

<Group Linguistic Bias Aware Neural Response Generation>

Jianan W., Xin W., Fang. L., Zhen X., Zhuoran W., Baoxun W.

SIGHAN 2018

Motivation:

Customizing diverse responses for same post based on different speaker attributes (e.g. gender).

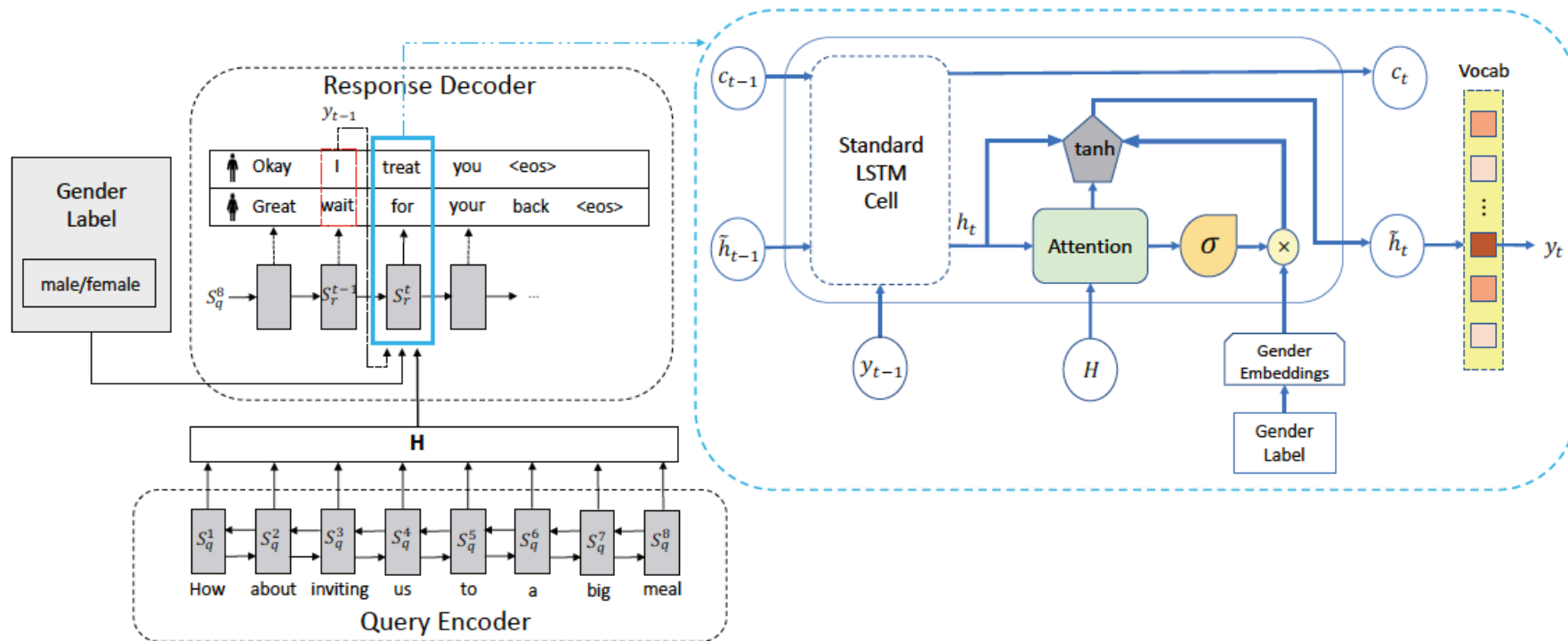
Contributions:

- Introducing linguistic biases of human group into response generator
- Proposing a novel neural component to dynamically introduce linguistic group bias through generation process

Model Architecture

Encoder-Decoder Structure:

1. Encoder: Standard BILSTM
2. Decoder: Unidirectional LSTM



Decoding Computation:

$$\tilde{h}_t = f(h_t, H, e^g) \quad \begin{array}{l} h_t \text{ is intermediate decoder hidden state, } H \text{ is from encoder and } e^g \\ \text{is gender embedding} \end{array}$$

$$f(h_t, H, e^g) = W_f[h_t, a_t, e^g \odot g_t] + b_f$$

$$g_t = \sigma(W_g a_t + b_g)$$

$$a_t = \sum_{j=1}^T \alpha_{tj} h_j$$

$$\alpha_{tj} = \frac{\exp(e_{tj})}{\sum_{k=1}^{T_x} \exp(e_{tk})}$$

$$e_{tj} = W_a[h_t, h_j] + b_a$$

Here h_t and h_j should be hidden states from encoder (sadly, not specified in original paper)

Dataset and Evaluation

Dataset: Data is from one of Chinese real-name social network sites (Not Released)

$4M$ (query, response) pairs without gender label, and $1M$ (query, response) pairs with gender label.

Evaluation: Human annotation (score from 0 to 1)

Metric: Attractive (+2), Neural (+1), Unsuitable (0)

Model Variations

1. S2S: Standard sequence-to-sequence model without attention
2. GLBA-Static: Proposed full model without gating on gender embedding
3. GLBA-Dyna: Proposed full model

Experiment Results

1. Average Human Evaluation Score

Method	Average Score
S2S	0.923
GLBA-Static	0.944
GLBA-Dyna	1.404

Method	Score		
	0	+1	+2
S2S	14.56%	78.56%	6.88%
GLBA-Static	17.56%	70.50%	11.94%
GLBA-Dyna	8.22%	43.11%	48.67%

2. Gender Consistency: Infer gender from generated response

Method	Accuracy
S2S	-
GLBA-Static	0.340
GLBA-Dyna	0.493

Case Study:

Query	Model		Reponse(Original)	Response(Translated)
我这早零下了。 It has been below zero degrees for long.	S2S		我这零下十几度。	It is around minus 10 degrees here.
	GLBA-Static	Female	可怜的孩纸。	Poor child...
		Male	这么冷。	That is so cold.
	GLBA-Dyna	Female	这么冷啊， 我们还没开始穿羽绒服呢。	Wow that is so cold. We do not even need to wear down jackets here.
		Male	这么冷，注意保暖。	That is so cold. Keep warm.

Query	Gender	Response
我这早零下了。	Female	这么冷啊我们还没开始穿羽绒服呢
	Male	这么冷注意保暖
OKOK, 俺还想晚上麻将。	Female	好滴好滴等你回来我们一起去玩
	Male	来吧热烈欢迎哥招待你
生病好难受啊, 健康的时候一定要快快乐乐的, 好好珍惜。	Female	嗯嗯知道了亲爱的
	Male	恩谢谢兄弟关心啊

Conclusion:

1. The model structure is rather straightforward
2. Introducing group linguistic bias into response generation process
3. Demonstrating response customization is reflected in key words, and proving that incorporating gender information locally is more effective than using it globally