Comparing Sentence-Based and Word-Based Semantic Space Representations to Brain Responses

Friederike Seyfried and Ping Li

July 14, 2020

Comparing sentence-based and word-based semantic space representations to brain responses

Abstract

Computational semantic space models have now been applied to sentences, but it is unclear whether they capture how the human brain represents sentences. Using fMRI we scanned adult readers reading expository texts and compared their brain responses to 3 semantic space vectors that modeled sentences either as combinations of words or as single units. We observe that computational semantic representations that are specifically designed to capture sentence content share information content with brain responses.

Introduction

To understand how information is processed by the brain, computational modeling of word semantics has been used in combination with brain imaging data. Moving onto larger units of language processing, it will be important to model sentence and discourse representations. In this emerging literature, sentences are sometimes modeled as a combinations of the word representations of the words in a sentences; for example, sentence semantic vectors are created by simply summarizing word vectors (i.e. Pennington, Socher, & Manning, 2014). However, sentences can also be viewed as units themselves and can be modelled without using word representations directly.

Sentence vectors created by computational networks perform well across tasks such as finding sentences with similar content (Reimers & Gurevych, 2019), but it is unclear if they resemble human sentence processing. In this study, we aim to explore whether sentence vectors can be used to study brain imaging data obtained from humans in a similar way as has been successfully done with word representations. Specifically, we compared the underlying information content of sentence vectors that use different approaches to model sentences to the brain responses from adults reading expository texts in a natural self-paced manner.

Materials and Methods

We tested 52 monolingual English-speaking adults who read 5 expository texts at a self-paced rate (see Hsu et al., 2019 for detailed procedure). While participants were reading both eyetracking data and fMRI data was acquired. fMRI data was acquired using a TR of 400 ms to collect one full brain volume (32 slices). The 5 texts participants read were between 28-31 sentences long and explained one concept, for example how electrical circuits work or if humans could live on Mars. The eyetracking data was used to determine word fixations during sentences for each participant to model brain responses to each sentence (fixation-related fMRI).

We used three sentence modeling approaches as models with which we compared the brain response to sentences. Global Vectors (GloVe, Pennington et al., 2014)is a model that was first developed for word semantic representation but can be used to represent sentence content by averaging the words of a sentence. GloVe is trained on co-occurrence corpora to obtain single word vectors. The word vectors can be summarized through different methods i.e. averaging, addition etc. to obtain a sentence vector that includes all words in a sentence (https://nlp.stanford.edu/projects/glove/). In this study, we used averaged word vectors.

Skipthought (Kiros et al., 2019) is based on a word semantic representation modeling approach as well (skip-gram) but this method models sentences as one unit. The model is trained to predict sentences that are left out (skipped) of a larger text context. BERT (Devlin, Chang, & Toutanova, 2018)is a recently developed successful word semantic model. BERT is a word based model that models word context in a bidirectional manner by masking preceeding or following words of the word that the algorithm is acquiring a representation for (Devlin et al., 2018). BERT has been adapted to be used on sentences by Reimers and Gurevych (2019, Sentence-BERT). Unlike GloVe, BERT is context-sensitive in extracting semantic representations, using the surrounding text to create word vectors. We obtained sentence vectors for the 5 texts participants read in the scanner from these 3 sentence representation models and compared each model to the brain responses in order to test if sentence vectors do capture how the brain represents sentence content.

We used representational similarity analysis (RSA, Kriegeskorte, 2008) to compare the information content of sentence vectors with the information content of the brain while reading the same sentences. RSA is a widely used method to study representations between different methods, for example neuroimaging and computational modeling. All analyses were carried out using the RSA toolbox for Matlab (Nili et al., 2014). The brain response to each sentence was compared to the response to each other sentence in the text to obtain a similarity score for each sentence pair and each sentence vector pairs analyzed in an analogous way.

We use a ROI-based approach focused on the left hemisphere and compared RDMs across 13 ROIs in the left hemisphere including Broca's area, the temporal lobes and the angular and supramarginal gyrus. A depiction of the ROIs used can be viewed in figure 1. The ROIs were calculated based on the Harvard-Oxford cortical atlas implemented in FSL (Desikan et al., 2006). RDMs were computed for each ROI based on the responses in the voxels in this ROI and then compared to each model RDM separately for each text resulting in 65 separate model to brain response comparisons.

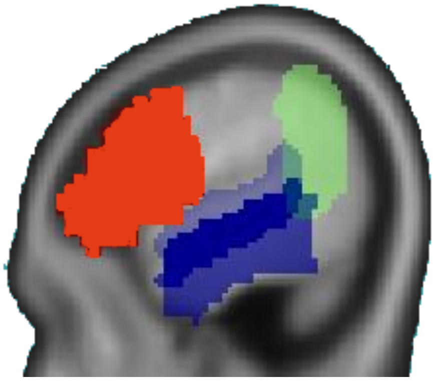

Figure 1. Regions of interest: There were 13 ROIs from three brain lobes. Two frontal lobe ROIs, the superior frontal lobe and Broca's area are shown in orange. There were four parietal ROIs (light green), the inferior parietal lobule, the precuneus, the supramarginal gyrus and the angular gyrus. Finally there were 6 temporal ROIs depicted in blue, the inferior, middle and superior temporal gyrus, the temporal fusiform gyrus and the temporal pole and the posterior inferior temporal gyrus.

Results

Across all 5 texts and 13 ROIs, BERT based RDMs were most similar to the RDMs obtained from brain responses (44 out of 65 comparisons). The Skipthought based RDMs performed second best, being the best fit in 21 cases. All correlations between BERT based RDMs and brain response and Skipthought based RDMs and brain responses reached significance. The GloVe based RDMs were less similar to the brain responses and failed to reach significance in 17 cases. There were no observable trends across ROIs with most ROIs showing highest similarity for the BERT based RDM. In 12 ROIs, the Skipthought based RDM was also observed to be most like the brain response, with the only ROI that was only correlated highest with one sentence vector type (BERT in this case) being the left angular gyrus.

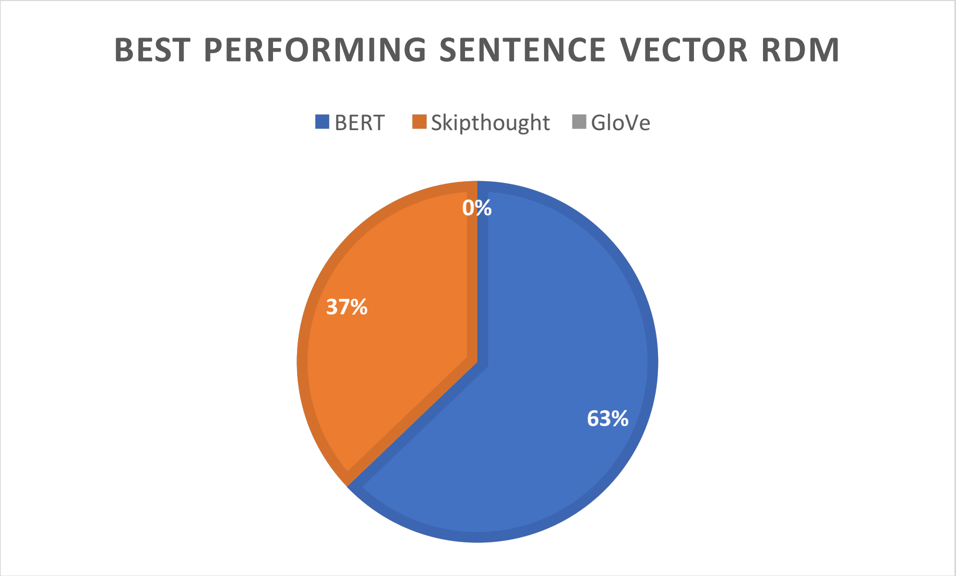

Figure 1. Comparison of the three NLP models compared to the brain data. Across 5 texts and 13 ROIs BERT was the most representationally similar model in 63% of cases, whereas Skipthought was the most similar in 37% of cases. The word based model GloVe was less similar to the representations obtained by using fMRI data.

Discussion and Conclusion

Natural language processing using computational modeling has moved from capturing semantic representations of words to larger linguistic units such as sentences (and discourse in rare cases; see Wehbe et al., 2014). It remains unknown whether computationally generated sentence vectors can capture how the human brain processes sentences, if the vectors are based on models that were trained on human texts but are not designed to model how humans process language. The preliminary results presented here support that models that are aiming to specifically represent sentences are successfully modeling aspects of human sentence processing. We observed that across 13 ROIs in the language network, BERT based sentence vectors trained using an adaptation of BERT specifically aiming at modeling sentences, more closely represent information content as brain responses. Skipthought based sentence vectors also seem to capture at least some aspects of the information content, but this model performed less accurately in comparison to the BERT model. Sentence vectors based on averaging word representations do not appear to represent information content that resembles brain response information content. This supports to the idea that sentence representations in the brain are more than a combination of word semantic representations.

Interestingly, we did not observe a differentiation across different parts of the language network in terms on what kind of information content is represented. It has been hypothesized that there is a division of labor across the language network with some regions being related to semantic processing of smaller units such as words, whereas other regions are more representative of syntactic processing (Friederici, Rüschemeyer, Hahne, & Fiebach, 2003). It has also been suggested that there are gradations of unit size that is processed across the language network, with units of processing becoming increasingly larger more anteriorly in the brain (Bornkessel-Schlesewsky, Schlesewsky, Small, & Rauschecker, 2015). There is also evidence that across the sentence, the brain responses could change depending on the contents of information processing and the level of depth in processing (Hsu, Clariana, Schloss, & Li, 2019). Our results do not show a difference in more word based vs more sentence based or smaller vs larger unit differentiation. However, this conclusion remains to be examined further, because our study only targets sentences and there is no direct comparison between sentence representation and word representation.

References

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., & Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: common computational properties. *Trends in Cognitive Sciences*, *19*(3), 142–150. http://doi.org/10.1016/j.tics.2014.12.008

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. http://doi.org/10.1016/j.neuroimage.2006.01.021

Devlin, J., Chang, M.-W., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv*, 1–14.

Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., & Fiebach, C. J. (2003). The Role of Left Inferior Frontal and Superior Temporal Cortex in Sentence Comprehension: Localizing Syntactic and Semantic Processes. *Cerebral Cortex*, 1–8.

Hsu, C.-T., Clariana, R., Schloss, B., & Li, P. (2019). Neurocognitive Signatures of Naturalistic Reading of Scientific Texts: A Fixation-Related fMRI Study. *Scientific Reports*, 1–16. http://doi.org/10.1038/s41598-019-47176-7

Kiros, R., Salakhuditnov, R., Zemel, R. S., Torralba, A., Urtasun, R., & Fidler, S. (2019). Skip-Thought Vectors. *ArXiv*, 1–9.

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 1–28. http://doi.org/10.3389/neuro.06.004.2008

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e1003553–11. http://doi.org/10.1371/journal.pcbi.1003553

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *ArXiv*, 1–12.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *ArXiv*, 1–11.

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A., & Mitchell, T. (2014). Simultaneously Uncovering the Patterns of Brain Regions Involved in Different Story Reading Subprocesses. *Plos One*, *9*(11), e112575–19. http://doi.org/10.1371/journal.pone.0112575