

Auto-Encoder 모델 기반 반도체 장비 고장 예측 연구

윤재화¹, 윤혜원¹, 김영진¹, 이장후¹, 안근진², 박승범³, 배병성⁴

¹호서대학교 전자융합공학부 학부생

²(주)시성 대표이사

³호서대학교 기술경영전문대학원 교수

⁴호서대학교 반도체공학과 교수

yunjaehwa92@gmail.com, hyewon08011@naver.com, kevin8445@naver.com,
alsxmdkdlm@naver.com, akj2995@daum.net, parks@hoseo.edu, bsbae3@hoseo.edu

Research on Semiconductor Equipment Failure Prediction Based on Auto-Encoder Model

Jaehwa Yun¹, Hye Won Yun¹, Yeong Jin Kim¹, Jang Hoo Lee¹, Guen Jin Ahn², Sungbum Park³, Byung Seong Bae⁴

¹Dept. of Electronic Convergence Engineering, Hoseo University

²Rain ICT, Co.

³Graduate School of Management of Technology, Hoseo University

⁴Dept. of Semiconductor Engineering, Hoseo University

요 약

반도체 장비는 여러 가지 요인(화재, 부품 고장 및 노후화 등)에 의해 가동이 중지되거나 더 이상 운용할 수 없는 경우가 자주 발생한다. 반도체 생산 중 장비의 고장은 전체 제조 공정을 중단시켜 기업의 생산 원가 비율을 상승시키는 원인이 된다. 이를 해결하기 위해 장비에 온도, 진동, 전력 센서를 부착하고, 각각의 센서에서 실시간으로 수집되는 데이터를 Auto-Encoder 모델로 분석하여 반도체 장비의 고장의 사전 예측 정확도를 향상시켰다.

1. 서론

반도체 장비는 각종 소재를 제작하거나 패키징 및 테스트 등에 있어서 필수적인 설비이다. 이러한 장비가 가동 중단으로 인해 발생하는 생산 초과 비용은 기업의 이익에 상당한 영향을 미치게 된다. 따라서 반도체 장비 고장을 미리 예측하여 그 발생 원인을 사전에 제거함으로써 가동 시간을 최대로 유지하고, 연계된 다른 장비에 미치는 영향 범위를 최소화할 수 있다. 특히 고장 원인의 범위는 장비 온도 상승, 화재, 전기적인 문제와 같은 장비 외적인 부분과 반도체 장비 부품의 마모나 고장 등 장비 내적인 부분과 같이 여러 각도에서 발생할 수 있다. 또한, 외적인 부분과 내적인 부분이 서로 연계되어 고장이 발생하는 경우도 배제할 수는 없다. 이번 연구는 이러한 외적인 부분과 내적인 부분을 동시에 관찰함으로써 고장을 사전에 예측하고, 고장의 원인을 사전에 대응하기 위한 방법을 제시한다.

센서 데이터를 실시간으로 수집하고, 수집된 데이터

를 동일한 시점의 시간 범위에서 동시에 접근하므로, 센서 간의 데이터 연관성에 대한 연구가 이루어져야 하기 때문에 데이터를 표준화하는 부분이 중요하다. 각각의 센서 데이터를 RAW 신호의 관점에서 바라보는 것이 아니라, 신호 데이터를 통계적인 관점과 시간적인 관점에서 분석하기 위해 FFT(Fast Fourier Transform) 분석 같은 스펙트럼 데이터로 전환하여 분석하고자 한다.

데이터 분석에 사용되는 기법은 결정 트리의 일종인 Xgboosting, Random Forest 같은 머신러닝 기법과 DNN(Deep Neural Network, 심층 신경망), RNN(Recurrent Neural Network, 순환 신경망), LSTM(Long Short-Term Memory, 장단기 메모리), Auto-Encoder 모델과 같은 딥러닝 기법을 사용하여 각 모델의 고장 예측 정확도를 비교 분석하고자 한다. 센서 데이터 수집, 데이터 표준화, 학습 데이터 제작, 시간 도메인 및 통계 도메인 변환, 머신러닝 및 딥러닝 모델 학습 및 평가 등의 순서로 연구가 진행된다.

2. 연구내용

그림 1과 같이 로터리 펌프에 진동 센서를 부착하고, 이를 신호 취득 및 전송하여 저장한 뒤 분석에 이용하였다. 그림 1에서 파란 화살표는 진동 센서를, 빨간 화살표는 온도 센서를 나타낸다. 온도 센서는 장비에서 발생하는 이상 마찰에 의한 발열을 모니터링할 수 있도록 하며, 진동과 함께 장비의 이상 여부를 잘 보여준다.



그림 1. 진동 센서 및 온도센서의 부착

데이터 수집 게이트웨이(엣지컴퓨터)에 수집된 데이터는 시간을 일치시켜 표준화를 진행하였고 모델 학습 및 평가를 위해 19개의 Feature 항목으로 변환하였다. 통계 속성 함수는 13개 Function 시간 도메인 즉 FFT 변환 Function은 6개 항목이다.

표 1. 19개 Feature Function.

Feature name	Function symbol	Feature name	Function symbol
Mean	$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$	Skewness	$\frac{1}{N} \sum_{i=1}^N \frac{(x_i - \bar{x})^3}{(N-1)(N-2)^{3/2}}$
Max	$\max(x_i)$	Standard Deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$
Peak	$P_m = \max(x_i)$	Power	$\sum_{i=1}^N x_i^2$
Peak to Peak	$P_k = \max(x_i) - \min(x_i)$	Mean of band Power Spectrum	$S_{\mu} = \frac{1}{n} \sum_{i=1}^n S(f)_i$
RMS	$\sqrt{\frac{\sum_{i=1}^N x_i^2}{N}}$	Max of band power spectrum	$S_{MAX} = \max(S(f)_i)$
Variance	$\sum_{i=1}^N \frac{(x_i - \bar{x})^2}{N-1}$	Sum of total band Power	$S_{SBP} = \sum_{i=1}^n S(f)_i$
Crest Factor	$\frac{P_m}{RMS}$	Variance of band power	$S_V = \frac{\sum_{i=1}^n (S(f)_i - S_{\mu})^2}{n-1}$
Pulse Indicator	$\frac{P_m}{\bar{x}}$	Skewness of band power	$S_K = \frac{1}{n} \sum_{i=1}^n \frac{(S(f)_i - S_{\mu})^3}{S_V^{3/2}}$
Margin	$ \frac{\sum_{i=1}^N x_i}{N} $	Kurtosis of band power	$S_K = \frac{1}{n} \sum_{i=1}^n \frac{(S(f)_i - S_{\mu})^4}{S_V^2}$
Kurtosis	$\sum_{i=1}^N \frac{(x_i - \bar{x})^4}{N-1}$		

학습 데이터 구성시 Feature에 대한 데이터를 구성하는 이유는 RAW 데이터에 대한 방대한 양의 데이터를 처리하기 위해 계산 부하를 최소화 하기 위한 과 FFT 처리를 통해 주파수 성분 관점에서 분석하

기 위함이다.

그림 2는 19개 항목에 대한 상관관계를 도식화 하여 설명하였다. 각 Feature들은 서로 간의 연관성이 높은 부분에 대해서 잘 표현하고 있다.

	MIN	MAX	MEAN	RMS	VAR	STD	POWER	PEAK	P2P	CREST FACTOR	SKW	KURTOSIS	MAX.F	SUM.F	MEAN.F	VAR.F	PEAK.F	SKW.F	KURTOSIS.F	label
MIN	1.00	-0.22	0.29	0.19	-0.45	-0.52	0.17	-0.22	-0.53	-0.51	-0.04	0.13	-0.15	0.17	0.24	-0.28	-0.18	-0.04	0.13	-0.28
MAX	-0.22	1.00	0.09	0.77	0.62	0.81	0.77	1.00	0.94	0.79	0.25	0.16	0.02	0.77	0.57	0.56	0.92	0.25	0.16	0.81
MEAN	0.29	0.09	1.00	0.99	0.47	0.44	0.90	0.98	0.90	0.13	-0.22	0.25	0.72	0.98	0.98	0.58	0.72	-0.22	0.25	0.27
RMS	0.19	0.77	0.99	1.00	0.60	0.56	1.00	0.77	0.81	0.23	-0.17	0.21	0.78	1.00	0.87	0.87	0.78	-0.17	0.21	0.34
VAR	-0.45	0.62	0.47	0.60	1.00	0.98	0.61	0.65	0.86	0.68	0.15	-0.17	0.72	0.61	0.41	0.83	0.72	0.15	-0.17	0.50
STD	-0.52	0.81	0.44	0.56	0.98	1.00	0.58	0.81	0.88	0.71	0.17	-0.20	0.71	0.58	0.37	0.89	0.71	0.17	-0.20	0.40
POWER	0.17	0.77	0.90	1.00	0.61	0.58	1.00	0.77	0.81	0.24	-0.17	0.20	0.78	1.00	0.86	0.89	0.78	-0.17	0.20	0.35
PEAK	-0.22	1.00	0.09	0.77	0.62	0.81	0.77	1.00	0.94	0.79	0.25	0.16	0.02	0.77	0.57	0.76	0.92	0.25	0.16	0.81
P2P	-0.53	0.94	0.50	0.81	0.88	0.88	0.61	0.94	1.00	0.85	0.23	0.09	0.85	0.61	0.42	0.73	0.85	0.23	0.09	0.83
CREST FACTOR	-0.51	0.79	0.13	0.23	0.68	0.71	0.24	0.78	0.81	1.00	0.55	0.06	0.65	0.24	0.16	0.48	0.55	0.55	0.06	0.82
SKW	-0.04	0.25	-0.22	-0.17	0.15	0.17	-0.17	0.25	0.23	0.56	1.00	-0.27	0.18	-0.17	-0.22	0.07	0.18	1.00	-0.27	0.29
KURTOSIS	0.13	0.16	0.25	0.21	-0.17	-0.20	0.20	0.16	0.09	0.05	-0.27	1.00	0.22	0.20	0.27	-0.05	0.22	-0.27	1.00	0.16
MAX.F	-0.15	0.02	0.72	0.78	0.72	0.71	0.78	0.92	0.85	0.65	0.18	0.23	1.00	0.78	0.78	0.85	1.00	0.18	0.22	0.80
SUM.F	0.17	0.77	0.98	1.00	0.61	0.58	1.00	0.77	0.81	0.24	-0.17	0.20	0.78	1.00	0.86	0.89	0.78	-0.17	0.20	0.35
MEAN.F	0.24	0.57	1.00	0.97	0.41	0.37	0.98	0.97	0.42	0.06	-0.22	0.27	0.78	0.98	1.00	0.69	0.78	-0.22	0.27	0.29
VAR.F	-0.19	0.14	0.58	0.67	0.82	0.80	0.69	0.74	0.73	0.48	0.07	-0.05	0.85	0.69	0.69	1.00	0.85	0.07	-0.05	0.50
PEAK.F	-0.15	0.02	0.72	0.78	0.72	0.71	0.78	0.92	0.85	0.65	0.18	0.23	1.00	0.78	0.78	0.85	1.00	0.18	0.22	0.80
SKW.F	-0.04	0.25	-0.22	-0.17	0.15	0.17	-0.17	0.25	0.23	0.56	1.00	-0.27	0.18	-0.17	-0.22	0.07	0.18	1.00	-0.27	0.29
KURTOSIS.F	0.13	0.16	0.25	0.21	-0.17	-0.20	0.20	0.16	0.09	0.05	-0.27	1.00	0.22	0.20	0.27	-0.05	0.22	-0.27	1.00	0.16
label	0.28	0.81	0.27	0.34	0.50	0.49	0.35	0.91	0.83	0.82	0.29	0.18	0.88	0.35	0.29	0.50	0.89	0.29	0.15	1.00

그림 2. 19개 항목간 상관관계

사전 고장 예측을 위해 데이터의 고장 발생일을 기준으로 1시간 전, 2시간 전, 3시간 전으로 데이터를 구분하고, 실제 발생 시간의 데이터는 데이터셋에서 제외하였다. 이를 통해 사전 고장 예측이라는 목표를 달성하기 위한 데이터셋을 구성하였다. 또한, 데이터를 조회 및 다운로드하여 분석할 수 있도록 모니터링 웹 시스템을 구축하였다.

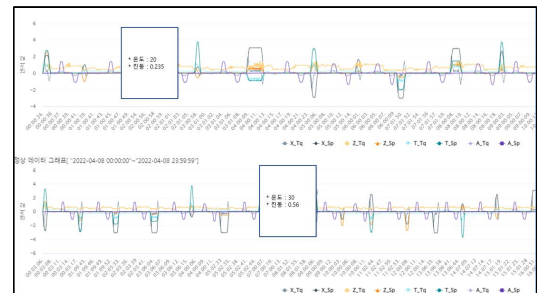


그림 3. 센서 데이터 조회 및 다운로드 시스템

그림 4는 통계 모델과 딥러닝 모델과의 학습 결과에 대한 비교 그림이다.

Table 1. Average Accuracy of Anomaly Detection prediction				
반도체 제조장비 A				
feature 19 dataset	온도, 진동, 전력 ALL	온도, 진동 전력 (-1시간전)	온도, 진동 전력 (-2시간전)	온도, 진동 전력 (-3시간전)
RandomForest	0.8220	0.9140	0.9070	0.9040
xgboosting	0.8888	0.9214	0.9298	0.8888
DNN	0.8700	0.9167	0.9060	0.8900
RNN	0.8500	0.9367	0.9230	0.9000
Auto-Encoder	0.9200	0.9500	0.9300	0.9120

그림 4. 5개 모델에 대한 정확도 평가표

3. 요약 및 향후 연구

5개의 모델 결과 중 Auto-Encoder 모델의 경우, 고장 발생 1시간 전의 데이터셋을 기준으로 학습한 결과의 테스트에서 고장 예측 정확도(정확도: 0.95, 1시간 전)가 높게 나왔다. 보통 Auto-Encoder는 RNN, LSTM 등의 시계열 모델 보다 예측 성능에 있어서 데이터 압축에 초점을 두고 있고 시간 의존성(Time Dependency)를 잘 포착하지 못하기 때문에 성능이 상대적으로 떨어지는 것으로 예상되어 이번 연구에서는 Auto-Encoder 모델에 LSTM 모델을 결합하여 시간 종속성 문제를 해결함으로써 예측 성능을 높인 부분이 잘 적용된 것 같다. 이번 실험을 통해, 반도체 장비에 부착된 온도, 진동, 전력 데이터 신호 중 고장 발생 1시간 전의 데이터를 대상으로 Auto-Encoder 모델을 통해 실제 반도체 장비에서 들어오는 센서 데이터를 추론할 경우, 고장 발생 예측 정확도를 측정할 수 있다는 결론에 이르렀다. 이러한 결과는 반도체 장비 고장을 1시간 전에 사전에 예측할 수 있다는 것을 보여주며, 장비 운영 담당자가 추론 결과에 대한 알람 메시지를 받고 고장 원인에 선제적으로 대응할 수 있음을 시사한다.

본 연구는 특정 반도체 장비를 대상으로 진행한 결과이기 때문에, 다른 반도체 장비에도 동일한 결과를 보장한다고는 할 수 없다. 따라서 다양한 반도체 장비에 대한 센서 데이터 분석 결과를 비교할 필요가 있다.

사 사

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음. This research was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (Ministry of Education-Ministry of Trade, Industry and Energy) (P0022185, Semiconductor Major Track)

참고문헌

- [1] Lee, J., Qiu, H., Yu, G., Lin, J. (2007) Rexnord :Technical Services, 'Bearing Data Set', IMS, University of Cincinnati, NASA Ames Prognostics Data Repository.
- [2] Park, S., Kim, J.W., Lee, C., Lee, J., Gil, H.B. (2012) Local Fault Detection Technique for Steel Cable using Multi-Channel Magnetic Flux Leakage Sensor, J. Comput. Struct. Eng. Inst. Korea, 25(4), pp.287~292.