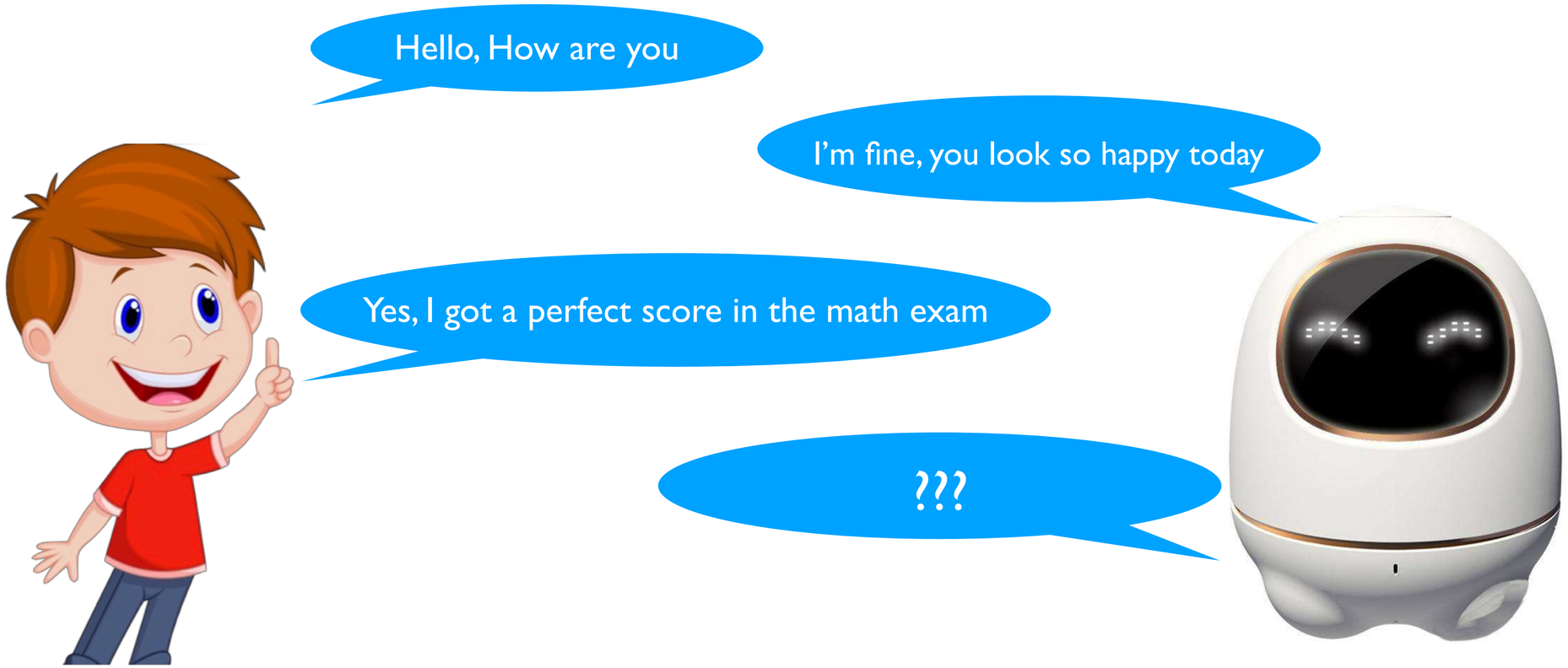
TripleNet: Triple Attention Network for Multi-Turn Response Selection in Retrieval-based Chatbots

Wentao Ma Yiming Cui Nan Shao Su He Wei-nan Zhang Ting Liu Shijin Wang Guoping Hu

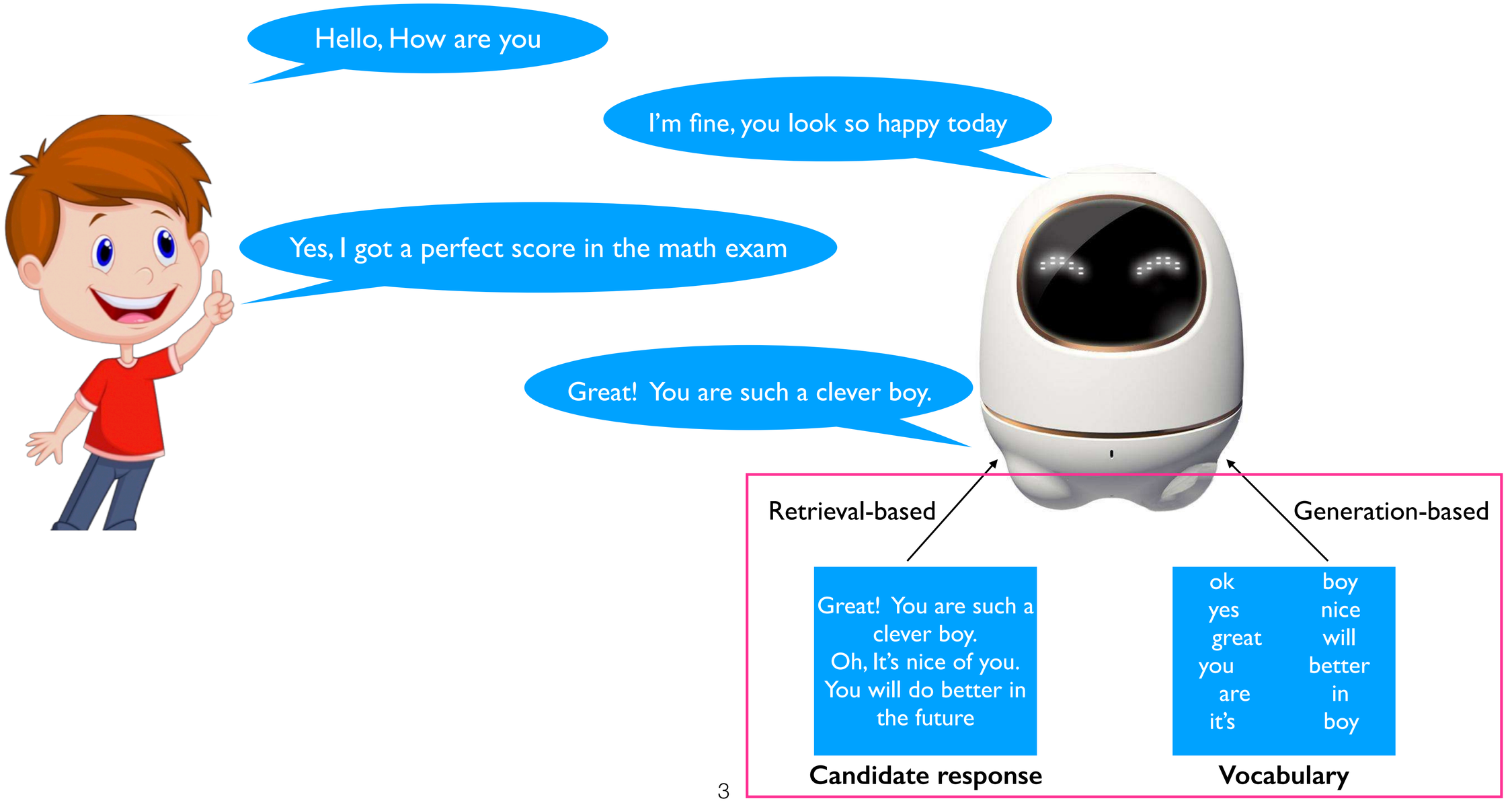
CoNLL 2019

State Key Laboratory of Cognitive Intelligence, iFLYTEK Research,
Research Center for Social Computing and Information Retrieval (SCIR),
iFLYTEK AI Research (Hebei)

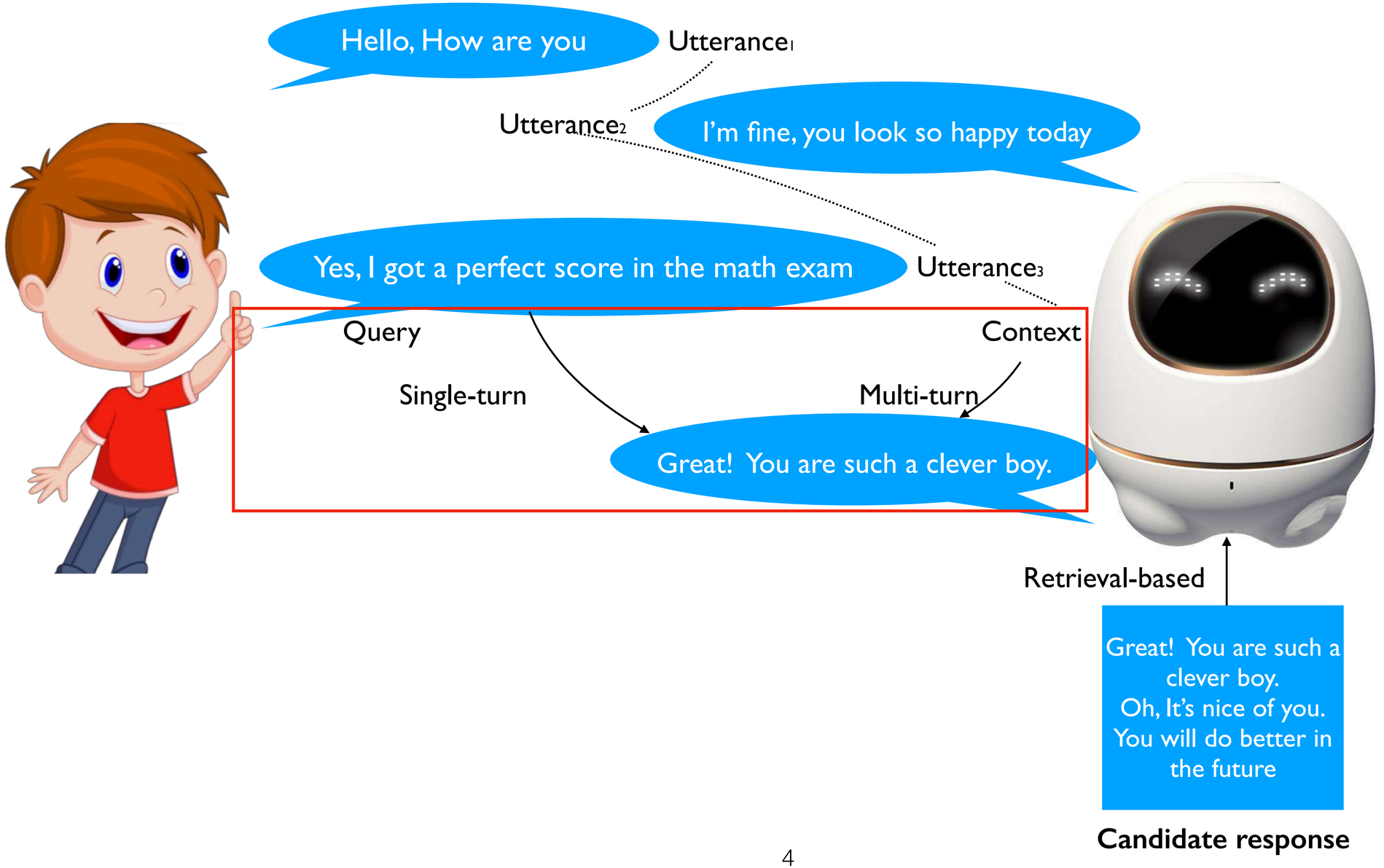
Background


Hello, How are you
I'm fine, you look so happy today
Yes, I got a perfect score in the math exam
???

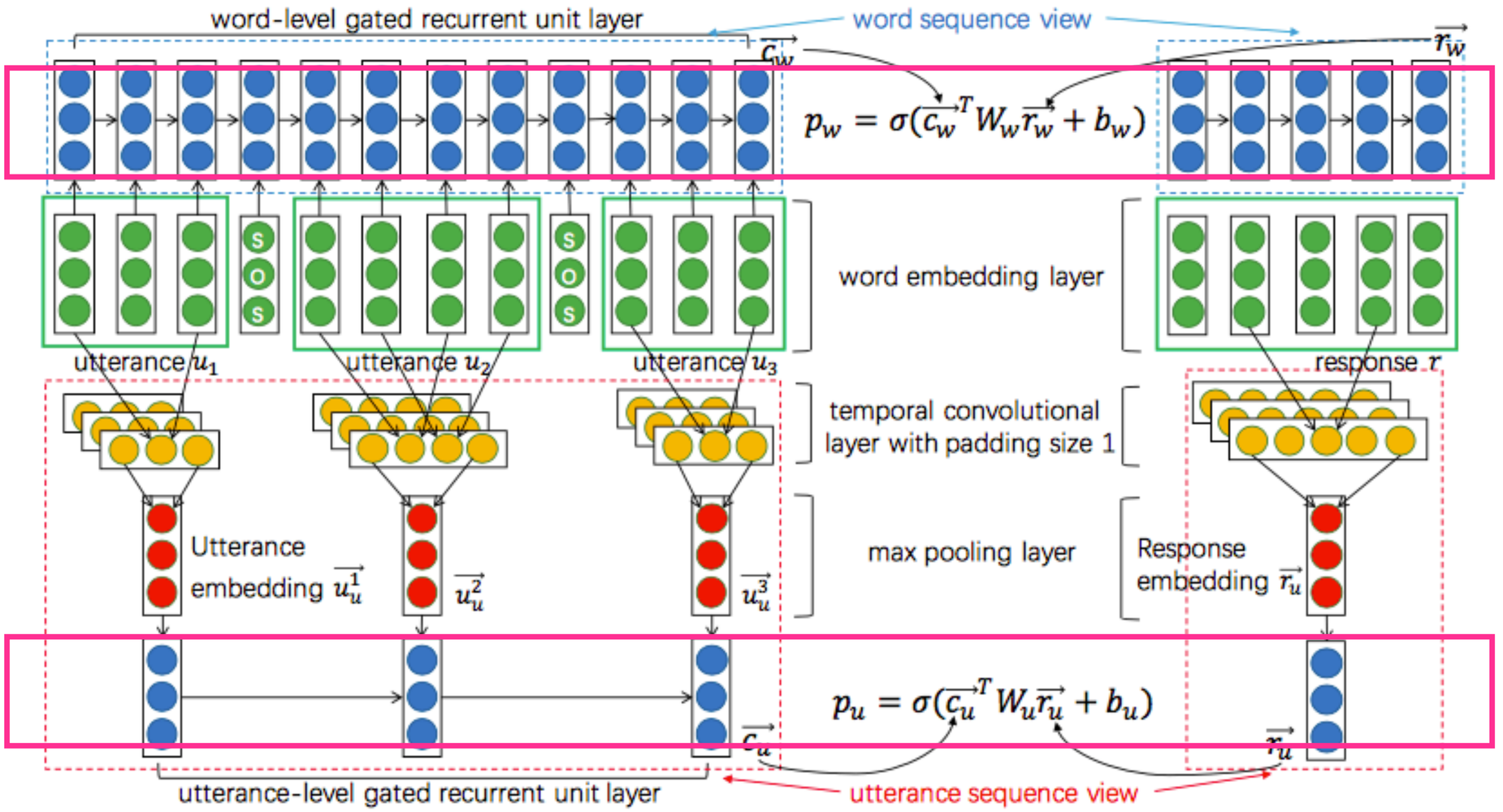
Background


Hello, How are you
I'm fine, you look so happy today
Yes, I got a perfect score in the math exam
Great! You are such a clever boy.
Retrieval-based
Generation-based
Great! You are such a clever boy.
Oh, It's nice of you.
You will do better in the future
ok yes great you are it's
boy nice will better in boy
Candidate response
Vocabulary

Multi-turn Response Selection

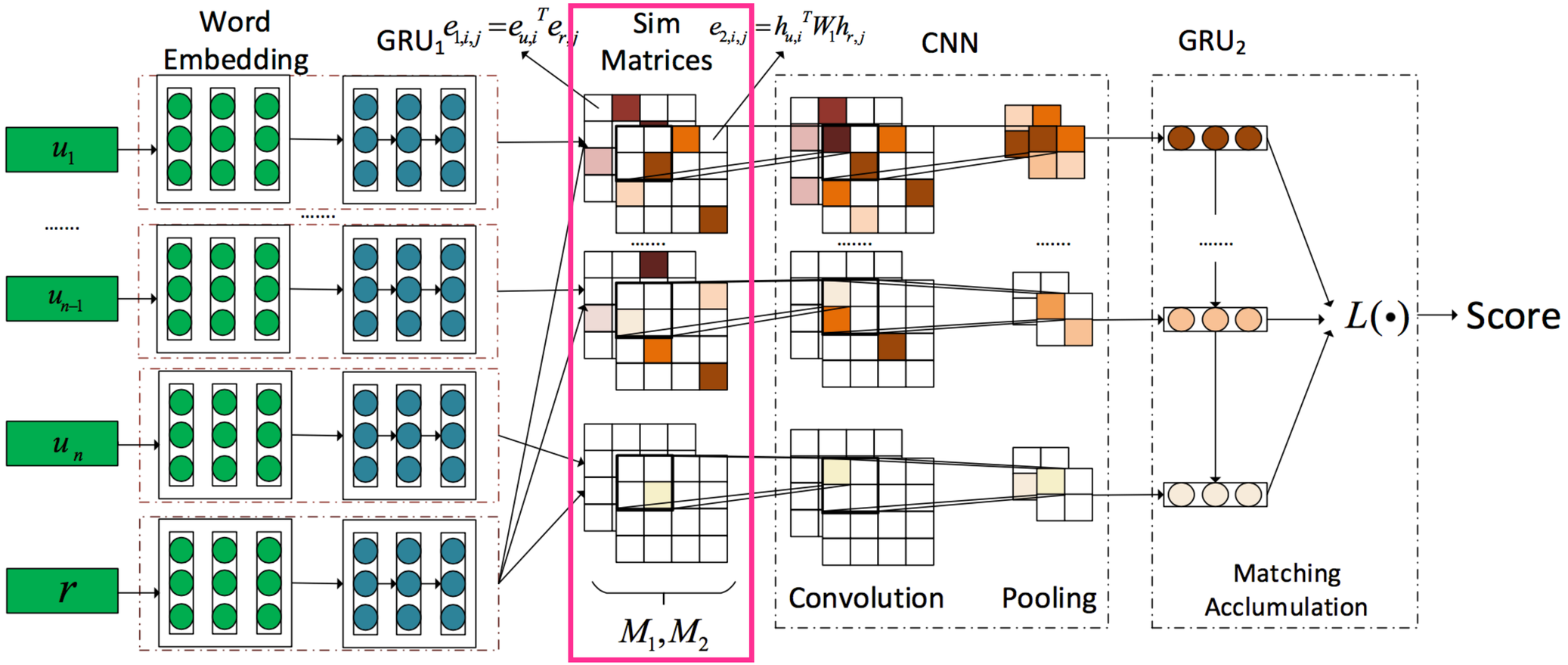


Related works



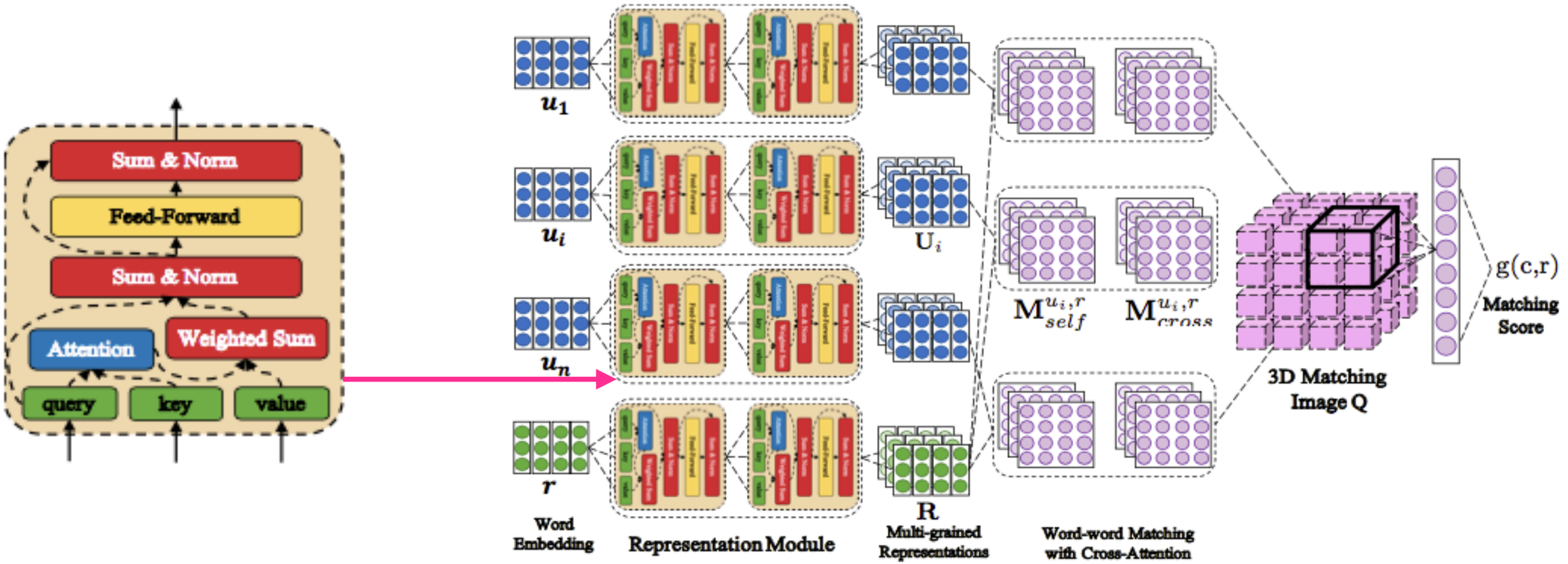
Multi-view (Zhou et al., 2016)

Related works



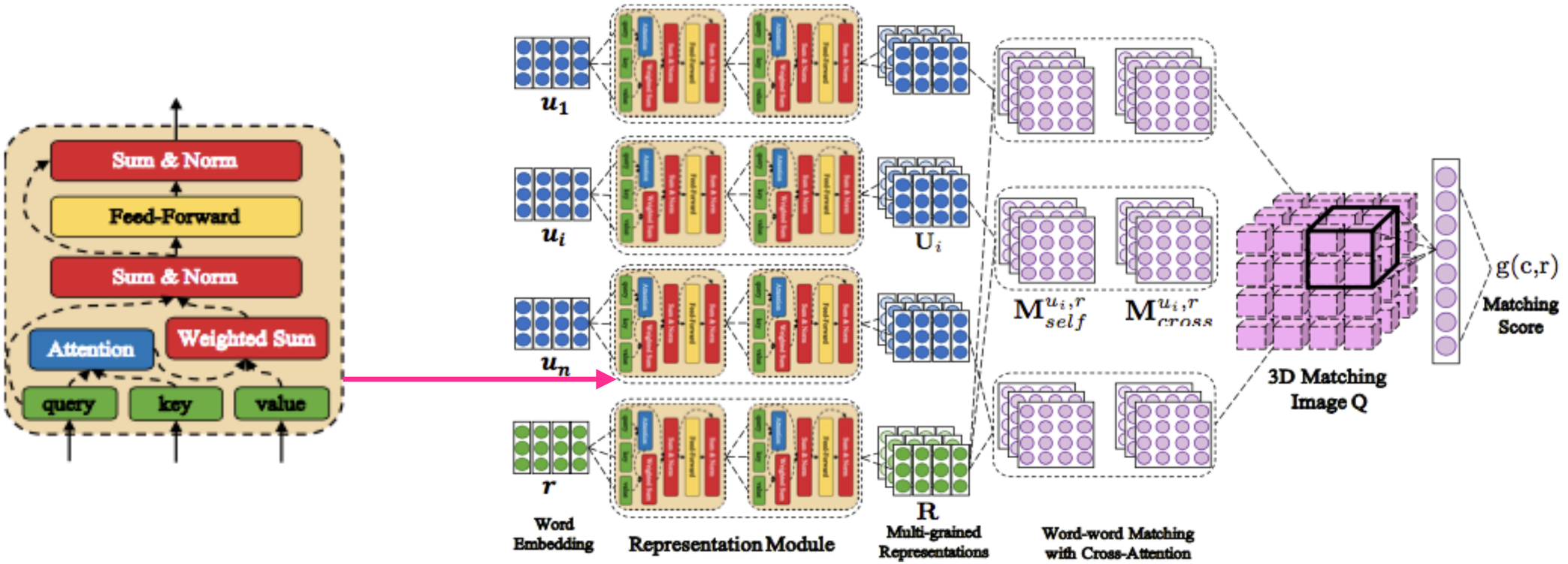
SMN(Wu et al., 2017)

Related works



DAM(Zhou et al., 2018)

Related works



DAM(Zhou et al., 2018)

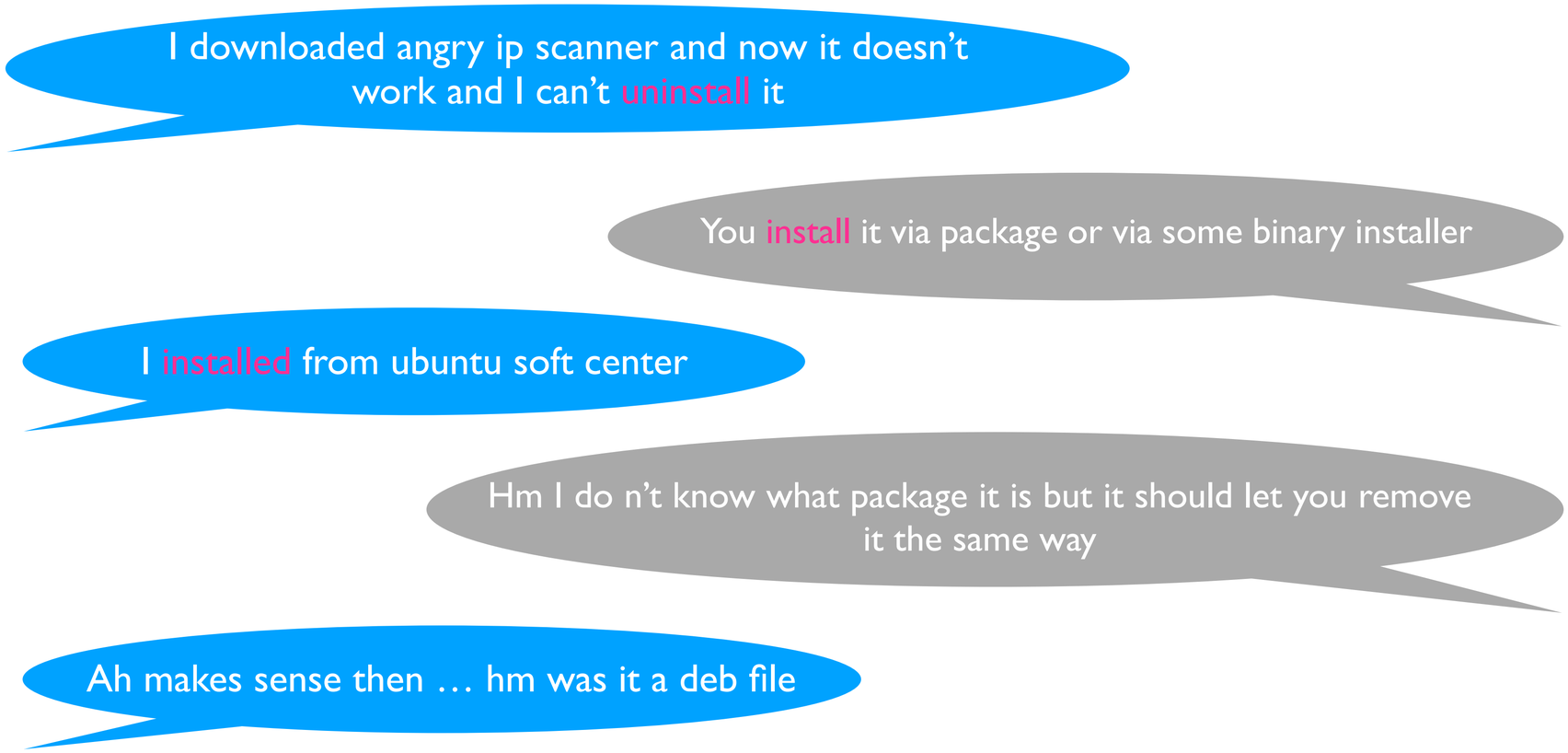
Build the representation and matching in multi-level

Treat all the utterances include query in the same way

Motivation

Does all the utterances (include query) in the context play the same role for multi-turn response selection?

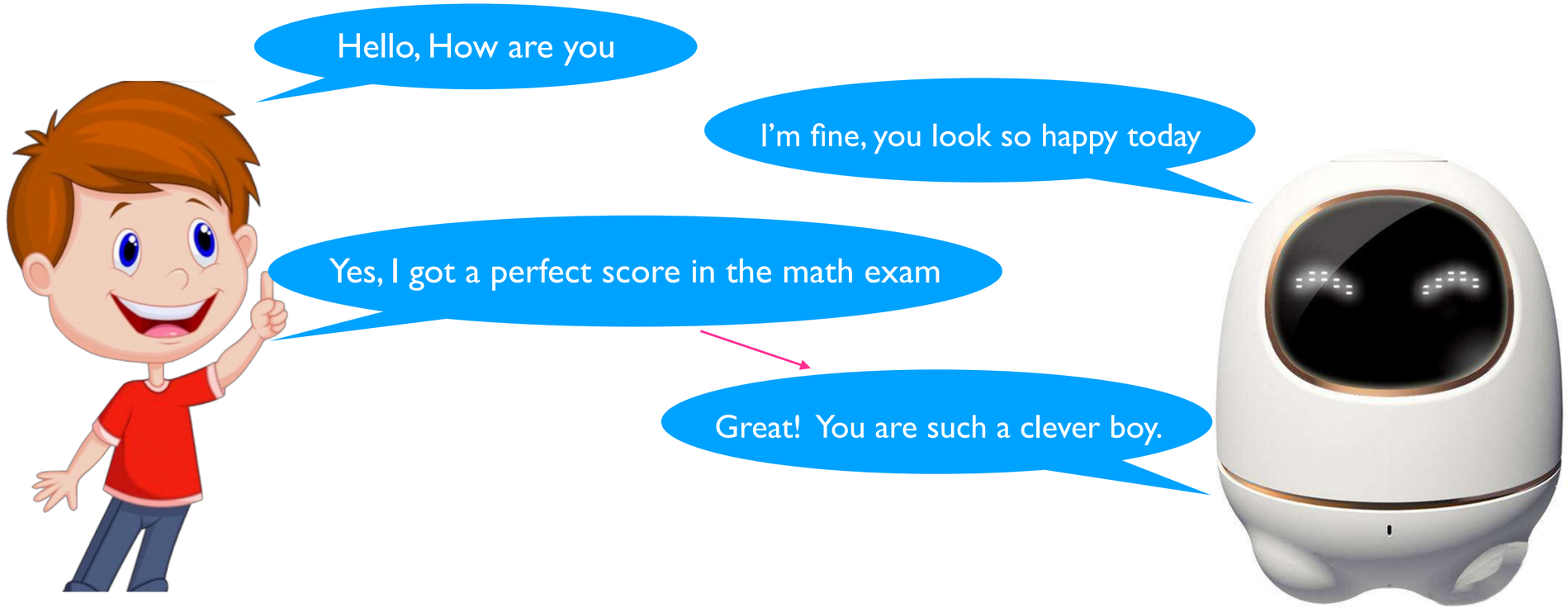
Motivation



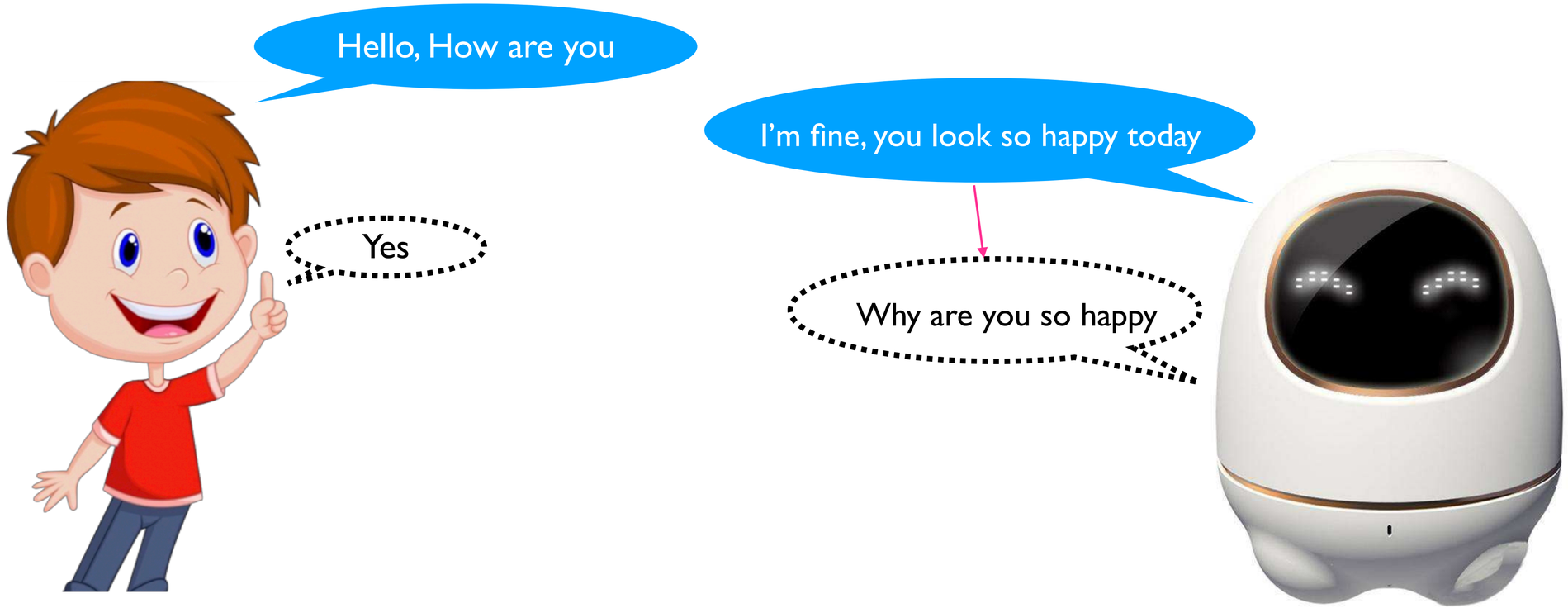
Candidate1: I think it was another format mayge sth starting with r

Candidate2: Thanks I appreciate it try sudo apt-get install

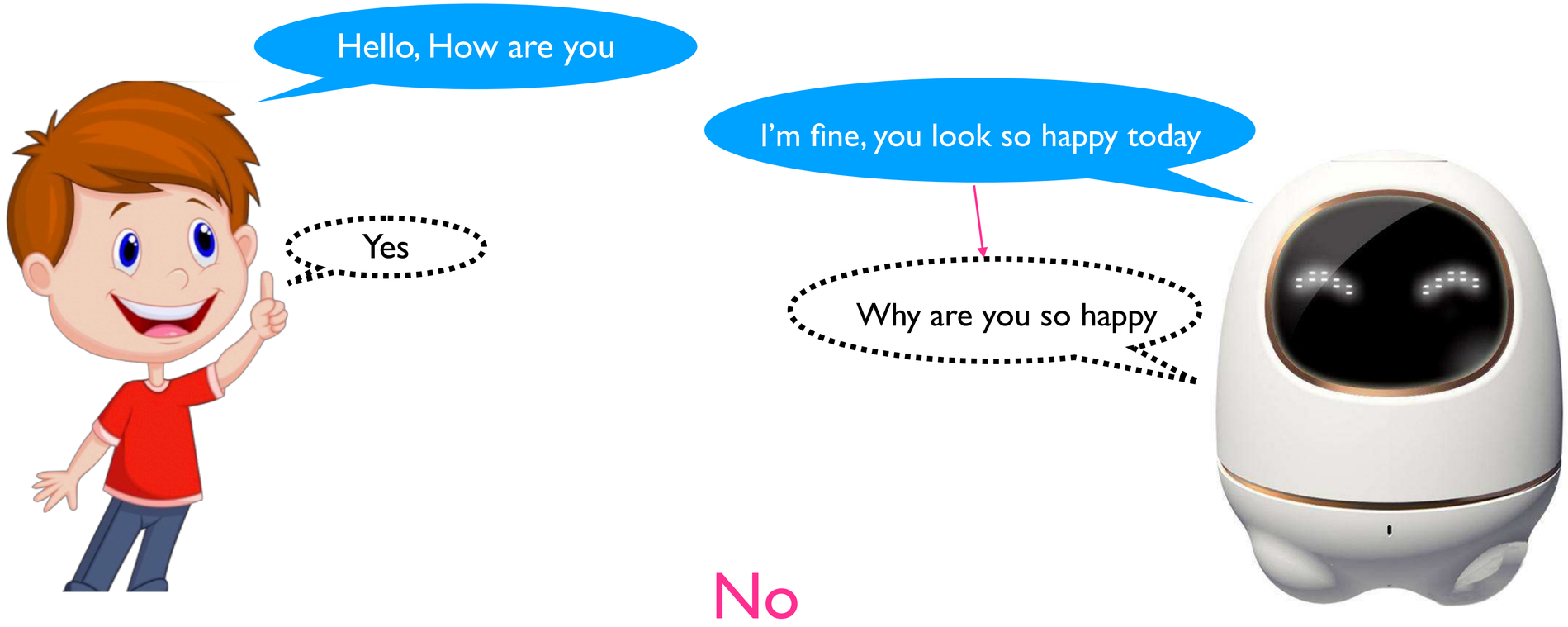
Motivation


Hello, How are you
I'm fine, you look so happy today
Yes, I got a perfect score in the math exam
Great! You are such a clever boy.

Motivation


Hello, How are you
Yes
I'm fine, you look so happy today
Why are you so happy

Motivation



The importance of different utterances depends on query

Motivation

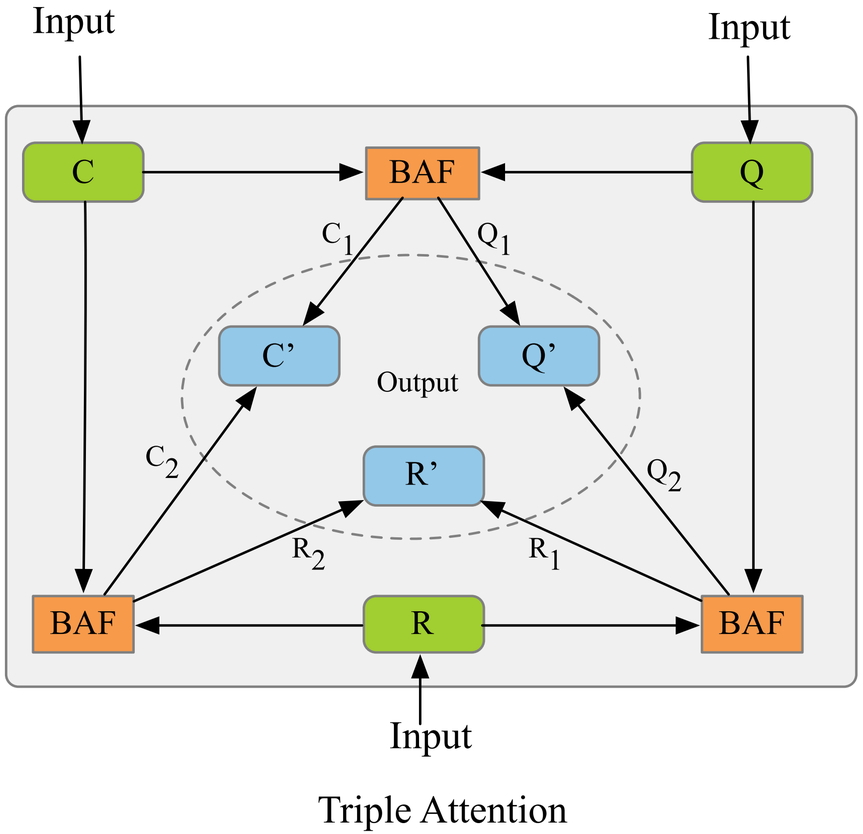
So we need to model the task by the triple <C,Q,R> instead of <C,R>

Model

How to model the relationships within the triple <C,Q,R> ?

Model

How to model the relationships within the triple <C,Q,R> ?



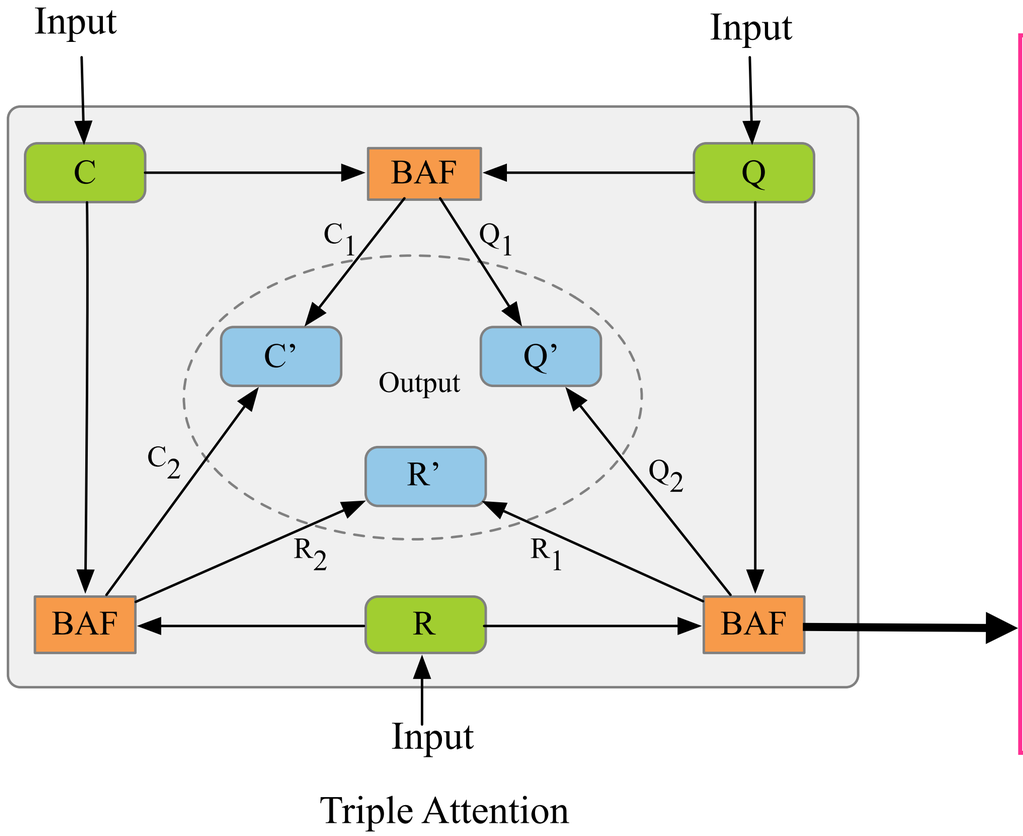
Triple Attention

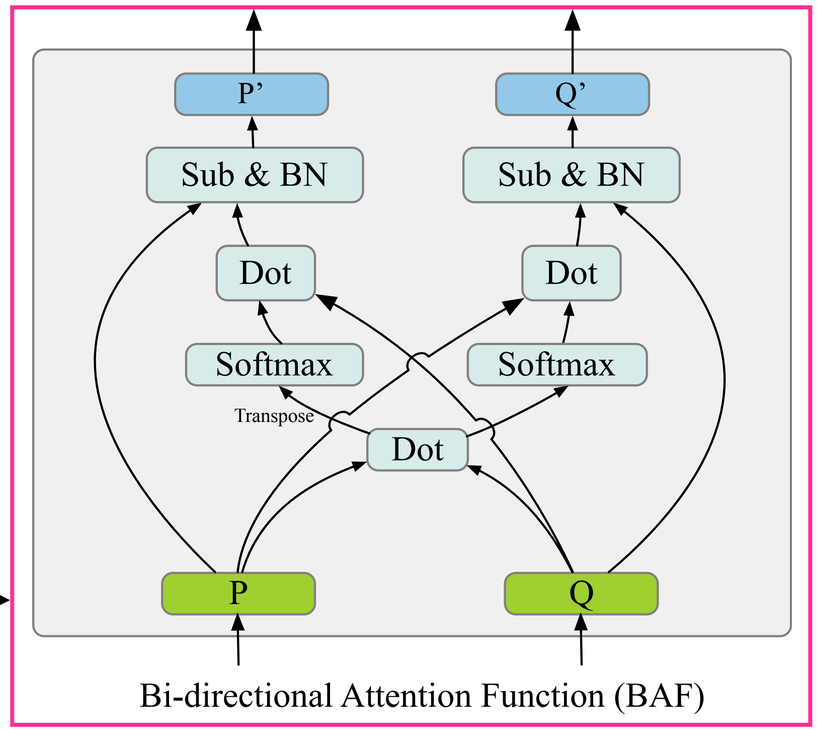
$$C_1, Q_1 = BAF(C, Q)$$
$$C_2, R_1 = BAF(C, Q)$$
$$Q_2, R_2 = BAF(C, R)$$
$$C' = BN(C_1 + C_2)$$
$$Q' = BN(Q_1 + Q_2)$$
$$R' = BN(R_1 + R_2)$$

Triple **Symmetrical** **Unchanged dimension**

Model

How to model the relationships within the triple <C,Q,R> ?





$$M_{pq} = P^T tanh(W_3 Q)$$
$$Att_{pq} = softmax(M_{pq})$$
$$Att_{qp} = softmax(M_{pq}^T)$$
$$P' = P - \tilde{Q};\ \ \tilde{Q} = QAtt_{pq};$$
$$Q' = Q - \tilde{P};\ \ \tilde{P} = PAtt_{qp};$$

Model

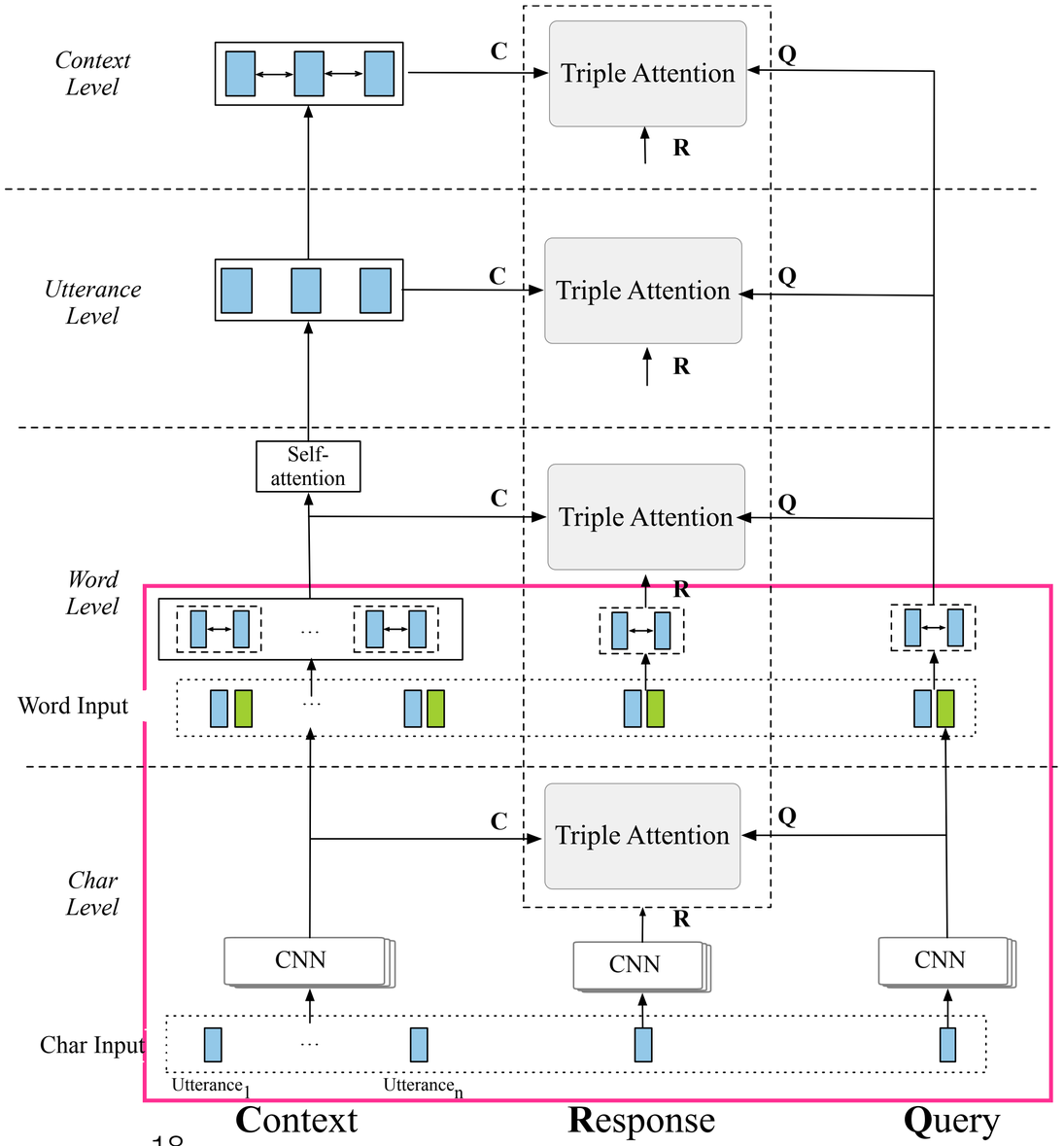
- Hierarchical representation
  - Char-level:

$$ch_{j,t} = tanh(W_1^j * x_{t:t+s_j-1} + b_1^j)$$
$$ch_j = MaxPooling_{t=0}^{L}[ch_{j,t}]$$

  - Word-level:

$$e(x) = [W_e \cdot x; ch(x); MF]$$

$$\overleftarrow{h(x)} = \overleftarrow{\text{LSTM}}(e(x))$$
$$\overrightarrow{h(x)} = \overrightarrow{\text{LSTM}}(e(x))$$
$$h(x) = [\overleftarrow{h(x)}; \overrightarrow{h(x)}]$$



Model

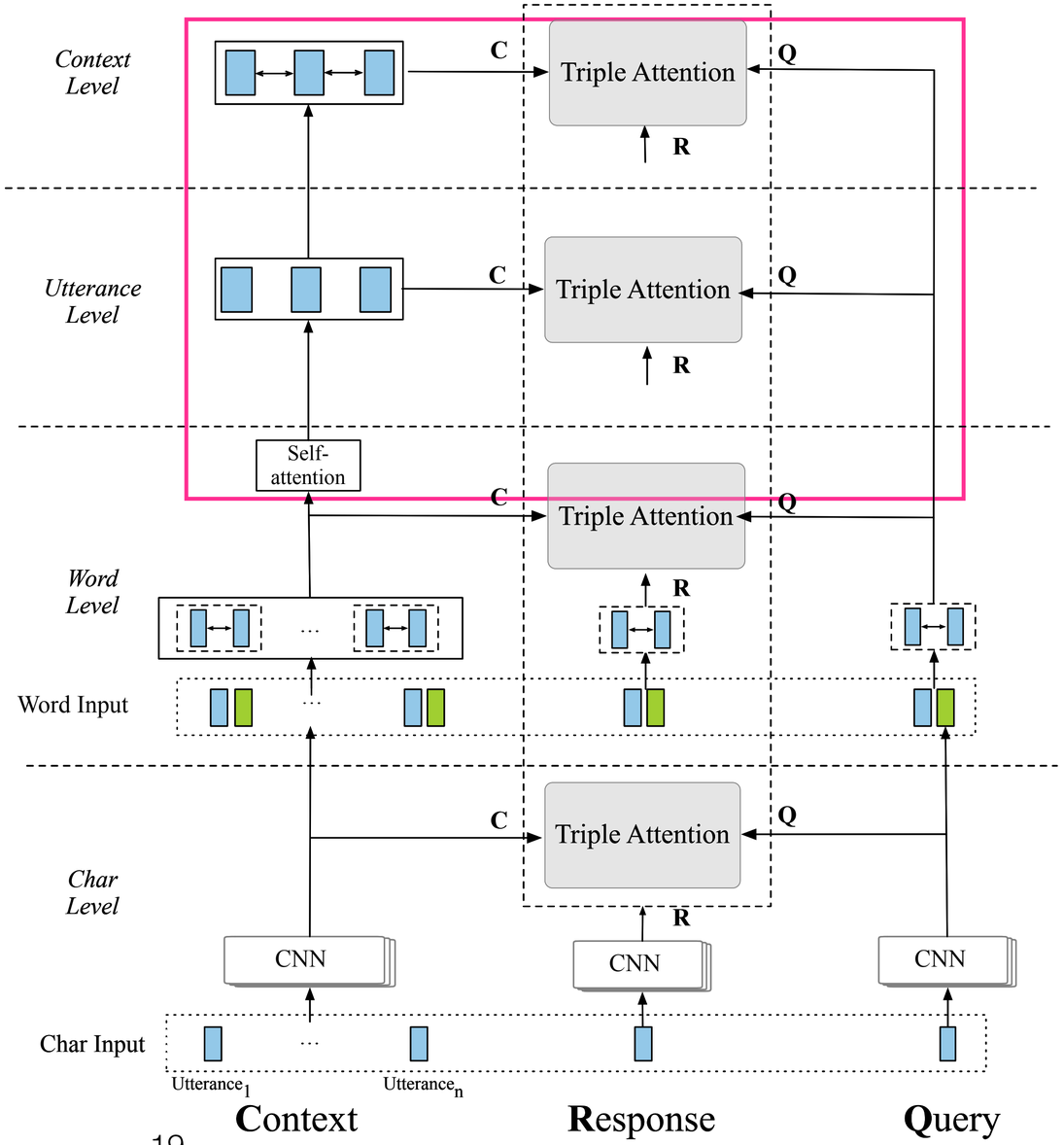
- Hierarchical representation
  - Utterance-level:

$$\alpha_i^k = softmax(W_3 tanh(W_2 h_{u_k}(i)^T))$$

$$u_k = \sum_{i=1}^{m} h_{u_k}^i \alpha_i^k$$

  - Context-level:

$$c_k = \text{Bi-LSTM}([u_k]_{k=1}^n)$$



Model

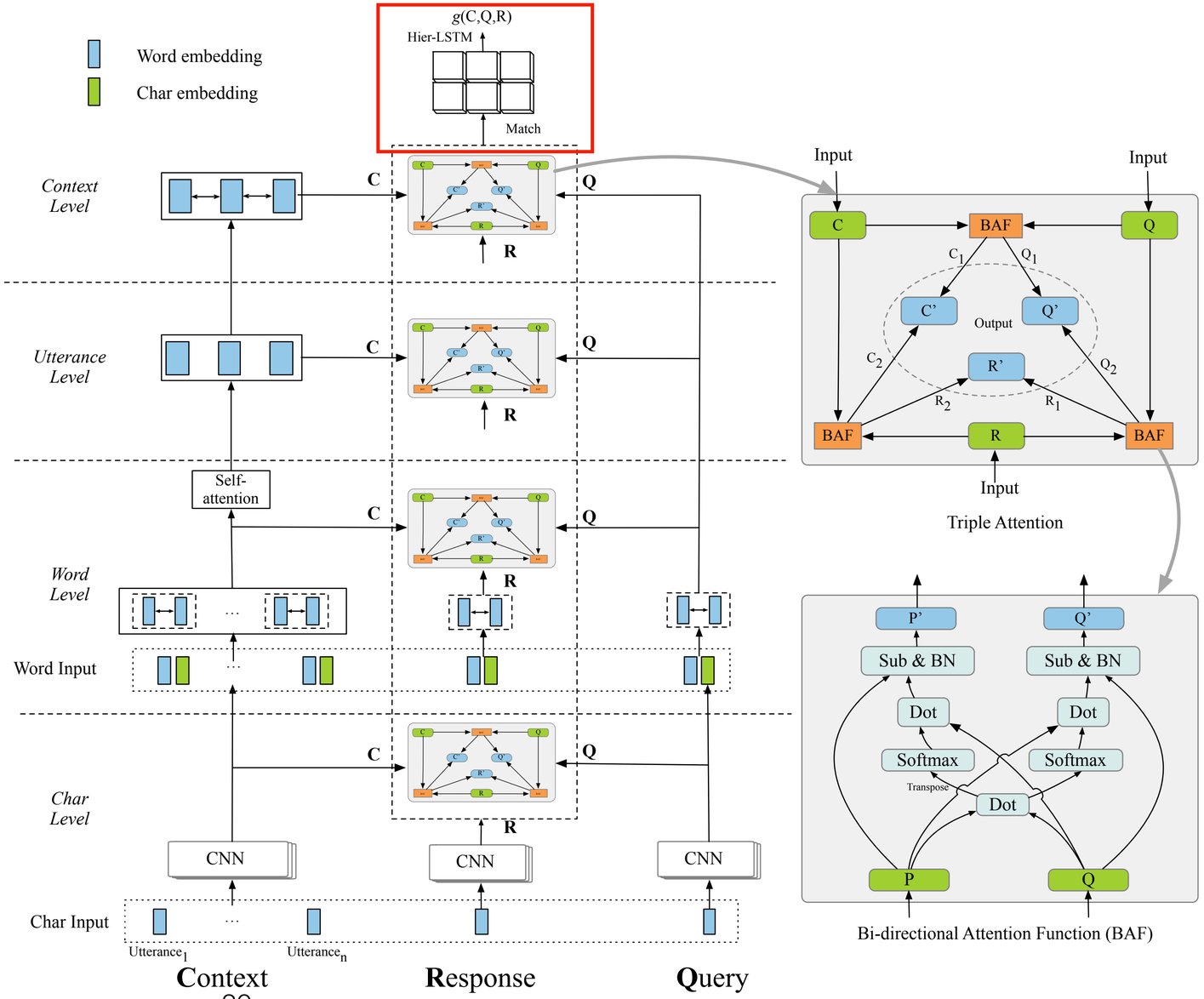
- Triple matching

$$\tilde{M}^1_{rc}(i,k,j) = cosine(ch'_r(i), ch'_{u_k}(j))$$
$$M^1_{rc}(i,k) = \max_{0<j<m} \tilde{M}_1(i,j,k)$$
$$M^1_{rq}(i,j) = cosine(ch'_r(i), ch'_q(j))$$
$$M_1 = [M^1_{rc}(i,k); M^1_{rq}(i,j)]$$

- Fusion and prediction

$$M = [M_1; M_2; M_3; M_4]$$
$$\tilde{m} = MaxPooling^{n+m}_{i=0}[\text{Bi-LSTM}(m_i)]$$
$$v = MaxPooling^{m}_{j=0}[\text{Bi-LSTM}(\tilde{m}_j)]$$

$$g(C,Q,R) = sigmoid(W_4 \cdot v + b_4)$$



Experiment

- Datasets
  - Ubuntu Dialogue Corpus (lowe et al., 2015)
    - Extract from Ubuntu chat logs, technical support for Ubuntu-related problem
    - Train/Dev/Test: 1/0.5/0.5 million session-response pairs
    - Evaluation Metrics: recall at position k in n candidates(Rn@K)
  - Douban Conversation Corpus (Wu et at., 2017)
    - Shares similar format with Ubuntu corpus but is open-domain in Chinese
    - Train/Dev/Test: 1million/0.5 million/10000 session-response pairs
    - Evaluation Metrics: mean average position (MAP), mean reciprocal recall (MRR) , Precision at position 1 (P@1), and Rn@K

Experiment

Overall result

		Ubuntu Dialogue Corpus				Douban Conversation Corpus					
		$R_2@1$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	MAP	MRR	P@1	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$
No-attention	DualEncoder	90.1	63.8	78.4	94.9	48.5	52.7	32.0	18.7	34.3	72.0
	MV-LSTM	90.6	65.3	80.4	94.6	49.8	53.8	34.8	20.2	35.1	71.6
	Match-LSTM	90.4	65.3	80.4	94.6	49.8	53.8	34.8	20.2	34.8	71.0
	DL2R	89.9	62.6	78.3	94.4	48.8	52.7	33.0	19.3	34.2	70.5
	Multi-View	90.8	66.2	80.1	95.1	50.5	54.3	34.2	20.2	35.0	72.9
	SMN	92.6	72.6	84.7	96.1	52.9	56.9	39.7	23.3	39.6	72.4
Attention-based	RNN-CNN	91.1	67.2	80.9	95.6	-	-	-	-	-	-
	DUA	-	75.2	86.8	96.2	*55.1*	59.9	42.1	24.3	*42.1*	*78.0*
	DAM	*93.8*	*76.7*	*87.4*	*96.9*	55.0	*60.1*	*42.7*	*25.4*	41.0	75.7
Our model	TripleNet	**94.3**	**79.0**	**88.5**	**97.0**	**56.4**	**61.8**	**44.7**	**26.8**	**42.6**	77.8
	TripleNet$_{elmo}$	95.1	80.5	89.7	97.6	60.9	65.0	47.0	27.8	48.7	81.4
	TripleNet$_{ensemble}$	95.6	82.1	90.9	98.0	63.2	67.8	51.5	31.3	49.4	83.2

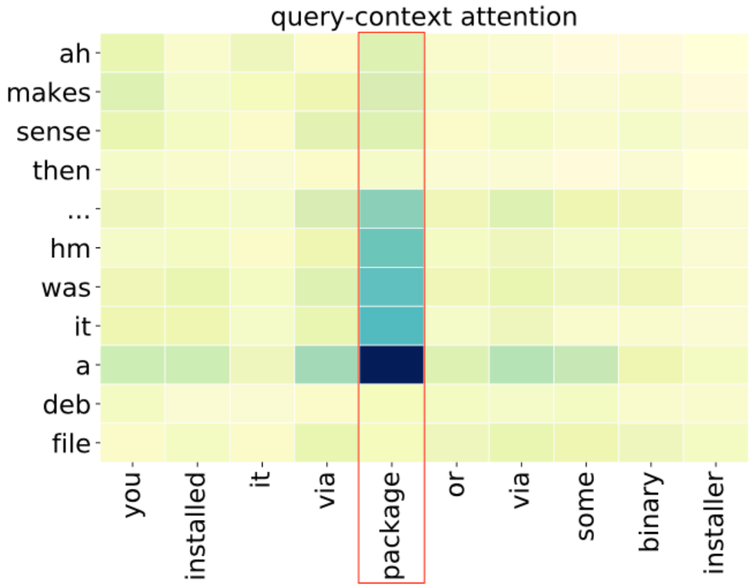
Experiment

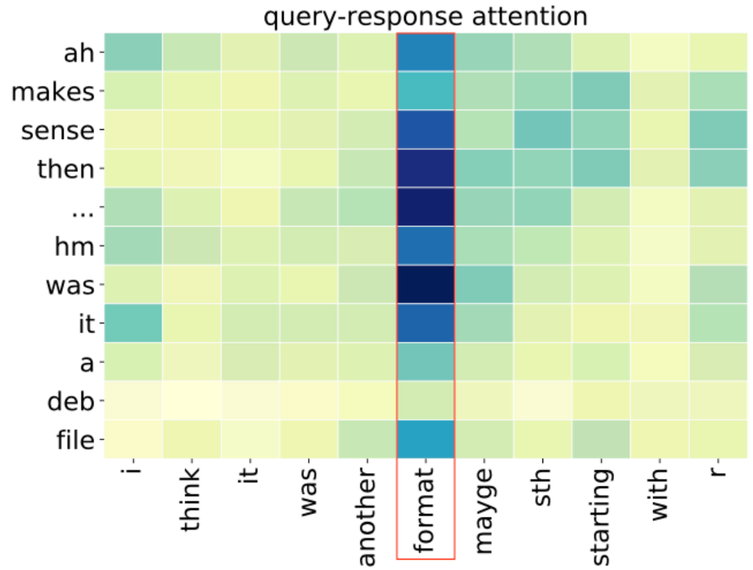
- Model Ablation
  - -TAM: remove triple attention and matching parts;
  - -A_{tri} : remove triple attention
  - -Query: remove query-related parts
  - -A: remove attention-related parts
  - -M: remove matching-related pats
  - -char: remove char-level calculation

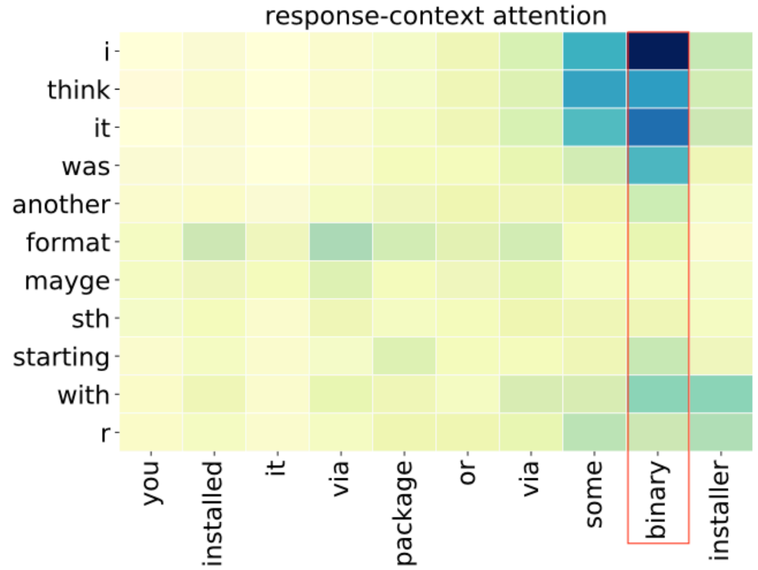
	$R_2@1$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$
TripleNet	94.3	79.0	88.5	97.0
-TAM	93.5	76.6	86.8	96.6
-A_{tri}	93.8	77.6	87.6	96.9
-Query	93.8	77.4	87.3	96.6
-A_{CR}	94.1	78.4	87.9	97.0
-A_{QR}	94.1	78.5	88.1	97.0
-A_{CQ}	94.3	78.7	88.3	97.0
-M_{CR}	93.7	76.9	87.0	96.7
-M_{QR}	94.4	78.5	88.1	97.1
-char	94.1	78.3	88.0	97.1
-word	94.3	78.5	88.2	97.0
-utterance	94.1	78.6	88.1	97.1
-context	94.0	78.4	88.0	97.0

Analysis

Utterance$_1$	A: I downloaded angry ip scanner and now it doesn't word and I can't uninstall it
Utterance$_2$	B: you install it via package or via some binary installer
Utterance$_3$	A: I installed from ubuntu soft center
Utterance$_4$	B: hm I do n't know what package it is but it should let you remove it in same way
Query(U_5)	A: ah makes sense then … hm was it a deb file
Response	B: I think it was another format mayge sth starting with r

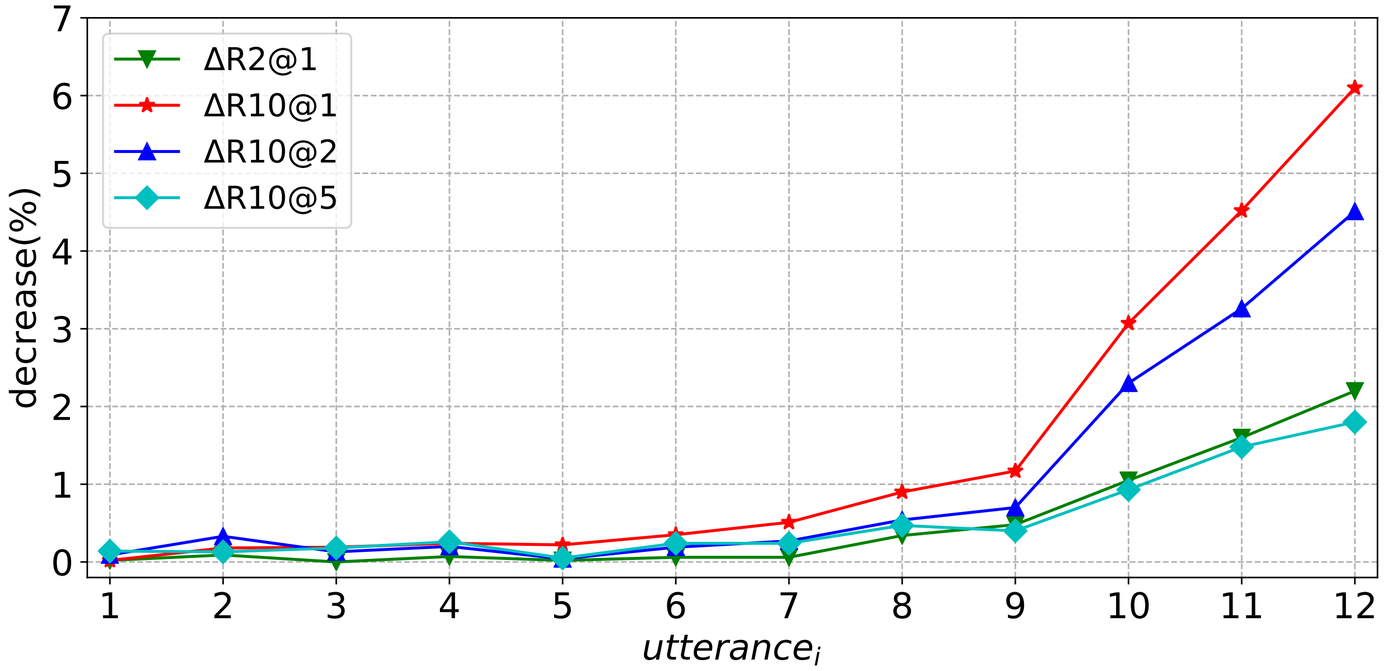






Discussion

- Importance of different utterance
  - Model: TripleNet-Query, remove query-related parts
  - We remove one of the utterance in context both in training and evaluation proces
  - Query: the 12th utterance, which is the last utterance in context



Summary

- We use a novel triple attention mechanism to model the relationships within <C,Q,R> instead of <C,R>;
- We propose a hierarchical representation module to fully model the conversation from char to context level;
- The experimental result on Ubuntu and Douban corpus show that TripleNet significantly outperform the state-of-art result.

What can you take away

- If you are interested in the retrieval-based chatbots, you can try to use the TripleNet to select the response in multi-turn conversation;
- Else if you are interested in some tasks which have three elements, you can try to use the triple attention to model the relationships within the triple;
- Else if you need to deal with Chinese NLP task, you can use the ELMo of Chinese version pretrained by Douban Conversation Corpus in tensorflow.
- Source code: https://github.com/wtma/TripleNet

Thank you
&
Question

Name: Ma Wentao

Email: wtma@iflytek.com