PLATO: Pre-trained Dialogue Generation Model with Discrete Latent Variable

Siqi Bao[*], **Huang He**[*], **Fan Wang, Hua Wu** and **Haifeng Wang**
Baidu Inc., China
{baosiqi, hehuang, wangfan04, wu_hua, wanghaifeng}@baidu.com

Motivation

- Pre-training models have been proved effective for a wide range of natural language processing tasks
- directly fine-tuning BERT on small conversation datasets has some problems:

➢the underlying pattern of dialogue generation is different from general text

➢training mode of unidirectional dialogue generation is distinct from bidirectional model

➢ there exists a one-to-many relationship in dialogue generation

framework

- large-scale Reddit and Twitter conversations are utilized to further pre-train the generation model
- to mitigate the difference in training mode, a flexible paradigm integrating unidirectional and bidirectional processing is employed in this work,
- a discrete latent variable is introduced to model the one-to-many relationship among utterances in conversations.

Framework

There are three elements : dialogue response r, latent variable K and context c. Z is A K-way categorical variable , $z \in [1, K]$, each value corresponding to a particular conversational intent.
. Given a pair of context and response, the underlying latent speech act can be estimated as $p(z|r,c)$

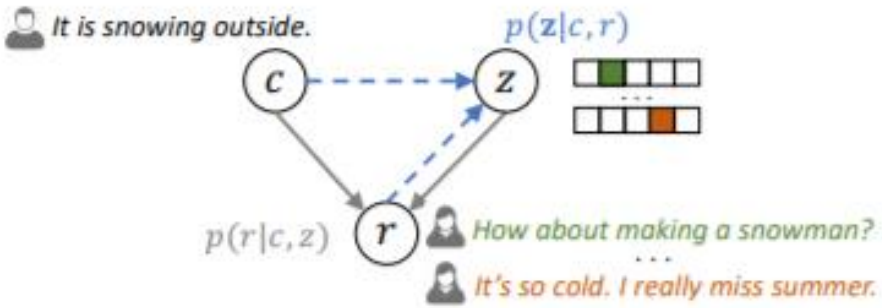


Figure 1: Graphical illustration of response generation (gray lines) and latent act recognition (dashed blue lines).

Framework

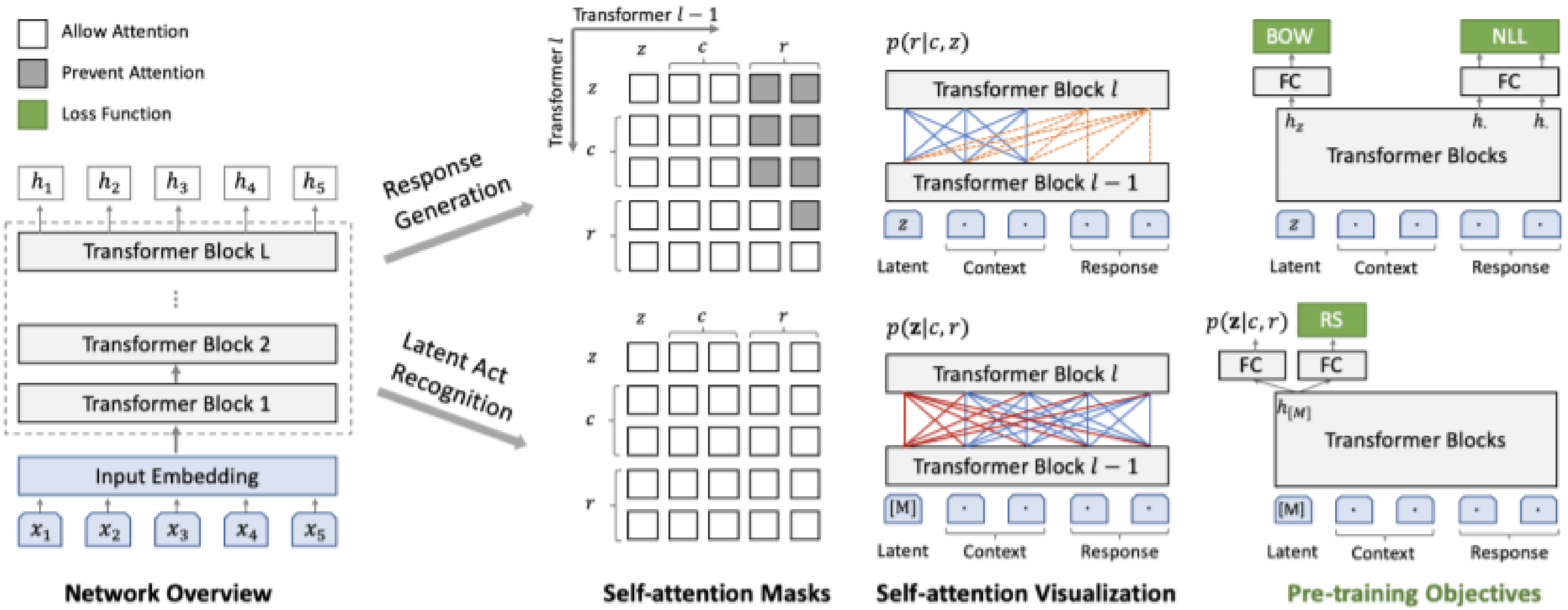


Figure 2: Architecture of dialogue generation with discrete latent variable. In self-attention visualization, red and blue lines denote bi-directional attention, and dashed orange lines denote uni-directional attention.

Model

- our pretraining of dialogue generation contains the following two tasks – response generation and latent act recognition.
- Input representation is the sum of four embeddings: corresponding token, role, turn and position embeddings.

➢Word embedding, for the latent variable .it also has an embedding matrix $E \in \mathbb{R}^{k \times D}$

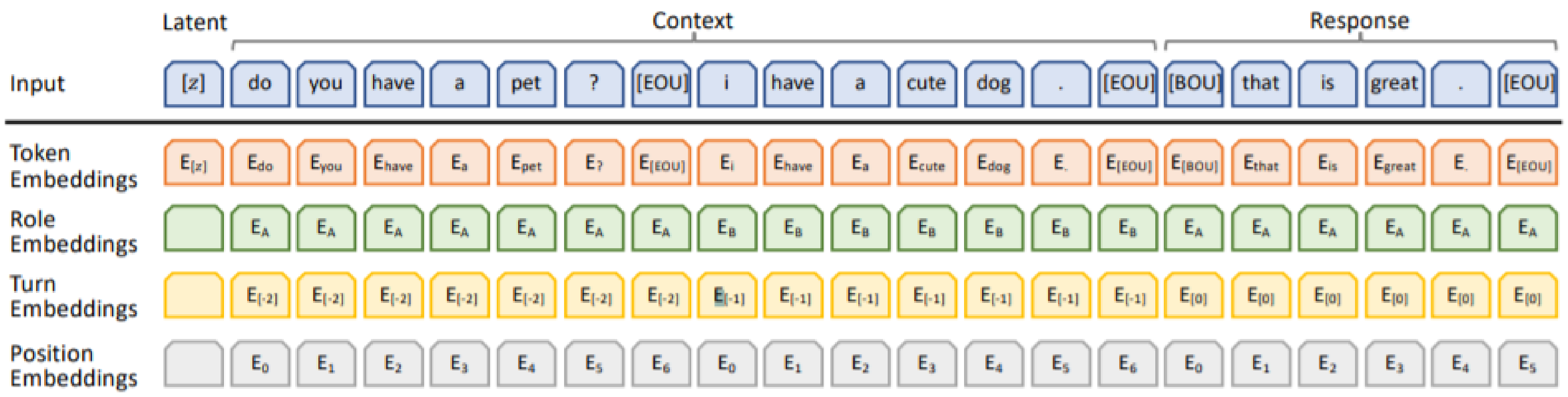
➢role embedding to differentiate the characters evolved in the conversation. E_A is the response, as well as the utterances generated by the same character ,E_B is for other characters E_C is for background knowledge

Model

- Input representation is the sum of four embeddings: corresponding token, role, turn and position embeddings.

➢there are multiturn utterances and we employ relative order in the assignment of turn embeddings. We use $E_{[0]}$ for turn embedding for this turn, $E_{[-1]}$ for last utterances, etc

➢Position embeddings are added according to the token position in each utterance.

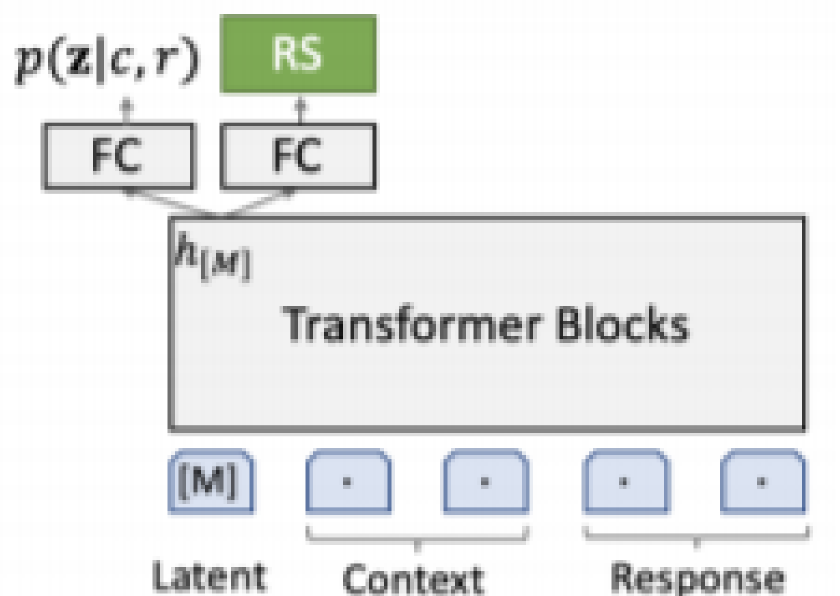


Response selection

- Response selection helps distinguish whether the response is relevant with the dialogue context and consistent with the background knowledge
- The positive sample comes from context and its corresponding response (c, r) ,negative sample comes from randomly selects response (c, r^{-}) ,

$$\mathcal{L}_{RS} = -\log p(l_r = 1|c, r) - \log p(l_{r^-} = 0|c, r^-)$$

$$p(l_r = 1|c, r) = \text{sigmoid}(W_3 h_{[M]} + b_3)$$



Response generation

- the response is generated conditioned on the latent variable and the context.
- NLLloss

$$\begin{aligned}\mathcal{L}_{NLL} &= -\mathbb{E}_{z\sim p(\mathbf{z}|c,r)} \log p(r|c,z) \\ &= -\mathbb{E}_{z\sim p(\mathbf{z}|c,r)} \sum_{t=1}^{T} \log p(r_t|c,z,r_{<t}) ,\end{aligned} \tag{2}$$

- Where z is the latent variable for this pair (c, r),

$$p(\mathbf{z}|c,r) = \mathrm{softmax}(W_1 h_{[M]} + b_1) \in \mathbb{R}^K$$

Response generation

- Besides NLLloss ,we also use the bag-of-words loss

$$\mathcal{L}_{BOW} = -\mathbb{E}_{z\sim p(\mathbf{z}|c,r)} \sum_{t=1}^{T} \log p(r_t|c,z)$$
$$= -\mathbb{E}_{z\sim p(\mathbf{z}|c,r)} \sum_{t=1}^{T} \log \frac{e^{f_{r_t}}}{\sum_{v\in V} e^{f_v}}$$

$$f = \mathrm{softmax}(W_2 h_z + b_2) \in \mathbb{R}^{|V|}$$

- the BOW loss discards the order of words and forces the latent variable to capture the global information of the target response.

Pretraing steps

- 1) latent act recognition

- Given a pair of context and target response pair, estimate the posterior distribution

- Randomly select r^{-} to calculate L_{RS}

- 2) Response Generation

-With the sampled latent value $\mathrm{z} \sim p(z|c,r)$， calculate L_{NLL} and L_{BOW}

- 3) Optimization

- Sum up to obtain L, update network parameters with back-propagation

Fine-tuning and Inference

- 1)Candidate Response Generation

Conditioned on each latent value $z \in [1, K]$, generate corresponding response r

2) Response Selection

Calculate the probability for each response $p(l_r = 1|c, r)$,and choose the one with the highest response

Dataset

- Persona –chat a knowledge grounded conversation dataset.
- Daily Dialog , which contains high-quality human conversations about daily life
- DSTC7-AVSD, in which the system need to generate an answer given dialogue context and background knowledge.(with caption and summary as

Dataset	Type	Knowledge	# Train	# Valid	# Test
Persona-Chat	Chit-chat with persona	Persona profiles	8,939 dialogues 131,438 turns	1,000 dialogues 15,602 turns	968 dialogues 15,024 turns
Daily Dialog	Chit-chat	N/A	11,118 dialogues 87,170 turns	1,000 dialogues 8,069 turns	1,000 dialogues 7,740 turns
DSTC7-AVSD	Conversational QA	Video caption & summary	7,659 dialogues 153,180 turns	1,787 dialogues 35,740 turns	1,710 dialogues 13,490 turns

Experiment

Dataset	Model	Automatic Evaluation			Human Evaluation			
		BLEU-1/2	Distinct-1/2	Knowledge R/P/F1	Fluency	Coherence	Informativeness	Overall
Persona-Chat	Seq2Seq	0.448 / 0.353	0.004 / 0.016	0.004 / 0.016 / 0.006	1.82	0.37	0.85	0.34
	LIC	0.405 / 0.320	0.019 / 0.113	0.042 / 0.154 / 0.064	1.95	1.34	1.09	1.29
	Our w/o Latent	**0.458** / **0.357**	0.012 / 0.064	0.085 / 0.263 / 0.125	1.98	1.36	1.04	1.30
	Our Method	0.406 / 0.315	**0.021** / **0.121**	**0.142** / **0.461** / **0.211**	**1.99**	**1.51**	**1.70**	**1.50**
Daily Dialog	Seq2Seq	0.336 / 0.268	0.030 / 0.128	-	1.85	0.37	0.44	0.33
	$iVAE_{MI}$	0.309 / 0.249	0.029 / 0.250	-	1.53	0.34	0.59	0.30
	Our w/o Latent	**0.405** / **0.322**	0.046 / 0.246	-	1.91	**1.58**	1.03	1.44
	Our Method	0.397 / 0.311	**0.053** / **0.291**	-	**1.97**	1.57	**1.23**	**1.48**

Human evaluation:randomly chooses 100 sentences, score the response quality on a scale of [0,1,2], from four aspects – fluency, coherence, informativeness and overall.

Experiment

Persona Profiles [P2]	Context
i love trying barbeque restaurants . i am an administrative assistant . i am looking forward to retiring in five years . i like ot play racquetball .	[P2] hi , do you like bbq ? [P1] hello yes i love bbq [P2] i love restaurants with bbq , they are a lot of fun . [P1] my poor toes are so webbed [P2] really ? afterwards , i go and play racquetball to loose some calories . [P1] cool i like to paint [P2] what do you do as a profession ? i work as administrative assistant [P1] i dig in people trash for things to build [P2] sounds unique but that is more of a hobby instead of a profession [P1] true . i paint for a living
Seq2Seq	that sounds like a lot of work . do you have a job ?
LIC	yeah it is a lot of fun
Our w/o Latent	do you have any hobbies ?
Our Method	that is cool , i am looking forward to retiring in 5 years