Eliciting Rich Positive Emotions in Dialogue Generation

Ziwei Gong [1] Qingkai Min[2] Yue Zhang[2]

[1]Columbia University [2]Westlake University

Motivation

Intentionality and effectiveness of conversation

In human communication theory, intentionality (intention of speakers) and effectiveness (effects of conversations) are key factors to a conversation, both of which can be exhibited by emotions. There has been research on dialogue systems for generating human-like, emotionally intelligent responses. However, existing work focuses on generating utterances with targeted emotion to express, yet few studies explore how one's emotion is affected by utterances, nor the intentionality of generated sentences.

One exception is **emotion elicitation**, which considers generating responses that elicit a pre-specified emotion in the other party. Though natural for humans to recognize and intentionally influence other's emotions, eliciting pre-specified emotions is challenging for dialogue models.

However, as shown in Figure 1, positive sentiment can include more fine-grained emotions such as "*Hopeful*", "*Joy*" and "*Surprise*", which can further serve to deepen the model's understanding of *effect*, if not *intention*. By incorporating more emotions in training, it ameliorates the performance in the elicitation of positive emotions. Besides, existing work is mostly based on small-scale human-annotated datasets, which limits its capacity of eliciting various emotions.

We fill this gap by proposing the first model for emotion elicitation that controls the generation of responses that elicit various pre-specified emotions.

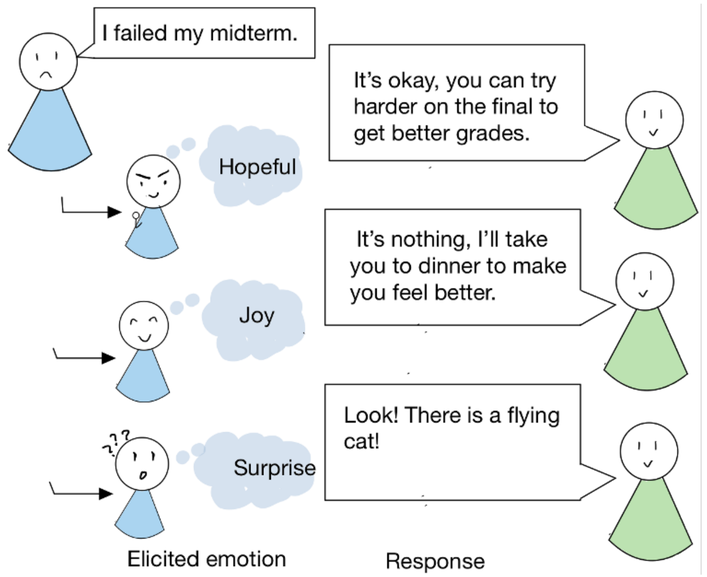

Figure 1. Examples of different responses that elicit different positive emotions.

Model

Effective EE-CVAE model that captures keys to elicite emotion.

Model Structure

Due to difficulties in annotation, we represent the elicited emotions using latent variables in order to take full advantage of the large-scale unannotated dataset, choosing Conditional Variational Auto-encoder (CVAE) as a backbone. Two discriminators are further used to control the generation of responses. The latent variable e is used to control the generation of the response. The latent variable z is separated from e to fully capture the elicited emotions

The overall structure of our model is shown in Figure 2. It can be seen as an extension of CVAE (in yellow) with a latent variable and two discriminators to elicit multiple emotions.

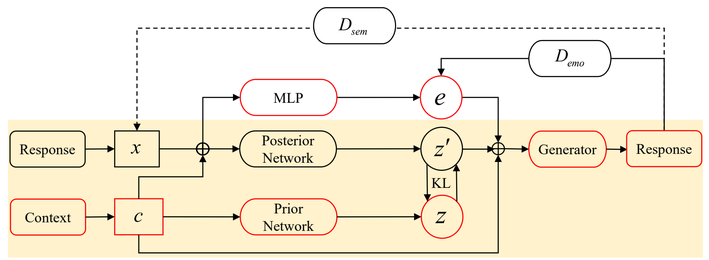

Figure 2. Illustration of our model. CVAE in yellow background. Red components are used for testing. Dashed arrow denotes a discriminator.

Experiments

For both baseline model EmpDG and our EE-CVAE modeel, we use more than 200k utterances from Friends and Open Subtitles datasets for pre-training the generator module, and a reconstructed MEmoR (TBBT) dataset to train the discriminators.

Model	TBBT - 9			
	PPL	Avg. len	KL	Acc.
EmpDG	667.4	8.7	-	
EmpDG$_{pre}$	462.2	9.2	-	0.290
Ours	196.4	14.3	25.9	
Ours$_{pre}$	91.5	13.2	14.0	0.448

Table 1. Results of models generation in comparison. "-" indicates not applicable, the average length for EmpDG is not reported because the generation results are unacceptable for most emotion categories. Human evaluations are conducted for selected models due to limited resources.

As shown in Table 1, our method indicates our model generates more fluent responses, possibly because the use of CVAE can be also more effective in isolating the influence of emotion signals. Compared with baseline, the rate of PPL reduction is markedly larger for ours when pre-training is added, which suggests that our CVAE structure benefits from wake-sleep learning and that more data allows the model to better separate the emotion signal e and the semantic context z.

In human evaluation, our model gives 44.8% accuracy, significantly better than 29.0% of the baseline. This shows the advantage of using a latent variable for modeling rich emotions, compared to hard-coding one emotion in a multi-encoder model. It also shows the effectiveness of our model in pretraining.

The Effect of Modeling Negative Emotions

Using all emotions in pretraining and finetuning produces the best performance in eliciting positive emotions.

	pretrain w/ all fine-tune w/ all	pretrain w/ all fine-tune w/ positive	pretrain w/ positive fine-tune w/ positive	Tie
Anticipation	.47	.32	.19	.02
Joy	.55	.215	.215	.02
Trust	.54	.17	.27	.02
All	.51	.25	.22	.02

Table 2. Results comparing three settings with the percentage of times one model is considered the best when eliciting different positive emotions.

Example generation

Context:Well, you be sure to let us know when you win the nobel prize for boysenberry.
Golden (anticipation): Hey. **EmpDG** (anticipation): yeah . **Ours** (joy): Oh , what a gentleman? **Ours** (trust): Wow , I really appreciate it.
Context: Aw, Amy, that was lovely. You know, this is fun. Let's do more. Someone else say something wonderful about me.
Golden (joy) Sheldon, I don't think everyone ... **EmpDG** (joy): What is great. **Ours** (joy) Oh, sure. Mmm. I told you, he's got too many. **Ours** (anticipation) And you.

Conclusions

We provided the first discussion on rich emotion elicitation in open-domain dialogue generation, incorporating various positive emotions with a framework that extends CVAE with a latent emotion variable equipped with two discriminators. Results show that rich emotion elicitation is a challenging task and our model gives more reliable utterances compared with a state-of-art model for single emotion elicitation, and introducing negative emotions in pretraining benefits the model's ability to elicit positive emotions.

zg2272@columbia.edu