

Pre-train and Plug-in: Flexible Conditional Text Generation with Variational Auto-Encoders

- 2020, Yu Duan - et.al

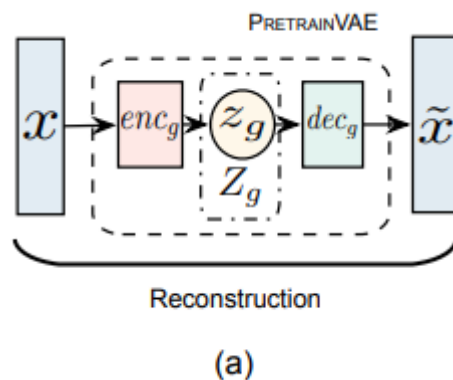
1. abstract

기존 conditional text generation의 경우, 새로운 condition이 주어지면 처음부터 다시 학습해야했음
그러나 우리는 Pre train 시킨 부분을 제외하고, 뒷 부분만 finetuning해서 빠르게 적용할 수 있다고 함

나의 주안점

Pretrain VAE 부분만 스윕 보자.

Encoder를 GRU로 구성하고, Decoder를 Transformer의 Encoder부분으로 구성



특이점이 posterior collapse를 방지하고자 WAE-GAN을 활용함

To address this, we utilize Wasserstein Autoencoder (WAE) (Tolstikhin et al., 2018) for PRETRAINVAE. Different from the original VAE, WAE encourages aggregated posterior distribution to be close to the prior, which is effective in alleviating the reconstruction problem of VAE (Tolstikhin et al., 2018). Specifically, we adopt WAE-GAN, a variant of WAE, which incorporates the merits of adversarial learning. During training, the encoder $enc_g(x) = q_g(z_g|x)$ encodes the text to the latent space and the decoder $dec_g(z_g) = p_g(x|z_g)$ reconstruct the text with the latent variable z_g . Thus, the loss function of PRETRAINVAE is formulated as:

$$\begin{aligned} \mathcal{L}_{\text{PRETRAINVAE}}(x) = & -\mathbb{E}_{q_g(z_g|x)}[\log p_g(x|z_g)] \\ & + \lambda D(Q(z_g), p(z_g)) \end{aligned} \quad (2)$$

where $Q(z_g) = \int q_g(z_g|x)p(x) dx$ is the aggregated posterior distribution; $p(z_g)$ is the prior normal distribution; D is the adversarial discriminator; λ is the coefficient hyper-parameter ($\lambda > 0$).

추가적으로 공부해 볼 만한 주제일 것 같음 - WAE-GAN

내 실험에 추가한 부분은 latent dim을 128로 크게 확장시켰고, 매 block마다 positional embedding을 더해줬다 (아무래도 시간의 의미를 더해주기 위해서 그런 것으로 사료됨)