Compositionality and Algorithmic Complexity in Neural Networks

Charles Torres[*1] and Richard Futrell[1]

[*]charlt4@uci.edu
[1]Language Science, University of California, Irvine, USA

1. Introduction

Language is compositional. One explanation for the existence of compositionality is that it is more compressible than a holistic system and has a simpler representation when making use of the *minimum description length* (MDL) principle (Brighton, 2002). The MDL principle is a learning procedure where a learner chooses the hypothesis which best optimizes the compressibility of the data and the hypothesis used to generate that data.

However, the MDL principle is sensitive to the hypothesis space under consideration as the computational formalism used can affect the learning procedure (McGregor, 2014). We investigate the compressibility of compositionality in one learning system: LSTMs. Previous work has shown that LSTMs prefer to learn compositional mappings from meaning to signal (Ren, Guo, Labeau, Cohen, & Kirby, 2019). We introduce a method which allows us to measure the complexity/accuracy tradeoff of LSTMs and use this to measure compositional and holistic mappings. We find that compositional languages have a more favorable tradeoff curve within LSTM architectures which may explain this preference.

2. Methods

Measuring the complexity of an LSTM could take a number of different forms. We consider the number of non-zero parameters in the network which measures the number of connections between neurons. We use a stochastic method of approximating L_0-regularization which pressures a network to zero-out parameters Louizos, Welling, and Kingma (2018). The severity of network pruning depends on the strength of the regularization term. See supplementary materials for implementation.

The task was to correctly learn a compositional or holistic language represented as a meaning-to-word mapping. Each meaning was a pair of the form (x, y) where $x, y \in \{1, 2, 3, 4\}$, and each word was an 8 character string from $\{a, b, c\}$. Compositional languages could be decomposed into two 4 character subwords mapping to meanings in x or y. LSTMs were trained using the cross-entropy objective. The entire language formed both the training and testing set

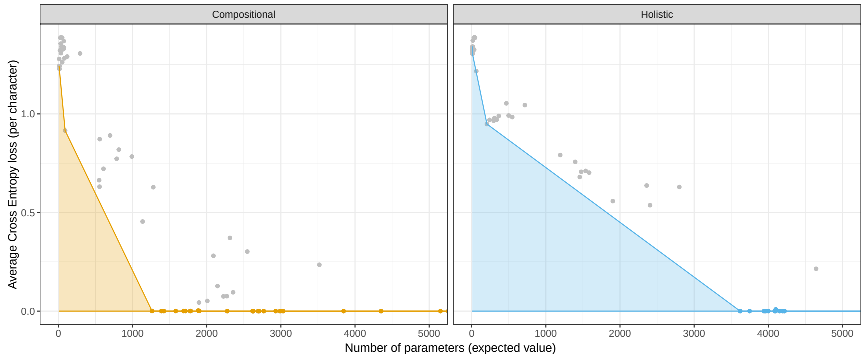

Figure 1. Area under the curve for the holistic and compositional languages. The area under the holistic curve is significantly larger owing to the greater complexity of networks which perform well on the task (colored points).

(memorization being desirable), and training was performed for 20,000 epochs or until performance did not improve for 50 epochs, then accuracy and number of non-zero parameters were stored. Training was repeated with different regularization penalties to yield 80 datapoints per language (for a total of 160). A concave curve was fit to the points representing the frontier of the complexity and accuracy of the LSTMs on the holistic and compositional objectives. The favorability of the tradeoff was measured as a difference in area under these curves.

3. Results

Results are visible in Figure 1. Holistic languages require more neural-network connections to learn accurately when compared with their compositional counterparts. This result is significant under a permutation test ($p < .001$).

4. Conclusion

Compositional languages are easier to represent than holistic languages in a neural network model, a fact which may explain their preferential learning as in Ren et al. (2019). Furthermore, that there is a correspondence between a “neural” representation and a “symbolic” one (as in Brighton (2002)) is interesting. LSTMs are not human brains, but these correspondences may hold in more biologically inspired models (like spiking neural networks), and we believe further investigations into the biases of different formalisms (LSTMs, brains, or otherwise) under simplicity constraints is desirable.

References

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial life*, *8*(1), 25–54.

Louizos, C., Welling, M., & Kingma, D. P. (2018). Learning sparse neural networks through l_0-regularization. In *International conference on learning representations.*

McGregor, S. (2014). Natural descriptions and anthropic bias: Extant problems in solomonoff induction. In *Language, life, limits: 10th conference on computability in europe* (pp. 293–302).

Ren, Y., Guo, S., Labeau, M., Cohen, S. B., & Kirby, S. (2019). Compositional languages emerge in a neural iterated learning model. In *International conference on learning representations.*