Jointly Optimizing Diversity and Relevance in Neural Response Generation

Xiang Gao, Sungjin Lee, Yizhe Zhang, Chris Brockett, Michel Galley, Jianfeng Gao, Bill Dolan

Microsoft Research, Redmond, WA

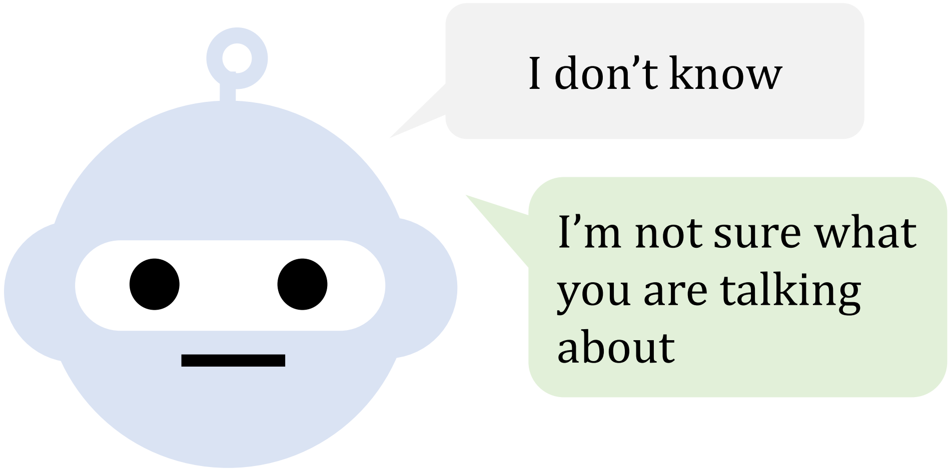
Paper: arxiv.org/abs/1902.11205

Code: github.com/golsun/**SpaceFusion**

Presented at NAACL-HLT 2019

Motivation

- 😎 Generate relevant and interesting response
- 😴 Vanilla Seq2Seq tend to be bland/generic
- Need to optimize diversity



Prior work

Decoding/ranking

- Li et al. (2016) rank beam search results by mutual information with the context.
- However requires a large beam width (e.g. 200).

Training/latent space

- Zhao et al. (2017) use a conditional VAE to model the discourse-level diversity
- However observe reduced relevancy*

*unless extra knowledge (dialog act) is provided

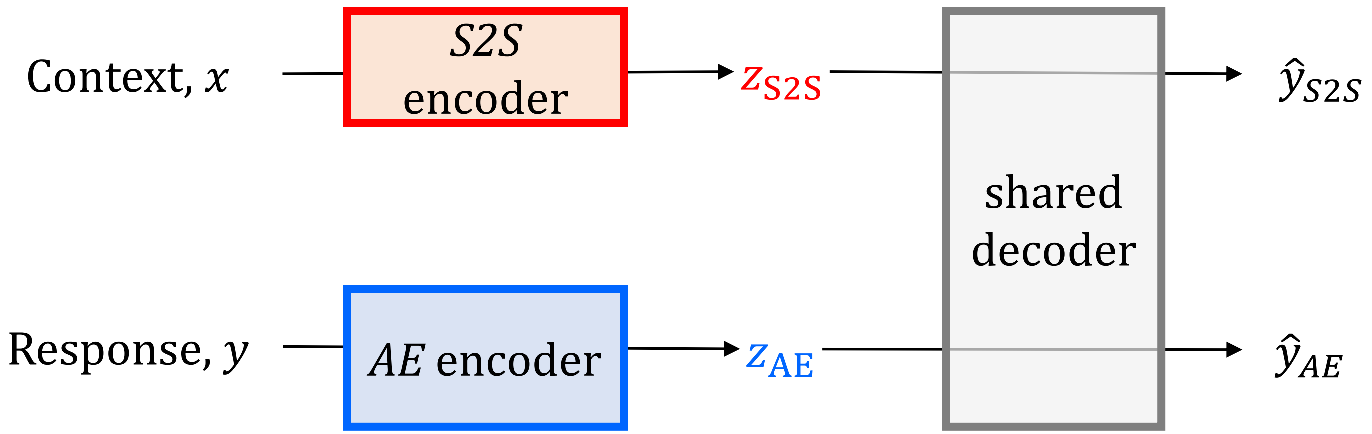
Idea

Jointly optimize diversity and relevancy at training by aligning two models

- **Sequence-to-Sequence (S2S)** → latent vector of *context*
- **Autoencoder (AE)** → latent vectors of multiple possible *diverse* responses

How to combine them in a shared latent space?

One easy way: a vanilla multi-task setting (Luan et al. 2017)

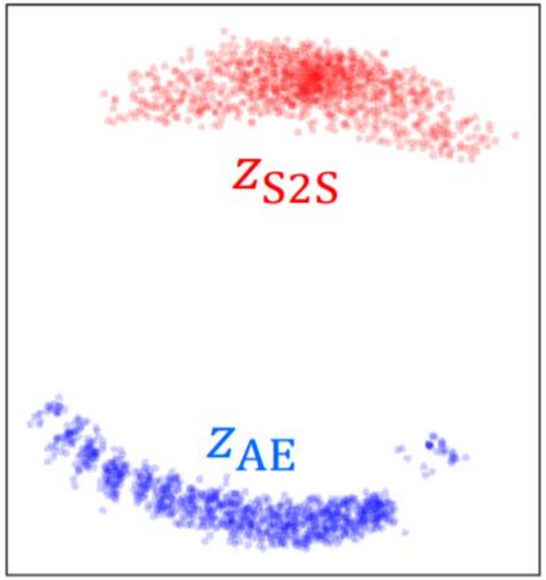


But…🤔 it's not easy to align S2S and AE

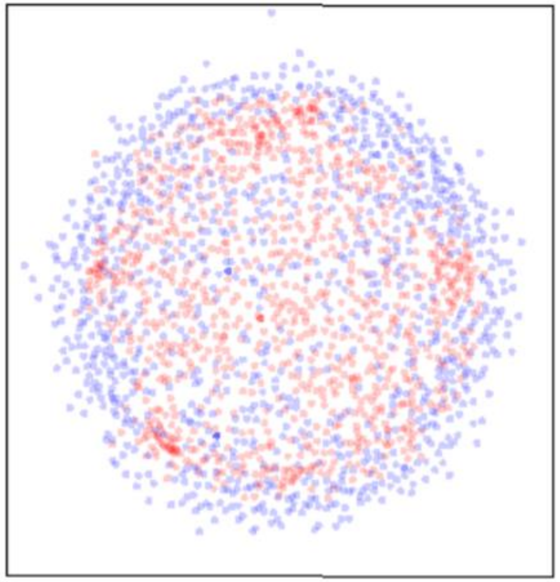
Though the decoder is shared, S2S and AE still form two separate clusters

- S2S not efficiently using what AE learned
- Un-desired property: big holes/gaps!

What a vanilla multi-task does



What we desire



MDS visualization of learned latent spaces. Left one is Luan et al. 2017; right one is ours

So we add regularization 🤝

- Pull S2S and AE dots closer to each other

$$\mathcal{L}_{\text{fuse}} = \underbrace{\sum_{i \in \text{batch}} \frac{d(z_{\text{S2S}}(x_i), z_{\text{AE}}(y_i))}{n}}_{\textit{cross-space distance}} - \underbrace{\sum_{i,j \in \text{batch}, i \neq j} \frac{d(z_{\text{S2S}}(x_i), z_{\text{S2S}}(x_j))}{n^2 - n} - \sum_{i,j \in \text{batch}, i \neq j} \frac{d(z_{\text{AE}}(y_i), z_{\text{AE}}(y_j))}{n^2 - n}}_{\textit{same-space distance}}$$

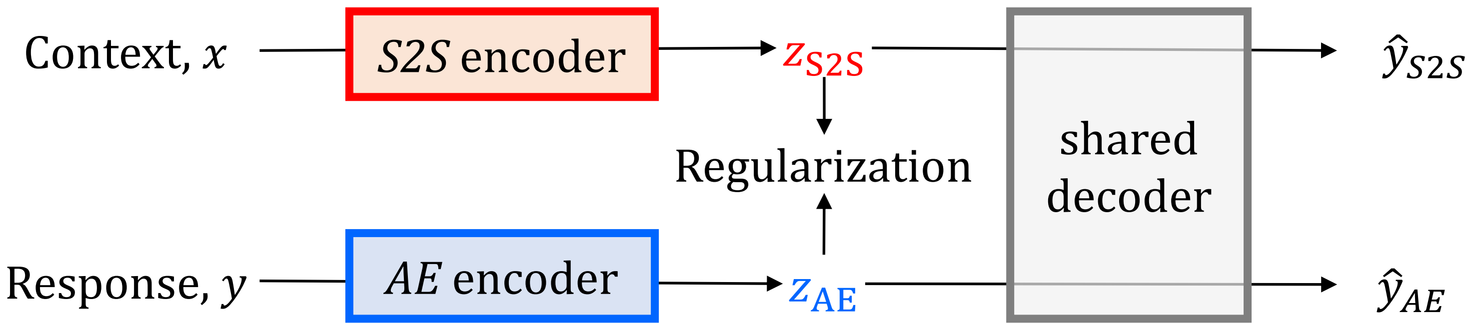
- Encourage a smooth transition between S2S and AE

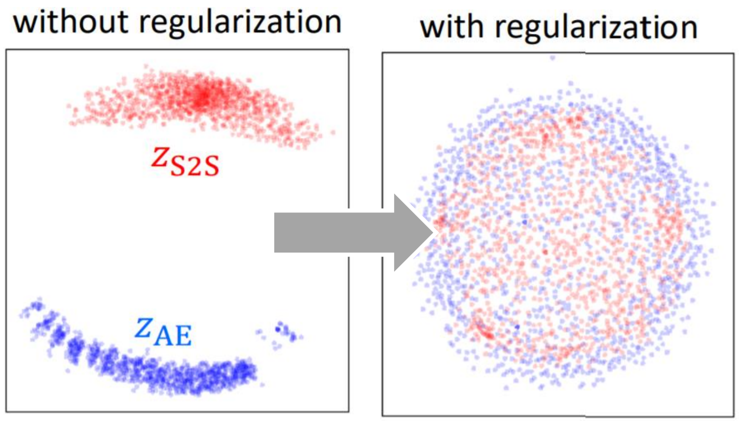
$$\mathcal{L}_{\text{interp}} = -\frac{1}{|y|} \log p(y | z_{\text{interp}})$$

where $z_{\text{interp}}(x_i, y_i) = u z_{\text{S2S}}(x_i) + (1-u) z_{\text{AE}}(y_i)$, x_i, y_i are paired context and response

- Finally combine them with vanilla multi-task loss

$$\mathcal{L} = \underbrace{\underbrace{\underbrace{-\frac{1}{|y|} \log p(y | z_{\text{S2S}})}_{\textit{Vanilla S2S}} - \frac{1}{|y|} \log p(y | z_{\text{AE}})}_{\textit{Vanilla MTask}} + \alpha \mathcal{L}_{\text{interp}} + \beta \mathcal{L}_{\text{fuse}}}_{\textit{SpaceFusion}}$$





Structured latent space

The regularization terms induce some desired ***structure*** of the latent space

Semantic → Geometry

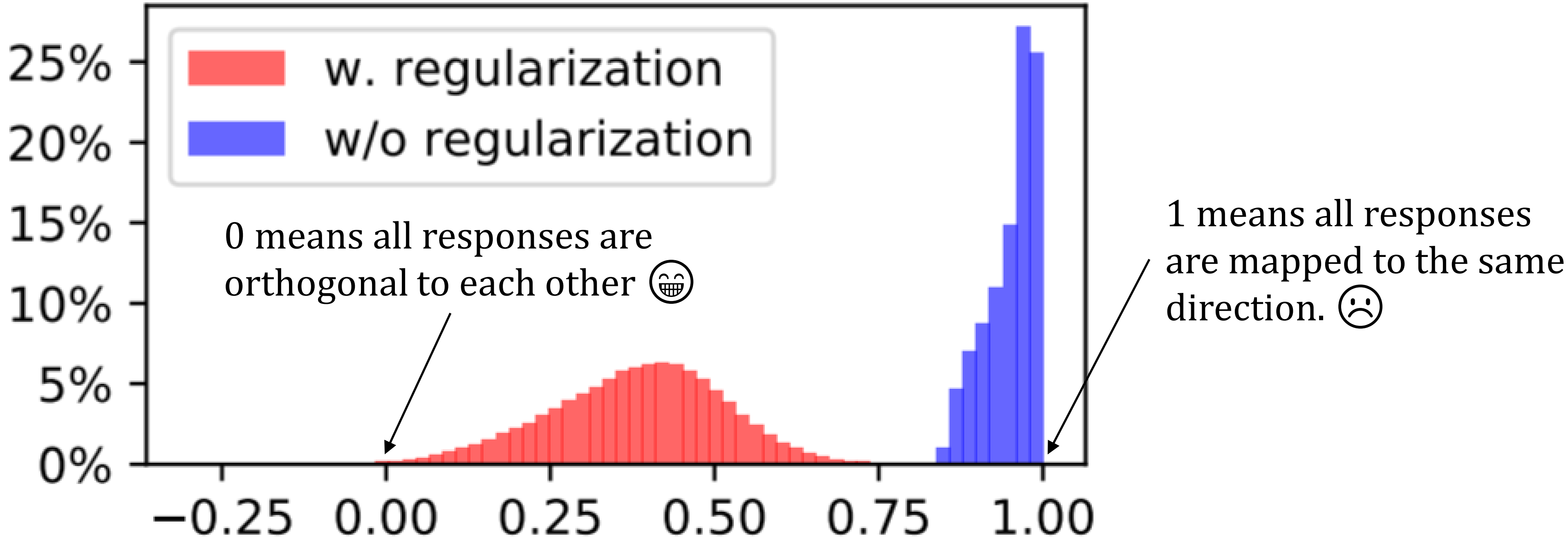
- **Diversity → direction:** as L_{interp} regularized semantic along a "line"
- **Relevancy → distance:** as L_{fuse} regularized distance

✓ Disentangle relevancy and diversity in the latent space
✓ Didn't impose any pre-defined distribution (e.g. Gaussian)

context x: Anyone want to start this game?					
response at $u = 0$: I would love to play this game.					
u	**towards "No I don't."**	u	**towards "when?"**	u	**towards "Yes I do."**
0.18	I am not interested in the game.	0.15	I'd be interested in the game	0.15	I'd love to play it.
0.21	I am not interested.	0.31	When is it?	0.27	Yes I do.
0.30	No I don't.	0.40	When will you?		
		1.00	When?		

Direction & diversity

SpaceFusion tend to map different possible responses to different direction



Cosine similarity between the direction towards different responses

Interpolation & smoothness

context x: Anyone want to start this game? towards one possible target y: Yes I do.			
u	**with regularization**	u	**without regularization**
0.00	I would love to play this game.	0.00	I would have to play with the game.
0.15	I would love to play it.	0.29	Dude, I know, but, or etc.
0.30	Yes I do	0.61	Op I was after though today
		0.85	I'm single :(though
		0.90	Yes I do.

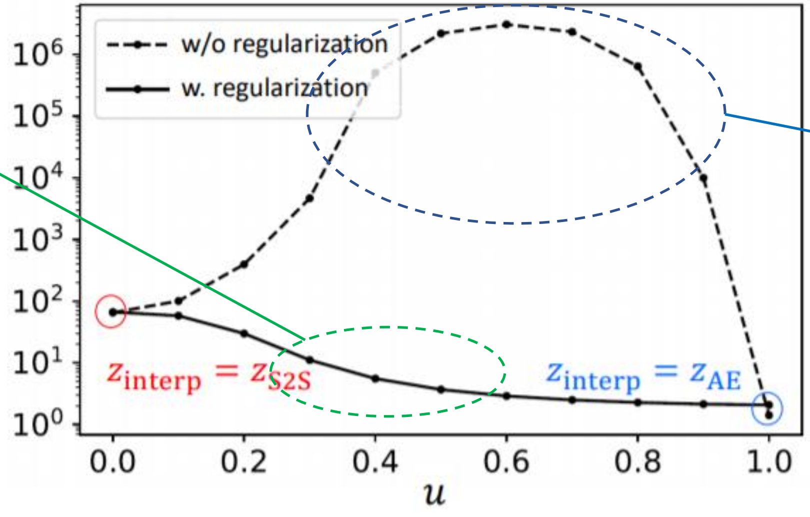
nicer interpolation 😁 →

Wired interpolation ☹

where $z_{\text{interp}}(x_i, y_i) = u z_{\text{S2S}}(x_i) + (1-u) z_{\text{AE}}(y_i)$, x_i, y_i are paired context and response

The perplexity of $y|z_{\text{interp}}$

likely (low perplexity) to generate target response → better interpolation



Not likely (high perplexity) to generate target response → wired interpolation

Experiments

- Contributed a large dataset from Reddit with natural multi-reference

	Switchboard	Reddit
train (x, y) samples	0.2M	7.3M
test (x, y) samples	5418	5000
ref. source	IR+filtering	natural
ref. availability	test only	train/vali/test
ref. per context	7.7	24.1

- Competitive baselines: CVAE+BOW (Zhao et al. 2017), MTask (Luan et al. 2017)

- **Human evaluation**

	relevance	interest	average
SPACEFUSION	**2.72**	**2.53**	**2.63**
CVAE+BOW	2.51	2.37	2.44
Multi-Task	2.34	2.14	2.24
human	3.59	3.41	3.50

Automatic evaluation

Following [Zhao et al. (2017)]

$$\text{Precision} = \frac{1}{N_r}\sum_{i=1}^{N_r} \max_{j\in[1,N_r]} \text{BLEU}(r_j, h_i)$$

Relevancy: Does a hypothesis "match" any reference?

$$\text{Recall} = \frac{1}{N_r}\sum_{j=1}^{N_r} \max_{i\in[1,N_r]} \text{BLEU}(r_j, h_i)$$

Diversity: Does all diverse references can be "matched" by some hypothesis?

dataset	model	Precision	Recall	F1
Switchboard	SPACEFUSION	**1.22**	**0.66**	**0.86**
	CVAE+BOW	0.76	0.57	0.65
	MTask	0.75	0.43	0.54
	S2S+Sampling	0.57	0.48	0.52
Reddit	SPACEFUSION	**0.40**	**0.26**	**0.31**
	CVAE+BOW	0.16	0.18	0.17
	MTask	0.31	0.18	0.23
	S2S+Sampling	0.10	0.11	0.11

$$\text{F1} = 2\,\frac{\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}}$$

Summary

We proposed a *regularized* multi-task learning approach, **SpaceFusion**, to jointly optimized relevancy and diversity by disentangling them in a structured latent spaces

- Align latent spaces from different models – potential for learning more universal representation
- Bring structure into latent spaces – interpretability and easier usage

Thank you!

SpaceFusion by Microsoft Research
Paper: arxiv.org/abs/1902.11205
Code: github.com/golsun/SpaceFusion
Presented at NAACL-HLT 2019