Memory Networks for Relation Extraction

曾双

2020.4.12

IJCAI 2017

Effective Deep Memory Networks for Distant Supervised Relation Extraction

Xiaocheng Feng, Jiang Guo, Bing Qin, Ting Liu, Yongjie Liu
SCIR, Harbin Institute of Technology, Harbin, China
{xcfeng, jguo, qinb, tliu, yongjieliu}@ir.hit.edu.cn

Knowledge Base Triples:

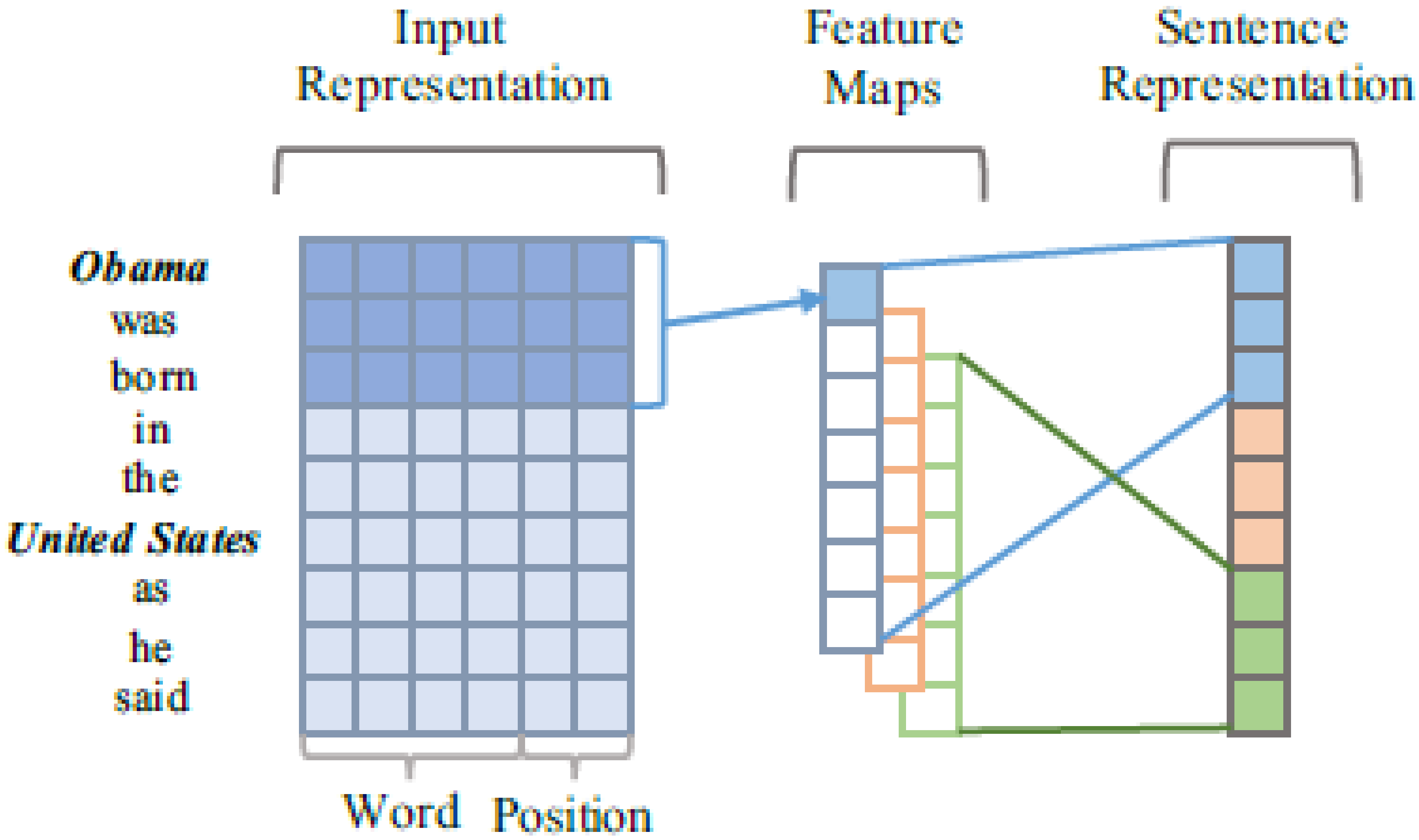
- Barack Obama, *EmployedBy*, United States
- Barack Obama, *BornIn*, United States

Latent Label	Sentence
EmployedBy	S1: United States President Barack Obama meets with NBA player LeBron James Today.
BornIn	S2: Obama was born in the United States just as he has always said.
---	S3: Obama ran for the united States Senate in 2004.

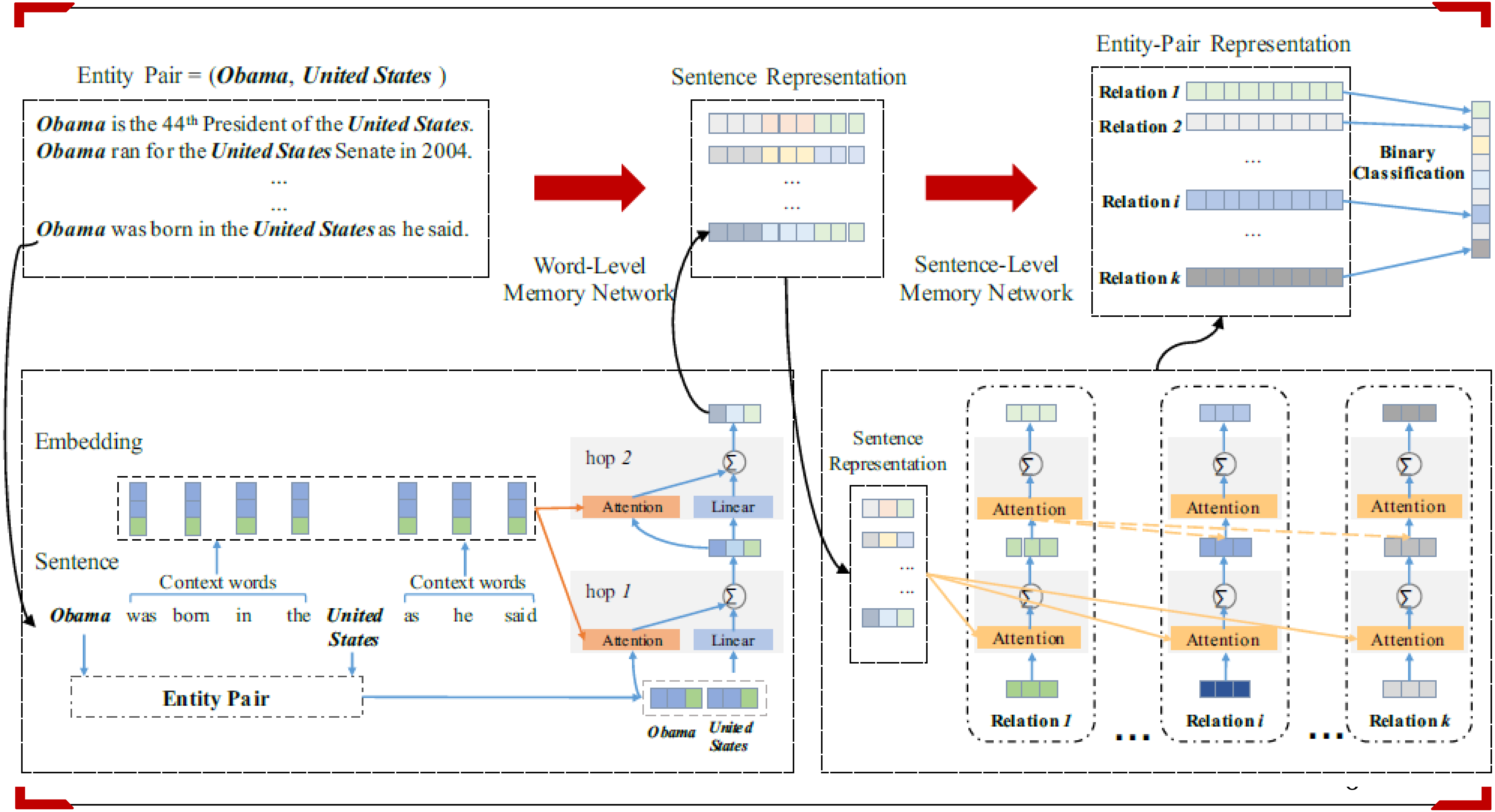
Figure 1: Training sentences generated through distant supervision for a knowledge base containing two facts.

Motivation

- Not all context words contribute equally to the inference of relation for an entity pair.
- There exists dependencies (e.g., entailment, conflict) between different relations, which is a crucial cue to infer some instances with implicit relation expression. For instance, if triple (A, capital, B) holds, another triple (A, contains, B) will hold as well.



Input Representation
Feature Maps
Sentence Representation
Obama
was
born
in
the
United States
as
he
said
Word
Position



Entity Pair = (Obama, United States)
Obama is the 44th President of the United States.
Obama ran for the United States Senate in 2004.
...
...
Obama was born in the United States as he said.
Word-Level Memory Network
Sentence Representation
Sentence-Level Memory Network
Entity-Pair Representation
Relation 1
Relation 2
Relation i
Relation k
Binary Classification
Embedding
Sentence
Context words
Context words
Obama was born in the United States as he said
Entity Pair
hop 2
hop 1
Attention
Linear
Obama
United States
Sentence Representation
Attention
Relation 1
Relation i
Relation k

Word Level Attention

$$g_i = \tanh(W_{word-att}[m_i; w_{eh}; w_{et}] + b_{word-att})$$

$$\alpha_i = \frac{\exp(g_i)}{\sum_{j=1}^{w} \exp(g_j)}$$

$$x = \sum_{i=1}^{w} \alpha_i m_i$$

Selective-Attention over Instances

$$z_i = x_i A v_{r_j}$$

$$\beta_i = \frac{\exp(z_i)}{\sum_{p=1}^{n} \exp(z_p)}$$

$$R_j = \sum_{i=1}^{n} \beta_i x_i$$

Selective-Attention over Relations

$$h_i = R_i B R_j$$

$$\gamma_i = \frac{\exp(h_i)}{\sum_{q=1}^{k} \exp(h_q)}$$

$$R_j^* = \sum_{i=1}^{k} \gamma_{ji} R_i$$

Feature-based methods

(1) **Mintz**: [Mintz *et al.*, 2009] proposed distant supervision paradigm and developed a multi-class logistic regression for classification. (2) **Multir** is a multi-instance learning method that was proposed by [Hoffmann *et al.*, 2011] with a deterministic "at-least-one" decision . (3) **MIML** [Surdeanu *et al.*, 2012] is a multi-instance multi-label approach for distant supervision using a graph model.

Neural-based methods

(1) **PCNN** [Zeng *et al.*, 2015] is a convolutional neural network based method for relation extraction. This method models overlapping relations by combining sentence-level relation extraction features into entity-pair-level results. (2) **ATT**: [Lin *et al.*, 2016] pointed out that distant supervision suffers from the entity pair wrong labeling problem. They developed a sentence-level attention model which can dynamically reduce the weights of those noisy instances and achieves state-of-the-art results.

Dataset	Sentences	Pos EPs	Neg EPs	relations
Training	112,941	4,266	61,460	26
Testing	152,416	1,732	91,842	26

Table 1: Statistics of the filtered NYT10 dataset, where EP denotes entity pair.

	Top 100	Top 200	Top 500	Average
Mintz	0.77	0.71	0.55	0.676
Multir	0.83	0.74	0.59	0.720
MIML	0.85	0.75	0.61	0.737
PCNN	0.84	0.77	0.64	0.750
ATT	0.86	0.80	0.68	0.780
DMN	**0.89**	**0.82**	0.68	**0.797**

Table 2: Precision values for the top 100, top 200, and top 500 extracted relation instances.

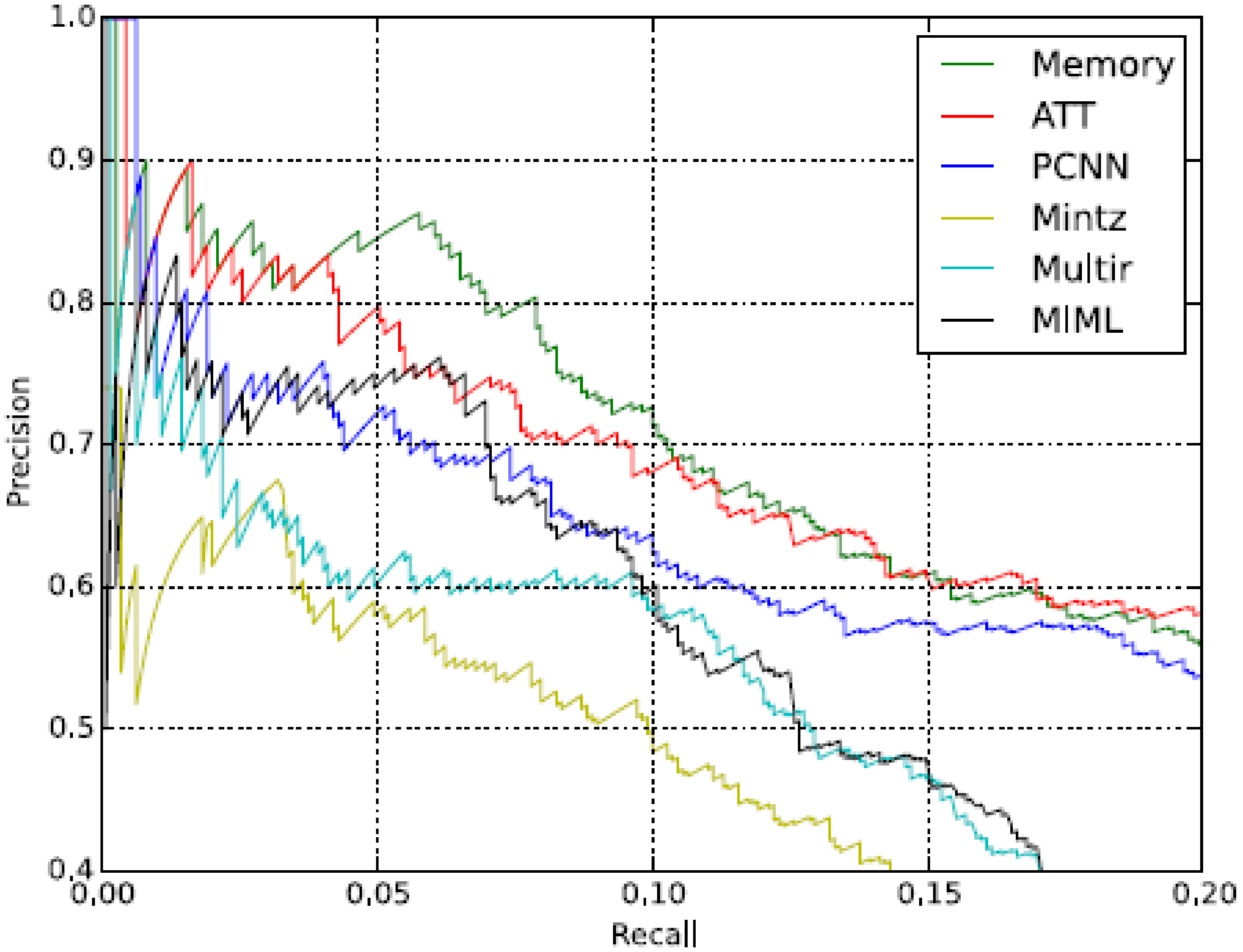


Figure 4: Precision-recall curves of various methods.

NAACL 2019

Relation Extraction with Temporal Reasoning Based on Memory Augmented Distant Supervision

Jianhao Yan[1], Lin He[1], Ruqin Huang[2], Jian Li[3]*, and Ying Liu[1]
[1]Institute for Network Sciences and Cyberspace, Tsinghua University
[2]Department of Computer Science and Technology, Tsinghua University
[3]Institute for Interdisciplinary Information Sciences, Tsinghua University
{yjh16, he-l14, hrq16}@mails.tsinghua.edu.cn
lijian83@tsinghua.edu.cn, liuying@cernet.edu.cn

Motivation

- DS RE : encoding and fusion
- The former encodes each instance into a low-dimensional representation.
- The latter combines representation of each instance. Then, their combination is used to predict the relation.

Motivation

- They all use a separate but identical encoding module among instances and introduce no difference temporally.
- They only adopt single step of fusion and introduce no sentence-level reasoning.

Motivation

- Angelina Jolie and Brad Pitt (using Wikidata). The knowledge base contains a factual relation of spouse between them with the valid period from August 2014 to September 2016.
- However, the extracted mention set contains instances about their marriage in 2014, as well as their divorce in 2016.
- Existing models may predict the relation of marriage since the instances may suggest a higher confidence for the relation of marriage.

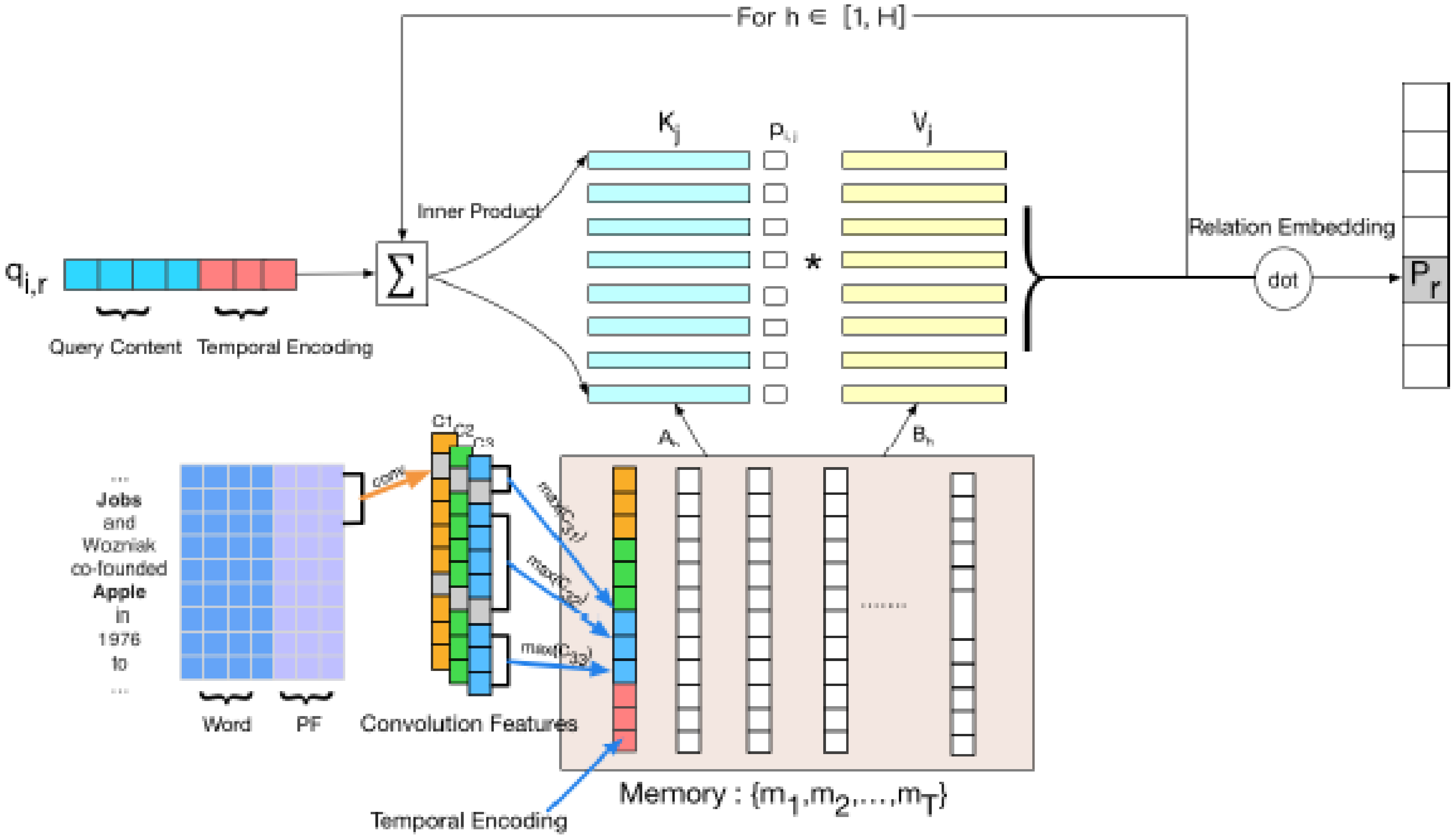


Figure 1: Overall TempMEM architecture

- The temporal encodings should comply with the chronological order of instances.
- The difference between two time spots determines the similarity between two temporal encodings.

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

$$\boldsymbol{m}_j = [\boldsymbol{O}_j; \lambda \cdot \boldsymbol{PE}(j)].$$

Query Construction

We construct each query with the guidance of the following intuition.

- Relation extraction within instances is equal to the query "what is the *relation* between *head* and *tail* at time spot t_i?".

$$\boldsymbol{q}_r = \boldsymbol{R}_r + (\boldsymbol{E}_{head} + \boldsymbol{E}_{tail}) * \boldsymbol{\Phi}_q,$$

$$\boldsymbol{q}_{r,i} = [\boldsymbol{q}_r; \lambda \cdot \boldsymbol{PE}(i)].$$

Memory Addressing In addressing, we compute the similarity between the query vector $\boldsymbol{q}_{i,r}$ and each candidate memory slot key $\boldsymbol{K}_j$. Note that the encoding output $\boldsymbol{m}_j$ is not in the same continuous space as the query vector. So, we adopt linear projections to both memory keys:

$$\boldsymbol{K}_j = \boldsymbol{A}_h^T \cdot \boldsymbol{m}_j, \tag{10}$$

where $\boldsymbol{A}_h \in \mathbb{R}^{D_m * D_r}$. Then, we compute the similarity score and importance probability using the bilinear form,

$$s_{i,j} = q_{i,r}^T \cdot \boldsymbol{W}_a \cdot \boldsymbol{K}_j, \tag{11}$$

$$p_{i,j} = \frac{exp(s_{i,j})}{\sum_{\hat{j}=1}^{M} exp(s_{i,\hat{j}})}, \tag{12}$$

Memory reading The value of each memory slot, which is also projected by an affine matrix $\boldsymbol{B} \in \mathbb{R}^{D_m * D_r}$, is read by computing the weighted sum over all memory slots with the importance probability derived in the addressing step:

$$\hat{q}_i = \sum_j p_{i,j} \boldsymbol{V}_j, \tag{14}$$

where $\boldsymbol{V}_j = \boldsymbol{B}_h^T \cdot \boldsymbol{m}_j$.

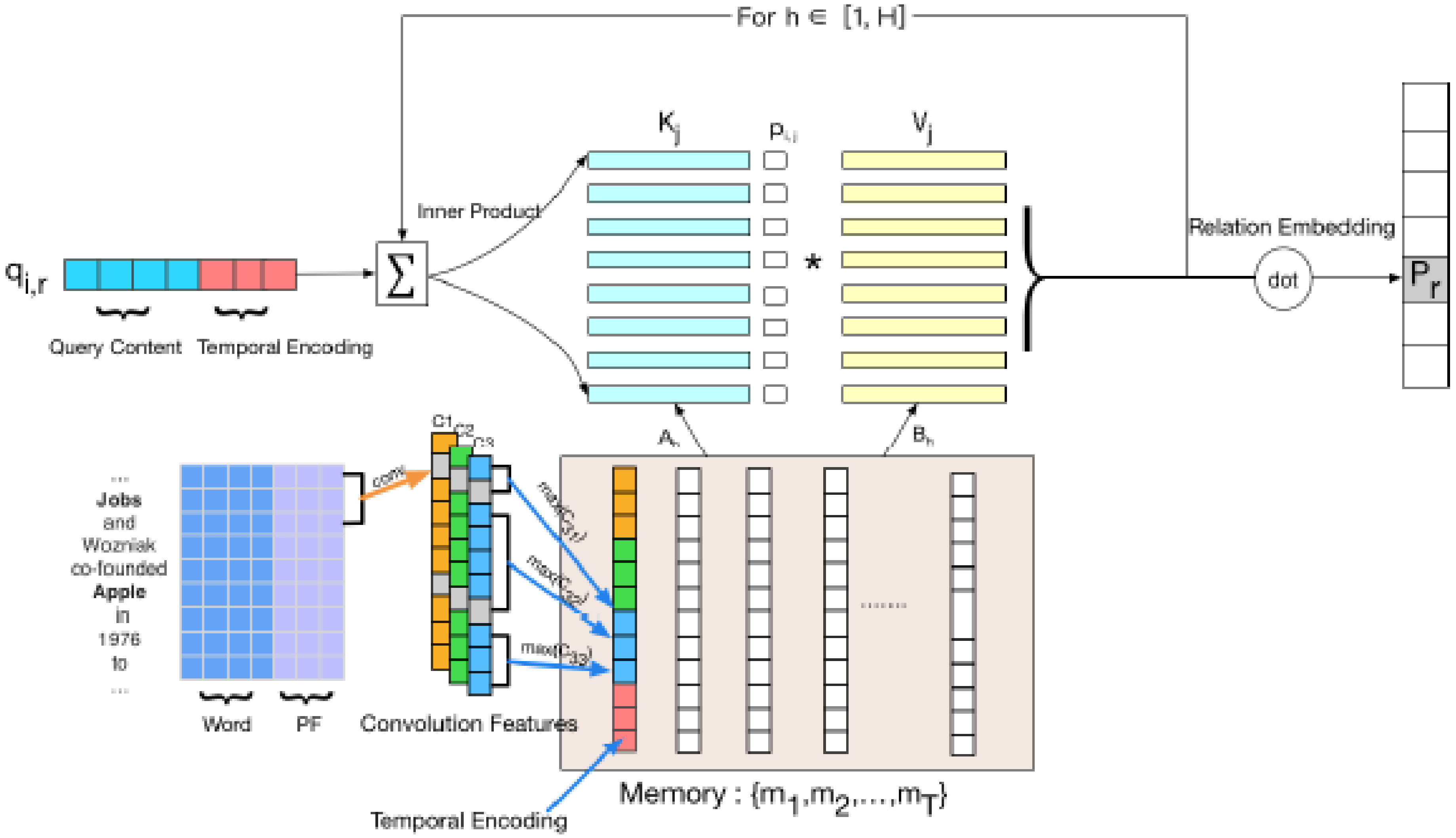


Figure 1: Overall TempMEM architecture

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

Table 2: Comparison with previous models. P@N_100/200/300 refers to the precision for the highest 100, 200 and 300 predictions in 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**

Table 3: Manual evaluation of Bag-level and Query-level F1 scores in WIKI-TIME.

id	Relation	Time Spot	Sentence
0	NA	1957-01-01	Her early career was marked by her collaboration with singer **Stelios Kazantzidis**.
1	NA	1960-01-01	... instances of **Marinella** in films of Greek cinema, from the 1960 by 1966 with **Stelios Kazantzidis** ...
2	Spouse	1964-05-07	**Marinella** married **Stelios Kazantzidis** on 7 May 1964 ...
3	Spouse	1964-05-07	**Stelios Kazantzidis** married **Marinella** on 7 May 1964 ...
4	NA	1966-09-01	In September 1966 he divorced **Marinella**...
5	NA	1968-01-01	Following **Marinella**'s departure Litsa Diamandi ...
6	NA	1968-01-01	**Marinella** sang on some songs ...
7	NA	1968-01-01	**Marinella** had an "answer back" to that latter song ...

Table 4: Aligned sentences of ⟨ Stelios Kazantzidis , Marinella ⟩

Thanks!