Adversarial Learning on the Latent Space for Diverse Dialog Generation

Kashif Khan*† , Gaurav Sahu*† , Vikash Balasubramanian*† , Lili Mou[ψ], Olga Vechtomova†

† University of Waterloo; [ψ]University of Alberta, Alberta Machine Intelligence Institute (Amii)

Objectives

Given a natural language query (with context):

- learn meaningful representation of text, and
- generate *diverse, relevant, & fluent* responses

Introduction

A dialog generation system has two main tasks:

- encoding the context in a conversation, and
- generating a response based on the given context.

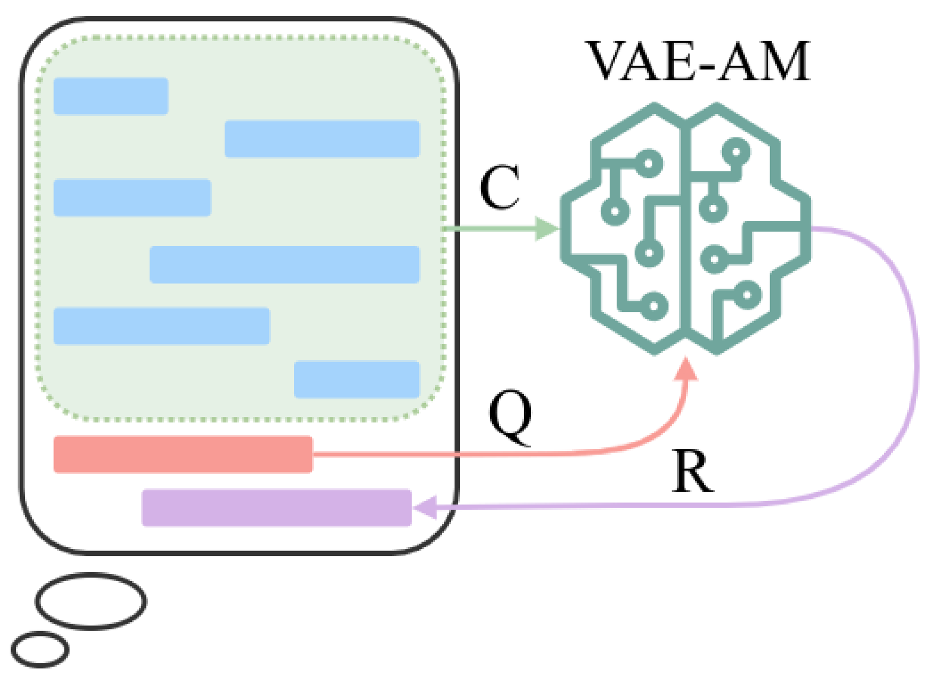


Figure: General pipeline. Q: Query, C: Context, R: Response

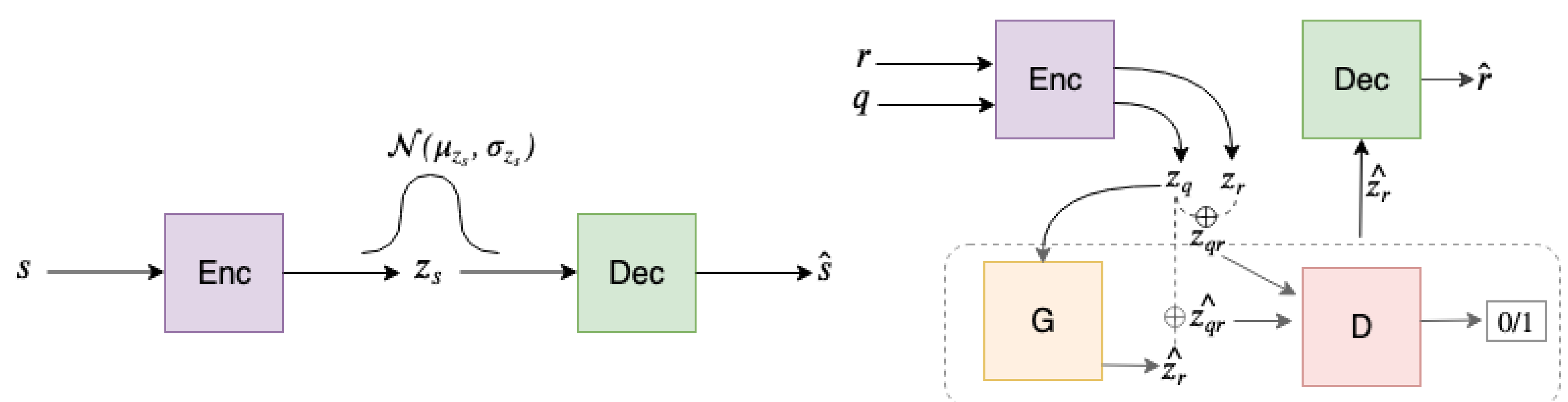
Approach

Proposed two-step approach for dialog-generation:

- **Step 1:** Train a **VAE** [1] to learn a real-valued vector representation of a generic sentence.
- **Step 2:** Train an **adversarial network** on the VAE's latent space $\boldsymbol{z}$ for dialog generation.

Given a query-response pair (q, r) in the training set, we use the trained encoder from Step 1 to obtain query and response latent variables $\boldsymbol{z}_q$ and $\boldsymbol{z}_r$ respectively. The adversarial generator G then learns to match $\hat{\boldsymbol{z}}_r = G(\boldsymbol{z}_q)$ and $\boldsymbol{z}_r$. Here, the adversarial discriminator D classifies $(\boldsymbol{z}_r, \boldsymbol{z}_q)$ versus $(\hat{\boldsymbol{z}}_r, \boldsymbol{z}_q)$.

Complete Architecture Diagram



(a) Step 1: Variational Autoencoder (b) Step 2: Adversarial network (dashed box)

Figure: Architecture of our proposed two-step training procedure. ⊕ denotes concatenation.

Important Result

Adversarial learning with MSE loss improves **relevance, fluency,** and **diversity** of responses in dialog.

Results

Model	BLEU			Diversity						Fluency
	Avg	Max	HM	Intra-1	Intra-2	Inter-1	Inter-2	ASL (14.43)	TTR	PPL
Single-turn results										
Seq2Seq	0.143	0.217	0.172	**0.99**	**0.99**	0.46	0.49	4.63	0.019	18.45
WED-S	0.215	0.357	0.268	0.94	**0.99**	0.48	0.74	10.42	0.034	33.91
DialogWAE	0.296	0.356	0.323	0.85	0.97	0.42	0.74	19.34	0.005	20
VAE-M (ours)	0.191	0.293	0.231	0.98	**0.99**	**0.5**	0.79	9.36	0.029	19.7
VAE-A (ours)	0.295	0.359	0.323	0.93	**0.99**	0.46	0.76	13.64	0.035	21.38
VAE-AM (ours)	**0.306**	**0.367**	**0.334**	0.91	**0.99**	0.46	**0.82**	16.90	**0.034**	**17.01**
Multi-turn results										
HRED*	0.232	0.232	0.232	0.94	0.97	0.09	0.10	10.1	-	-
CVAE*	0.222	0.265	0.242	0.94	0.97	0.18	0.22	10.0	-	-
CVAE-CO*	0.244	0.259	0.251	0.82	0.91	0.11	0.13	11.2	-	-
VHCR*	0.266	0.289	0.277	0.85	0.97	0.42	0.74	16.9	-	-
DialogWAE	0.279	0.365	0.316	0.79	0.92	0.35	0.68	19.84	0.007	161.86
VAE-AM (ours)	**0.314**	**0.371**	**0.340**	**0.847**	**0.98**	**0.41**	**0.73**	15.3	**0.036**	**119.39**

Table: Dialog generation results on the de-duplicated DailyDialog dataset [2]

Loss Functions

- **Step 1:** KL divergence loss for the VAE is given by (λ_{KL} balances the two terms):

$$J_{\mathrm{AE}}(\boldsymbol{\theta}_{\mathrm{Enc}}, \boldsymbol{\theta}_{\mathrm{Dec}}) = -\mathbb{E}_{q_E(\boldsymbol{z}|s)}[\log p(s|\boldsymbol{z})] + \lambda_{\mathrm{KL}}\mathrm{KL}(q_E(\boldsymbol{z}|s)||p(\boldsymbol{z})) \quad (1)$$

- **Step 2:** Overall loss for Step 2 is given as:

$$J = J_{\mathrm{CGAN}} + \gamma J_{\mathrm{MSE}} \text{ where,}$$
$$J_{\mathrm{MSE}} = ||\boldsymbol{z}_r - \hat{\boldsymbol{z}}_r||^2$$
$$J_{\mathrm{CGAN}} = \min_G \max_D \mathbb{E}_{(\boldsymbol{z}_q, \boldsymbol{z}_r)\sim\mathcal{D}_{\mathrm{train}}}[\log D(\boldsymbol{z}_r, \boldsymbol{z}_q) + \log(1 - D(G(\boldsymbol{z}_q), \boldsymbol{z}_q))] \quad (2)$$

Conclusion

- We propose an effective two-stage model for dialog generation
- We make use of sentence representations learned by a VAE and train a adversarial network on VAE's latent space to generate diverse responses given a query and context
- We observe that our model outperforms existing state-of-the-art approaches by generating more diverse, fluent, and relevant sentences.

References

[1] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *ICLR*, 2014.

[2] Hareesh Bahuleyan, Lili Mou, Hao Zhou, and Olga Vechtomova. Stochastic Wasserstein autoencoder for probabilistic sentence generation. In *NAACL-HLT, Volume 1*, pages 4068–4076, 2019.

Acknowledgements

NSERC (Grant Nos. RGPIN-2019-04897 and RGPIN-2020-04465); Amii Fellow Program; Canadian CIFAR AI Chair Program; Compute Canada.