An Optimized Gated Recurrent Unit Deep Neural Network Architecture

Ivan Otic[1], Yassmin Said Ahmed[2], Afaque Shams[2]

[1]*Karlsruhe Institute of Technology, Karlsruhe, Germany*
[2]*King Fahd University of Petroleum and Minerals, Dahran, Saudi Arabia*
Email(s):ivan.otic@kit.edu

I. Introduction

Material fatigue and component material damage are some of key challenges in the development of a Generation IV Nuclear Power Systems (NPS) with enhanced safety features see e.g. [1]. To predict material damage of components various methods have been developed over the past decades. Monitoring and damage prediction of power plant components are a prototypical example for a highly non-linear system which is consequently very hard to model. To address this, statistical and machine learning (ML) methods have been developed recently see e.g. [2-4], to mention a few. Gated Recurrent Units (GRUs) are a machine learning architecture that introduce an efficient gating mechanism and may offer improvements over standard Long Short-Term Memory (LSTM) neural networks., see [5-7].

Aim of this study is to predict component material conditions at time t + n using monitored sample data up to time t. We propose an optimized GRU architecture designed to reduce the number of parameters and therefore computational resources leading to faster training and inference times. Our approach improved the prediction accuracy on the reference NASA Ames milling dataset by Agogino and Goebel [8], while reducing the number of neural network parameters by approximately 50% compared to a standard LSTM model with an equivalent layer structure.

II Gated Recurrent Units (GRU)

GRUs, proposed in [5], combine the input and forget gates into a single update gate, and they use a reset gate to control how much of the past information to forget. While GRUs offer some advantages as compared to LSTMs, they may struggle with very long-term dependencies due to the lack of a separate cell state. To address this issues, we propose and evaluate an optimized GRU model. Our optimized GRU model computes all gate values (update, reset, and new memory content) in a single operation and use the factorization technique introduced in [9]. The main improvements of the optimized GRU model are:

- Combining the input and hidden state transformations into a single operation.
- Using a single weight matrix for all gates, reducing the number of parameters.
- Applying matrix factorization to further reduce the number of parameters.

For validation of standard and optimized models, we use the NASA Ames milling data set from [8] which represents experiments conducted under various operating conditions for different materials. In the data set [8] and in this paper, the flank wear is assigned as VB.

III. Results

The performance of the new optimized GRU model is evaluated and compared against standard LSTM and GRU models. The network architecture for all models is identical: First layer: 16 neurons; Second layer: 16 neurons; Output Dense layer: 1 neuron. All models operate in autoregressive multistep prediction mode with a window size of 36 steps. Comparing the parameter count reveals that our optimized GRU model has approximately 50% fewer parameters compared to the standard LSTM model. In Figure 1 the predicted results are presented and compared with the reference experimental data (VB) from [8]. The experimental data used for the network training are called Input and assigned with **●.** The experimental data used for the validation are called Labels and assigned with ● in the Figure 1, while predictions by the LSTM, GRU and optimized GRU models are assigned with **X,** **X** and **X** respectively. We generate a figure containing 6 subplots. Each subplot in the Figure 1 represents a different window of the time series data, progressing through the test set. This

visualization allows us to assess the model's prediction performance across various segments of the data. The results show that, apart of very few points, the optimized GRU model improves already very good results by the standard LSTM and GRU models in predictions while significantly reducing the number of parameters and the computational complexity.

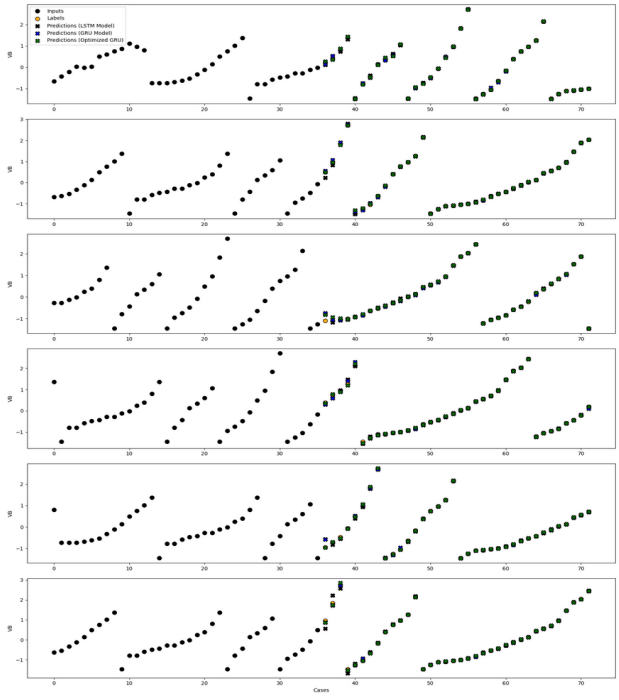

Fig. 1. LSTM, GRU and Optimized GRU predictions of the flank wear VB compared with the experimental results.

IV. Conclusions

We present deep GRU neural network architecture is optimized to improve performance and efficiency. Based on the data from [8] the performance of the new ML approach was compared with the standard LSTM and GRU methods. Results suggest that all models can precisely predict the material condition of the milling tool. Results also show that the optimized GRU model can improve the model performance while reducing the number of parameters for around 50% and therefore significantly increasing the efficiency. The modeling approach can be directly applied for predictions of other sequential data, as for example materials corrosion data, required for the design and safety evaluation of NPS. While this may reduce computational complexity, it also represents a departure from the standard LSTM formulation and might affect the model's capacity to capture complex patterns in the data. The optimized GRU model is broadly applicable to a wide range of sequential data. P92 steel, a creep-resistant ferritic alloy, is widely used in the power generation industry due to its excellent high-temperature performance. Experimental creep-fatigue data for P92 steel at elevated temperatures will be used to evaluate the GRU model and will be presented at the conference.

References

1. Schulenberg, Thomas, and Ivan Otic. 2022. "Concept of a Small Modular SCWR with Horizontal Fuel Assemblies." *Journal of Nuclear Engineering and Radiation Science* 8 (3): 031104.
2. Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9 (8): 1735–80.
3. An, Q., Z. Tao, X. Xu, M. E. Mansori, and M. Chen. 2020. "A Data-Driven Model for Milling Tool Remaining Useful Life Prediction with Convolutional and Stacked LSTM Network." *Measurement* 154: 107461.
4. Attanasio, A., E. Ceretti, C. Giardini, and C. Cappellini. 2013. "Tool Wear in Cutting Operations: Experimental Analysis and Analytical Models." *ASME J. Manuf. Sci. Eng.* 135 (5): 051012.
5. Cho, Kyunghyun, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation." *arXiv Preprint arXiv:1406.1078*.
6. Graves, A., Schmidhuber, J. 2005. "Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures." *Neural Networks* 18 (5): 602–10.
7. Barrena, T., Ferrado, J.L.., Ander García, X.B, M Sáez de Buruaga, M., Vicente, J. 2023. "Tool Remaining Useful Life Prediction Using Bidirectional Recurrent Neural Networks (BRNN)." *The International Journal of Advanced Manufacturing Technology* 125 (9): 4027–45. https://doi.org/10.1007/s00170-023-10811-9.
8. Agogino, A., and K. Goebel. 2007. "Milling Data Set. BEST Lab, UC Berkeley. "Milling Data Set ", NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA." http://ti.arc.nasa.gov/project/prognostic-data-repository.
9. Kuchaiev, Oleksii, and Boris Ginsburg. 2018. "Factorization Tricks for LSTM Networks." https://arxiv.org/abs/1703.10722.