Integrating Order Information and Event Relation for Script Event Prediction

Zhongqing Wang, Yue Zhang and Ching-Yun Chang

EMNLP, 2017

Disha Jindal(djjindal@seas.upenn.edu)

Feb-20-2019

Penn Engineering

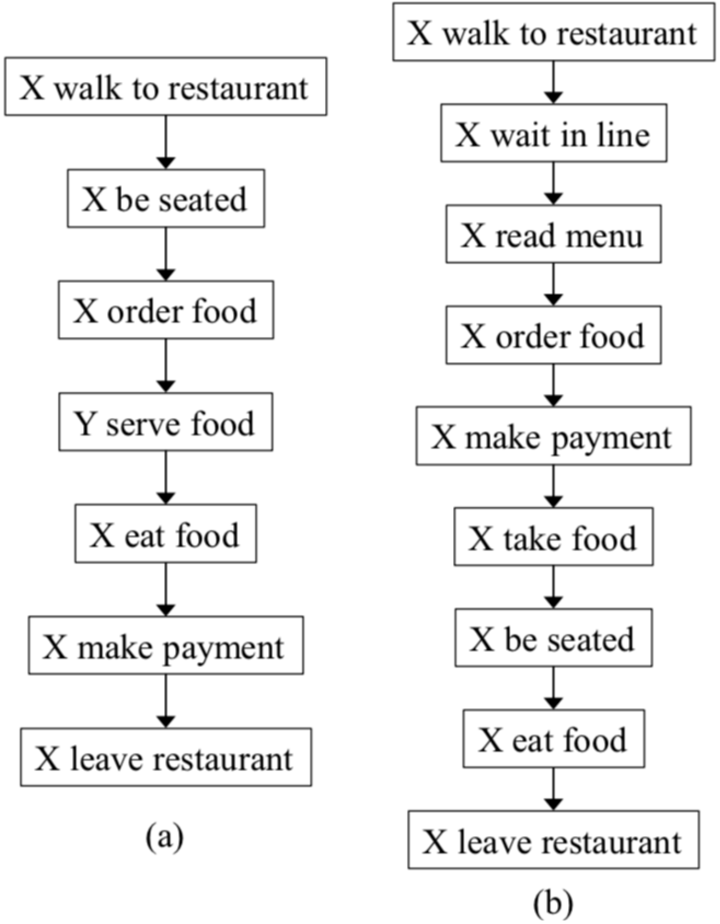
Motivation

Event chain:

Frequently recurring sequence of events with *partial orders*.

Script:

Participant Roles(“customer”, “waiter”, and “table”) + Event chain



Motivation

Challenges:

- Script knowledge is assumed to be part of the common ground.
- We do not mention events which can easily be inferred by the addressee.

"get me a piece of cake"

A text understanding system that does not have access to **script knowledge** will probably not be able to draw any inference or the series of events that took place.

Penn Engineering

Problem Space

Event Extraction and Representation

- ⟨arg_1, relation, arg_2⟩
- $v(e_s, e_o, e_p)$

Script Representation

- *Paraphrase Sets*
- *Narrative Chain*
- *Narrative Event Evolutionary Graph*

Event Modeling

(Focus of this paper)

- *Strong Order Learning*
- *Event - Pair Learning*
- *Combination*

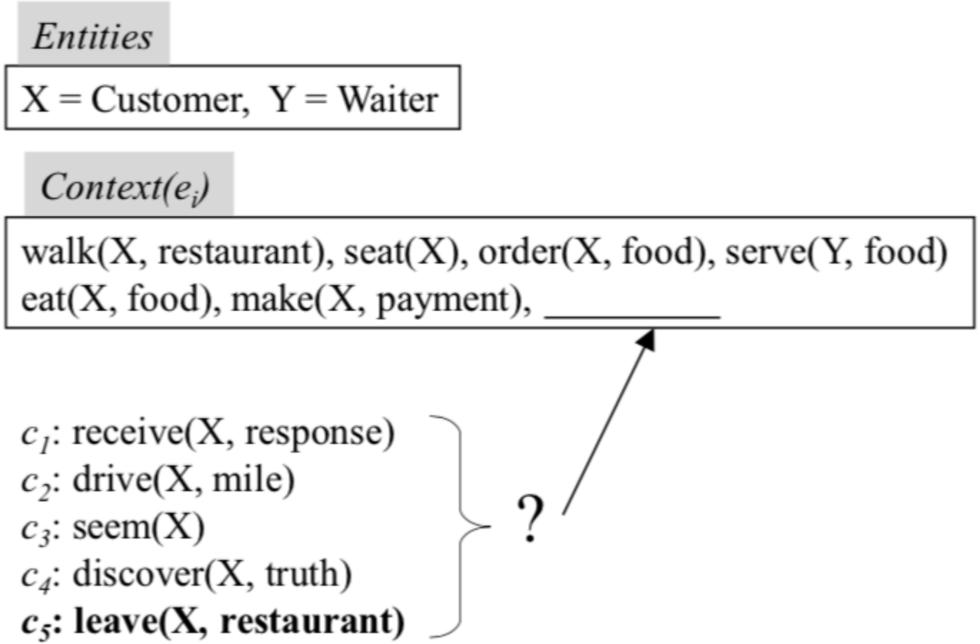
Event Evaluation

- *Narrative Cloze Test*
- *Adversarial Narrative Cloze Test*
- *MCNC Test*
- *Story Cloze Test*

Penn Engineering

Problem

Given a chain of narrative events $e_1, e_2, \ldots, e_{n-1}$ and five candidate events, the task is to predict the most likely next event e_n



Contents:

- Event and Candidate Extraction
- Previous work
- Novel aspects/contributions of this paper
- MemNet architecture
  - Event representation
  - Modeling Temporal Orders
  - Modeling Pairwise Event Relations
- Results and Analysis
  - Influence of different event structures
  - Influence of different network configurations
- Assumptions/Scope of Improvement
- Extensions/Thoughts

Event Extraction

Text: Robbers made a big score, fleeing after stealing more than $300,000 from Wells Fargo armored-truck guards who were servicing the track's ATMs, the Police Department said. The two Wells Fargo guards reported they were starting to put money in the clubhouse ATM when a man with a gun approached and ordered them to lie down. . .

Entity mentions: {*Wells Fargo armored-truck guards*, *The two Wells Fargo guards*, *they*, . . .}

Predicate events: service(x_0, ATMs), report(x_0), put(x_0, money, in clubhouse), lie+down(x_0), . . .

- POS Tagging and Dependency Parsing (C&C tools)
- Phrase Structure Parsing and Coreference Resolution(OpenNLP3)

Event Extraction:

- Each time an entity is an argument to a verb
- Predicative adjectives where an entity is an argument to the verb be or become

  John was upset ⇒ *be(x_0, upset)*

- To mitigate the over-emphasis on frequent predicates, filtered events by creating a *stopevent list.*

Candidate Extraction

Input: Event Chain, 5 randomly ordered candidates, c0,…,c4
Output: Most Likely Candidate

Primary: G&C16 (Granroth-Wilding and Clark 2016)

- New York Times portion of the Gigaword 2003 corpus
- Training , Test , Development Split: {1,500,000; 10,000;1,000}

Second benchmark: C&J08 (Chambers and Jurafsky 2008)

- News stories from 2001 corpus
- Documents: 69, Multiple choice event chains: 346

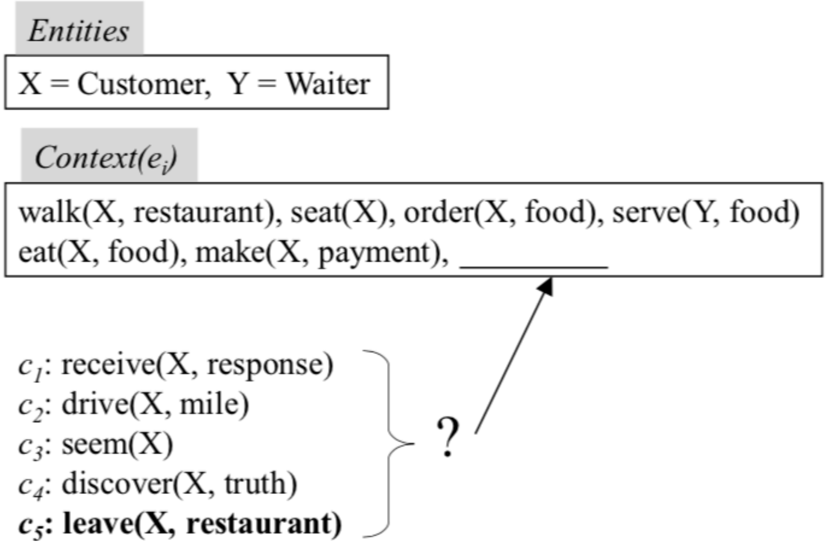
Candidate Extraction:

- 1 is observed and 4 are sampled at random from elsewhere in the corpus.
- Protagonist is replaced by the protagonist of the current chain
- Other entities are replaced by randomly chosen entities from the same document as the current chain

Previous SoTA: Event-Comp and RNN

Same event representation as that of MemNet which is $v(e_s, e_o, e_p)$

- Event-Comp
  - What Happens Next? Event Prediction Using a Compositional Neural Network Model by Granroth-Wildin and Clark 2016
  - Pair-wise modeling
  - Equal weightage to all event pairs
  - Does not consider temporal ordering of events. e.g. if, presented with (die, subj), it can suggest (live, subj) as the next event, simply because the two often co-occur



- RNN
  - Learning Statistical Scripts with LSTM Recurrent Neural Networks by Pichotta and Mooney 2016
  - Sequence/Strong-order modeling
  - Does not consider event pair relations
  - Given the flexible order of event chains, it overfits

Contributions of this work

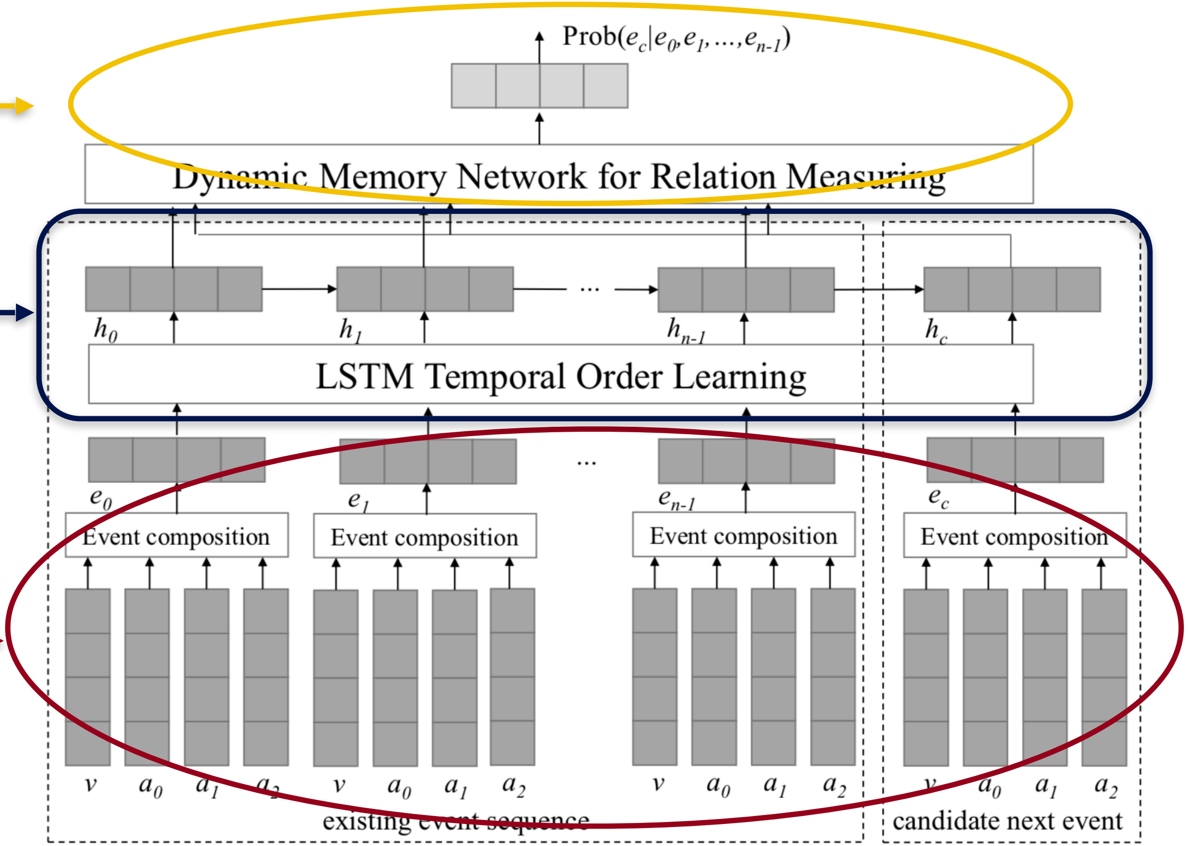
- Comparison
  - Between pair-based and sequence-based learning methods
- Paper introduces MemNet
  - A novel dynamic memory network model, which combines the advantages of both LSTM temporal order learning and traditional event pair coherence learning
- Reported best results in the standard MCNC(multi-choice narrative cloze) test

MemNet Architecture

Pair-wise relatedness score
b\w candidate event
& the context event chain

Temporal
order encoding
of events

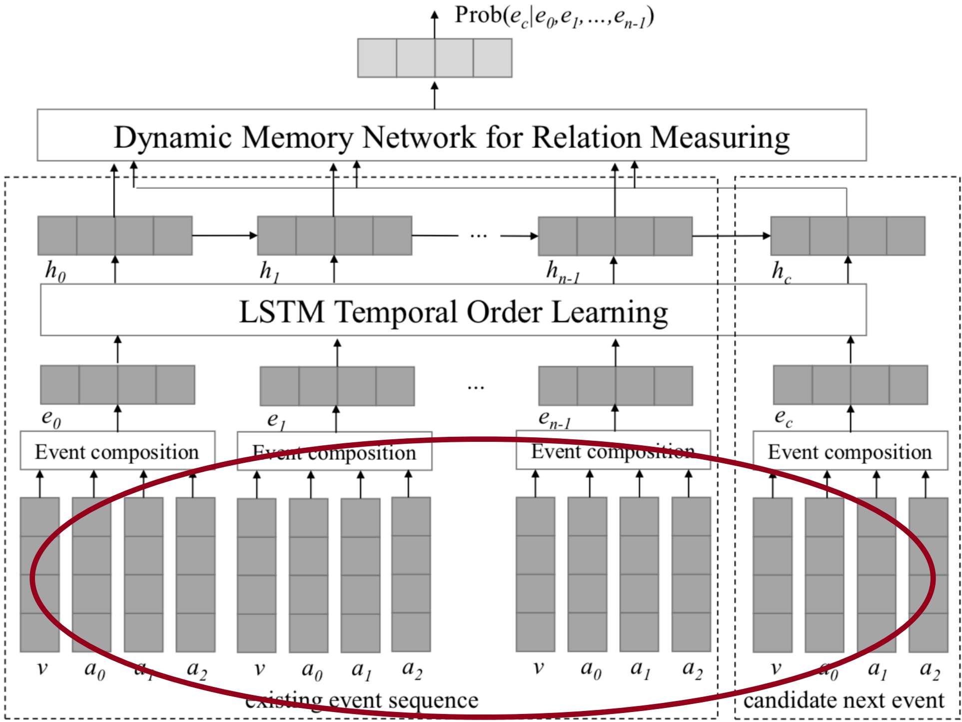
Event
Embedding
Composition



MemNet Architecture: Event Representation

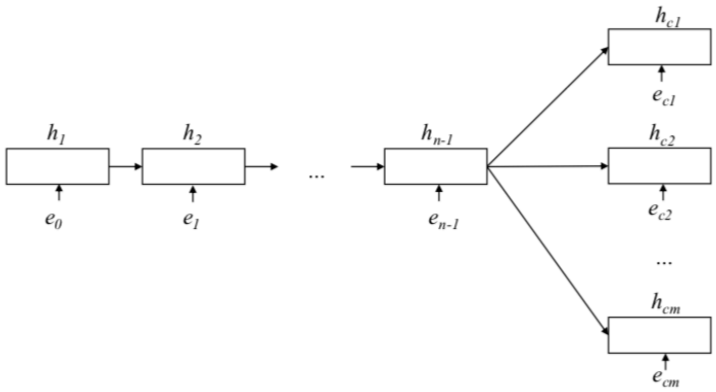
- Set of events: events $e_1, e_2, \ldots, e_{n-1}$
  - *$v(a_0, a_1, a_2)$*
  - *verb(subject, direct object, prepositional object)*
  - *bring{John, Marry, to the restaurant}*
- Word vectors are trained using the *skip-gram model*
- Event vectors are trained using *composition*

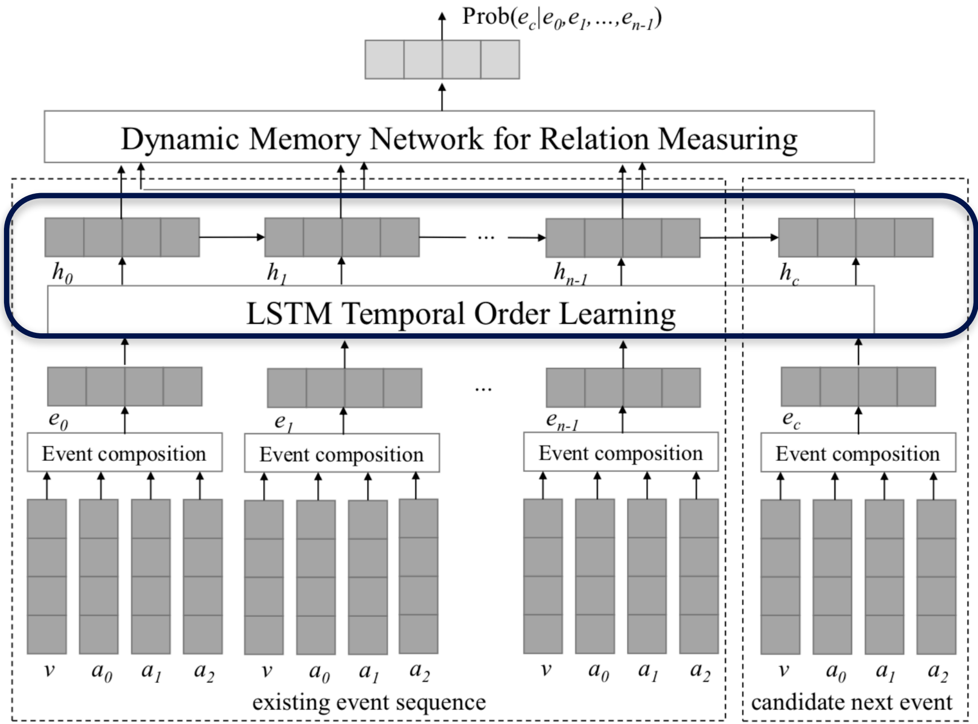
$$e(e) = \tanh(W_e^{v} \cdot e(v) + W_e^{0} \cdot e(a_0) + W_e^{1} \cdot e(a_1) + W_e^{2} \cdot e(a_2) + b_e)$$



MemNet Architecture: Modeling Temporal Orders

- Set of event candidates: $e^1_c, e^2_c, \ldots, e^m_c$
- Initial hidden state is randomly initialized
- $h_c = \mathrm{LSTM}(e(e_c), h_{n-1})$





MemNet Architecture: Modeling Pairwise Event Relations

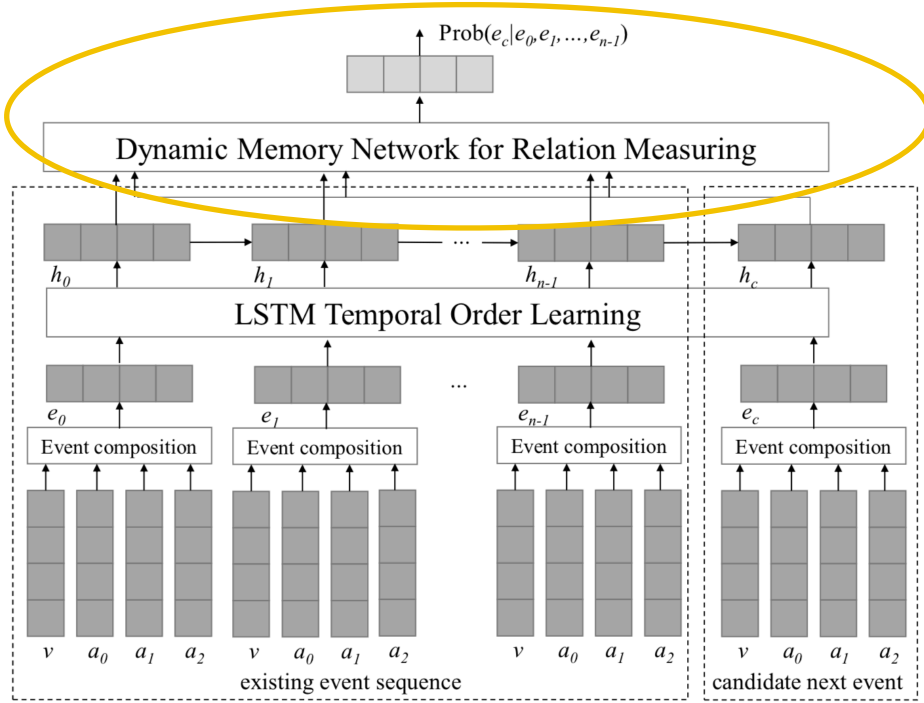
1. Siamese Network

$$s_i = \text{sigmoid}(W_{si}h_i + W_{sc}h_c + b_s),$$

$$s = \frac{\sum_{i=1}^{n-1} s_i}{n-1}$$

Problem: Equal importance to each event on the chain

Events: *"wait in queue"*, *"getting seated"* and *"order food"*
Candidate: *"eat food"*



MemNet Architecture: Modeling Pairwise Event Relations

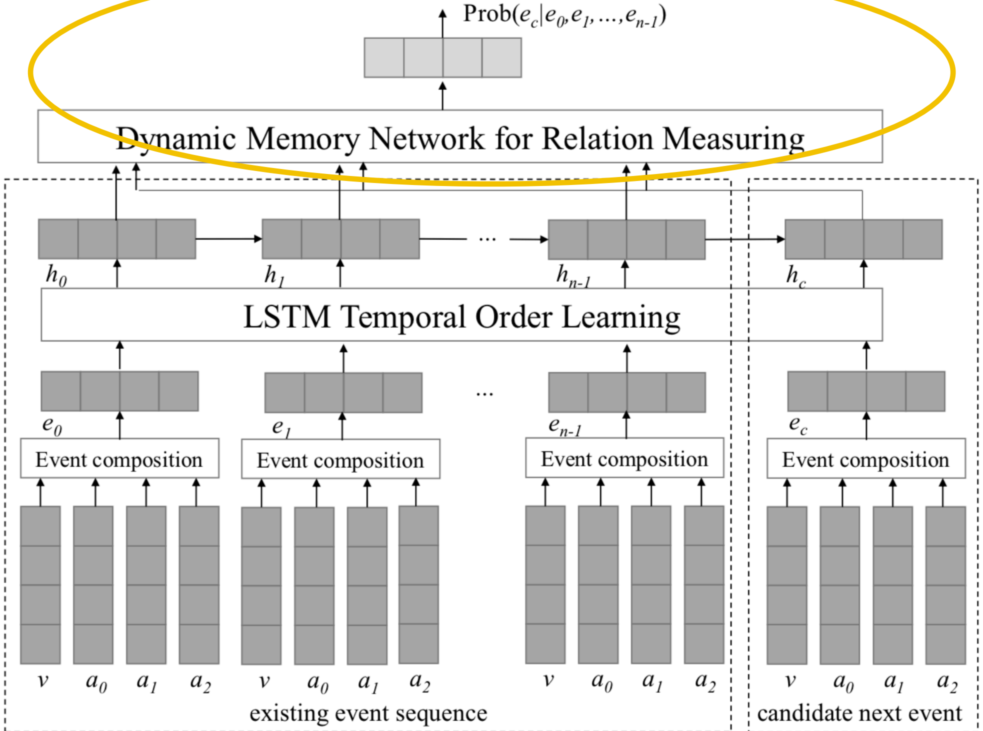
2. **Attention**: Relative importance of each existing event according to the subsequent event candidate

$$s_i = \text{sigmoid}(W_{si} h_i + W_{sc} h_c + b_s),$$

$$u_i = \tanh(W_{ei} h_i + W_c h_c + b_u)$$

$$\alpha_i = \frac{\exp(u_i)}{\sum_j \exp(u_j)}$$

$$s = \sum_{i=1}^{n-1} \alpha_i \cdot s_i$$



Penn Engineering

MemNet Architecture: Modeling Pairwise Event Relations

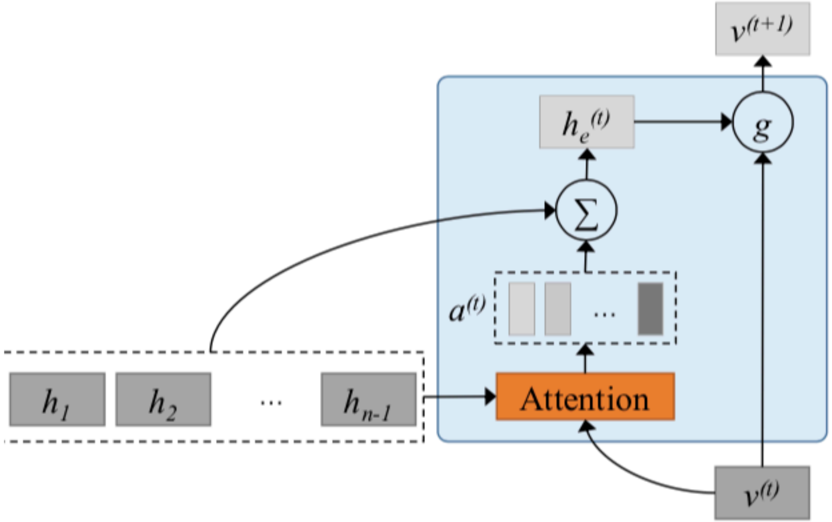
3. Deep Memory Network

- Refines event weight and event relations by recurrently modeling more abstract representations of the scenario to infer deep semantic information.
- Multiple dynamic computational layers (hops)
- Consolidated representation of context event chain is represented by h_e
- h_c and h_e are integrated to deduce a deeper representation of the full event chain hypothesis
- The intuition is that it triggers an iterative attention process which allows the model to condition its attention on the inputs and the result of previous iterations.

$$h_e = \sum_{i-1}^{n-1} \alpha_i \cdot h_i$$

$$z = \sigma(W_z h_e^t + U_z v^t)$$
$$r = \sigma(W_r h_e^t + U_r v^t)$$
$$\hat{h} = \tanh(W h_e^t + U(r \odot v^t))$$
$$v^{t+1} = (1 - z) \odot v^t + z \odot \hat{h}$$

- Convergence: $|v^{t+1} - v^t| < \mu$



Penn Engineering

Experiments: Parameters

- Optimizer: AdaGrad
- Regularization: L2
- Word vectors dimension: 300
- LSTM hidden layer size: 128
- Memory network threshold(μ) : 0.1

Results: Comparison with other models

Chambers and Jurafsky (2008): Event pair relations based on PMI

Jans et al. (2012): Event pair relations based on skip bigram probabilities.

Granroth and Clark(2016): Event pair relations based on scores using a Siamese network

Pichotta and Mooney (2016): Modeled event chains

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**

Analysis: Impact of Different Event structures

- It shows the relative importance of each component.
- It demonstrates the "*central role of the verb*" in denoting an event.

Method	Acc. (%)
MemNet	**54.36**
-*verb*	42.63
-(a_0, a_1)	52.32
-(a_0)	53.43
-(a_1)	53.57
-(a_2)	54.02

Event Structure

$v(a_0, a_1, a_2)$

Penn Engineering

Analysis: Impact of Different Modules

Influence of Temporal Order

-LSTM(**51.72**) vs MemNet(**54.36**)

-Attention, -LSTM(**48.26**) vs -Attention(**50.76**)

Influence of Event-Pair Modeling

LSTM-only(**46.72**) vs -Attention, -LSTM(**48.26**)

Influence of Attention

-Attention(**50.76**) vs -Hop(**52.03**)

-Attention, -LSTM(**48.26**) vs -Hop, -LSTM(**50.65**)

Influence of Multi-Hop Deep Memory Network

-Hop(**52.03**) vs MemNet(**54.36**)

-Hop, -LSTM(**50.65**) vs -LSTM(**51.72**)

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

Granroth and Clark(2016)

Pichotta and Mooney(2016)

Penn Engineering

Conclusions

- Calculated event pair relation using LSTM hidden states having encoded temporal orders.
- A dynamic memory network to automatically induce event weights for events.
- Outperformed SoTA event pair models and event chain models

Assumptions/Scope of Improvement

- Narrative order is same as that of temporal order
- The dataset used is noisy due to the automatic extraction process and the random sampling of confounders
  - Dataset quality assurance by human annotation
  - Random Sampling can be replaced by some sort of adversarial technique such as SWAG.
- Word embedding are trained using the *Skip-gram* algorithm. More expressive contextual models might give better results
- This paper uses Narrative chain representation for event prediction. There are other representations such as narrative event graphs which claim to capture dense connection information and semantic relations among events

Extensions/Thoughts…

What all can it predict ?

We got seated, and had to wait for 20 minutes. Then, the waiter brought the ...
We ordered, and had to wait for 20 minutes. Then, the waiter brought the ...

I ordered a medium sirloin steak with fries. Later, the waiter brought ...

- steak I had ordered,
- the steak,
- our food, or
- it.

Kevin was **robbed** by Robert, _but_ the police mistakenly **arrested** _him_.

Extensions/Thoughts…

- Do these systems model and infer structurally simpler events?
  - Verb based events enough?
  - More semantic abstraction?
    - Discourse markers
    - Disambiguation of semantic frames
    - Shallow linguistic Features, Semantic features, Script Features and Temporal Features
- To what extent do we have to provide the the explicit syntactic dependencies as a mediating representation for these event-inferring systems?
- Extrinsic Evaluation?
  - Narrative generation system.
  - Coreference Resolution
  - APT(Advanced Persistent Threats) attack: Where a user takes a sequence a actions to make a consistent attack. Might help in improving accuracy of network defense.