

컨셉 드리프트를 고려한 조기탐지 및 해석 프레임워크

강민정¹, 오수빈¹, 이상민²

¹ 광운대학교 인공지능응용학과 석사과정

² 광운대학교 인공지능응용학과 교수

minjkang0901@gmail.com, sboh0020@gmail.com, smlee5679@gmail.com

A Framework for Early Detection and Interpretation of Concept Drift

Min-Jung Kang¹, Su-Bin Oh¹, Sang-Min Lee²

¹Dept. AI Applications, Kwangwoon University

²Dept. of AI Applications, Kwangwoon University

요약

본 연구는 반도체 제조 과정에서 생산 가용 능력이 저하되는 시점을 조기 탐지하기 위한 프레임워크를 제안한다. 이를 위해 데이터 패턴의 불규칙한 변동이 잦은 환경에서 모델의 재학습 없이 최적의 성능을 유지할 수 있도록 온라인 학습 방식을 활용하였다. Augmented Dicky-Fuller test를 통해 데이터의 정상성 여부를 검정하고, 데이터에 변화가 있을 경우 학습 모델은 지속적으로 업데이트된다. 특히, 상한 재공재고는 생산량과 직결되는 주요 지표로써, 낮게 예측된 시점에서 주요 원인 변수를 파악하는 것이 중요하다. 따라서 정확도와 효율성 측면에서 다른 모델 대비 가장 우수한 성능을 보였던 제안 기법에 shapley additive explanations(SHAP)을 적용하여 생산 저하 시 문제가 되는 원인 변수를 분석하고자 하였다.

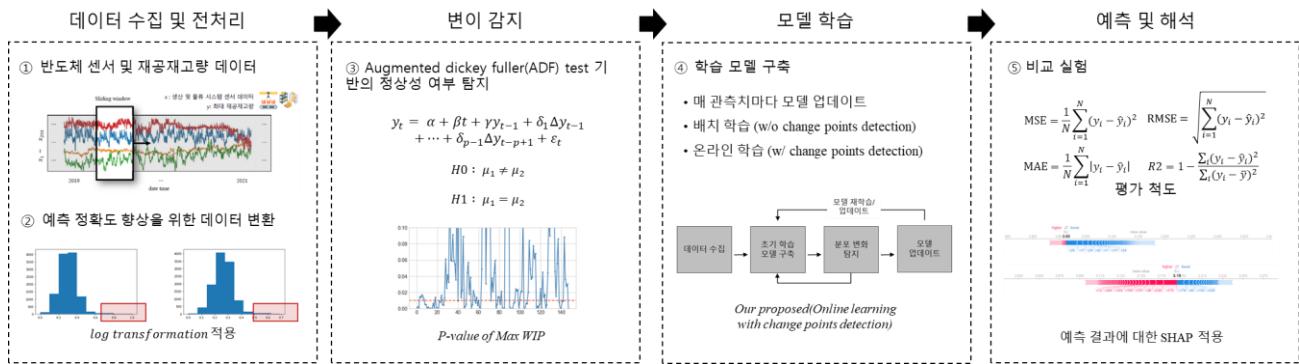
1. 연구 배경

반도체 제조는 미세 공정에 따른 리드타임의 증가와 대규모 생산 시스템 유지를 위해 제조 환경 관리가 필수적이다. 특정 제품군에서는 수요 감소로 인해 재고가 증가하며 공급 초과를 불러일으킬 수 있다. 따라서 변동성이 유연하게 대응할 수 있는 생산 시스템과 적절한 생산능력 확인이 필수적이다. 이를 위해 반도체 회사들은 라인 최적화와 공정 전환을 실시하며 다양한 지표를 모니터링 하고 있다. 그 중에서 재공재고(Work-In-Process, WIP) 수준은 반도체 제조 과정의 생산성을 나타내는 중요 지표로 사용된다[1]. 최대 재공재고는 반도체 생산에 필요한 인프라 시스템 가용 능력과 직결되기에 정확한 재공재고 추정은 필수적이다. 그러나 제조 환경은 시간에 따라 변이하는 특성을 갖는다. 불규칙한 데이터 변이가 발생할 경우 기존 구축된 학습모델의 성능은 저하된다[2].

본 연구에서는 이러한 배경 하에 데이터 분포 변화

를 감지하는 가설 검증 기반 방법론과 자가학습이 가능한 프레임워크를 도입하여 최적의 학습모델을 제안한다. 제안 기법은 첫째로, 시계열 데이터의 정상성 평가에 사용하는 augmented Dicky-Fuller(ADF) test를 통해 단위근 내 데이터의 변이를 탐색한다[3]. 두 번째로 데이터 변이 시점을 감지할 때마다 주어진 데이터 내에서 최상의 일반화 성능을 유지하기 위해 모델 자가학습 방식을 활용한다. 셋째로, 구축한 학습 모델에 model agnostic 한 shapley additive explanations(SHAP)을 통해 정확한 최대 재공재고 추정뿐만 아니라 인프라 시스템 내 저하 시 문제가 되는 병목 자원을 파악할 수 있다.

본 논문은 다음과 같이 구성된다. 2 장에서는 본 연구의 전반적인 방법론에 대해 상세하게 설명하고, 3 장에서는 실험 설정 및 예측 모델의 성능을 비교 분석한다. 4 장에서는 예측 결과에 대한 원인 변수를 해석하며 마지막으로 5 장에서는 본 연구의 의미와 한계점 그리고 향후 연구 방향에 대해 논의한다.



(그림 1) 전체 실험 프레임워크

2. 제안 방법론

2.1 실험 아키텍처

그림 1은 본 논문에서 제안하는 전체 프레임워크이다. 첫째, 반도체 공장의 물류, 생산 센서 데이터 및 재공재고량 데이터를 수집하였다. 데이터마다 스케일이 달라 전처리 과정을 진행하였으며, 둘째, 정상성 검정을 위해 ADF test를 진행하였다. ADF test를 통해 학습 모델의 업데이트 여부를 결정하게 된다. 셋째, 분포 변화시에도 강건한 재공재고량 예측을 위해 온라인 학습 방식을 활용하였다. 넷째, 최적의 학습 모델을 선정하기 위하여 제안하는 프레임워크를 포함한 다양한 학습 모델의 실험 결과를 확인하였다. 마지막으로 해석 모델인 SHAP을 통해 예측 결과의 원인 변수를 분석하였다.

2.2 데이터 수집 및 전처리

이 연구는 2020년 1월부터 2021년 4월까지 반도체 공장에서 매시간 수집한 데이터를 사용하였다. 사용한 데이터는 비식별화 처리된 공장 내의 재공재고량과 생산 및 물류 센서 정보를 포함하고 있다. 학습에 사용된 데이터에서 예측 대상인 종속 변수(Y)는 최대 재공재고량으로 설정되었다. 따라서, 센서와 물류에 해당하는 293개의 데이터를 독립변수(X)로 하여 각 시점의 최대 재공재고량을 예측하는 모델을 구축하고자 하였다.

개별 종속 변수들 사이에 스케일 차이가 있어 학습에 영향을 미칠 수 있기 때문에 이를 고려하여 최소-최대 스케일 변환(min-max transformation)과 로그 변환(log transformation)의 전처리 과정을 진행하였다.

2.3 데이터 변이 감지

학습 모델을 구축하기 전 데이터 분포 변화 여부를 검정하기 위해 ADF test를 활용하였다. ADF test는 시계열 데이터의 정상성을 평가하는데 활용되는 가설검정 방법으로, 특정 window size에서 얻은 p-value 값을 이용한다. 본 연구에서는 최대 재공재고량의 유의수준과 p-value 값을 비교하여 해당 window size 내에서의 정상성 여부를 판단하였다.

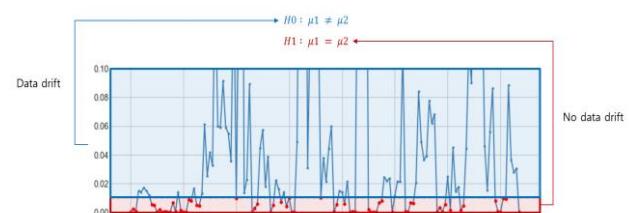
$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (1)$$

ADF 테스트는 자기 회귀 모델(autoregressive model)로 Dickey-Fuller test에 기반으로 한다. 수식 (1)에서 α 와 β 는 상수 값과 trend 계수를 의미하고, p 는 lag 차수를 나타낸다. ADF test의 귀무가설은 단위근이 존재함을 나타내며, 대립가설은 시계열 데이터가 정상성을 만족함을 주장한다.

$$H_0 : \mu_1 = \mu_2 \quad (2)$$

$$H_1 : \mu_1 \neq \mu_2 \quad (3)$$

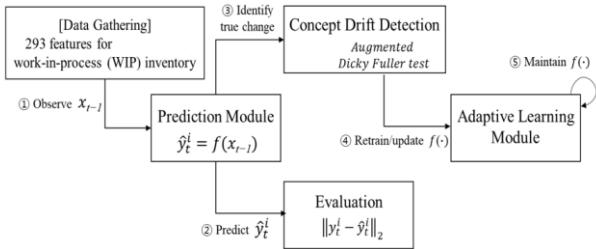
가설검정에서 p-value 값이 설정한 유의수준보다 크면 H_0 를 채택하여 비정상 패턴으로 판단한다. 반대의 경우에는 H_1 를 채택하여 정상 패턴으로 간주한다. 탐지된 비정상성 패턴이 존재하는 데이터의 경우는 모델 업데이트 시 사용된다.



(그림 2) ADF test 기반 데이터 변이 탐지

2.4. 모델 구축

본 연구에서는 ADF test를 통해 데이터의 변동이 존재할 경우에 학습 모델이 업데이트 되는 모델을 제안한다. 그림 3은 최대 재공재고량을 예측하는 점진적 학습 방식의 구조도를 나타낸다.



(그림 3) 제안 기법 구조도

3. 실험 결과

3.1 평가 척도

본 연구에서는 평균 제곱 오차(mean squared error; MSE), 평균 제곱근 오차(root mean squared error; RMSE), 평균 절대 오차(mean absolute error, MAE), 그리고 결정 계수(r-squared, R2)를 이용하여 사용된 학습 모델들의 회귀 예측 성능을 측정하였다. 이들은 모델의 정확도(accuracy)를 판단하는 지표로 활용되었다.

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

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

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (6)$$

$$R2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (7)$$

또한, 훈련 시간(m/s)과 업데이트 횟수를 비교 모델과의 효율성(efficiency) 측면에서 비교할 수 있는 지표로 활용하였다.

3.2 실험 세팅

제안 방법론의 성능을 입증하기 위해 총 세 가지 모델을 비교 실험하였다. 첫 번째로, 온라인 업데이트 모델은 최신 데이터셋에 대해 변이 감지 없이 매번 업데이트된다. 이 방법은 데이터가 실시간으로 입력되면서 즉시 학습이 가능하므로 작은 양의 데이터셋

으로도 점진적인 학습이 가능하지만, 노이즈에 대한 민감성이 문제가 될 수 있다. 두 번째로, 배치 학습 모델은 주어진 전체 데이터를 sliding window 만큼 옮겨가며 일정 크기의 배치 크기만큼의 데이터가 쌓이면 업데이트하는 방식이다. 본 실험에서는 배치 사이즈는 506(3 주), sliding window 는 72(3 일)로 설정하였다. 업데이트된 모델은 다음 시점의 최대 재공재고량 값을 예측한다. 이 방식은 데이터 변이 여부에 관계 없이 매번 모델을 업데이트 한다는 단점이 있다. 세 번째로, 본 연구에서 제안하는 기법은 ADF test를 활용하여 데이터의 정상성 여부를 판단하고 비정상 패턴일 경우에만 점진적으로 모델 업데이트를 수행한다. 배치학습과 달리 데이터의 변이가 있을 경우에만 업데이트 되기 때문에 높은 정확도와 함께 훈련시간 감소라는 이점이 있다.

비교 실험은 특정 시점 이전의 데이터에 대해 구축된 동일한 초기 학습 모델을 기반으로 진행되었다. Fair 한 비교 분석을 위해 세 가지 기법 모두 해당 초기 학습 모델을 기반으로 점진적 학습이 이루어진다.

3.3 실험 결과

표 1은 3 가지 모델을 정확도의 측면에서 비교 실험한 결과이다. 모든 예측 성능 평가 척도에서 다른 기법 대비 변이 감지를 기반으로 자가학습 하는 제안 방법이 가장 좋은 결과를 보인 것을 확인할 수 있다.

<표 1> 최대 재공재고량 예측 정확도 비교

	MSE	MAE	RMSE	R2
Online update	0.00005	0.0044	0.0074	0.39
Batch learning	0.00004	0.005	0.006	0.55
Proposed	<u>0.00004</u>	<u>0.0043</u>	<u>0.0065</u>	<u>0.7</u>

<표 2> 학습 모델 훈련 속도, 업데이트 수 비교

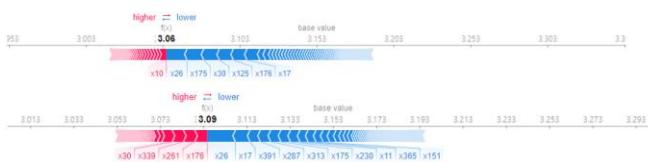
	Online update	Batch learning	Proposed
Training time(m/s)	568.543	80.361	<u>46.207</u>
Update counts	1113	145	<u>86</u>

또한 제안 기법의 정확도뿐만 아니라 효율성의 측면에서의 성능을 검증하고자 하였다. 표 2는 기법 별 학습 시간, 업데이트 횟수에 대한 실험 결과이다. 온라인 업데이트의 경우 매번 점진적인 학습을 진행하

기 때문에 568.844(m/s)로 가장 많은 시간이 소요되는 것을 확인할 수 있다. 제안 방법은 비정상 패턴의 경우에만 모델이 업데이트 되기 때문에 86 번의 업데이트만 진행되며 46.207(m/s)로 가장 짧은 학습시간이 소요됨을 확인할 수 있다. 따라서 제안 기법이 정확도 및 효율성 측면에서 모두 좋은 성능을 보이는 것을 확인할 수 있다.

4. 해석 결과

블랙박스 모델의 특성상 모델이 도출한 예측 결과를 명확하게 해석하기는 어렵다. Model-agnostic 한 해석기법 중 하나인 shapley additive explanations(SHAP)은 shapley value 를 활용하여 예측 결과에 대한 해석이 가능하다. 그림 4, 5 는 특정 관측치에 대한 shapley value 를 1 차원 평면으로 정렬하여 시각화한 것이다. Base value 는 학습 데이터셋에서 모델의 평균 예측값을 나타내며, output value 는 모델의 예측 값이다. 빨간색 부분은 예측 값을 증가시키는 변수들의 영향도를 나타내고, 반대로 파란색 부분은 예측 값을 감소시키는 변수들의 영향도를 나타낸다. 그래프 내 arrows 는 변수 영향도를 나타내며 크기가 클수록 많은 영향을 끼친 변수라고 해석할 수 있다.



(그림 4) 최대 재공재고량이 낮게 예측된 관측치의 force plot



(그림 5) 최대 재공재고량이 높게 예측된 관측치의 force plot

그림 4, 5 는 본 연구에서 사용된 데이터셋에 대해서 최대 재공재고량 값이 높게 예측된 샘플과 낮게 예측된 샘플에 대한 force plot 이다. 제안 기법을 통해 도출된 다음 시점의 예측 값에 대해 원인 변수를 추적할 수 있다. 예측 값이 낮은 경우가 최대 재공재고량이 낮은 시점에 해당하기 때문에 관심 있는 관측치이다. 따라서, base value(=0.153) 대비 shapley value 가 낮은 그림 4 의 경우, 학습 모델의 예측 값을 저하시

키는데 가장 크게 기여하는 변수는 두 관측치에서 공통적으로 feature 26 인 것을 확인할 수 있다.

5. 결론

본 연구에서 데이터의 변동이 있을 때 최소한의 업데이트로 최적의 예측 정확도를 유지할 수 있는 프레임워크를 제안한다. 데이터 변이를 감지하기 위해 모델을 구축하기 전에 시계열 데이터의 정상성을 평가하는 ADF test 를 활용하였다. 가설 검정을 통해 데이터 분포 변화를 탐지하고, 변화가 감지된 경우에만 모델이 지속적으로 업데이트 된다. 데이터의 변이 여부와 관계없이 업데이트를 진행하는 예측 모델과의 비교를 통해 정확도뿐만 아니라 훈련 속도와 업데이트 수의 효율성 면에서 성능을 입증하였다.

최대 재공재고량을 정확하게 예측하는 것도 중요하지만 낮게 예측된 경우 원인 변수를 역추적 하는 것이 중요하다. 블랙박스에 해당하는 딥러닝 모델의 특성을 고려하여 model-agnostic 방식의 SHAP 모델을 적용하였다. 가장 좋은 성능을 보였던 제안 기법에 대해 SHAP 을 활용하여 원인 변수를 해석하고자 하였다. Shapely value 가 낮은 관측치에 대하여 학습 모델의 예측 값을 저하시키는데 영향을 미친 변수를 확인할 수 있다. 제안하는 프레임워크는 미래 시점에 대한 정확한 재공재공 예측뿐만 아니라 인프라 시스템 내 병목 자원을 조기 탐지할 수 있다.

하지만, 성능 평가 시에는 보다 추가적인 비교 모델을 구축하여 실험 결과에 대한 검증이 필요하다. 특히 종속 변수에 대한 변이뿐만 아니라 다양한 센서 데이터에 대한 변이 감지가 고려되어야 한다. 이러한 한계점을 극복하기 위해 종속 변수에 대한 데이터 분포 감지를 통해 보다 강건한 학습 모델을 구축하는 확장 연구를 수행할 계획이다.

참고 문헌

- [1] Lee, S., Kim, H. J., & Kim, S. B. (2020). Dynamic dispatching system using a deep denoising autoencoder for semiconductor manufacturing. *Applied Soft Computing*, 86, 105904.
- [2] Liu, Y., Wu, H., Wang, J., & Long, M. (2022). Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in Neural Information Processing Systems*, 35, 9881-9893.
- [3] Fowler, C., Cai, X., Baker, J. T., Onnela, J. P., & Valeri, L. (2022). Testing unit root non-stationarity in the presence of missing data in univariate time series of mobile health studies. *arXiv preprint arXiv:2210.04998*.