東北大學
Northeastern University

BERT-BC: A Unified Alignment and Interaction Model over Hierarchical BERT for Response Selection

Zhenfei Yang, Beiming Yu, Yuan Cui, Shi Feng, Daling Wang, Yifei Zhang

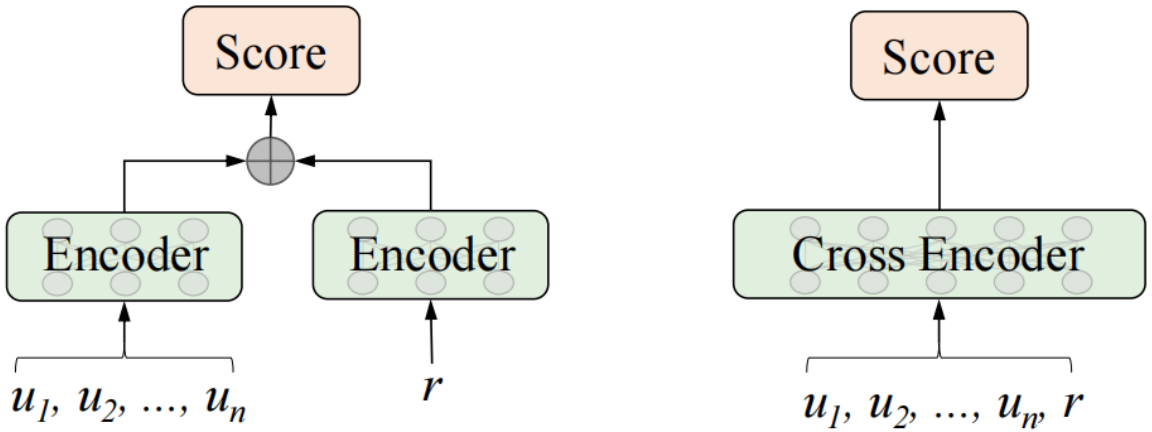
Northeastern University, Shenyang, China
LREC-COLING 2024, **Torino**

Content

01 / Introduction

Representation-based Bi-Encoder and Interaction-based Cross-Encoder

- Representation-based Bi-Encoder
  - The context and replies are encoded into semantic representations using Bi-Encoder respectively, and the similarity scores are computed afterwards.
  - Ignoring the multiple utterance dependency and logical reasoning relationship between response and context.
- Interaction-based Cross-Encoder
  - Cross-Encoder directly feed the concatenation of context and response into the pre-trained model for interactive inference.
  - Ignoring the comprehensively independent representation modeling of context and response.



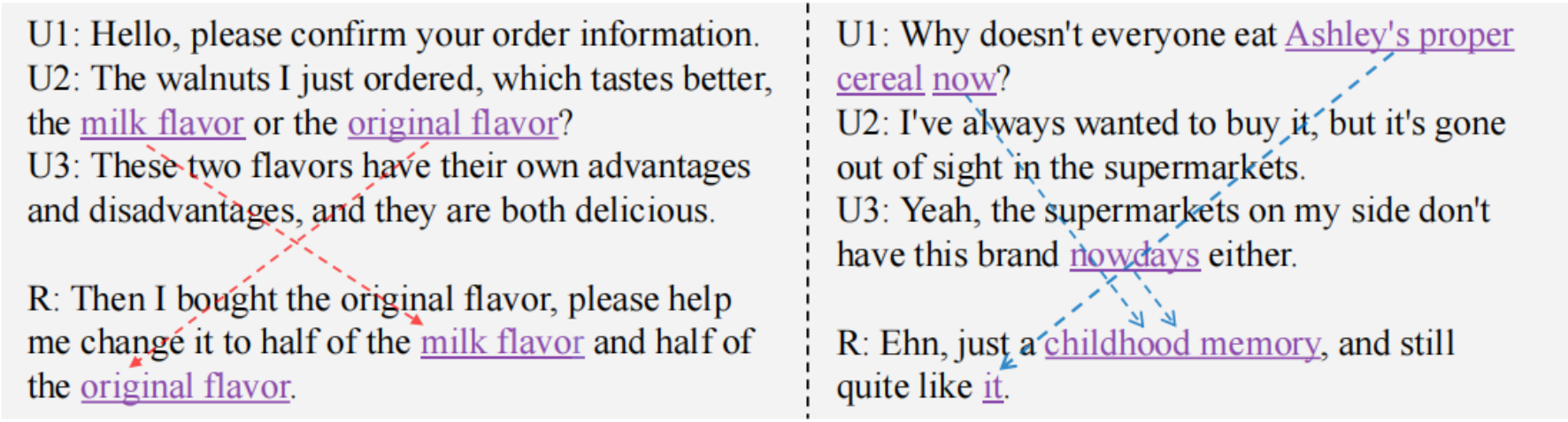
(a) Bi-Encoder (b) Cross-Encoder

$u_1, u_2, ..., u_n$ denotes the context and r denotes the response

01 / Introduction

Semantically Related Sample and Conversationally Related Sample.

- The Bi-Encoder model is computationally fast with low cost and performs better for the **semantically related samples** (Figure (a)) with **high keyword co-occurrence and semantic approximation** between context and response.
- The Bi-Encoder cannot solve **conversationally related samples** that require **conversational-level understanding** and **relational reasoning**.
- Most conventional methods leverage simple heuristics to construct negative samples, which makes it challenging to distinguish stronger distractors in realistic scenarios.



(a) semantically related sample (b) conversationally related sample

01 / Introduction

Our Contribution

- We propose a pretrained response selection model BERT-BC, which combines the representation-based Bi-Encoder and the interaction-based Cross-Encoder.
- We devise a multiple contrastive learning method to enhance the Bi-Encoder's ability to learn **semantically related samples** by align context and response, and propose a hard negative resampling strategy to enhance the Cross-Encoder's interaction ability to learn **conversationally related samples**.
- The empirical results show that our approach can achieve new state-of-the-art performance on three benchmark datasets(*i.e.*, Ubuntu, Douban, E-Commerce).

Content

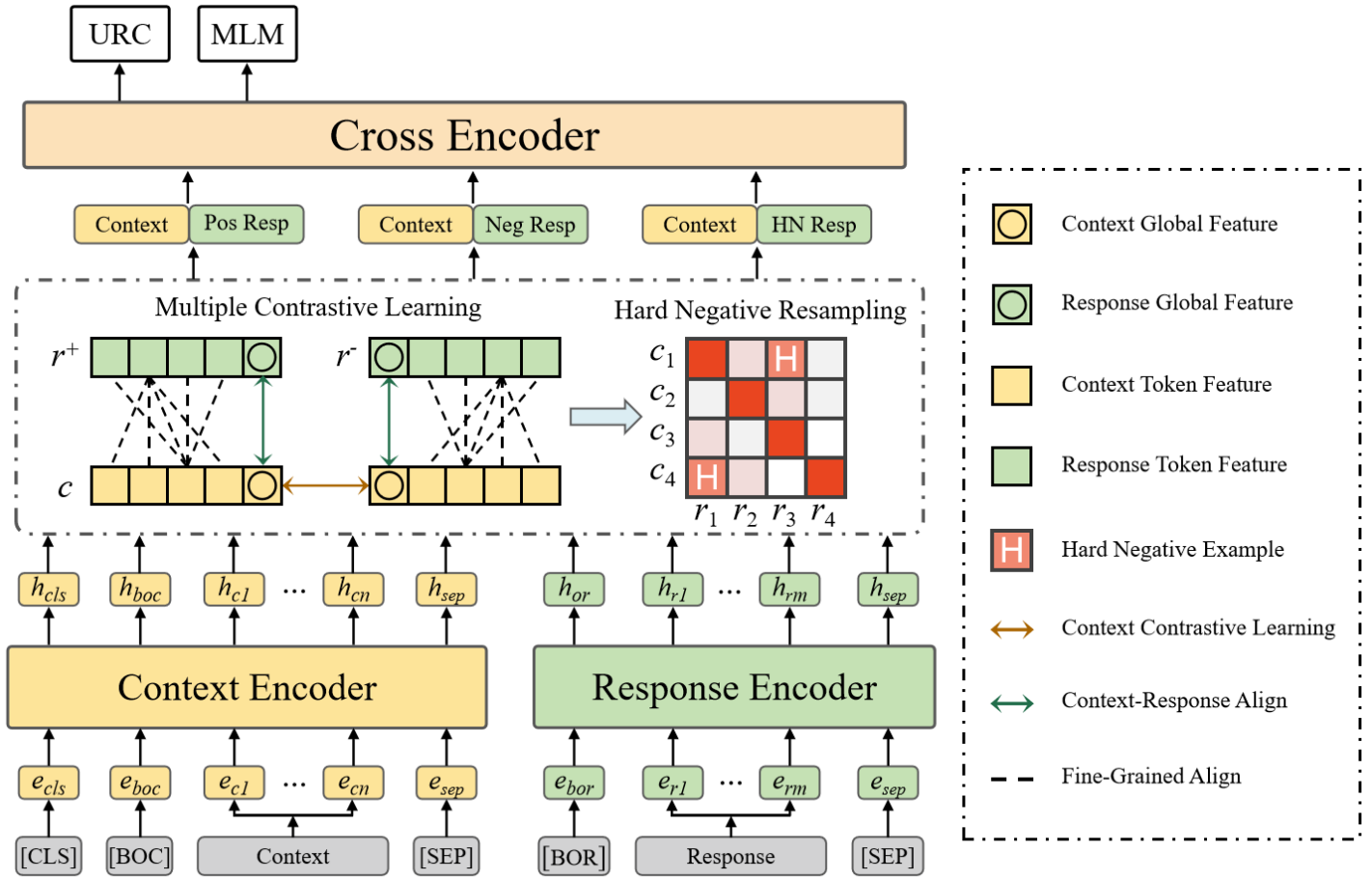
02 / Method

Problem Formulation

Assume that given a conversation dataset consisting of a triplet $D = (c_i, r_i, y_i)_{(i=1)}^{N}$. $c_i = \{u_{i,1}, u_{i,2}, \ldots, u_{i,m}\}$ denotes dialogue history utterances; m indicates the number of utterances in the context; r_i denotes a candidate response; y_i is a label, when $y_i = 1$, r_i is a suitable response about c_i and $y_i = 0$ otherwise. The purpose of the response selection task is to learn a matching model $g(\cdot, \cdot)$. For a given context-response pair (c_i, r_i), the matching scores of c_i and r_i are obtained by $g(c_i, r_i)$.

02 / Method

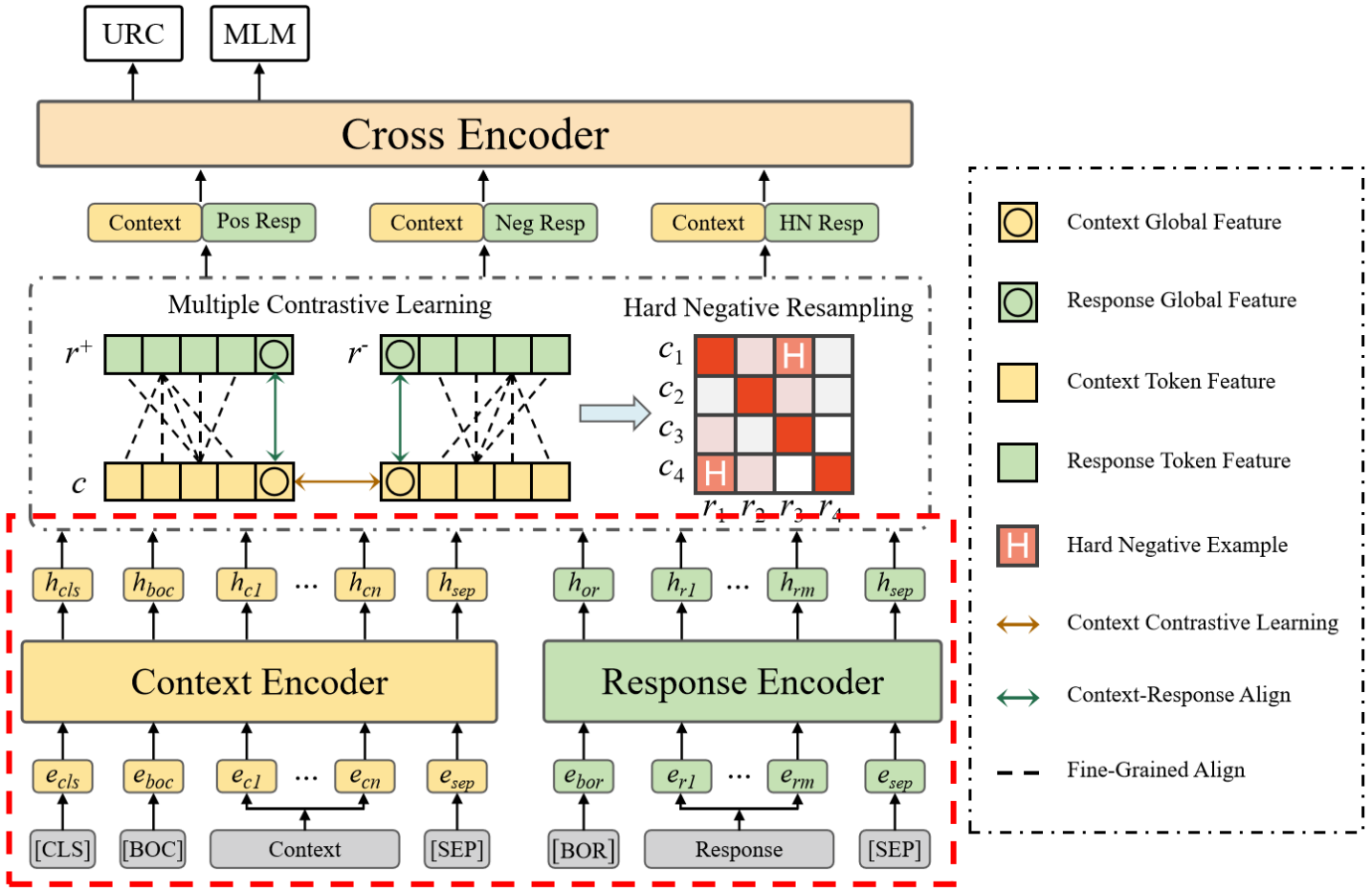
Model Framework



- Context-Response Representation
  - Low-layer BERT Bi-Encoder
  - Multiple Contrastive Learning
- Dialogue Interactive Reasoning
  - High-layer BERT Cross-Encoder
  - Hard Negative Resampling

02 / Method

Low-layer BERT Bi-Encoder

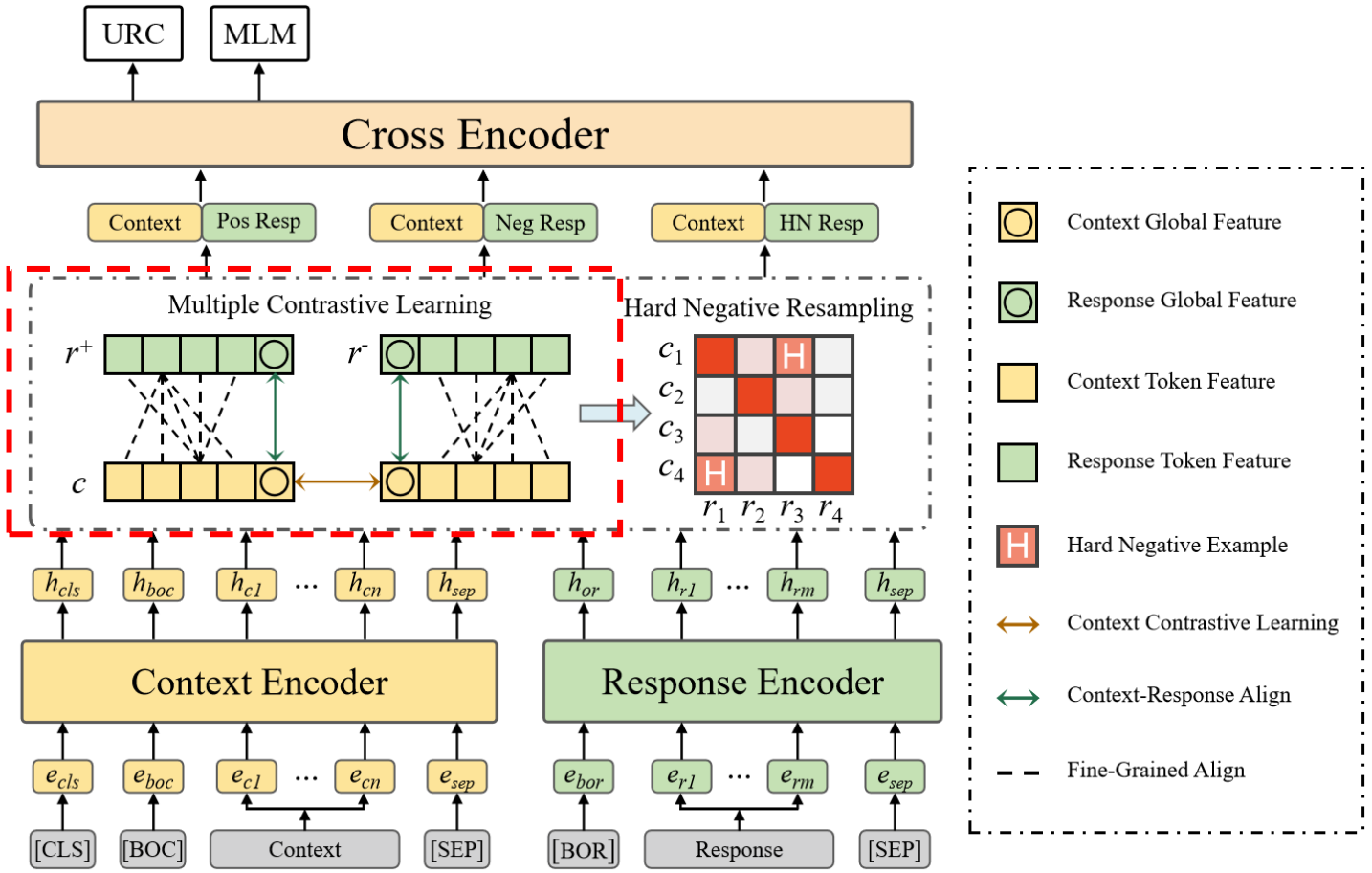


基于对比学习预训练回复检索模型

- Context-Response Representation
  - Low-layer BERT Bi-Encoder
  - Multiple Contrastive Learning
- Dialogue Interactive Reasoning
  - High-layer BERT Cross-Encoder
  - Hard Negative Resampling

02 / Method

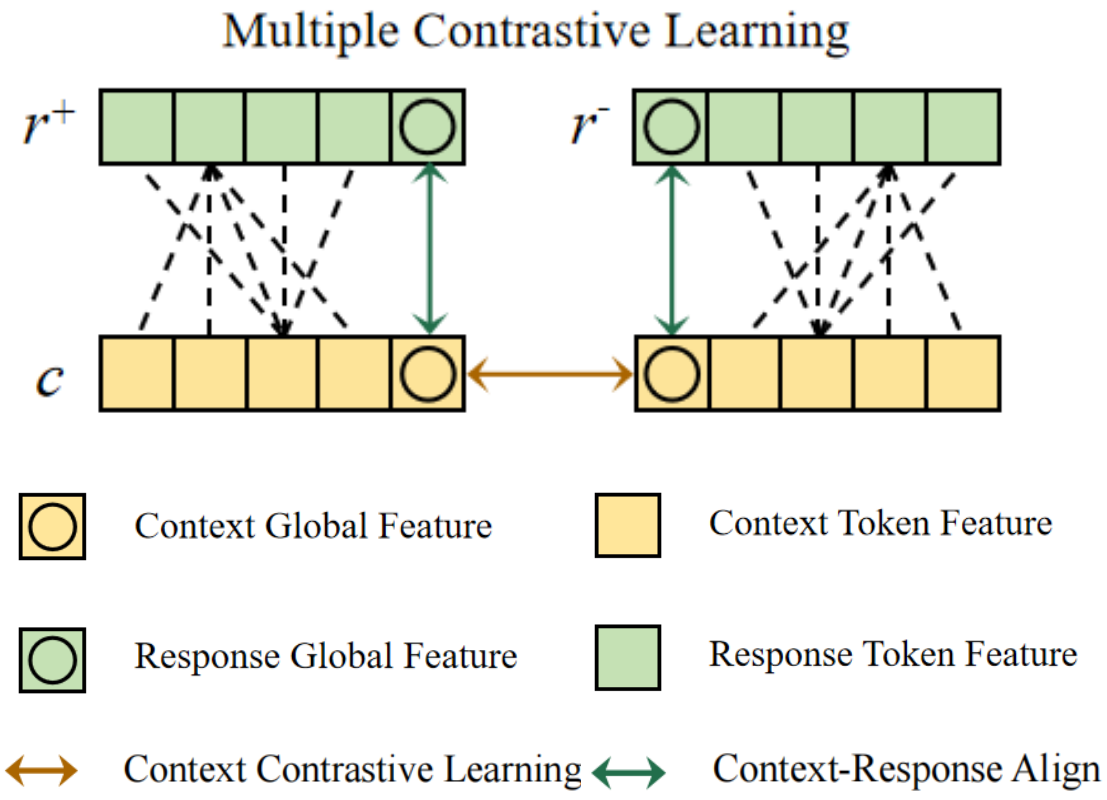
Multiple Contrastive Learning

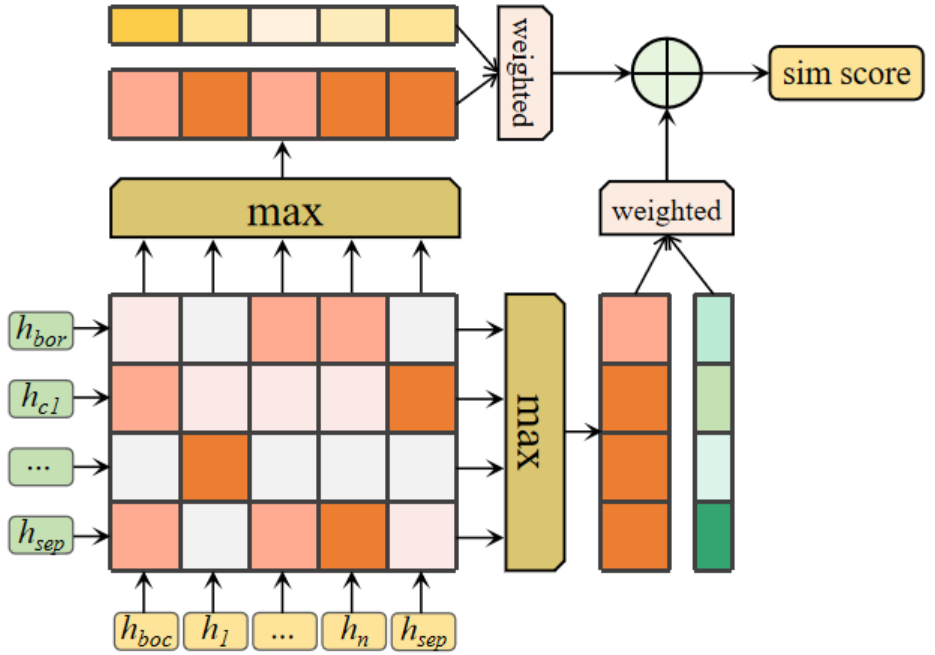


基于对比学习预训练回复检索模型

- Context-Response Representation
  - Low-layer BERT Bi-Encoder
  - Multiple Contrastive Learning
- Dialogue Interactive Reasoning
  - High-layer BERT Cross-Encoder
  - Hard Negative Resampling

Multiple Contrastive Learning

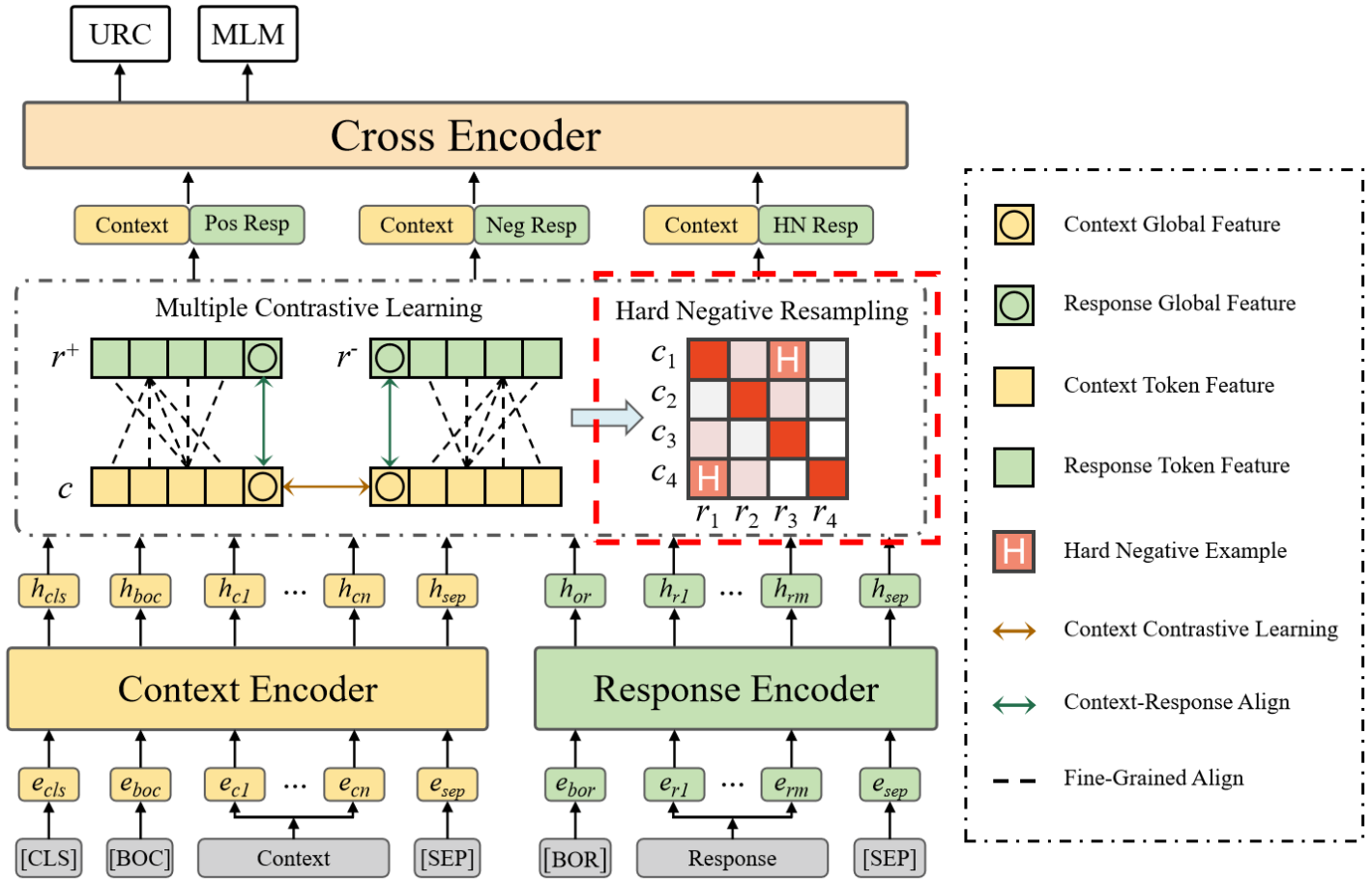




Fine-grained Alignment

02 / Method

Hard Negative Resampling

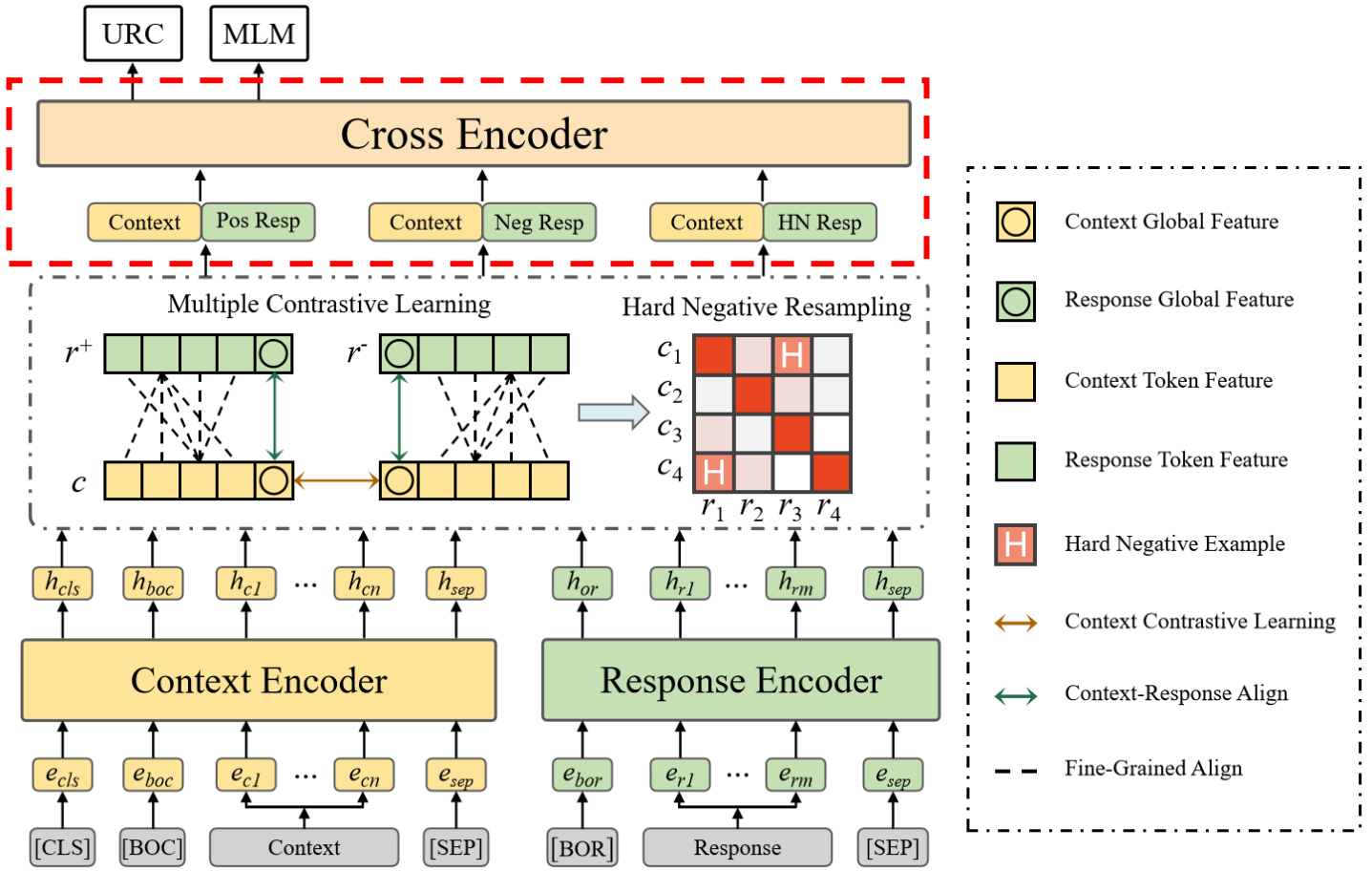


基于对比学习预训练回复检索模型

- Context-Response Representation
  - Low-layer BERT Bi-Encoder
  - Multiple Contrastive Learning
- Dialogue Interactive Reasoning
  - High-layer BERT Cross-Encoder
  - Hard Negative Resampling

02 / Method

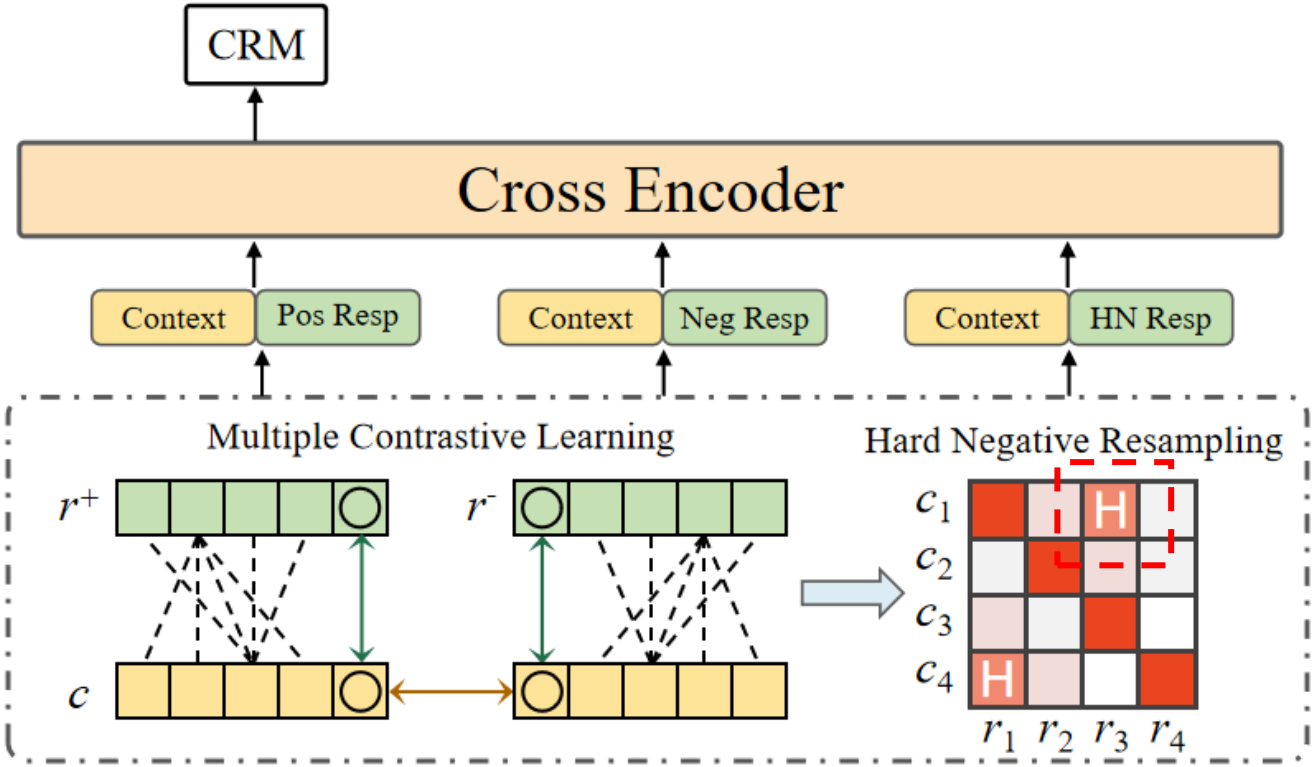
模型概述



基于对比学习预训练回复检索模型

- Context-Response Representation
  - Low-layer BERT Bi-Encoder
  - Multiple Contrastive Learning
- Dialogue Interactive Reasoning
  - High-layer BERT Cross-Encoder
  - Hard Negative Resampling

Total Loss



$$L = L_{MCL} + L_{CRM} + L_H$$

Content

03 ／ Experiment

Datasets

- Ubuntu
- Douban
- E-Commerce

Dataset		Train	Valid	Test
Ubuntu	#pairs	1M	500K	500K
	pos:neg	1:1	1:9	1:9
Douban	#pairs	1M	50K	6670
	pos:neg	1:1	1:1	1.2:8.8
E-Commerce	#pairs	1M	10K	10K
	pos:neg	1:1	1:9	1:9

Table 1: Corpus statistics of datasets

Baseline Models

- Single-turn matching models
  TF-IDF、RNN、CNN
- Multi-turn matching models
  SMN、DAM、MSN
- Pretrained models
  BERT、Poly-Encoder、UMS、
  BERT-FP、BERT-TAP、Uni-Encoder(previous SOTA)

03 / Experiment

Experimental Results

Models	Ubuntu			Douban						E-commerce		
	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	MAP	MRR	$P@1$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$
TF-IDF	0.410	0.545	0.708	0.331	0.359	0.180	0.096	0.172	0.405	0.159	0.256	0.477
RNN	0.403	0.547	0.819	0.390	0.422	0.208	0.118	0.223	0.589	0.325	0.463	0.775
CNN	0.549	0.684	0.896	0.417	0.440	0.226	0.121	0.252	0.647	0.328	0.515	0.792
SMN	0.726	0.847	0.961	0.529	0.569	0.397	0.233	0.396	0.724	0.453	0.654	0.886
DAM	0.767	0.874	0.969	0.550	0.601	0.427	0.254	0.410	0.757	0.526	0.727	0.933
MSN	0.800	0.899	0.978	0.587	0.632	0.470	0.295	0.452	0.788	0.606	0.770	0.937
BERT	0.808	0.897	0.975	0.591	0.633	0.454	0.280	0.470	0.828	0.610	0.814	0.973
PolyEncoder+FP*	0.884	0.950	0.991	0.617	0.664	0.498	0.316	0.492	0.844	0.914	0.965	0.995
UMS_{BERT+}*	0.875	0.942	0.988	0.625	0.664	0.499	0.318	0.482	0.858	0.762	0.905	0.986
BERT-FP*	0.911	0.962	0.994	0.644	0.680	0.512	0.324	0.542	0.870	0.870	0.956	0.993
BERT-TAP*	0.912	0.966	0.994	0.644	0.684	0.511	0.323	0.548	0.853	0.926	0.980	0.998
Uni-Encoder*†	0.916	0.965	0.994	0.648	0.688	0.518	0.327	0.557	0.865	-	-	-
BERT-BC(ours)	**0.924**	**0.968**	**0.995**	**0.665**	**0.701**	**0.538**	**0.356**	**0.565**	**0.870**	**0.957**	**0.981**	**0.998**

Table 2: Evaluation results on Ubuntu, Douban, and E-Commerce datasets. * denotes pre-train on corresponding dialogue corpus. † denotes previous state-of-the-art model.

Ablation Study

Method						Metric			
Base	CRA	FGA	CCL	HNR	DDP	R_{10}@1	R_{10}@2	R_{10}@5	MAP
√						0.641	0.824	0.970	0.777
√	√					0.826	0.934	0.989	0.898
√		√				0.837	0.935	0.985	0.903
√	√	√				0.846	0.945	0.990	0.910
√	√	√	√			0.855	0.942	0.993	0.914
√	√	√	√	√		0.905	0.960	0.988	0.943
√	√	√	√	√	√	**0.957**	**0.980**	**0.998**	**0.974**

CRA: Context-Response Alignment

FGA: Fine-Grained Alignment

CCL: Context Contrastive Learning

HNR: Hard Negative Resampling

DDP: Dialogue Domain Pretrain

04 / Further Analysis

Impact of Contrastive Learning at Different Layer

Dataset	Layer	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	MAP
E-Commerce	3-layer	0.752	0.891	0.983	0.85
	6-layer	0.836	0.936	0.993	0.903
	9-layer	**0.855**	**0.942**	**0.993**	**0.914**
	11-layer	0.822	0.931	0.986	0.849
Douban	3-layer	0.296	0.48	0.822	0.604
	6-layer	**0.299**	**0.49**	0.835	**0.610**
	9-layer	0.283	0.484	**0.840**	0.601
	11-layer	0.258	0.429	0.781	0.560

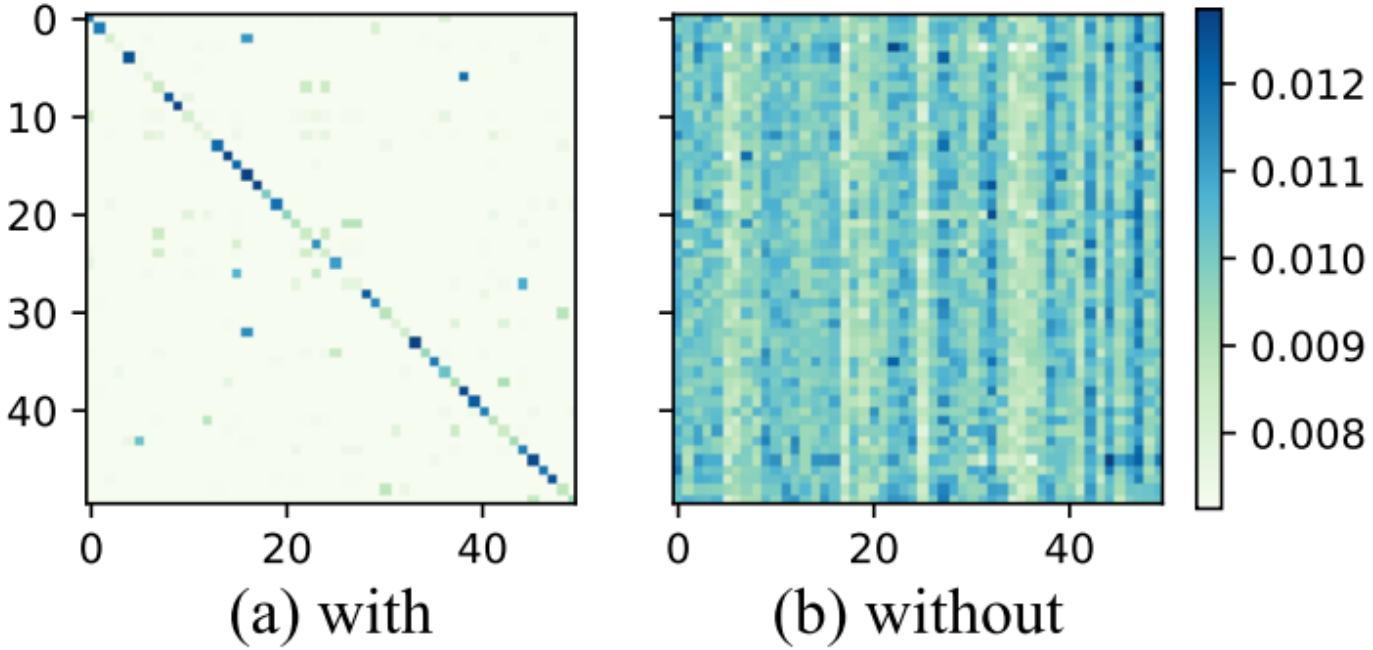
04 / Further Analysis

Effectiveness of HNR Strategy

Method	R_{10}@1	R_{10}@2	R_{10}@5	MAP
Base	0.641	0.824	0.970	0.777
Base + MCL	0.855	0.942	**0.993**	0.914
Base + MCL + HNR	**0.905**	**0.960**	0.988	**0.943**
Base + MCL + Random	0.869	0.951	0.989	0.921
Base + MCL + CUR	0.892	0.952	0.991	0.935
HCL(ACL21)	0.721	0.896	**0.993**	-

04 / Further Analysis

Visualization of Alignment Matrix



04 / Further Analysis

Case Study

Context	**Customer:** Hello, I haven't received my deliver parcels, any news yet? **Service Staff:** What province are you from? **Customer:** I am in Xinjiang, It's been two days and I still haven't received the parcels. **Service Staff:** Today I'll help you to urge the courier. **Customer:** Let me know when the time comes, I'm waiting to use it! **Service Staff:** Well, I got it. **Customer:** Otherwise, I will deal with this matter as not received the goods.
BERT	Well, we will send you S.F. Express.
MCL	Usually the goods will be delivered to Xinjiang within three days.
Ours	Sorry, I'll help you to urge the courier company.
Ground Truth	Sorry, I'll help you to urge the courier company.

Table 8 Case Study

05 / Conclusion

Conclusion

- In this paper, we propose a response selection model BERT-BC that combining Bi-Encoder and Cross-Encoder with three contrastive learning mechanisms and a hard-negative resampling strategy.
- The experimental results demonstrate the superiority of our proposed BERT-BC model in the response selection task.

Future Work

- In future work, we consider introducing common-sense knowledge in response selection.

THANKS FOR WATCHING