Variational Neural Conversational Model

Chao-Ming Yen [*1] **Yikang Li** [*1] **Xupeng Tong** [*1]

Abstract

To be added...

1. Introduction

Conversation modeling is a famous task that allows machine to generate reasonable responses according to the sentence it is shown. Previously, fair amount of works have done.

In this project, we plan to improve the model performance based on previous works by incorporating latent information in the model by discovering several existing in variational methods. Especially, we are interested in RNN based variational autoencoder, that can seamlessly concatenate the seq2seq model with fine tuned regularizations.

To be added......

2. Related Works

2.1. Neural Conversational Model

Sequence To Sequence model is first introduced in (Cho et al., 2014), and since then, has become the standard model for dialogue systems and machine translation. It consists of two RNNs (Recurrent Neural Network) : An Encoder and a Decoder. The encoder takes a words sequence as input and processes one word at each time step. The objective is to convert symbol sequence into a fixed size feature vector that encodes the important information in the sequence while losing the redundant or unneccesary information.

(Vinyals & Le, 2015) (Cho et al., 2014)

2.2. Auto-Encoding Variational Bayes

(Kingma & Welling, 2013)

Given an observed variable x, VAE introduces a continuous latent variable z, and assumes that x is generated from z

$$p(x, z) = p(x|z)p(z)$$

The prior over the latent random variables, $p(z)$, is always chosen to be a simple Gaussian distribution and the conditional $p(x|z)$ is an arbitrary observation model whose parameters are computed by a parametric function of z. In VAE, $p(x|z)$ is typically parameterized function approximator such as a neural network. While latent random variable models of the form given in Eq. (3) are not uncommon, endowing the conditional $p(x|z)$ as a potentially highly non-linear mapping from z to x is a rather unique feature of the VAE.

The generative model $p(x|z)$ and inference model $q(z|x)$ are trained jointly by maximizing the variational lower bound with respect to their parameters, where the integral with respect to $q(z|x)$ is approximated stochastically. The gradient of this estimate can have a low variance estimate, by reparametrizing $z = \mu + \sigma \odot \epsilon$

$$ELBO_i(\lambda) = E_{q\lambda(z|x_i)}[\log p(x_i|z)] - KL(q\lambda(z|x_i)||p(z))$$

To be added and elaborated......

2.3. Variational Recurrent Neural Network

Earlier works in (Chung et al., 2015) introduced high-level random latent variables to recurrent neural network (RNN), empowering the model to be able to capture even higher variabilities sequential dataset such as natural speech. Differed from variational auto-encoders (VAE) used for the cases of non-sequential dataset, where latent random variables were designed to capture the variations in the observed variables. In VRNN, the recurrent network has a VAE for each time step, and these VAEs are conditioned on hidden state variable, such that

$$\mathbf{x_t}|\mathbf{z_t} \sim \mathcal{N}(\mu_{\mathbf{x,t}}, \mathbf{diag}(\sigma^2_{\mathbf{x,t}})), \mathbf{where}[\mu_{\mathbf{x,t}}, \sigma^2_{\mathbf{x,t}}] = \varphi^{\mathbf{dec}}_\tau(\varphi^{\mathbf{z}}_\tau(\mathbf{z_t}), \mathbf{h_{t-1}})$$

extract sequential features, and hidden states of RNN can be updated using recurrence equation

$$\mathbf{h_t} = \mathbf{f}_\theta(\varphi^{\mathbf{x}}_\tau(\mathbf{x_t}), \varphi^{\mathbf{z}}_\tau(\mathbf{z_t}), \mathbf{h_{t-1}})$$

[*]Equal contribution [1]Carnegie Mellon University, USA. Correspondence to: Xupeng Tong <xtong@andrew.cmu.edu>.

Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 2017. JMLR: W&CP.

2.4. Generative Adversarial Network

(Goodfellow et al., 2014)

(Li et al., 2017)

Generative adversarial network is a framework containing a generative network and a discriminative network. The idea of GAN comes from zero-sum game in game theory, where two players (networks) compete against each other. The generative network is taught to map from a latent space to a particular data distribution of interest, and the discriminative network is simultaneously taught to discriminate between samples from the true data distribution and synthesized samples produced by the generator.

In order to learn generator's distribution p_g over data x, GAN introduces a prior noise on input noise variables $p_z(z)$, then represent a mapping to data space as $G(z;\theta_g)$ and denote the probability that samples come from real data instead of generator as $D(x)$. The discriminator is trained to maximize the probability of assigning the correct label to both real samples and generated samples. i.e. $\log D(x)$. The generator is trained simultaneously to maximize discriminator's error, or equally minimize $\log(1 - D(G(z)))$. Overall, we are optimizing the adversarial loss:

$$
\begin{aligned}
&\min_G \max_D V(D, G) \\
=&E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))]
\end{aligned}
$$

Further, the author shows that with purely back-propagation, the algorithm can achieve global optimality, which means p_g converges to p_{data}.

The idea of GAN has enjoyed great success in computer vision in terms of generating images that look authentic to human observers. What's more, recent researcher also apply GAN to the field of dialogue generation. They first pre-train the generative model by predicting target sequences given the conversation history using a SEQ2SEQ model with attention mechanism. They also pre-train the discriminator and conduct data processing to improve response quality. In addition to adversarial training, they also proposed a model for adversarial evaluation that uses success in fooling an adversary as a dialogue evaluation metric.

3. Datasets

We will tested our model on the OpenSubtitles dataset (Tiedemann, 2009). This is a dataset that incorporated movie subtitles with sentences uttered by characters. The model will be trained to predict the next sentence given the previous one, for every consecutive sentences, and each sentence will be used both for context and as target.

4. Plan

4.1. Incorporating latent variables in the training of Seq2Seq model

There are various ways that allow us to improve our model with seq2seq

Following the work in (Zhang et al., 2016), which introduces a variational model for neural machine translation that incorporates a continuous latent variable z to model the underlying semantics of sentence pairs, we can also apply it to our neural conversation model that uses the same seq2seq model.

We can also apply the method proposed by (Chung et al., 2015), that explicitly models the dependencies between latent random variables across subsequent timesteps.

4.2. Incorporating latent information unsupervisely as the input to Seq2Seq model

Since sometimes, incorporating the latent variable into the training process directly may be hard. We consider an alternative approach that, instead of learn the latent variable through an end to end one way pass method. We can train a Variational Recurrent Auto-Encoder (Fabius & van Amersfoort, 2014) for each sentences first. VRAE is a variational autoencoder that can be used for the unsupervised learning on time series data, mapping the time series data to a latent vector representation.

By appending the latent vector representation of each sentences along with the vector encoded by the seq2seq encoder, we naturally incorporate latent information of the sentence and that could serve as the input to be fed into the decoder phase.

4.3. Improving the attention alignment model by variational inference

One potential issue with this seq2seq model is that a neural network needs to be able to compress all the necessary information of a source sentence into a fixed-length vector. To allow the decoder access to the input more directly, an attention mechanism was introduced in (Bahdanau et al., 2014).

The affect of the alignment model has become one of the most important feature of state-of-art sequence to sequence models. By incorporating latent variables in this particular part through variational method may gives us a boost in the model performance.

4.4. Adversarial Variational Inference

Improving the above mentioned variational method with Generic Adversarial Network (GAN) as proposed in

(Mescheder et al., 2017)

References

Bahdanau, Dzmitry, Cho, Kyunghyun, and Bengio, Yoshua. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

Cho, Kyunghyun, Van Merriënboer, Bart, Gulcehre, Caglar, Bahdanau, Dzmitry, Bougares, Fethi, Schwenk, Holger, and Bengio, Yoshua. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

Chung, Junyoung, Kastner, Kyle, Dinh, Laurent, Goel, Kratarth, Courville, Aaron C, and Bengio, Yoshua. A recurrent latent variable model for sequential data. In *Advances in neural information processing systems*, pp. 2980–2988, 2015.

Fabius, Otto and van Amersfoort, Joost R. Variational recurrent auto-encoders. *arXiv preprint arXiv:1412.6581*, 2014.

Goodfellow, Ian, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron, and Bengio, Yoshua. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.

Kingma, Diederik P and Welling, Max. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Li, Jiwei, Monroe, Will, Shi, Tianlin, Ritter, Alan, and Jurafsky, Dan. Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547*, 2017.

Mescheder, Lars, Nowozin, Sebastian, and Geiger, Andreas. Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks. *arXiv preprint arXiv:1701.04722*, 2017.

Tiedemann, Jörg. News from opus-a collection of multilingual parallel corpora with tools and interfaces. In *Recent advances in natural language processing*, volume 5, pp. 237–248, 2009.

Vinyals, Oriol and Le, Quoc. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

Zhang, Biao, Xiong, Deyi, Su, Jinsong, Duan, Hong, and Zhang, Min. Variational neural machine translation. *arXiv preprint arXiv:1605.07869*, 2016.