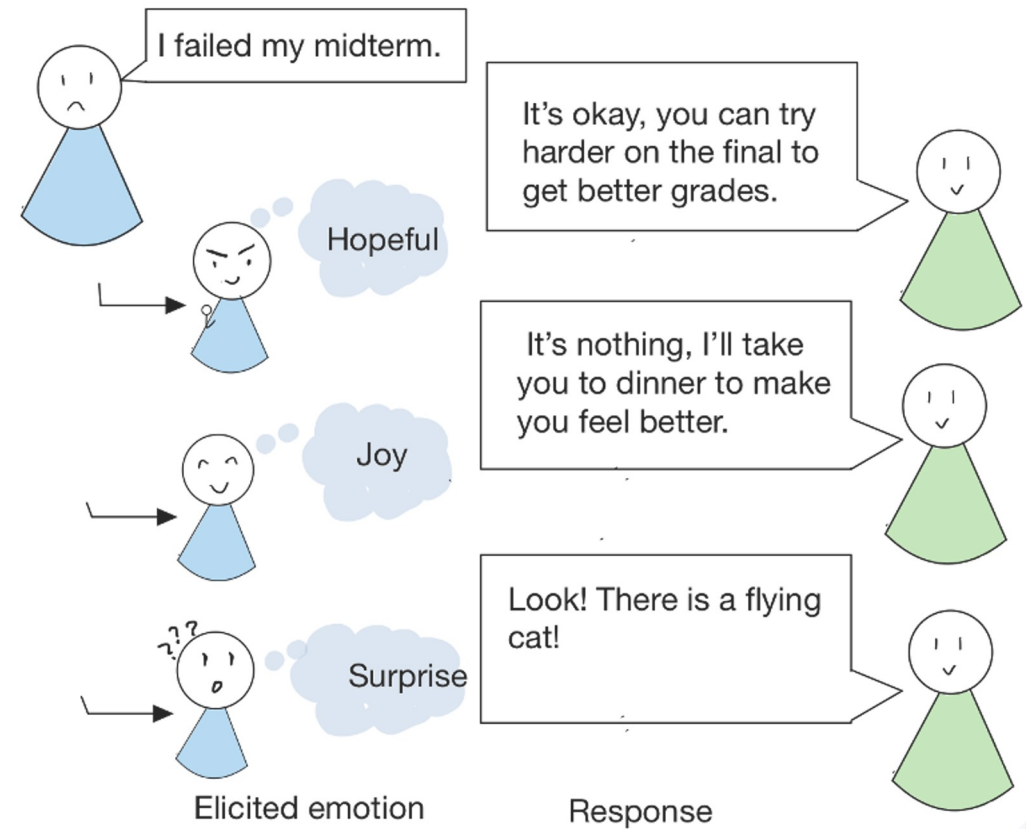
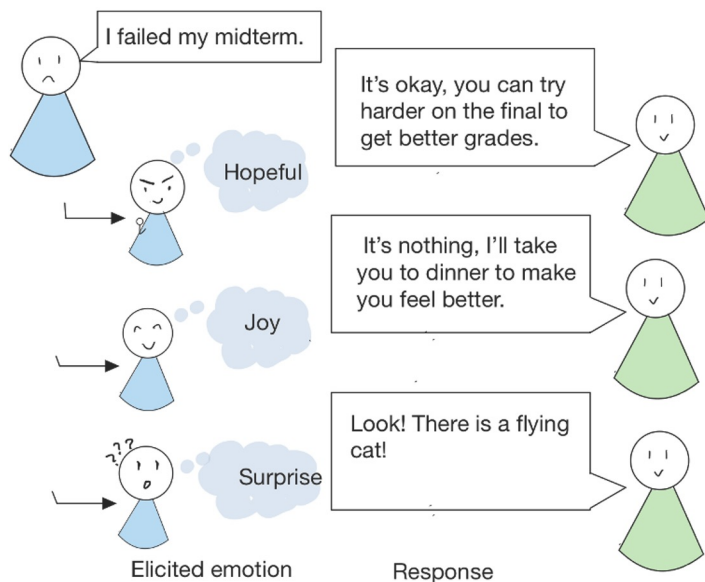


Eliciting Rich Positive Emotions in Dialogue Generation

- Sara Gong, Spring 2023

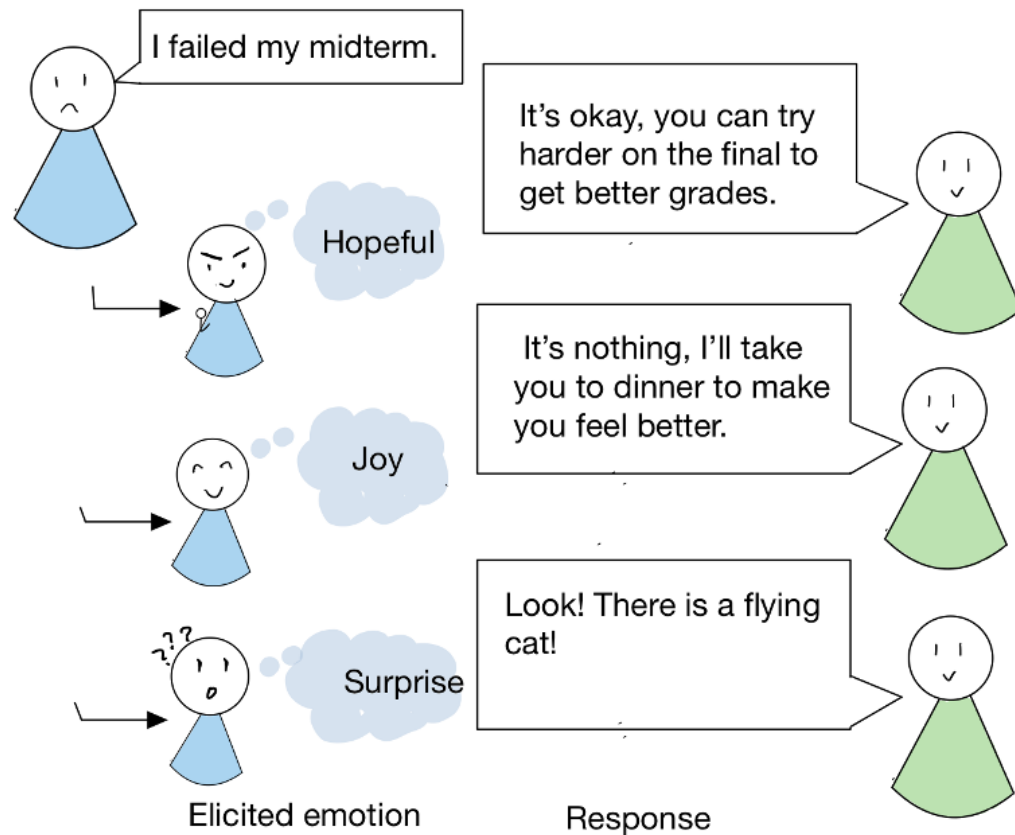


Related Work



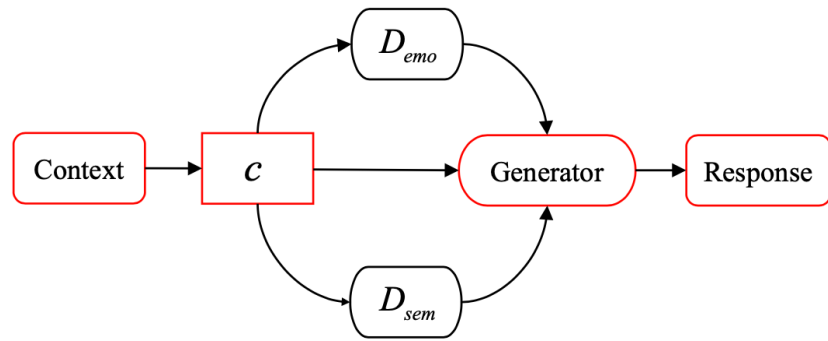
- Emotion in Generation
- Emotion Elicitation
 - statistical response generator (Hasegawa et al., 2013)
 - Hierarchical Recurrent Encoder Decoder model (Serban et al., 2016) extended with a separate layer of emotion modules (Lubis et al., 2018)
 - encoder-decoder adversarial model with two discriminators to increase emotion-awareness or empathetic dialogue generation (Li et al., 2020)
- Style Transfer
 - control text over multiple styles in generation while preserving the original content
 - Using Variational Autoencoder (VAE) and wake-sleep learning procedure (Fu et al., 2018; Tikhonov et al., 2019; Fei et al., 2020)
- Conditional Variational Autoencoder
 - CVAE is an extension of VAE, which has been used for dialogue generation (Chen et al., 2019) by introducing a latent variable to capture discourse-level variations (Zhao et al., 2017).

Motivations



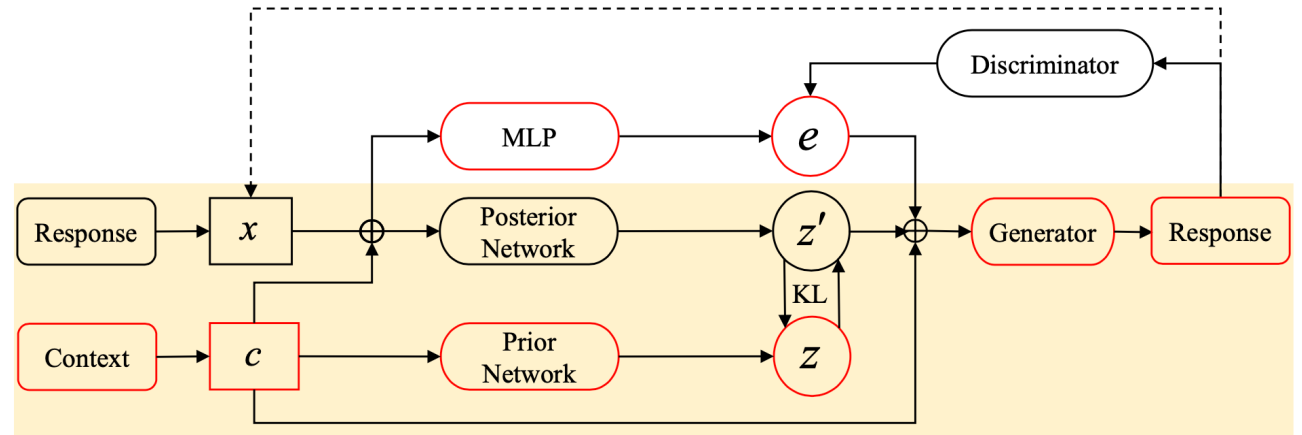
- Key factors to a conversation (in human communication theory):
 - **intentionality** (intention of speakers) and **effectiveness** (effects of conversations) (Littlejohn and Foss, 2010; Lindquist et al., 2015; Morick, 1971),
 - both exhibited by emotions (Dezecache et al., 2013).
- Current work on emotion elicitation focuses on positive sentiment
- However, 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**.
- Small-scale human-annotated datasets, which limit the capacity of eliciting various emotions.

Model Comparison



(a) Baseline mode: EmpDG

Single emotion category



(b) Our EE-CVAE model.

Multiple emotion categories

Model Detail

- CVAE for Dialogue Generation (yellow background)
- Adding Emotion Elicitation Function
- augment CVAE with a latent variable e , which is used to control the generation of a response together with the unstructured variable

$$\begin{aligned} \mathcal{L}_{\text{VAE}}(\theta, \phi) = & \mathbb{E}_{q_{\phi}(z|c,x)q_{\phi}(e|c,x)} [\log p_{\theta}(x|z, c, e)] \\ & - \text{KL}(q_{\phi}(z|c, x) \| p_{\theta}(z|c)) \\ & \leq \log p(x|c), \end{aligned}$$

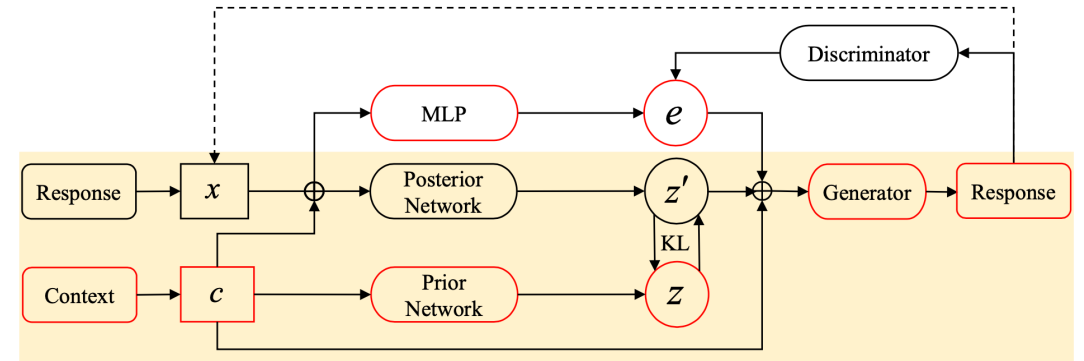
- a discriminator D is used to force the generator to produce coherent emotions

$$\mathcal{L}_{\text{Attr},e}(\theta) = \mathbb{E}_{p(z)p(e)} \left[\log q_D(e | \tilde{G}_{\tau}(z, e)) \right]$$

- Similarly, the variational encoder is reused to separate unrelated attributes from e by forcing them to be fully captured by z . It can be considered as another discriminator E :

$$\mathcal{L}_{\text{Attr},z}(\theta) = \mathbb{E}_{p(z)p(e)} \left[\log q_E(z | \tilde{G}_{\tau}(z, e)) \right].$$

- Combining, we have $\min \mathcal{L}_G = \mathcal{L}_{\text{VAE}} + \lambda_e \mathcal{L}_{\text{Attr},e} + \lambda_z \mathcal{L}_{\text{Attr},z}$.



(b) Our EE-CVAE model.

Training illustration of our model. Red components are used for testing. CVAE in yellow background. Dashed arrow denotes a discriminator.

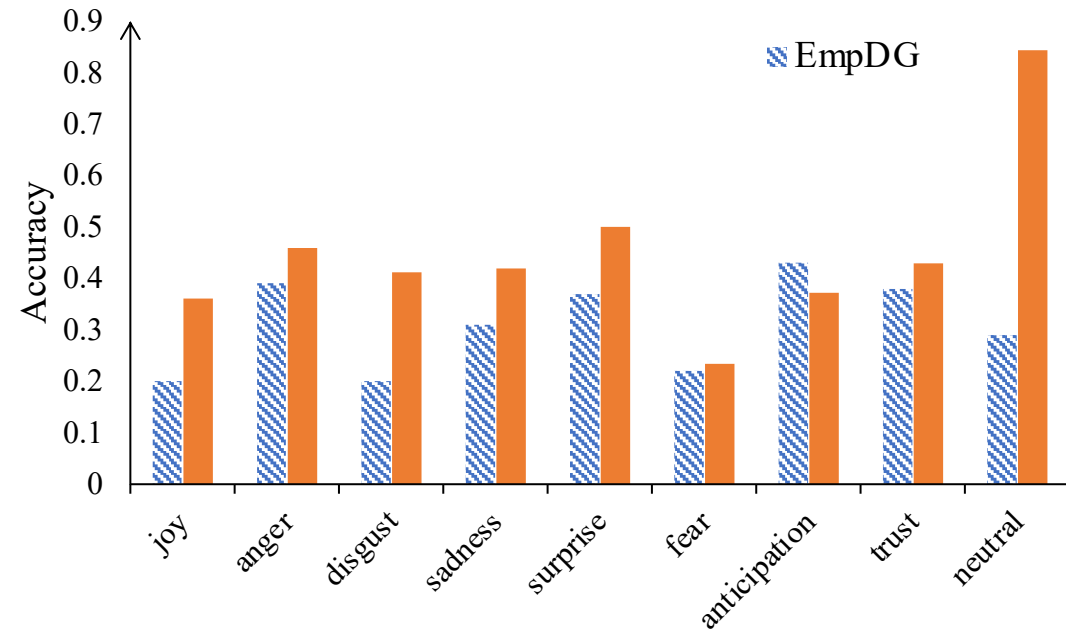
Dataset

- Reconstructed the multi-modal MEmoR dataset (Shen et al., 2020) to fit our emotion elicitation task and conducted human evaluation to validate the usability in a single modality. (annotator agreement of 80% accuracy (Cohen's $\kappa = 0.491$))
- The reconstructed corpus has 22,732 utterances
 - Split the data in training (18,943), dev (1,894), and test (1,894).
- Pretrain: we use more than 200k utterances from the Friends (Zahiri and Choi, 2017) and Open Subtitles2 datasets

Results

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.



Comparison of accuracy across 9 emotions.

Sample generations

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.

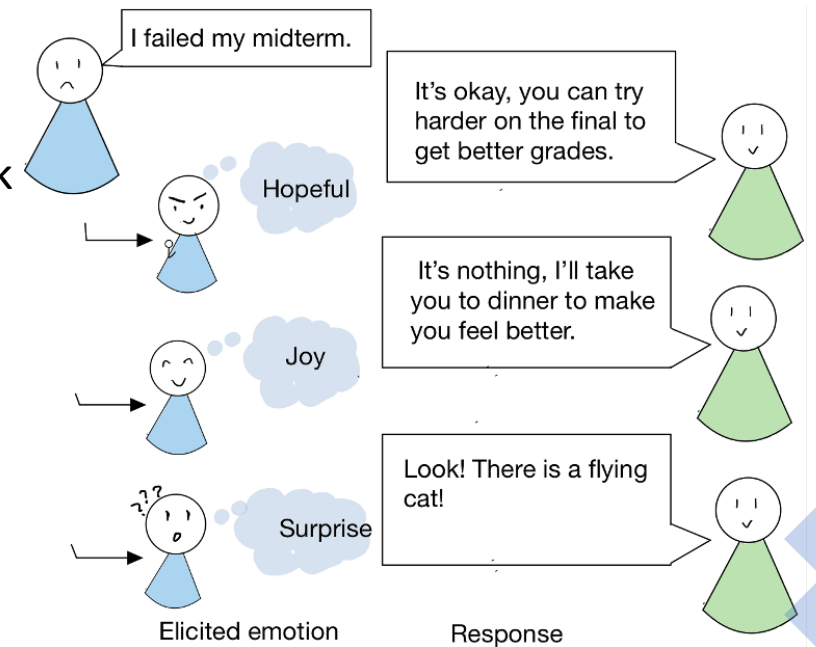
The Effect of Modeling Negative Emotions


- Results comparing three settings with the percentage of times one model is considered the best when eliciting different positive emotions.
- Setting 1: modeling all emotions in pretraining and fine-tuning.
- Setting 2: modeling all emotions in pretraining, fine-tuning with only positive emotions.
- Setting 3: modeling only positive emotions in pretraining and fine-tuning.
- Using all emotions in pretraining and finetuning produces the best performance in eliciting positive emotions.

	Setting1	Setting2	Setting 3	Tie
Anticipation	.47	.32	.19	.02
Joy	.55	.215	.215	.02
Trust	.54	.17	.27	.02
All	.51	.25	.22	.02

Conclusions and Future Directions

- Using **all** emotions in pretraining and finetuning produces the best performance in eliciting positive emotions.
- Results show the advantage of using a latent variable for **modeling rich emotions**, compared to hard-coding one emotion in a multi-encoder model.
- The effectiveness of our model in **pretraining**.
- Our results show that rich emotion elicitation is a challenging task for current neural models and there is a need for more effective few-shot learning.





Thank you! Questions?

