A computational knowledge engine for human neuroscience

Ellie Beam (ebeam@stanford.edu), Stanford University, Department of Psychiatry & Behavioral Sciences

1 Introduction

Mental functions were largely defined in psychology before their biological basis was known. Here, deep learning is applied to assess how well mental functions in neuroimaging article texts predict spatial locations of brain activity. Several approaches to representing the semantic content of articles are compared, including term occurrences, embeddings trained with the global vectors (GloVe) algorithm [1], and long short-term memory (LSTM) encodings [2]. The results support the use of averaged GloVe embeddings as the language model for this task. Further gains in predictive performance are seen when training on article full texts instead of titles alone.

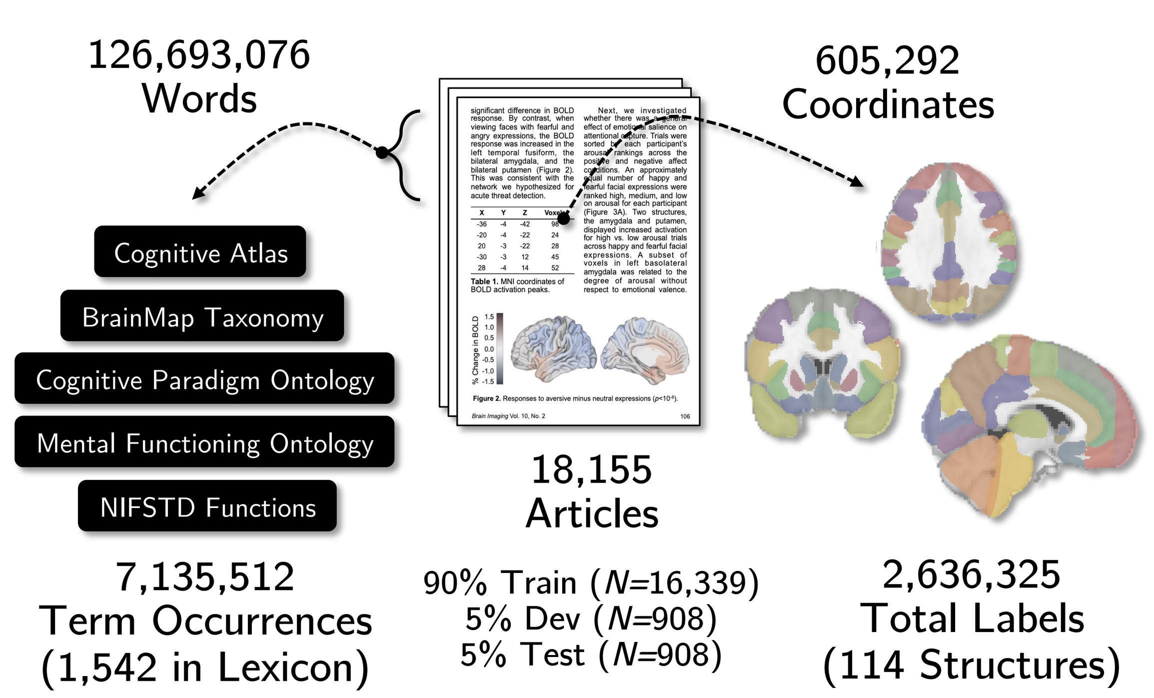
2 Dataset

Text Inputs

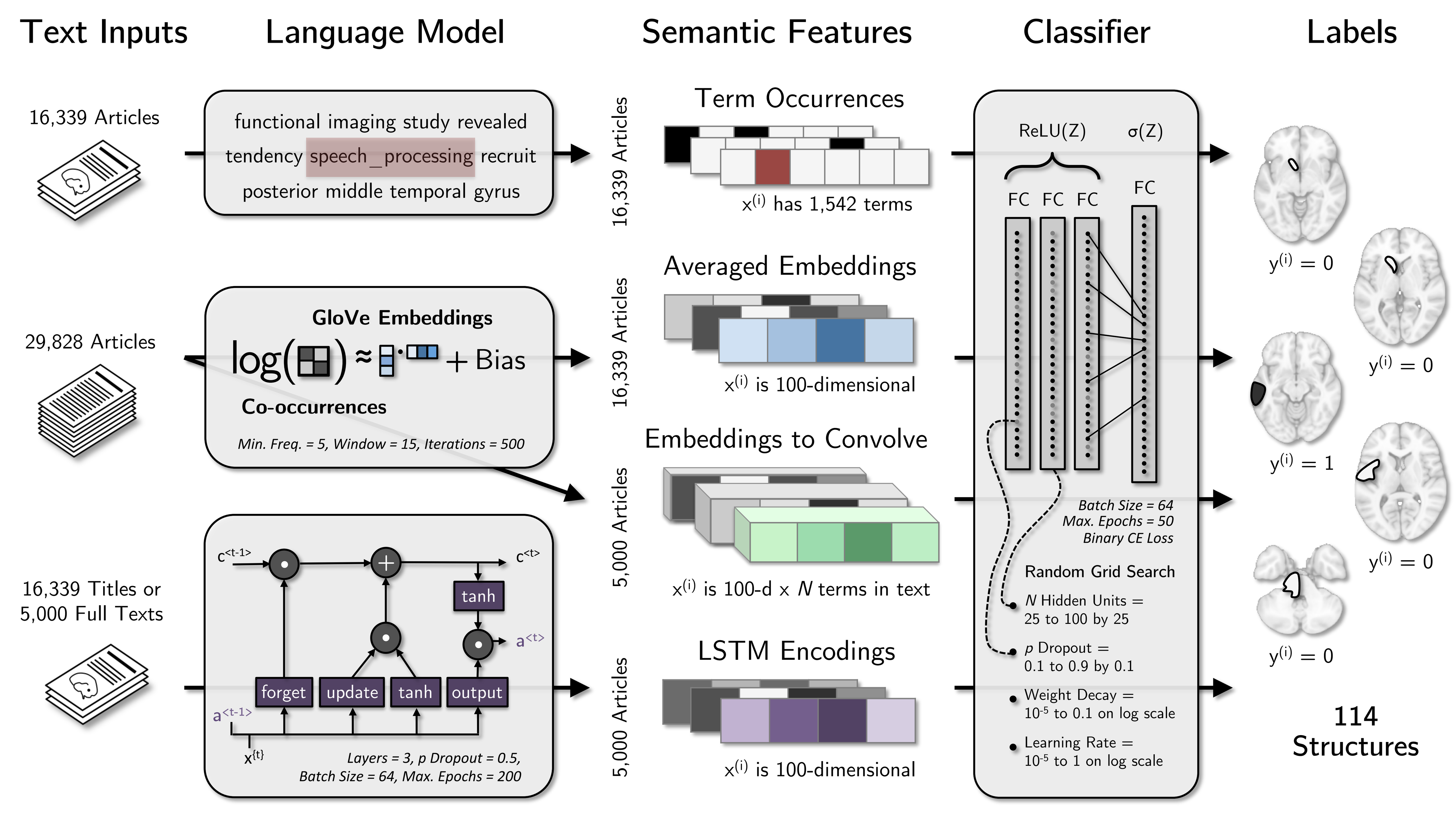
Studies were combined from BrainMap [3], Neurosynth [4], and accessible journals. Either the title or full text was taken as input. Preprocessing included lowercasing, lemmatization with WordNet, stop word removal, and combination of *n*-grams from a lexicon of terms for mental functions.

Neural Outputs

Spatial coordinates represent locations in the brain that were found to be statistically related to mental function by fMRI or PET imaging. Coordinates reported in Talairach space were converted to MNI, then mapped to a whole-brain neuroanatomical atlas.

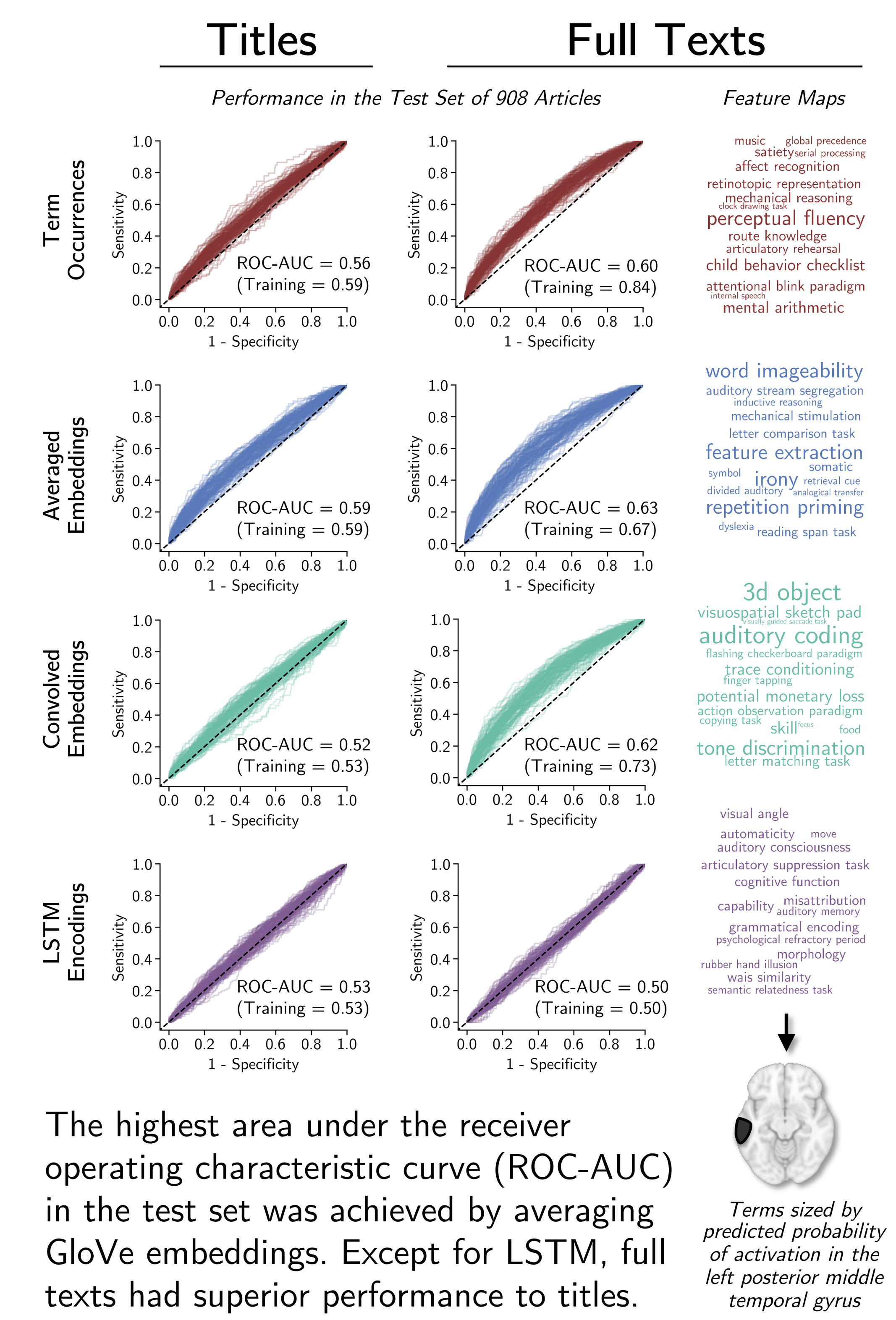


3 Features & Models



The semantic content of articles was represented by applying one of four language models to terms for mental functions in article titles or texts. Language model outputs served as inputs to a neural network classifier predicting whether a coordinate was reported in each structure of a neuroanatomical atlas.

4 Results

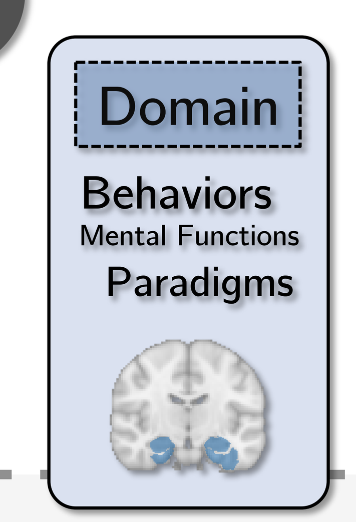


The highest area under the receiver operating characteristic curve (ROC-AUC) in the test set was achieved by averaging GloVe embeddings. Except for LSTM, full texts had superior performance to titles.

5 Discussion

- Models predicting neuroimaging study data from mental functions in article texts were able to achieve higher ROC-AUC than expected by chance
- The top models had feature maps that captured expected predictive relationships (e.g., posterior temporal lobe has a well established role in language processing, and it was strongly predicted by language terms)
- Embedding-based models better generalized to the test set than the term occurrence model, likely due to their more compact feature space
- LSTM underperformed other models and was unique in performing better with titles than full texts, suggesting that semantic representation is not improved by long-range dependencies across texts (which may be 1000s of words long)

6 Future Work



- Improve the convolutional and LSTM encodings of articles through architecture and hyperparameter search
- Generate "domains" of mental function and brain circuitry by unsupervised learning, then assess their reproducibility with the classification procedure developed here

[1] Pennington, J., Socher, R., and Manning, C. "GloVe: global vectors for word representation." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532-1543, 2014.

[2] Jain, S. and Huth, A.G. "Incorporating context into language encoding models for fMRI." *Neural Information Processing Systems* 32, pp. 1-10, 2018.

[3] Fox, P.T. and Lancaster, J.L. "Mapping context and content: the BrainMap model." *Nat. Rev. Neurosci.* 3, pp. 319-321, 2002.

[4] Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., and Wager, T.D. "Large-scale automated synthesis of human functional neuroimaging data." *Nat. Methods* 8, pp. 665-670, 2011.