A Study on a Correlation between a Predictive Model of Motion Pictures Imitating the Predictive Coding of the Cerebral Cortex and Brain Activity

Chihiro Fujiyama[1]**(fujiyama.chihiro@is.ocha.ac.jp), Shinji Nishimoto**[234] **(nishimoto@nict.go.jp),**

Satoshi Nishida[23]**, (s-nishida@nict.go.jp) Hideki Asoh**[5]**(h.asoh@aist.go.jp), Ichiro Kobayashi**[1] **(koba@is.ocha.ac.jp)**

[1]Graduate School of Humanities and Sciences, Ochanomizu University, Otsuka2-1-1, Bunkyo, Tokyo 112-8610, Japan

[2] Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), Yamadaoka 1-4, Suita, Osaka 565-0871, Japan

[3]Graduate School of Frontier Biosciences, Osaka University, Yamadaoka 1-3, Suita 565-0871, Japan

[4]Graduate School of Medicine, Osaka University, Yamadaoka 2-2, Suita 565-0871, Japan

[5]National Institute of Advanced Industrial Science and Technology (AIST), Aomi 2-3-26, Koto, Tokyo 135-0064 Japan

Abstract

In recent years, deep neural networks inspired by the notion of predictive coding have been shown to make accurate predictions of future frames. In this study, we focus on a predictive neural network, one of such implementations to evaluate the relationship between natural and artificial neural networks. By using PredNet, a predictive neural architecture, we show that representations extracted from the architecture are correlative with brain activities evoked by natural movie stimuli. Our result gives a verification result on the theoretical hypothesis of predictive coding.

Keywords: predictive coding; natural stimuli; fMRI; decoding; deep neural networks

Introduction

Predictive coding is a theoretical hypothesis which posits that the brain continually predicts incoming sensory stimuli (Rao & Ballard, 1999; Friston, 2005). In recent years, great strides have been made in exploring deep neural networks with architecture explicitly inspired by the notion of predictive coding, and they have been shown to accurately predict future frames. In this study, we focus on PredNet (Lotter, Kreiman, & Cox, 2016), one of such predictive neural architectures and evaluate the correlation between representations acquired by the architecture and human brain activities evoked by natural movie stimuli. This is an elemental approach with the aim of making a suggestion about whether predictive coding functions in human brain. Moreover, we attempt to generate images by using representations estimated from brain activities.

PredNet

PredNet (Lotter et al., 2016) is a deep predictive neural network constructed by imitating the concept of predictive coding. Given a current frame, it predicts a future frame. PredNet consists of stacked layers of the same architectural modules. Each module has four parts: a convolutional input layer, a recurrent convolutional (SHI et al., 2015) representation layer, a convolutional prediction layer, and an error representation. In each module, a representation layer holds history of past frames to generate a prediction and a prediction layer generates a prediction of what the module input will be on the next frame. An error representation is computed by comparison of the generated prediction and the output of an input layer which processes the input into the module and it is then passed to the next module as well as a representation layer in the same module. In PredNet, predictions generated in upper modules are passed to lower modules, and error representations computed in lower modules are passed to upper modules. The outline of PredNet is illustrated in Figure 1.
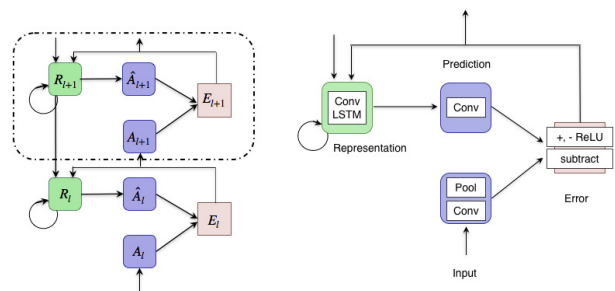

Figure 1: PredNet model. Left: Information flow within two layers. Each module consists of an input layer (A_l), a representation layer (R_l), a prediction layer ($\hat{A}_l$), and an error representation (E_l). Right: Implementation of each module for video prediction.

Experiments

We investigate into the correlation between representations acquired at the representation layers in PredNet and human brain activities evoked by natural movie stimuli. First, we train PredNet with image sequences extracted from natural movies. Next, we train ridge regressions in which brain activity data is the explanatory variable and each representation is the objective variable. Finally, we evaluate the correlation between actual representations and representations estimated by using the regressions. The relationship between brain activities and representations in PredNet is illustrated in Figure 2

Experimental Setup

We trained PredNet on the same natural movies as ones used in (Nishimoto et al., 2011) and preprocessed the movies by extracting static image sequences with 10fps and downsampling to 160 × 120 pixels. We followed the settings shown in
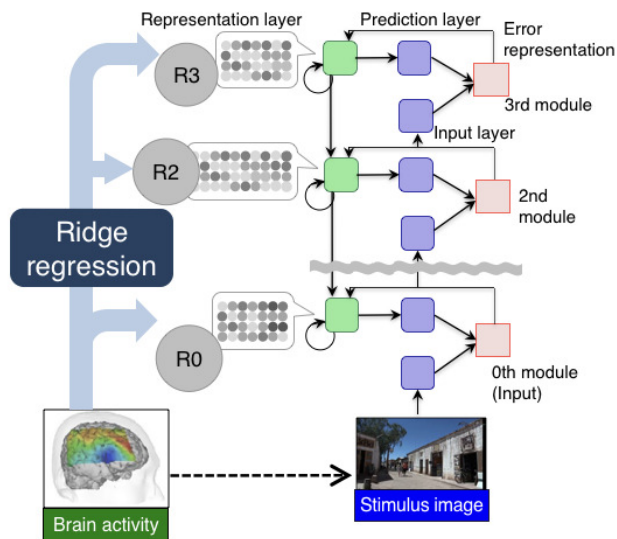
Figure 2: Relationship between representations in PredNet and brain activities.

Table 1 for training PredNet. The model was trained to minimize the sum of mean squared error between actual and predicted frames over ten time steps.

Table 1: Architectural configurations of PredNet.

# Layers	4
Size of convolutional filter	3×3 (for all convolutions)
# Channels	From lower module, 3, 48, 96, 192
Optimization algorithm	Adam (Kingma & Ba, 2014) (α=0.001, β_1=0.9, β_2=0.999)
learning rate decay	After the midpoint α=0.0001

As representations acquired by PredNet, we used representations in representation layers in zeroth (input), second, and third modules. In this paper, we refer to each of them as R0, R2, and R3. As for the first hidden layer R1, we have not applied ridge regression to that, because of the problem of considerably high dimensions for the computation resources – it has 230,400 dimensions. As the brain activity data we used in the experiments, we employed the brain activity data of subjects stimulated by natural movies, the BOLD signal observed by functional magnetic resonance imaging (fMRI), which has 65,665 voxels corresponding to the cerebral cortex part among all 96×96×72 observed voxels. We trained ridge regressions on 4,497 representation-brain activity pairs, and evaluated it on 300 pairs.

Correlation between representations and brain activities

Table 2 shows correlation coefficients between the actual representations acquired during a prediction task and the corresponding representations of R0, R2, and R3, estimated

from brain activities by using ridge regressions, respectively. As λ, a weight for regularization term in ridge regressions, we applied $\left[0.50, 1.0, 5.0, 10.0, 1.0 \times 10^2, 1.0 \times 10^3, 1.0 \times 10^4, 2.5 \times 10^4, 5.0 \times 10^4, 1.0 \times 10^5, 1.0 \times 10^6, 1.0 \times 10^7 \right]$.

The correlation coefficient between R0 estimated from brain activities and actual R0 is approximately 0.31. On the other hand, correlation coefficients at R2 and R3, which are in upper modules, are much lower.

Generating images from brain activities

We attempted to generate images using representations of R0, R2, and R3 which are estimated from brain activities. Figure 3 shows examples of generated images. For comparison, we additionally display examples of images generated from representations replaced with 0, or random values [-1.0, 1.0].
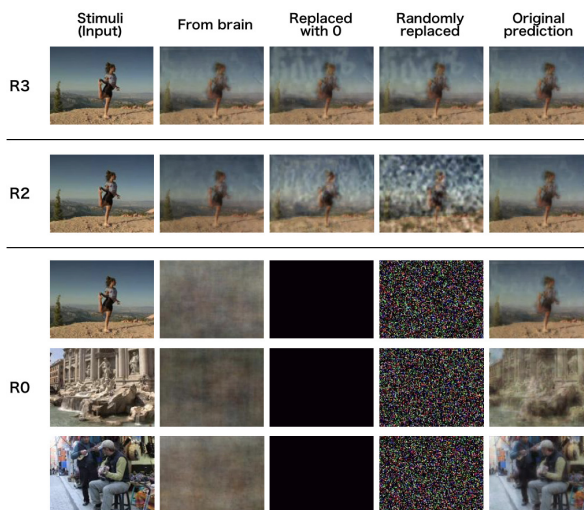

Figure 3: Examples of images generated from various representations. From left, stimulus (input) image at the time step, images generated using representations estimated from brain activities with ridge regressions, images generated from representations replaced with 0, images generated from representations replaced with random values [-1.0, 1.0], original predicted images generated during a plain prediction task.

As for R3, in the case of replacing the values of the representations with either 0 or random values, we see that almost similar image is generated in both cases. Even in the case of giving different input to the representation layer, there is not so big difference in the generation result, therefore, we do not assume that the representation layer of the third module functions well in the PredNet. On the other hand, in the case of replacing the values of R3 with the values estimated from the brain activity data, compared to the cases of replacing with 0s and random values, we see that images similar to the ones originally generated by PredNet are generated. Though the images generated by replacing the feature values of R3 with 0s and random values are similar to each other, we have actually confirmed that there is slight difference between them. It

Table 2: Correlation between the actual representations and the representations estimated from brain activities.

Representations	λ	\\\\	Correlation coefficient										
			0.5	1.0	5.0	10.0	100.0	1K	10K	25K	50K	100K	1M	10M
R0			0.2490	0.2491	0.2495	0.2499	0.2570	0.2868	0.3125	**0.3174**	0.3152	0.3055	0.2384	0.1652
R2			0.06695	-	-	-	-	-	-	0.1448	0.1562	0.1662	**0.1804**	0.1671
R3			0.06984	0.06986	0.06998	0.07013	0.07253	0.08299	0.09355	0.09867	0.1022	**0.1040**	0.09288	0.07160

can be thought that the images generated with the estimated values becomes similar to the ones generated by PredNet because of the parameters which makes them slightly different. However, overall, we think that R3 does not influence a generated image even if replacing its feature values. In other words, the information from R3 does not contribute significantly for image generation in the framework for PredNet. As for R2, in the case of replacing the feature values with estimated values, compared to the cases with 0s and random values, we see that the images similar to the ones generated with PredNet are generated. As for R0, because it does not directly receive the information for correction from error signals, an image is generated with only the estimated feature values from brain activity data, therefore, the generated images become less visible. In each layer of any of R0, R2, and R3, compared to the images generated by replacing its feature values with either 0s or random values, we see that the generated images with estimated values from brain activity data tend to be similar to the ones generated by PredNet, furthermore, by the fact that this phenomenon is getting remarkable as the layers get shallow from R3 to R0, we also see that the contribution of the feature values of each layer for generating an image gets higher as the layer gets lower. Furthermore, Figure 4 shows examples of images generated from the feature values of R0 estimated by using brain activity data which were used for training a ridge regression. Compared to the images generated with the brain activity data for evaluation, the images generated with the data for training are similar to the ones by PredNet. This leads to the fact that using ridge regression is appropriate to some extent as a model to train the correspondence relation between the data and the feature values of hidden layers of PredNet. On the other hand, because the difference between visibility of generated images with the brain activity data for evaluation and the ones with the data for training is remarkable, we can say that there is still space to improve the regression model as a generalization model. The reason why it was difficult to train the correspondence relation between brain activity data and the feature values of each layer as a generalization model with high accuracy would be that the dimensions of each representation is considerably high as R0, R2, R3 are 57,600, 115,200, 57,600, respectively.

Conclusion

In this study, as an attempt to quantitatively evaluate the correlation between the brain activities and the behavior of PredNet which is a prediction model developed with deep neural networks, we have implemented PredNet (Lotter et al., 2016) and
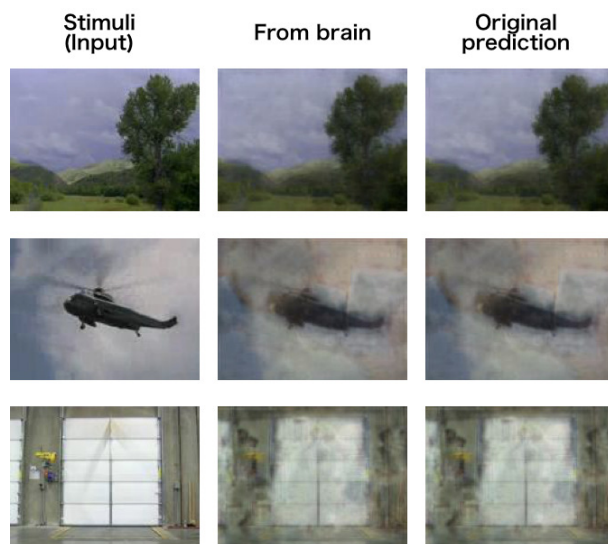

Figure 4: Examples of images generated with brain activity data for training. Left: stimulus (input) image at the time step. Middle: images generated using representations estimated from brain activities in training data for ridge regressions. Right: original predicted images generated by PredNet.

investigated the correlation between brain activity data and the representation values at each hidden layer of PredNet. Furthermore, we have evaluated the functions of each representation layer of PredNet by generating images with the estimated feature values of each module of PredNet from brain activity data. Through these experiments, as a result, we have confirmed that there is significant correlation between human brain activity data and the feature values of R0 in PredNet. This fact may be an evidence that there is natural prediction activity in the human brain. As future work, we will work on raising the prediction accuracy of both PredNet and ridge regression. Moreover, in order to estimate the corresponding relation between brain activity and representation by a more general model with high accuracy, we would like to work on reducing the dimensions of representations of each hidden layer in PredNet by means of an auto-encoder, and dealing with brain activity considering the region of interest of the cerebral cortex.

References

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815–836.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lotter, W., Kreiman, G., & Cox, D. D. (2016). Deep predictive coding networks for video prediction and unsupervised learning. *CoRR*, *abs/1605.08104*. Retrieved from http://arxiv.org/abs/1605.08104

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, *21*(19), 1641–1646.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extraclassical receptive-field effects. *Nature neuroscience*, *2*(1), 79.

SHI, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., & WOO, W.-c. (2015). Convolutional lstm network: A machine learning approach for precipitation nowcasting. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems 28* (pp. 802–810). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/5955-convolutional
-lstm-network-a-machine-learning-approach
-for-precipitation-nowcasting.pdf