An Attention-based Model for Joint Extraction of Entities and Relations with Implicit Entity Features

ADVISOR: JIA-LING, KOH

SOURCE: WWW 2019

SPEAKER: SHAO-WEI, HUANG

DATE: 2019/09/20

OUTLINE

- Introduction
- Method
- Experiment
- Conclusion

INTRODUCTION

- Extract entities and their semantic relations from an unstructured input sentence.

(Ex):
"Donald Trump is the 45th and current president of the United States"

(Donald Trump, President-Of, United States)

Eentiy 1 Semantic relation Eentiy 2

INTRODUCTION

Two categories method of RE

- **Pipelined models：**
  - Identify the entity pair first. And then predict the relations between them.

(Donald Trump, United States) → (President-Of)

Result have an impact on it

- **Joint models(this paper use)：**
  - Identify the entity pair and relations at the same time.

INTRODUCTION

Problem definition

Input sentence:	Donald	Trump	is	the	45th	and	current	President	of	the	United	States
Tags:	B-PR-1	E-PR-1	O	O	O	O	O	O	O	O	B-PR-2	E-PR-2

Final Results: (Donald Trump, President-Of, United States)

B：Begin

(Have Begin, Inside, End, Single four type)

PR：President-Of

(Predfined 24 relation type)

1：First entity

Other

E：End

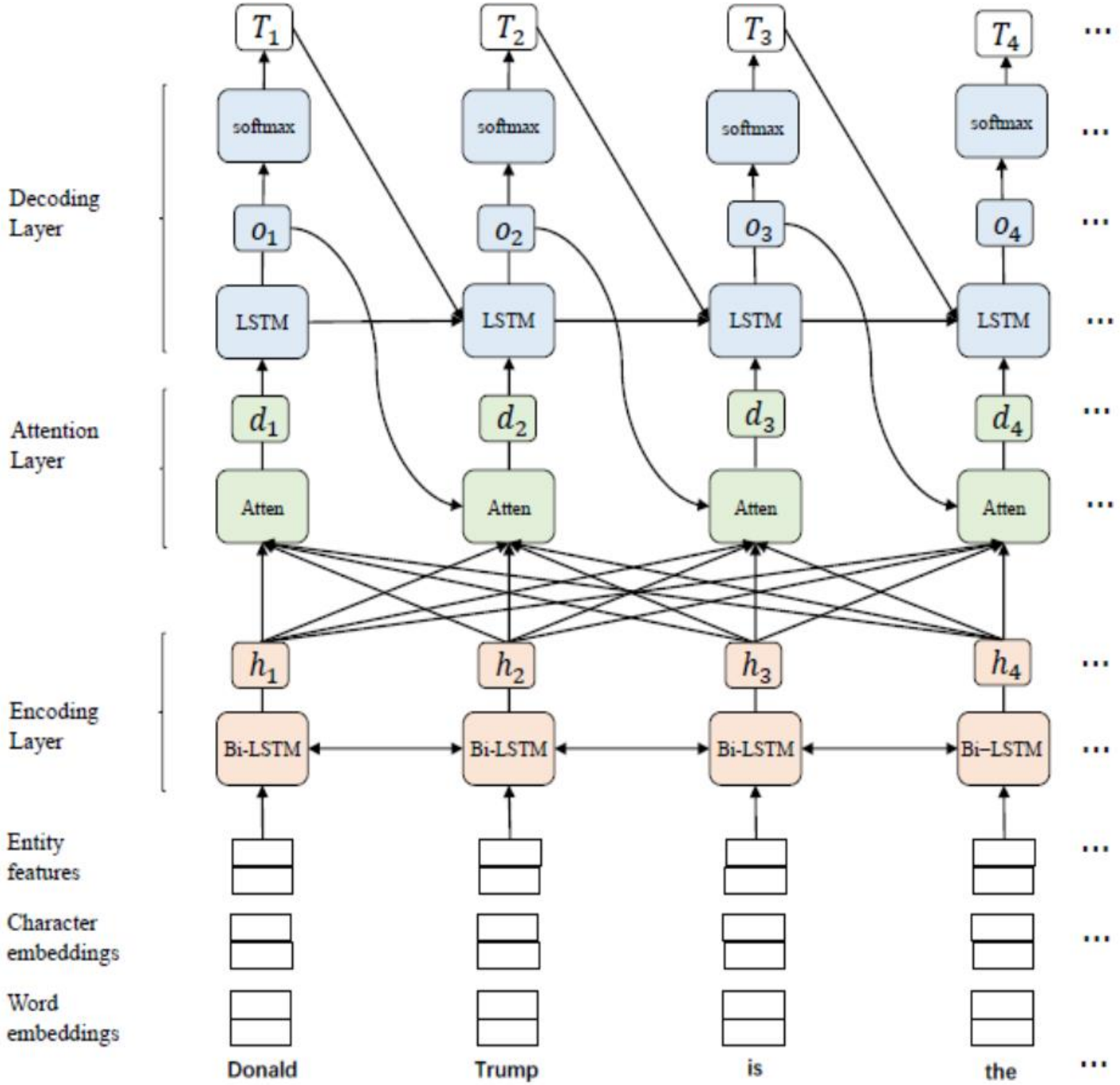
PR：President-Of

2：Second entity

OUTLINE

- Introduction
- Method
- Experiment
- Conclusion

FRAMEWORK


Decoding Layer
Attention Layer
Encoding Layer
Entity features
Character embeddings
Word embeddings
softmax
LSTM
Atten
Bi-LSTM
T_1 T_2 T_3 T_4
o_1 o_2 o_3 o_4
d_1 d_2 d_3 d_4
h_1 h_2 h_3 h_4
Donald
Trump
is
the

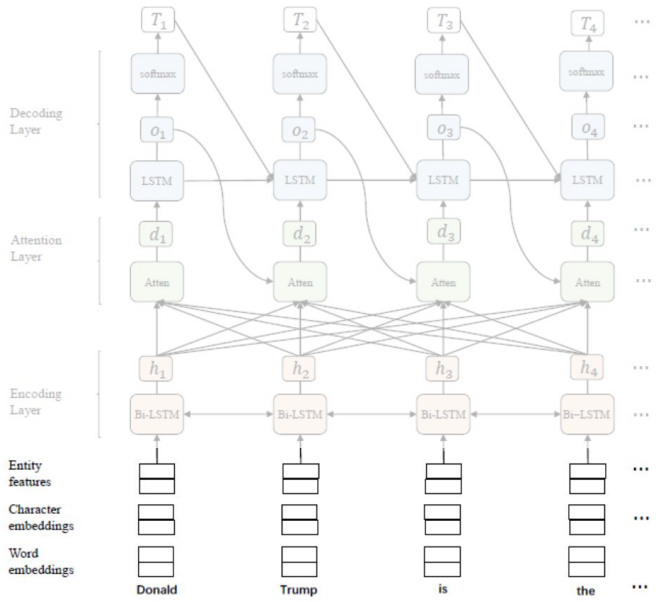
METHOD Features(Word embedding & character embedding)

➢Word embedding :

- Pre-train the embedding for words.

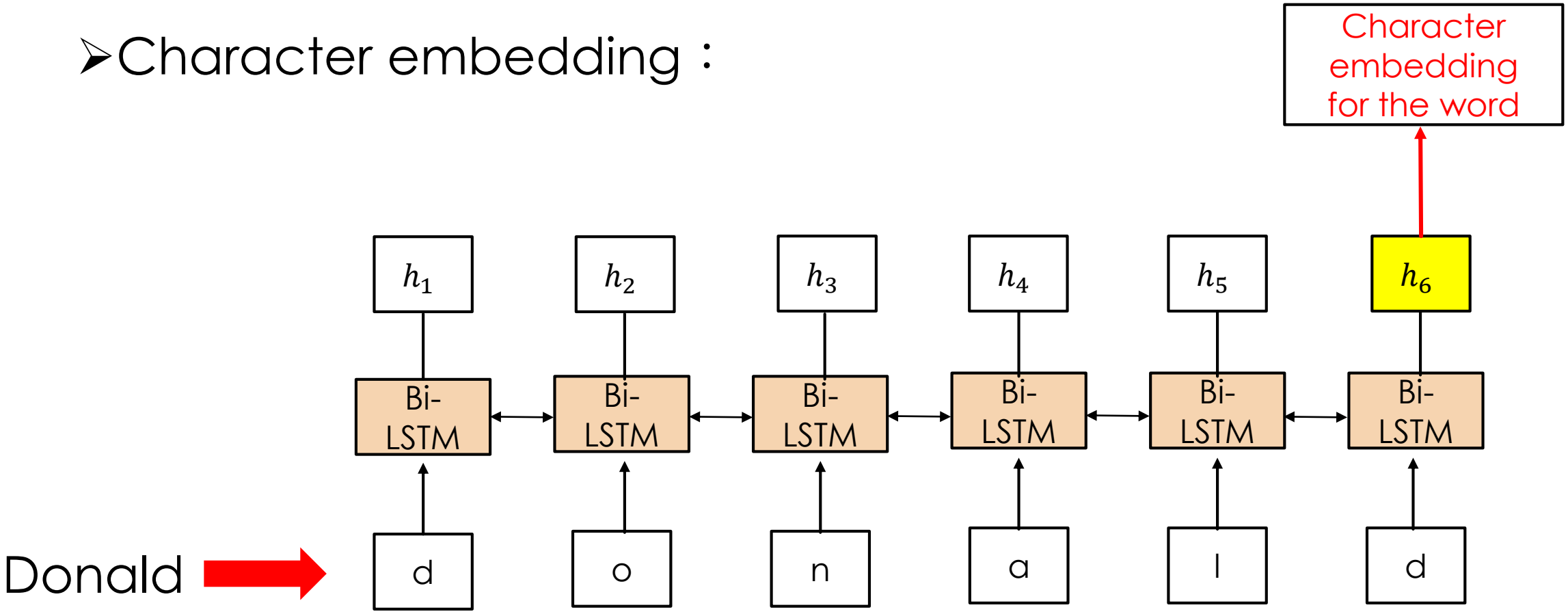
➢Character embedding :

- Each word is broken up into individual characters.
- Each characters are mapping to their embedding $(c_1, c_2, \ldots, c_L)$.
- Adopt Bi-LSTM to generates the character embedding for the word.



METHOD Features(Word embedding & character embedding)

➢Character embedding :

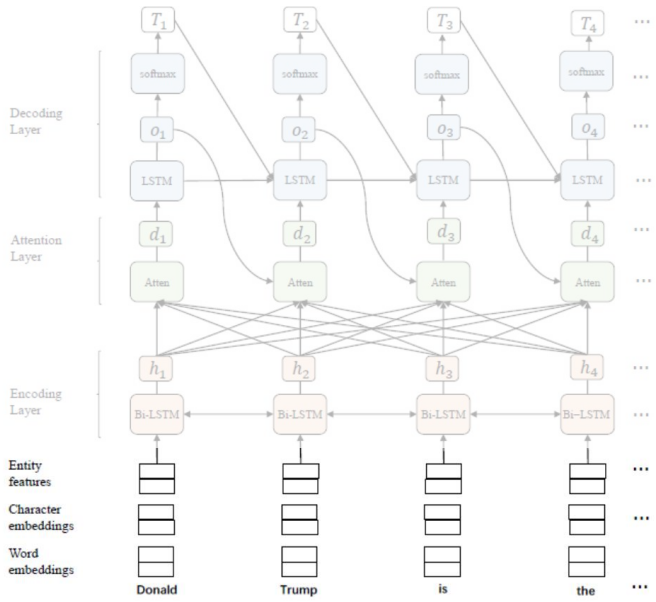


METHOD Features(Implicit features)

➢Implicit entity feature：

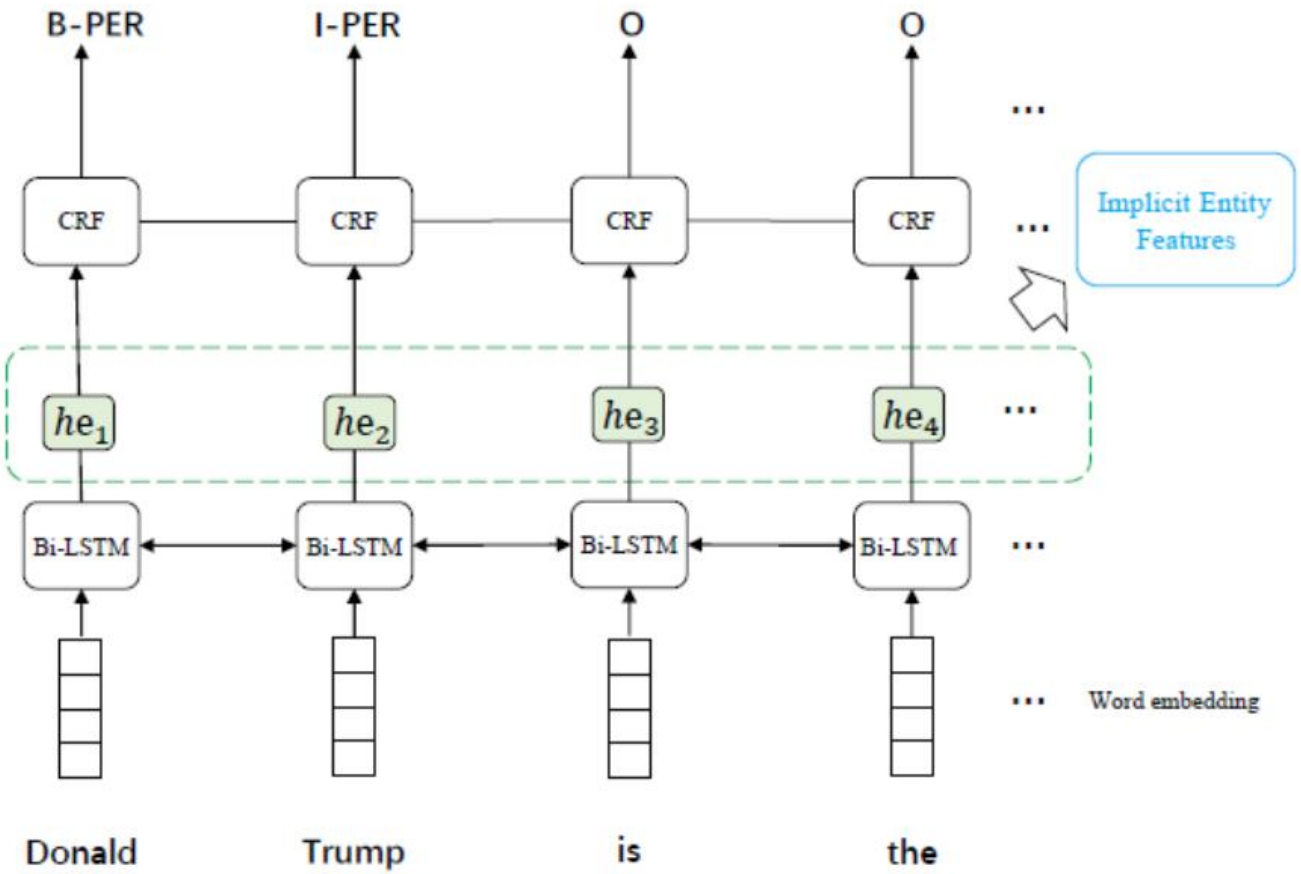
- Pre-train an model on an existing NER dataset.

- Feed the input sentence(Danald Trump is the……) into this model.
- The hidden vectors are entity features.



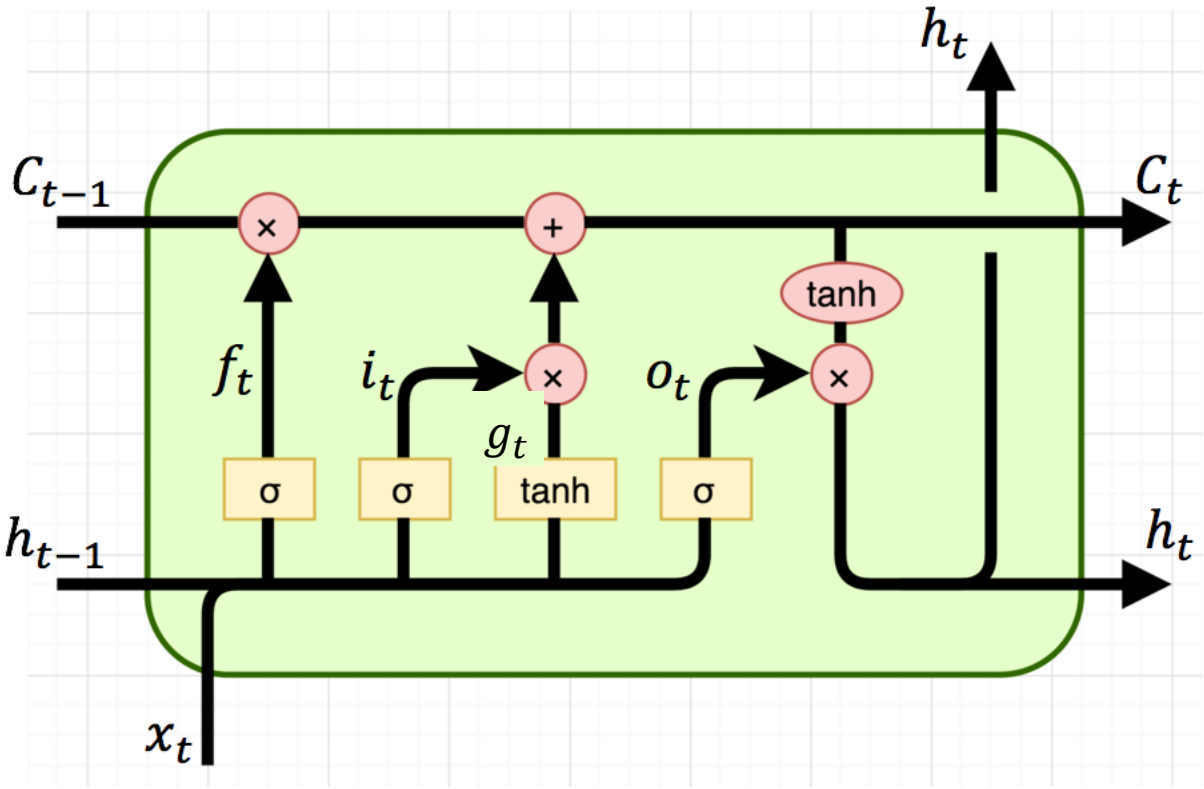
METHOD Features(Implicit features)

➢Implicit entity feature：



METHOD Encoder layer

➢LSTM：



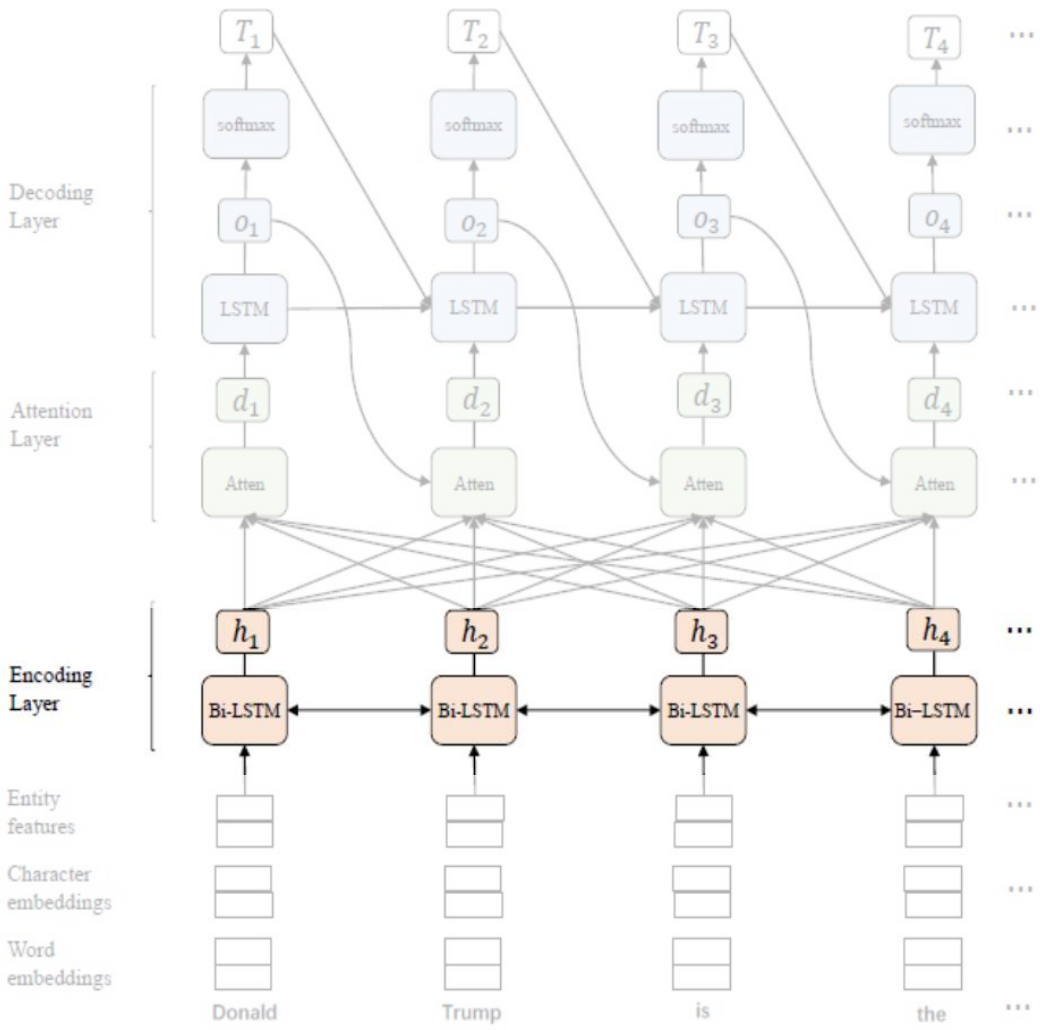
$$i_t = \sigma(W_{xi}x_t + b_{xi} + W_{hi}h_{t-1} + b_{hi})$$
$$f_t = \sigma(W_{xf}x_t + b_{xf} + W_{hf}h_{t-1} + b_{hf})$$
$$g_t = tanh(W_{xg}x_t + b_{xg} + W_{hg}h_{t-1} + b_{hg})$$
$$o_t = \sigma(W_{xo}x_t + b_{xo} + W_{ho}h_{t-1} + b_{ho})$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$
$$h_t = o_t \odot tanh(c_t)$$

https://www.itread01.com/content/1545027542.html

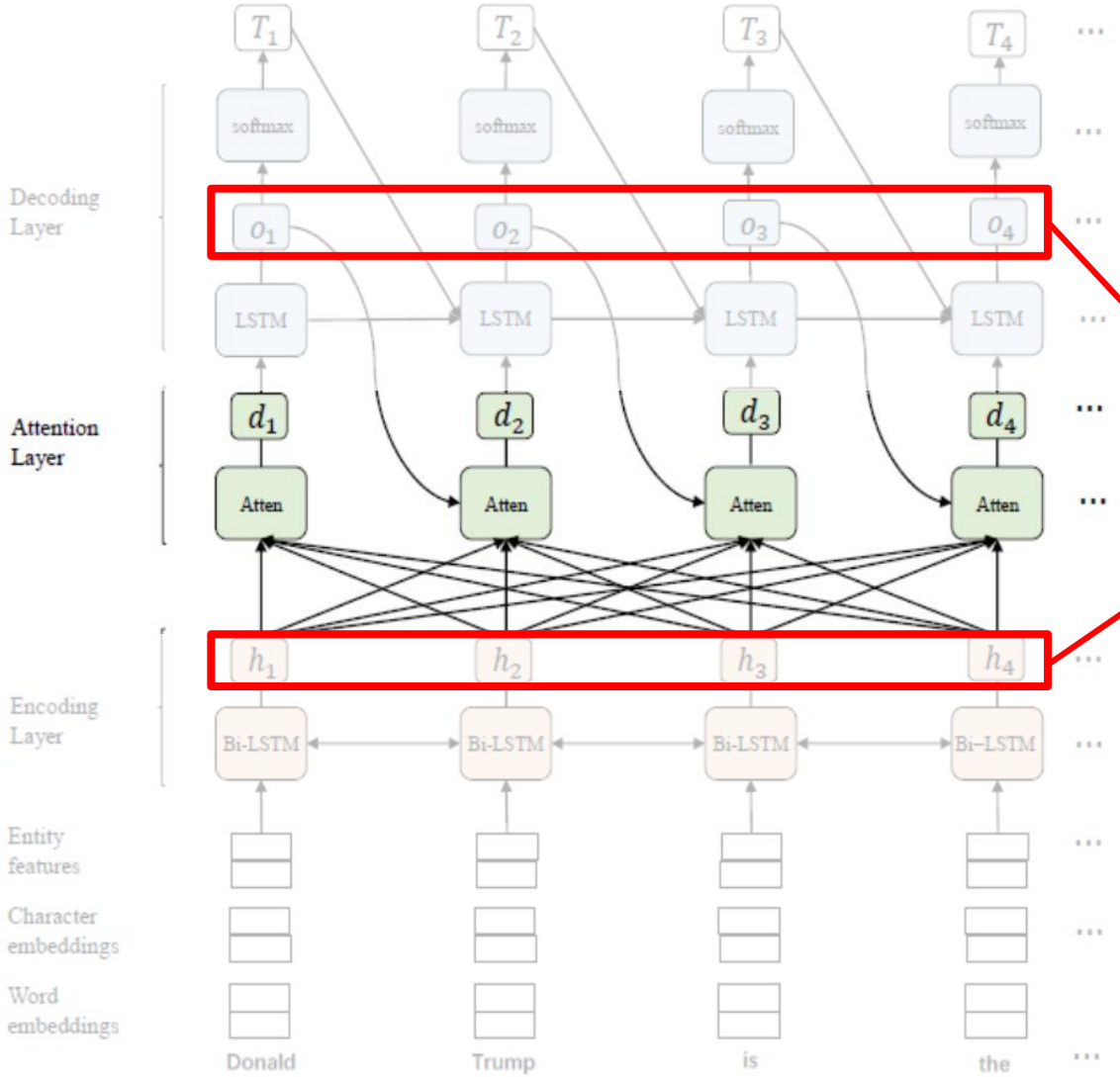
METHOD Encoding layer

➢Encoding layer：

- This layer receives three vectors(concate) as input.
- Use Bi-LSTM to computes the t step hidden state h_t.



METHOD Attention layer



The input of attention layer

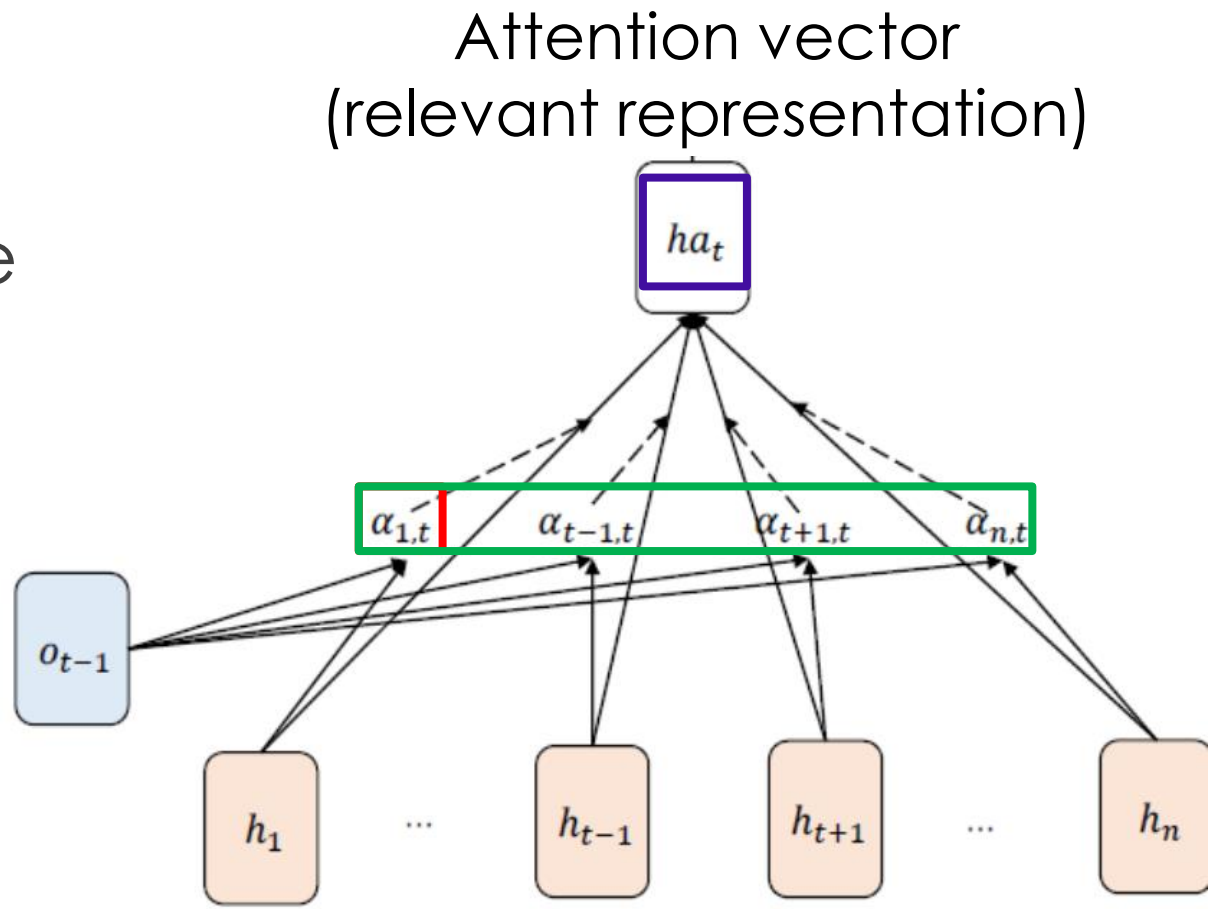
METHOD Attention layer

➢ **Tag aware attention :**

- Allow the model to select the relevant parts of the sentence for the prediction of the tag.
- $e_{j,t} = h_j \cdot o_{t-1}$

$$\alpha_{j,t} = \frac{exp(e_{j,t})}{\sum_{k=1}^{t-1} exp(e_{k,t}) + \sum_{k=t+1}^{n} exp(e_{k,t})}$$

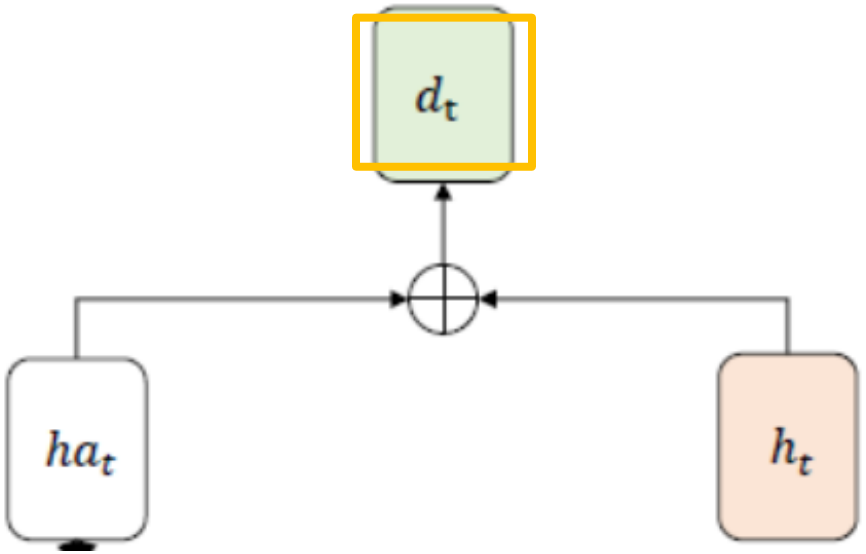
$$ha_t = \sum_{j=1}^{t-1} \alpha_{j,t} h_j + \sum_{j=t+1}^{n} \alpha_{j,t} h_j$$



METHOD Attention layer

➢ **Fusion gate：**

- When predicting the tag of a word, the gate allows to trade off the information used from ha_t and h_t.



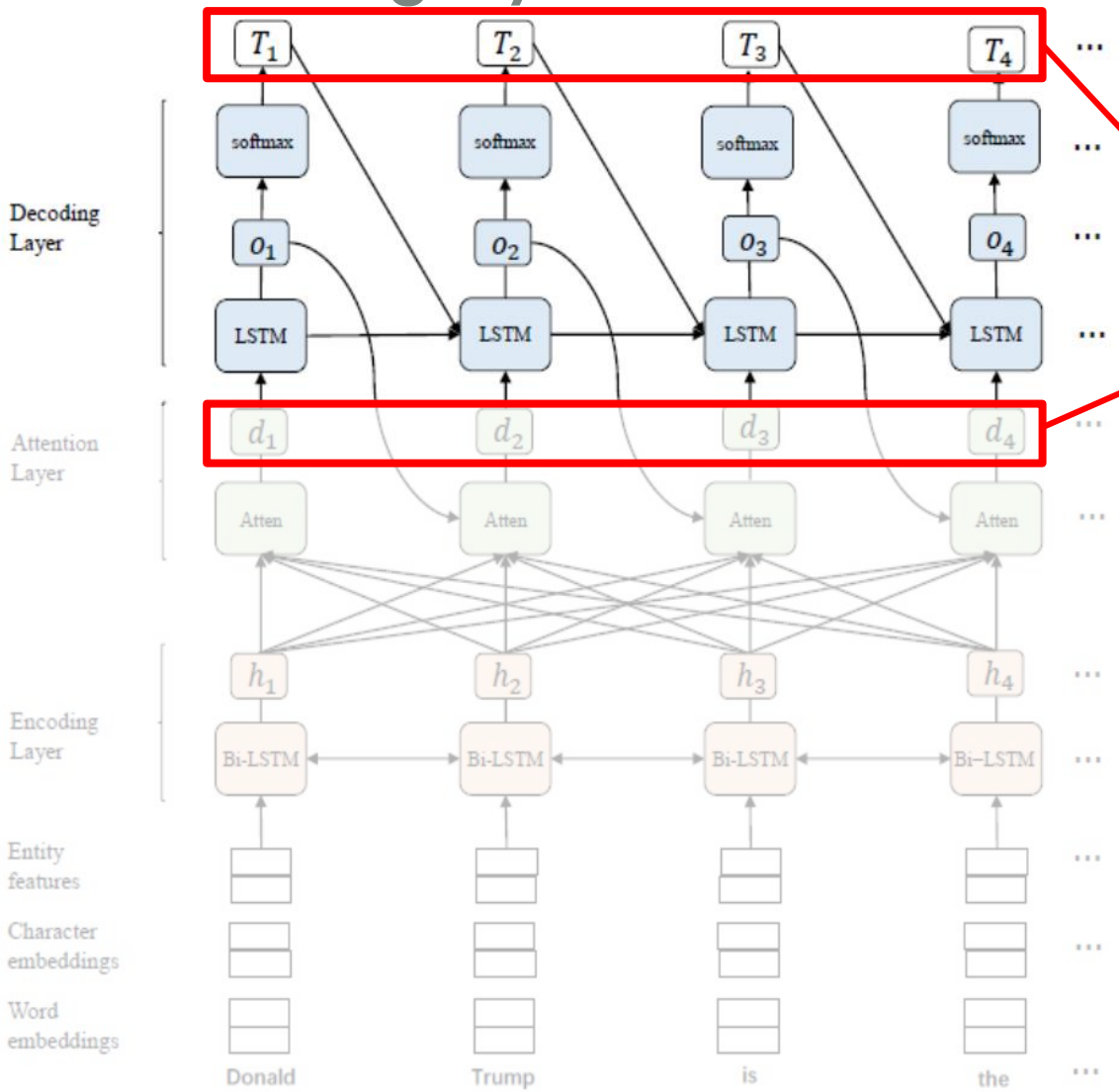
Weight

- $\boxed{g_t} = \sigma(W_g^{(3)} tanh(W_g^{(1)} ha_t + W_g^{(2)} h_t))$

$\boxed{d_t} = g_t ha_t + (1 - g_t) h_t$

Output of attention layer

METHOD Decoding layer



The input of decoding layer

METHOD Decoding layer

➢Decoding layer：

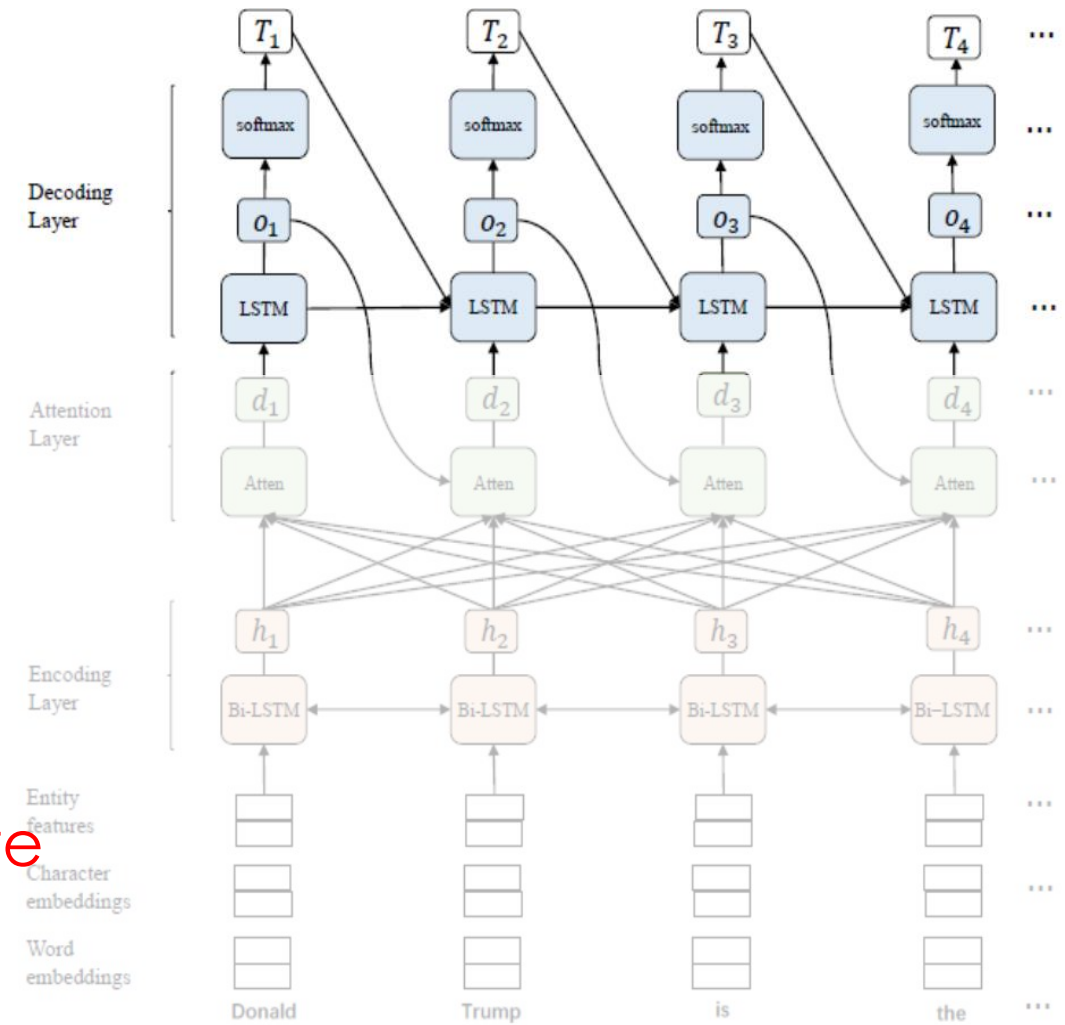
- Adopt LSTM to generate vectors representing the output states.

$$\overrightarrow{o_t} = LSTM_{decoder}(d_t + W_T T_{t-1}, \overrightarrow{o_{t-1}})$$

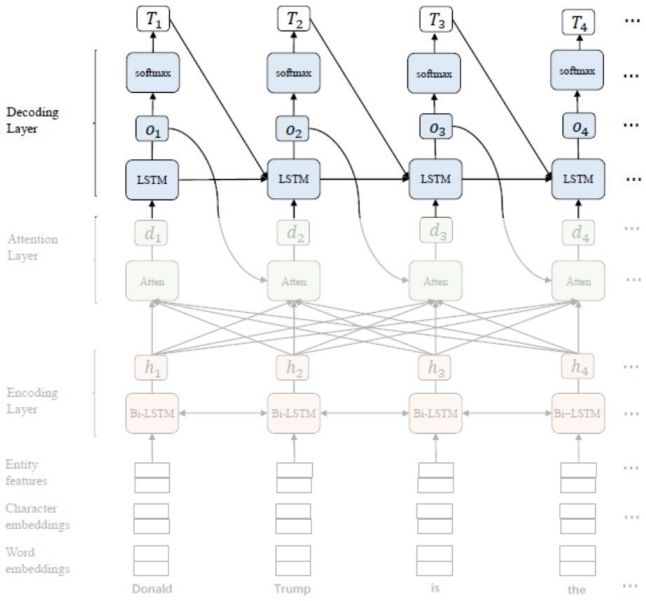
t-th output of Attention layer

t-1-th tag embedding

t-1-th hidden state of LSTM



METHOD Decoding layer



➢Decoding layer(continue) :

- Adopt a softmax classifier to compute entity tag probabilities.

$$s_t = W_s o_t + b_s$$

$$p(tag_t | S, tag_{t-1}) = \frac{exp\ (s_t)_{tag_t}}{\sum_{tag'_t \in TAG} exp\ (s_t)_{tag'_t}}$$

- Objective function :

$$I(O) = \begin{cases} 1, & if\ tag = 'O' \\ 0, & if\ tag \neq 'O' \end{cases}$$

$$L = max \sum_{m=1}^{|\mathbb{D}|} \sum_{t=1}^{L_m} (\boxed{log(p(tag_t^m | S_m, tag_{t-1}^m))} \cdot \boxed{I(O)} + \alpha \cdot log(p(tag_t^m | S_m, tag_{t-1}^m)) \cdot (1 - I(O)))$$

OUTLINE

- Introduction
- Method
- Experiment
- Conclusion

EXPERIMENT

Dataset

- **NYT :**
  - 353000 triplets in the training data and 3880 triplets in the testing data.
  - Relation type in the dataset is 24.

EXPERIMENT

- Dimension sizes：

Model	Network Structure	Size
Our Model	Word Embedding	300
	Char Embedding	20
	$LSTM_{char}$	25
	Entity Features	100
	$LSTM_{encoder}$	300
	$LSTM_{decoder}$	600
	Tag Embedding	50
Entity Features Generation Model	Word Embedding	300
	$LSTM_{entity}$	50

EXPERIMENT

- *A triplet is regarded as correct when all correct.*

➢ Comparison with baselines：

	Method	Prec.	Rec.	F1
Pipelined	FCM	0.553	0.154	0.240
	DS+logistic	0.258	0.393	0.311
	LINE	0.335	0.329	0.332
Joint	MultiR	0.338	0.327	0.333
	DS-Joint	0.574	0.256	0.354
	CoType	0.423	0.511	0.463
End-to-end	LSTM-LSTM-Bias	0.615	0.414	0.495
	Transition-Based	0.643	0.421	0.509
	Our model	0.640	0.464	**0.538**

EXPERIMENT

- *A triplet is regarded as correct when all correct.*

➢ Ablation results：

Method	Prec.	Rec.	F1
Our model	0.640	0.464	**0.538**
-Implicit Entity features	0.604	0.456	0.521
-Attention Layer	0.601	0.450	0.515
-Fusion Gate+Concat	0.599	0.469	0.526

EXPERIMENT

- Ablation results on triplet elements：

Elements	Method	Prec.	Rec.	F1
(E1, R)	Our Model	0.684	0.515	**0.588**
	-Implicit Entity Features	0.670	0.509	0.579
	-Attention Layer	0.669	0.501	0.573
(E2, R)	Our Model	0.663	0.499	**0.569**
	-Implicit Entity Features	0.642	0.488	0.555
	-Attention Layer	0.633	0.475	0.543
(E1, E2)	Our Model	0.667	0.501	**0.572**
	-Implicit Entity Features	0.646	0.490	0.558
	-Attention Layer	0.658	0.493	0.564

EXPERIMENT

- Visualization of attention weights：

Chad Hurley , the co-founder and chief executive of YouTube , said the company was still working on its filtering technology

Word: Chad Tag: B-PC-1

Chad Hurley , the co-founder and chief executive of YouTube , said the company was still working on its filtering technology

Word: Hurley Tag: B-PC-2

Chad Hurley , the co-founder and chief executive of YouTube , said the company was still working on its filtering technology

Word: YouTube Tag: S-PC-2

OUTLINE

- Introduction
- Method
- Experiment
- Conclusion

CONCLUSION

- Propose an attention-based model enhanced with implicit entity features for the joint extraction of entities and relations.
- This model can take advantage of the entity features and dose not need to manually design them.
- Design a Tag-Aware attention mechanism which enables our model to select the informative words to the prediction.