Position-Aware Scaled Feedforward Networks: Analysis and Empirical Validation

Aardvark

October 23, 2025

Abstract

We present a systematic evaluation of Position-Aware Scaled Feedforward Networks (PASFN), analyzing its effectiveness across multiple experimental configurations. Our results show consistent improvements over baseline models while maintaining computational efficiency.

1 Introduction

Recent work has highlighted the importance of feedforward network design in transformer architectures. Our work provides comprehensive empirical analysis of position-aware scaling in feedforward layers.

2 Related Work

Our work builds upon gated feedforward networks and position-aware architectures. We compare with recent approaches including Position-Infused FFNs and Rotary FFNs.

3 Method

PASFN modifies the standard feedforward computation by introducing learned position scaling before the gated transformation. The scaling factor s has shape [1,1,d] where d is the model dimension.

4 Experiments

4.1 Setup

We evaluate on FineWeb using:

- Model sizes: 134M, 355M parameters

- Training: 640 Chinchilla-optimal steps
- Baseline: SwiGLU with identical hyperparameters

4.2 Results

Our method shows consistent improvements across model sizes while maintaining computational efficiency.

5 Limitations

Key limitations include:

- Diminishing returns at larger scales
- Requires careful learning rate tuning

6 Conclusion

PASFN shows consistent improvements in our experiments, though benefits are modest and context-dependent.

References

[1] Vaswani et al. NeurIPS 2017.

[2] Shazeer. arXiv:2002.05202.

[3] Tolstikhin et al. NeurIPS 2021.