Variational Memory Encoder-Decoder

Hung Le, Truyen Tran, Thin Nguyen and Svetha Venkatesh

{lethai, truyen.tran, thin.nguyen, svetha.venkatesh}@deakin.edu.au

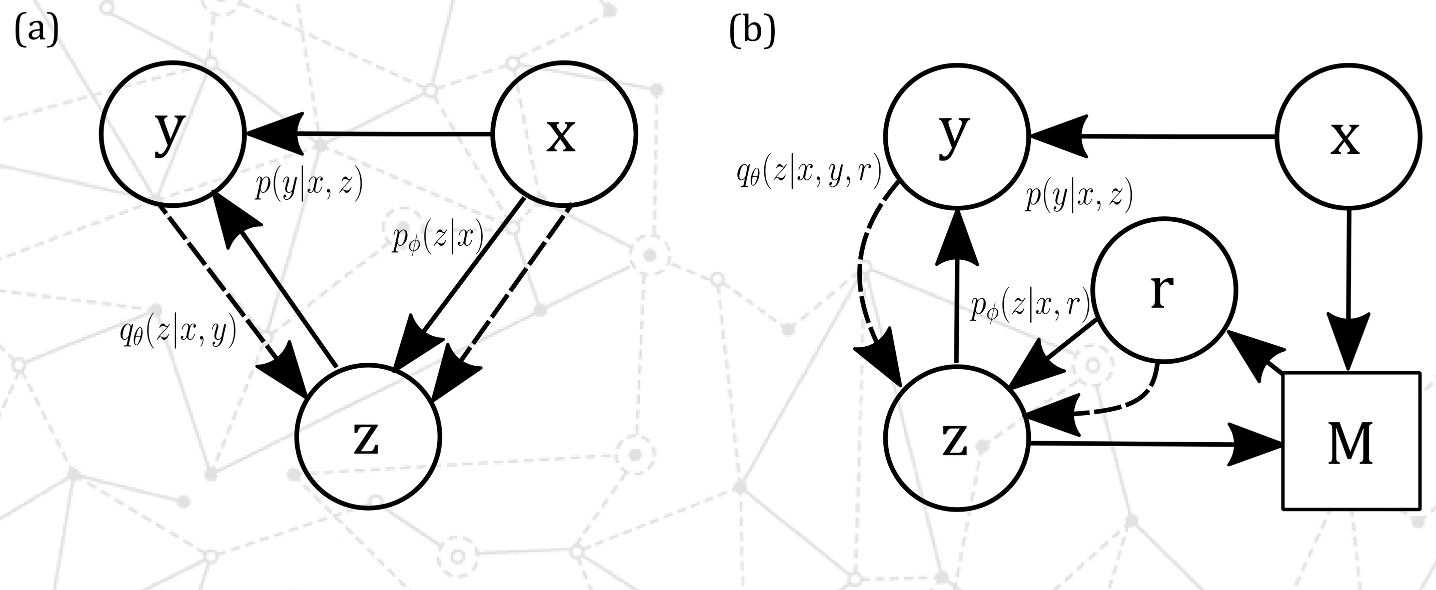
DEAKIN UNIVERSITY AUSTRALIA Worldly

Introduction

Introducing variability while maintaining coherence is a core task in learning to generate utterances in conversation. Standard neural encoder-decoder models and their extensions using conditional variational autoencoder often result in either trivial or digressive responses. To overcome this, we explore a novel approach that injects variability into neural encoder-decoder via the use of external memory as a mixture model, namely Variational Memory Encoder-Decoder (VMED).

Methods

Built upon CVAE [1] and partly inspired by VRNN [2], we introduce Variational Memory Encoder-Decoder (VMED). With an external memory module, VMED explicitly models the dependencies between latent random variables across subsequent timesteps. Unlike the VRNN which uses hidden values of RNN to model the latent distribution as a Gaussian, our VMED uses read values r from an external memory M as a Mixture of Gaussians (MoG) to model the latent space.



Graphical Models of the vanilla CVAE (a) and our proposed VMED (b)

VMED formulations

- The prior:

$$p_\phi(z_t|x,r_{t-1}) = \sum_{i=1}^{K} \pi_t^{i,x}(x,r_{t-1}^i)\,\mathrm{N}\left(z_t;\mu_t^{i,x}(x,r_{t-1}^i),\sigma_t^{i,x}(x,r_{t-1}^i)\mathbf{I}\right)$$

- The posterior:

$$q_\theta(z_t|x,r_{t-1},y_{\le t}) = \mathrm{N}\left(z_t;\mu_t^{x,y}(x,r_{t-1}^i,y_{\le t}),\sigma_t^{x,y}(x,r_{t-1}^i,y_{\le t})\mathbf{I}\right)$$

- The KL divergence:

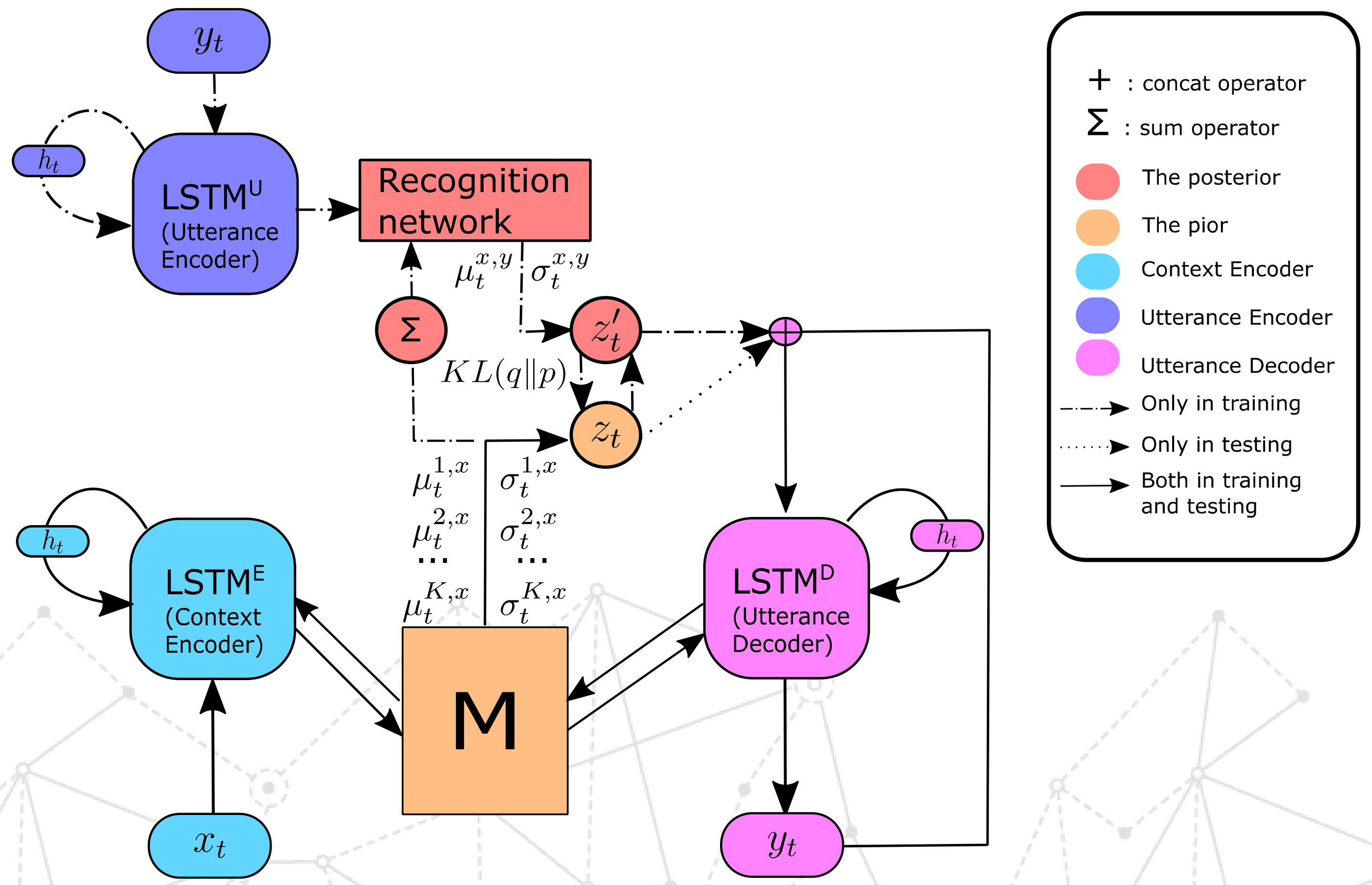
$$KL(q_\theta\|p_\phi) \le D_{var}(q_\theta\|p_\phi)$$

$$= \sum_{i=1}^{K} \pi_t^{i,x}(x,r_{t-1}^i)\exp\left(-\mathrm{KL}\left(\mathrm{N}(\mu_t^{i,x},\sigma_t^{i,x}\mathbf{I})\|\mathrm{N}(\mu_t^{x,y},\sigma_t^{x,y}\mathbf{I})\right)\right)$$

- Loss function:

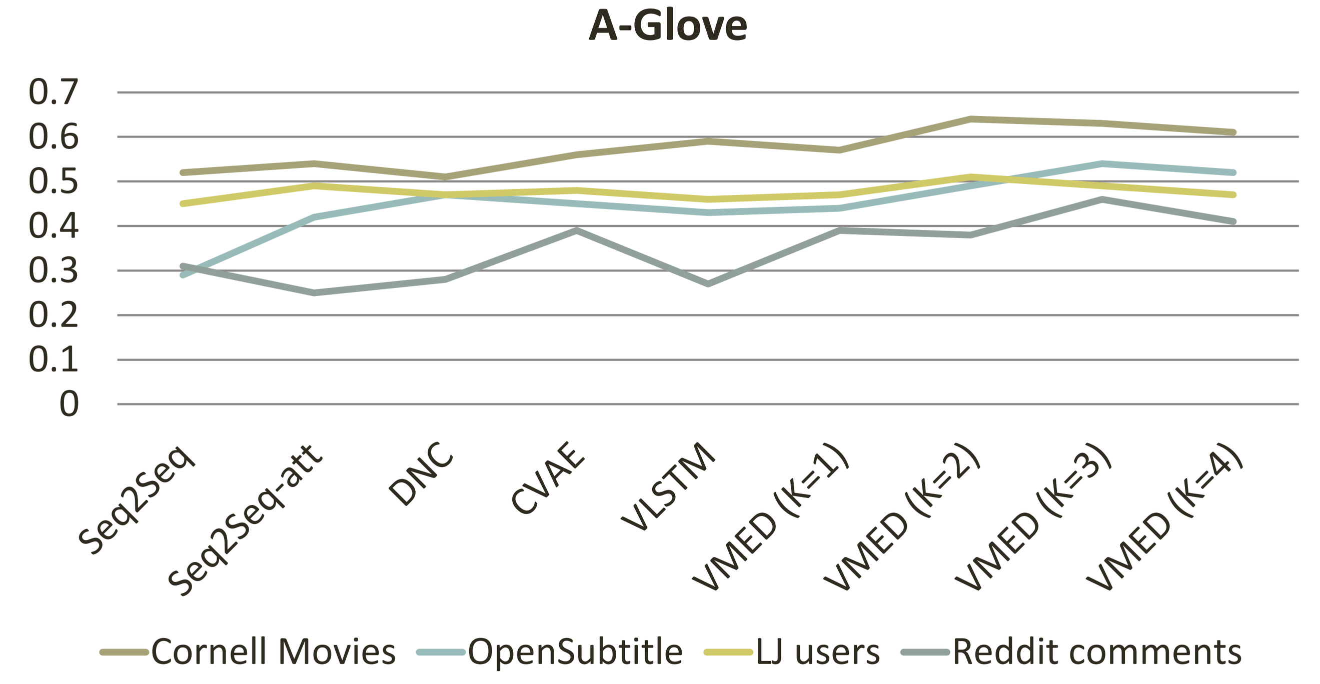
$$L = \sum_{t=1}^{T} log\left(D_{var}(q_\theta\|p_\phi)\right) + \sum_{t=1}^{T}\sum_{l=1}^{L} log\left(p\left(y_t|x,z_{\le t}^{(l)}\right)\right)$$

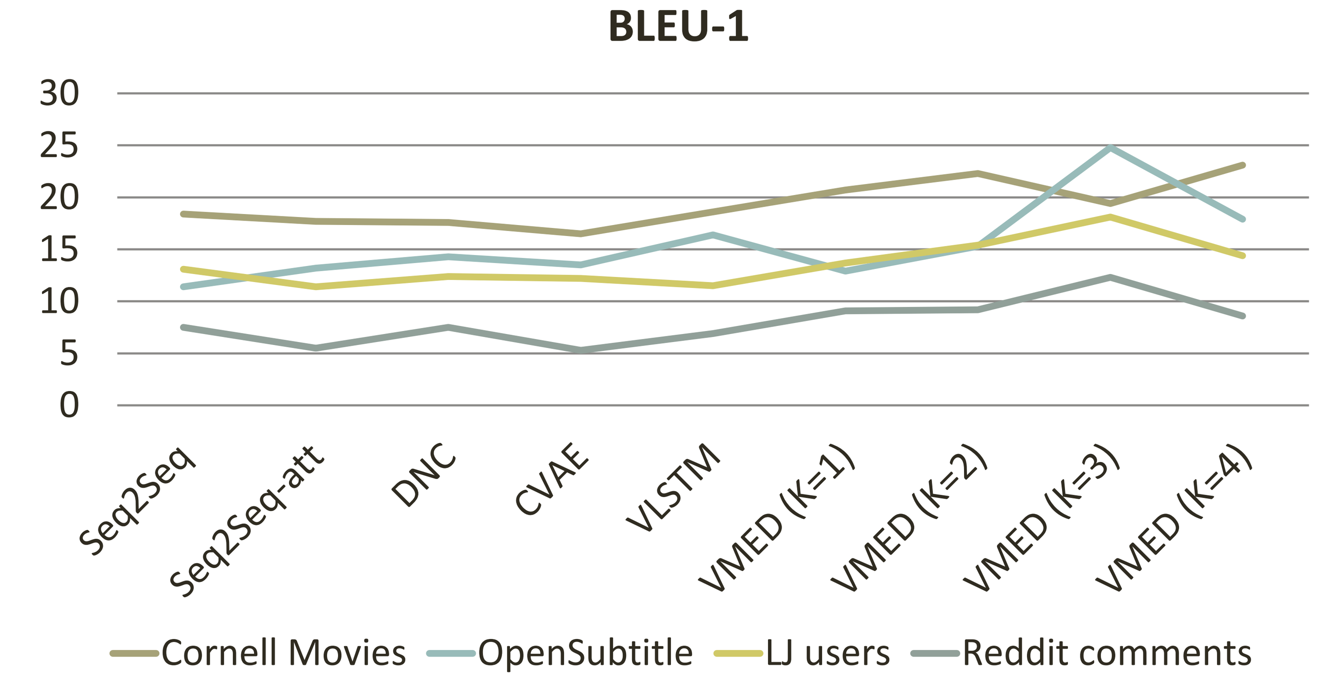
Training

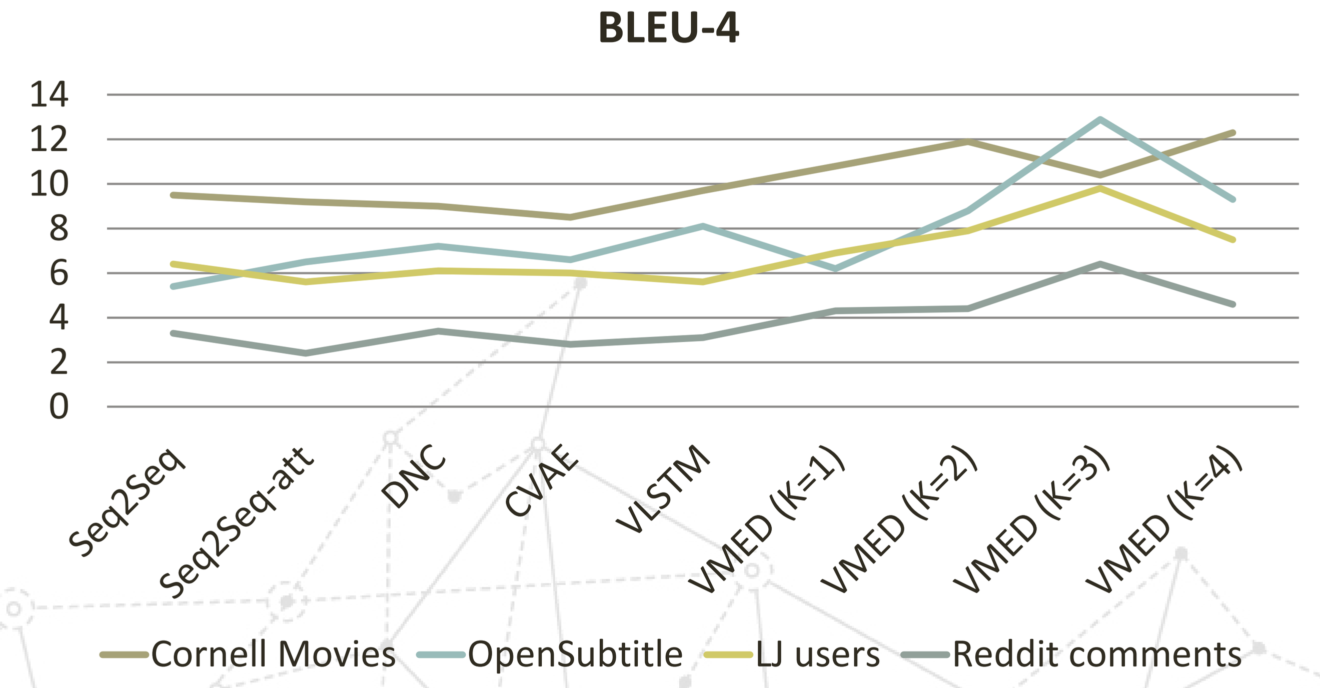


Training and testing of VMED

Results







Case study

Reddit comment: What actor will win an Oscar in the next 10 years ?

Seq2Seq: Colin
Seq2Seq-att: Liam Neeson
DNC: Tom Gyllenhaal
CVAE: Daryl and Aaron /*/ Carefully count Alfred Deniro /*/ Ponyo Joker possible
VLSTM: Michael Bullock /*/ Michael /*/ Michael De
VMED (K=3): Edward or Leo Dicaprio goes on /*/ Dicaprio will /*/ Dicaprio Tom has actually in jack on road

Reddit comment: What is your favorite scene in film history ? Mine is the restaurant scene in the Godfather.

Seq2Seq: The scene in
Seq2Seq-att: The final
DNC: The scene in
CVAE: Inception god! Not by a shark /*/ Amour great /*/ Pro thing you know 3 dead
VLSTM: The scene in /*/ The of a dead /*/ The sky in scene
VMED (K=3): The opening scene from history movie /*/ The scene in a shot nights! Robin movie /*/ The psycho scene in fight from

References

[1] Zhang et al., Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In Proceedings of the Annual Meeting of the Association for Computational Linguistics 2017

[2] Chung et al., A recurrent latent variable model for sequential data. In Advances in Neural Information Processing Systems, 2015