Compositional generalization through meta sequence-to-sequence learning

Brenden M. Lake
New York University & Facebook AI Research

facebook Artificial Intelligence

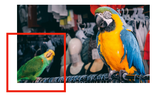
Abstract

People can learn a new concept and use it compositionally, understanding how to "dax twice" after learning how to “dax." In contrast, powerful sequence-to-sequence (seq2seq) neural networks fail such tests of compositionality, especially when composing new concepts together with existing concepts. I show how memory-augmented neural networks can be trained to generalize compositionally through meta seq2seq learning. In this approach, models train on a series of seq2seq problems to acquire the compositional skills needed to solve new seq2seq problems. Meta se2seq learning solves several of the SCAN tests for compositional learning and can learn to apply implicit rules to variables.

Systematic compositionality

Human language and thought are characterized by systematic compositionality — the algebraic capacity to understand and produce novel combinations from known components.

Learning a new verb
“This is how you photobomb”



Can you then…
“…photobomb twice?"
“…photobomb while jumping?”
“…photobomb vigorously?”

Learning a new verb
“This is how you dax”

Can you then…
“…dax twice?"
“…dax while jumping?”
“…dax like you mean it?”

Fodor & Pylyshyn (1988), Marcus (2003), and others have argued that standard neural networks cannot capture systematic compositionality. What about contemporary deep neural networks?

A testbed for compositional generalization

The SCAN challenge for compositional generalization (Lake & Baroni, 2018): Can a model generalize to "jump twice" after learning how "walk twice," "run twice," and "jump?"

Training

“walk” ➔ WALK

“walk twice” ➔ WALK WALK

“walk around left” ➔ LTURN WALK LTURN WALK LTURN WALK LTURN WALK

“turn left and walk opposite twice” ➔ …

“run” ➔ RUN

“run twice” ➔ RUN RUN RUN

“run around left” ➔ LTURN RUN LTURN RUN LTURN RUN LTURN RUN

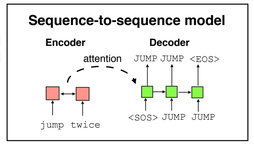
"turn left and run opposite twice” ➔ …

“jump” ➔ JUMP

Test

“jump twice” ➔ ?

“turn left and jump opposite twice” ➔ ?



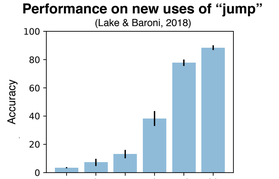
SCAN details

- ~21,000 commands in total
- Primitives (6): walk, look, run, jump, turn left, turn right, each with a corresponding action
- Modifiers (8): twice, thrice, around left, opposite left, etc.
- Conjunctions (2): and, after

Results

- Standard nets have difficulty incorporating a new primitive (jump) and aren’t systematic: e.g., responding incorrectly on all conjunctions of “jump thrice” or “jump", while ironically succeeding on “jump thrice and jump"

Performance on new uses of “jump”
(Lake & Baroni, 2018)



Meta sequence-to-sequence (meta seq2seq) learning

- Meta seq2seq learning: Training on a series of seq2seq problems (episodes) to acquire the compositional skills needed for solving new seq2seq problems
- The approach demands structured, memory-augmented architectures such that new problems can be solved using external memory and recurrent dynamics, rather than weight updates

Mutual exclusivity task

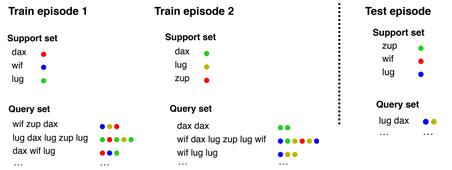
This task requires two abilities inspired by human generalization patterns (Lake, Linzen, & Baroni, 2018)
1) Learning isolated symbol mappings, and then generalizing to symbol sequences
2) Using mutual exclusivity (ME) to resolve unseen mappings: assuming that if an object has one label, it does not need another label (standard deep nets struggle to learn a ME bias; Gandhi & Lake, 2019)

Meta-training

- 10,000 episodes with permuted meanings
- 4 possible inputs: dax, wif, lug, zup
- 4 possible outputs: ●,●,●,●

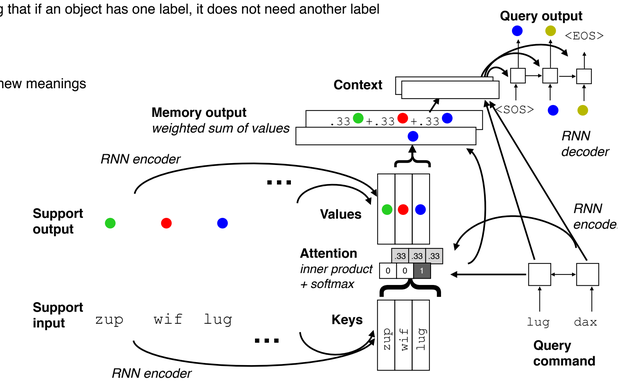
Test

- Novel episodes with new meanings (permutations)

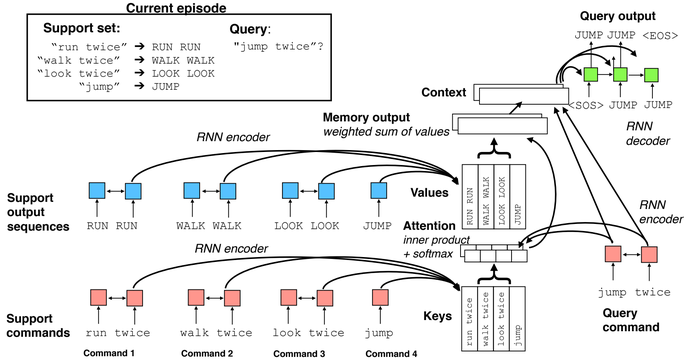


Results

- 100.0% (n=5) performance on quickly acquiring and using new mappings, using just memory and activation dynamics



SCAN task for compositional generalization



Meta-training

- 10,000 episodes encouraging the network to learn new meanings for primitive commands (permuted meanings for ‘jump’,‘run’, ‘walk’, ‘look’) and use them compositionally to answer queries
- Episodes have 20 support and 20 query commands sampled from SCAN (Loss computed on query)

Results

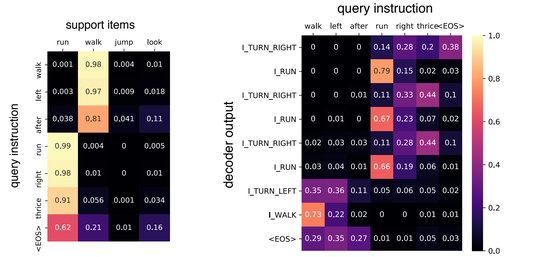
- Evaluated on new episodes with original mappings (held-out; see diagram)
- Support commands: only “jump”, “walk”, “run”, and “look”
- Query commands: “jump twice", "turn left and jump opposite thrice", "jump around twice"…
- Near perfect performance shows that meta se2seq learning can learn new meanings and use them compositionally, learning to apply implicit rules to variables.

Model	Accuracy (n=5)
meta seq2seq	**99.95%**
-without memory	0.03%
-without decoder attention	10.32%
-without support loss	5.43%

Interpreting the meta seq2seq model

Current episode

Support set:		Query:
“run”	➔ RUN	“walk left after run right thrice”?
“walk”	➔ WALK	
“look”	➔ LOOK	
“jump”	➔ JUMP	



Attention over the support items (columns) while reading query (rows)

Decoder attention over input command words (columns) while producing output actions (rows)

Conclusions

- People are skilled compositional learners while standard neural networks are not
- The meta seq2seq approach can learn to generalize compositionally, exploiting the algebraic structure of a domain to help understand novel utterance
- Rather than attempting to solve commands like “jump around right twice and walk thrice” by comparing surface level patterns in the training corpus, meta seq2seq learns to treat the instruction as a template, “**x** around right twice and **y** thrice” with **x** and **y** as variables
- We are currently extending the approach to large-scale compositional word learning in language modeling
- By facilitating the acquisition of compositional skills, the meta seq2seq framework holds promise for building more human-like and more capable AI systems

References

Jerry Fodor and Zenon Pylyshyn. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28:3–71, 1988.
Kanishk Gandhi and Brenden M Lake. Mutual exclusivity as a challenge for neural networks. *arXiv preprint*, 2019.
Brenden M Lake and Marco Baroni. Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks. In *International Conference on Machine Learning (ICML)*, 2018.
Brenden M Lake, Tal Linzen, and Marco Baroni. Human few-shot learning of compositional instructions. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 2019.
Gary F Marcus. *The Algebraic Mind: Integrating Connectionism and Cognitive Science.* MIT Press, Cambridge, MA, 2003.