

제지공정 시계열 데이터의 이상 원인 자동 해석을 위한 LLM 기반 설명 생성 모델 연구

이서영, 장여진, 길기훈, 백민우, 유지오, 이상금*

*국립한밭대학교

{syoun2353, jyeoj251, minegihun, bmw5779, uzo7383}@gmail.com, *sangkeum@hanbat.ac.kr

An LLM-Based Explanation Generation Model for Automatic Interpretation of Anomaly Causes in Papermaking Process Time-Series Data

Seoyoung Lee, Yeojin Jang, Gihun Gil, Minu Baek, Jio Yoo, Sangkeum Lee*

*Hanbat National Univ.

요약

제지공정은 다수의 공정 변수가 상호작용하는 복잡한 시계열 시스템으로, 이상 구간 발생 시 원인을 직관적으로 해석하는 데 어려움이 있다. 기존의 설명 가능 인공지능(XAI) 기법은 예측 결과에 대한 변수 기여도를 정량적으로 제공할 수 있으나, 수치 중심의 결과로 인해 공정 맥락을 반영한 원인 설명에는 한계가 있다[2]. 본 논문에서는 제지공정 시계열 데이터에서 예측 오차를 기준으로 식별된 이상 구간에 대해 SHAP 기반 변수 영향 분석 결과를 구조화하여 LLM(Large Language Model)을 활용한 자연어 이상 원인 설명을 자동 생성하는 방법을 제안한다. 이를 위해 공정 데이터를 정형 입력 구조로 구성하고 구조적 프롬프트를 적용한다. 실험 결과, 제안 방법은 기존 설명과의 비교를 통해 BERTScore 기준 F1-score 0.832를 나타내며, 동일한 이상 시점에 대해 의미적으로 일관된 자연어 설명을 생성한다.

I. 서론

제지공정에서는 스팀 압력, 수분, 평량과 같은 다수의 공정 변수들이 상호작용하는 복잡한 구조를 가지며, 이들 변수 간의 상관관계를 고려한 데이터 기반 분석과 이상 원인 해석은 에너지 효율 향상과 품질 유지를 동시에 달성하기 위한 핵심 과제로 제시되고 있다[1]. 이러한 환경에서는 공정 조건 변화가 누적되며 다양한 이상 현상이 발생하지만, 실제 현장에서는 이상 발생 원인을 설명하기 어려운 한계가 존재한다. 최근에는 LSTM(Long Short-Term Memory)과 같은 시계열 예측 모델을 활용하여 공정 변수의 시간적 의존성을 반영하고, 예측 오차를 기반으로 이상 구간을 식별하는 접근을 활용한다. 또한 SHAP(SHapley Additive exPlanations)와 같은 설명 가능 인공지능(XAI) 기법을 통해 예측 결과에 대한 변수 기여도를 정량적으로 분석하려는 시도가 이루어지고 있다. 그러나 이러한 기법들은 주로 수치나 시각화 중심의 결과를 제공하여, 공정 운영자가 이상 원인을 공정 맥락에서 직관적으로 이해하는 데에는 여전히 한계가 존재한다[2].

본 연구에서는 LSTM 기반 예측 모델을 통해 식별된 이상 구간에 대해 SHAP 기반 변수 영향 분석을 수행하고, 그 결과를 구조화된 입력 형태로 구성하여 대규모 언어 모델(LLM)을 활용한 자연어 기반 이상 원인 설명 생성 방법을 제안한다. 본 방법은 예측 오차 형상에 기여한 주요 공정 변수를 중심으로, 수치 기반 XAI 분석 결과를 공정 맥락을 반영한 자연어 설명으로 변환함으로써 제지공정 이상 원인 해석의 이해 가능성을 향상시킨다. 제안 방법은 공정 데이터를 정형 입력 구조로 재구성하고, 구조적 프롬프트를 통해 “이상 원인 - 근거 데이터 - 영향 변수 - 권장 조치” 형태의 일관된 설명을 생성하도록 설계되었다.

II. 본론

II-1. 이상 시점 정의 및 선정

본 연구에서는 실제 스팀 사용량과 예측값 간의 절대 오차를 기반으로 이상 시점을 정의하며, 그중 오차가 최대인 지점을 대표 이상 시점으로 선정하여 변수 기여도 분석 및 설명 생성의 기준으로 활용한다.

II-2. 이상 시점 시계열 분석

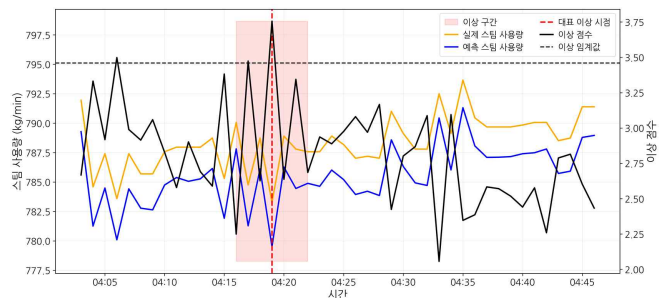


Figure 1. 이상 구간이 포함된 스팀 사용량 시계열 및 이상 점수

Figure 1은 대표 이상 시점을 중심으로 실제 스팀 사용량과 예측값 간의 오차의 시계열 변화를 분석하여, 이상 현상이 시간 구간에 걸쳐 확인된다. 대표 이상 시점을 전후한 구간에서 이상 점수가 임계값을 초과하며, 실제 값과 예측값 간의 오차가 확대되는 경향이 나타난다. 이를 통해 이상 현상이 단일 시점이 아닌 일정 시간 구간에 걸쳐 나타내며, 이후 분석에서는 이 대표 이상 시점을 기준으로 변수 기여도 분석과 이상 원인 설명 생성을 수행한다.

II-3. 대표 이상 구간의 변수 영향 분석 및 설명 입력 구성

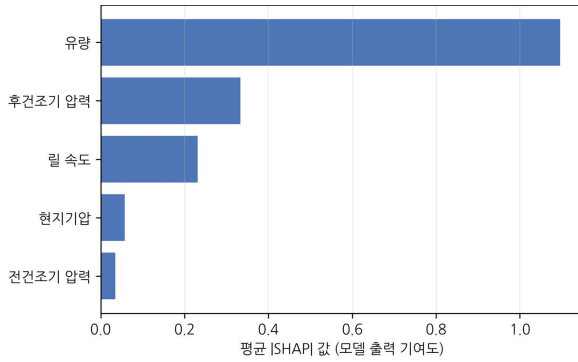


Figure 2. 대표 이상 구간 평균 기반 변수 기여도

대표 이상 구간에서 예측 오차 형성에 기여한 주요 공정 변수를 SHAP 기반으로 분석하여, 이후 LLM 기반 이상 원인 설명 생성의 입력으로 활용한다. Figure 2는 해당 구간에 대해 산출된 변수 기여도를 나타낸다. 본 분석은 예측 모델 관점에서 이상 구간 동안 어떤 입력 변수가 스팀 사용량 예측에 상대적으로 큰 영향을 미쳤는지를 파악하는 데 목적이 있으며, 단일 시점의 일시적 변동이나 노이즈 영향을 완화하고 구간 전반에 걸쳐 반복적으로 나타나는 변수 영향 패턴을 반영한다.

이를 위해 대표 이상 구간 내 시점들을 대상으로 각 변수의 SHAP 값 절대값을 계산한 후, 구간 전체에 대한 평균값을 산출한다. 분석 결과, 대표 이상 구간에서는 유량 변수가 가장 높은 평균 SHAP 값을 나타내며 예측값 형성에 지배적인 영향을 미친 변수로 확인된다. 다음으로 후건조기 압력과 릴 속도가 상대적으로 유의미한 기여도를 보이며, 현지기압(hPa) 및 전건조기 압력의 영향도는 제한적인 수준으로 나타났다. 이는 해당 이상 현상이 다수의 변수 변화가 동시에 작용한 결과라기보다는, 특정 공정 변수들을 중심으로 예측 모델의 출력이 크게 변동한 구간임을 시사한다. 이러한 변수 영향 분석 결과를 바탕으로, 예측 모델 관점에서 의미 있는 영향을 미친 상위 변수들만을 선별하여 LLM 설명 입력으로 구성하였다. 이를 통해 설명의 간결성을 유지하는 동시에, 데이터 기반 신뢰성을 확보한다.

II-4. 구조적 프롬프트 기반 Qwen LLM 이상 원인 설명 시나리오

앞선 시계열 분석 및 변수 영향 분석 결과를 기반으로, 대표 이상 시점의 이상 원인을 LLM을 통해 자연어로 설명한다. 본 연구에서 LLM은 이상 탐지나 예측을 수행하는 모델이 아니라, 앞선 시계열 분석 및 변수 영향 분석 결과를 해석·전달하는 설명 계층으로 활용한다. 이를 통해 수치 및 시각화 중심의 XAI 결과를 공정 맥락을 반영한 자연어 설명으로 확장하고자 하였으며, 사전 학습된 Qwen2.5 3B Instruct 모델을 활용한다.

LLM의 입력은 대표 이상 시점의 실제 스팀 사용량, 예측값, 절대 오차와 함께, SHAP 기반 변수 영향 분석을 통해 도출된 주요 공정 변수들로 구성하였다. 모든 입력 정보는 Key - Value 형태로 정형화하여 제공하였으며, 입력 데이터 외 정보의 개입을 방지하기 위해 구조적 프롬프트(Structured Prompting)를 적용하였다[3]. 설명 출력은 이상 원인, 근거 데이터, 영향 변수, 권장조치의 네 가지 항목으로 제한한다.

구조적 프롬프트를 적용한 Qwen 기반 LLM은 대표 이상 시점에서 실제 스팀 사용량이 예측값보다 크게 나타났으며, 유량과 릴 속도를 주요 영향 변수로 식별하였다. 이를 통해 구조적 프롬프트 기반 LLM이 앞선 분석 결과에 기반하여 산업 공정 이상 시점에 대해 자연어 원인 설명을 생성한다.

II-5. 설명 품질 평가

LLM이 생성한 이상 원인 설명의 품질을 평가하기 위해, 대표 이상 시점에 대해 수행된 SHAP 기반 변수 영향 분석 결과를 토대로, 연구자가 주요 영향 변수와 예측 오차 특성을 요약한 기준 설명(reference)을 작성하고 이를 비교 기준으로 활용한다. 동일한 대표 이상 시점에 대해 LLM이 생성한 5개의 이상 원인 설명 문장을 기준 설명과 1:1로 비교하여 BERTScore를 산출한다. 각 설명 쌍에서 계산된 점수의 평균값을 설명 품질 지표로 사용함으로써, LLM이 생성한 설명이 기준 설명의 핵심 정보를 의미적으로 얼마나 일관되게 반영하는지를 평가한다.

Metric	Score
Precision	0.864
Recall	0.803
F1-score	0.832

Figure 3. BERTScore 기반 LLM 설명 품질 평가 결과

Figure 3은 LLM이 생성한 이상 원인 설명과 기준 설명 간의 의미적 유사도를 BERTScore로 평가한 결과를 나타낸다. BERTScore는 사전 학습된 언어 모델 임베딩을 기반으로 설명 간 핵심 정보의 의미적 유사성을 평가하는 지표이다. 평가 결과, Precision은 0.864, Recall은 0.803, F1-score는 0.832로 나타났으며 이는 LLM이 SHAP 기반 변수 영향 분석을 통해 도출된 주요 공정 변수와 이상 특성을 기준 설명과 개념적으로 일관되게 반영하여 설명을 생성한다.

III. 결 론

본 연구에서는 제지공정 시계열 데이터에서 스팀 사용량의 실제값과 예측값 간 오차를 기반으로 이상 시점을 정의하고, 해당 이상 현상에 대해 데이터 기반 자연어 설명을 자동 생성하는 LLM 활용 방법을 제안한다. 대표 이상 시점을 중심으로 시계열 분석과 SHAP 기반 변수 기여도 분석을 적용하며, 이를 통해 예측 모델 관점에서 영향도가 높은 주요 공정 변수를 식별하였다. 식별된 변수 정보와 예측 오차 관련 데이터를 정형화하여 구조적 프롬프트 기반 Qwen 계열 LLM의 입력으로 제공한 결과, LLM은 입력 데이터에 근거한 일관된 이상 원인 설명을 생성하였으며, 설명 품질 평가를 통해 의미적 일관성을 정량적으로 검증한다. 이러한 결과는 LLM이 이상 탐지나 예측을 수행하지 않더라도, 기존 분석 결과를 해석하고 전달하는 설명 생성 도구로서 효과적으로 활용될 수 있음을 보여준다. 본 연구는 시각화 중심의 기존 XAI 기법이 갖는 해석상의 한계를 보완하여, 산업 공정 이상 분석 결과를 도메인 맥락에 맞는 자연어 설명으로 확장할 수 있는 가능성을 제시한다.

참 고 문 헌

- [1] Lee, Sangkeum, et al. "Factory Energy Management by Steam Energy Cluster Modeling in Paper-Making." Proceedings of the 2023 11th International Conference on Smart Grid (icSmartGrid), Paris, France. 2023.
- [2] Lundberg, S. M. and Lee, S.-I., "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems (NeurIPS), 2017.
- [3] Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., and Ba, J., "Large Language Models Are Human-Level Prompt Engineers," Proceedings of the International Conference on Learning Representations (ICLR), 2023. (arXiv:2211.01910)