이상치 탐지 성능 향상을 위한 MI-FGSM 데이터 증강 기법

장현석[1], 정민성[1], 이충호[2], 허태욱[2] *이상금[1]

*국립한밭대학교[1], 한국전자통신연구원[2]

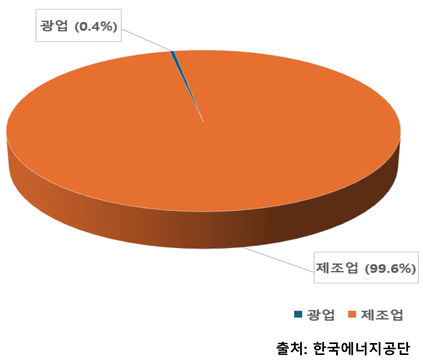
논문 바로가기

EcoAI Lab

Abstract

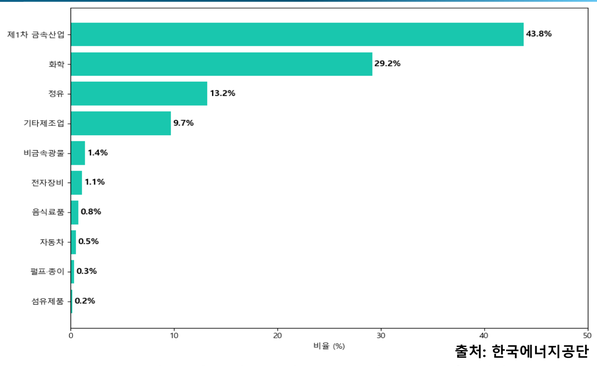
- (기후문제 및 이상치 탐지 필요성) 화공 산업은 국내 온실가스 배출의 상당 부분을 차지하며, 효과적인 온실가스 배출량 감축과 기후변화 대응을 위해 신뢰성 높은 전력 데이터 분석과 이상치 탐지가 필수적임.
- (클래스 불균형) 데이터 분석 결과, 정상 데이터 대비 이상치가 적은 클래스 불균형이 존재하였고 이는 탐지 결과의 신뢰성을 저해함.
- (데이터 증강 적용) 학습 데이터를 MI-FGSM에 적용해 생성된 적대적 샘플을 통해 이상치를 충당하고 두 클래스간 균형을 조절함.
- (성능 분석) MI-FGSM 파라미터인 최대 교란 크기 ϵ과 모델의 탐지 성능 지표인 F1 점수간 상관관계를 분석함.
- (ϵ 임계값 도출) ϵ과 F1 점수의 상관관계 그래프를 통해 F1 점수의 성능이 낮아지는 ϵ의 임계값을 도출하고 데이터 증강을 위한 MI-FGSM의 ϵ 허용 범위에 대한 가이드라인을 제시함.

Introduction & Research Background



출처: 한국에너지공단

그림1. 산업부문 내 업종별 온실가스 배출 비율



출처: 한국에너지공단

그림2. 제조업 내 업종별 온실가스 배출 비율

– (그림1) 산업 부문 내 제조업 99.6%, 광업 0.4%로 제조업이 매우 큰 비중을 점유함.

– (그림2) 제조업 중 금속, 화학, 정유, 기타 제조업이 총 95.9%를 차지하며, 그 중 화학 산업이 29.2%를 차지함.

LSTM-AE & MI-FGSM

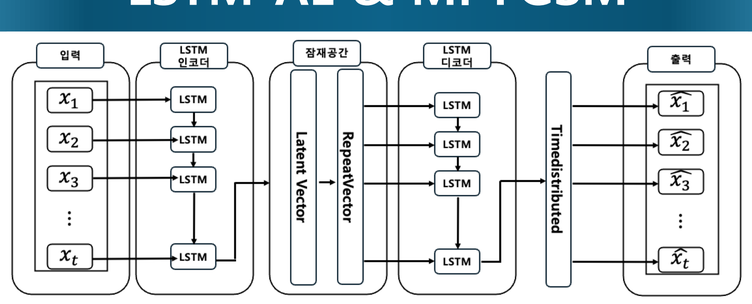


그림3. LSTM-AE 모델 구조

- (그림3) LSTM(Long Short – Term Memory)과 AE(AutoEncoder)의 결합 모델인 LSTM-AE 구조도이며, 각 층의 역할은 다음과 같음.
  - (LSTM 인코더/디코더) 시계열 데이터의 시간 의존성을 반영하여 데이터 포인트를 순차적으로 처리함.
  - (잠재공간) 인코더에서 데이터의 핵심 정보를 벡터 형태로 변환하고, RepeatVector에서 디코더의 입력 형태로 변환함.
  - (

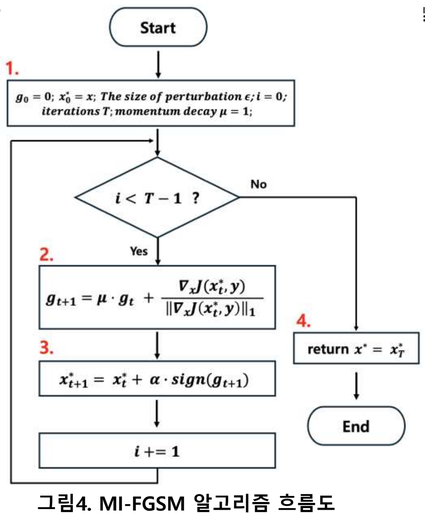


그림4. MI-FGSM 알고리즘 흐름도

력 데이터 형태와 동일하게 변환함.

- (1단계) 모멘텀 벡터 g_0와 적대적 샘플 x_0^*, 교란 크기 ϵ, 반복 횟수 T, 모멘텀 계수 μ, 카운터 i 선언함.
- (2단계) 시점 t의 모멘텀에 μ를 곱한 값과 비용 함수 J의 적대적 샘플 x_t^*와 정답 레이블 y에 대한 입력 데이터 x의 그래디언트 방향 성분을 더해 $t+1$ 시점의 모멘텀 g_{t+1}을 구함.
- (3단계) 시점 t의 적대적 샘플 x_t^*에 g_{t+1}의 방향으로 α만큼 이동한 값을 더해 $t+1$ 시점의 적대적 샘플 x_{t+1}^*을 생성함.
- 2,3 단계를 카운터의 조건 만족 시까지 반복함
- (4단계) T번의 업데이트 과정으로 갱신된 적대적 샘플 x_T^*을 반환.

실험 및 결과

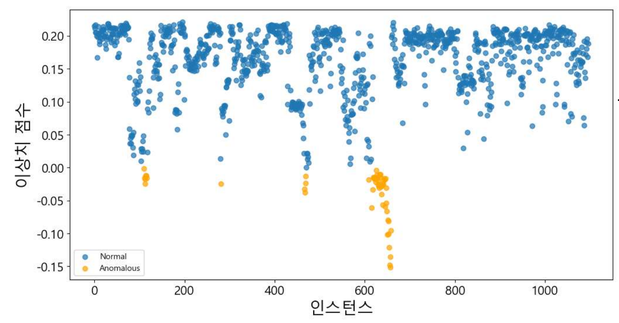


그림5. Isolation Forest 이상치 분류 결과

- (데이터셋)(그림5) 2020년 1월부터 2022년 12월까지 15분 단위로 수집된 전력 데이터 중 점유율이 높은 기업의 데이터를 사용함. Isolation Forest 분류 결과, 전체 1096개의 데이터 중 55개의 이상치가 존재함.

– (전처리)데이터를 [0, 1]로 정규화하고, 이상치 증강을 위해 학습 데이터를 MI-FGSM에 적용해 적대적 샘플을 생성함. 이를 활용해 테스트셋의 클래스 불균형을 완화함. 이때 ϵ은 재구성 오차의 임계값과 탐지 성능에 영향을 미쳐, 정규화된 데이터 스케일에 따라 ϵ을 미세한 범위인 [0.5, 0.25, 0.1, 0.05, 0.01]에서 조정하며 탐지 결과를 비교함.

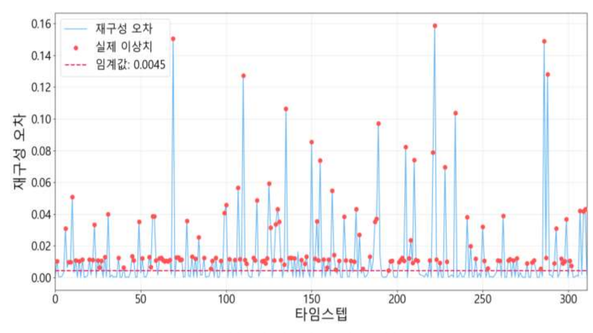


그림6. 이상치 탐지 재구성 오차 분포(ϵ = 0.5)

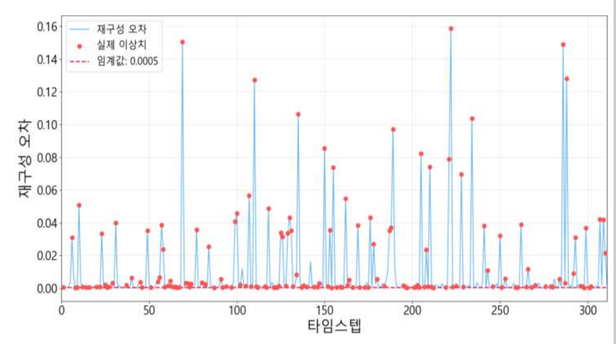


그림7. 이상치 탐지 재구성 오차 분포(ϵ = 0.01)

– (그림6) ϵ = 0.5일 때 MI-FGSM으로 생성된 적대적 샘플은 대부분 재구성 오차의 임계값보다 큼. 기존 이상치 대비 정상 데이터에 근사한 분포를 보이며, F1 점수는 0.9778로 높은 성능이 나옴.

– (그림7) ϵ = 0.01일 때의 재구성 오차는 그림 6와 달리, 적대적 샘플의 분포가 임계값과 현저하게 높은 유사성을 보임. 또한, F1 점수는 0.6907로 ϵ = 0.5일 때 대비 0.2871이 감소함.

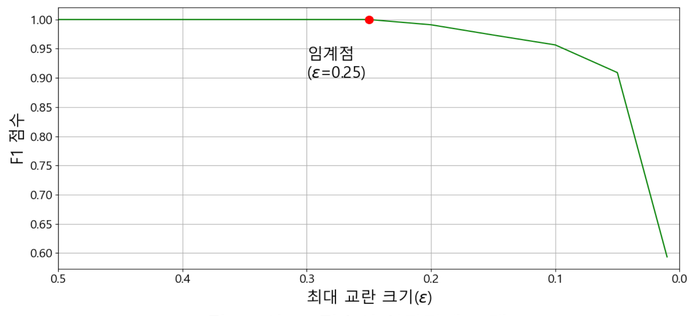


그림8. ϵ와 F1 점수의 상관관계 그래프

– (그림8) 빨간색 점에 해당하는 ϵ = 0.25 지점부터 ϵ 값이 작아짐에 따라 F1 점수가 낮아지며, 0에 수렴할수록 급격하게 감소함. ϵ이 크면 노이즈의 교란 범위가 넓어져 이상치 탐지가 용이하지만, ϵ이 작아질수록 교란 범위가 좁아져 정상 데이터와의 구분 경계가 모호해짐. 따라서, F1 점수가 낮아지는 시작점인 ϵ = 0.25를 실험 데이터의 임계값으로 선정함.

결론

– (연구 목표) 화공 산업 전력 데이터의 클래스 불균형을 해소하고, LSTM-AE의 이상치 탐지 성능에 대한 신뢰성을 향상하고자 MI-FGSM 기반 데이터 증강 방법론을 제안함.

– (실험 과정) 화공 산업 내 점유율이 높은 기업 데이터를 전처리 후, ϵ = [0.5, 0.25, 0.1, 0.05, 0.01] 범위에서 F1점수를 비교함.

– (실험 결과) ϵ = 0.5에서 0.9778의 F1 점수를 달성한 반면, ϵ이 감소됨에 따라 F1 점수가 낮아지는 경향을 보였고, ϵ = 0.01에서는 F1 점수는 0.6907임.

– (시사점) ϵ과 F1 점수의 상관관계를 분석한 결과, ϵ이 임계값보다 낮아 교란 범위가 과도하게 좁을 경우 적대적 샘플과 정상 데이터간 구분이 불분명하여 탐지 성능이 저하됨.

향후 연구

– (일반화 검증) 다양한 데이터셋에서 MI-FGSM 기반 데이터 증강을 적용해 일반화를 검토할 필요가 있음.

– (적대적 학습 적용) 적대적 학습과 이상치 탐지 과정을 하나의 파이프라인으로 통합하는 End-to-End 구조로 탐지 성능을 최적화하고자 함.

– (산업체별 특성 기반 최적화) 다양한 산업체의 고유한 특성에 적합하도록 모델을 조정·적용하여, 산업체의 적용가능성 및 운영 효율 향상에 기여함.