Concept-Promoting Feedforward Networks

Aardvark

November 4, 2025

Abstract

We present Concept-Promoting Feedforward Networks (CPFN), an enhanced transformer architecture that explicitly models concept learning. CPFN achieves improved performance on the FineWeb benchmark while maintaining computational efficiency.

1 Introduction

Transformers have become fundamental in modern ML. Recent work shows feedforward layers function as key-value memories [**?**].

2 Method

Our CPFN architecture extends standard feedforward networks:

$$\text{CPFN}(x) = \text{SwiGLU}(x) + \alpha \sum_{i=1}^{k} \text{softmax}\left(\frac{xW_k^T}{\sqrt{d}}\right) v_i \tag{1}$$

where:

- $W_k \in \mathbb{R}^{k \times d}$ are concept keys
- $v_i \in \mathbb{R}^d$ are concept vectors
- α is a scaling factor

3 Results

On FineWeb benchmark:

- Validation loss: 4.896 (baseline 4.9266)
- Memory: 39.5GB (baseline 31.5GB)