Predictive Coding algorithms induce brain-like responses in Artificial Neural Networks

Anonymous authors
Double blind review

Abstract

This study investigates whether predictive coding (PC) inspired deep neural networks can serve as biologically plausible models of the brain. We compared two PC-inspired training objectives - a predictive and a contrastive approach - to a supervised baseline, using a simple recurrent neural network (RNN) architecture. Our results show that, compared to Supervised or Untrained models, the PC-inspired models exhibited more key signatures of PC. This includes mismatch responses (MMR), formation of prior expectations, and learning of semantic representations. These findings indicate that PC-inspired models can capture important computational principles of predictive processing in the brain, and serve as a promising foundation for building biologically plausible artificial neural networks.

Keywords: predictive coding; deep neural networks; brain modeling; mismatch response; prior expectations; semantic representations

Introduction

This study investigates whether predictive coding (PC) inspired deep neural networks can serve as biologically plausible models of the brain. Current deep neural network (DNN) models, while powerful (Richards et al., 2019; Schrimpf et al., 2020), are not well-aligned with our understanding of biological neural networks (Albada et al., 2022; Bengio et al., 2016; Kietzmann et al., 2018; Pulvermüller et al., 2021; Salvatori et al., 2023; Stork, 1989). PC provides a promising framework for building deep neural networks that capture important computational principles of the brain (Hinton, 2022; Lotter et al., 2020; Millidge et al., 2020, 2023; Whittington & Bogacz, 2017).

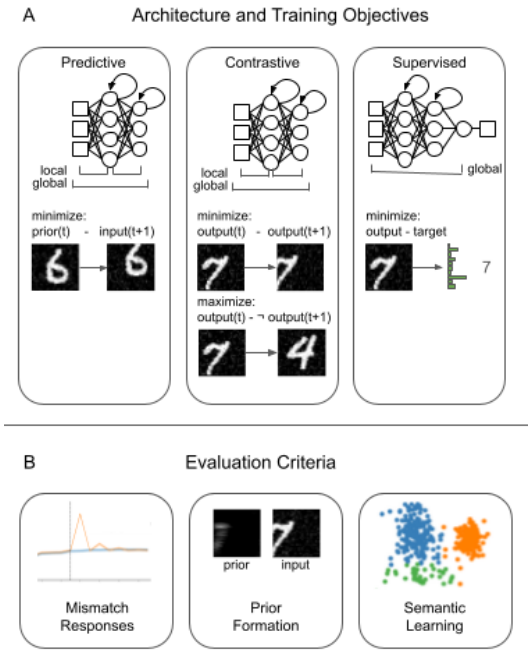

Figure 1: *A. Network Conditions: Illustration of the RNN architecture used and optimization objectives for the Predictive, Contrastive, and Supervised conditions. The Predictive and Contrastive objectives can be trained locally, layer-by-layer, unlike the global backpropagation used for the Supervised objective. B. Evaluation Criteria: The dependent variables used to assess hallmarks of predictive coding: mismatch responses (MMR), prior expectations, and learned representations.*

Methods

We compared two PC-inspired training objectives: a predictive approach (inspired by Lotter et al., 2020) where the network tries to predict the future input, and a contrastive approach where expected future stimuli are contrasted from unexpected ones (inspired by Hinton, 2022). We implemented both approaches in a simple recurrent neural network

(RNN) architecture and compared them to a typical Supervised and an Untrained baseline. Each learning condition was trained on moving image series based on the MNIST dataset. We evaluated the models on key signatures of PC, including the generation of MMR (deviations in neural activity to unexpected stimuli), formation of prior expectations (the network forming accurate priors of future inputs), and the learning of semantic information (the network's capacity to encode complex stimulus information without explicit supervision). Each metric was evaluated using independent-sample T or Z-tests on an unseen test dataset of 6000 image sequences. This allowed us to assess the extent to which these models exhibit behaviors that align with predictive processing (see Figure 1).

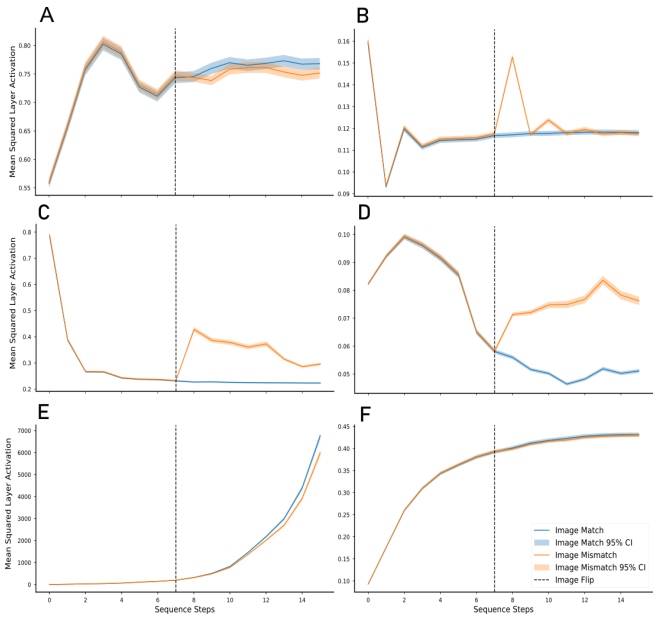

Figure 2: *The mismatch responses (MMR) for the main network conditions: A. Predictive global; B. Predictive local; C. Contrastive global; D. Contrastive local; E. Supervised; F. Untrained. Each subplot depicts the mean squared average activation over time, separated for expected and unexpected input changes. The shaded areas represent 95% confidence intervals.*

Results

Our results show that the PC-inspired models, especially the locally trained predictive model, exhibited these PC-like behaviors better than the supervised or untrained models. For MMR, the Contrastive global (T_{11998} =-41.58, p<1e-5), Contrastive local (T_{11998}=-30.00, p<1e-5), and Predictive local (T_{11998}=-40.37, p<1e-5) conditions showed significant deviations, while Predictive global, Supervised, and Untrained did not (see Figure 2). In prior formation, Predictive global (r=0.81, Z_{11998}=61.08, p<1e-5) and Predictive local (r=0.40, Z_{11998}=23.05, p<1e-5) exhibited strong correlations between prior and next input, unlike other conditions. For learning, all models performed significantly better than an Untrained network. Predictive global had the highest accuracy (0.39, T_{11998}=7.58, p<1e-5), followed by Supervised (0.37, T_{11998}=5.31, p<1e-5), Contrastive global (0.37, T_{11998}=5.18, p<1e-5), Contrastive local (0.35, T_{11998}=3.65, p=0.00389), and Predictive local (0.35, T_{11998}=3.06, p=0.03331).

Discussion

The findings suggest that PC-inspired models, especially the locally predictive network, showed the most biologically plausible PC-like behavior, including clear priors, robust MMR, and meaningful semantic representations. Contrastive models exhibited strong MMR but lacked explicit priors, while Supervised and Untrained controls did not show clear MMR nor prior expectations. These results confirm that simple PC-based objectives can yield models that not only perform well in unsupervised tasks but also align closely with theoretical and empirical hallmarks of biological PC (Garrido et al., 2009; Hodson et al., 2024). This work contributes to our understanding of the relationship between artificial and biological neural networks, and highlights the potential of PC-inspired algorithms for advancing brain modeling as well as brain-inspired machine learning (Friston, 2018; Millidge et al., 2022; Spratling, 2017).

References

Albada, S. J. van, Morales-Gregorio, A., Dickscheid,

T., Goulas, A., Bakker, R., Bludau, S., Palm, G., Hilgetag, C.-C., & Diesmann, M. (2022). *Bringing Anatomical Information into Neuronal Network Models* (Vol. 1359, pp. 201–234). https://doi.org/10.1007/978-3-030-89439-9_9

Bengio, Y., Lee, D.-H., Bornschein, J., Mesnard, T., & Lin, Z. (2016). *Towards Biologically Plausible Deep Learning* (arXiv:1502.04156). arXiv. https://doi.org/10.48550/arXiv.1502.04156

Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, *21*(8), 1019–1021. https://doi.org/10.1038/s41593-018-0200-7

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Hinton, G. (2022). *The Forward-Forward Algorithm: Some Preliminary Investigations* (arXiv:2212.13345). arXiv. https://doi.org/10.48550/arXiv.2212.13345

Hodson, R., Mehta, M., & Smith, R. (2024). The empirical status of predictive coding and active inference. *Neuroscience & Biobehavioral Reviews*, *157*, 105473. https://doi.org/10.1016/j.neubiorev.2023.105473

Kietzmann, T. C., McClure, P., & Kriegeskorte, N. (2018). *Deep Neural Networks in Computational Neuroscience* (p. 133504). bioRxiv. https://doi.org/10.1101/133504

Lotter, W., Kreiman, G., & Cox, D. (2020). A neural network trained for prediction mimics diverse features of biological neurons and perception. *Nature Machine Intelligence*, *2*(4), 210–219. https://doi.org/10.1038/s42256-020-0170-9

Millidge, B., Salvatori, T., Song, Y., Bogacz, R., & Lukasiewicz, T. (2022). *Predictive Coding: Towards a Future of Deep Learning beyond Backpropagation?* (arXiv:2202.09467). arXiv. https://doi.org/10.48550/arXiv.2202.09467

Millidge, B., Tang, M., Osanlouy, M., & Bogacz, R. (2023). *Predictive Coding Networks for Temporal Prediction* (p. 2023.05.15.540906). bioRxiv. https://doi.org/10.1101/2023.05.15.540906

Millidge, B., Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). *Activation Relaxation: A Local Dynamical Approximation to Backpropagation in the Brain* (arXiv:2009.05359). arXiv. https://doi.org/10.48550/arXiv.2009.05359

Pulvermüller, F., Tomasello, R., Henningsen-Schomers, M. R., & Wennekers, T. (2021). Biological constraints on neural network models of cognitive function. *Nature Reviews Neuroscience*, *22*(8), 488–502. https://doi.org/10.1038/s41583-021-00473-5

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., … Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*, *22*(11), 1761–1770. https://doi.org/10.1038/s41593-019-0520-2

Salvatori, T., Mali, A., Buckley, C. L., Lukasiewicz, T., Rao, R. P. N., Friston, K., & Ororbia, A. (2023). *Brain-Inspired Computational Intelligence via Predictive Coding* (arXiv:2308.07870). arXiv. https://doi.org/10.48550/arXiv.2308.07870

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., Kar, K., Bashivan, P., Prescott-Roy, J., Geiger, F., Schmidt, K., Yamins, D. L. K., & DiCarlo, J. J. (2020). *Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?* (p. 407007). bioRxiv. https://doi.org/10.1101/407007

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92–97. https://doi.org/10.1016/j.bandc.2015.11.003

Stork. (1989). Is backpropagation biologically plausible? *International 1989 Joint Conference on Neural Networks*, 241–246 vol.2. https://doi.org/10.1109/IJCNN.1989.118705

Whittington, J. C. R., & Bogacz, R. (2017). An Approximation of the Error Backpropagation Algorithm in a Predictive Coding Network with Local Hebbian Synaptic Plasticity. *Neural Computation*, *29*(5), 1229–1262. https://doi.org/10.1162/NECO_a_00949