

질소화합물 배출량 예측 모델

발표자: 하현수

팀원: 하현수(조장), 박민상, 이종민, 임채원

Github: <https://github.com/1000mmoney/1team>

1. 안전 관련 머신러닝 모델 개발 관련 요약

이 프로젝트는 가스터빈의 배출가스 데이터를 분석하고 예측하기 위해 머신러닝 모델을 개발하는 것을 목표로 한다.

가스터빈의 배출가스는 환경 및 안전과 직결되는 중요한 요소로, 이 데이터를 통해 배출가스의 수준을 예측함으로써 환경 보호와 기계의 효율성을 개선할 수 있다.

데이터셋에는 가스터빈의 작동 조건과 배출가스 농도가 포함되어 있으며, 이 정보를 바탕으로 예측 모델을 학습시키고 성능을 평가한다.

2. 개발 목적

- a. 머신러닝 모델은 가스터빈의 배출가스를 예측하는 데 활용된다.
이는 가스터빈의 안전성과 효율성을 높이기 위한 목적을 가지고 있으며, 예측된 배출가스 데이터를 통해 적절한 유지보수 및 운영 조치를 취할 수 있다.
- b. 모델 개발의 의의는 가스터빈의 운영 조건에 따른 배출가스의 변화를 예측하여 환경 보호와 기계의 수명을 연장하는 데 있다.
예측된 결과를 기반으로 가스터빈의 운전 조건을 조정하고, 배출가스의 최소화를 통해 환경적 가치를 창출할 수 있다.
- c. * 독립변수 : 가스터빈의 운영 상태를 나타내는 변수들
(예: 압력, 온도, 유량 등)
* 종속변수 : 배출가스 농도
(예: NOx, CO 등)

3. 배경지식

- a. 가스터빈의 배출가스는 환경 오염의 주요 원인 중 하나다.
질소화합물(NO_x)은 일산화질소(NO), 이산화질소(NO_2)와 같은 기타 질소산화물을 통칭하는 용어이며, 질소화합물 (NO_x)는 공기 중으로 배출되었을 때 햇빛의 광화학 반응을 통해 미세먼지와 오존 등을 생성하며 EURO-7 과 대기환경보전법 및 사업장 대기오염 총량관리제 등을 통해 심각성이 대두된다
화력발전소, 소각로, 건설현장의 덤프트럭 등 인위적으로 발생하는 질소화합물은 대기오염의 원인이 되며 현장 작업자의 호흡계 질환, 심혈관계 질환을 유발한다.
따라서, 배출가스를 효율적으로 관리하고 예측하는 것은 환경 보호와 공공 건강을 위해 중요하다.
- b. 머신러닝 모델은 대량의 데이터를 분석하여 패턴을 학습하고 예측하는 기술이다.
물리적 모델을 기반으로 한 질소화합물(NO_x) 배출량은 연료 및 공기의 유체거동과 화학반응을 동시에 고려한 모델이 필요하기 때문에 실시간 질소화합물(NO_x) 배출량 모니터링은 한계를 지닌다.
따라서 Python 를 활용하여 데이터를 통해 인공지능을 Machine Learning 으로 학습시켜 질소화합물(NO_x)배출량을 실시간으로 산출하여 모니터링할 수 있는 프로그램을 개발이다.
이 프로젝트에서는 가스터빈의 운영 데이터를 기반으로 배출가스를 예측하기 위해 회귀 분석을 포함한 여러 머신러닝 알고리즘을 활용한다.
모델의 정확도를 높이고, 예측의 신뢰성을 확보하는 것이 주요 목표이다.

4. 개발 내용

- a. 데이터에 대한 구체적 설명 및 시각화
 - i. * **데이터 개수** : 36,733 개의 샘플
* **데이터 속성** : 10 개 변수 (운영 조건, 가스터빈 상태 등)
 - ii. 압력과 온도가 배출가스 농도에 미치는 영향을 분석하여 주요 영향을 미치는 요소를 식별한다.

- b. 데이터에 대한 설명 이후, 어떤 것을 예측하고자 하는지 구체적으로 설명
 - i. * **독립 변수** : 가스터빈의 운전 조건 (온도, 압력 등)
 - * **종속 변수** : 질소화합물 배출가스 농도
- c. 머신러닝 모델 선정 이유
 - i. 터빈에서 나온 물질들이 질소화합물과 얼마나 관계가 있는지 확인하기 위해 선형 회귀 등의 모델을 선택했다.
각 모델의 특성과 성능을 고려하여 최적의 예측 결과를 도출한다.
 - ii. 다양한 모델을 선정하여 성능을 비교함으로써, 가장 정확하고 신뢰할 수 있는 예측 모델을 식별한다.
각 모델의 예측 성능을 평가하여 최적의 모델을 결정한다.
- d. 사용할 성능 지표
 - i.
 - * **MAE(Mean Absolute Error)** : 예측값과 실제값 간의 평균 절대 오차
 - * **RMSE (Root Mean Squared Error)** : 예측값과 실제값 간의 평균 제곱근 오차
 - * **MSE (Mean Squared Error)**: 예측값과 실제값 간의 평균 제곱 오차
 - * **Accuracy**: 분류 문제의 정확도
 - * **오차행렬**: 예측의 정확도와 오류를 시각화
 - ii. 이 지표들은 예측의 정확성과 신뢰성을 평가하기 위한 기준으로, 다양한 측면에서 모델의 성능을 평가할 수 있다.

5. 개발 결과

- a. 성능 지표에 따른 머신러닝 모델 성능 평가
 - i. 수치 자료 및 시각화 자료를 사용
 - * **MAE, RMSE, MSE**: 각 모델의 평균 절대 오차, 평균 제곱근 오차, 평균 제곱 오차
 - * **Accuracy, 오차행렬**: 모델의 정확도와 예측 오류를 시각화한 자료
 - ii. 여러 모델의 성능을 비교하여 가장 우수한 예측 성능을 보이는 모델을 선정한다.

- b. 모델의 성능 결과를 분석하여 예측의 정확성, 신뢰성, 그리고 각 모델의 장단점을 설명한다.

6. 결론

- a. 가스터빈 배출가스 예측을 위한 다양한 머신러닝 모델을 개발하고 평가한다.
최종적으로 가장 높은 성능을 보인 모델을 선정하였으며, 이 모델은 가스터빈의 안전성과 효율성을 높이는 데 기여할 수 있다.
- b. 여러 모델을 비교함으로써 배출가스의 예측 정확도를 높이고, 위해노출 농도에 따른 인체에 미치는 영향을 예측할 수 있다.
- c. 모델의 한계로는 데이터의 품질과 양, 그리고 특정 환경 변수에 대한 반응 예측의 어려움이 있을 수 있다.
이를 보완하기 위해 추가적인 데이터 수집과 모델 개선이 필요할 수 있다.
- d. 따라서 이 모델은 미세먼지 등의 비산물질 발생 사업장에서 데이터를 추출하여 배출가스의 예측정확도를 상승시켜 작업자의 인체에 미칠 수 있는 영향을 예측하고 이에 따른 보호조치 및 환경조성 등을 기대 할 수 있기 때문에 해당 작업에 활용 될 수 있다.
- e. 작업장 환경에 대한 실시간 데이터를 통해 질소산화물의 배출량을 정확하게 예측하여 근로자의 건강을 보호하고, 쾌적한 작업환경을 조성하여 산업재해 예방의 기여할 수 있을 것으로 예상할 수 있다.
- f. 질소화합물 뿐 아니라 건설분야에서 소요되는 에너지와 이산화탄소 등 다른 오염물질의 배출량 모니터링에 활용 가능할 것으로 보인다.