22 April 2025 · CC-BY 4.0 Qeios

Peer Review

Review of: "Predictive Coding Explains Asymmetric Connectivity in the Brain: A Neural Network Study"

Abhranil Das[1]

1. Psychology, University of Texas at Austin, United States

Summary of the paper

The brain uses top-down predictions to make sense of noisy or incomplete sensory input. But anatomical studies show that feedback in the cortex is often asymmetric, i.e. it doesn't neatly mirror the feedforward pathways. Why would the brain be wired this way, and does this kind of asymmetry actually help with sensorimotor tasks?

- The authors used two CNNs, long and short, trained to predict grasp locations on objects using input RGB + depth images.
- The network output 6 maps for grasp parameters, including grasp quality.
- They added predictive feedback connections to these CNNs, mimicking how the brain might use top-down signals to clean up noisy input.
- At each time step, deeper layers try to predict and correct errors in earlier layers.

With this set-up they explored the questions:

- does predictive feedback help with noisy input?
- does it matter where in the network feedback comes from or goes to?
- does the distance of the feedback connection (short, medium, long-range) affect performance?
- does it help more if feedback comes from high-level or low-level parts of the model?

Experiment 1: does feedback help with noisy input?

- With no noise: feedback didn't help (sometimes even slightly worsened performance).

 doi.org/10.32388/GDIMB5

- With medium or high noise: feedback clearly improved performance. The model's grasp predictions got better over time with feedback, especially as noise increased.

Experiment 2: varying feedback distance

- Kept the target layer constant (early in the network).
- Found that medium-range feedback improved performance the most.
- Short-range and long-range were less effective, or only worked under specific noise conditions.

Experiment 3: varying feedback target layer

- This reversed the design: now feedback always comes from the same place (deeper in the network), but targets different layers.
- Again, medium-range connections worked best overall.

Experiment 4: Varying abstraction level of the feedback source layer

- Feedback is now of the same length, but originates from an early, intermediate or deep layer
- Feedback from early layers worked best.
- This contradicts some previous studies where deeper layers were more helpful, but those used attention-based feedback, not predictive coding.

Conclusions

- Feedback is most effective when it's not too close (lacks context) or too far (too abstract), and comes from an early-ish layer.
- This matches with brain areas like V6A, which sit mid-way in the dorsal stream and are known to integrate both vision and motor planning.

Feedback

The following are my main critiques and questions on this paper:

- The language of the entire paper could be made more simple and lucid, while conveying the same information. This is especially true of the abstract.
- Page 4: '*to facilitate interpolation... from the finite set of ground truth labels*': it is unclear what this means.

doi.org/10.32388/GDIMB5

- Tables 1-2 could probably be moved out of the flow of the main paper, into an appendix and supplementary info.
- Why did you choose a short-backbone and a long-backbone CNN model?
- In the text, only the experimental setups are described one after the other first, but no results. The results are in a different section. But the accompanying figures show the network architectures and results together. This is a bit of a confusing organization. In the text, it may be better to have each experiment described together with its results.
- This paper describes predictive coding for a passive or static task (generating grasp parameters for an input image). Real-world grasping is a dynamic process that involves real-time predictive coding as part of a real-time sensorimotor loop. The role and outcome of predictive coding in such a real-time loop could be expected to be quite different from what it is on this passive, static task. Some discussion about how the findings here would relate to real-world dynamic tasks would be helpful.
- The different red arrows of fig. 2 etc., indicating short/medium/long feedback, are hard to tell apart.
- Numeric tables such as table 3 are hard to parse and make sense of, compared to say, figures.
- In the discussion: *'too short a feedback loop may lack the broader contextual information needed to mitigate noise effectively, while longer loops may carry overly abstract predictions.'* This is an important point that the paper is speculating on. This paper would become a lot more important if it undertook some more investigations to verify if this is indeed the case.
- In the discussion: *'medium-range feedback benefits noise clean-up the most when it originates from layers that are proximal to the input, suggesting that relatively early representations are most effective in removing noise.'* Any interpretations of why early representations would be most effective in removing noise?
- In the discussion, it says that the U-shape may emerge in the long backbone CNN if fed with higher noise images. Why don't you run this experiment to know for sure, if it's not too much additional work?

Declarations

Potential competing interests: No potential competing interests to declare.