

SHAP 기반 Subpopulation 분석과 머신러닝을 활용한 OLED 증착 공정 PPA 편차 원인 인자 도출 방법론

양상식, 서동주
삼성디스플레이

sangsik.yang@samsung.com, dongjoo.seo@samsung.com

요 약

본 연구는 OLED 증착 공정에서 발생하는 PPA(Pixel Position Accuracy) 편차의 원인 인자를 체계적으로 도출하기 위해 XGBoost 와 LightGBM 기반 예측 모델을 구축하고 SHAP 을 활용한 분석 방법론을 제안하였다. 증착 공정의 FMM(Fine Metal Mask) ID, 온도, Tact, Glass, 증착기 인자 등 정형화된 테이블 데이터를 활용해 불량률 예측 모델을 학습시킨 후, SHAP Feature Importance 를 기반으로 Subpopulation 분석 및 공정 인자 그룹화를 수행하여 불량 원인을 도출하였다. 실제 양산 데이터를 활용한 결과, 상위 4% 순위 내에서 정확한 원인 인자 식별과 그룹별 평균 중요도에서 유의미한 증가를 확인함으로써 제안 방법론의 효과성을 검증하였다. 이는 엔지니어의 경험 의존적 분석에서 벗어나 데이터 기반의 신속하고 정확한 불량 원인 규명을 가능하게 한다.

I. 서 론

본 연구는 OLED 증착 공정에서 발생하는 픽셀 위치 정확도(PPA) 편차의 주요 원인을 도출하기 위해 머신러닝 기반 분석 방법론을 제안하였다. OLED 제조 공정은 R/G/B 서브픽셀별로 정밀한 증착이 요구되며, FMM 형상, 온도, 증착 Tact Time 등 복합적인 설비 상태 변수가 불량 발생에 영향을 미친다. 기존에는 엔지니어의 경험에 의존하는 분석 방식으로 인해 분석 속도와 정확도가 개인 역량에 크게 좌우되는 한계가 존재하였으며, 이를 개선하기 위해 본 연구는 트리 기반 모델과 SHAP 을 결합한 접근법을 적용하여 데이터 기반의 체계적인 원인 분석을 수행하였다.

SHAP(SHapley Additive exPlanations)은 게임 이론의 샤플리 값(Shapley Value)을 기반으로 하여 머신러닝 모델의 각 입력 특성이 예측 결과에 기여한 개별 영향도를 수량화하는 모델 무관한 해석 기법이다. 주어진 예측 $f(x)$ 에 대해 각 특성 i 의 SHAP 값 ϕ_i 는 다음과 같이 정의된다.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S))$$

N 은 전체 특성 집합, S 는 특성 부분집합, f_S 는 S 에 속한 특성만을 사용한 모델의 기여도를 나타낸다. SHAP은 샤플리 값의 공리인 효율성, 대칭성, 영향력 부재, 그리고 선형성을 만족하며, 각 특성의 기여를 더하여 기준값과의 차이로 예측값을 완전히 분해할 수 있는 장점이 있다. SHAP은 모델 예측 결과의 특성 중요도를 수치화하여 해석 가능성을 확보하는 데 효과적이며, 제조업 및 반도체 공정에서 불량 예측 및 설비 진단에 널리 활용되어 왔다[1,2]. 특히 Causal SHAP은 인과 관계를 반영한 보다 정확한 설명을 목표로 하나, 정확한 인과 그래프 구축의 어려움, 계산 복잡도 증가, 동적 변수 처리 한계 등의 문제로 산업 현장 적용에 제약이 있어[3], 본 연구는 이러한 한계를 고려하여 SHAP 기반 분석에 초점을 맞추며, 실제 증착 데이터를 활용해 제안 방법론의 효과성을 검증함으로써 데이터 기반 공정 최적화에 기여하고자 한다.

II. 본론

증착 공정에서 생성되는 데이터의 수집, 전처리부터 모델학습 및 불량 원인 분석을 Figure 1의 Flow를 따라

진행하였다. 실시간 데이터 수집을 위해 Dataiku 시스템을 활용하여 최근 1개월간의 증착 공정 내(In-Fab) 데이터를 원인 인자로, PPA 편차에 의한 불량률을 결과 인자로 수집하였다. 이후 XGBoost 및 LightGBM 모델을 적용해 불량률 예측 모델을 학습하고, 학습된 모델에 SHAP Tree Explainer를 적용하여 SHAP Feature Importance를 계산하였다. 불량률의 통계적으로 유의미한 차이를 기준으로 증착주기 및 FMM 교체주기 기반 Subpopulation을 'REFERENCE' 그룹과 'COMPARE' 그룹으로 분류한 후, 원인 인자를 FMM ID, 온도, Tact Time, Glass, 증착기 인자로 구분하고, 각 인자별 평균 SHAP Feature Importance를 산출하였다.

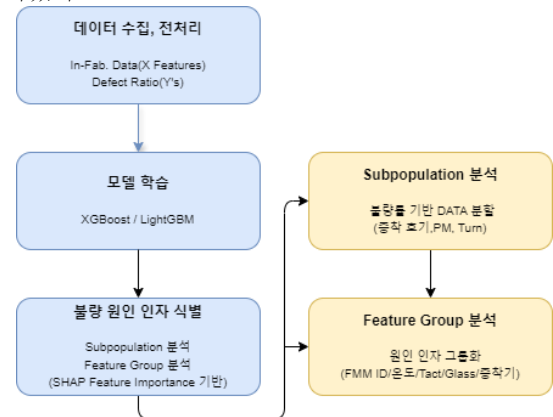


Figure 1. 불량 분석 Flow.

학습에 적용된 학습 인자는 Table 1과 같다. 분석 방법론의 강건성을 확인하기 위하여 XGBoost와 LightGBM을 적용하였으며, Tree Size와 정규화 강도를 Split하여 모델의 Capacity가 원인계 예측에 미치는 영향도 변화를 검증하였다. 그 외 인자는 모델 기본값을 적용하였다.

학습에 적용한 데이터는 실 양산 데이터를 적용하였으며 Subpopulation은 특정 FMM이 적용된 시점을 기준으로 설정하였다. Table 2에 학습된 DATA 및 REFERENCE/COMPARE Subpopulation 및 원인계별 인자 수를 정리하였다.

XGBoost			
Parameter	reference	weak_reg.	strong_reg.
lambda (L2)	1	0.1	10
max_depth	6	8	4
LightGBM			
Parameter	reference	weak_reg.	strong_reg.
lambda_l2 (L2)	0	0	10
max_depth	6	8	4
num_leaves	31	90	10

Table 1. 모델 학습 Parameter
학습 DATA

Number of features	학습 DATA Rows
123	3106
Subpopulation	
REFERENCE	COMPARE
171	149
인자 Group	
FMM ID	19
Tact	2
온도	35
증착기	22
모델	12
Glass	33

Table 2. 학습 DATA

Figure 2 와 3 에 각각 XGBoost/Strong Reg. 모델의 그룹별 평균, 상위 10 개 인자의 Feature Importance 를 나타내었다. 실제 불량은 COMPARE 구간의 Tact 변동 및 특정 ID 의 FMM 에 의하여 발생하였으며, Figure 2 에서 Tact 그룹의 Feature Importance 가 타 그룹 대비하여 큰 Importance 를 얻었으며, REFERENCE 대비하여 COMPARE 에서 두배 수준의 증가를 보여준다. 상위 중요도 기준 상위 10 개 인자 확인시 불량 발생을 유발한 FMM ID 인자와 Tact 인자가 1, 4 위로 나타남을 확인할 수 있었다. 이것은 해당 인자의 수준이 높은, 특정 ID FMM 이 반영되었거나 Glass to Glass 시간이 증가된 경우 불량률이 높은 현상을 반영한 결과이다.

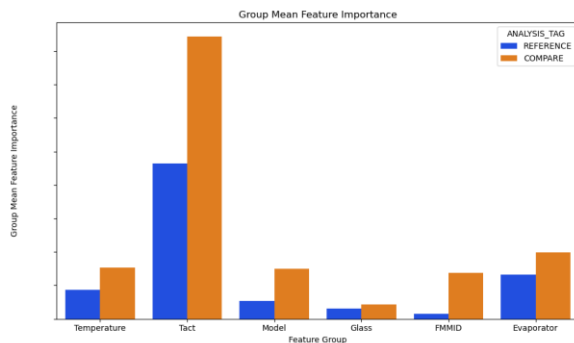


Figure 2. 그룹별 평균 Feature Importance

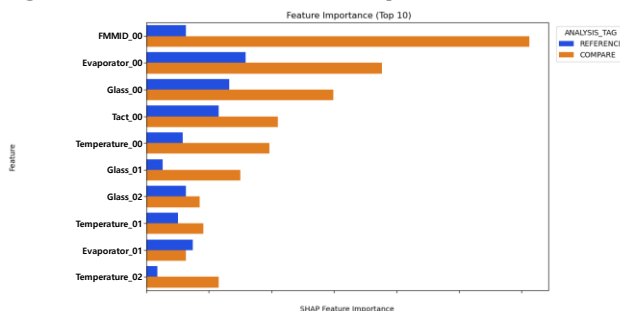


Figure 3. Top 10 Feature Importance

제안된 모델링 및 분석 방법론이 실 제품생산에서 적용될 경우 데이터의 Capacity 가 변화하는 상황에서도

안정적인 예측성능을 확보하는 것이 중요하며, 이를 검증하기 위하여 모델 Type/Capacity 에 따른 불량 인자 예측성능을 Table 3 에 나타내었다. XGBoost 의 경우 모델 Capacity 변화에 둔감한 성능을 보여주었으며, LightGBM 의 경우 XGBoost 대비하여 검출 성능이 떨어지는 사례가 발생하였다. 이는 LightGBM 이 특성간 상호작용을 학습하는 과정에서 노이즈를 과적합하여 나타났다고 추정되며, XGBoost 의 기본 parameter 가 보수적이고 안정적인 성능을 목표로 설계된 점에서 기인한다고 판단할 수 있다[4,5].

모델	조건	FMM ID 인자 순위	Tact 인자 순위
XGBoost	reference	1	5
	strong reg.	1	4
	weak reg.	1	4
LightGBM	reference	6	7
	strong reg.	1	5
	weak reg.	1	5

Table 3. 불량 원인 인자 순위

III. 결론

본 연구는 OLED 증착 공정의 픽셀 위치 정확도(PPA) 편차 원인을 규명하기 위해 서브모집단 기반 피쳐 중요도 분석을 수행하였다. 트리 기반 모델(XGBoost, LightGBM)을 활용해 전체 데이터와 그룹별 특성 중요도를 비교한 결과, XGBoost 의 경우 상위 4%(Top 5 이내) 중요 변수 내에서 불량 원인을 효과적으로 식별했으며, 불량 그룹에서 해당 변수의 중요도가 유의미하게 증가함을 확인하였다. 이는 기존 경험적 분석 대비 데이터 기반의 객관적 근거를 제공하며, SHAP 을 통한 해석 가능성 확보로 제조 공정 최적화에 기여할 수 있음을 보여준다.

참 고 문 헌

- [1] Lundberg et al. "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems (NeurIPS), 30, 2017.
- [2] Lee et al. "SHAP Value-Based Feature Importance Analysis for Short-Term Load Forecasting," J. Electr. Eng. Technol. 18, pp579- 588 2023.
- [3] Heskes et al. "Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models," NeurIPS, 33, 2020.
- [4] Chen et al. "XGBoost: A Scalable Tree Boosting System," KDD, 2016.
- [5] Guolin Ke et al. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," Advances in Neural Information Processing Systems (NeurIPS), 30, 2017.