Exploring Story Generation with Multi-task Objectives in Variational Autoencoders

Zhuohan Xie, Trevor Cohn, Jey Han Lau
School of Computing and Information Systems
The University of Melbourne

THE UNIVERSITY OF MELBOURNE

Summary

- we combine BERT and GPT-2 to build domain-specific VAE for story generation
- we propose an approach to incorporate the latent variable into the VAE's decoder
- we introduce two auxiliary objectives to encourage the latent variable to capture topic information and discourse relations
- we experiment with several story datasets and show that our enhanced VAE produces higher quality latent variables and generates stories with better quality-diversity trade off compared to GPT-2

Problems

- Current pretrained languages can generate fluent sentences, but usually does not address the diversity issue
- VAE is able to generate diverse meaningful sequences with the power of a tractable latent space
- VAE models only memorise local information but suffer from loss of global features

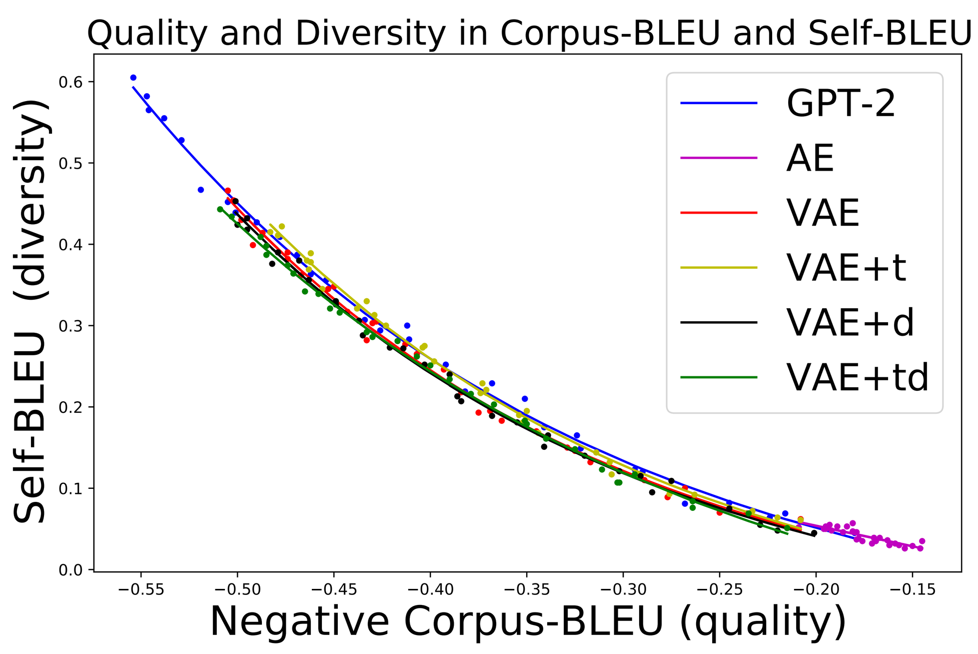
Dataset

Collection	Len	Train	Dev	Test
APNEWS	138	46.4K	1.9K	1.8K
Reuters	88	7.8K	2K	1K
ROC	60	88K	5K	2K
WP	110	2.95M	2K	2K

Table 1: Average length and number of documents in APNEWS, Reuters, ROC and WritingPrompts (WP) Dataset

Quality and Diversity Evaluation

- We adapt temperature sweep and use top-p sampling with varying p values
- The figure shows that the VAEs generally achieve a better trade off than fine-tuned GPT-2
- AE is not able to generate high quality stories under our tested p values and produces a curve near the bottom right corner

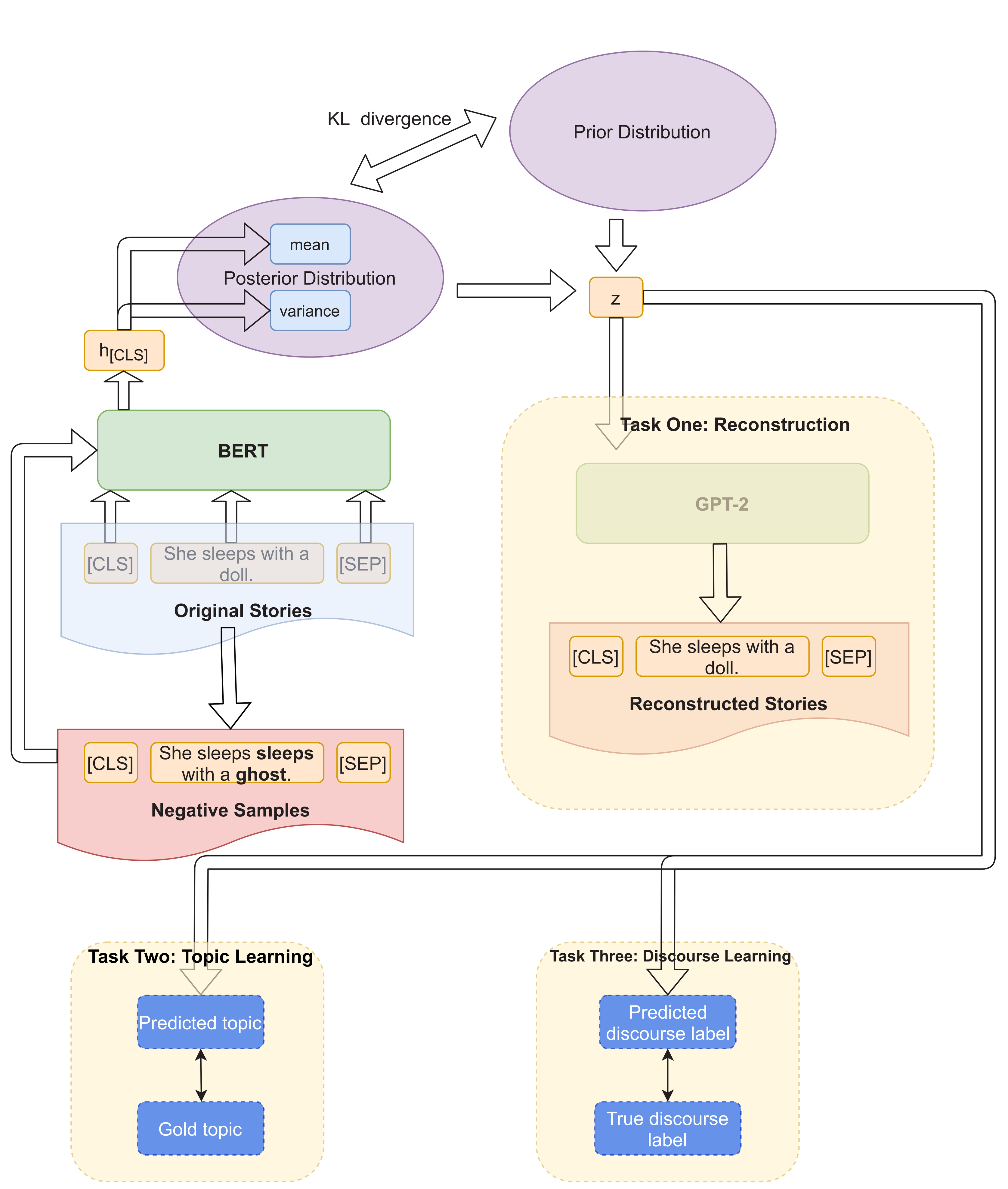


Topic Enhancement

Model	μ	z
AE	0.702	0.699
VAE	0.446	0.436
VAE+t	0.691	0.583

Table 3: Topic classification accuracy using mean of the posterior distribution μ and the latent variable z on Reuters.

Framework



The figure demonstrates our proposed VAE model.

- We use use the [CLS] token in BERT to represent the whole story and add two linear layers on top to compute the mean (μ) and standard deviation (σ) of the latent variable z
- To incorporate the latent variable z into the GPT-2 decoder, we append the latent variable as prefix token at the beginning of input sequence
- The two additional objectives train latent variable to predict the story topic and distinguish between negative samples vs. original stories

Global Features Enhancement Evaluation

- Table 2 presents the predicted discourse scores on a set of generated stories
- Stories with high discourse scores are generally coherent, while stories with low scores often have logical or repetition problems
- Table 3 shows topic classification accuracy using mean of the posterior distribution μ and the latent variable z
- Our topic-enhanced VAE is indeed able to capture much of the topic information, producing a better topic classification accuracy compared to vanilla VAE

Score	Story	Issue
0.83	[MALE] went fishing . he was excited about the trip . he saw a big fish . he was excited to get it . he caught a huge fish .	
0.81	[FEMALE] was nervous for her first day of school . she was nervous because she was so new to school . [FEMALE] was scared to be in the classroom . the teacher introduced her to other students . [FEMALE] was very excited to learn about her new class .	
0.40	[MALE] received a call from his boss . he had a promotion . **he took it . he took it anyway . he got it .**	repeat and incoherent
0.32	[MALE] grew up on a farm . [MALE] wanted to grow vegetables . **he was tired of them . [MALE] bought carrots .** he then grew vegetables .	incoherent

Table 2: Predicted discourse scores using the discourse-enhanced VAE.