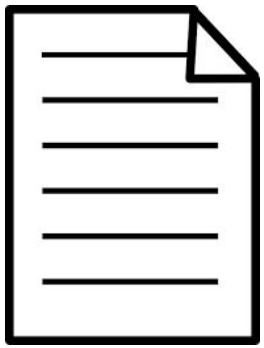
Integrating Order Information and Event Relation for Script Event Prediction

Zhongquing Wang, Yue Zhang,
and Ching-Yun Chang

Presented by James Finch

Task



Syntactic parsing

Entities

X = Customer, Y = Waiter

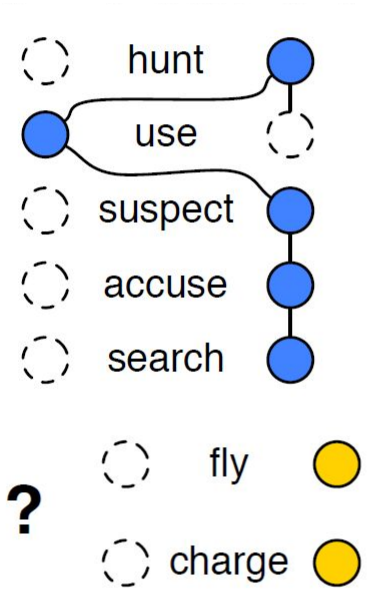
Context(e_i)

walk(X, restaurant), seat(X), order(X, food), serve(Y, food)
eat(X, food), make(X, payment), _____

c_1: receive(X, response)
c_2: drive(X, mile)
c_3: seem(X)
c_4: discover(X, truth)
c_5: leave(X, restaurant)

?

Previous Work - Chambers and Jurafsky (2009)



1. Calculate PMI between candidate events
    - Counts of times two event verbs appear together in the same coreference chain
2. Score of candidate event's coherence with the current cluster of events
3. Add events by max scoring candidate up to max of 7 events
4. Run a separate SVM classifier to order events after clustering to get final event chain

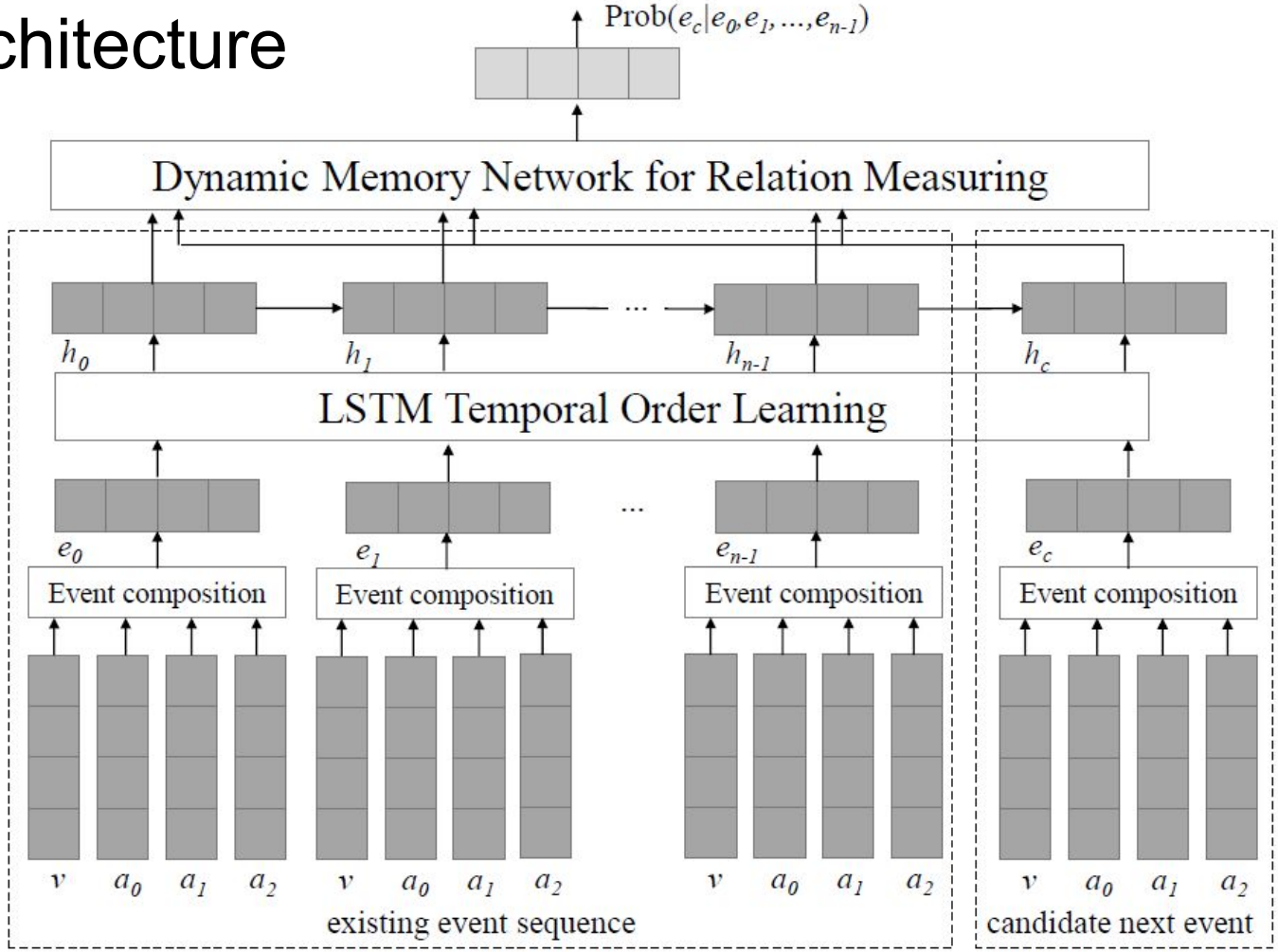
Differences w.r.t. Chambers and Jurafsky (2009)

- Interested in script event prediction
  - “Order” means order appearing in text
  - Don’t care about event chains themselves (no custering)
- Word embeddings reduce data sparsity issue
- Neural approach improves prediction performance

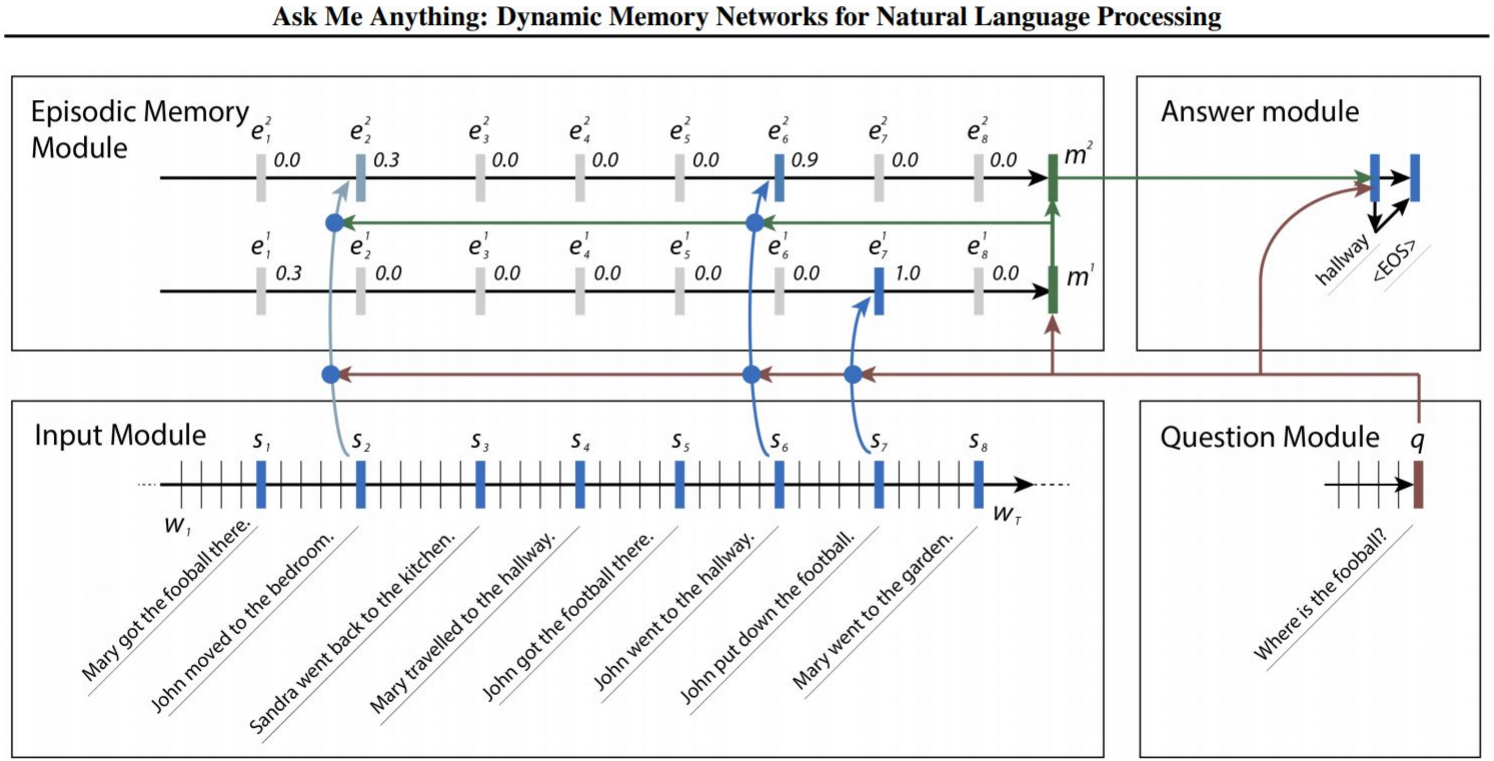
Approach

- Use LSTM to capture & predict sequences
- Sophisticated attention mechanism
- Unsupervised (gigaword corpus)

System architecture



Dynamic Memory Network (Kumar et al. 2015)

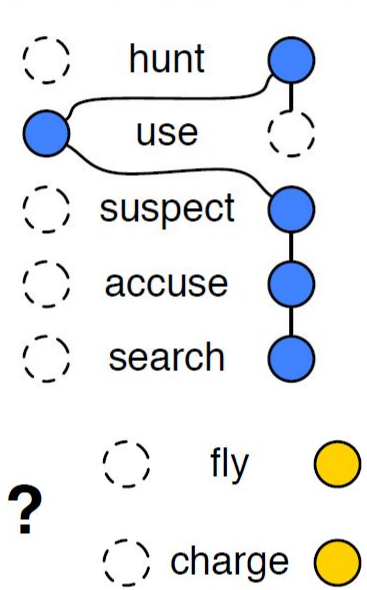


Approach Motivation

- Relevant events could come from anywhere in history
- DMN implicitly “clusters” using attention mechanism
- Use initial LSTM to model event order

Evaluation and Results

Evaluation on
Multiple-Choice Cloze Task:



Trained on NYT portion of Gigaword corpus (unsupervised) with 1,500,000 event chains

Two evaluation datasets:

1. G&C16 - 10,000 test cases
2. C&J08 - 346 test cases in 69 documents

Results

PMI: Chambers and Jurafsky (2008)

Bigram: another statistical baseline

Event-Comp: Word embedding to relatedness classifier (between two events)

RNN: similar to their implementation but taking away the DMN

Method	G&C16	C&J08
PMI	30.52	30.92
Bigram	29.67	25.43
Event-Comp	49.57	43.28
RNN	45.74	43.17
MemNet	**55.12**	**46.67**

Network Structure Analysis

-Hop: no DMN

-Attention: no DMN or attention

-LSTM: directly to DMN

-Hop, -LSTM: attention only

-Attention, -LSTM: just WE and FFN

LSTM-Only: LSTM

Method	**Acc. (%)**
MemNet	**54.36**
-Hop	52.03
-Attention	50.76
-LSTM	51.72
-Hop,-LSTM	50.65
-Attention,-LSTM	48.26
LSTM-Only	46.72

Table 2: Analysis of network structure.

References

Chambers, Nathanael, and Dan Jurafsky. 2008. “Unsupervised Learning of Narrative Event Chains.” In *Proceedings of ACL-08: HLT*, 789–797. Columbus, Ohio: Association for Computational Linguistics. http://www.aclweb.org/anthology/P/P08/P08-1090.

Chambers, Nathanael, and Dan Jurafsky. 2009. “Unsupervised Learning of Narrative Schemas and Their Participants.” In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 602–610. Suntec, Singapore: Association for Computational Linguistics. http://www.aclweb.org/anthology/P/P09/P09-1068.

Granroth-Wilding, Mark, and Stephen Clark. 2016. “What Happens Next? Event Prediction Using a Compositional Neural Network Model.” In *Thirtieth AAAI Conference on Artificial Intelligence*. https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11995.

Kumar, Ankit, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. 2015. “Ask Me Anything: Dynamic Memory Networks for Natural Language Processing.” *arXiv:1506.07285 [Cs]*, June. http://arxiv.org/abs/1506.07285.

Wang, Zhongqing, Yue Zhang, and Ching-Yun Chang. 2017. “Integrating Order Information and Event Relation for Script Event Prediction.” In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 57–67. Copenhagen, Denmark: Association for Computational Linguistics. https://www.aclweb.org/anthology/D17-1006.