Miniproject

NX-414 - Brain Like Computing and Intelligence

Clémentine NAÏM
Maria YUFFA

Introduction This report builds on research demonstrating that weighted sums of neuronal firing rates in the inferior temporal cortex predict core object recognition performance in humans [1]. We extend this study by comparing task-driven and data-driven modeling approaches to predict average neural firing rates using image datasets from both nonhuman primates and humans.

Methods and Models Our approach encompasses both linear (including ridge regression and PCA for dimensionality reduction) and nonlinear models (utilizing Convolutional Neural Networks). We optimized our models using cross-validation and grid-search techniques to find the best parameters. Specifically, for the task-driven model, we used PCAs of activations from different layers of the pre-trained ResNet50 model, which showed the highest explained variance. For the data-driven approach, we implemented a CNN with variable layers and batch sizes, further tuning parameters such as learning rate and weight decay for optimal performance. The key metrics for model evaluation were explained variance, correlation between predicted and actual neural activities, and Mean Squared Error (MSE). Note, that we took the mean value of each metric for each neuron.

Experimentation We experimented with various models including using PCA activations from VGG19 (features.28 and classifier.3, this choice reflects the belief that middle layers perform the best) and co-training with neural activations and object imagery by using pre-trained ResNet50. However, these methods did not improve results. The co-trained model had very low variation with very stochastic loss over epochs, meaning that the model struggled to fit complex behaviour and required more training steps (trained on 5) and, hence, more computational resources. Performing experimentation with CNN, we optimized CNN models by varying layers and batch sizes, and fine-tuning learning rates and weight decay for L2 regularization. The final model we obtained was a 3 layers CNN with a 32 batch size, a learning rate of 0.001 and a weight decay of 0.001. But the results we obtained were not great according to the explained variance metric (0.2100), even with the low loss (0.1076).

Results and Conclusion Overall, the best performing model is the Ridge, trained on layer 3 PCA activations of pre-trained ResNet50, with a mean explained variance of 0.406 and correlation of 0.628 on validation set.

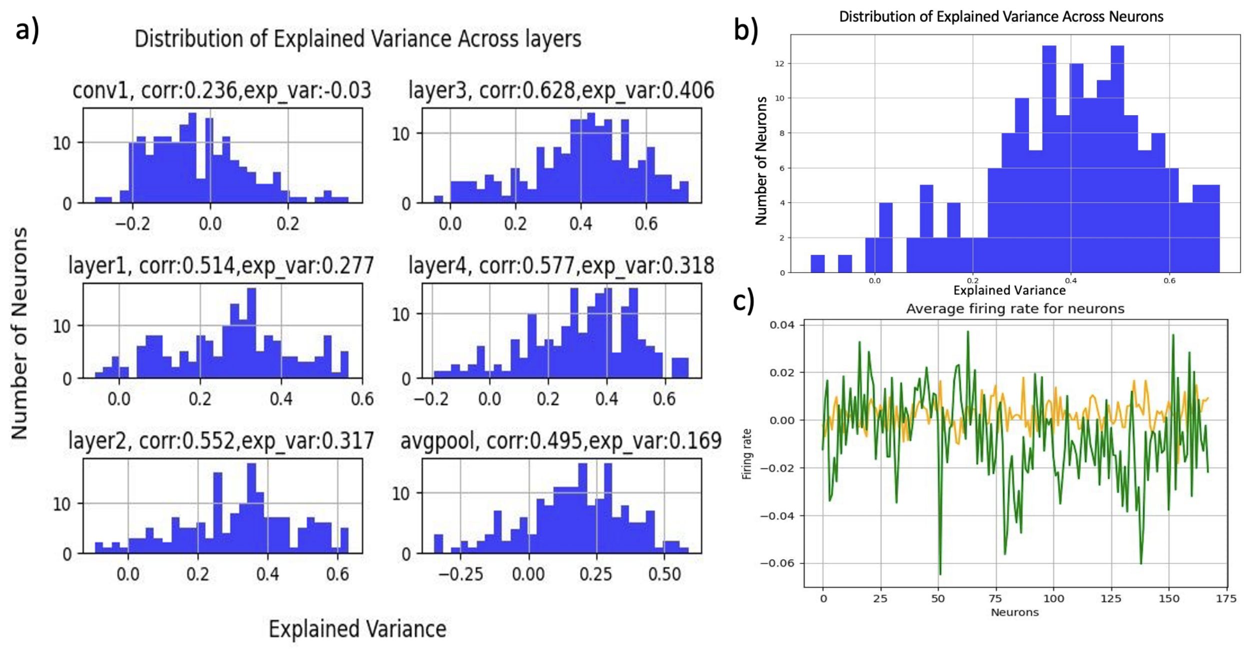

Figure 1: a) Explained Variance across layers for neural activity predicted by ResNet50 model; b) Distribution of Explained Variance across neurons trained on the PCAs of activations extracted from feauture.28 layer of VGG19; c) Average target vs predicted firing rates from co-trained neural network.

Model	Strategy	Explained variance	Correlation	MSE loss
Linear Regression	—	-1.1615	0.1465	0.2857
Ridge Regression	—	-1.1574	0.1467	0.2852
Ridge Regression	Cross-validation	-0.3913	0.1718	0.1894
ResNet50 (pre-trained)	PCA of activations (best)	0.4059	0.6923	0.0733
ResNet50 (random)	PCA of activations (best)	0.1562	0.4184	0.1132
VGG19 (pre-trained)	PCA of activations (features.28)	0.3955	0.6243	—
CNN (2 layers)	—	0.1701	—	0.1468
CNN (3 layers)	—	0.2100	—	0.1076
Co-trained Neural Network	using pre-trained ResNet50	0.0044	0.0551	0.1398

Table 1: Performance of different models.

References

[1] Najib J. Majaj et al. "Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance". In: *The Journal of Neuroscience* 35.39 (2015), pp. 13402–13418. DOI: 10.1523/JNEUROSCI.5181-14.2015. URL: https://doi.org/10.1523/JNEUROSCI.5181-14.2015.