Natural Language Generation for Operations and Maintenance in Wind Turbines

Joyjit Chatterjee* and Nina Dethlefs
Department of Computer Science & Technology
University of Hull
United Kingdom
j.chatterjee-2018@hull.ac.uk

Abstract

Wind energy is one of the fastest-growing sustainable energy sources in the world but relies crucially on efficient and effective operations and maintenance to generate sufficient amounts of energy and reduce downtime of wind turbines and associated costs. Machine learning has been applied to fault prediction in wind turbines, but these predictions have not been supported with suggestions on how to avert and fix faults. We present a data-to-text generation system using transformers to produce event descriptions from SCADA data capturing the operational status of turbines and proposing maintenance strategies. Experiments show that our model learns feature representations that correspond to expert judgements. In making a contribution to the reliability of wind energy, we hope to encourage organisations to switch to sustainable energy sources and help combat climate change.

1 Introduction

Machine learning can play an integral role in tackling climate change by ensuring greater reliability of variable energy, such as wind energy (Rolnick et al., 2019). There is a growing uptake of machine learning in the wind industry to predict operational anomalies (Zaher et al., 2009; Abdallah et al., 2018). The average wind turbine, for example, suffers from a downtime of 1.6 hours every 1.5 days (Peters et al., 2012), leading to losses of up to 1,600 USD per day (Milborrow, 2018). There is currently a paucity of intelligent decision support systems which cannot only predict the occurrence of an impending fault but also generate a human-intelligible diagnosis of its cause(s). Data-to-text generation has been explored for domains such as weather forecast generation (Sripada et al., 2004; Dethlefs and Turner, 2017; Gkatzia et al., 2017), spatial navigation (MacMahon et al., 2006), sports commentaries (Chen et al., 2010; Mei et al., 2016), amongst others. In this paper, we aim to establish the feasibility of applying deep learning and natural language generation to the wind domain in order to generate informative event descriptions of alarms using SCADA data and historical logs of alarm messages. We hope that our initial study can pave the way towards AI-based intelligent decision support systems for operations and maintenance for wind turbines, making them a more reliable energy source and contributing to the uptake and sustainability of wind energy on the whole.

2 Data Description and Preprocessing

As input features, we used SCADA data from an operational turbine,[1] rated at 7MW, including 102 features from the turbine logs on electrical, temperature and pressure readings, operational sensors as well as meteorological data. As outputs, we use human-authored event descriptions for 26 discrete alarm types. See Table 1 for an example of a data structure and error description, where features

[1] Platform for Operational Data (POD) Disseminated by ORE Catapult: https://pod.ore.catapult.org.uk

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

1-n are sensor readings. A challenge with the original dataset is a substantial class imbalance across alarm types. For instance, the *Pitch System Fatal Error* event accounted for 5,050 cases owing to the fairly common pitch angle disorientation in turbines, while *HPU 2 Pump Active For Too Long* only accounted for 2,525 cases. To avoid a biased generation policy that favours majority classes, we use Synthetic Minority Oversampling Technique (SMOTE) Chawla et al. (2002) to address class imbalance by generating more samples of a certain type while preserving the original data distribution. Ultimately, we obtained 500 examples for each alarm type in an overall dataset of 13,000 samples.

Time Stamp	**Feature 1** (X_0)	**Feature 2** (X_1).....	**Feature n** (X_n)	**Event Description**
dd/mm/yyyy hh:mm:ss	2.104	0.890	8.124	Turbine Operating Normally
dd/mm/yyyy hh:mm:ss	1.245	3.753	9.509	Pitch System Fatal Error

Table 1: Example of an input data structure and corresponding alarm log.

3 Learning models and experiments

We model our NLG system as an encoder-decoder architecture that takes as input a sequence of SCADA features and outputs an alarm event description in natural language. We opt for a transformer architecture for this task in the light of recent results that have shown that a transformer's multi-head attention principle to compute attention weights over sequences in a single iteration can make the model both substantially faster on long sequences as well more accurate in some cases as outputs do not depend on the sequential order of input processing (Vaswani et al., 2017; Devlin et al., 2019). For our experiments, we use a **Transformer** with 8 multi-head attention heads, model size of 64, and 3 dense layers for each head. We use two baselines: (1) **Seq2Seq** is an encoder-decoder model with an LSTM, 200-dimensional word embeddings, 64 hidden neurons, a learning rate of 0.001, dropout of 0.1, and Adam optimisation; and (2) **Seq2Seq (Att)** in a Seq2Seq model with LSTM and Luong attention Luong et al. (2015) using the same hyperparameters as above. For attention, we use the *concat* score function to compute the alignment vectors, alongside the *dot* and *general* functions. We split our dataset in a 80%-20% ratio into training and test data and use a batch size of 32. All models were trained over 200 epochs.

4 Results

Table 2 shows results in terms of BLEU scores as well as human ratings and computation time.

Model	**BLEU-4**	**Semantic similarity**	**Fluency**	**Computation time**
Seq2Seq	0.454 ± 0.220	3.65 (4)	3.14 (3)	**1.42 min**
Seq2Seq (Att)	0.443 ± 0.226	3.59 (4)	3.17 (3)	4.25 min
Transformer	$\mathbf{0.492 \pm 0.196}$	**3.96 (4)**	**3.36 (4)**	3.30 min

Table 2: Results in terms of BLEU score (with standard deviation), human ratings (median ratings shown in parentheses) and computation time. The best performing model is shown in bold-face.

Objective Evaluation We can see that our **Transformer** clearly outperforms the other two models in terms of BLEU scores, where **Seq2Seq** scores slightly higher than **Seq2Seq (Att)**, likely due to the relatively short sequences (5.49 words on average with a max sequence length of 14 words in the training data). We also observe that **Seq2Seq** is the fastest model, followed by the **Transformer**. All computation times were obtained with NVIDIA Tesla K80 GPUs on Google's Compute Engine.

Subjective Evaluation To confirm our objective metrics and assess the semantic correctness of generated outputs as well as fluency, a human rating study was conducted via Amazon Mechanical Turk (AMT). Humans were asked to assign ratings on a 1-5 Likert scale for semantic similarity, where 1 is *not similar at all* and 5 is *identical*. Similarly, for fluency, 1 is *not fluent at all* and 5 is *human fluency*. We generated 200 random messages from the test set per model and asked 87 human judges to rate them, leading to 1200 ratings in total. We can see from Table 2 that the **Transformer** outperforms its baselines in terms of semantic similarity, arguably generating the most correct event descriptions overall. It also scores highest on the fluency metric. We can see the BLEU ranking confirmed with **Seq2Seq** being ranked higher for *semantic similarity* than **Seq2Seq (Att)**, while the models score similarly in terms of *fluency* with **Seq2Seq (Att)** slightly ahead.

References	1. (DEMOTED) Gearbox oil tank 2 level shutdown. 2. Wind direction transducer error 1&3. 3. Sub pitch priv fatal error has occurred more than 3 times in 86,400 Seconds.
Seq2Seq	1. Demoted oil tank shutdown. 2. Wind direction transducer error . 3. Error occurred more than fatal in $NUM seconds.
Seq2Seq (Att)	1. Demoted gearbox oil tank under pressure full brake. 2. Pitch system fatal error. 3. Sub pitch priv critical error has occurred more than $NUM times in $NUM seconds.
Transformer	1. Demoted gearbox oil tank level shutdown. 2. Wind direction transducer error . 3. Sub pitch priv fatal error has occurred more than $NUM times in $NUM seconds.

Table 3: Generated messages for each model, with remarks about their viability.

Error and output analysis Table 3 shows example messages generated with each of our models alongside ground-truth references. For **Seq2Seq** we can see that while reasonable outputs are generated for the first two situations, the model recognises an error in the last situation but is not able to generate a coherent message. The **Seq2Seq (Att)** model generates coherent messages in all cases, but refers to the wrong sub-component in the first example, confusing the oil tank with the yaw brake. Similarly the second example confuses the pitch system with the wind direction transducer. The final message is acceptable but highlights an error as “critical" rather than “fatal". Finally, our **transformer** generates the most semantically correct and coherent messages but still misses out on nuances such as the exact tank that is being shut down.

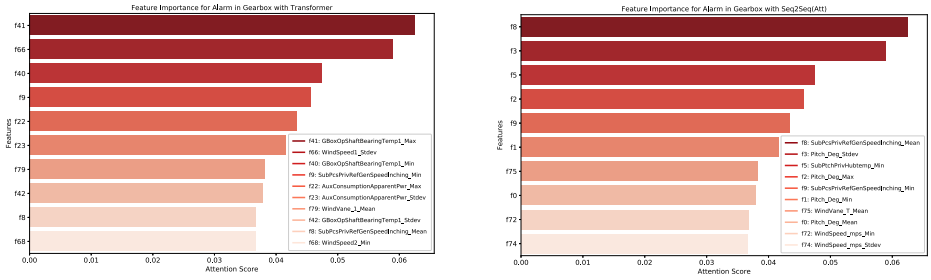

Figure 1: Feature importance plot for an anomaly alarm in gearbox obtained with **Transformer** model (left) and with a **Seq2Seq (Att)** model (right).

Figure 1 compares the features learnt by our **Transformer** and **Seq2Seq (Att)** model in terms of the top 10 for the example of a gearbox alarm. According to the **Transformer** *GBoxOpShaftBearingTemp1_Max* and *GBoxOpShaftBearingTemp1_Min* are highly-ranked, which can likely be attributed to overheating of the high speed gearbox shaft bearings and the gearbox housing, thus resulting in an alarm for shutting down the oil tank due to an increasing gearbox temperature. In contrast, feature scores for **Seq2Seq (Att)** give a fair sense of what is leading to the gearbox alarm. However, the relevance of the features is less effective. Specifically, *SubPcsPrivRefGenSpeedInching_Mean* signifies a high speed generator inching problem, an indirect consequence leading to contact with the generator but does not hint primarily at the root cause of the problem.

5 Conclusion

We have presented a novel application scenario and preliminary results for neural data-to-text generation in the wind industry to assist engineers in understanding the context of an impending fault and potentially prevent it. We have found that transformer networks hold promise for this task and that their attention weights are closely aligned with expert judgements. We look in future to generate messages that are longer and more informative than the data we have presented. We envisage the embedding of our data-to-text component into a dialogue system to offer interactive decision support for wind turbine maintenance, contributing to the reliability and uptake of wind energy.

References

D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, A. Luccioni, T. Maharaj, E. D. Sherwin, S. K. Mukkavilli, K. P. Körding, C. Gomes, A. Y. Ng, D. Hassabis, J. C. Platt, F. Creutzig, J. Chayes, and Y. Bengio, "Tackling climate change with machine learning," *CoRR*, vol. abs/1906.05433, 2019. [Online]. Available: http://arxiv.org/abs/1906.05433

A. Zaher, S. McArthur, and D. Infield, "Online wind turbine fault detection through automated scada data analysis," *Wind Energy*, vol. 12, pp. 574–593, 2009.

I. Abdallah, V. Dertimanis, H. Mylonas, K. Tatsis, E. Chatzi, N. Dervilis, K. Worden, and E. Maguire, "Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data," in *Safety and Reliability – Safe Societies in a Changing World, Proceedings of the European Safety and Reliability Conference*, Trondheim, Norway, June 2018, pp. 3053–3061.

V. Peters, A. Ogilvie, and C. Bond, "Wind plant reliability benchmark," *Continuous Reliability Enhancement for Wind (CREW) Database, Sandia National Laboratories*, Sep 2012. [Online]. Available: https://energy.sandia.gov/wp-content//gallery/uploads/Sandia-CREW-2012-Wind-Plant-Reliability-Benchmark-Presentation.pdf

D. Milborrow, "At the tipping point: 2017 wind cost analysis," *Wind Power Monthly*, Feb 2018. [Online]. Available: https://www.windpowermonthly.com/article/1455361/tipping-point-2017-wind-cost-analysis

S. G. Sripada, E. Reiter, I. Davy, and K. Nilssen, "Lessons from deploying nlg technology for marine weather forecast text generation," in *Proceedings of the 16th European Conference on Artificial Intelligence*, ser. ECAI'04, 2004, pp. 760–764.

N. Dethlefs and A. Turner, "Deep text generation - Using hierarchical decomposition to mitigate the effect of rare data points," in *Proceedings of Language, Data and Knowledge (LDK)*, Galway, Ireland, 2017.

D. Gkatzia, O. Lemon, and V. Rieser, "Data-to-text generation improves decision-making under uncertainty," *IEEE Computational Intelligence Magazine,*, vol. 12(3), 2017.

M. MacMahon, B. Stankiewicz, and B. Kuipers, "Walk the Talk: Connecting Language Knowledge, and Action in Route Instructions," in *Proc. of National Conference on Artificial Intelligence (AAAI)*, Boston, Massachusetts, 2006.

D. Chen, J. Kim, and R. Mooney, "Training a Multilingual Sportscaster: Using Perceptual Context to Learn Language," *Journal of Artificial Intelligence Research*, vol. 37, pp. 397–435, 2010.

H. Mei, M. Bansal, and M. Walter, "What to talk about and how? selective generation using lstms with coarse-to-fine alignment," in *Proceedings of the 16th Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2016.

N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, 2002.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 5998–6008. [Online]. Available: http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://www.aclweb.org/anthology/N19-1423

T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 1412–1421. [Online]. Available: https://www.aclweb.org/anthology/D15-1166