

역전파 신경망과 통계적 처리를 이용한 공정 데이터 분류

김성모, 김병환  
 세종대학교 전자공학과

Process Data Classification Using Backpropagation  
 Neural Network and Statistical Processing

Sung-Mo Kim, Byungwhan Kim  
 Sejong University, Electronic Engineering

**Abstract** - 역전파 신경망과 데이터분포 특성을 고려한 새로운 알고리즘을 개발하였으며, 이를 플라즈마 데이터의 분류에 응용하였다. 데이터 분포는 통계적인 평균치와 표준편차를 이용하여 특징지었으며, 바이어스인자를 이용하여 9 종류의 데이터를 발생하였다. 각 데이터에 대하여 은닉층의 뉴런수를 변화시키며, 바이어스와 뉴런수에 따른 모델성능을 평균학습시간 (ATT), 평균 예측정확도 (APA), 최적예측정확도 (BPA), 그리고 분류정확도 (CA) 측면에서 세분하여 분석하였다. ATT와 APA에 대해서는 최적화된 학습인자와 데이터 분류인자가 일치하였고, BPA와 CA는 일치하지 않았다. 두 인자간의 상호작용을 동시에 최적화함으로써 완전 분류를 달성하였다.

1. 서 론

공정 모델은 경제적인 공정개발, 즉 공정 최적화에 소요되는 시간과 경비를 줄이기 위해 절실히 요구되고 있다. 반도체 공정 모델은 주로 물리적 모델에 의해 개발이 되어오고 있으나, 계산에 소요되는 시간이 엄청나고, 또한 모델개발 과정에 수반되는 가정들의 존재로 그 예측의 신뢰도가 높지 않다는 단점이 있다. 이에 반해, 신경망은 실험데이터를 기초로 용이하게 개발될 수 있으며, 그 예측정확도는 매우 우수해 공정 최적화에 효과적으로 응용할 수 있다. 신경망중 역전파 신경망이 가장 많이 응용이 되었으며, 복잡한 플라즈마 방전[1]과 플라즈마를 이용한 식각공정의 모델링에 관계하는 물리적 매커니즘의 해석 [2]에 중요하게 적용된 바가 있다. 역전파 신경망 모델의 예측성능은 다수의 학습인자에 의해 결정이 되며, 여기에는 학습허용도, 초기웨이트의 크기, 활성화함수의 종류와 경사, 그리고 은닉층의 뉴런수가 포함된다. 대부분의 응용에 있어, 마지막 인자인 뉴런수를 조정하는데 그치고 있으며, 다른 인자의 개별적인 그리고 상호간의 영향을 고려한 모델 개발 기법이 발표된 바 있다 [3-4]. 하지만 학습인자의 조정에 의해 개발되는 모델의 예측성능은 그 한계가 있으며, 또 다른 한계는 실험공간의 일부영역에 대한 예측성능의 저하에서 찾을 수 있다. 이는 학습에 이용되는 데이터에 대한 연구, 즉 데이터의 분포에 따른 예측성능의 변화에 대한 연구를 요구한다. 또한 데이터 분류와 학습인자간의 최적화 기법도 아울러 요구된다.

본 연구에서는 데이터 특성을 통계적인 평균치와 편차로 수량화하였으며, 평균치에 바이어스 인자가 곱해진 편차를 가중치로 인가하여 분류하였다. 각 분류된 데이터에 대해서 학습인자 중 뉴런수를 변화시키며, 모델의 예측과 분류성능을 살펴보았다. 초기 웨이트의 불규칙적인 분포를 고려하여 주어진 조건에서 총 200개의 모델을 발생하였으며, 이를 ATT, APA, BPA, CA 측면에서 고찰하였다. 분류에 이용된 데이터는 광도파로 제작에 이용된 실리카 박막의 식각공정 중에 수집[5] 되었으며, 식각인자 중 산화막 프로파일에 대한 데이터를 이용하였다.

2. 본 론

2.1 실험 데이터와 분류

그림 1은 식각공정을 수행한 장비의 유도결합형 플라즈마를 이용한 장비의 개요도 이다. 본 식각장비를 이용하여 CF<sub>4</sub>/CHF<sub>3</sub> 가스플라즈마하에서 실리카 박막을 식각하였다. 공정에 이용된 공정변수는 총 4종류이며, 여기에는 소스전력 바이어스 전력, 그리고 두 종류의 가스(CHF<sub>3</sub>와 CF<sub>4</sub>)이다. 부분인자 실험 계획표에 의거하여, 총 여덟 번의 실험이 수행되었고 여기에 중앙점에 해당하는 실험을 첨가하여, 학습데이터를 구성하였다. 모델의 성능을 검증하고자 추가적으로 여섯 번의 실험이 수행되었으며, 따라서 이용된 데이터의 총 수는 17개이다. 프로파일의 각도는 SEM(scanning electron microscopy)를 이용하여 추출하였다. 학습인자는 다시 통계적인 평균치와 편차로 수량화하였으며, 데이터 분포 결정인자는 다음과 같이 설정하였다.

$$\text{결정인자} = \text{평균치} \pm (\text{편차} \times \text{바이어스인자}) \quad (1)$$

바이어스 인자는 -0.4에서 +0.4까지 0.1간격으로 변화시켰으며, 이로 인해 총 9종류의 데이터 set을 준비하였다. 각기 다르게 분류된 데이터 set에 대해서 학습인자중 은닉층의 뉴런수 모델성능ATT, APA, BPA와 CA로 나누어 각기 HN과 바이어스에 따른 변화를 고찰한다.

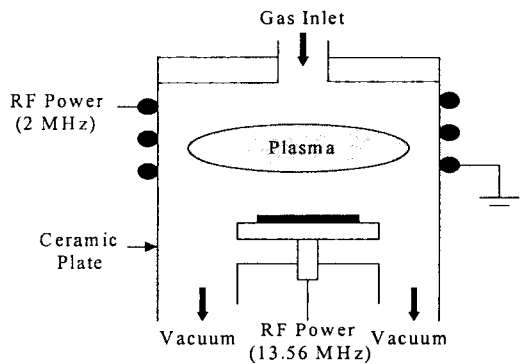


그림 1. 식각장비의 개요도

2.2 역전파 신경망

역전파 신경망 (BPNN) [6]은 플라즈마 공정데이터 해석에 가장 많이 응용되고 있는 구조이다. BPNN은 주로 은닉층이 하나인 3개의 층으로 구성되며, 학습은 일반화된 델타 규칙을 이용하여 진행이 되고, 이를 지배하는 규칙은 다음과 같이 표현된다.

$$W_{i,j,k}(m+1) = W_{i,j,k}(m) + \eta \Delta W_{i,j,k}(m) \quad (2)$$

여기서  $W_{i,j,k}$  는  $(k-1)$ 층의  $j$ 번째 뉴런과  $k$ 층의  $i$ 번째 뉴런사이의 연결강도 이고,  $\Delta W_{i,j,k}$  는 학습편차를 줄이기위해 인가되는 웨이트의 변이이며, 다음과 같이 정의 된다.

$$\Delta W_{i,j,k} = \frac{\partial E}{\partial W_{i,j,k}} \quad (3)$$

$\eta$ 은 학습률을 지칭한다. BPNN의 학습에는 많은 학습인자가 관여하고 있으며 [3-4], 본 연구에서는 이중 은닉층의 뉴런수를 변화시켰다. 특히 초기웨이트의 불규칙성을 고려하여 주어진 뉴런수와 바이어스인자에 대해 200개의 모델을 발생하여 분석하였다.

## 2.3 모델 성능별 분석

### 2.3.1 평균학습시간(ATT)

그림 2는 바이어스와 HN에 따른 ATT를 도시하고 있다. 그림 2에서 볼 수 있듯이 ATT는 데이터 분포에 따라 어떤 전환점을 기준으로 정 반대의 특성을 보이고 있다. 모든 HN에 대해서 그 전환점은 모두 바이어스가 0인 지점에서 발생하고 있다. HN이 3인 경우 데이터 분포에 따라, 특히 음의 바이어스 값에 대해서 ATT의 특성이 크게 영향을 받고 있다. 즉 ATT는 바이어스가 -0.4에서 0으로 증가하면서 일반적으로 증가하고 있으나 0.1이상의 바이어스 값에 대해서는 그 변화가 미미하다. 한편, 양의 바이어스에 대한 ATT는 음의 바이어스에 대한 ATT보다 적다. HN이 4인 경우 전환점 전후의 바이어스에 따른 ATT의 변화는 없으며, 그 수치도 HN이 3인 경우와는 달리 서로 비슷하다. 전환점 전후의 이러한 양상은 HN이 5이상인 경우에도 나타나고 있다. 하지만 HN이 3과 4인 경우와 비교할 때, ATT의 수치는 전환점을 기준으로 반대적인 양상을 보이고 있다. 한 예로, HN이 5인 경우 양의 바이어스에 대한 ATT는 음의 바이어스에 대한 ATT보다 더 증가되었다. 이상으로부터 HN을 변화시킬 때 ATT에서의 전환점은 HN이 4에서 발생하고 있음을 알 수 있다. 다음에 고정된 바이어스에 대한 HN의 영향을 살펴본다. Fig. 1에서 알 수 있듯이 고정된 바이어스에 대해서 HN의 수를 증가할 때 ATT는 감소하고 있다. 이러한 바이어스가 -0.2, -0.1 0에서 HN이 3에서 4로 증가할 때 가장 두드러지게 나타난다. 그림에서 알 수 있듯이, 최적의 ATT는 HN이 8이고 바이어스가 -0.4에서 발생하였으며, 그 수치는 285이다.

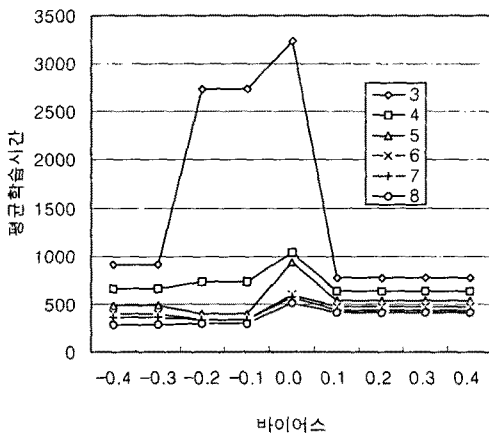


그림 2. 바이어스와 뉴런수에 따른 예측정확도

### 2.3.2 평균예측 정확도(APA)

그림 3은 APA를 바이어스와 HN의 수에 따라 도시하고 있다. HN이 3과 4인 경우 바이어스가 -0.4에서 -0.3으로 증가하였을 때 APA는 감소하였으나, 바이어스가 -0.3에서 -0.1까지 증가할 때는 APA가 함께 증가하고 있다. 이러한 현상은 HN이 5이상인 경우에도 공통으로 나타나고 있다. 바이어스가 0.2이상인 경우에 대해서는 모든 HN에 대해서 APA의 경향이 일치한다. 모든 HN에 대해서 APA의 전환점은 0.1 또는 0.2에서 발생하고 있다. 다음에 고정된 바이어스에 대해서 HN을 변화시킬 때의 특성을 살펴본다. 바이어스가 -0.4인 경우 HN을 증가하면 APA는 일관성 있게 감소한다. 그러나 바이어스가 -0.3인 경우 HN이 5에서 전환점이 발생하며 HN을 6이상으로 증가할 때 APA의 변화는 미미하다. 바이어스가 -0.2, -0.1, 0인 경우 HN의 수를 변화시키더라도 APA의 변화는 미미하다. 그러나 바이어스를 0.1에서 0.4로 증가시킬 때 HN의 증가에 따라 APA는 일관성 있게 증가한다. 이는 바이어스가 -0.4인 경우와 대조적이다. 이상으로부터 양과 음의 바이어스에 대한 HN의 영향이 크게 다르다는 것을 알 수 있다. 결론적으로 고정된 HN에 따른 APA의 바이어스에 따른 변화는 비슷하지만 고정된 바이어스에 따른 HN의 영향은 바이어스에 따라 매우 달랐다. 최적의 APA는 HN이 8이고 바이어스가 -0.3과 -0.4에서 발생하였으며 그 수치는 0.315이다. 앞서의 ATT분석과 비교할 때 최적화된 학습인자 데이터 분류인자 값이 일치하는 점이 흥미롭다.

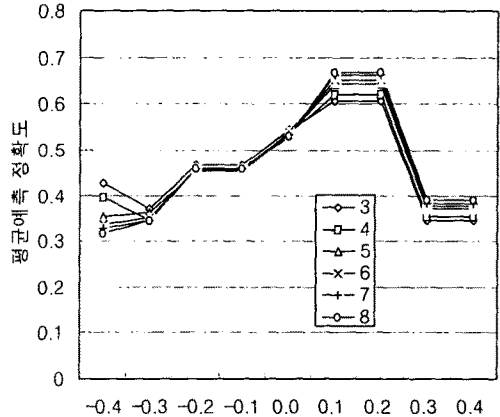


그림 3. 바이어스와 뉴런수에 따른 예측정확도

### 2.3.3 최적예측 정확도(BPA)

그림 4는 BPA를 바이어스와 HN의 수에 따라 도시하고 있다. 예측정확도는 다음의 수식에 의해 계산을 하였다.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}} \quad (4)$$

여기서  $y_i$ 와  $\hat{y}_i$ 는 실제 측정치와 모델로부터의 예측치를 지칭하고,  $n$ 은 테스트 데이터의 수이다. 본 연구에서의  $n$ 은 6이다. 그림 4의 각 점은 주어진 학습조건에서 예측 모델을 200개 발생하고, 그 중 결정된 최적의 모델에 대한 BPA와 일치한다. HN이 3인 경우 바이어스가 -0.4에서 0.2까지 증가할 때, 정확도는 일관성 있게 증가한다. 바이어스가 0.3이상으로 증가 할 때는 급격히

감소하며, 따라서 -0.2에서 전환점이 있음을 알 수 있다. HN이 4이상인 경우 -0.3에서 전환점이 발생하며, -0.2이상인 바이어스에 대해서는 HN이 3인 경우와 비슷하다. 결국 모든 HN에 대해서 0.3에서 그 전환점이 발생한다. 또한 모든 HN에 대해서 BPA는 -0.3에서 발생하며, 이는 일반적으로 사용하는 데이터 (즉 바이어스가  $[-0.4, 0]$ 의 범위에 대해서 HN의 변화에 따라 BPA는 불규칙적으로 변화하였으며, 이 같은 현상은 최근의 보고와 일치한다 [4]. 바이어스가 0.1과 0.2인 경우 HN의 증가에 따라 정확도는 비교적 일관성있게 증가하였고, 0.3과 0.4인 경우에는 그 변화가 미미하였다. BPA는 HN이 3이고 바이어스가 -0.4인 경우 발생하였으며, 그 수치는 0.0358이었다.

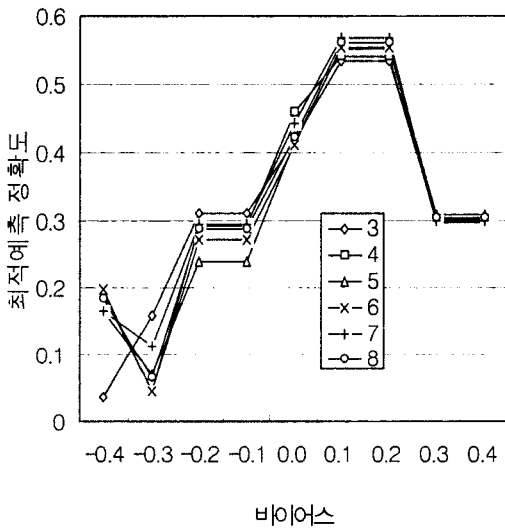


그림 4. 바이어스와 뉴런수에 따른 예측정확도

### 2.3.4 분류정확도(CA)

표 1은 CA를 바이어스와 HN의 수에 따라 도시하고 있다. 표 안의 각 데이터는 앞서 결정된 최적의 모델에 대한 CA이다. 고정된 HN에 대해서 CA의 분포는 불규칙적이다. 최적 CA는, 수치적으로는 6개에 해당. 총 4개의 조건에서 결정이 되었으며, 이들의 공통점은 바이어스가 -0.3에서 발생한다는 사실이다. 앞서 BPA에 대응하는 CA는 표 1로부터 5임을 알 수 있다. 이는 BPA를 기초로 결정된 최적의 모델이 최적의 CA를 보이지 않는다는 것을 단적으로 입증한다. 오히려, 바이어스를 기초로 데이터를 분류하였을 때는 앞서 진술한 바와 같이 여러 곳에서 최적의 CA를 달성하였는바, 이는 통계적인 CA의 중요성을 의미한다. 그림 4와 표 1로부터 BPA와 최적의 CA는 HN이 4이고 바이어스가 -0.3일 때 언어짐을 알 수 있다.

## 3. 결 론

데이터의 통계적인 분포를 고려하여 BPNN의 분류 성능을 최적화하는 알고리즘을 개발하였다. 데이터의 평균치와 표준편차를 계산하였으며, 바이어스 인자를 표준편차에 인가하여 다양한 학습과 테스트 데이터를 발생하였다. 바이어스 인자와 학습인자 (온닉 뉴런수)의 상호

작용에 따른 BPNN의 성능변화를 살펴보았으며, 성능은 ATT, APA, BPA 그리고 CA로 세분화하여 살펴보았으며 이 들간의 연관성도 살펴보았다. 각 평가인자들은 데이터 분류인자 바이어스와 HN의 수에 따라 각기 다양한 변화를 보였으며, 비교적 작은 바이어스에 대해서 최적의 예측성능과 분류성능을 가진 모델을 개발할 수 있었다. 그리고 바이어스와 뉴런수의 변화에 따라 APA와 BPA 그리고 BPA와 CA의 경향이 일치하지 않았다. 결론적으로 데이터의 통계적인 분류와 BPNN의 학습인자를 동시에 최적화 함으로써 최적의 예측과 분류 성능을 지닌 모델을 개발할 수 있었다.

표 1 바이어스와 뉴런수에 따른 분류정확도

	바이어스			Hidden Neuron		
	3	4	5	6	7	8
-0.4	5	5	5	5	5	5
-0.3	5	6	6	5	6	6
-0.2	3	5	5	3	4	3
-0.1	3	5	5	3	4	3
0	3	5	5	5	5	5
0.1	5	4	3	4	4	3
0.2	5	4	3	4	4	3
0.3	5	4	5	4	4	5
0.4	5	4	5	4	4	5

### 감사의 글

본 연구는 IMT-2000 출연금 기술개발 사업에 의해 지원되었으며, 이에 한국보건산업진흥원에 감사를 드립니다.

### [참고 문헌]

- [1] B. Kim and G. T. Park, "Modeling plasma equipment using neural networks," *IEEE Trans. Plasma Sci.*, vol. 29, no. 1, 8-12, 2001.
- [2] B. Kim, S. M. Kong, and B. T. Lee, "Modeling SiC etching in  $C_2F_6/O_2$  inductively coupled plasma etching," *J. Vac. Sci. Technol. A*, vol. 20, no. 1, 146-152, 2002.
- [3] B. Kim and G. S. May, "An optimal neural network process model for plasma etching," *IEEE Trans. Semicond. Manufact.*, vol. 7, 12-21, 1994.
- [4] B. Kim and S. Park, "An optimal neural network plasma model: a case study," *Chemom. Intell. Lab. Syst.* vol. 65, 39-50, 2001.
- [5] B. Kim, J. H. Sun, C. J. Choi, D. D. Lee, and Y. S. S, "Use of neural networks to model low temperature tungsten etch characteristics in  $SF_6$  plasma," *J. Vac. Sci. Technol. A*, vol. 18, no.2, 417-422, 2000.
- [6] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature*, vol. 323, 533 (1986).