<Polite Dialogue Generation Without Parallel Data>

Tong Niu, Mohit Bansal

UNC Chapel Hill

TACL 2018

Outline:

1. Introduction
2. Methodologies
3. Experiments
4. Conclusion

Introduction:

1. Generating stylistic language is crucial to developing engaging and convincing conversational agents.
2. The main purpose is to generate semantically-fluent and contextually-relevant response based on conversation while maintaining desirable style.
3. The main challenge is lack of parallel dataset (regular-to-stylistic pairs).

Methodology:

1. To conquer the problem of lacking parallel dataset, the authors propose three weakly supervised approaches.
2. The main pipeline consists of two parts: (1) style-classifer and sequence to sequence model.
3. The style-classifer is used to influnce and encourage stylistic dialogue generation while seq2seq paradigm takes care of fluency and context relevance. Jürgen

Model Overview Outline

1. Style-Classifier
2. Fusion Model
3. Label-fine-tuning (LFT) Model
4. Polite-RL Model

Politeness Classification Model

The classification model is composed of LSTM and CNN and the corresponding structure is shown in the right side.

The BiLSTM encoder first extracts compute hidden states $h_{1:n}$ of input sequence $x_{1:n}$. Then a CNN with T kernels are used to compute feature maps of output hidden states. After applying maxpooling, T feature maps are transformed to a vector z with fixed size T. Finally, a softmax classifier is used to determine whether the input sequence is polite or rude.

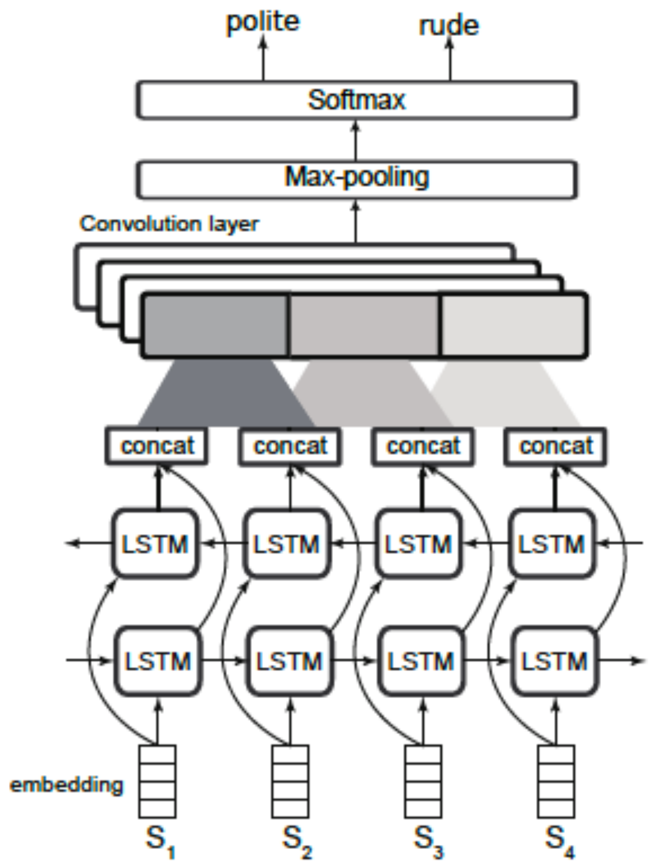

Figure 1: Our LSTM-CNN politeness classifier.

Fusion Model

1. Train a language model solely on polite utterance dataset
2. Train a conventional seq2seq model
3. During inference time, use pretraiend language model to rerank output of seq2seq model.

$$p_t = \alpha\, p_t^{\mathrm{S2S}} + (1-\alpha)\, p_t^{\mathrm{LM}}$$

Weakness:
1. The pretrained language model is not aware of conversation context, therefore the generated response maybe irrelevant.
2. The model does not learn to be polite during training phase, but is forced to be polite during decoding phase.

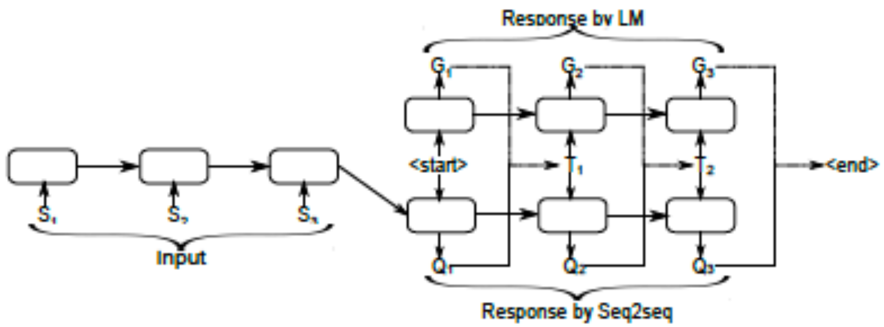

Figure 2: Fusion model: the output probability distributions of the decoder and the polite-LM are linearly mixed to generate the final decoded outputs.

Label-Fine-Tuning Model

1. Append a polite label vector at the beginning of source sequence. The vector is trainable.
2. To control the politeness level, the vector is scaled by the politeness score of target response.
3. During training, the score is got by feeding ground-truth response through the pretrained classifier.
4. During inference, the the score is manually provided ranging from 0.0-1.0 (0.0-0.5: rule, 0.5: neutral, 0.5-1.0: polite).

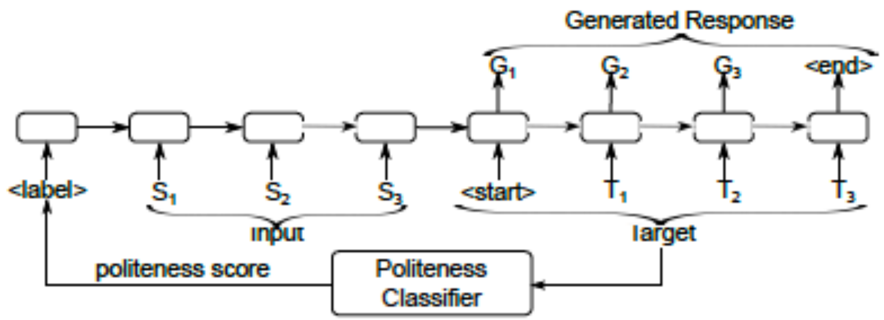

Figure 3: Label-Fine-Tuning model: during training, the embedding of the prepended label is scaled by the style classifier's continuous score on the ground-truth (target) sequence. During testing, we scale the embedding of the label by the desired (continuous) politeness score of the generated response.

Polite Reinforcement Learning Model

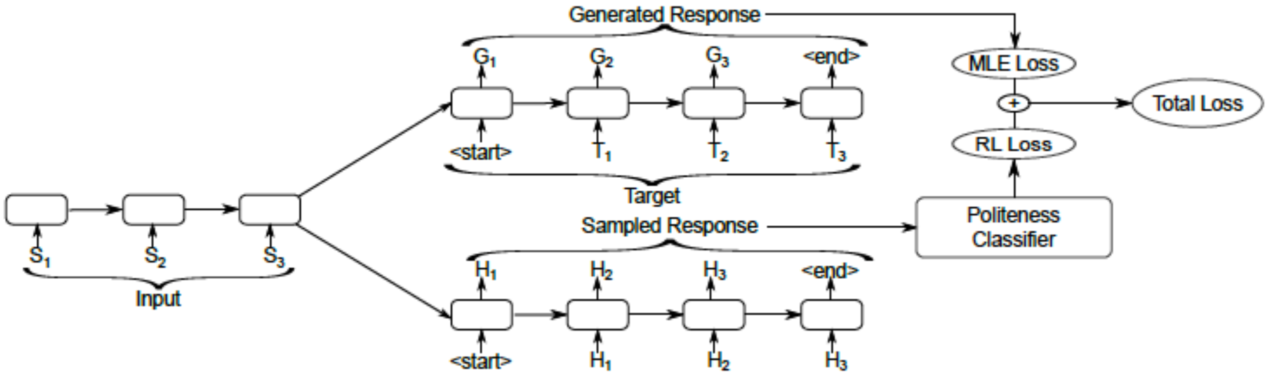

Figure 4: Polite-RL model: upper-right shows max-likelihood (ML) training with generated and ground-truth target sequences; lower-right shows RL training with a randomly sampled response generated by the model and the reward it generates after getting fed into the style classifier. Note that the attention mechanism is not shown here for clarity.

The training objective consists of two terms, first is conventional MLE and second is RL loss.

$$L = L_{\mathbf{ML}} + \beta L_{\mathbf{RL}} \qquad L_{\text{ML}} = -\sum_{t=1}^{n} \log p(y_t^* | y_1^*, ..., y_{t-1}^*, x). \qquad L_{\text{RL}} = -(R - R_b) \sum_{t=1}^{n} \log p(y_t^s | y_1^s, ..., y_{t-1}^s, x)$$

y^* is ground-truth target and y^s is sampled sequence during training phase.

Datasets:

1. To pretrain politeness classifier: Stanford Politeness Corpous
2. To train dialogue model: MovieTriples Dataset (each instance has the form of X-Y-X), where X and Y are two characters. The task is to generate the last response.

Evaluation:

1. Human: Evaluate politeness level on five point scale
2. Automatic: BLEU is used as a complementary evaluation for dialogue quality

Model Variations:

1. Retrival: First get most similar responses X_2 to the context (X_1, Y), and rerank the results based on their politeness using classifier.
2. Generic-10: First choose top 10 most polite and generic responses from Stanford Politeness Corpus. Then rerank the response for given context based on TF-IDF similarity.
3. Seq2seq: Standard RNN-based generation model with attention.
4. Fusion
5. LFT: Lable-Fine-Tuning
6. Polite-RL

Experiment Results:
The test contains 300 randomly selected contexts.
Quality: Context-relevance

	Politeness	Quality	Difference
Retrieval	3.57	3.15	0.42
Generic-10	**3.66**	2.99	0.67
Seq2seq	3.11	3.42	0.31
Fusion	3.23	3.05	0.18
LFT	3.63	3.39	0.24
Polite-RL	3.50	**3.54**	**0.04**

Table 3: MTurk human evaluation results on politeness level and dialogue quality (as well as the absolute value difference between the two, to show balance) of the Retrieval Models, Seq2seq and the three proposed generative models (avg. of two annotators is shown here). Top results are boldfaced.

Evaluation of 3-way LFT Model

1. Continuous-LFT: Generating polite, neutral and rude responses using scaling score of 1.0, 0.5 and 0.0

	Continuous-LFT
Polite	3.70
Neutral	3.15
Rude	1.19

Output Examples

Input	X: you 're sweet to say so . Y: pretty song .
Retrieval	X: is there a song you do n't know ?
Generic-10	X: can you clarify ?
S2S	X: ⟨person⟩ , i know .
Fusion	X: you 're a good man .
LFT	X: i love this song .
RL	X: you sound like a goddess .
Input	X: i find it hard to believe myself . Y: his lips , unused as they are to such a contortion , twisted and stretched . into a smile .
Retrieval	X: thank you . and always smile .
Generic-10	X: nice .
S2S	X: how can you be so sure ?
Fusion	X: you 're a good man , mr .
LFT	X: your lips are well .
RL	X: your lips are so beautiful .

Visualization of Polite-RL Reward

Heatmap of each sampled token, when computing the RL-reward. The heatmap reflects the absolute magnitude of derivative on each dimension.

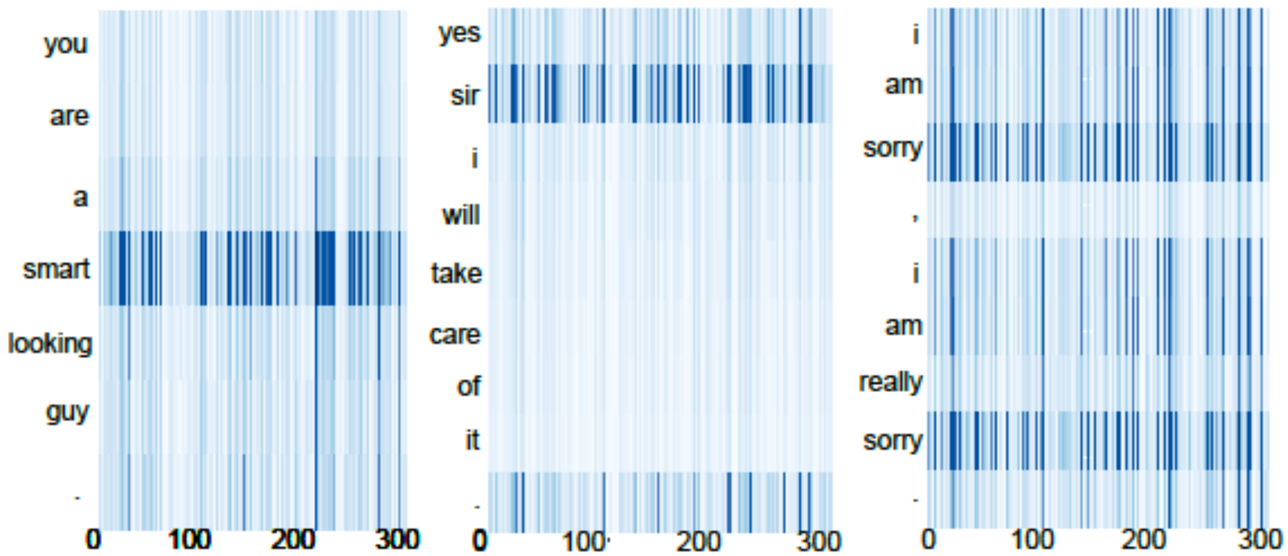

Figure 5: Saliency heatmaps of the classifier's attention (reward for sampled responses in Polite-RL model).

Contributions of this paper

1. Design three simple yet effective approaches for learning stylistic dialogue generation via unparalleled data.
2. Present comprehensive experiment results to justify their claims.

Future Work

1. Their approaches do not take care of the problem of response diversity. One possible direction is using skeleton-then-generation framework (CD, NAACL 2019).
2. Extend their approach to generate responses with more than two attributes (e.g. male, female and neutral).