国际人工智能会议
AAAI 2021 论文北京预讲会

Stylized Dialogue Response Generation Using Stylized Unpaired Texts

Equal Contribution

Yinhe Zheng, Zikai Chen, Rongsheng Zhang, Shilei Huang, Xiaoxi Mao, Minlie Huang

Tsinghua University

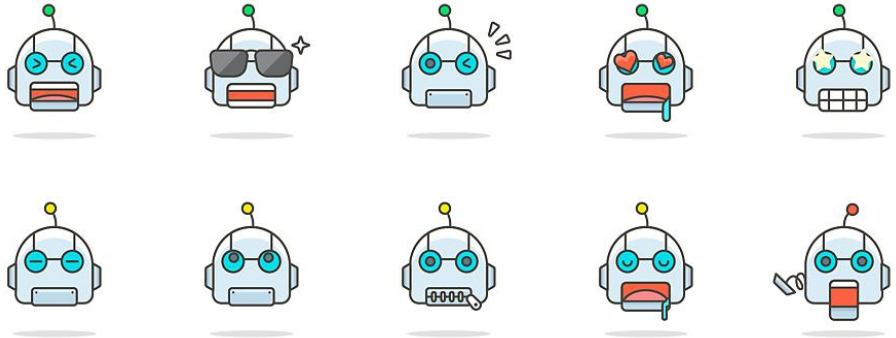
Samsung Research

Fuxi AI Lab, NetEase Inc.

Background

- Text style is an interesting phenomenon to model

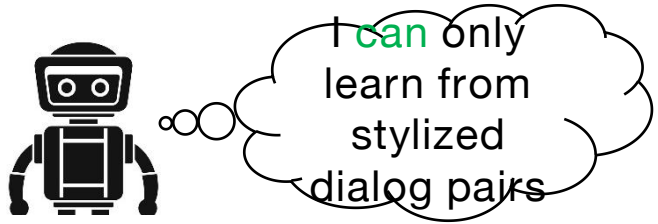
- Stylized dialogue systems are attractive to users



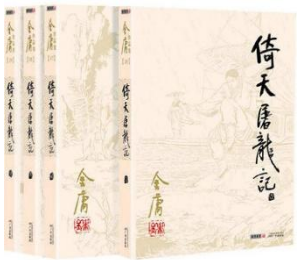
Motivation

- Most existing stylized dialog models need to train with stylized dialog pairs

<Post, Stylized Response>

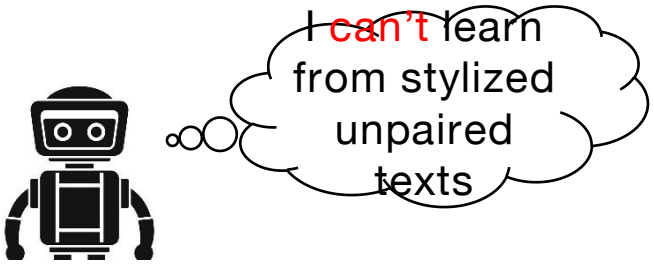


- However, most textual features are embedded in unpaired texts, (e.g. Novels)



“…忽见东首火把照耀，有七八人手执兵器，快步奔来。张杨二人忙在草丛中躲起。那干人奔到邻近…”

Stylized Unpaired Texts



Task Setting

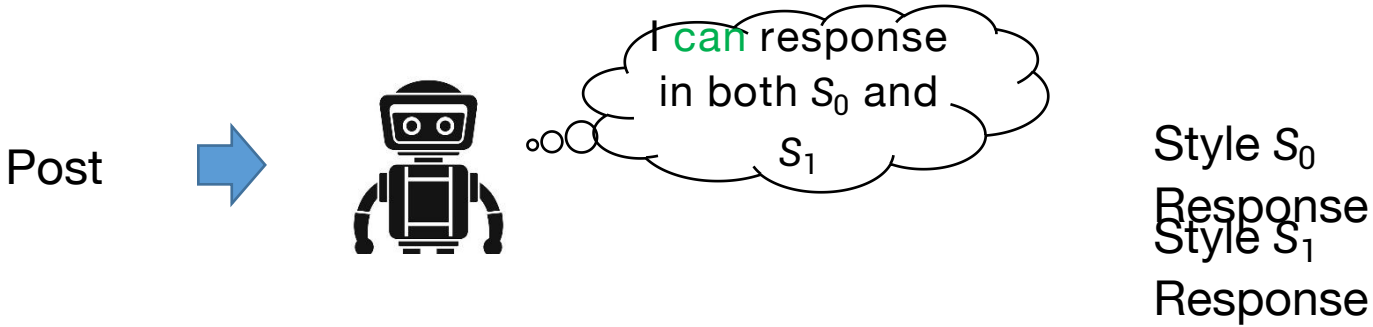
- Input

  Dialogue pairs in style S_0: $\mathcal{D}_p = \{x_1, y_1|, x_2, y_2|, \ldots x_N, y_N|\}$

  Unpaired texts in style S_1: $\mathcal{D}_t = \{t_1, t_2, \ldots, t_M\}$

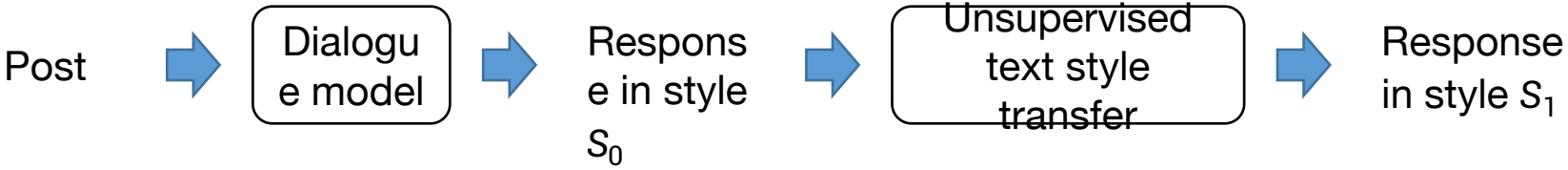
- Resulting model

  A dialogue mode that can produce both S_0 and S_1Responses

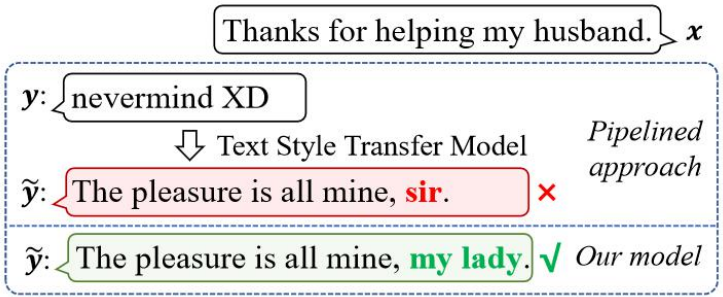


More on Task Setting

- Note that: This task can be tackled using existing unsupervised text style transfer model



- But this may lead to in-consistent responses

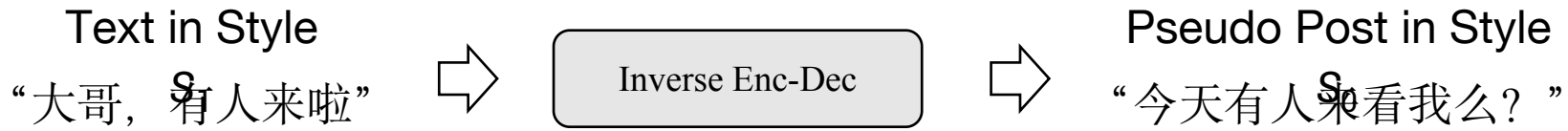


Contribution

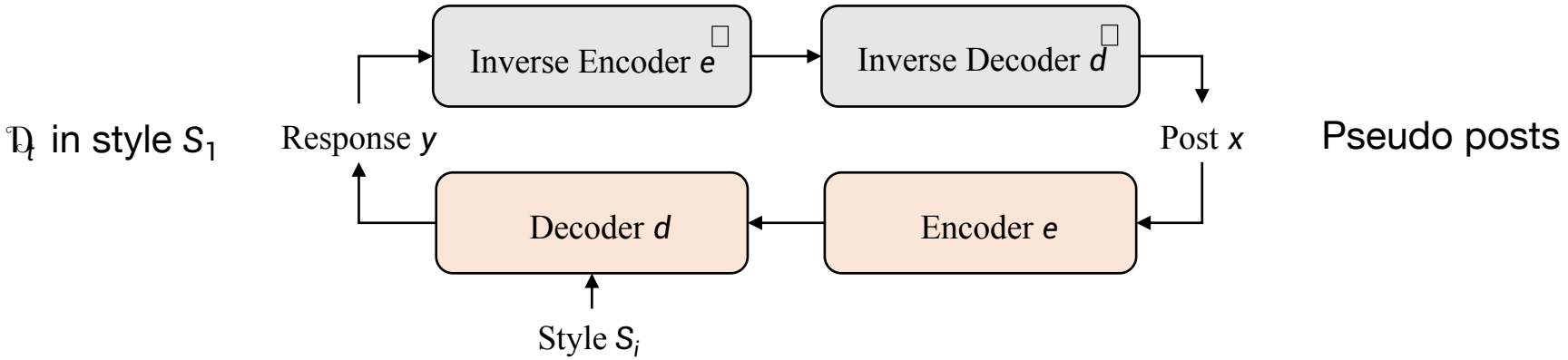
- A novel method is proposed to build a stylized dialogue model that can capture stylistic features embedded in unpaired texts. Specifically:
  - An inverse dialogue model is introduced to generate stylized pseudo dialogue pairs, which are further utilized in a joint training process.
  - An effective style routing approach is devised to intensify the stylistic features in the decoder.
- Automatic and human evaluations on two datasets show that our method outperforms competitive baselines with a large margin in producing stylized and coherent dialogue responses.

Inverse dialogue model

- Takes in a response, produce a post



- Use produced pseudo dialogue pairs to train the stylized dialogue model

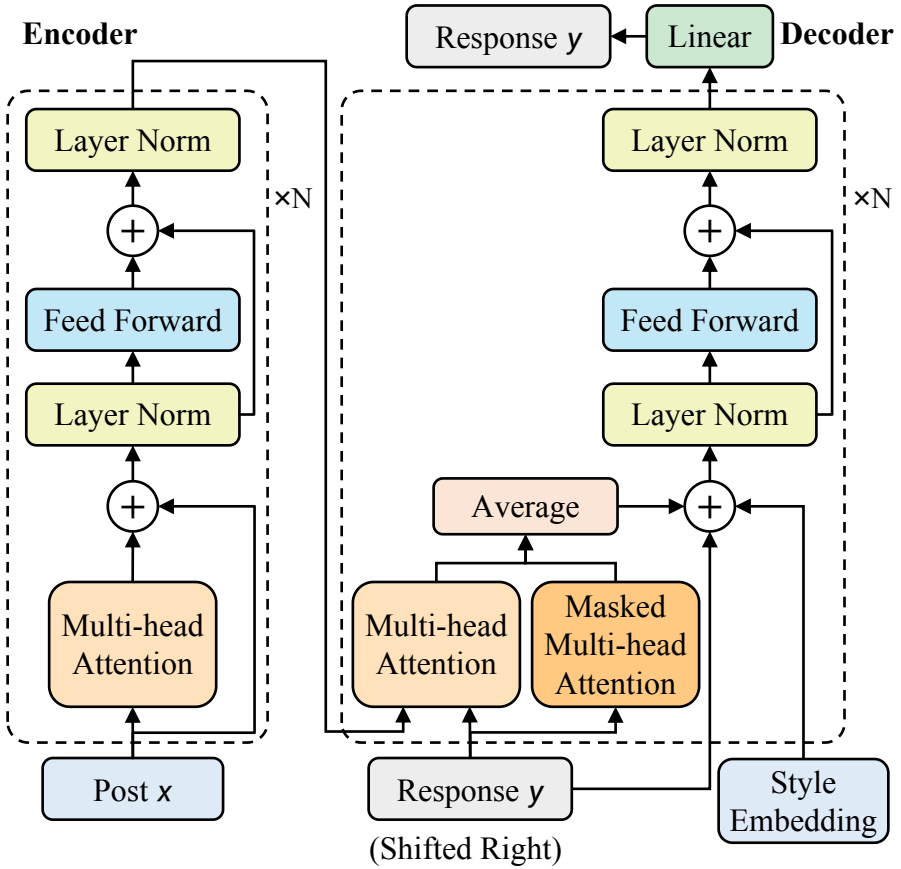


Style routing approach

- Add style embedding at the end of each attention block

quer y key valu e

$$\boldsymbol{R}_{prev} = \mathrm{MMHA}[\boldsymbol{e_w}(y_p), \boldsymbol{e_w}(y_p), \boldsymbol{e_w}(y_p)],$$
$$\boldsymbol{R}_{post} = \mathrm{MHA}[\boldsymbol{e_w}(y_p), \boldsymbol{e}(x), \boldsymbol{e}(x)],$$
$$\boldsymbol{R}_{avg} = (\boldsymbol{R}_{prev} + \boldsymbol{R}_{post})/2.$$
$$\boldsymbol{R}_{merge} = \boldsymbol{R}_{avg} + \boldsymbol{e_s}(S_i).$$



Joint Training

- Train the inverse dialogue model and stylized dialogue model interactively

Post to response loss

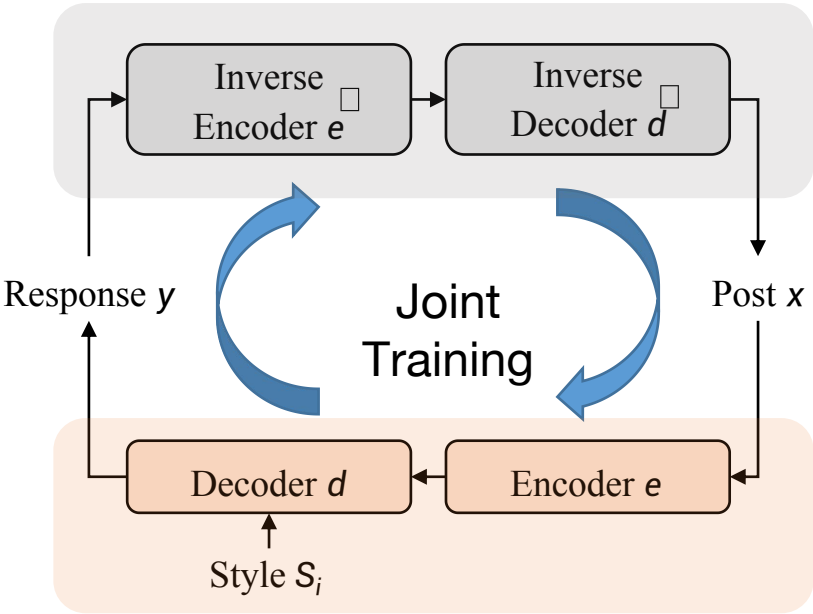
$$\mathcal{L}_{p2r} = \mathop{\mathbb{E}}_{\langle x,y\rangle\sim\mathcal{D}_p} -\log p_d(y|\boldsymbol{e}(x), S_0),$$

Response to post loss

$$\mathcal{L}_{r2p} = \mathop{\mathbb{E}}_{\langle x,y\rangle\sim\mathcal{D}_p} -\log p_{\hat{d}}(x|\hat{\boldsymbol{e}}(y)).$$

Inverse dialogue loss

$$\mathcal{L}_{inv} = \mathop{\mathbb{E}}_{\substack{t\sim\mathcal{D}_s,\\ x'\sim p_{\hat{d}}(x|\hat{\boldsymbol{e}}(t))}} -\log p_d(t|\boldsymbol{e}(x'), S_1),$$



Experiments

- Chinese data: Weibo dialogue (S_0), Jinyong novel (S_1)
- English data: Reddits informal dialogue (S_0), Formal English writing (S_1)

Response in Style S_1

Model	WDJN Dataset								TCFC Dataset							
	BLEU-1,2		Dist.	BERT	SVM	Flu.	Coh.	Style	HAvg.	BLEU-1,2		Dist.	BERT	SVM	Flu.	Coh.	Style	HAvg.
SLM	2.90	0.37	26.6	26.7	40.7	1.96*	1.52	0.37	0.79	12.6	0.99	42.5	85.6	87.2	1.90*	0.89	1.78	1.36
SRL	2.53	0.33	40.4	36.2	43.2	1.83	1.52	0.39	0.82	7.83	0.70	42.7*	47.6	53.5	1.76	0.72	1.25	1.09
SFusion	3.84	0.20	33.1	8.24	19.8	1.63	0.69	0.40	0.67	5.51	0.28	**61.0**	21.9	39.0	1.47	0.56	1.17	0.90
S2S+BT	6.22	0.68	30.7	66.0	83.6	1.89	1.53*	0.63	1.09	12.1	1.25	42.0	86.3	86.8	1.58	0.72	1.66	1.14
S2S+CT	11.3	0.62	32.4	72.3	76.4	0.45	0.19	**1.50**	0.38	8.05	0.64	60.9	67.7	67.8	0.37	0.12	0.64	0.24
S2S+PTO	3.57	0.44	32.9	35.1	43.3	1.82	1.54*	0.35	0.75	9.55	0.84	34.5	28.6	50.3	0.35	0.26	0.39	0.32
Ours	**13.6**	**1.53**	**42.8**	**78.3**	**89.3**	**1.96**	**1.60**	1.16	**1.48**	**15.1**	**1.71**	43.4	**97.3**	**96.1**	**1.90**	**1.01**	**1.89**	**1.46**
Human	N/A		49.3	80.1	85.4	1.93	1.60	1.53	1.67	N/A		62.7	89.6	85.8	1.91	1.18	1.83	1.56

Response in Style S_0

Model	WDJN Dataset								TCFC Dataset							
	BLEU-1,2		Dist.	BERT	SVM	Flu.	Coh.	Style	HAvg.	BLEU-1,2		Dist.	BERT	SVM	Flu.	Coh.	Style	HAvg.
S2S	8.50	2.42	35.1	97.0	93.0	1.96*	1.73	1.86	1.85*	6.92*	0.61*	54.8	70.1*	60.9	1.82*	1.16*	1.68*	1.50*
SFusion	8.65	0.82	35.3	**99.9**	**99.2**	1.41	0.74	1.92*	1.16	4.61	0.22	**62.8**	**70.3**	**61.1**	1.57	0.76	**1.77***	1.19
Ours	**11.6**	**2.93**	**39.0**	93.5	89.2	**1.97**	**1.85**	**1.93**	**1.92**	**6.96**	**0.67**	49.4	69.4	59.2	**1.85**	**1.16**	1.70	**1.51**
Human	N/A		56.4	97.9	94.4	1.89	1.86	1.98	1.91	N/A		72.6	72.0	72.1	1.76	1.19	1.76	1.52

Conclusion

- An inverse dialogue model is introduced in our method to produce stylized pseudo dialogue pairs
- Automatic and manual evaluation shows that our method outperforms competitive base-lines in producing coherent and style-intensive responses.

Codes and data are coming soon:
https://github.com/silverriver/Stylized_Dialog

国际人工智能会议
AAAI 2021 论文北京预讲会

THANKS

2020.12.19