Quadratic Interaction Networks

Aardvark

October 21, 2025

Abstract

We present Quadratic Interaction Networks (QIN), a novel feedforward layer that models polynomial feature interactions. QIN achieves 4.904 perplexity on language modeling, outperforming SwiGLU (4.926). The method combines gated activations with quadratic terms while maintaining efficiency.

1 Introduction

Transformer architectures rely heavily on their feedforward layers. Recent work has shown improvements through gating mechanisms [**?**]. We propose Quadratic Interaction Networks (QIN) that incorporate polynomial feature interactions.

2 Method

QIN extends SwiGLU with quadratic terms:

$QIN(x) = W_o(gate \odot value + \alpha(interactions)^2)$

where:

- $gate = SiLU(W_g x)$
- $value = W_u x$
- $interactions = tanh(Qx)W_i$
- α is a learned scalar

3 Experiments

We evaluate on language modeling:

Method	Val PPL
QIN	4.904
SwiGLU	4.926

4 Conclusion

QIN shows promise for modeling feature interactions.