每周汇报

蔡碧瑜

2022/04/07

论文列表

No.	title	Publication Source	Year
1	Deep Fusion LSTMs for Text Semantic Matching	ACL	2016
2	Bi-Directional Attention Flow for Machine Comprehension	ICLR	2017
3	**Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling**	**EMNLP**	2019

/01

Deep Fusion LSTMs for Text Semantic Matching (2016.ACL)

Deep Fusion LSTMs for Text Semantic Matching (2016.ACL)

$$X = x_1, x_2, \cdots, x_m$$

$$Y = y_1, y_2, \cdots, y_n$$

define $\Rightarrow$

A matching vector $\mathbf{h}_{i,j}$

$$\mathbf{h}_{i,j}(X,Y) = \mathbf{h}_{i,j}(X|Y) \oplus \mathbf{h}_{i,j}(Y|X)$$

$$(\mathbf{h}_{i,j}^{(x)}, \mathbf{c}_{i,j}^{(x)}) = \mathbf{LSTM}(\mathcal{H}_{i,j}, \mathbf{c}_{i-1,j}^{(x)}, \mathbf{x}_i)$$
$$(\mathbf{h}_{i,j}^{(y)}, \mathbf{c}_{i,j}^{(y)}) = \mathbf{LSTM}(\mathcal{H}_{i,j}, \mathbf{c}_{i,j-1}^{(y)}, \mathbf{x}_j)$$

Deep Fusion LSTMs for Text Semantic Matching (2016.ACL)

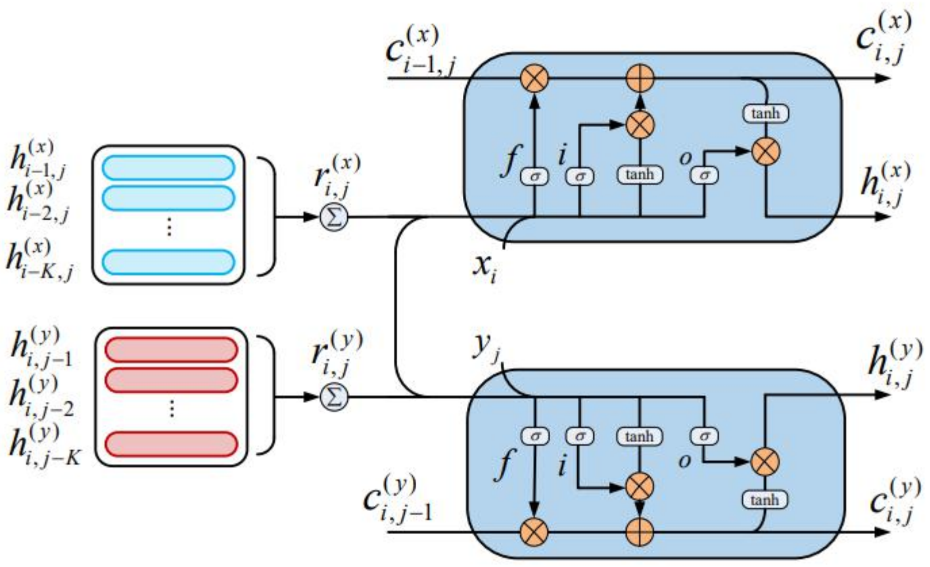


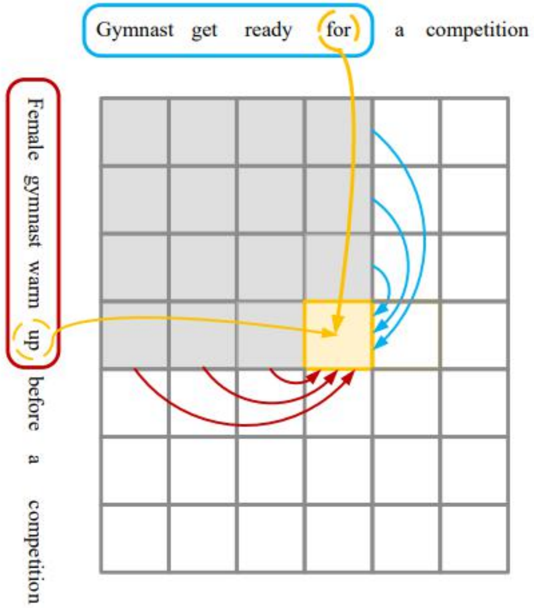
Figure 2: Illustration of DF-LSTMs unit.

$$(\mathbf{h}_{i,j}^{(x)}, \mathbf{c}_{i,j}^{(x)}) = \mathbf{LSTM}(\mathcal{H}_{i,j}, \mathbf{c}_{i-1,j}^{(x)}, \mathbf{x}_i)$$

$$(\mathbf{h}_{i,j}^{(y)}, \mathbf{c}_{i,j}^{(y)}) = \mathbf{LSTM}(\mathcal{H}_{i,j}, \mathbf{c}_{i,j-1}^{(y)}, \mathbf{x}_j)$$

$$\mathbf{M}_{i,j}^{(x)} = \{\mathbf{h}_{i-K,j}^{(x)}, \dots, \mathbf{h}_{i-1,j}^{(x)}\} \qquad \mathbf{r}_{i,j}^{(x)} = \mathbf{a}_{i,j}^{(x)} \mathbf{M}_{i,j}^{(x)}$$

$$\mathbf{M}_{i,j}^{(y)} = \{\mathbf{h}_{i,j-K}^{(y)}, \dots, \mathbf{h}_{i,j-1}^{(y)}\} \qquad \mathbf{r}_{i,j}^{(y)} = \mathbf{a}_{i,j}^{(y)} \mathbf{M}_{i,j}^{(y)}$$



$$\mathcal{H}_{i,j} = \mathbf{r}_{i,j}^{(x)} \oplus \mathbf{r}_{i,j}^{(y)}$$

Figure 3: Illustration of unfolded DF-LSTMs.

/02

Bi-Directional Attention Flow for Machine Comprehension (2017. ICLR)

Bi-Directional Attention Flow for Machine Comprehension (2017. ICLR)

$$\mathbf{S}_{tj} = \alpha(\mathbf{H}_{:t}, \mathbf{U}_{:j}) \in \mathbb{R}$$

$$\alpha(h, u) = w_{(S)}^{T}[h; u; h \odot u]$$

$$\mathbf{H} \in \mathbb{R}^{2d \times T}$$
$$\mathbf{U} \in \mathbb{R}^{2d \times J}$$

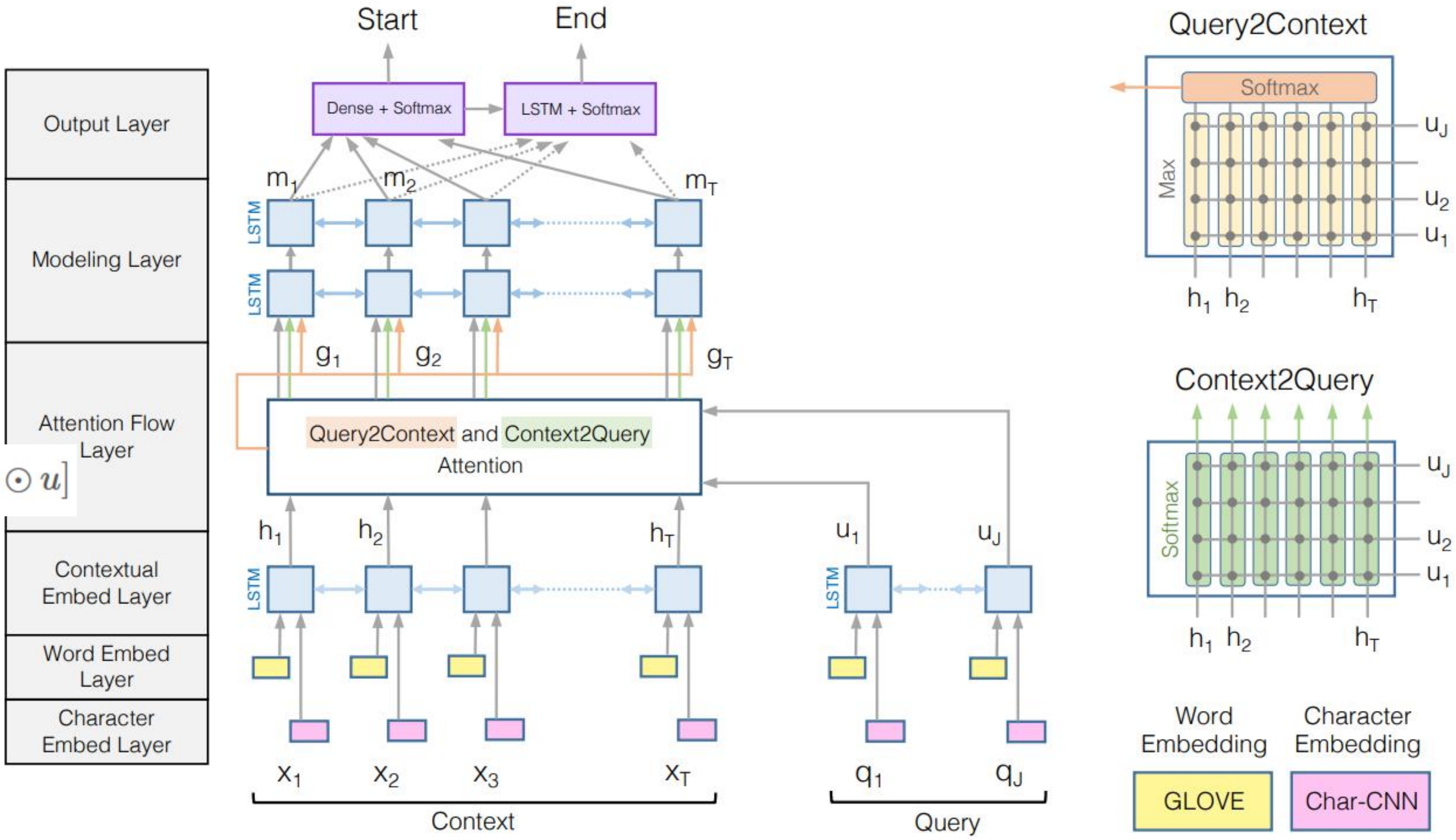


Figure 1: BiDirectional Attention Flow Model *(best viewed in color)*

Bi-Directional Attention Flow for Machine Comprehension (2017. ICLR)

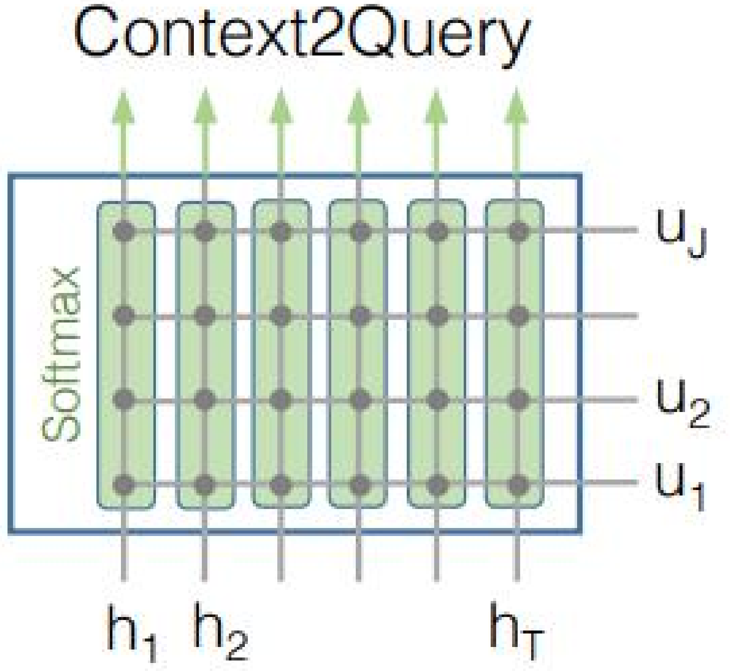
Context-to-query(C2Q)

- which query words are most relevant to each context word

$$\mathbf{S}_{tj} = \alpha(\mathbf{H}_{:t}, \mathbf{U}_{:j}) \in \mathbb{R}$$

$$\mathbf{a}_t \in \mathbb{R}^J \qquad \mathbf{a}_t = \mathrm{softmax}(\mathbf{S}_{t:})$$

$$\tilde{\mathbf{U}}_{:t} = \sum_j \mathbf{a}_{tj} \mathbf{U}_{:j}$$



Bi-Directional Attention Flow for Machine Comprehension (2017. ICLR)

Query-to-context(Q2C)

- which context words have the closest similarity to one of the query words and hence critical for answering the query

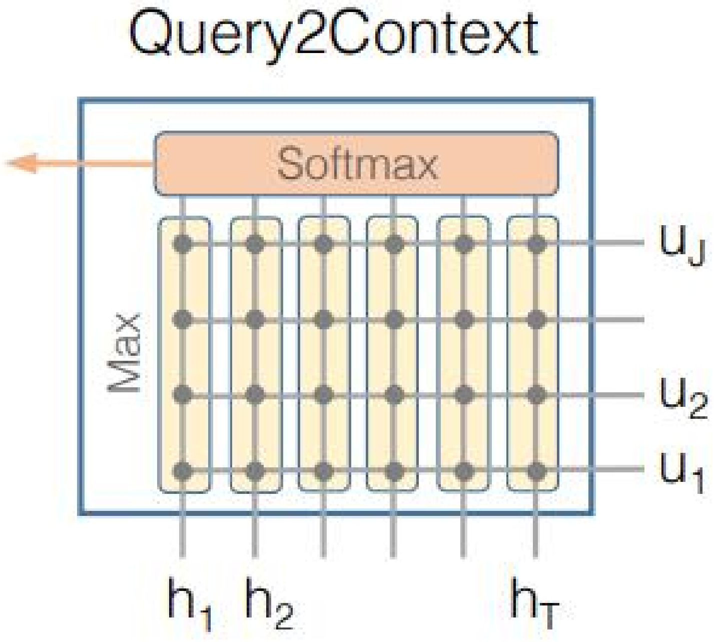
$$\mathbf{S}_{tj} = \alpha(\mathbf{H}_{:t}, \mathbf{U}_{:j}) \in \mathbb{R}$$

$$\mathbf{b} = \mathrm{softmax}(\max_{col}(\mathbf{S})) \in \mathbb{R}^T$$

$$\tilde{\mathbf{h}} = \sum_t \mathbf{b}_t \mathbf{H}_{:t} \in \mathbb{R}^{2d}$$

tile T times across the column

$$\tilde{\mathbf{H}} \in \mathbb{R}^{2d \times T}$$



Bi-Directional Attention Flow for Machine Comprehension (2017. ICLR)

$$\mathbf{G}_{:t} = \boldsymbol{\beta}(\mathbf{H}_{:t}, \tilde{\mathbf{U}}_{:t}, \tilde{\mathbf{H}}_{:t}) \in \mathbb{R}^{d_{\mathbf{G}}}$$

$$\beta(h, \tilde{u}, \tilde{h}) = [h; \tilde{u}; h \circ \tilde{u}; h \circ \tilde{h}] \in \mathbb{R}^{8d \times T}$$

Bi-Directional Attention Flow for Machine Comprehension (2017. ICLR)

$$\mathbf{p}^1 = \mathrm{softmax}(\mathbf{w}^\top_{(\mathbf{p}^1)}[\mathbf{G};\mathbf{M}])$$

$$\mathbf{p}^2 = \mathrm{softmax}(\mathbf{w}^\top_{(\mathbf{p}^2)}[\mathbf{G};\mathbf{M}^2])$$

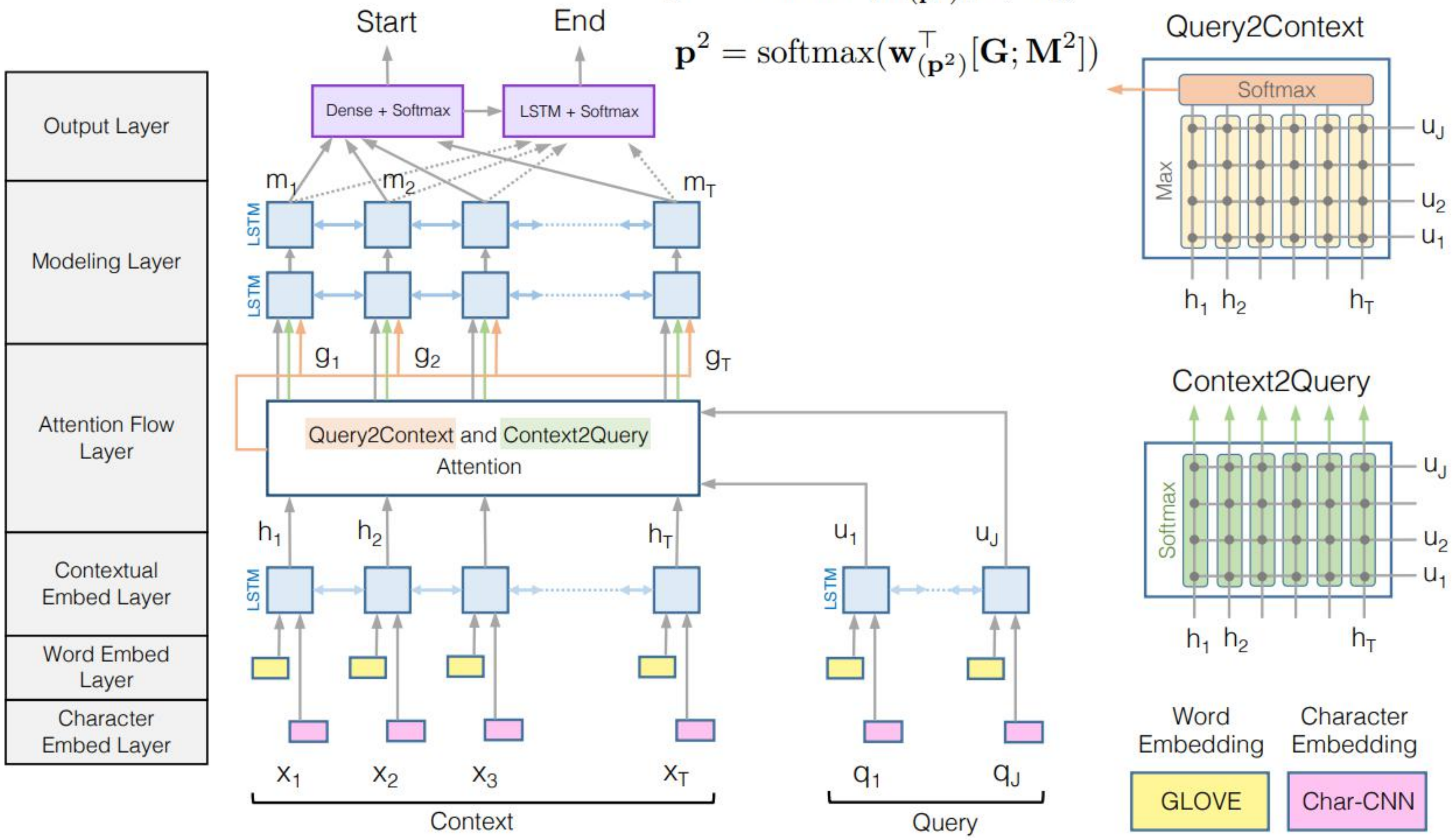


Figure 1: BiDirectional Attention Flow Model *(best viewed in color)*

/03

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)

content

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)

- Information retrieval $\Rightarrow$ • Relevance matching
- NLP problems $\Rightarrow$ • Semantic matching

$\Downarrow$

HCAN (Hybrid Co-Attention Network)

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)

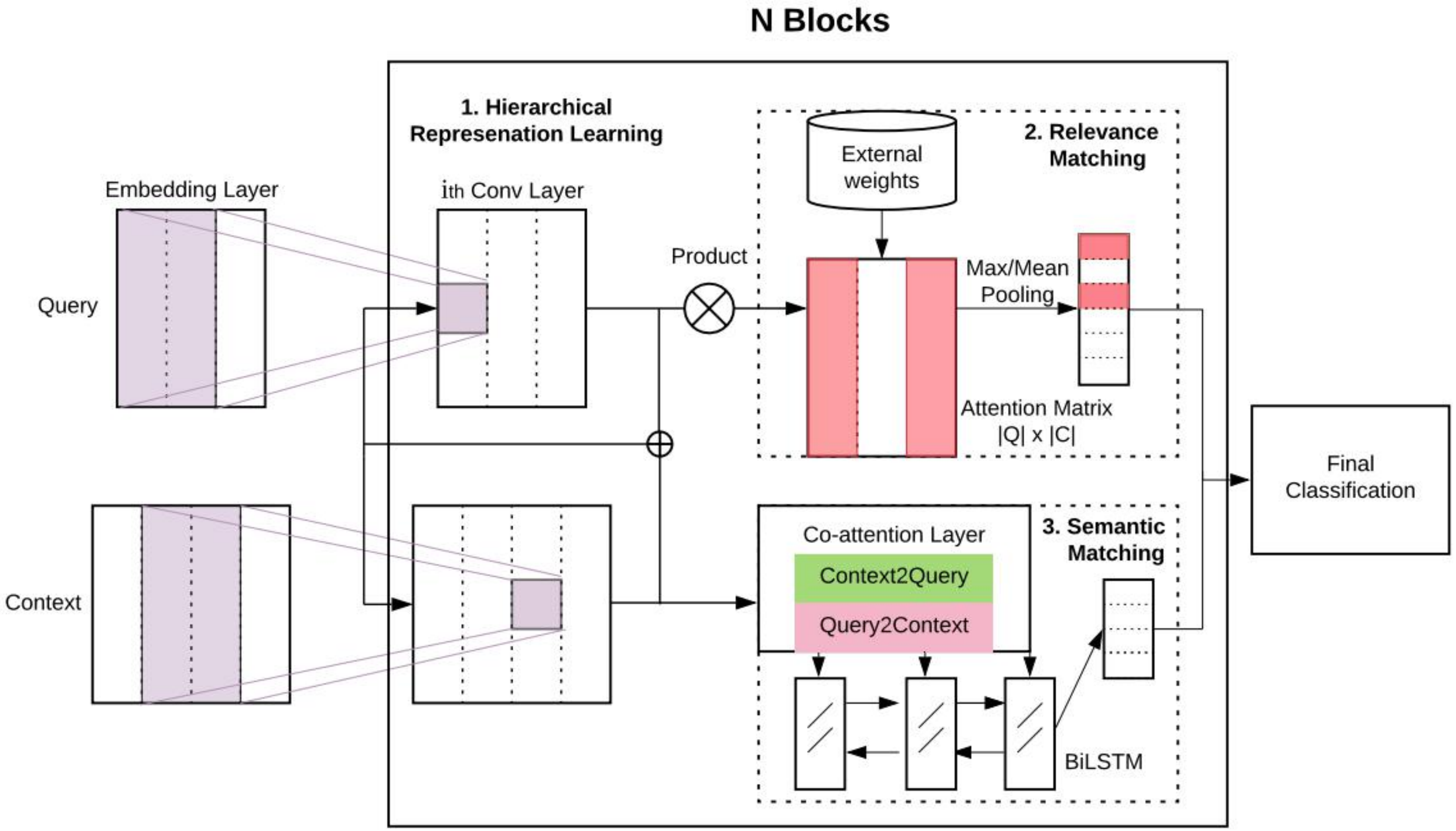
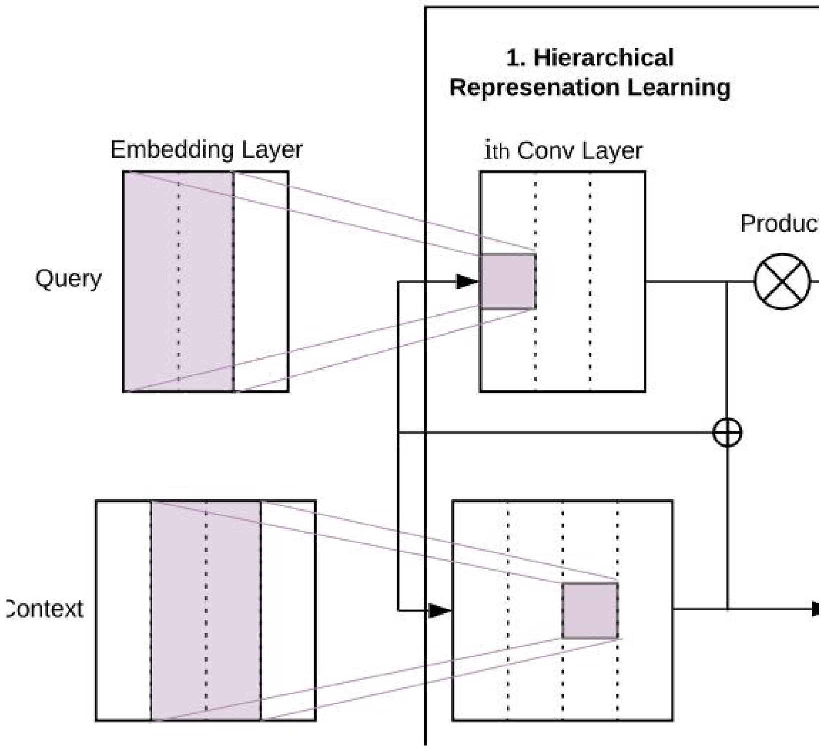


Figure 1: Overview of our Hybrid Co-Attention Network (HCAN). The model consists of three major components: (1) a hybrid encoder module that explores three types of encoders: *deep*, *wide*, and *contextual*; (2) a relevance matching module with external weights for learning soft term matching signals; (3) a semantic matching module with co-attention mechanisms for context-aware representation learning.

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)



$$\{w_1^q, w_2^q, ..., w_n^q\} \qquad \mathbf{Q} \in \mathbb{R}^{n \times L}$$

$$(q, c)$$

$$\{w_1^c, w_2^c, ..., w_m^c\} \qquad \mathbf{C} \in \mathbb{R}^{m \times L}$$

Deep Encoder

$$\mathbf{U}^h = \mathrm{CNN}^h(\mathbf{U}^{h-1}), h = 1, \ldots, N$$

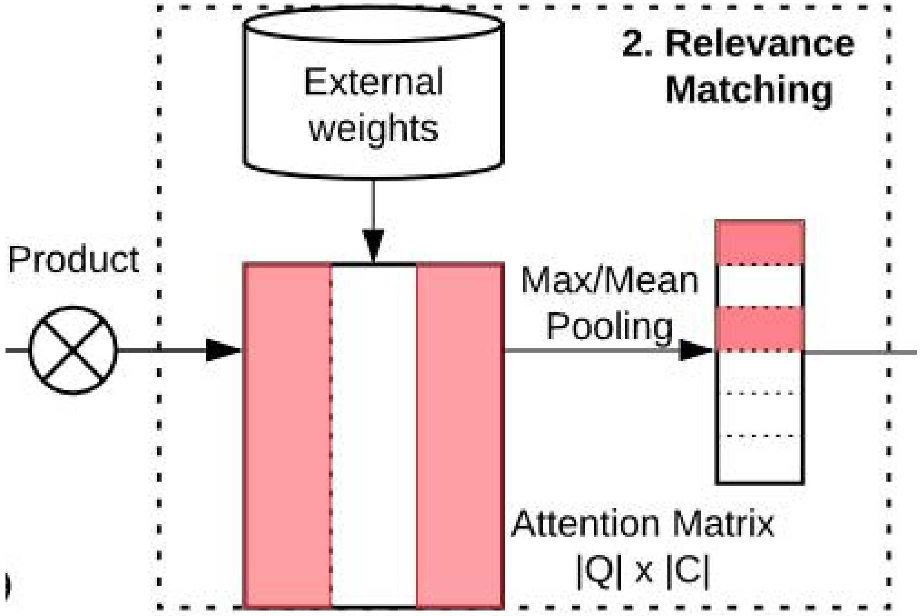
Wide Encoder

$$[k, k+1, ..., k+N-1]$$

Contextual Encoder

$$\mathbf{U}^h = \mathrm{BiLSTM}^h(\mathbf{U}^{h-1}), h = 1, \ldots, N$$

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)



$$Max(\mathbf{S}) = [\max(\tilde{\mathbf{S}}_{1,:}), ..., \max(\tilde{\mathbf{S}}_{n,:})],$$
$$Mean(\mathbf{S}) = [mean(\tilde{\mathbf{S}}_{1,:}), ..., mean(\tilde{\mathbf{S}}_{n,:})],$$
$$Max(\mathbf{S}), Mean(\mathbf{S}) \in \mathbb{R}^n$$

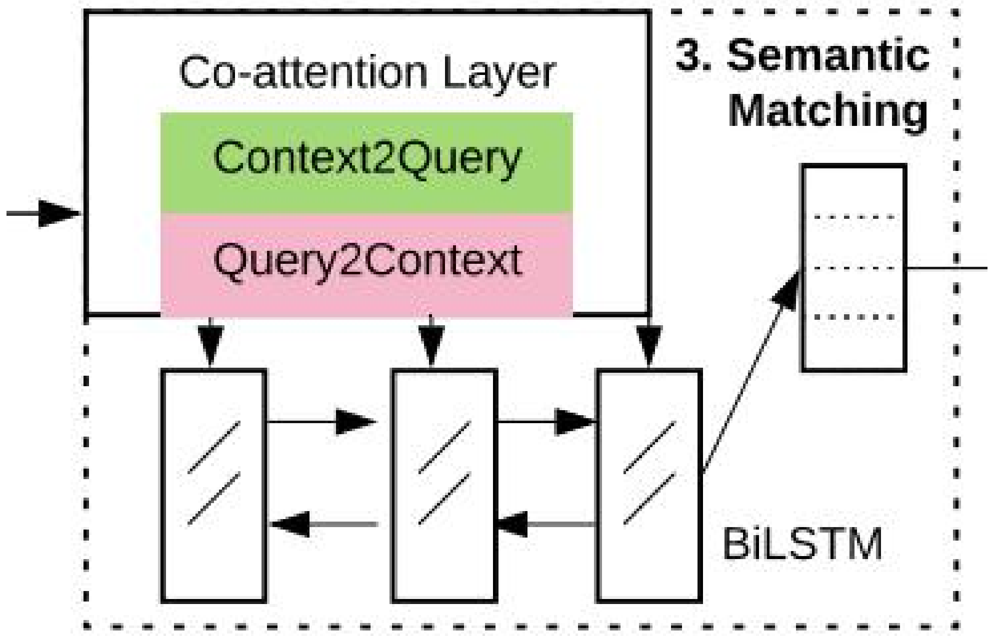
$$\mathbf{o}_{RM} = \{wgt(q) \odot Max(\mathbf{S}), wgt(q) \odot Mean(\mathbf{S})\}$$
$$\mathbf{O}_{RM} \in 2 \cdot \mathbb{R}^n,$$

$$\mathbf{S} = \mathbf{U}_q {\mathbf{U}_c}^T, \mathbf{S} \in \mathbb{R}^{n \times m}$$

$$\Big\Downarrow \text{ Softmax}$$

$$\tilde{\mathbf{S}}$$

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)

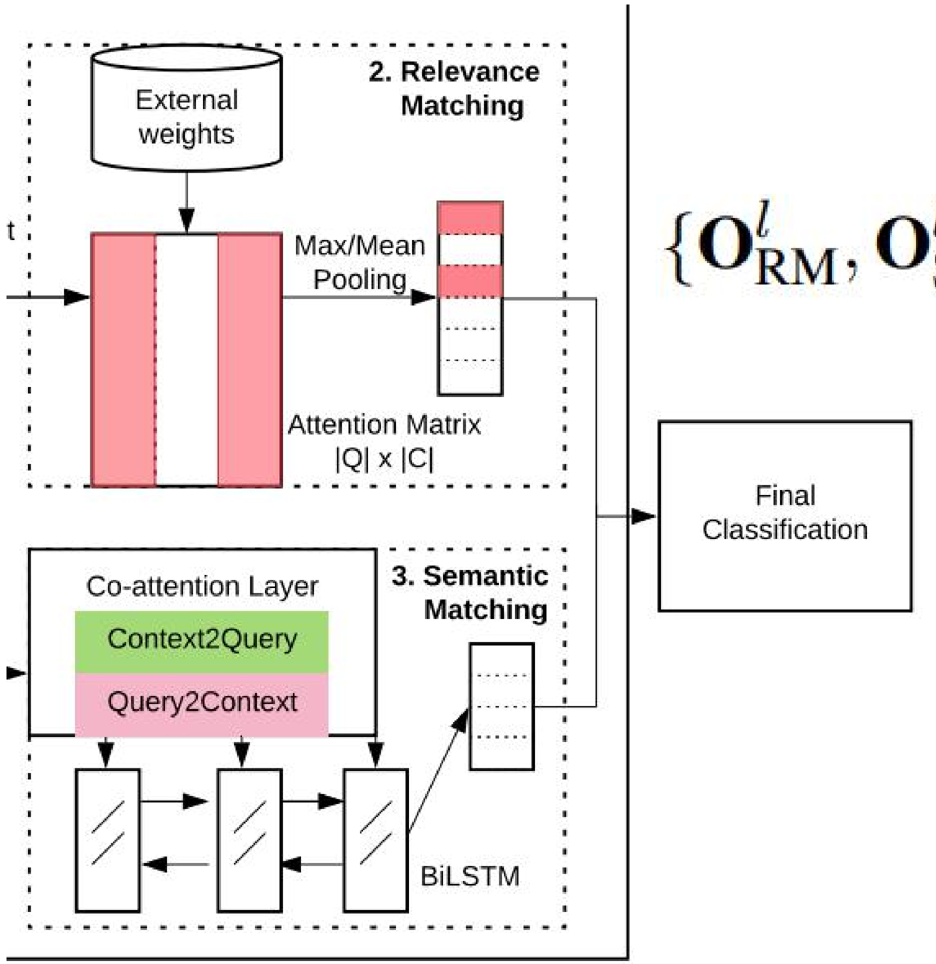


$$\mathbf{A} = REP(\mathbf{U}_q\mathbf{W}_q) + REP(\mathbf{U}_c\mathbf{W}_c) + \mathbf{U}_q\mathbf{W}_b\mathbf{U}_c^T$$
$$\mathbf{A} = softmax_{\text{col}}(\mathbf{A})$$
$$\mathbf{A} \in \mathbb{R}^{n\times m}$$

$$\tilde{\mathbf{U}}_q = \mathbf{A}^T\mathbf{U}_q$$
$$\tilde{\mathbf{U}}_c = REP(\max_{\text{col}}(\mathbf{A})\mathbf{U}_c)$$
$$\tilde{\mathbf{U}}_q \in \mathbb{R}^{m\times F}, \tilde{\mathbf{U}}_c \in \mathbb{R}^{m\times F}$$

$$\mathbf{H} = [\mathbf{U}_c; \tilde{\mathbf{U}}_q; \mathbf{U}_c \otimes \tilde{\mathbf{U}}_q; \tilde{\mathbf{U}}_c \otimes \tilde{\mathbf{U}}_q]$$
$$\mathbf{O}_{\text{SM}} = \text{BiLSTM}(\mathbf{H})$$
$$\mathbf{H} \in \mathbb{R}^{m\times 4F}, \mathbf{O}_{\text{SM}} \in \mathbb{R}^d$$

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)



$$\{\mathbf{O}^l_{\mathrm{RM}}, \mathbf{O}^l_{\mathrm{SM}}\}$$

$$\begin{aligned} o =& \mathit{softmax}(\mathrm{MLP}(\{\mathbf{O}^l_{\mathrm{RM}}, \mathbf{O}^l_{\mathrm{SM}}\})), \\ l =& 1, 2, ..., N \quad \text{and} \quad o \in \mathbb{R}^{\|class\|} \end{aligned}$$

$$L = -\sum_{(o_i, y_i)} \log o_i[y_i],$$

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)

- Answer Selection → TrecQA
- Paraphrase Identification → TwitterURL
- Semantic Textual Similarity → Quora
- Tweet Search → TREC Microblog

Model	TrecQA		TwitterURL	Quora
	MAP	MRR	macro-F1	Acc
InferSent	0.521	0.559	0.797	0.866
DecAtt	0.660	0.712	0.785	0.845
ESIM_{seq}	0.771	0.795	**0.822**	0.850
ESIM_{tree}	0.698	0.734	-	0.755
$\text{ESIM}_{seq+tree}$	0.749	0.768	-	0.854
PWIM	0.739	0.795	0.809	0.834
State-of-the-Art Models				
Rao et al. (2016)	**0.780**	0.834	-	-
Gong et al. (2018)	-	-	-	**0.891**
BERT	0.838	0.887	0.852	0.892
Our Approach				
RM	0.756	0.812	0.790	0.842
SM	0.663	0.725	0.708	0.817
HCAN	0.774	**0.843**	0.817	0.853

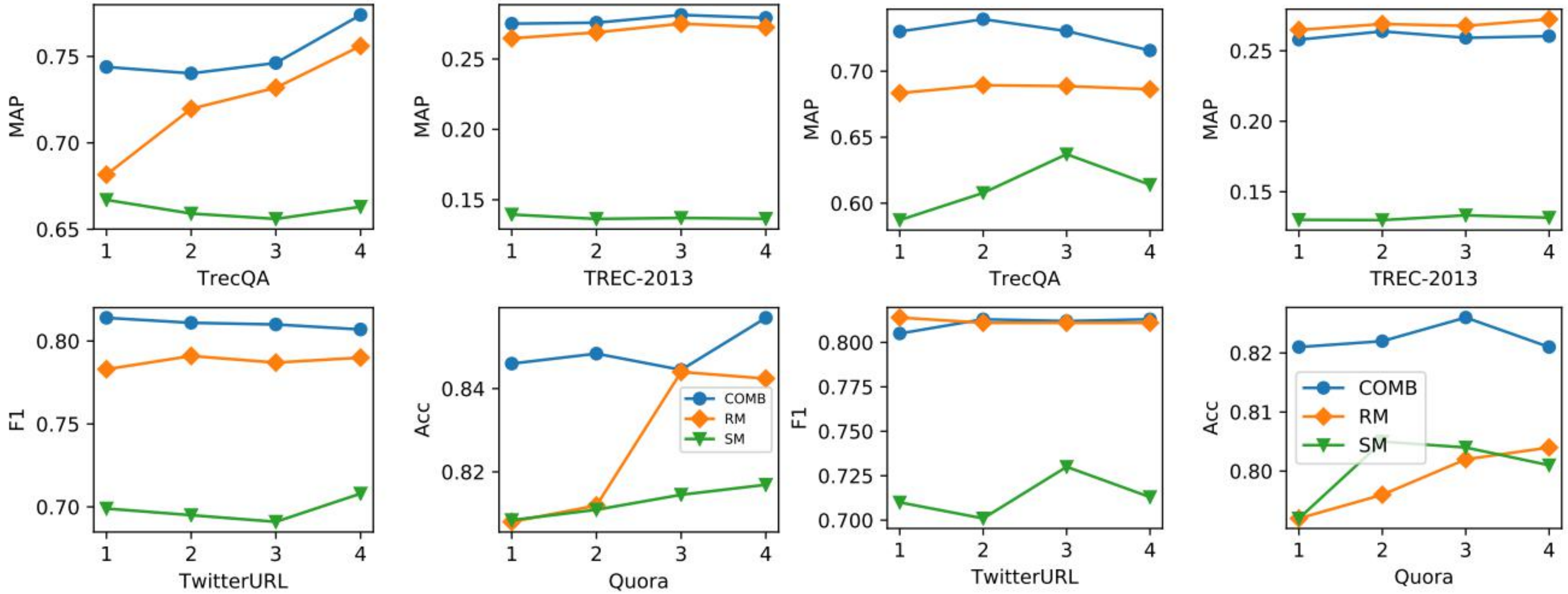
Model	TREC-2013		TREC-2014	
	MAP	P@30	MAP	P@30
QL	0.2532	0.4450	0.3924	0.6182
RM3	0.2766	0.4733	**0.4480**	0.6339
L2R	0.2477	0.4617	0.3943	0.6200
Neural Baselines				
DUET	0.1380	0.2528	0.2680	0.4091
DRMM	0.2102	0.4061	0.3440	0.5424
K-NRM	0.1750	0.3178	0.3472	0.5388
PACRR	0.2627	0.4872	0.3667	0.5642
BERT	0.3357	0.5656	0.5176	0.7006
Our Approach				
RM	0.2818	0.5222	0.4304	0.6297
SM	0.1365	0.2411	0.2414	0.3279
HCAN	**0.2920**	**0.5328**	0.4365	**0.6485**

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)

Encoder	Model	TrecQA		TwitURL	Quora	TREC-2013		TREC-2014	
		MAP	MRR	macro-F1	Acc	MAP	P@30	MAP	P@30
Deep	RM	0.756	0.812	0.790	0.842	0.282	0.522	0.430	0.630
	SM	0.663	0.725	0.708	0.817	0.137	0.241	0.241	0.328
	HCAN	**0.774**	0.843	**0.817**	**0.853**	**0.292**	**0.533**	**0.437**	**0.649**
Wide	RM	0.758	0.806	0.790	0.830	0.278	0.510	0.421	0.617
	SM	0.673	0.727	0.719	0.811	0.138	0.247	0.247	0.336
	HCAN	0.770	**0.847**	0.795	0.843	0.285	0.524	0.435	0.642
Contextual	RM	0.690	0.736	0.811	0.804	0.272	0.503	0.417	0.613
	SM	0.668	0.735	0.730	0.805	0.133	0.256	0.242	0.324
	HCAN	0.739	0.790	0.815	0.826	0.285	0.524	0.434	0.635

Table 4: Evaluation of different encoders in Sec. 2.1 (best numbers on each dataset are bolded).

Bridging the Gap Between Relevance Matching and Semantic Matching for Short Text Similarity Modeling (EMNLP.2019)



(a) Deep encoder

(b) Contextual encoder

Figure 2: Model effectiveness with different numbers of encoder layers.

Thanks!