Learning to Select Context in a Hierarchical and Global Perspective for Open-domain Dialogue Generation

ID: 2518

Lei Shen[1,2,*], Haolan Zhan[2,*], Xin Shen[3], Yang Feng[1,2],
[1]Institute of Computing Technology, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences, [3]Australian National University

Introduction

Background

Open-domain multi-turn conversations mainly have three features, which are hierarchical semantic structure, redundant information, and long-term dependency. Grounded on these, selecting relevant context becomes a challenge step for multiturn dialogue generation. However, existing methods cannot differentiate both useful words and utterances in long distances from a response. Besides, previous work just performs context selection based on a state in the decoder, which lacks a global guidance and could lead some focuses on irrelevant or unnecessary information. In this paper, we propose a novel model with hierarchical self-attention mechanism and distant supervision to not only detect relevant words and utterances in short and long distances, but also discern related information globally when decoding. Experimental results on two public datasets of both automatic and human evaluations show that our model significantly outperforms other baselines in terms of fluency, coherence, and informativeness.

	Example 1
Utterance 1	你好，送的卡物流怎么查呢？ (Hello, how to query the logistics information of gift cards?)
Utterance 2	体验卡对吗？ (The trial cards, right?)
Utterance 3	对。 (Yes.)
Response	您可以关注下微信公众号“卡助手”，公众号内可以查询哦。 (Please follow the official account “Card Assistant” in WeChat, and then you can query it inside.)
	Example 2
Utterance 1	有人吗? (Hello, anybody here?)
Utterance 2	亲，有什么问题我可以帮您处理或解决呢? (Dear, what can I do for you?)
Utterance 3	我这个订单有保险吗，帮我查一下。 (Is my order insured? Please check it out for me.)
Response	购买的时候没有就没有哦。 (No insurance if it is not attached when you buy.)

Table 1: A motivation case.

Motivation & Key Idea

Motivation

To enpower dialogue generation model not only to detect relevant words and utterances in short and long distances, but also to discern related information globally when decoding.

Key ideas

- We propose **HiSA-GDS**, a modified Transformer model with Hierarchical Self-Attention and Globally Distant Supervision.
- Experimental results on two public datasets along with further discussions show that HiSA-GDS significantly outperforms other baselines and is capable to generate more fluent, coherent, and informative responses.

Method

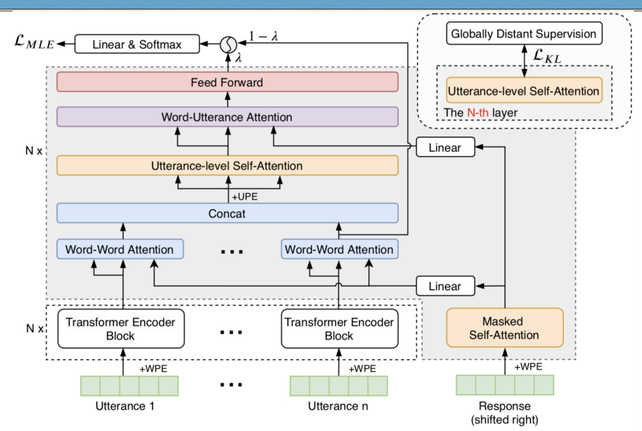

Figure 1: Architecture of our proposed model - HiSA-GDS.

Experiments - Case Study

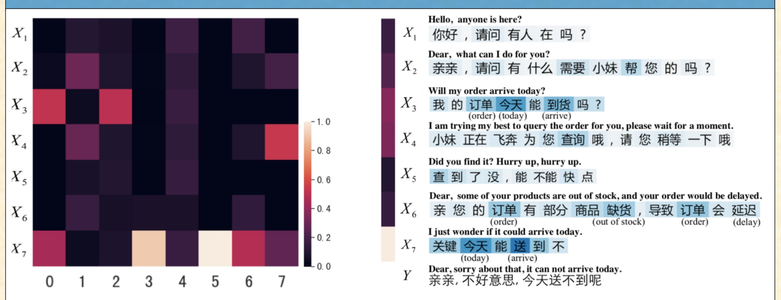

Figure 2: Left: Utterance-level multi-head attention visualization of HiSA-GDS in the word-utterance attention layer. 0 to 7 are the index of each head. Right: Word-level attention visualization in the word-word attention layer.

Experiments - Evaluation

Comparison

Model	Ubuntu						JDDC					
	B-2	D-2	Avg	Ext	Gre	Coh	B-2	D-2	Avg	Ext	Gre	Coh
S2SA [16]	0.896	6.104	46.323	28.851	39.209	48.117	4.233	3.609	53.901	36.493	37.578	46.176
HRED [1]	3.853	6.661	57.972	34.007	41.462	63.173	9.405	11.762	63.191	46.714	43.295	57.183
VHRED [17]	3.677	8.098	57.251	32.024	41.808	61.464	6.367	15.184	62.436	43.337	41.787	63.924
Static [4]	1.581	3.586	51.055	36.193	**53.983**	69.748	2.285	3.738	60.820	38.047	35.367	65.938
HRAN [18]	3.880	7.402	56.763	33.501	41.584	67.635	5.962	**16.365**	63.064	43.439	42.389	62.391
Transformer [10]	3.697	7.278	53.463	36.353	42.763	69.970	5.389	5.185	68.336	48.284	41.103	67.485
ReCoSa [6]	3.872	9.406	59.368	35.834	41.835	71.922	5.962	6.594	61.085	41.473	42.942	71.374
HiSA	4.021	9.598	63.527	36.208	40.598	72.261	6.986	14.804	66.103	43.715	45.081	73.286
HiSA-GDS	**7.351**	**10.934**	**68.283**	**41.468**	50.382	**75.823**	**7.127**	15.823	**73.952**	**52.502**	**49.477**	**74.281**

Table 2: Automatic evaluation results on Ubuntu and JDDC (%). The metrics BLEU-2, Distinct-2, Average, Extrema, Greedy and Coherence are abbreviated as B-2, D-2, Avg, Ext, Gre, and Coh, respectively.

Dataset	Model	HiSA-GDS vs.			kappa
		Win	Loss	Tie	
Ubuntu	S2SA [16]	58%	12%	30%	0.468
	HRED [1]	46%	19%	35%	0.531
	VHRED [17]	48%	20%	32%	0.493
	Static [4]	51%	17%	32%	0.596
	HRAN [18]	42%	9%	49%	0.424
	Transformer [10]	44%	19%	37%	0.474
	ReCoSa [6]	40%	6%	54%	0.528
JDDC	S2SA [16]	53%	24%	23%	0.547
	HRED [1]	56%	16%	34%	0.468
	VHRED [17]	52%	19%	29%	0.453
	Static [4]	48%	11%	41%	0.518
	HRAN [18]	50%	22%	28%	0.495
	Transformer [10]	51%	29%	20%	0.447
	ReCoSa [6]	45%	27%	28%	0.461

Table 3: Human evaluation between HiSA-GDS and other baselines.

Conclusion

In this paper, we propose a novel model for open-domain dialogue generation, HiSA-GDS, which conducts context selection in a hierarchical and global perspective. The hierarchical self-attention is introduced to capture relevant context at both word and utterance levels. We also design a globally distant supervision module to guide the response generation at decoding. Experiments show that HiSA-GDS can generate more fluent, coherent, and informative responses.