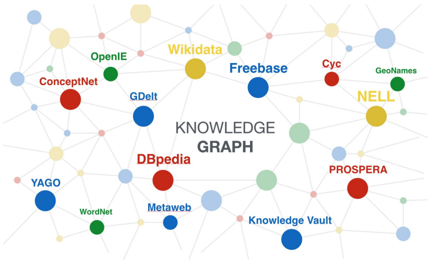
Relation Extraction with Temporal Reasoning Based on Memory Augmented Distant Supervision

Jianhao Yan

Tsinghua University

Background

- Distantly Supervised (DS) Relation Extraction



aligning

Motivation

$\langle$Angelina Jolie, Brad Pitt$\rangle$

Jolie had fallen in love with **Pitt** during filming of *Mr. & Mrs. Smith* (2005)

Pitt and **Jolie** announced their engagement in April 2012 after seven years together.

Jolie and **Pitt** were married on August 23, 2014, in a private ceremony in Château Miraval, France.

On September 19, 2016, **Jolie** filed for divorce from **Pitt**, citing irreconcilable differences.

Motivation

Dating	**Jolie** had fallen in love with **Pitt** during filming of *Mr. & Mrs. Smith* (2005)
Engagement	**Pitt** and **Jolie** announced their engagement in April 2012 after seven years together.
Marriage	**Jolie** and **Pitt** were married on August 23, 2014, in a private ceremony in Château Miraval, France.
Divorce	On September 19, 2016, **Jolie** filed for divorce from **Pitt**, citing irreconcilable differences.

Motivation

Dating

Jolie had fallen in love with **Pitt** during filming

Engag

nent in

Model has no clue how to predict relation !!!

3, 2014,

M

val,

Divorce

, 2016, **Jolie** filed for divorce

from ing irreconcilable differences.

Motivation

- 2005 — **Dating**: **Jolie** had fallen in love with **Pitt** during filming of *Mr. & Mrs. Smith* (2005)
- 2012 — **Engagement**: **Pitt** and **Jolie** announced their engagement in April 2012 after seven years together.
- 2014 — **Marriage**: **Jolie** and **Pitt** were married on August 23, 2014, in a private ceremony in Château Miraval, France.
- 2016 — **Divorce**: On September 19, 2016, **Jolie** filed for divorce from **Pitt**, citing irreconcilable differences.

Task Definition

- Introduce temporal information in both the **relation labeling** and the **mention set**.

Traditional DS:

$$\mathrm{P}(r \mid S = \{s_1, s_2, \ldots, s_T\})$$

DS with Temporal Reasoning:

$$\mathrm{P}\left(r_{t_i} \middle| S = \{(s_1, t_1), (s_2, t_2), \ldots, (s_T, t_T)\}, t_i\right)$$

Where $t_1 \leq t_2 \leq \ldots \leq t_T$

Task Definition

- Introduce temporal information in both the **relation labeling** and the **mention set**.

Traditional DS:

$$P(r \mid S = \{s_1, s_2, \dots, s_T\})$$

DS with Temporal Rea…

$$P(r_{t_i} \mid S = \{(s_1, t_1), (s_2, t_2), \dots, (s_T, t_T)\}, t_i)$$

Where $t_1 \leq t_2 \leq \dots \leq t_T$

New dataset WIKI-TIME!

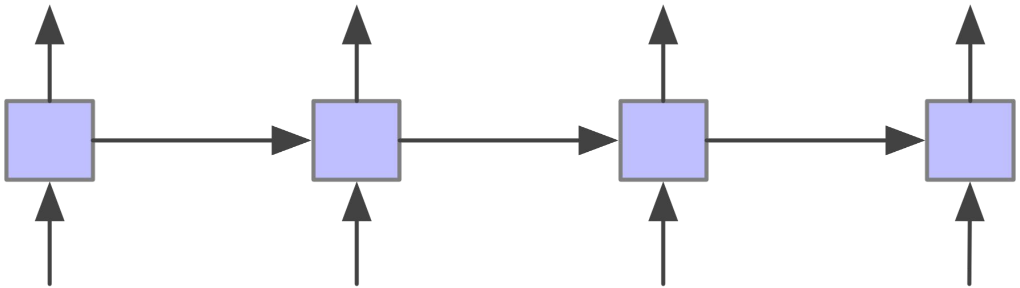
Method

$$P\left(r_{t_i} \middle| S = \{(s_1, t_1), (s_2, t_2), \ldots, (s_T, t_T)\}, t_i\right)$$

$$\text{Where } t_1 \le t_2 \le \ldots \le t_T$$

- The task is similar to the classical sequence labeling problem.

- So LSTM or LSTM + CRF?



Method

$$P(r_{t_i} \mid S = \{(s_1, t_1), (s_2, t_2), \ldots, (s_T, t_T)\}, t_i)$$

$$\text{Where } t_1 \leq t_2 \leq \ldots \leq t_T$$

- The task is similar to the classical sequence labeling task.

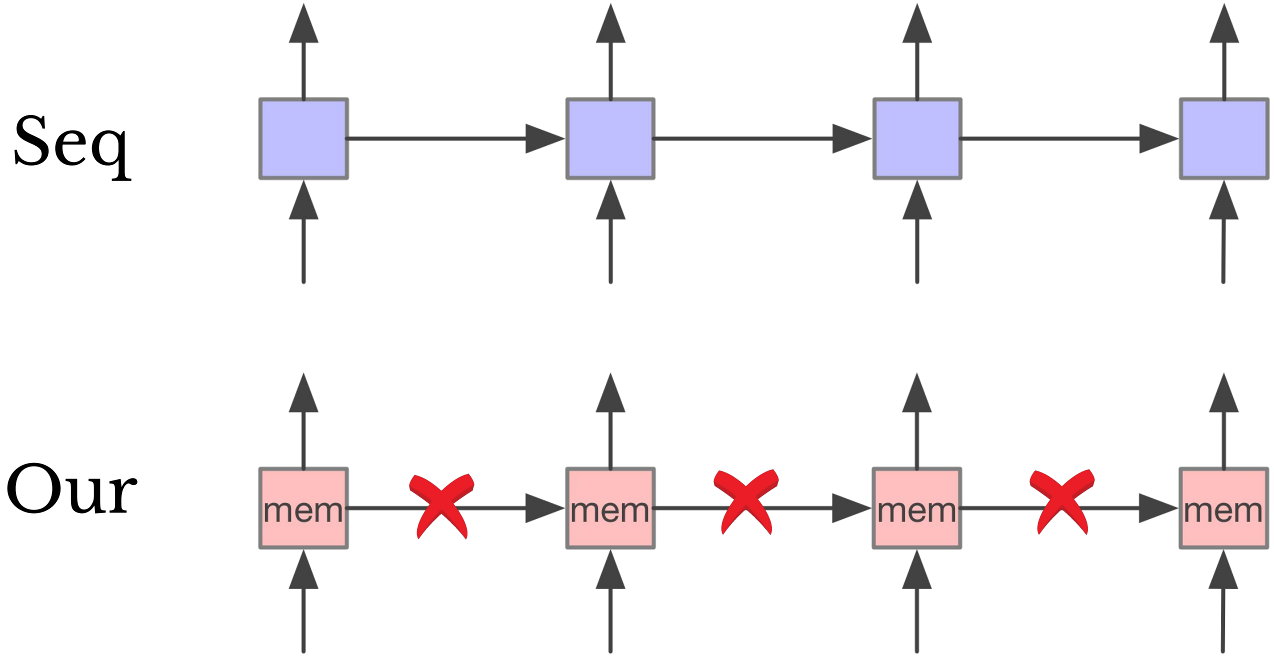
- So LSTM or LSTM + CRF?

No ! Noisy input will destroy sequence model !

Method

$$P\left(r_{t_i} \middle| S = \{(s_1, t_1), (s_2, t_2), \dots, (s_T, t_T)\}, t_i\right)$$

$$\text{Where } t_1 \leq t_2 \leq \dots \leq t_T$$



Motivation

Sentence-level memory nets.

- Make use of *supporting instances*.
  - In traditional DS, models always use attention and other techniques to denoise.
  - However, there are sentences that are not direct positive examples for the given relation, but can provide supporting evidence

Method

Alleviate the **hard** order dependency effect in sequence model.

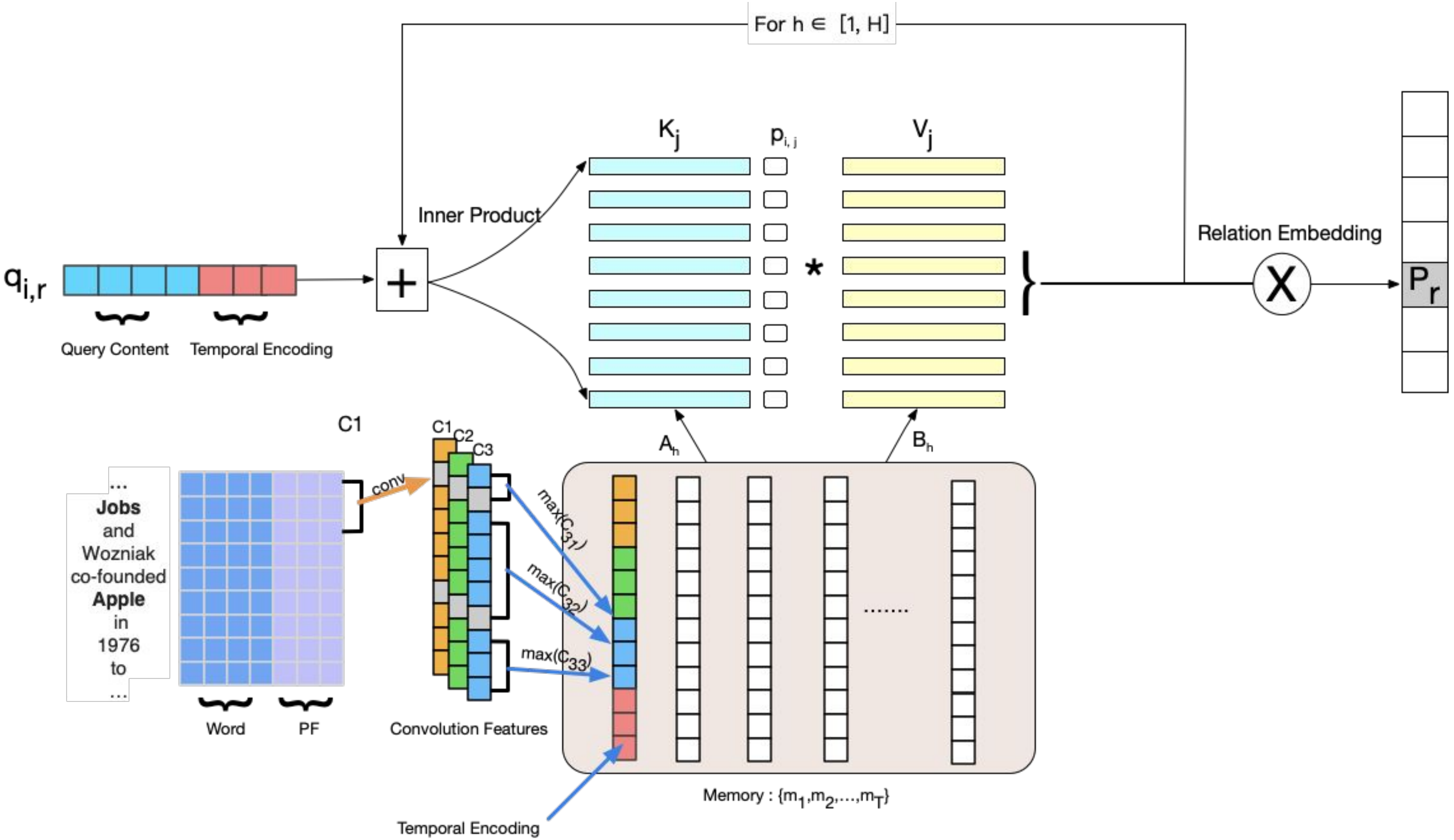
- Construct a query sequence on each mention time spot of mention set.
- Use Memory Nets to introduce sentence-level temporal reasoning.

Method

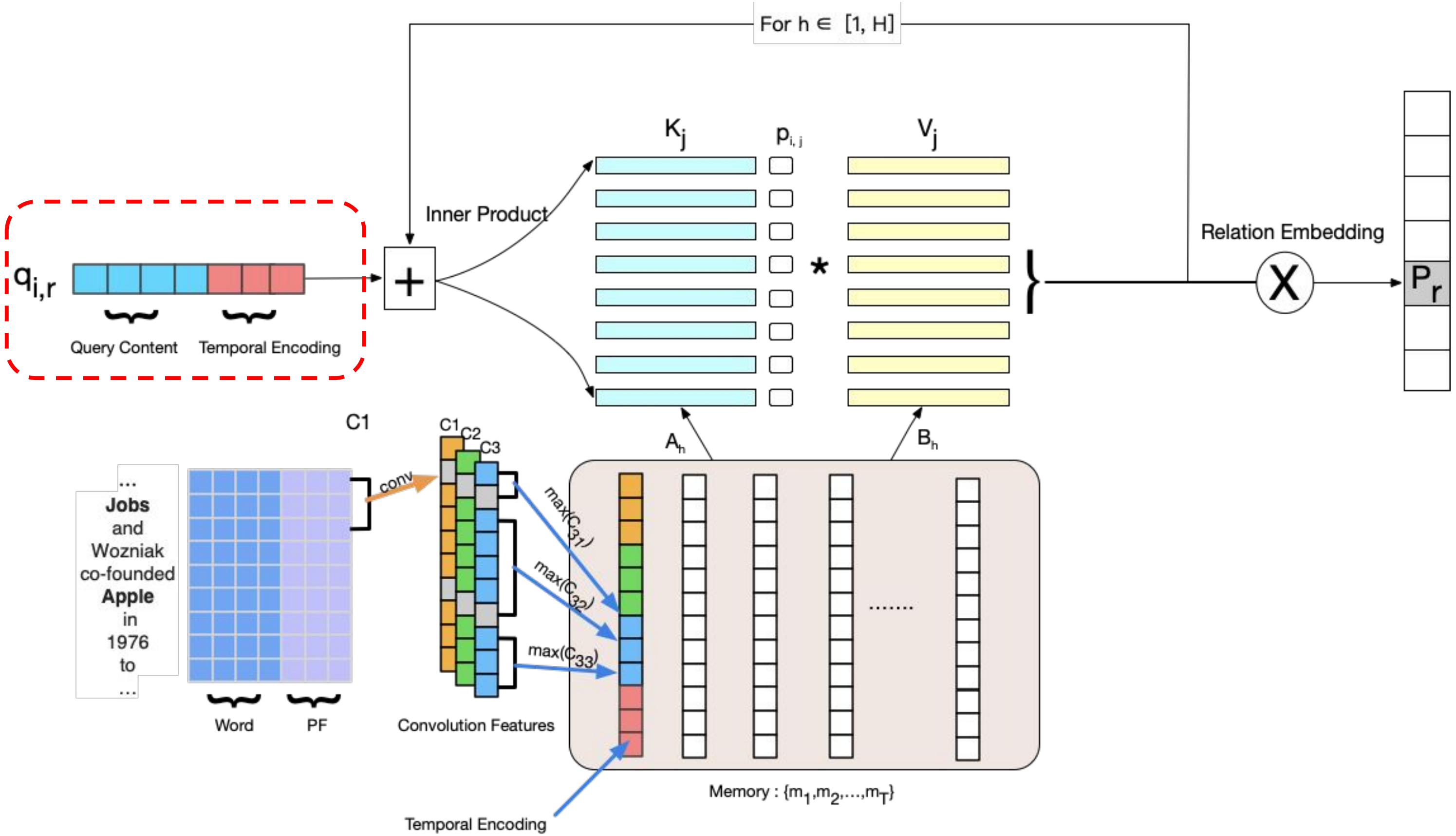
Alleviate the hard order dependency effect in sequence model.

- **Construct a query sequence on each mention time spot of mention set.**
- Use Memory Nets to introduce sentence-level temporal reasoning.

Method


For h ∈ [1, H]
K_j
$p_{i,j}$
V_j
Inner Product
Relation Embedding
$q_{i,r}$
Query Content
Temporal Encoding
P_r
C1
C1
C2
C3
conv
max(C_{31})
max(C_{32})
max(C_{33})
A_h
B_h
...
Jobs
and
Wozniak
co-founded
Apple
in
1976
to
...
Word
PF
Convolution Features
Memory : {m_1,m_2,...,m_T}
Temporal Encoding

Method



For h ∈ [1, H]
K_j
$p_{i,j}$
V_j
Inner Product
*
Relation Embedding
X
P_r
$q_{i,r}$
Query Content
Temporal Encoding
C1
C1
C2
C3
conv
...
Jobs
and
Wozniak
co-founded
Apple
in
1976
to
...
Word
PF
Convolution Features
$\max(C_{31})$
$\max(C_{32})$
$\max(C_{33})$
A_h
B_h
....
Memory : $\{m_1, m_2, \ldots, m_T\}$
Temporal Encoding

Temporal Encoding (TE)

- Several constraints for TE.
  - Should comply with the chronological order of instances.
  - Encoding similarity is only decided by the difference between two time spots.

$$PE(j) = \begin{cases} sin(j/10000^{d/d_m}) & if\ d\%2 = 0 \\ cos(j/10000^{(d-1)/d_m}) & if\ d\%2 = 1 \end{cases}$$

[Vaswani et al., 2017]

Query Construction

- 4 key variables for RE. $(relation, e_1, e_2, t)$

Query Construction

- 4 key variables for RE. $(relation, e_1, e_2, t)$
- We split each query into content and temporal encoding.

$$q_r = R_r + (E_{e_1} + E_{e_2}) * \Phi_q$$

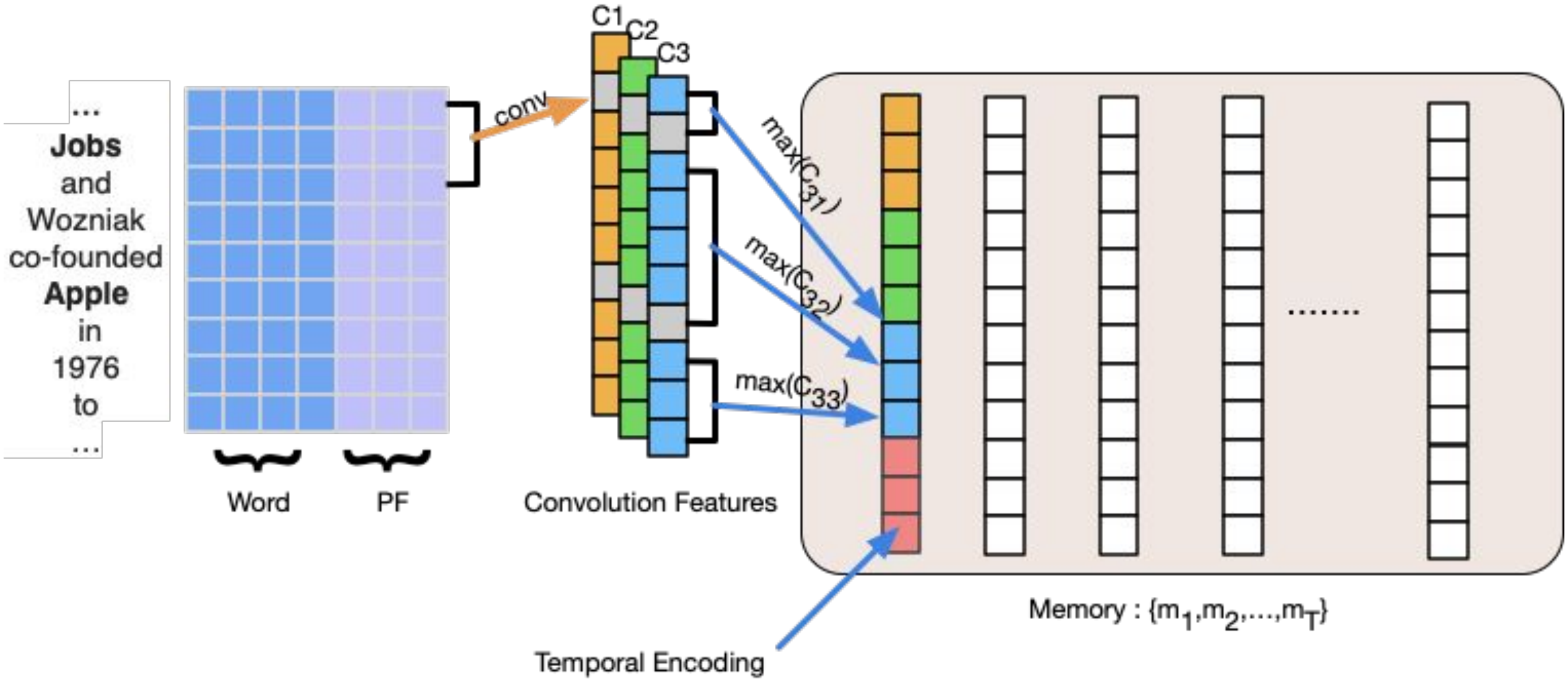
$$q_{r,i} = [q_r; \lambda \cdot TE(i)]$$

Method

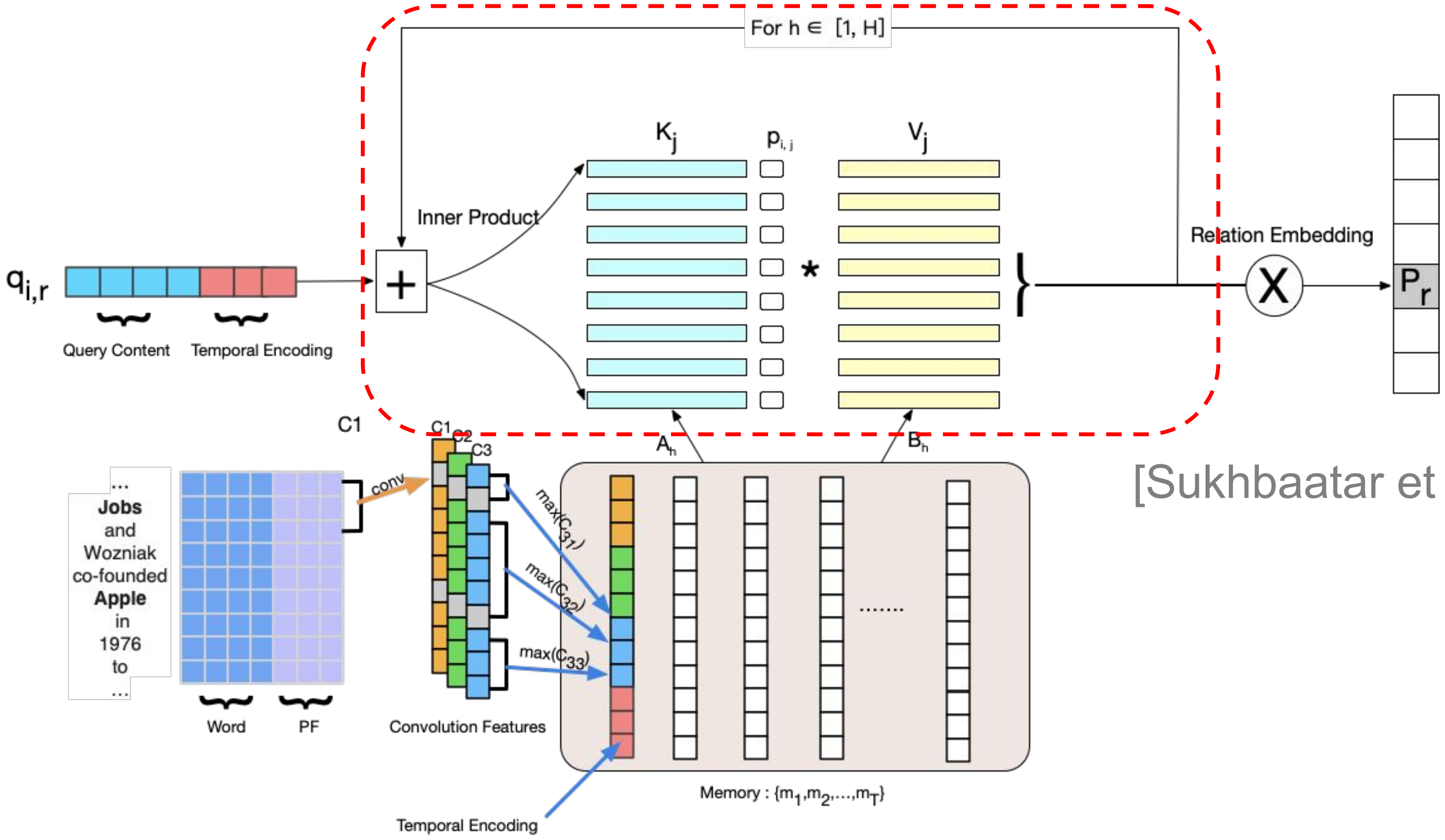
Alleviate the hard order dependency effect in sequence model.

- Construct a query sequence on each mention time spot of mention set.
- **Use Memory Nets to introduce sentence-level temporal reasoning.**

Memory Encoding



Method



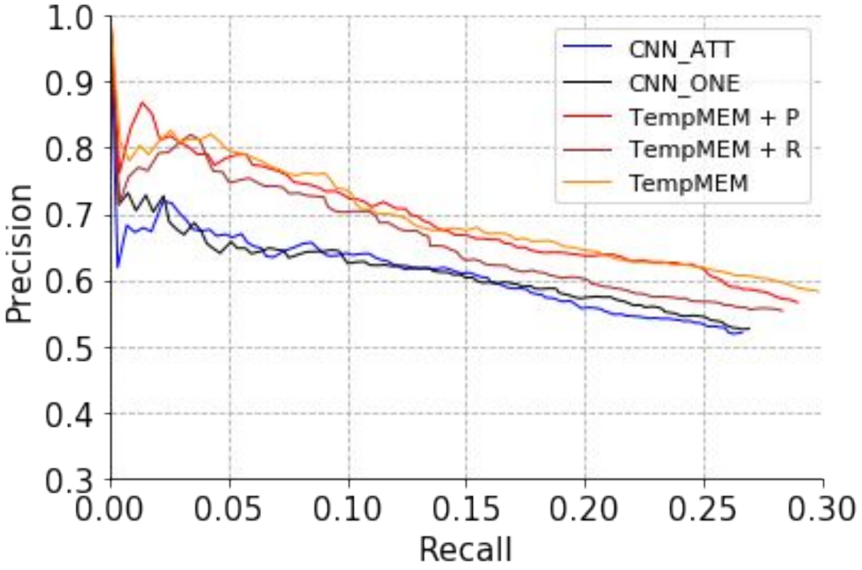
[Sukhbaatar et al., 2015]

Optimization

- Query-level Cross Entropy by SGD.

$$J(\theta) = \sum_{s=1}^{N_s} \sum_{i=1}^{T} y_t \cdot \log p(\hat{y}_t | S_s, \theta, t_i)$$

Experiment Results(WIKI-TIME)



PR curves on WIKI-TIME

Method	P@N_100	P@N_200	P@N_300
CNN_ATT	67.33	67.66	66.45
CNN_ONE	70.3	68.66	65.78
TempMEM	81.18	**82.09**	**78.41**
TempMEM+R	79.21	78.61	75.42
TempMEM+P	**81.19**	79.1	77.41

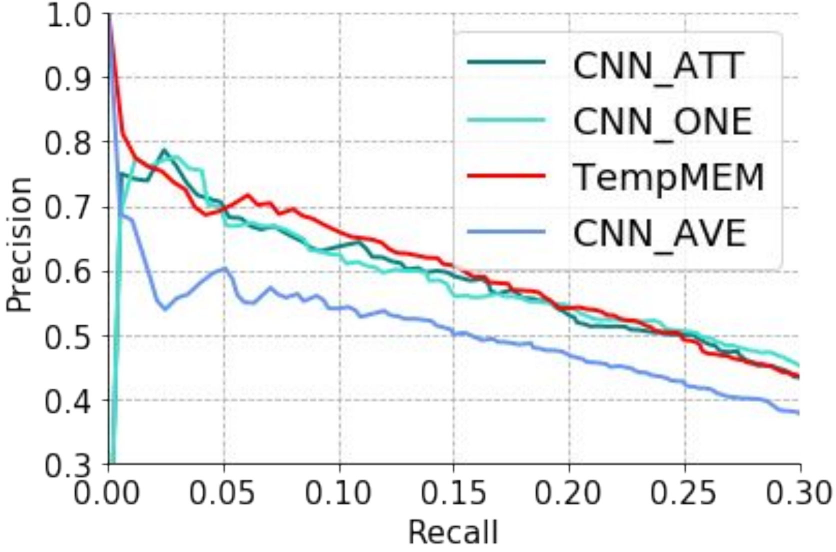
Automatic results on WIKI-TIME

Method	Bag-level F1	Query-level F1
CNN_ATT	39.66	-
CNN_ONE	40.15	-
TempMEM	47.88	54.75
TempMEM+R	46.76	47.83
TempMEM+P	**54.86**	**60.01**

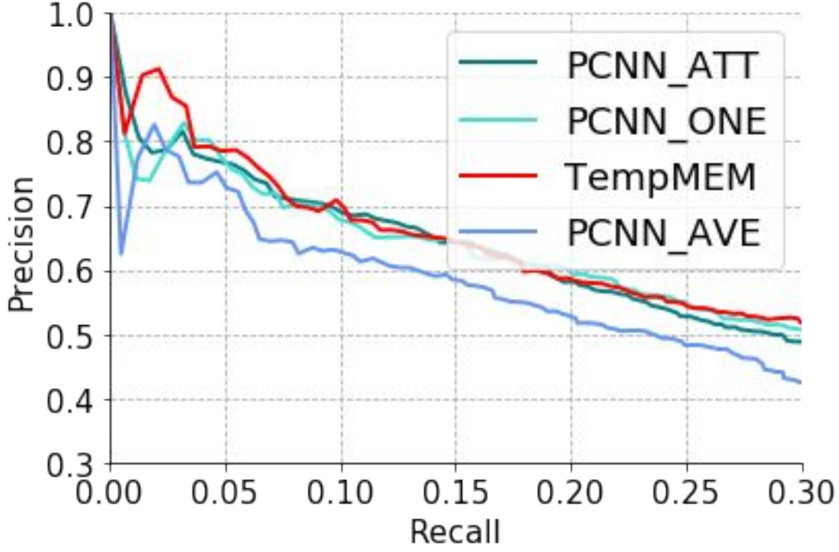
Manual results on WIKI-TIME

Experiment Results(NYT-10)

NO TEMPORAL INFO!



PR curves with CNN



PR curves with PCNN

To Sum Up

- Temporal reasoning task
  - Newly developed dataset WIKI-TIME
- TempMEM model with temporal encoding and sentence-level reasoning
- Experimental results on WIKI-TIME & NYT-10 prove our model achieves better performance.

Thanks.

elliottyan37@gmail.com