Attention-Based Deep Learning Methods for Predicting Gas Turbine Emissions

Rebecca Potts and Georgios Leontidis

School of Natural and Computing Sciences & Interdisciplinary Centre for Data and AI, University of Aberdeen, Aberdeen, UK

Abstract

Predictive emissions monitoring systems (PEMS) for gas turbines are critical for monitoring harmful pollutants being released into the atmosphere, while reducing the use of expensive continuous emissions monitoring systems (CEMS) which require daily maintenance to achieve accurate readings. We consider two attention-based deep learning models, FT-Transformer and SAINT, and compare with classical tree-based XGBoost to predict emissions from gas turbines. We find that the attention-based models outperform XGBoost for both prediction tasks, i.e. carbon monoxide (CO) and nitrogen oxides (NOx).

1 Introduction

A predictive emissions monitoring systems (PEMS) model is trained on past data, utilising process parameters such as temperatures and pressures, and uses real-time data to generate estimations for emissions. Gas turbine process data can be considered tabular data. Classical methods such as XGBoost (Chen and Guestrin [2016]) and CatBoost (Prokhorenkova et al. [2018]) have shown excellent results in the tabular domain, and are often seen as the standard solution for structured data problems.

Previous comparisons between neural networks (NNs) and classical methods such as XGBoost for tabular regression generally conclude that classical methods match or outperform the NN based models. For example, Kossen et al. [2021] find that taking the entire dataset as input and using self-attention between datapoints is outperformed by CatBoost and only matches the performance of XGBoost. Grinsztajn et al. [2022] find that classical tree-based methods outperform the NN models, even on numerical only datasets, whereas a survey by Borisov et al. [2021] concludes that deep learning tabular methods outperform classical when a dataset consists of mostly numerical features.

Previous machine learning methods used for gas turbine PEMS include Vanderhaegen et al. [2010] and Si et al. [2019] who compared different configurations of NN based models, Cuccu et al. [2017] compared 12 different machine learning algorithms and Azzam et al. [2018] explored a genetic algorithm to tune the hyperparameters of a NN and support vector machines for their PEMS.

Kaya et al. [2019] collected a novel PEMS dataset to develop a benchmark PEMS using an extreme learning machine (ELM) regressor using three fusion strategies: averaging, random forest, and basic ELM. We use this dataset to utilise attention-based deep learning methods for PEMS, the first time these have been used for this problem.

We compared two attention-based deep learning models, FT-Transformer (Gorishniy et al. [2021]) and SAINT (Somepalli et al. [2021]), against the tree-based XGBoost and the three ELM fusion strategies seen in Kaya et al. [2019] to predict gas turbine emissions.

2 Methodology

FT-Transformer (Gorishniy et al. [2021]) is an adaptation of the Transformer architecture for tabular data. Transformers (Vaswani et al. [2017]) utilise multi-head self-attention to jointly attend to information to determine which inputs are the most important at a time. In FT-Transformer, input features are embedded via the Feature Tokenizer module and a stack of Transformer layers are applied to all embedded input features, such that every Transformer layer operates on the feature level of one object.

SAINT (Somepalli et al. [2021]), the Self-Attention and Intersample Attention Transformer, is a hybrid deep learning approach for tabular data problems. Attention is performed over both rows and columns, whereby the self-attention attends to individual features within each data sample, and intersample attention relates each row to other rows in the table.

The gas turbine emissions dataset is publicly available and consists of five years of data with 36,733 instances. It has nine input parameters, including temperature and pressure, and two target variables, NOx and CO. It consists of only numerical features and no categorical.

We split the dataset into train, valid and test sets as in the original work (Kaya et al. [2019]), with XGBoost utilising only train and test sets due to the nature of XGBoost. A batch size of 32 was used with a learning rate of 0.0001. The optimiser AdamW was used with MSE loss function. For the multi-head self-attention, 8 attention heads were used for both FT-Transformer and SAINT.

3 Results and Discussion

Overall mean absolute error (MAE) results are shown in Table 1 for test set data. Attention-based models outperform XGBoost and the ELM fusion schemes for both CO and NOx. FT-Transformer records the best results for CO, while SAINT marginally records the best performance for NOx, with very similar results by FT-Transformer.

Table 1: MAE results for attention-based methods FT-Transformer and SAINT compared to tree-based XGBoost and best results of three fusion schemes from Kaya et al. [2019]. Best results shown in bold.

	CO	NOx
	MAE	MAE
FT-Transformer	**0.50**	2.57
SAINT	0.84	**2.51**
XGBoost	1.15	11.65
Averaging	1.05	7.91
Meta-ELM	1.26	24.05
Random Forest	0.93	11.29

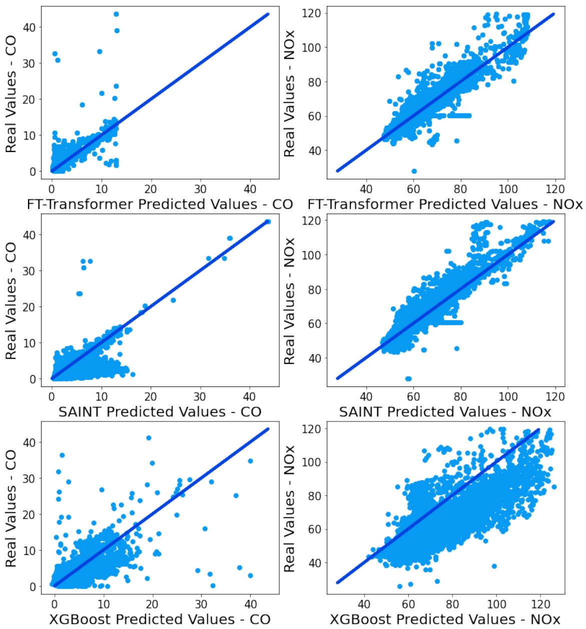

Figure 1: Scatter plot of test set predictions for FT-Transformer, SAINT and XGBoost against real sensor readings for CO and NOx.

Test set predictions for CO and NOx are plotted against corresponding real values in Figure 1, where perfect results would lie on the identity line. The attention-based methods have less spread compared to XGBoost for NOx, showing their superior correlation.

4 Conclusion

We applied and compared two attention-based deep learning models, FT-Transformer and SAINT, and the classical tree-based method XGBoost to predict emissions for a gas turbine dataset and found that the attention-based deep learning models outperformed XGBoost for MAE values, as well as outperforming the original baseline results for this dataset.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council [EP/W522089/1].

References

M. Azzam, M. Awad, and J. Zeaiter. Application of evolutionary neural networks and support vector machines to model nox emissions from gas turbines. *Journal of environmental chemical engineering*, 6(1):1044–1052, 2018.

V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, and G. Kasneci. Deep neural networks and tabular data: A survey. *arXiv preprint arXiv:2110.01889*, 2021.

T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

G. Cuccu, S. Danafar, P. Cudré-Mauroux, M. Gassner, S. Bernero, and K. Kryszczuk. A data-driven approach to predict nox-emissions of gas turbines. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 1283–1288, 2017. doi: 10.1109/BigData.2017.8258056.

Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko. Revisiting deep learning models for tabular data. *Advances in Neural Information Processing Systems*, 34:18932–18943, 2021.

L. Grinsztajn, E. Oyallon, and G. Varoquaux. Why do tree-based models still outperform deep learning on tabular data? *arXiv preprint arXiv:2207.08815*, 2022.

H. Kaya, P. Tüfekci, and E. Uzun. Predicting co and nox emissions from gas turbines: novel data and a benchmark pems. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(6): 4783–4796, 2019.

J. Kossen, N. Band, C. Lyle, A. N. Gomez, T. Rainforth, and Y. Gal. Self-attention between datapoints: Going beyond individual input-output pairs in deep learning. *Advances in Neural Information Processing Systems*, 34:28742–28756, 2021.

L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, 2018.

M. Si, T. J. Tarnoczi, B. M. Wiens, and K. Du. Development of predictive emissions monitoring system using open source machine learning library – keras: A case study on a cogeneration unit. *IEEE Access*, 7:113463–113475, 2019. doi: 10.1109/ACCESS.2019.2930555.

G. Somepalli, M. Goldblum, A. Schwarzschild, C. B. Bruss, and T. Goldstein. Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv preprint arXiv:2106.01342*, 2021.

E. Vanderhaegen, M. l. Deneve, H. Laget, N. Faniel, and J. Mertens. Predictive emissions monitoring using a continuously updating neural network. In *Turbo Expo: Power for Land, Sea, and Air*, volume 43970, pages 769–775, 2010.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.