Seq2Seq Models with Dropout can Learn Generalizable Reduplication

Linguistics Department - University of Massachusetts Amherst
Brandon Prickett, Aaron Traylor, and Joe Pater
SIGMORPHON 2018

1) Introduction

- The debate between connectionist and symbolic theories of grammar has largely revolved around morphology (Rumelhart and McClelland 1986; Pinker and Prince 1988).
- This includes reduplication, with many claiming that connectionist models without explicit, algebraic variables cannot represent reduplicative patterns.
  - ⇒ *Example of reduplication from Karao (from Štekaurer et al. 2012):*

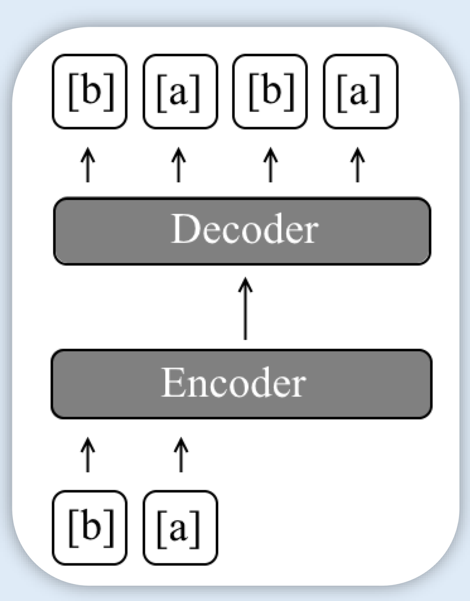
    man**ba**kal ⟶ man**baba**kal

    'fight each other (2 people)' 'fight each other (>2 people)'
  - ⇒ *Example of reduplication using algebraic variables (where α stands for the reduplicated stem):*

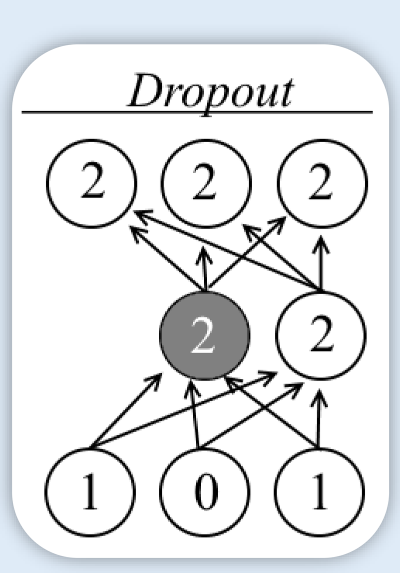
    $$\alpha \longrightarrow \alpha\alpha$$
- Marcus et al. (1999) showed that simple, variable-free networks could not generalize reduplication, whereas infants could.
  - ⇒ See Gasser (1993), Berent (2013), and Tupper and Shahriari (2016) for more discussion on this.
- Here we apply a state-of-the-art neural network with no explicit variables to the problem of reduplication and show that it succeeds where simpler neural networks failed.

2) The Model

- We used a Sequence-to-Sequence architecture (Seq2Seq; Sutskever et al. 2014).
  - ⇒ Models string-to-string mappings where the input and output have independent lengths.
  - ⇒ Performs well at other morphological tasks (Cotterell et al. 2016) and correlates well with human behavior (Kirov 2017; Kirov & Cotterell 2018).



- The model used Long Short-Term Memory (LSTM; Hochreiter and Schmidhuber 1997).
  - ⇒ Allows a model to keep track of which features of the input string are most helpful for predicting long-term patterns in the output.
  - ⇒ Prevents vanishing gradients (Bengio et al. 1994) and increases models' representational power (Levy et al. 2018).
- We also included simulations with and without Dropout to test its effect on the model's generalization (Srivastava et al. 2014).



  - ⇒ When using dropout, a random subset of the network's units won't activate, regardless of their input.
  - ⇒ This causes the model to find a more general solution.

3) Simulation Design

- Berent (2013) describes three different scopes of generalization for reduplication-like patterns:

	i	e	o	a
p	pi	pe	po	pa
b	bi	be	bo	ba
t	ti	te	to	ta
d	di	de	do	da
n	ni	ne	no	na

Novel Syllable/Word

	i	e	o	a
p	pi	pe	po	pa
b	bi	be	bo	ba
t	ti	te	to	ta
d	di	de	do	da
n	ni	ne	no	na

Novel segment

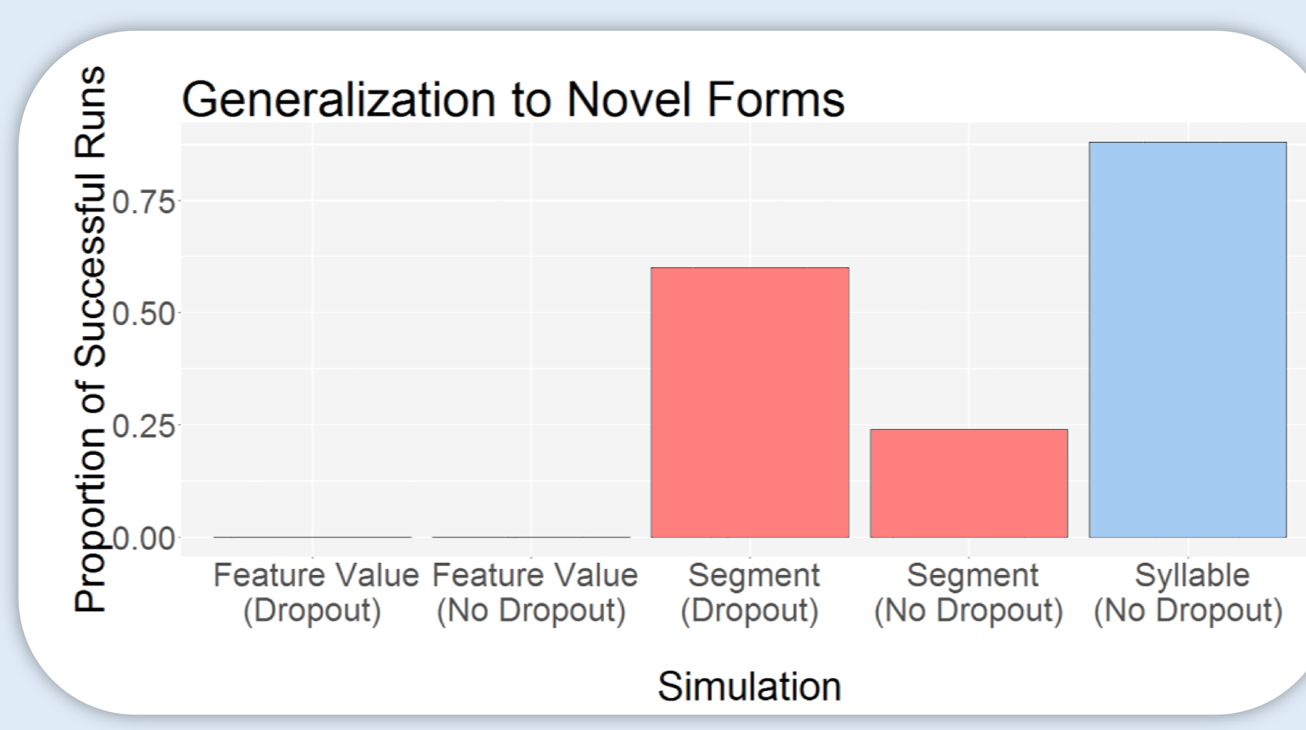
	i	e	o	a
p	pi	pe	po	pa
b	bi	be	bo	ba
t	ti	te	to	ta
d	di	de	do	da
n	ni	ne	no	na

Novel feature value

- To test our model's scope of generalization, we used randomly-produced toy languages and withheld data from training that represented each of the three scopes.
  - ⇒ Each language had an inventory of 40 segments, with the withheld segments being randomly chosen in each simulation (except for the novel-feature-value simulations, in which [n] was always withheld).

4) Results

- The model successfully learned how to map stems in the training data to their reduplicated forms for all of the simulations.
- Dropout increased the model's scope of generalization from novel syllables to novel segments.



5) Discussion

- Without dropout, the model could generalize to novel syllables.
- Dropout increased the model's scope of generalization, but regardless of dropout, generalizing to novel feature values seems to be out of the model's reach.
  - ⇒ But do humans generalize to novel feature values? This is unclear, based on the data presented by Marcus et al. (1999) and Berent (2013).
- These results suggest that variables may not be necessary to model human generalization of reduplication.

Acknowledgments

The authors would like to thank the members of the UMass Sound Workshop, the members of the UMass NLP Reading Group, Tal Linzen, and Ryan Cotterell for helpful feedback and discussion. Additionally, we would like to thank the SIGMORPHON reviewers for their comments. This work was supported by NSF Grant #BCS-1650957.

References

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166. **Berent, I. (2013).** The phonological mind. *Trends in Cognitive Sciences*, *17*(7), 319–327. **Cotterell, R., Kirov, C., Sylak-Glassman, J., Yarowsky, D., Eisner, J., & Hulden, M. (2016)**. The SIGMORPHON 2016 shared task—morphological reinflection. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology* (pp. 10–22). **Hochreiter, S., & Schmidhuber, J. (1997)**. Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. **Kirov, C. (2017)**. Recurrent Neural Networks as a Strong Domain-General Baseline for Morpho-Phonological Learning. In *Poster presented at the 2017 Meeting of the Linguistic Society of America*. **Kirov, C., & Cotterell, R. (2018)**. *Recurrent Neural Networks in Linguistic Theory: Revisiting Pinker & Prince (1988) and the Past Tense Debate*. **Levy, O., Lee, K., FitzGerald, N., & Zettlemoyer, L. (2018)**. Long Short-Term Memory as a Dynamically Computed Element-wise Weighted Sum. *ArXiv Preprint ArXiv:1805.03716*. **Marcus, G. (2001)**. *The algebraic mind*. Cambridge, MA: MIT Press. **Marcus, G., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999)**. Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80. **Rumelhart, D., & McClelland, J. (1986)**. On learning the past tenses of English verbs. In J. McClelland & D. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2: Psychological and Biological Models, pp. 216–271). The MIT Press. **Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014)**. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958. **Štekauer, P., Valera, S., & Körtvélyessy, L. (2012)**. *Word-formation in the world's languages: a typological survey*. Cambridge University Press. **Sutskever, I., Vinyals, O., & Le, Q. V. (2014)**. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104–3112). **Tupper, P., & Shahriari, B. (2016)**. Which Learning Algorithms Can Generalize Identity-Based Rules to Novel Inputs? *ArXiv Preprint ArXiv:1605.04002*.