파라미터와 학습 파라미터의 최적화에 사용하였다. 반도체 공정중의 플라즈마 화학 증착법에 의한 실리콘 산화막의 증착 조건과 특성을 측정한 데이터를 사용하여 신경망을 학습하였다. 신경망의 입력 변수는 개스량, 온도, 압력, RF Power이며, 출력값은 산화막의 permittivity, refractive index, residual stress, uniformity, impurity concentration 등이다. 신경망을 학습시킬 데이터를 얻기 위해서 central composite 실험계획법에 의해서 실험하여 데이터를 취득하였다. 신경망은 학습오차와 예측오차를 최소화하도록 최적화 하였으며, 이를 위하여 유전알고리즘을 사용하였다. 심플렉스 알고리즘을 사용하였을 때보다 학습오차는 약 10% 감소하였으며 예측오차는 약 65%의 성능 향상을 이루었다.

2. 실험계획법과 신경망 모델

2.1 PECVD 산화막 특성 측정

실리콘 산화막은 Plasma-Therm 700 시리즈를 사용하여 증착하였으며 N<sub>2</sub>O, 2% SiH<sub>4</sub>를 사용하였고 표 1에서 보인바와 같이 증착조건을 변화화하면서 central composite 실험계획법에 의해서 실험하였다.

표 1 Deposition Parameters

Parameter	Range
Substrate Temperature	200 - 400 °C
Pressure	0.25 - 1.8 torr
RF Power	20 - 150 watt
2% SiH <sub>4</sub> in N <sub>2</sub> Flow	200 - 400 sccm
N <sub>2</sub> O Range	400 - 900 sccm

약 5µm의 산화막을 4" (100) 방향의 실리콘 웨이퍼 위에 증착 시켰다. 13.56MHz의 RF Power를 사용하였으며 전극간의 간격은 2.29cm이고 전극의 직경은 11 인치 이다. Central composite 실험계획법에 의해 만들어진 29개의 증착조건에 의해서 실험된 데이터는 신경망의 학습에 사용되었고 이외에도 8개의 다른조건에서 증착을 실시하여 여기서 얻은 데이터는 예측오차를 측정하는데 사용하였다. 증착후에는 Metricon 2010 prism coupler를 사용하여 산화막의 두께와 refractive index를 측정하였고 이를 웨이퍼의 다섯군데서 측정하여 uniformity를 계산하였다. Flexus F2320을 사용하여 증착된 산화막의 stress를 측정하였다. H<sub>2</sub>O, SiOH와 같은 불순물을 측정하기 위해서 Perkin-Elmer 1600 FTIR을 사용하였고, 산화막의 permittivity를 계산하기 위해서 parallel-plate capacitor를 만들어서 Keithley 590 CV analyzer와

HP 4275 LCR meter를 사용하여 capacitance를 측정하였다.

### 2.2 신경망의 구조적 파라미터와 학습 파라미터

신경망은 잡음이 섞인 비선형의 입출력 데이터를 사용하여 입출력 관계를 찾아내는 학습 능력을 갖고 있다. 일반적으로 많이 사용되는 학습 알고리즘은 feed forward error backpropagation(FFEBP) 알고리즘이다. 신경망의 입력층의 뉴런들은 입력 데이터를 받아들이며 출력층의 뉴런들은 신경망에서 계산된 결과를 외부에 알려주는 역할을 한다. 은닉층은 외부에는 노출되지 않는 부분이나 실제 데이터를 처리하고 학습하는데 매우 중요한 역할을 하는 부분이다. FFEBP 알고리즘에서는 이러한 은닉층의 뉴런에 의해서 여러 개의 상층된 입출력 매핑을 해결하는 역할을 하고 있다 [7]. 은닉층의 뉴런의 수가 많으면 복잡한 입출력 관계를 설정할 수 있으나 경우에 따라서는 over train 되는 경우가 발생하여 입출력 관계를 일반화하지 못하는 문제가 발생할 수 있다.

Backpropagation은 gradient descent 기법을 사용하는데 누적된 오차함수  $E$ 를 편미분 하여 신경망의 weight를 체계적으로 변화시켜 입출력 관계를 학습시킨다. Weight의 변화량을  $\Delta w_{ijk}$ 라고 했을 때 다음의 식에 의해서 weight가 변화한다.

$$\Delta w_{ijk}(n+1) = -\eta \frac{\partial E}{\partial w_{ijk}} + \alpha \Delta w_{ijk}(n) \quad (1)$$

여기서  $i$ 는  $k$ 층의 노드를 나타내고  $j$ 는  $(k-1)$ 층의 노드를 나타내며  $w_{ijk}$ 는 두 노드간의 weight를 나타낸다.  $\eta$ 는 학습율이라 하며 보통 0과 1사이의 값을 갖는다. 학습율이 크면 학습 속도는 빨라지지만 오차함수의 최적점 근방에서 수렴하지 않으며, 또한 너무 작게 하면 이러한 문제를 해결할 수 있지만 학습속도가 너무 느리게 된다. 식(1)의 우변의 두 번째 항을 모멘텀항이라 하며  $\alpha$ 값은 0과 1사이의 값을 갖게 된다. 모멘텀항을 추가함으로써 신경망을 학습시킬 때 학습공간에서 학습이 진행되는 방향으로의 진행을 도우며 최적점 근방에서의 수렴을 도와준다. 학습율과 모멘텀 이외에도 신경망의 성능에 영향을 미치는 것은 학습오차이다. 학습오차가 작으면 주어진 데이터에서는 정확하게 예측하지만 그 이외의 데이터에서는 일반적으로 예측기능이 떨어지게 된다.

## 3. 신경망의 최적화

### 3.1 심플렉스 방법을 이용한 신경망의 최적화

신경망을 최적화 하기 위해서 진술한 학습파라미터 뿐만 아니라 구조적 파라미터도 최적화 하여야만 한다. 즉 은닉층의 뉴런 수, 학습율, 모멘텀, 학습오차 등의 4가지 파라미터 값을 최적화 해야 한다. 이러한 파라미터들은 신경망의 학습오차 뿐만 아니라 예측오차도 작게 하는 방향으로 설정되어야 한다. 심플렉스 방법을 사용하여 위해서 4개의 파라미터들의 값을  $2^4$  factorial 실험 계획법에 의해서 표2와 같이 변화 시키거면서 그에 따른 신경망의 성능을 조사하였다.

표 2 Ranges of Neural Network Parameters

Parameters	Range	Default Value
No. of Hidden Neurons	3 - 9	6
Learning Rate	0.05 - 0.5	0.275
Momentum	0.35 - 0.95	0.65
Training Tolerance	0.01 - 0.13	0.07

이를 위한 성능지표(PI)는 다음과 같이 설정하였다.

$$PI = K_1 \sigma_t^2 + K_2 \sigma_p^2 \quad (2)$$

여기서  $\sigma_t$ 는 신경망의 학습오차에 대한 RMS 값이고,  $\sigma_p$ 는 신경망의 예측오차에 대한 RMS 값이다. 학습오차를 줄이는 것 보다 예측오차를 줄이는 것이 더 중요하므로  $K_1=1, K_2=10$ 으로 조정하였다. Nelder와 Mead의 심플렉스 알고리즘을 사용하기 위해서 RS/Discover [8]를 사용하여 식(2)의 성능지표를 최소화하는 각 파라미터 값을 조사하였다. 심플렉스 알고리즘은 매우 강한 탐색 알고리즘이며, 특히 변수의 수가 5개를 넘지 않을 경우 매우 효과적인 알고리즘이다. 심플렉스 방법은 원래는 regular simplex인 경우 탐색 방법의 성능이 탁월하였으나 Nelder와 Mead에 의해서 심플렉스가 regular가 아닐 경우에도 적용할 수 있게 제안되었다 [9]. 본 심플렉스 방법을 사용하여 신경망의 파라미터들을 최적화 하였을 때의 성능과 이를 유전알고리즘으로 최적화 하였을 때의 성능을 뒤에 비교하여 놓았다.

### 3.2 유전 알고리즘을 이용한 신경망의 최적화

유전 알고리즘(GA)은 Holland에 의해서 주창된 탐색 알고리즘으로서 [6] reproduction, crossover, mutation 등과 같은 연산을 통하여 탐색 영역에서 최적의 값을 찾아내는 방법이다. 유전 알고리즘은 비정형의 넓은 탐색 영역에서 국부 최적점에 빠져들지 않고 빠르게 전역 최적점을 찾는 능력을 갖고 있으며, 또한 다른 'hill-climbing' 방법들과는 달리 함수를 미분할 필요 없이 목적함수의 값만을 요구하는 강력한 탐색 알고리즘이다. 유전 알고리즘을 사용하기 위해서는 후보 최적점을 이진 스트링으로 변환해야 하며 다음과 같은 4가지 사이클을 반복하여 최적 스트링(최적점)을 찾는다.

- (i) 스트링 군집 발생
- (ii) 스트링에 대한 목적함수 평가
- (iii) 최적 스트링의 선택
- (iv) 유전 연산을 통한 새로운 군집 생성

군집의 각 요소들은 스트링으로 코딩된 후 유전 연산을 거쳐서 각각의 적합도를 검사한다. 적합도 검사에서 우수한 값을 얻은 개체는 새로운 군집 생성을 위한 후보로 선택된다. 선택은 선택전략에 따라 적합하게 판명된 개체가 선택되게 되며 선택된 개체는 새로운 군집을 형성하게 된다. 본 연구에서 사용한 선택전략은 elitist roulette wheel selection이다. 유전 알고리즘에서 사용한 적합도 판별 함수는 다음과 같다.

$$F = \frac{1}{1 + PI} \quad (3)$$

100 세대 동안 반복 계산 후  $F$  값이 최대가 되는 것을 최적해로 선정하였다. 유전 알고리즘을 이용한 신경망 최적화 방법을 도식적으로 표현하면 그림1과 같다.

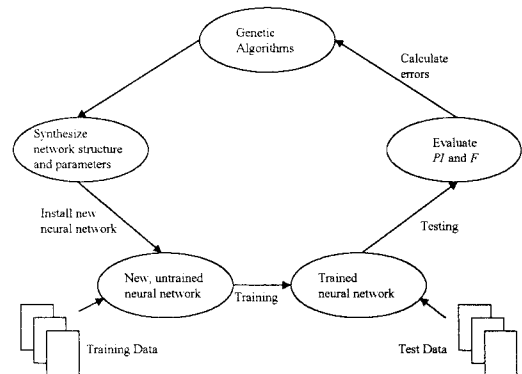


그림 1 Block Diagram of Genetic Optimization

#### 4. 결과 및 토의

신경망 파라미터들을 심플렉스 알고리즘과 유전 알고리즘을 사용하여 최적화 하였을 때의 결과를 표3과 표4에 나타내었다. 표3은 심플렉스 알고리즘으로 최적 파라미터를 찾았을 때의 결과이고, 표4는 유전 알고리즘으로 최적 파라미터를 찾았을 때의 결과인데, 모멘텀이나 설정 학습오차 값은 거의 비슷하나 뉴런의 개수와 학습률은 차이가 있음을 알고 있다. 하지만 표5와 표6을 보면 그 결과에 현격한 차이를 볼 수 있다. 표5와 표6은 심플렉스 알고리즘과 유전 알고리즘으로 찾은 최적 파라미터를 사용하여 신경망을 구성하였을 경우의 학습오차와 예측오차 이다. 학습오차의 경우 심플렉스 알고리즘에 의해 찾은 최적 파라미터를 사용했을 때나 유전 알고리즘을 사용했을 때 우열을 가리기 힘들게 나왔으나 예측오차를 보면 유전 알고리즘에 의해 찾은 최적 파라미터를 사용하여 구성한 신경망의 성능이 평균 66% 정도 향상됨을 볼 수 있다.

표 3 Network Parameters Optimized by Simplex Search

PECVD Response	Hidden Neurons	Momentum	Learning Rate	Training Tolerance
Permittivity	9	0.40	0.50	0.01
Ref. Index	9	0.40	0.50	0.01
Stress	6	0.40	0.05	0.13
Non-Uniformity	6	0.40	0.50	0.13
Impurity Concentration	6	0.40	0.05	0.07

표 4 Network Parameters Optimized by Genetic Algorithms

PECVD Response	Hidden Neurons	Momentum	Learning Rate	Training Tolerance
Permittivity	7	0.41	0.19	0.01
Ref. Index	7	0.40	0.37	0.08
Stress	7	0.39	0.07	0.06
Non-Uniformity	4	0.43	0.06	0.11
Impurity Concentration	4	0.37	0.08	0.07

표 5 Training Error Comparison of Simplex and GAs Network Optimization

PECVD Response	Simplex	GAs	%Improvement
Permittivity	0.0578	0.0110	80.94
Ref. Index	0.0232	0.0822	-71.76
Stress	0.0500	0.0571	-12.47
Non-Uniformity	0.1146	0.1099	4.09
Impurity Concentration	0.0951	0.0841	7.35

#### 5. 결론

Backpropagation 신경망은 많은 경우의 프로세스 모델 방법에 성공적으로 사용되었지만 신경망의 성능은 구조적 파라미터와 학습 파라미터의 값에 의해 결정된다. 본 연구에서는 유전 알고리즘에 기초한 최적화 방법을 사용하였을 때 최적의 파라미터 값들을 찾아냄을 알 수 있었다.

표 6 Prediction Error Comparison of Simplex and GAs Network Optimization

PECVD Response	Simplex	GAs	%Improvement
Permittivity	0.1788	0.0363	79.68
Ref. Index	0.2158	0.0591	72.61
Stress	0.9659	0.4815	50.16
Non-Uniformity	0.1361	0.0246	81.92
Impurity Concentration	0.0964	0.0795	17.54

#### (참 고 문 헌)

- [1] C. Himmel, and G. May, "Advantages of Plasma Etch Modeling Using Neural Networks Over Statistical Techniques," *IEEE Trans. Semi. Manufac.*, vol. 6, pp. 103-111, May, 1993.
- [2] S. Han, M. Ceiler, S. Bidstrup, P. Kohl, and G. May, "Modeling the Properties of PECVD Silicon Dioxide Films Using Optimized Back-Propagation Neural Networks," *IEEE Trans. Comp. Pack. Manufac. Tech. Part A*, vol. 17, no. 2, pp. 174-182, June, 1994.
- [3] F. Nadi, A. M. Agogino, and D. A. Hodges, "Use of Influence Diagrams and Neural Networks in Modeling Semiconductor Manufacturing Processes," *IEEE Trans. Semi. Manufac.*, vol. 4, no. 1, pp. 52-58, Feb. 1991.
- [4] B. Irie and S. Miyake, "Capabilities of Three-layered Perceptrons," *IEEE International Conference on Neural Networks*, pp. 641-648, 1988.
- [5] B. Kim and G. May, "An Optimal Neural Network Process Model for Plasma Etching," *IEEE Trans. Semi. Manufac.*, vol. 7, pp. 12-21, Feb., 1994.
- [6] J. H. Holland, *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI, University of Michigan Press, 1975.
- [7] M. Gutierrez, J. Wang, and R. Grondin, "Estimating Hidden Units for Two-Layer Perceptrons," *Proc. IEEE Internatl Conf. on Artificial Neural Networks*, 1989, pp. 120-124.
- [8] *RS/Discover User's Guide*, BBN Software Products, 1992.
- [9] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal*, vol. 7, pp. 308-313, 1965.