LREC-COLING 2024

UniRetriever: Multi-task Candidates Selection for Various Context-Adaptive Conversational Retrieval

Hongru Wang, Boyang Xue, Baohang Zhou, Rui Wang,
Fei Mi, Weichao Wang, Yasheng Wang, Kam-Fai Wong
{hrwang, kfwong}@se.cuhk.edu.hk

https://rulegreen.github.io/

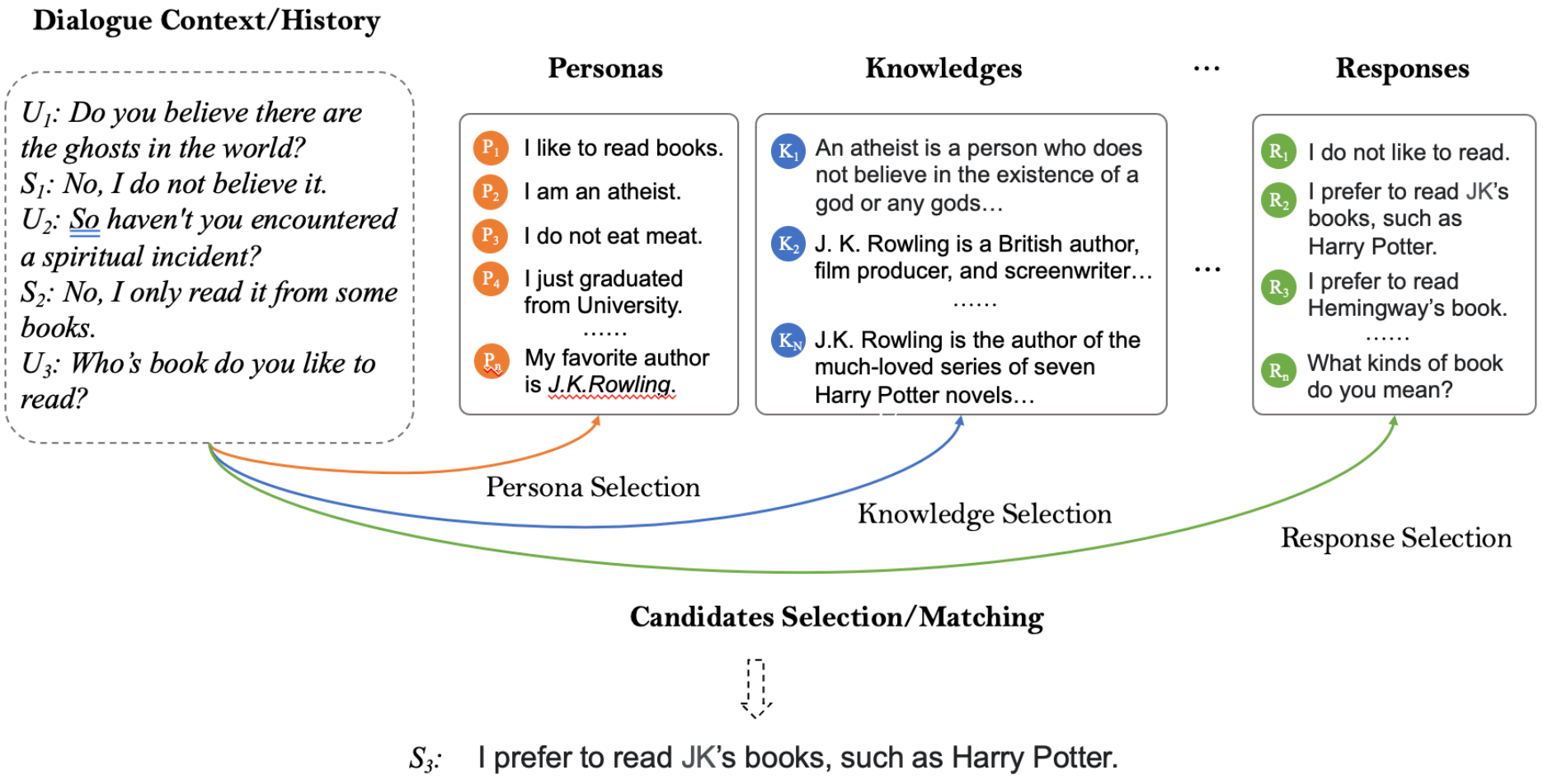
Overview

- Introduction
- Related Work
- UniRetriever
- Experiments
- Conclusion and Future Directions

Introduction

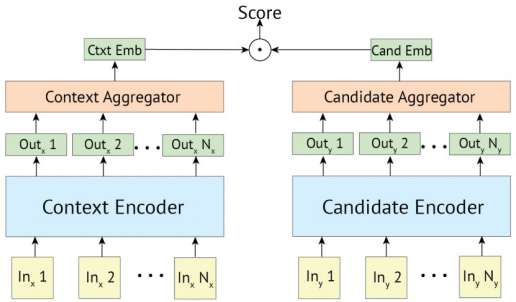
- Dialogue System requires access to various external knowledge sources to deliver **reliable, informative, personalized, and helpful responses**, depending on which sources are invoked.
- Independently training one retriever for one source bring **unnecessary computing cost**, and sometimes there may be complementary effects between different sources.

Introduction

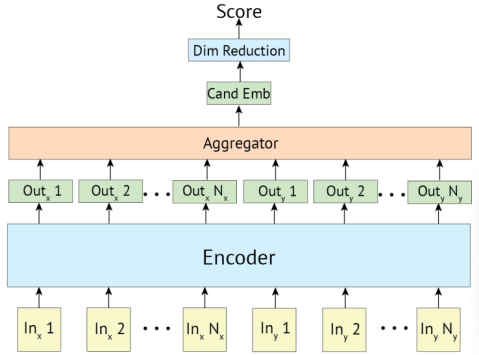


Related Work (Architecture)

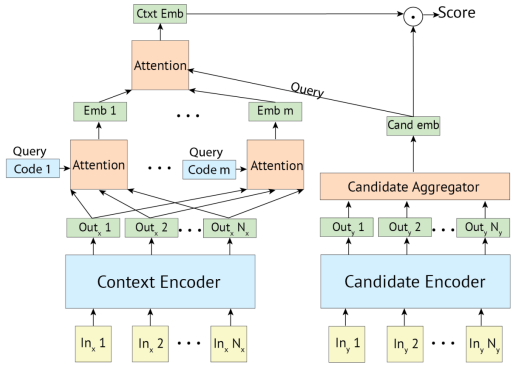
- Bi-encoder
  - Sentence-BERT



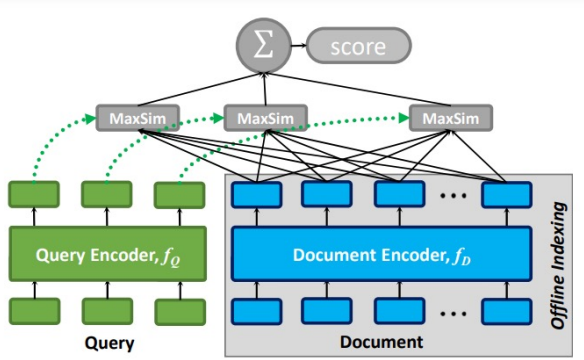
- Cross-encoder
  - BERT
  - RoBERTa
  - ……



- Some Variants



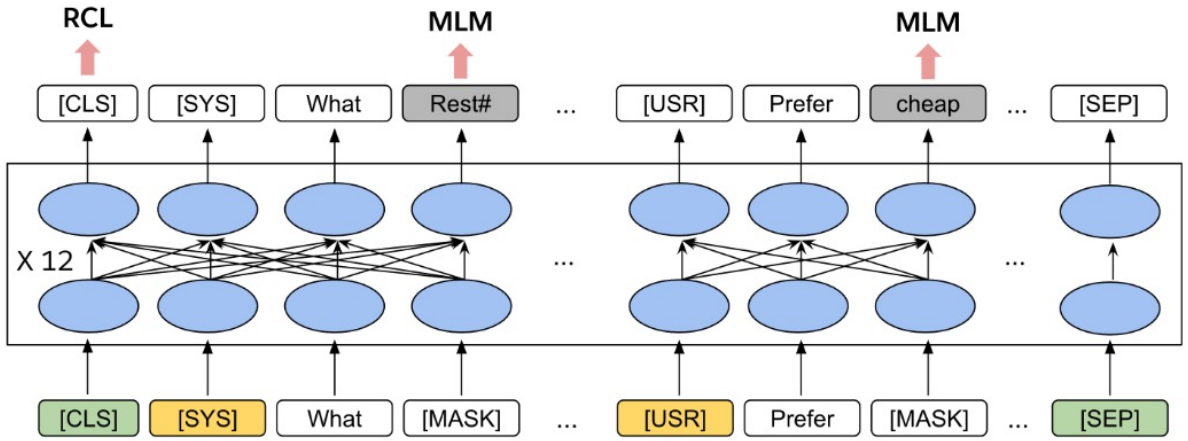
Poly-encoder



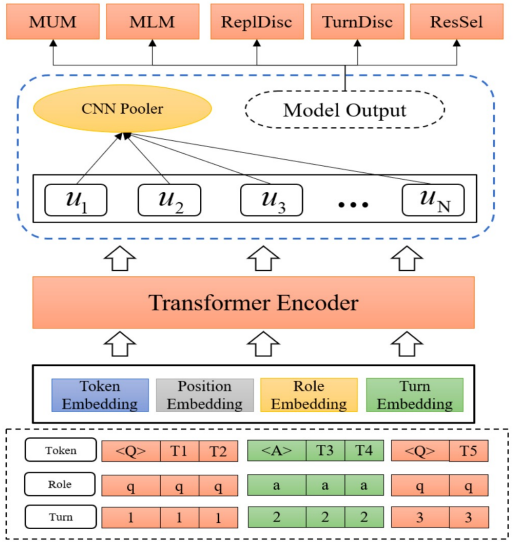
ColBERT

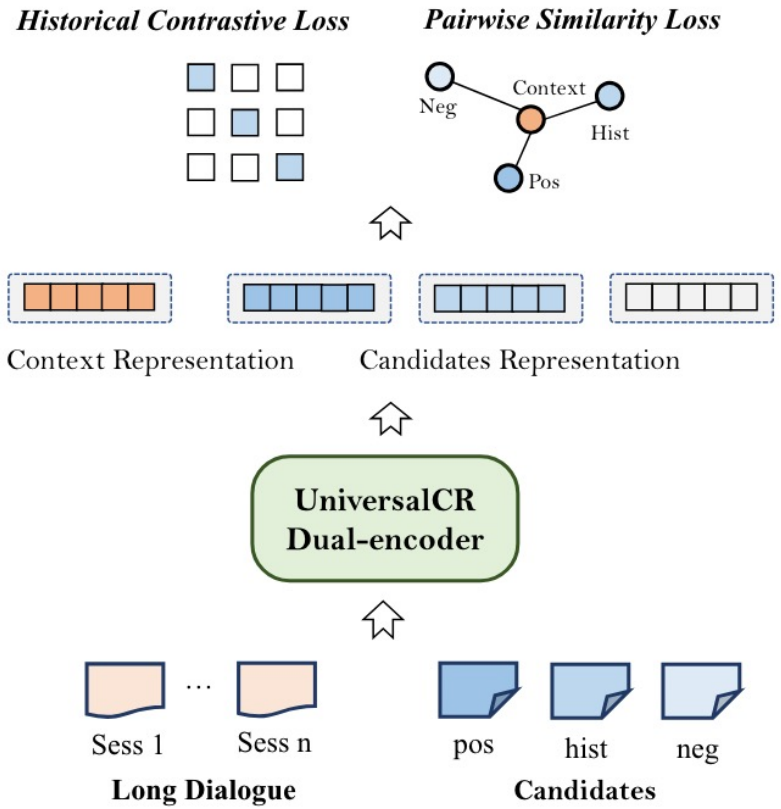
Related Work (Loss)

- Pre-train
  - Tod-BERT
  - DialogueBERT



- Pre-training Objectives
  - MLM
  - RCL
  - MUM
  - ReplDisc
  - ……





- Context-Adaptive Encoder
  - Handle long dialogue such as multi-session
  - Capture more contextual information in long distance
- Candidate Encoder
  - Encode different candidates for different knowledge sources, i.e., memory, persona.
  - Using positive, historical, and negative candidates to capture subtle differences.

Context-adaptive Encoder

- Encode each utterance in the dialogue

$$h_i = \textbf{Enc}(utter)$$

- Using the last user utterance in current session to extract all previous related utterances

$$H_{prev} = TopK_{h_j \in D_{prev}}(sim(h_t^u, h_j))$$

- Learn contextualized embeddings of multi-turn dialogues

$$h_{hist} = \textbf{Attn}(u_t, H_{hist}, H_{hist}) \qquad [H_{prev}; H_{curr}] \text{ to form } H_{hist}$$

- Final representation of context using last user utterance and previous historical embedding

$$h_d = \lambda * h_{hist} + (1 - \lambda) * h_t^u$$
$$\lambda = \sigma(\mathbf{w} * [h_{hist}; h_t^u])$$

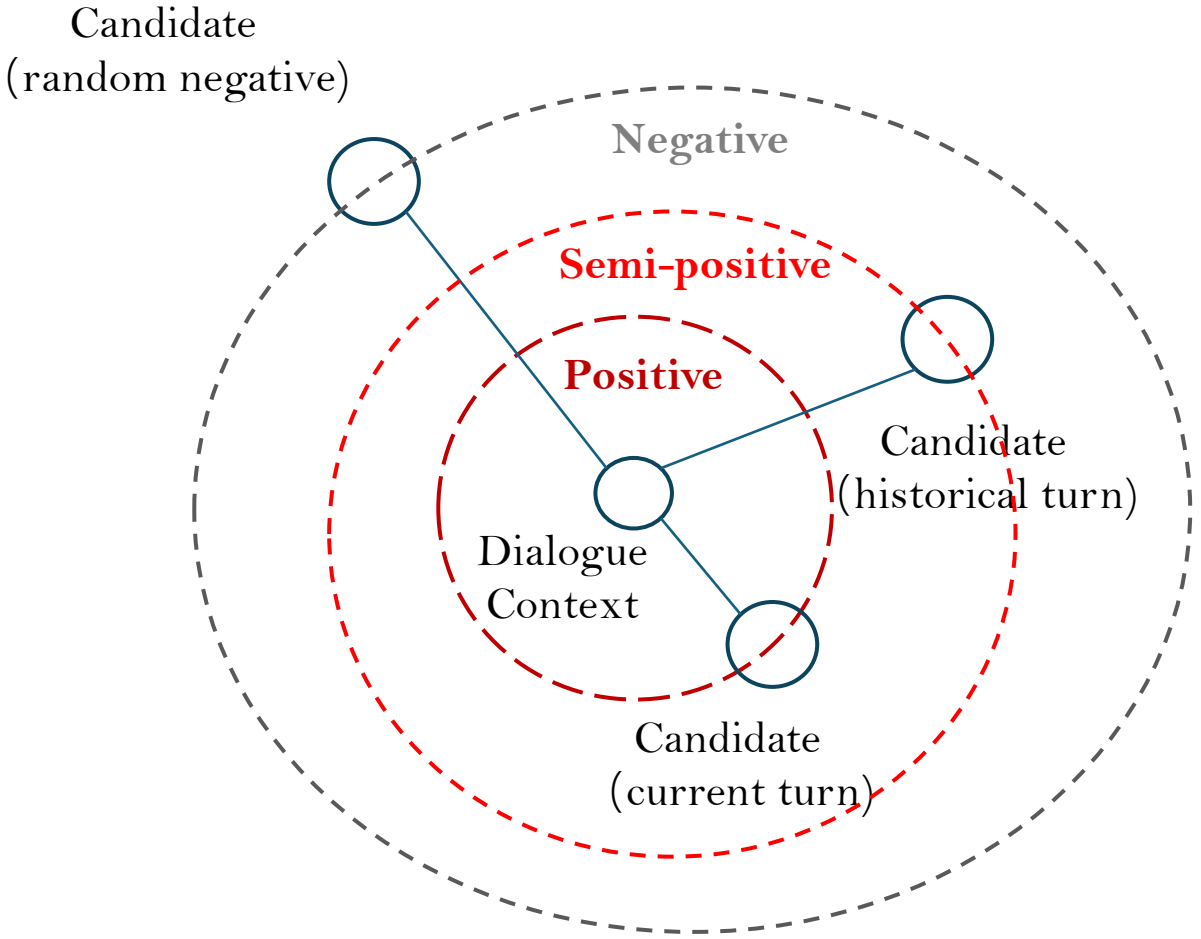
Candidate Encoder

- Encode different candidate from different knowledge sources

$$c_i = \textbf{Enc}(cand)$$

[CLS] [CANDIDATES] w_1, w_2, ..., w_n in which [CANDIDATES] can be re- placed by any candidate selection task indicator token, such as [PERSONA] [KNOWLEDGE] [RESPONSE]

Loss Design



D (pos, context) < D (hist_pos, context)
< D (neg, context)

Circle Loss!

Loss Design

- **Historical Contrastive Learning.** Considering the historical candidates as semi-hard negative samples.

$$\mathcal{L}_{hist} = \frac{e^{sim(h_d^i, c_i^+)}}{\sum_{j \in \mathcal{B}} e^{sim(h_d^i, c_j^+)} + e^{sim(h_d^i, c_i^-)}}$$

- **Pair-wise Similarity Loss.**

$$\mathcal{L}_{pair} = log[1 + \sum_{i=1}^{K} e^{\gamma(s_{neg}^i - s_{hist}^i)} + \sum_{j=1}^{L} e^{\gamma(s_{hist}^j - s_{pos}^j)}]$$

Experiment

- **Three** candidate selection tasks
  - Persona selection
  - Knowledge selection
  - Response selection
- **Six** datasets
  - Three are used in main experiments
  - Three are used to evaluate the generalization capability

Datasets	#Train	#Dev	#Test	#All
DuLeMon	28,243	1,993	2,036	30,202
KBP	4,788	589	584	5,961
Dusinc	2,565	319	359	3,243
KiDial	21795	2813	2580	27188
Diamante	29,758	2,548	2,556	34,862
KdConv	26,038	3,759	3,968	33,765
All	113187	12021	12083	137291

Results

Model	Persona Sel.			Knowledge Sel.			Response Sel.		
	R@1	R@5	MRR	R@1	R@5	MRR	R@1	R@5	MRR
BM25	0.06	0.38	9.49	0.35	1.45	28.37	0.59	1.35	30.72
DPR	7.38	17.62	-	29.18	57.91	-	11.85	36.35	-
MultiCPR	10.70	17.27	-	41.45	58.81	-	9.65	19.15	-
RocketQAv1	21.39	52.02	34.23	37.86	65.91	42.86	21.46	71.20	41.92
SentenceBERT	18.99	47.64	32.91	43.57	86.59	61.13	35.45	73.59	51.86
Bi-Encoder	26.79	56.13	40.46	80.08	98.88	86.14	52.93	90.73	68.69
Poly-Encoder 16	26.26	55.76	40.08	79.11	99.11	85.61	51.17	**91.51**	67.83
Poly-Encoder 32	25.73	55.76	39.80	78.76	98.91	85.46	50.67	90.88	67.49
RocketQAv2	21.87	50.80	34.27	34.62	61.64	39.71	21.87	70.23	42.31
UniversalCR$_{single}$	28.43	55.17	40.99	85.65	98.72	90.69	57.20	85.68	69.29
UniversalCR$_{full}$	**28.73**	**55.99**	**41.47**	**86.67**	**99.22**	**91.45**	**59.94**	87.09	**71.73**

Table 2: The performance of our proposed model and baselines on dataset DuLemon (Xu et al., 2022), KiDial, and Diamante (Lu et al., 2022a), correspond to persona selection, knowledge selection, and response selection. UniversalCR$_{sigle}$ simply fine-tune our model on each dataset instead of all in UniversalCR$_{full}$.

Model	P.R@1	K.R@1	R.R@1
UniversalCR$_{full}$	28.73	86.67	59.94
– *context enc.*	21.10	83.64	53.68
– $\mathcal{L}_{pair}$	28.52	85.27	56.65
– $\mathcal{L}_{hist}$	20.14	43.64	34.12

Table 3: Ablation Study. The - *context enc.* stands for considering all utterances in dialogue history by using a mean representation, and - $\mathcal{L}_{pair}$ and - $\mathcal{L}_{hist}$ means removing the corresponding loss constraint.

Model	P.R@1	K.R@1	R.R@1
full concatenation	31.41	89.34	65.34
context enc.	28.73	86.67	59.94

Table 4: The performance of different ways to process the dialogue history in which the full concatenation can be viewed as our theoretical upper bound.

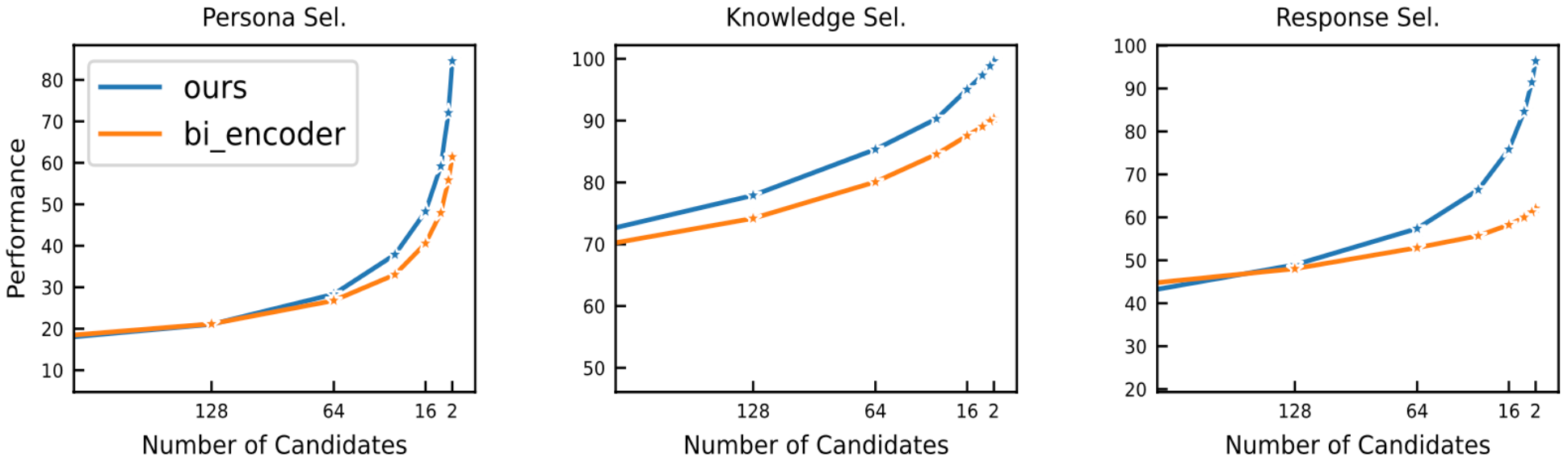


Figure 5: The Performance of UniversalCR$_{full}$ on different selection tasks: persona selection, knowledge selection, and response selection, with the number of candidates ranging from 256 to 2.

Conclusion

- We propose a universal conversational retrieval framework, unifying three dominant candidate selection tasks: *persona selection*, *knowledge selection*, and *response selection*, in one framework while keeping the bottleneck layer as a single dot-product with a fixed size to achieve the balance of effectiveness and efficiency.

- We design one context-adaptive encoder and two carefully crafted loss constraints to address lengthy dialogue and capture subtle differences across various candidates respectively.

- We conduct extensive experiments to demonstrate the superiority of our proposed framework on six datasets in both supervised and unsupervised settings. Besides that, we offer an in-depth analysis of various candidate pool sizes and different context processing methods. These findings suggest a promising path toward building a robust and universal dialogue retrieval framework.

Thank You

Q&A

Homepage

Zhihu

WeChat