Generative Models for Text

Agustinus Kristiadi

University of Bonn

12 Dec 2017

Outline

Introduction
Quick Introduction to Generative Models

Controlled Generation of Text
Introduction

Conclusion

Generative Model

- Recall Bayes' Rule:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

- Generative Model: modeling $P(X)$.
- Important models: GAN, VAE.
- In our work we will be based on VAE.

Autoencoder (AE)

- Neural nets that take input X and to reconstruct it, i.e. outputting $\hat{X}$

Variational Autoencoder (VAE)

- Constraint the hidden layer into some distribution (of latent variable).
- Force the hidden layer to match our prior distribution, e.g. standard normal.
- Objective: maximize (reconstruction + hidden unit regularization)

$$\max_{\theta} \mathcal{L}(\theta; X) = \mathbb{E}_{q(z|X)}[\log p(X|z)] - D_{KL}[q(z|X) \| p(z)]$$

- In practice: $q(z|X) = E(X)$ and $p(X|z) = D(z)$, where $E(X)$ and $D(z)$ are neural nets.

Benefits of VAE

- Latent variables nicely contained in $N(0, I) \implies$ can be nicely interpolated.
- Generating data X becomes possible: $z \sim N(0, I); X = D(z)$.

Toward Controlled Generation of Text (Hu, 2017)

- Extending VAE model
  - Use LSTM-RNNs as encoder and decoder.
  - Add another neural net to enforce conditional attribute constraint.
- Enables us to condition text generation.
  E.g. generate text with past tense and positive sentiment:
  "this was spectacular , i saw it in theaters twice".

Architecture

- We are optimizing:

$$\min_{\theta_G, \theta_E} \mathcal{L}_{VAE} + \lambda \mathcal{L}_{attr}$$

$$\min_{\theta_D} \mathcal{L}_D = \mathcal{L}_s + \lambda_u \mathcal{L}_u$$

where $\mathcal{L}_{attr}$ is loss function for conditional attribute constraint.

Algorithm

Example Expected Results

Conclusion

- VAE is a useful modification of original autoencoder.
- We can extend VAE to also learn conditional constraint.
- We can generate text with desired properties based on the conditional constraint.