Better Conversations *by* Modeling, Filtering, and Optimizing *for* Coherence and Diversity

Xinnuo Xu, Ondřej Dušek, Ioannis Konstas and Verena Rieser

Interaction Lab, Heriot-Watt University

Introduction

- **New GloVe-based measure of coherence** *(Coh)* between dialogue context and generated response
- Filtering corpora using *Coh* to get more coherent & diverse training data
- **New Generator based on conditional Variational Autoencoders** (cVAE)
  - *Coh* used as latent variable
  - **Novel** *Coh*-based **context gate** to improve coherence & diversity while generating

Measure of Coherence

Coh = cosine similarity over averaged GloVe embeddings between context vs. response

Coh	Dialogue context	Response
≥0.9	yeah? ## you're new in town, right?	you call this a town?
≥0.9	to find the president.	the president, he lives!
0.7-0.9	call your grandfather to pick me up.	grandpa 's not here.
0.7-0.9	kitchen. ## definitely the kitchen.	specifically the stove.
0.4-0.6	so the boy becomes a man	it's amazing.
0.4-0.6	can we stop at the drug store?	oh, uh, don't worry.
<0.4	mag? ## thank you.	i'm going to go for a walk.
<0.4	i dont want it. ## why ## ok	you want the car.

Data and Pre-processing

- **OST**: OpenSubtitles corpus with automatic dialogue turn segmentation (Lison and Meena, 2016).
- **fOST**: OST + *Coh*-based filtering
  - Only data where $Coh > 0.68$, i.e. $> \overline{Coh} + 2 \cdot \sigma$

Dataset	Coh	D-1%	D-2%	D-Sent%
OST	0.390	14.3	57.9	83.8
fOST	0.801	15.5	62.9	89.3

Conclusions

- Explicity **modelling & optimizing for coherence** significantly **improves dialogue response generation**
- Our cVAE-based models consistently outperform competitive baselines
- Our **cVAE-based models produce better responses** even when trained on noisy, incoherent data

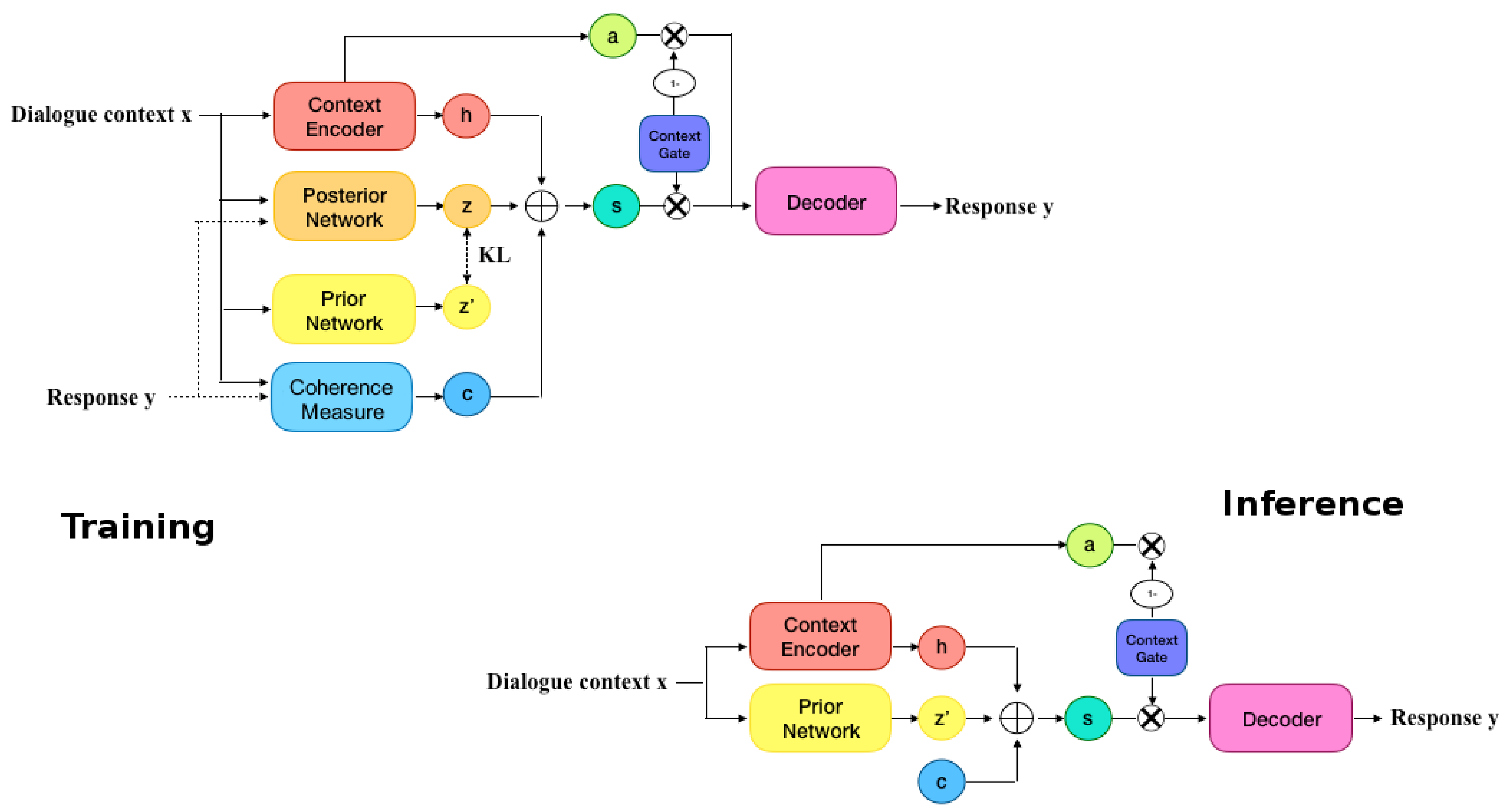
Code available at:

```
https://github.com/XinnuoXu/CVAE_Dial
```

Our Model: cVAE + Context Gate

Enhancements over seq2seq model with attention:

- Conditioning on context generally: latent variable *z*
- Conditioning on *Coh* explicitly: latent variable *c*
- Context gate: balance between words generated so far & previous context



Combined training objective:

- Generation loss (MLE): response should be close to ground truth
- Coherence loss: match coherence signal given by *c*
- Diversity loss (cVAE): responses should be diverse so they can reproduce *z*

Experiments

Evaluation metrics:

- BLEU: word overlap, *Coh*: coherence (GloVe-based)
- D-1, D-2, D-Sent: The proportion of distinct unigrams, bigrams, and sentences in the outputs

Model variants

- cVAE-XGate: our context gate
- cVAE-CGate: context gate variant of Tu et al. (2017)

Results – models trained on OST, evaluated on fOST (hardest)

Model	BLEU%	Coh	D-1%	D-2%	D-Sent%
Seq2seq	0.86	0.284	3.6	14.6	29.4
MMI (Li et al. 2015)	0.89	0.278	3.7	15.3	31.5
cVAE-CGate	2.25	0.422	**5.4**	**28.2**	81.0
cVAE-XGate	**2.41**	**0.434**	4.8	23.4	**84.0**

Generation output examples

- comparing MMI (B-MMI) and cVAE-XGate (B-XGT), B-GT is the ground truth response

	Dialogue context	Response
1	A: i have an audition at 4:00. B: apparently now i'm cooking dinner. A: how did this all happen?	B-GT: you volunteered. B-MMI: i don 't know. B-XGT: well , i'm going to have a dinner.
2	A: was it what you wanted? B: no. A: i wanted an airplane.	B-GT: you like model planes? B-MMI: i wanted to know. B-XGT: i wanted a helicopter.
3	A: great. B: thanks so much. A: vanilla decaf latte?	B-GT: yeah, a vanilla decaf latte. B-MMI: no. B-XGT: yeah , that 's a good coffee

EPSRC
Engineering and Physical Sciences Research Council

Acknowledgements

This research is supported by the Engineering and Physical Sciences Research Council (EPSRC), as part of the CDT in Robotics and Autonomous Systems at Heriot-Watt University and The University of Edinburgh.

Contact Details

Email: xx6, o.dusek, i.konstas, v.t.rieser@hw.ac.uk

Web: www.edinburgh-robotics.org

Twitter: @EDINrobotics