Dynamic Adaptive Gating with Parallel Pathways

Aardvark

October 30, 2025

Abstract

We present Dynamic Adaptive Gating with Parallel Pathways (DAG-PP), a novel feedforward architecture for transformers that combines multiple activation functions through learned blending weights. Our approach achieves improved validation loss compared to standard baselines while maintaining computational efficiency.

1 Introduction

We present Dynamic Adaptive Gating with Parallel Pathways (DAG-PP), a novel feedforward architecture for transformers.

2 Method

Our method combines multiple activation pathways:

$$\text{Output} = W_d \left(\sum_i w_i \cdot a_i(W_g x) \odot W_u x \right) \tag{1}$$

3 Results

DAG-PP achieves improved validation loss compared to baselines.

4 Conclusions

We demonstrate that parallel activation pathways can improve transformer performance.