이상탐지서비스 알고리즘 소개 및 기업 적용 사례

이재훈 연구위원
LG전자 AI빅데이터담당
jae.h.lee@lge.com

2023년 대한산업공학회/한국경영과학회 춘계공동학술대회

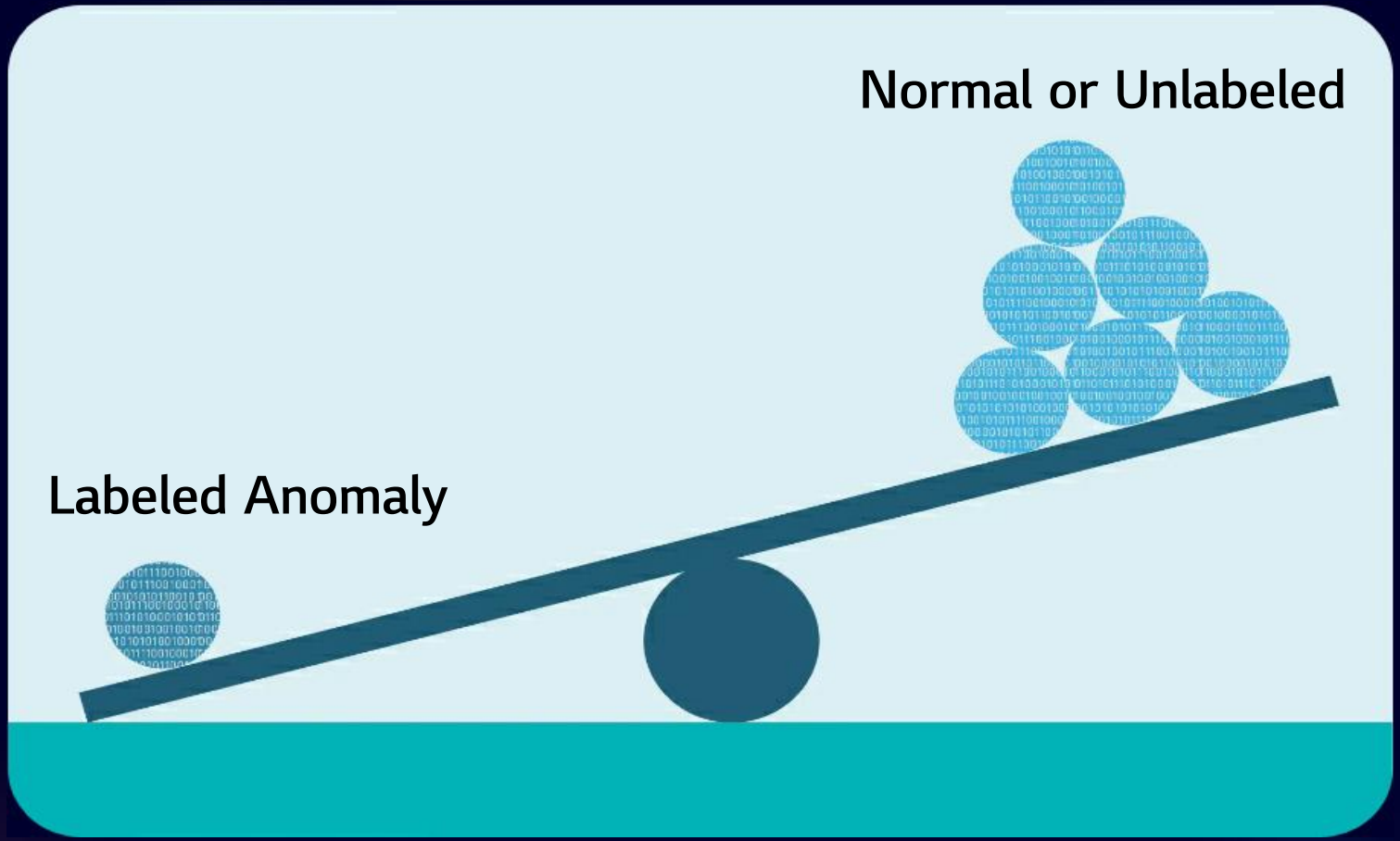
Time Series Anomaly Detection

✓ 서비스가 확산될수록 여러 지표를 동시에 모니터링할 필요성 증대

✓ 제조, 생산, SCM, 고객, 제품, 환경 등 수많은 영역에서 이상감지 활용

- 공장 설비 예지 보전
- 네트워크 침입 감지
- 부품 수급량 이상 감지
- VOC 이슈 비윤 이상 감지
- 고객만족도지수 이상 감지
- 환경(온습도 등) 이상 감지

Challenge 1. Lack of Labels



Challenge 1. Lack of Labels

1 Unsupervised Learning

- 모든 데이터를 OK로 간주
- 사용 데이터 수 증가
- 이상 판정 기준 모호

2 Semi-supervised Learning

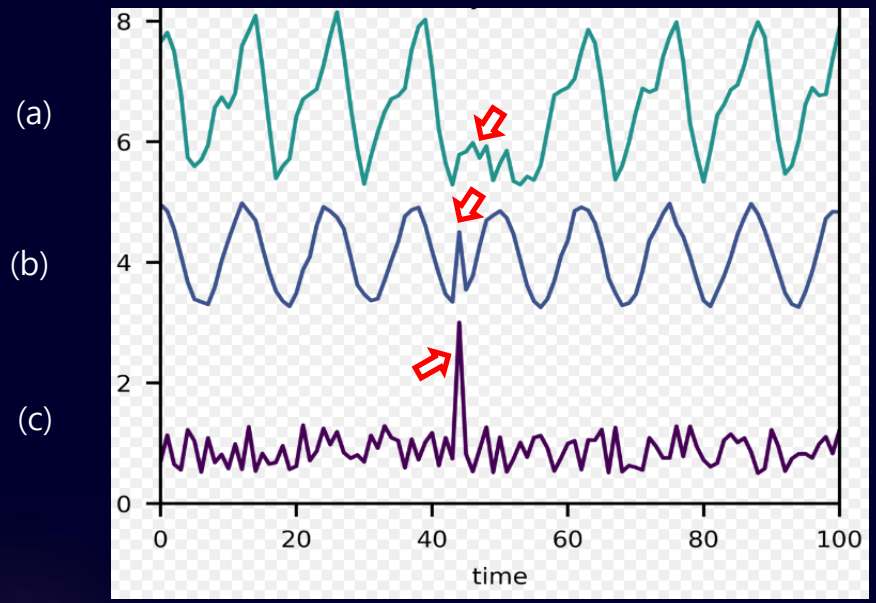
- 자기 자신의 값을 정답으로 간주
- AutoEncoder, USAD 등

3 Supervised Learning

- 오버샘플링
- 다운샘플링
- Transfer Learning

Challenge 2. Diversity of Anomaly Types

✓ 도메인 특성과 필요에 따라 다양한 형태의 Anomaly 존재



Contextual Collective Anomaly

Contextual Point Anomaly

”Out-of-range” Point Anomaly

Multivariate Anomaly?

Challenge 2. Diversity of Anomaly Types

Point Anomaly Detection

- Local Outlier Factor (LOF)
- Dynamic Threshold
- Isolation Forest
- K-Nearest Neighbors (KNN)
- Robust Random Cut Forest
- One-Class SVM
- ...

Contextual Anomaly Detection

- Robust Random Cut Forest
- Spectral Residuals
- Matrix Profile
- Hierarchical Temporal Memory
- DeepAnT
- LSTM AutoEncoder
- ...

Multivariate Anomaly Detection

- AutoEncoder
- LSTM AutoEncoder
- USAD
- DBSCAN
- One-class SVM
- Robust Random Cut Forest
- TadGAN
- ...

Challenge 3. Automation & Optimization

Best Model Selection

- 사용자의 필요와 Anomaly의 유형에 따라 Candidate 모델을 학습한 후 최적 모델 선택

1

Anomaly Type Detection

- Time Series의 특성과 Anomaly의 유형 감지
- 사용자의 필요가 고려되어야 함

2

3

Hyper-parameter Optimization

- 선택된 모델이 Anomaly를 잘 탐지하도록 Hyper-parameter 최적화

Challenge 3. Automation & Optimization

Best Model Selection

- 최적 모델 선택 어떻게 할 것인가?
- 모델의 성능을 어떻게 정의할 것인가?

1

Anomaly Type Detection

- 사용자 필요에 따른 Anomaly 정의가 선행되어야 해서 완전 자동화 어려움

2

3

Hyper-parameter Optimization

- 어떻게 최적화할 것인가?
- 모델의 성능을 어떻게 정의할 것인가?

Challenge 4. Efficiency

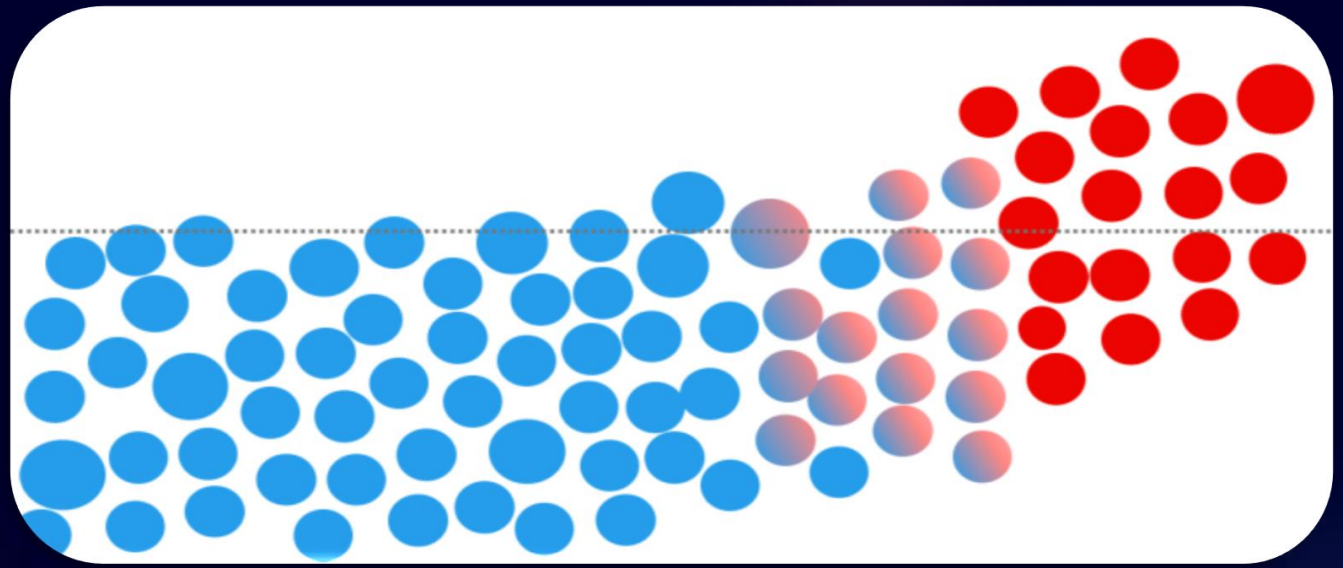
✓ 실시간으로 많은 시계열 변수를 동시 관리해야 함

✓ 자동화된 Workflow 구축 필요

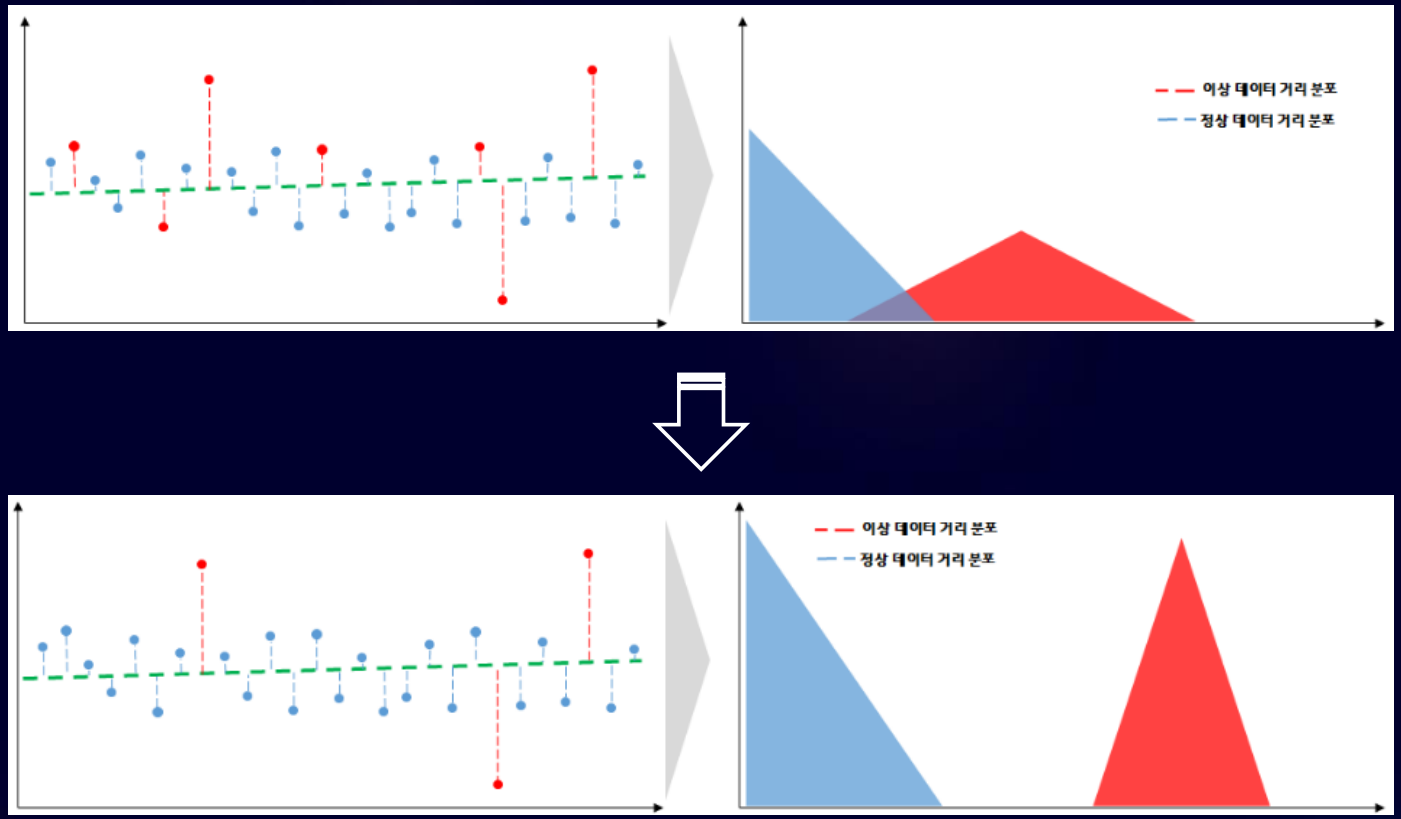
✓ 실시간 학습 및 추론 필요

Challenge 5. Concept Drift

- ✓ 시간 또는 환경의 변화에 따라 데이터의 분포가 바뀜
  - ✓ 연속 학습 (Continual Learning) 필요
  - ✓ 또는 조건부 재학습 필요



Object Function Based Optimization



Object Function 정의

- ✓ How to select best model?
- ✓ How to optimize hyper-parameters?
- ✓ Anomaly예측값과 Normal예측값과의 거리가 클수록 Object Function 증가

$$C(Y_A, Y_N, N_P) = D(Y_A, Y_N) - C_P(N_P)$$

Y_A : Anomaly로 예측한 데이터

Y_N : Normal로 예측한 데이터

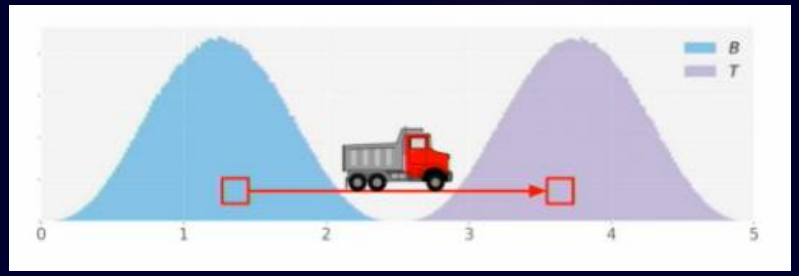
N_P : 특정 거리 조건을 만족하는 Anomaly 데이터 수

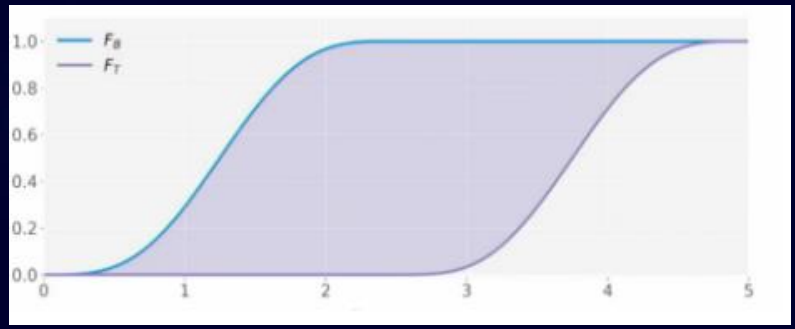
C : Distance Cost 함수

C_P : Penalty Cost 함수

Earth Mover Distance

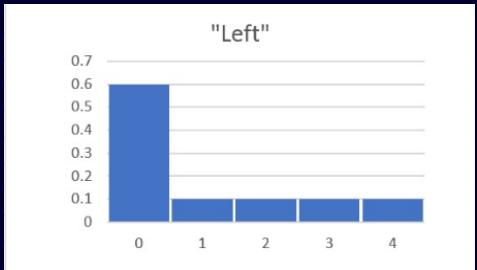
- ✓ 두 분포의 차이를 나타내는 거리 함수
- ✓ 이 산을 저 산으로 옮기는 데 필요한 일의 양을 거리로 계산

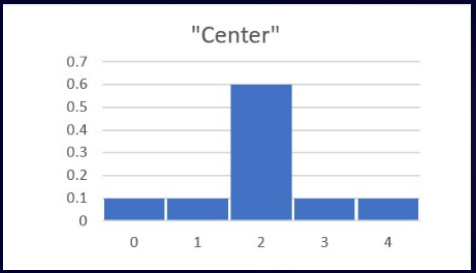


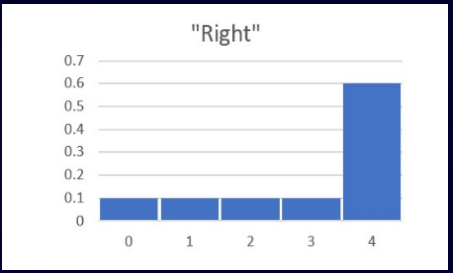


Earth Mover Distance

✓ 두 분포 간의 겹치는 영역이 작은 경우 Earth Mover Distance가 합리적

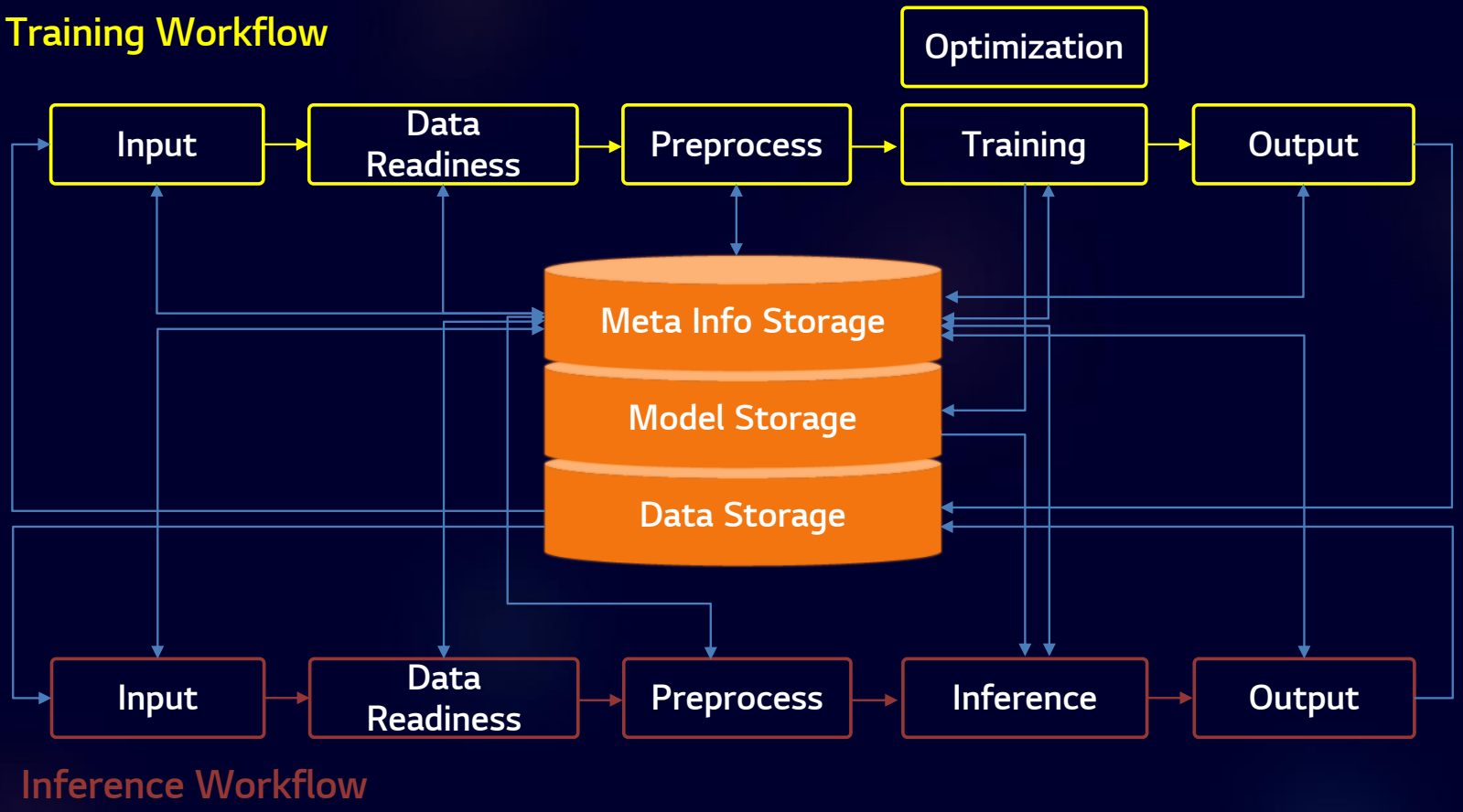






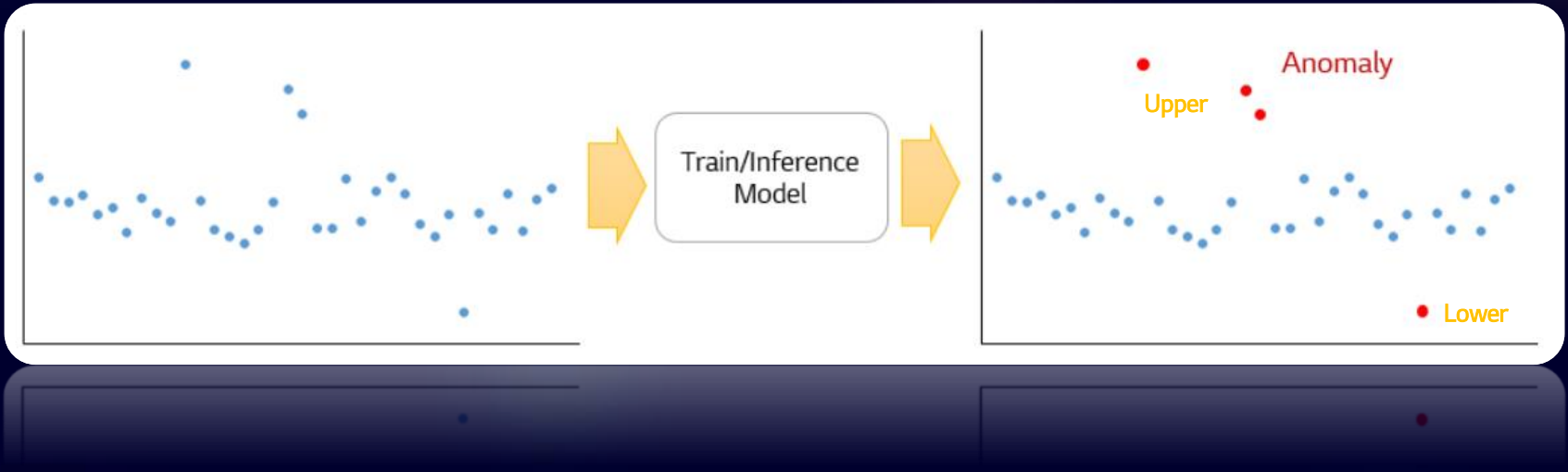
Distribution Comparison	Kullback-Leibler	Jensen-Shannon	Earth Mover
Left vs Center	1.79	0.45	1.0
Center vs Right	1.79	0.45	1.0
Left vs Right	1.79	0.45	√ 2.0

Modeling Workflows For Stream Data



Our Point Anomaly Detection Algorithm

- ✓ 실시간 스트리밍 point 데이터가 정상/ 이상인지 탐지하는 알고리즘
- ✓ Univariate Time Series Data을 대상으로 함



Our Point Anomaly Detection Algorithm

Automatic HPO

- Anomaly threshold 포함하여 모든 파라미터 자동화
- Bayesian -based 하이퍼파라미터 최적화

Diverse Candidate Models

- 통계적 방법론: Exponential Smoothing, Dynamic Threshold
- 머신러닝 방법론: Spectral Residual, Robust Random Cut Forest

Best Model selection

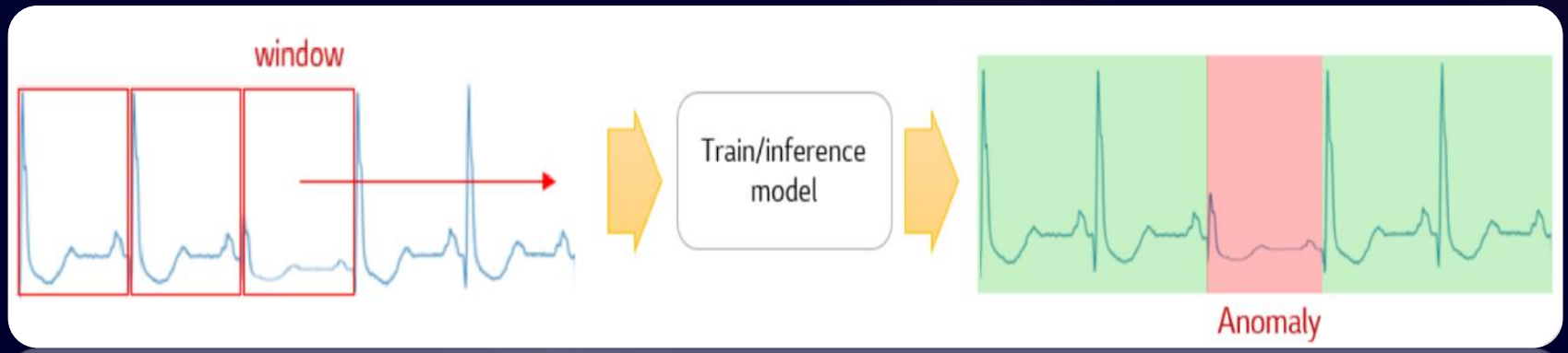
- 자체 목적함수 (Object function)기반으로 최적 모델을 자동으로 선택
- 수동으로 candidate 모델 제한 가능

Customizable detection

- Anomaly 방향 설정 가능 (Upper / Lower / Both)
- 특정 파라미터 수동 설정 가능

Our Contextual Anomaly Detection Algorithm

- ✓ 실시간 스트리밍 데이터가 구간별 정상/이상인지 판별
- ✓ Time Series 정상 패턴을 학습하고 이에서 벗어난 이상 패턴을 탐지



Our Contextual Anomaly Detection Algorithm

Automatic HPO

- 모든 파라미터 자동화
- Anomaly Score 분포 활용
- Bayesian -based 하이퍼파라미터 최적화

Continual Learning

- 초기 Training 이후 점진적으로 연속학습
- Concept Drift 대응 가능

Tumbling / Sliding window

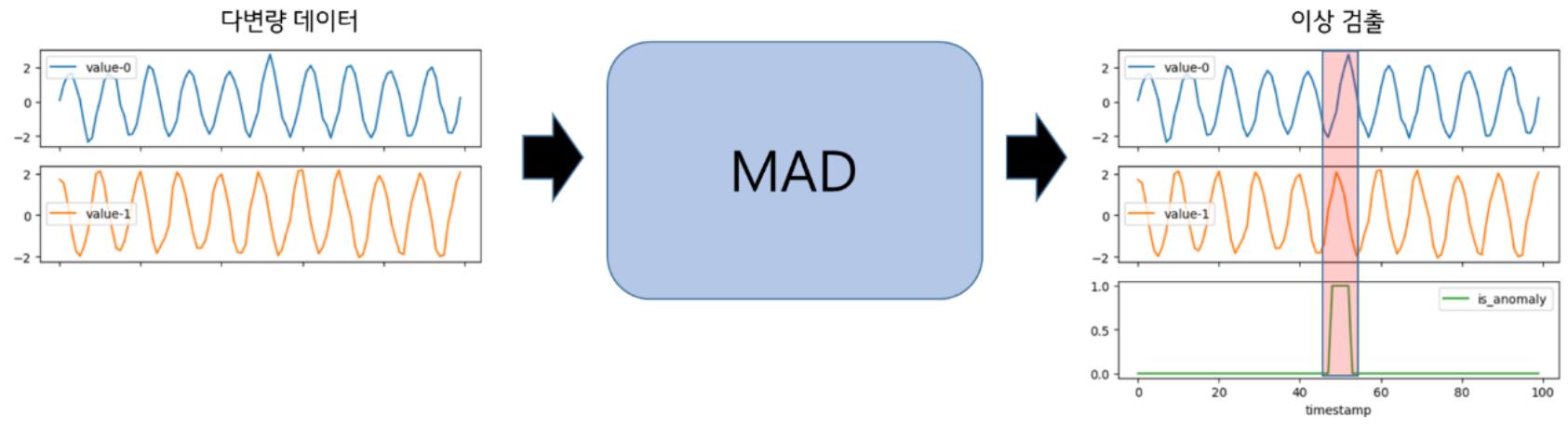
- Tumbling Window: Overlapping없이 이상 패턴 구간 탐지 가능
- Sliding Window: 한 point씩 구간 Sliding하여 이상 탐지 가능

Adaptive Threshold / Window

- Anomaly Score 분포에 기반한 최적의 threshold 제공
- 효율적인 정상 분포 학습을 위해 매번 모델 업데이트마다 Window 조정

Our Multivariate Anomaly Detection Algorithm

- ✓ 여러 변수를 모니터링하면서 이상 또는 불량을 조기 검출
- ✓ Time Series 정상 패턴을 학습하고 이에서 벗어난 이상 패턴을 탐지



Our Multivariate Anomaly Detection Algorithm

Automatic HPO

- 모든 파라미터 자동화
- Semi-supervised 학습 기반 최적화
- Bayesian -based 하이퍼파라미터 최적화

LSTM Autoencoder

- Long-term Dependency 시계열 특성 활용
- 조기 이상진단 (Early Sensing) 가능

Neural Architecture Optimization

- 데이터에 따라 최적화된 네트워크 구조를 통해 Overfitting / Underfitting 방지
- X인자 갯수 및 Window size 고려

Adaptive Threshold / Window

- Look Back Window Size 자동 조정
- Anomaly Score기반의 자동 Threshold 설정

이상진단 서비스 적용 사례

✓ Oil Pikup 유량 검사 공정

1 **모니터링**
- 이상발생률 평균 0.02%

2 **이상 감지**
- 이상발생률 5배 증가
- 금형 개선조치

3 **조치 결과**
- 이상발생률 평균 0.03%

이상진단 서비스 적용 사례

✓ Harness 압입 공정 개선

1 **모니터링**

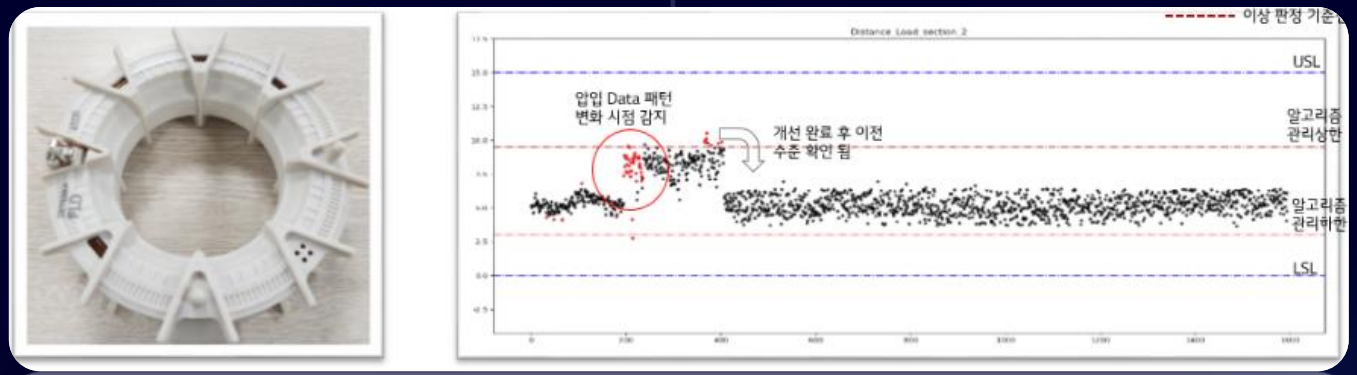
- 압입력 데이터 모니터링

2 **이상 감지**

- 규격 선 내 이상 감지 (압입력 증가)
- 신규 금형 제작

3 **조치 결과**

- 압입력 수준 회복



AWS MarketPlace 출시