Generalized Schema Learning in Neural Networks

Catherine Chen, Qihong Lu, Andre Beukers, Chris Baldassano, and Kenneth Norman

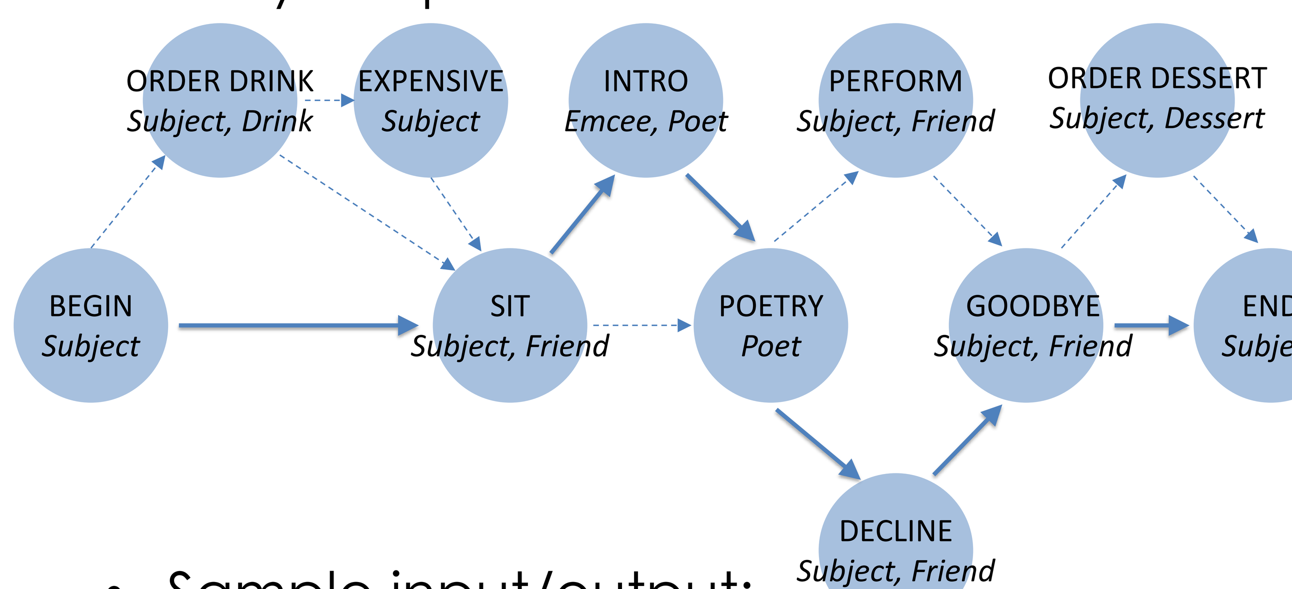
PRINCETON UNIVERSITY

Schema Learning

- Humans learn and apply schemas to understand the world.
- Past work: neural networks can perform role-filler binding when explicitly told what filler information to maintain [5].
- *Can neural networks learn role-filler regularities and generalize to new fillers?*

Approach

- Test networks' ability to perform role-filler binding: associate concrete "fillers" with abstract "roles" using context.
- Construct examples using Coffee Shop World [2]: a program that stochastically generates stories based on pre-defined statistics.
  - Story Graph:

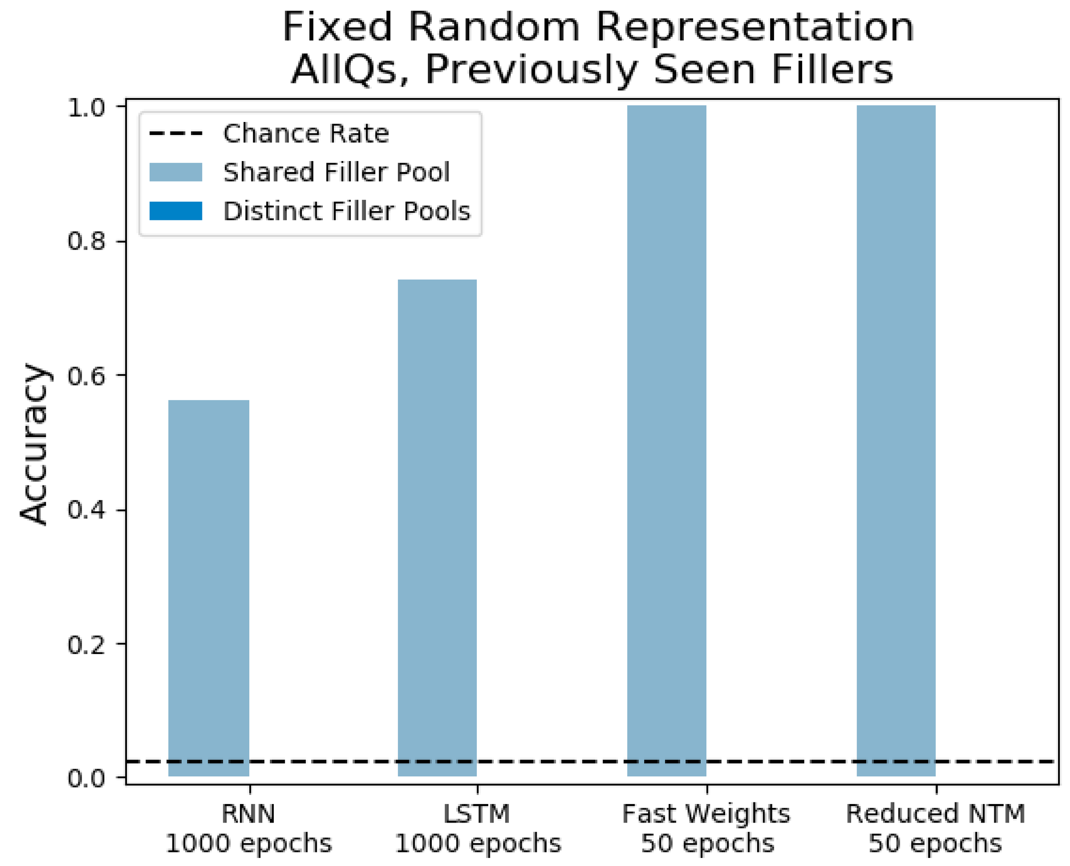


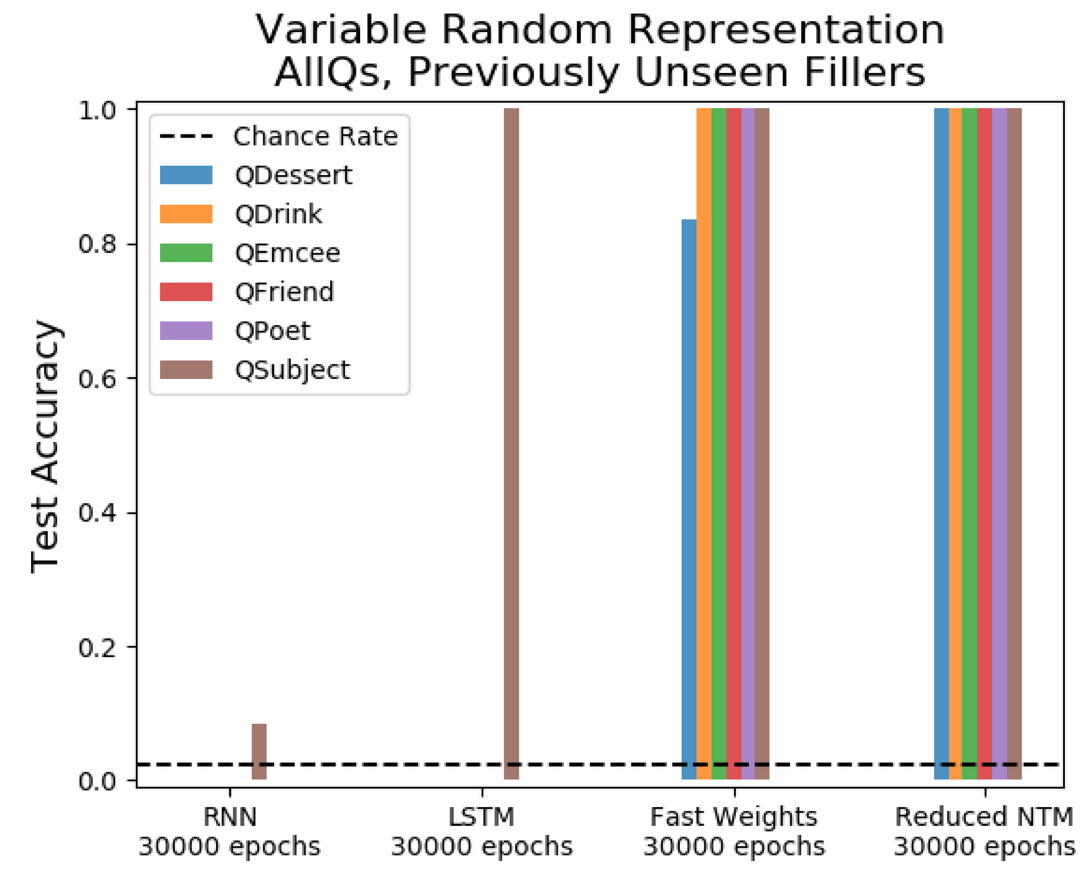
  - Sample input/output:

    BEGIN Amy SIT Amy Bob INTRO Cal Deb POETRY Deb DECLINE Amy Bob GOODBYE Amy Bob END Amy ? QPoet / Deb

- Train networks with four types of memory architectures:
  - RNN: Standard recurrent neural network.
  - LSTM [4]: RNN with gated hidden state.
  - Fast Weights [1]: RNN with associative memory matrix.
  - Reduced NTM [3]: RNN with external memory buffer.
- Decode network activity.

Results





Experiments 1 & 2: Small Set of Train Fillers.

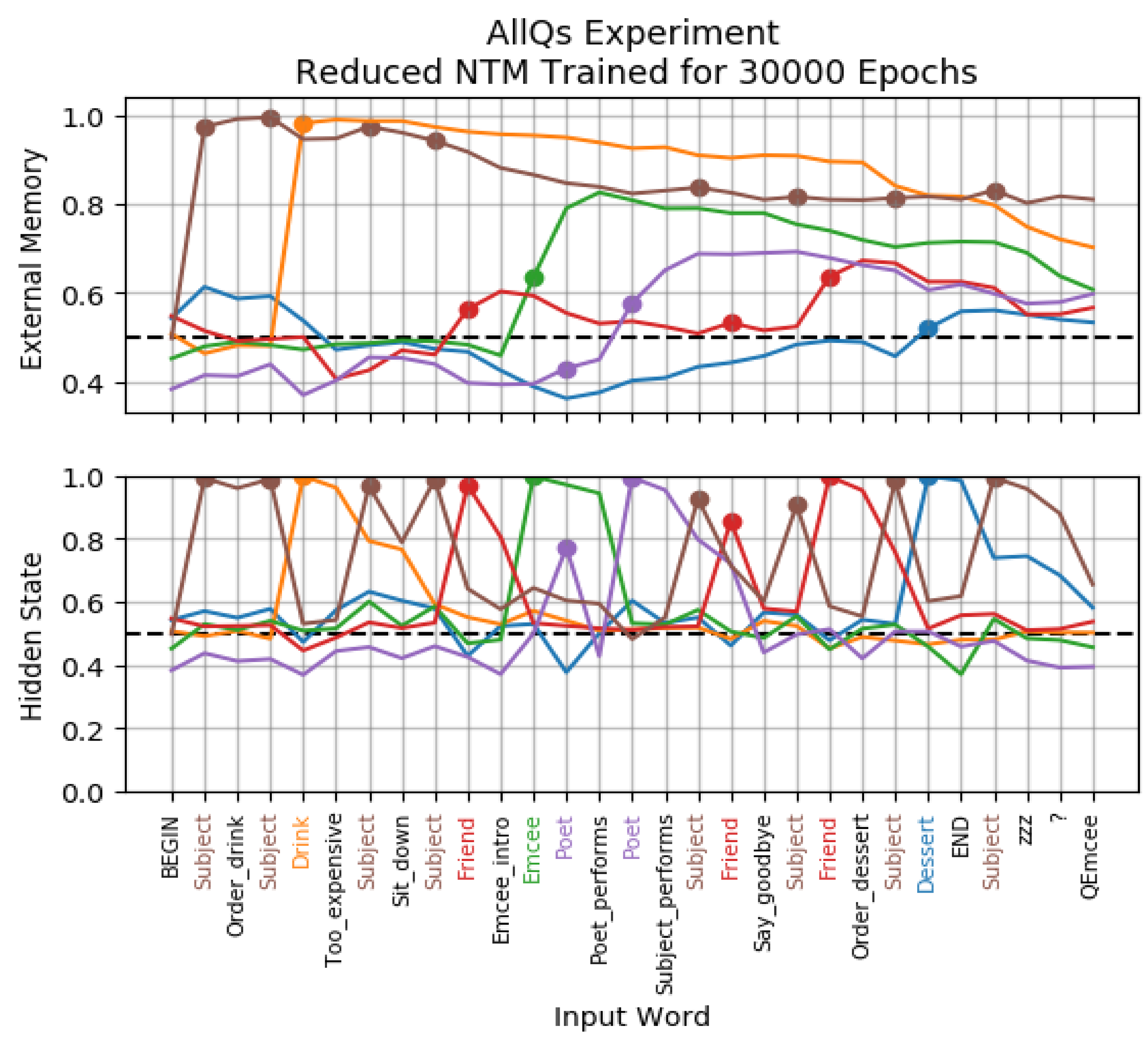
- Experiment 1: *Shared filler pools*: Test and train stories generated using the same fillers.
  - All architectures learn role-filler binding if fillers were present during training (even if story is a new path through the graph).
- Experiment 2: *Distinct filler pools*: Test and train stories using different, non-overlapping fillers.
  - All networks fail role-filler binding on stories with new fillers.

Experiment 3: Large Set of Train Fillers

- *Variable Random Inputs*: Networks must generalize for train and test stories.
  - Fillers were random vectors, different in each story.
- All architectures reach above-chance test accuracy → perform some amount generalization if forced to do so during training.
- *Split-task error analysis*:
  - LSTM and RNN learn to generalize only on the easiest task.
  - Reduced NTM and Fast Weights networks learn to solve all six tasks.

Experiment 4: Decoding Analysis

- *Method*: Record network states during input sequences and use ridge regression to decode fillers from network states.
- Scores for the hidden state peak when filler appears in input and then fall.
- Scores for enhanced memory components rise when network receives the relevant filler in the input sequence, and remain high.
- Scores from the reduced NTM are shown here as an example.



Conclusions

Generalizing role-filler binding to previous unseen fillers:

- Depends on the breadth of train fillers.
- Depends on memory architectures: simpler RNNs learn only the simplest task.

Decoding analysis gives insight into how networks learn to solve tasks, and how they might use enhanced memory to solve more complex tasks.

References

1. Ba, J., Hinton, G., Mnih, V., Leibo, J. Z., & Ionescu, C. Using Fast Weights to Attend to the Recent Past. (2016).
2. Coffee Shop World. (n.d.). (https://github.com/PrincetonCompMemLab/narrative)
3. Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwiska, A., . . . Hassabis, D. Hybrid computing using a neural network with dynamic external memory. *Nature* 538, 471-476. (2016).
4. Hochreiter, S., & Schmidhuber, J. Long short-term memory. *Neural Computation* 9, 1735-1780. (1997).
5. St. Cy, M. F., & McClelland, J. L.. Learning and applying contextual constraints in sentence comprehension. *Artificial intelligence* 46, 217-257. (1990).

Funding was provided by a grant from Intel Labs and by a Multi-University Research Initiative grant to KAN (ONR/DoD N00014-17-1-2961). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research or the U.S. Department of Defense.