Deep Learning for Brain Encoding and Decoding

Subba Reddy Oota[1,5,6], **Jashn Arora**[2], **Manish Gupta**[2,3], **Raju S. Bapi**[2], **Mariya Toneva**[4]
[1]Inria Bordeaux, France, [2]IIIT Hyderabad, India, [3]Microsoft, Hyderabad, India, [4]Princeton Neuroscience Institute, USA
[5]LaBRI - Laboratoire Bordelais de Recherche en Informatique, [6]IMN - Institut des Maladies Neurodégénératives Bordeaux, France
subba-reddy.oota@inria.fr, jashn.arora@research.iiit.ac.in,
gmanish@microsoft.com, raju.bapi@iiit.ac.in, mt6949@princeton.edu

Keywords: brain encoding; brain decoding; cognitive neuroscience; deep learning.

Introduction

How does the brain represent different modes of information? Can we design a system that can automatically understand what the user is thinking? We can make progress towards answering such questions by studying brain recordings from devices such as functional magnetic resonance imaging (fMRI). The brain encoding problem aims to automatically generate fMRI brain representations given a stimulus. The brain decoding problem is the inverse problem of reconstructing the stimuli given the fMRI brain representation. Both the brain encoding and decoding problems have been studied in detail in the past two decades and the foremost attraction of studying these solutions is that they serve as additional tools for basic research in cognitive science and cognitive neuroscience. Recently, inspired by the effectiveness of deep learning models for natural language processing and computer vision, such models have been applied for neuroscience as well. In this tutorial, we plan to discuss different kinds of stimulus representations, and popular encoding and decoding architectures in detail. The tutorial will provide a working knowledge of the state of the art methods for encoding and decoding, a thorough understanding of the literature, and a better understanding of the benefits and limitations of encoding/decoding with deep learning.

Encoding models that accurately predict brain activity have several practical applications in evaluation and diagnosis of neurological conditions and thus also help designing therapies for brain damage. Invertible encoding models enable principled formulation of brain decoding models which in turn are useful for designing brain-machine or brain-computer interfaces. Recent advances in the use of pretrained deep network models enable us to use them as priors for brain decoding tasks. Deep learning models are useful for improving accuracy but also offer the flexibility of decoding across a gamut of tasks and domains.

Tutorial outline

We present the overall tutorial outline with time slots (adding up to 3 hours) in this section.

Intro to Brain encoding and decoding [30 min]

- Intro to Brain Encoding/Decoding and applications
- Intro to popular datasets: Text Words, Sentences, Paragraphs (Harry Potter Story, ZUCO EEG, Question-Answering MEG); Visual (Binary visual patterns, Natural Images (Vim-1), BOLD5000, Algonauts and SS-fMRI); Videos (BBC's Doctor Who, Japanese Ads, Pippi Langkous, Algonauts); Multimodal (Words + line drawing of concept named by each word, The Moth Radio Hour (Jain & Huth, 2018), Pereira, Narratives).

Stimulus Representations [30 min]

- Text Stimuli: Syntactic (corpus co-occurrence counts, topic models, syntactic features and discourse features), Semantic (word embedding methods, sentence representation models, recurrent neural networks and Transformer methods), Experiential attributes (Rated on 0-6 scale or binary).
- Visual Stimuli: Visual field filter banks, Gabor wavelet pyramid; Convolutional neural networks; Concept recognition models.
- Audio Stimuli: Phoneme rate and presence of phonemes.

Brain Encoding [60 min]

- Linguistic Encoding in the Brain: Words represented as vectors in an embedding space and mapped to neural activation; Words augmented with context using sequence models; Sentence representations based on Transformer models (Schrimpf et al., 2021); Modulation of word meaning representations based on task settings (Toneva, Stretcu, Póczos, Wehbe, & Mitchell, 2020); Disentangling lexical, syntactic, and semantic representations (Caucheteux, Gramfort, & King, 2021).
- Visual Encoding in the Brain: Representations derived from object classification models (Kubilius et al., 2019); Representations extracted from models of computer vision tasks (Wang, Tarr, & Wehbe, 2019); Representation of temporal dynamics of vision.

Brain Decoding [60 min]

- Decoder Architectures: Ridge regression models; Most informative voxels selection; Fully connected layer, MLPs; Representational similarity analysis; Classifiers like SVMs and Gaussian Naïve Bayes.
- Decoding task settings: Decoding to a vector representation using single-mode stimuli or multi-modal stimuli (Pereira et al., 2018); Relationship between various natural language understanding tasks and the brain

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

decoding task (Gauthier & Levy, 2019); Reconstruct the visual stimuli (Beliy et al., 2019); Reconstructing personal imagined experiences; Application-based decoding like predicting individuals' engineering exam results, reflecting whether current thoughts are detailed, correspond to the past or future, are verbal or in images, object recognition.

- Metrics: 2 versus 2 (2V2) Accuracy; Pearson Correlation; RSM; R^2 score, MSE; Pairwise Accuracy; Rank based Accuracy; Classification rate.

Target audience and prerequisites

The tutorial should be beneficial for researchers from both academia and industry. Specifically, academics with interests in cognitive modeling and brain mapping will benefit from the broad perspective of how deep learning methods can be applied for these purposes. Researchers from companies who innovate and design novel interfaces and hardware subsystems for the next generation of virtual, augmented, and mixed reality experiences will find the tutorial exciting. Researchers working on innovative user interfaces like wearable devices, neuro-development of emotion processing and regulation, and human perception are also expected to find the tutorial relevant. Expected pre-requisites include basic maths, machine learning and basic deep learning concepts.

Presenter experience and interests

The tutorial presenters represent great diversity with respect to academic as well as industry affiliations, multi-geography and different career stages. We hope that the tutorial will attract both industry as well as academic participation.

Subba Reddy Oota is a PhD Student at Inria Bordeaux, France. He received his Masters from IIIT Hyderabad in 2016. His research interests are in the areas of language analysis in the brain, brain encoding, and decoding. He has presented several research papers in refereed conferences like NAACL, Cogsci, WACV, IJCNN, ICDAR, and ICONIP.

Jashn Arora is an Undergraduate Student at IIIT Hyderabad. His research interests are in the areas of Natural language Processing, analysis of language representation in the brain, and brain encoding and decoding. He has experience conducting tutorials and lab sessions as a teaching assistant for various courses at IIIT Hyderabad.

Manish Gupta is a Principal Applied Researcher at Microsoft, India. He is also an Adjunct Faculty at IIIT Hyderabad and a visiting faculty at the Indian School of Business (ISB). He received his Masters from IIT Bombay in 2007 and his Ph.D. from UIUC in 2013. His research interests are in the areas of deep learning, web mining, and neuroscience. He has extensive experience offering tutorials at top conferences like CIKM'13, WWW'14, SIGIR'15, CIKM'20, IJCAI'20, ECML/PKDD'21, and WSDM'22.

Raju S. Bapi is a professor and head of the Cognitive Science Lab, IIIT Hyderabad. Earlier, he was a professor at Univ. of Hyderabad, India; EPSRC Research Fellow at Univ. of Plymouth, UK; and a Researcher at ATR Research Labs, Kyoto, Japan. He has over 20 years of teaching and research experience in AI, Machine Learning, Neural Networks, and Cognitive Science. He has a PhD (Computer Science) from Univ. of Texas, Arlington. He is a senior member of IEEE, ACM, Society for Neuroscience, and Cognitive Science Society.

Mariya Toneva is a postdoctoral fellow at the Neuroscience Institute at Princeton University and a tenure-track faculty at the Max Planck Institute for Software Systems, starting in Fall 2022. She obtained her PhD from Carnegie Mellon University in a joint program between Machine Learning and Neural Computation in 2021. Her research focuses on building computational models of language processing in the brain that can also improve natural language processing systems. Her work has appeared at CogSci, ICLR, NeurIPS, and SNL, and she has organized workshops and symposia at ICLR and SNL.

References

Beliy, R., Gaziv, G., Hoogi, A., Strappini, F., Golan, T., & Irani, M. (2019). From voxels to pixels and back: Self-supervision in natural-image reconstruction from fmri. *NeurIPS*, *33*, 6517–6527.

Caucheteux, C., Gramfort, A., & King, J.-R. (2021). Disentangling syntax and semantics in the brain with deep networks. In *ICML* (pp. 1336–1348).

Eickenberg, M., Gramfort, A., Varoquaux, G., & Thirion, B. (2017). Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, *152*, 184–194.

Gauthier, J., & Levy, R. (2019). Linking artificial and human neural representations of language. *EMNLP-IJCNLP*, 529-539.

Jain, S., & Huth, A. G. (2018). Incorporating context into language encoding models for fmri. In *NeurIPS-32* (pp. 6629–6638).

Kubilius, J., Schrimpf, M., Kar, K., Rajalingham, R., Hong, H., Majaj, N., ... others (2019). Brain-like object recognition with high-performing shallow recurrent anns. *NeurIPS*, *33*, 12805–12816.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., ... Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature Communications*, *9*(1), 1–13.

Schrimpf, M., Blank, I., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., ... Fedorenko, E. (2021). The neural architecture of language: Integrative reverse-engineering converges on a model for predictive processing. *PNAS*, *Vol*(118), 45.

Schwartz, D., Toneva, M., & Wehbe, L. (2019). Inducing brain-relevant bias in natural language processing models. *NeurIPS*, *33*, 14123–14133.

Toneva, M., Stretcu, O., Póczos, B., Wehbe, L., & Mitchell, T. M. (2020). Modeling task effects on meaning representation in the brain via zero-shot meg prediction. *NeurIPS*, *34*.

Toneva, M., & Wehbe, L. (2019). Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). *NeurIPS*, *33*, 14954–14964.

Wang, A., Tarr, M., & Wehbe, L. (2019). Neural taskonomy: Inferring the similarity of task-derived representations from brain activity. *NeurIPS*, *33*, 15501–15511.