

통계적 처리를 이용한 일반화된 회귀 신경망의 분류성능의 최적화

김근호, 김병환
 세종대학교 전자공학과

Optimization of Generalized Regression Neural Network
 Using Statistical Processing

Geun-Ho Kim, Byunwhan Kim
 Sejong University, Electronic Engineering

Abstract - 일반화된 회귀 신경망 (GRNN)을 이용하여 플라즈마를 분류하는 새로운 알고리즘을 보고한다. 데이터분포를 통계적인 평균치와 표준편차를 이용하여 특징지었으며, 바이어스 인자를 이용하여 9 종류의 데이터를 발생하였다. 각 데이터에 대하여 GRNN의 학습인자를 최적화하였으며, 모델성능은 예측과 분류 정확도로 나누어 바이어스와 학습인자의 함수로 분석하였다. 바이어스는 모델성능에 상당한 영향을 주었으며, 학습인자와의 상호작용을 통하여 완전 분류를 이루었다.

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 모델링과 관계하는 물리적 메커니즘의 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 일반화된 회귀 신경망 [5]을 학습데이터의 분류에 응용한다. 이를 위하여 학습데이터의 특성을 통계적인 평균치와 편차로 수량화하였으며, 평균치에 바이어스 인자가 곱해진 편차를 가중치로 인가하여 분류하였다. 각 분류된 데이터에 대해서 학습인자인 "spread"를 변화시키며, 모델의 예측과 분류성능을 살펴보았다. 모델성능은 예측정확도와 분류정확도측면에서 고찰하였다. 분류에 이용된 데이터는 광도파로 제작에 이용된 실리콘 박막의 식각공정중에 수집 [6] 되었으며, 식각인자중 산화막 프로파일에 대한 데이터를 이용하였다.

2. 본 론

2.1 데이터 수집과 분류

그림 1은 데이터 수집을 위해 식각공정을 수행한 장

비의 개요도이다. 식각공정은 ICP 식각장비를 이용하여 CF₄/CHF₃ 가스 플라즈마하에서 실리콘 박막을 식각하였다. 공정에 이용된 공정변수는 총 4종류이며, 여기에는 소스전력 바이어스 전력, 그리고 두 종류의 가스(CHF₃와 CF₄)이다. 부분인자 실험계획표에 의거하여, 총 8번의 실험의 수행이 되었고 여기에 중앙점에 해당하는 실험을 첨가하여, 학습데이터를 구성하였다. 모델의 성능을 검증하고자 추가적으로 6번의 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.

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

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(HN)만을 변화시키며, 모델성능을 평가하였다. 뉴런수는 3에서 8까지 하나씩 증가시켰다. 모델성능은 예측정확도와 분류정확도로 나누어 각기 spread와 바이어스에 따른 변화를 고찰한다.

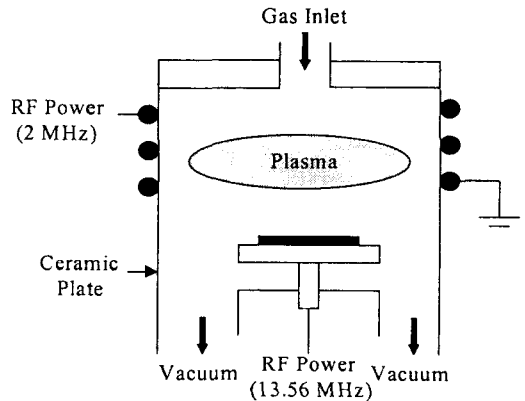


그림 1. 식각장비의 개요도

2.2 일반화된 회귀 신경망

일반화된 회귀 신경망 (generalized regression neural network) [5]의 예측치는 다음과 같이 표현된다.

$$\hat{Y} = \frac{\sum_{i=1}^n y_i \exp[-D(x, x_i)]}{\sum_{i=1}^n \exp[-D(x, x_i)]} \quad (1)$$

여기서 함수 D는

$$D_x(x, x_i) = \sum_{j=1}^n \left(\frac{x_j - x_{ij}}{\sigma_j} \right)^2 \quad (2)$$

로 표현되고, 변수 n과 p는 각각 독립변수와 학습 패턴의 수를 지칭한다. 그리고 변수 σ 는 폭 (spread) 변수 (또는 smooth factor)라 칭하며, 이 폭이 작은 경우, 좋은 패턴 인식을 이룰 수 있으나, 모델의 일반화 능력을 저하시키며, 폭이 넓을 경우 반대의 현상이 발생한다. 따라서 최적의 일반화 능력을 얻기 위해 이 값의 조정이 필요하다.

GRNN은 총 4개의 층, 즉 입력층, 패턴층, 합층, 그리고 출력층으로 구성된다. 입력층의 뉴런수는 독립변수 (예 공정 변수)의 수와 일치하며, 패턴층의 뉴런수는 학습패턴의 수와 일치한다. 합층은 두 개의 뉴런으로 구성되며, 처음의 뉴런은 패턴층의 모든 출력을 모두 합해, 식 (1)의 분자 항을 계산하며, 다른 하나의 뉴런은 분모 항을 계산한다. 입력층과 패턴층간의 하중치 (W_p)는 입력패턴 (X)에 의해 결정되며, 이를 표현하면,

$$W_p = X^T \quad (3)$$

패턴층의 하나의 뉴런은 합층의 두 개의 뉴런에 연결되며, 패턴층의 i번째의 뉴런과 합층 첫 뉴런간의 연결 하중치는 y_i 가 된다. 이 i번째의 뉴런과 합층의 다른 하나의 뉴런과의 연결 하중치는 1이 된다. 합층과 출력층간의 하중치 (W_s)는 y_i 와 1에 의해 다음과 같이 결정된다.

$$W_s = [Y \text{ ones}] \quad (4)$$

출력층에서는, 단순히 합층의 두 뉴런의 출력을 나누어 식(1)의 예측치를 출력한다.

2.3 분류 성능의 최적화

2.3.1 예측정확도에의 영향

그림 2는 GRNN의 예측성능을 spread와 양의 바이어스의 함수로 도시하고 있다. 예측정확도는 다음의 수식을 이용하여 계산하였다.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}} \quad (5)$$

여기서 y_i 와 \hat{y}_i 는 실제 측정치와 모델로부터의 예측치를 지칭하고, n은 테스트 데이터의 수이다. 본 연구에서의 n은 6이다. 바이어스가 0일 때 예측 정확도는 spread가 0.5에서 0.9까지 증가함에 따라 계속 감소한다. 한편, 0.9이상으로 spread를 증가시킬 때에는 예측정확도가 점진적으로 증가한다. 결국, 가장 우수한 모델은 spread가 0.9에서 결정된다. 해당 모델의 예측정확도는 0.451이다. 바이어스를 0.1이상으로 증진할 때는 바이어스가 0인 경우와 같이 작은 spread에 대해서는 그 값이 증가될 때 급격히 감소한다. 그러나 보다 값이 큰 spread에 대해서는 spread의 증가에 관계없이 예측정확도의 변화는 미미하다. 대략 1.0보다 작은 spread에 대해서 바이어스가 증가할 때의 예측정확도의 분포는 불규칙적이다. 이러한 현상은 1.0보다 큰 spread에 대해서도 나타난다. 즉 바이어스의 예측정확도에의 영향은 불규칙적임을 알 수 있다. 한편 각 바이어스에 대해서 얻은 최적의 모델을 비교해 볼 때 바이어스가 0인 경우 얻은 모델이 가장 우수하였으며, 이는 양의 바이어스를 기준으로 데이터를 분류하였을 때 모델 정확도를 향상시키지 못한다는 것을 단적으로 보여준다.

그림 3은 GRNN의 예측성능을 spread와 음의 바이어스의 함수로 도시하고 있다. 바이어스가 -0.1인 경우 최적의 모델은 spread가 1.0일 때 얻어진다. 해당 모델

의 예측정확도는 0.423이며, 바이어스가 0인 경우 얻은 모델과 비교해 볼 때 6.2%정도의 향상된 정확도를 보였다. 바이어스가 -0.3인 경우, spread가 1.3에서 최적의 모델이 발생하였으며, 그 정확도는 0.361이고, 이는 바이어스가 -0.1인 경우의 모델에 비해서, 19.9% 향상된 정확도이다. 최적의 spread보다 spread를 증가할 때, 예측정확도는 점진적으로 증가하고 있다. 그림 3에 나타나 있듯이, 바이어스가 -0.4일 때 매우 특이한 현상이 발생한다. 작은 spread에 대해서 그 값을 증가할 때, 예측정확도는 증가한다. 이는 앞서의 경우와 반대되는 현상이다. 이에 따라 spread가 0.5에서 최적의 모델이 결정되었으며, 모델의 정확도는 0.126이다. 이는 바이어스가 0인 경우에 비해서 그 정확도가 72.0% 향상된 정확도이다. 작은 spread에 대해서 음의 바이어스 절대치가 증가함에 따라 예측정확도는 계속해서 감소하고 있다. 이러한 현상은 큰 spread에 대해서도 나타나고 있으며, 이는 양의 바이어스에 대해서 관찰되지 않은 현상이다. 결국 기준데이터를 음의 바이어스를 기초로 하여 분류하였을 때, 예측정확도는 바이어스와 spread에 따라 큰 변화를 보였으며, 그 단적인 예가 바이어스가 -0.4에 해당하는 경우이다. 최적의 모델은 바이어스가 -0.4, 그리고 spread가 0.5에서 결정되었다.

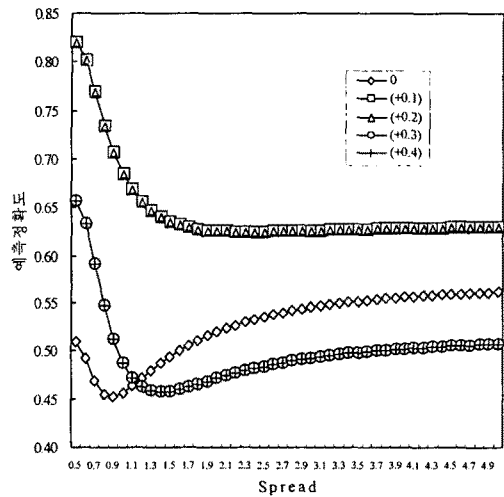


그림 2. 양의 바이어스와 spread에 따른 예측정확도

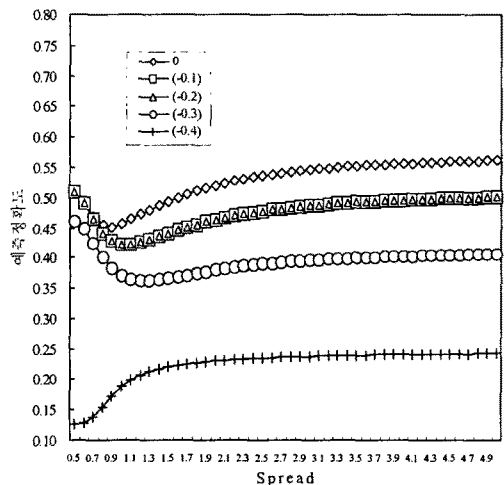


그림 3. 음의 바이어스와 spread에 따른 예측정확도

2.3.2 분류정확도에의 영향

그림 4는 GRNN의 분류성능을 spread와 양의 바이어스의 함수로 도시하고 있다. 바이어스가 0인 경우 분류성능은 spread의 변화에 따라서 불규칙적으로 변하고 있다. Spread가 0.5와 1사이, 그리고 1.5와 2.4사이에 분포할 때 최고의 분류성능을 보이고 있으며, 그 수치는 5개이다. Spread를 2.4이상으로 증진하였을 때는 분류성능이 급격히 저하된다. 바이어스가 0.1과 0.2인 경우 분류성능은 매우 저하되며, 모든 spread에 대해서 2로서 일정하다. 한편 바이어스가 0.3과 0.4로 증가될 때, 분류성능은 다소 증가되나, 그 수는 4로서 일정하다. 모든 spread에 대해서 양의 바이어스의 변화에 따른 분류성능의 변화는 불규칙적임을 그림 4로부터 알 수 있다.

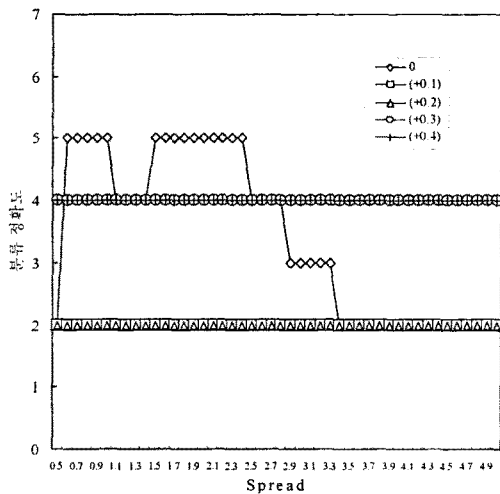


그림 4. 양의 바이어스와 spread에 따른 분류정확도

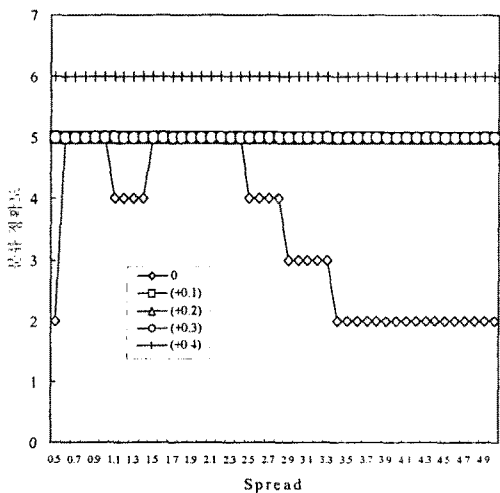


그림 5. 음의 바이어스와 spread에 따른 분류정확도

그림 5는 GRNN의 분류성능을 spread와 양의 바이어스의 함수로 도시하고 있다. 바이어스가 -0.1, -0.2, -0.3인 경우 분류성능은 5로서 모든 spread에 대해서 일정하다. 비록 그림 3으로부터, 예측정확도는 spread

의 변화에 따라 변하였지만, 분류성능을 향상시키지는 못하였다. 하지만 바이어스가 -0.4인 경우 분류성능은 6이며, 따라서 완전한 분류를 달성하였다. 이는 그림 3의 결과로부터 예측된 결과이다. 그림 3에서 바이어스가 -0.4일 때 모델의 예측정확도는 다른 바이어스에 해당하는 모델에 비해서 훨씬 우수하였기 때문이다. 그림 3와 그림 5로부터 결국 최적의 모델은 spread가 0.5에서 결정된다. 이 spread에 대해서 최적의 예측정확도와 완전 분류를 달성하였기 때문이다. 결국 바이어스의 변화에 따라 분류성능은 상당한 영향을 받았으며, 바이어스가 0인 기준 데이터를 제외하고는 spread 변화에 따른 영향은 없었다.

3. 결 론

데이터의 통계적인 분포를 고려하여 GRNN의 분류성능을 최적화하는 알고리즘을 개발하였다. 데이터의 평균치와 표준편차를 계산하였으며, 바이어스 인자를 표준편차에 인가하여 다양한 학습과 테스트 데이터를 발생하였다. 각 데이터에 대해 GRNN의 학습인자인 spread를 변화시켜 모델의 예측과 분류성능을 계산하였다. 예측성능은 양의 바이어스에 대해서는 기준 데이터인 바이어스가 0인 경우에 비해 예측과 분류성능을 저하시키었지만, 음의 바이어스에 대해서는 두 성능을 개선하였다. 특히 예측성능의 경우 바이어스가 -0.4인 경우 비약적인 향상이 있었으며, 완전 분류를 달성하였다. 결론적으로 주어진 데이터와 GRNN의 학습인자를 동시에 최적화함으로써 예측과 분류성능이 비약적으로 향상된 모델을 개발할 수 있었다.

감사의 글

본 연구는 IMT-2000 출연금 기술개발사업에 의해 지원되었으며, 이에 한국보건산업진흥원에 감사함을 느낍니다.

[참 고 문 헌]

- [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] D. F. Specht, "A generalized regression neural network," *IEEE Trans. Neural Networks*, vol. 2, no. 6, 568, 1991.
- [6] 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.