Variational Neural Conversational Model

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

Abstract

Dialogue generation or conversation model has been offered with great promise in the recent year thanks to the development of sequence to sequence model proposed in 2014. Sequence to sequence model can be seen as a special member of the encoder-decoders family that utilizes RNN (recurrent neural network) to learn the conditional distribution of a target sentence given a source sentence with end to end optimization. Variational autoencoder (VAE) is a very promising model that neatly combines the strength of deep learning and variational Bayesian methods with reparameterization and well crafted objective function for optimization. In this project, we want to explore the strength of both the seq2seq model and variational methods including variational autoencoder in the application of dialogue generation task. We are also interested in incorporating the generic adversarial network (GAN) to our model enlightened by the recent research in combining VAE with GAN. Finally, we will discuss the attention mechanism associate with seq2seq model and ways we can improve it.

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 (VAE), that can seamlessly concatenate the seq2seq model with fine tuned regularizations.

In this proposal, we will first introduce some related works in the recent years from sequence to sequence model to the well known variational autoencoder. How do people bridge them, and what is the existing methods we can do that? Beyond that, we've also covered the review in generative adversarial network (GAN) and its application in improving the performance of variational autoencoder from the recent progress.

In the plan part, we propose four potential approaches through four different perspectives that might potentially improve the task in dialogue generation.

Firstly, can we use use the existing recurrent variational method (Chung et al., 2015) in training of our seq2seq model? Or can we borrow the ideas from the works done in machine translation (Zhang et al., 2016) and see how they deal with the variational seq2seq? Secondly, can we unsupervisely learn a dense vector with sequence input (Fabius & van Amersfoort, 2014) and use that encoded vector, along with the vector encoded by seq2seq model, to the decoder of seq2seq model simultaneously? Thirdly, can we further improve the attention mechanism with variational inference? Lastly, can we use adversarial training of our variational model with the recent progress on that? Our project will basically around answering these four questions and explore the best potential of variational method in neural conversation model.

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 (Vinyals & Le, 2015) 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.

Neural Conversational Model has been greatly improved a lot recently along with the development of seq2seq model,

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

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

thanks to the attention mechanism (Bahdanau et al., 2014). Instead of generating each word in the target sequence all depending on a single vector encoded by the encoder, attention mechanism introduces an alignment model so each word in the target sequence can be produced by the linear combination of all the intermediate output in the encoding phase.

2.2. Auto-Encoding Variational Bayes

Variational autoencoder (VAE) (Kingma & Welling, 2013) has successfully injected the probabilistic flavor in the basic autoencoder by reparameterization and reconstruction of the outputs as probabilistic random variables within a model and approximate objective function that can conducted end to end training.

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, $p(x|z)$ plays a role as parameterized function approximator (neural network). 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 $z = \mu + \sigma \odot \epsilon$

We can formulate the above problem as minimizing the KL divergence of these two distributions, however it is generally hard to actually compute it. Alternatively, VAE chooses to optimize some thing that is equivalent to the KL up to an added constant,

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

called Evidence Lower BOund (ELBO).

With the perspective from bayesian statistics, the encoder becomes a variational inference network, mapping observed inputs to its approximate posterior distributions over the latent space, while the decoder works as a generative network that maps arbitrary latent coordinates back to distributions over the original space.

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

$$\mathbf{x_t}|\mathbf{z_t} \sim \mathcal{N}(\mu_{\mathbf{x,t}}, \mathbf{diag}(\sigma^{\mathbf{2}}_{\mathbf{x,t}})$$

where,

$$[\mu_{\mathbf{x,t}}, \sigma^{\mathbf{2}}_{\mathbf{x,t}}] = \varphi^{\mathbf{dec}}_{\tau}(\varphi^{\mathbf{z}}_{\tau}(\mathbf{z_t}), \mathbf{h_{t-1}})$$

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

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

This leads to parameterized generative model with the factorization

$$p(\mathbf{x_{\leq T}}, \mathbf{z_{\leq T}}) = \prod_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{p}(\mathbf{x_t}|\mathbf{z_{\leq t}}, \mathbf{x_{<t}})\mathbf{p}(\mathbf{z_t}|\mathbf{x_{<t}}, \mathbf{z_{<t}})$$

and the factorization of inference model as

$$q(\mathbf{z_{\leq T}}|\mathbf{x_{\leq T}}) = \prod_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{q}(\mathbf{z_t}|\mathbf{x_{\leq t}}, \mathbf{z_{<t}})$$

, both factorization equations are used in learning phase to obtain the timestep-wise variational lower bound by the likelihood approach.

In these works, the application of VRNN to natural speech generation and handwriting generation demonstrated a significant performance boost compared to the results from traditional RNN, based upon experiments over well-known datasets such as Blizzard, TIMIT and Accent. This inspired us about the possibility that if same random latent variable can be introduced the enhance RNN when it comes to handle chat generating tasks.

2.4. Generative Adversarial Network

Generative adversarial network (Goodfellow et al., 2014) 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, the adversarial loss we are optimizing could be wrote as ,

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

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.

Recently, works like adversarial autoencoders (AAE) (Mescheder et al., 2017) integrates the power of GAN with another famous generative model variational autoencoder (VAE) to perform variational inference by matching the aggregated posterior of the hidden code vector of the autoencoder with an arbitrary prior distribution, have shown a very interesting and promising aspect in this model in its application to the variational inference .

3. Datasets

We will test our model on the OpenSubtitles dataset (Tiedemann, 2009). This dataset has included movie subtitles with sentences uttered by characters, since the taking of the characters in the dataset is not so clear, every two consecutive sentences will be treated as training data we have. Our model will be trained to predict the next sentence given the previous one, for every sentence pairs in the training data, so each sentence will be used both for context and as target.

The performance of dialogue generation will be scored by BLEU by the sentences generated on the testing data.

4. Plan

4.1. Incorporating latent variables in the training of Seq2Seq model

Since variational inference can model complex multimodal distributions, and the underlying true data distribution of natural language itself might consists of multimodal conditional distributions, it is natural to extend models in RNN with variational inference.

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.

Besides the work mentioned above, variational recurrent neural network (VRNN) is probabily the first approach in capturing the latent temporal dependencies of RNN using variational inference (Chung et al., 2015). The model explicitly models the dependencies between latent random variables across subsequent timesteps. To the best of our knowledge, VRNN have not been applied to the seq2seq model.

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 features 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. Incorporating variational inference with a well defined end-to-end model is generally hard, coordinate descent that bridges the end to end training and the variational inference might be one possible approach that we will try giving the attention more probabilistic sensing.

4.4. Adversarial Variational Inference

Improving the above mentioned variational method with GAN is generally a very interesting approach as proposed in (Mescheder et al., 2017). After all the experiment if we have succeeded as mention above, we will make a bold step in the adversarial training of our model.

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.

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.