

Peer Review

Review of: "Superposition in Transformers: A Novel Way of Building Mixture of Experts"

Yijie Bei¹

1. Duke University, United States

Overall, this is an interesting method for merging models. However, there are some points that are unclear to me:

1. The paper mentioned that one can either hard merge two models or retain the autoencoder and b-spline. Does retaining the b-spline mean that during inference time you would need to infer on both models to get the hidden states from both models?
2. It is unclear to me why the autoencoder is necessary, and it would be great to see it in an ablation study.
3. It is unclear to me whether the presented results are done on the hard merged model or the alternative method mentioned. I imagine the quality difference could be significant.
4. It is unclear to me how this method compares to existing methods (e.g., those available through mergekit).

Declarations

Potential competing interests: No potential competing interests to declare.