

보일러 효율 영향 인자 확인 및 최적화 효율 계산 인공지능 모델 개발

박주열 권보민 김희성 전동환 아스카로바 장을므르자^o 정설영 김치관
경북대학교 컴퓨터학부, 대림로얄엔피
qkrjuyeol@gmail.com rnjsqhalsdl@gmail.com gim864830@gmail.com jdh7351@gamil.com janyla.ask03@gmail.com
snowflower@knu.ac.kr dlrkim@naver.com

Development of an AI Model for Identifying Factors Affecting Boiler Efficiency and Calculating its Optimal Efficiency

Juyeol Park, Bomin Kwon, Heeseong Kim, Donghwan Jeon,
Askarova Zhanylmyrza^o, Seolyoung Jeong, Chigwan Kim
School of Computer Science and Engineering, College of IT Engineering, Kyungpook National University,
Daelim Royal EnP

요 약

현재 산업용 보일러의 데이터를 분석하기 위해 사용하는 효율 계산 공식만으로는 보일러의 효율을 더 높이기엔 한계가 있다. 이 한계를 극복하기 위해 우리는 현장에서 3톤 산업용 보일러가 2개월 간 가동한 데이터를 사용하였고, Random Forest 알고리즘과 XGBoost 알고리즘을 활용한 모델의 피쳐 중요도로 데이터를 분석하였다. 또한, Random Forest Regresssion 모델의 예측값을 차분 진화 알고리즘 모델의 손실함수로 사용하여 최종적으로 최적의 효율과 해당하는 인자들의 값을 추측해보았다.

1. 서 론

산업용 보일러는 다양한 산업 현장에서 열에너지를 제공하는 핵심 설비로서, 그 효율성은 생산성과 경제성에 큰 영향을 미친다. 높은 보일러 효율은 에너지 비용 절감뿐 아니라, 탄소 배출 저감에도 기여함으로써, 비용 측면뿐 아니라 환경적 측면에서도 매우 중요하다. 현재는 보일러의 효율을 평가하고 최적화하는 방식으로 KS B 6205 공식을 사용하고 있다. 그러나 이러한 이론적 접근 방식은 실제 현장의 동적 요인과 물리적 한계를 반영하지 못한다. 이에 따라, 현재의 방식으로는 더 높은 효율성을 달성하는데 한계에 직면하게 되었다.

본 연구에서는 이러한 문제를 해결하기 위해 머신러닝과 딥러닝 기반의 효율 예측 모델을 개발하고자 한다. KS B 6205 공식에서 다루지 않는 다양한 변수들을 포괄적으로 반영하고, 실제 데이터로부터 학습된 모델을 통해 더 높은 수준의 예측을 목표로 한다. 실제 현장에서 관측된 데이터를 기반으로 보일러 효율에 영향을 미치는 주요 요인들을 분석하고, 이를 최적화하는 인공지능 모델을 제안한다.

연구의 결과는 산업용 보일러의 효율성을 개선하는 데 실질적인 도움을 줄 뿐 아니라, 탄소 배출량 감소에도 기여할 수 있을 것으로 기대된다. 머신러닝과 딥러닝을

적용한 예측 모델은 보일러 운영의 효율성 최적화에 있어서 기존의 한계를 극복하는 새로운 접근법이 될 수 있으며, 이를 통해 산업 현장의 에너지 관리에 중요한 혁신을 제공할 가능성이 있다.

2. 관련 연구

산업용 보일러는 석유화학, 발전 등 다양한 산업 현장에서 에너지를 발생시키는 중요한 설비 중 하나이다. 이러한 보일러는 여러 건물 등에도 포함되어 있으며, 에너지 사용에 큰 영향을 주기 때문에 가동 효율의 중요성이 크다. Nemitallah Medhat A 등[1]은 인공 신경망과 유전 알고리즘 등 AI와 ML을 활용하여 보일러 최적화 가능성을 탐구하고 있다. Vladislav Kovalnogov 등[2]은 버너의 효율성을 높이기 위해 Random Forest 방법을 사용하였다. Pauli Virtanen 등[3]은 파이썬의 다양한 오픈 소스 라이브러리를 활용하여 실험적인 계산을 하는 것에 대해 말하고 있다.

위의 연구들은 보일러의 실제 가동 데이터를 바탕으로 하고 있지 않으며, 여러 인공지능 알고리즘을 사용하여 예측, 분석한다는 연구 방향은 같지만 보일러 시스템의 전체적인 운전 최적화를 목표로 하고 있지 않다. 본 연구는 기존의 연구들이 다루지 못한 다양한 조건과 운용

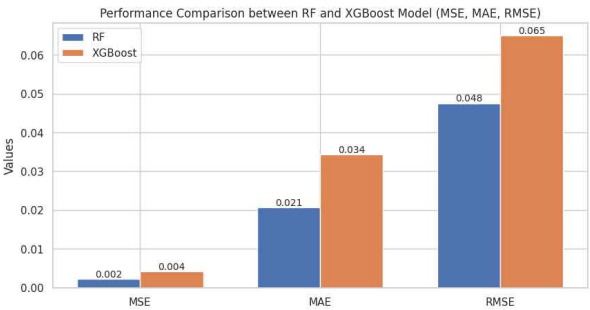


그림 1 Ramdom Forest 모델과 XGBoost 모델의 성능 비교 (MSE, MAE, RMSE)

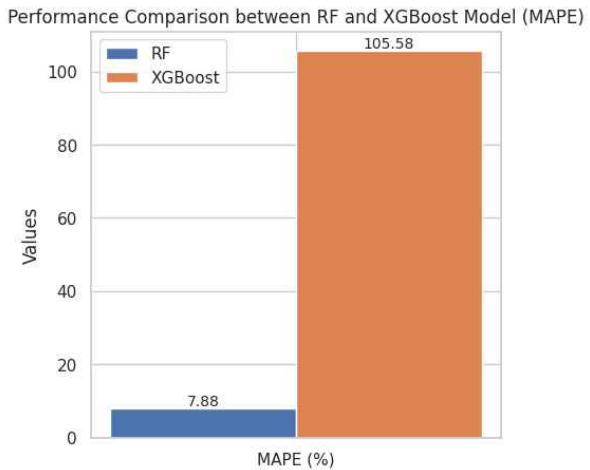


그림 2 Ramdom Forest 모델과 XGBoost 모델의 성능 비교 (MAPE)

데이터를 고려하여 보다 실용적인 모델을 개발하고자 한다.

3. 본 론

3.1 Random Forest 알고리즘 및 XGBoost 알고리즘 기반 모델 설계

XGBoost 알고리즘과 Random Forest 알고리즘은 예측 성능이 높고 다양한 문제에 효과적으로 적용될 수 있다. 또한, 두 알고리즘 모두 트리를 기반으로 하여 정형 데이터를 처리하기에 적합하다. 높은 성능의 모델을 만들기 위해 두 알고리즘을 사용하여 인공지능 모델을 만들고 그 결과와 성능을 비교했다.

XGBoost는 Gradient Boosting 알고리즘을 확장하여 성능과 효율성을 향상시킨 알고리즘이다. Gradient Boosting은 순차적으로 트리를 추가하여 예측 성능을 향상시키는 학습 방법이다. 정규화, 병렬 처리, 조기 종료 등 여러 최적화 기법을 기존의 Gradient Boosting 알고리즘에 사용하여 향상된 성능을 보여준다. 각 트리는 이전 트리의 오차를 줄이기 위해 학습되며, 손실 함수를 최소화하는 방향으로 트리를 추가한다.

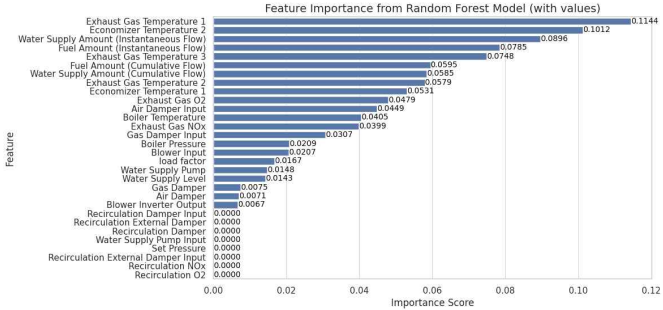


그림 3 Random Forest 모델을 통해 얻은 보일러 효율 인자 중요도

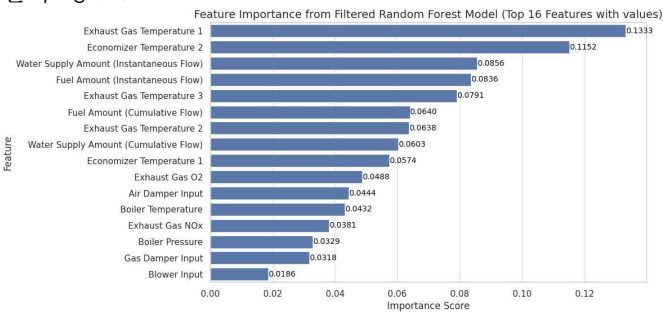


그림 4 상위 중요도 인자를 Random Forest 모델을 통해 재학습시킨 모델의 인자 중요도

Random Forest는 의사결정나무를 배경(Bagging) 방식으로 만든 알고리즘으로, 부트스트랩을 통해 다양한 서브 데이터셋을 생성한다. 여러 개의 의사결정나무가 각각의 데이터셋을 학습하고 결과를 취합하여 단일 의사결정나무의 과적합 문제를 해결할 수 있는 알고리즘이다. Random Forest는 높은 안정성과 예측력을 보이며, 특히 다중공선성 문제를 효과적으로 다룬다.

예측 모델의 성능을 높이기 위해 하이퍼파라미터 튜닝 방법으로 랜덤 서치(Random Search) 기법을 활용하였다.

그림 1와 그림 2는 여러 성능 평가 지표에 대해 두 모델을 비교한 결과이다. MSE, MAE, RMSE 세 가지 지표에서 Random Forest 모델의 성능이 근소하게 앞선다. 하지만 MAPE 값에서 Random Forest 모델보다 XGBoost 모델의 성능이 크게 떨어져 이후의 연구에서 Random Forest 모델을 기반으로 인공지능 모델 개발을 진행하였다.

3.2 상위 중요도 인자 Random Forest 모델 재학습

Random Forest 모델 학습 결과를 통해 피쳐 중요도를 분석하였다. 분석 결과는 그림 3에 나타난다. 그림 3의 결과에서 영향력 있는 인자값에 집중하기 위해 임계값을 0.01로 설정하여 중요도가 임계값 이상인 인자를 색출하였다. 색출한 인자를 사용하여 Random Forest 모델을 재

학습하였다. 재학습을 통해 그림 4에서와 같이 상위 중요도 피쳐의 중요도를 다시 뽑아냈다.

이 재학습 과정에서는 이미 최적의 하이퍼파라미터가 결정된 상태이므로 추가적인 하이퍼파라미터 탐색 과정을 생략할 수 있었다. 따라서 학습 속도가 크게 향상되었고, 중요한 피쳐만을 활용함으로써 밀도 높은 예측을 할 수 있었다.

3.3 Differential Evolution 알고리즘 기반 모델 설계

보일러 효율을 예측하는 회귀 모델을 활용하여 보일러 효율 영향 인자를 최적화하기 위해 진화적 최적화 알고리즘 모델의 사용을 고려했다.

진화적 최적화 알고리즘은 생물의 진화 원리를 바탕으로 최적의 해를 탐색하는 기법으로 선택, 교차, 변이, 대체 과정을 반복하여 loss function의 값이 최소가 되도록 만든다. 유전 알고리즘은 loss function의 유형에 대해 크게 구애받지 않기 때문에 Random Forest Regressor 모델이 계산한 예측값을 점수로 활용하는 현 상황에 가장 적합한 모델이라고 판단되었다.

현재 입력값으로 주어지는 값이 모두 실수이기 때문에 진화적 최적화 알고리즘 중 실수형 매개변수 최적화에 특화된 알고리즘인 차분 진화(Differential Evolution - DE) 알고리즘을 기반으로 모델을 설계하였다. 각 컬럼의 bound를 전처리 과정을 거친 원본 데이터에서의 각 컬럼의 최소값과 최대값으로 설정하였고 차분 진화 알고리즘은 loss function을 최소화하기 때문에 매개변수 조합에 대한 Random Forest Regressor 모델의 효율 예측값의 음수를 loss function으로 사용하여 진행하였다.

그림 5는 해당 과정을 거쳐 나온 최적의 효율 인자의 값과 효율 예측값이다. DE를 사용하여 재학습된 RF모델의 효율성 결과가 약 95%로 나타났다.

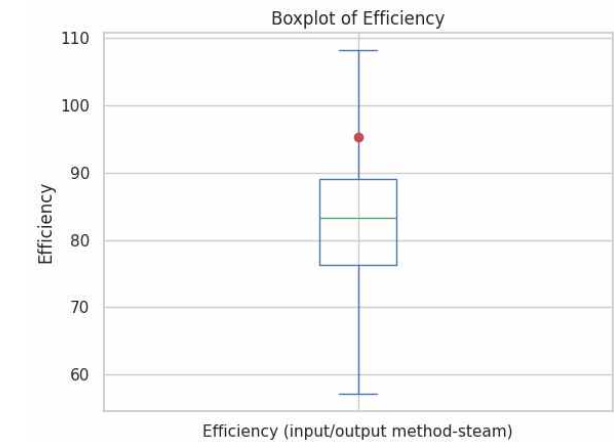


그림 5 최적화된 조합에서 예측된 효율성 값

3.4 역전처리 및 복원

최적화된 효율 인자 값과 효율 값은 전처리 과정을 거친 데이터이므로, 사용자에게 제공하기 전에 원래 스케일로 복원하는 과정이 필요하다. 이를 통해 데이터는 원래의 스케일로 복원되며, 사용자는 전처리 전의 해석 가능한 값으로 결과를 확인할 수 있다.

이 과정이 올바르게 수행될 수 있도록, 전처리 시 사용된 모든 컬럼이 포함된 데이터프레임을 0으로 초기화하여 새롭게 생성하고 최적화된 효율 인자와 효율 값을 해당 컬럼에 삽입한 이후 스케일 복원을 수행하였다. 이를 통해 데이터는 원래의 스케일로 성공적으로 복원되었으며, 사용자는 전처리 전의 해석 가능한 값으로 결과를 확인할 수 있었다.

4. 결 론

3톤 용량의 산업용 보일러의 가동 효율을 KS B 6205 공식으로 단순히 계산하는 것에 그치지 않고, RF모델로 효율에 영향을 주는 주요 인자를 추출하고 DE 모델로 피쳐값을 예측하였다. 연구의 결론과 한계점을 정리하면 다음과 같다.

- 1. 보일러를 가동하는 동안 RF모델로 효율에 영향을 미치는 값은 KS B 6205의 공식에 포함되는 값 말고도 다양하게 존재했다.
- 2. RF모델로 추출한 피쳐 중요도로 보일러 효율에 영향을 주는 인자를 판단했을 때, 18개의 인자가 보일러의 효율에 주로 기여한다는 것을 알 수 있었다.
- 3. DE 모델을 통해 주요 인자들과 효율 값을 예측한 결과 현재 3톤 보일러의 최대 효율보다 높은 효율을 나타내는 인자값의 조합을 찾을 수 있었다.
- 4. DE 모델을 통해 최대의 보일러 효율을 내는 인자들을 찾았지만, 현실의 물리적 한계로 인해 실제로 효율 향상에 기여하지 못할 수 있다는 한계가 있다.

참고문헌

[1] Nemitallah Medhat A ,Nabhan Mohammad A ,Alowaifeer Maad,Haeruman Agus,Alzahrani Fahad,Habib Mohamed A ,Elshafei Moustafa,Abouheaf Mohammed I ,Aliyu Mansur,Alfarraj Motaz. “Artificial intelligence for control and optimization of boilers’ performance and emissions: A review“, Journal of Cleaner Production, 417, 10, 138109, 2023

- [2] Vladislav Kovalnogov, Ruslan Fedorov, Vladimir Klyachkin, Dmitry Generalov, Yulia Kuvayskova, Sergey Busygin., “Applying the Random Forest Method to Improve Burner Efficiency” , Mathematics, 10, 12, 2143, 2022
- [3] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, SciPy 1.0 Contributors, “SciPy 1.0: fundamental algorithms for scientific computing in Python” , Nature Methods, 17, 3, 261-272, 2020.

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음.