Comparison Between Variational Autoencoder and Encoder-Decoder Models for Short Conversation

Shin Asakawa  
Center for Information Sciences, Tokyo Women’s Christina University, Zempukuji, 2, 6, 1  
Nerima, Tokyo 1678585, Japan

Takashi Ogata  
Faculty of Software and Information Science, Iwate Prefectural University  
Takizawa, Iwate, 020-0693, Japan  
E-mail: asakawa@ieee.org, t-ogata@iwate-pu.ac.jp

Abstract

We provide a point of view concerning generative models such that they could deal with short conversation. These include the standard recurrent neural network language, sequence to sequence, vector embedding, and variational autoencoder models. These models seem to be possible candidates to describe such conversations, there are several differences among them.

Keywords: neural network language model, recurrent neural networks, vector embedding, variational autoencoder, sequential data processing, variational inference, narrative generation.

1. Introduction

Conversation includes understanding and generation. Therefore, modeling of conversation play an important role in natural language processing (NLP) and machine intelligence. Although previous approaches exist, they are often restricted to specific domains (e.g., booking an airline ticket) and require hand-crafted rules. In this paper, we will provide a brief survey about recent progress concerning sentence generation. First the recurrent neural network language models (RNNLM) will be introduced (section 2). Then we will make brief explanations about recent models: vector embedding models (section 3), variational autoencoder (VAE) models (section 4), and combining models between RNNLM and VAE. In section 6, we will discuss possible contributions of the models cited here to narrative studies.

2. Recurrent Neural Network Models

RNNLMs (RNNLMs) represent the state of the art in unsupervised generative modeling for NLP modeling. In supervised settings, RNNLMs decode conditioned on task specific features in machine translation1,11. Standard RNNLMs can predict each word of a sentence conditioned on the previous word and an evolving hidden state. RNNLMs can generate sentences word-by-word based on an evolving distributed state representation. RNNLMs can make their internal representation probabilistic models without any significant independent assumptions. However, by breaking the model structure down into a series of next-step predictions, the RNNLM does not expose an interpretable representation of global features like topic or of high-level syntactic properties. Vinyals and Le12 showed RNNLM (Recurrent Neural Neural Network Language Model) could deal with conversation (Fig. 1).

© The 2017 International Conference on Artificial Life and Robotics (ICAROB 2017), Jan. 19-22, Sapporo Convention Center, Miyazaki, Japan
4. Variational autoencoder models

The variational autoencoder (VAE) is a generative model that is based on a regularized version of the standard autoencoder. This model imposes a prior distribution on the hidden codes \( z \) which enforces a regular geometry over the codes and makes it possible to draw proper samples from the model using ancestral sampling. The VAE models the autoencoder architecture by replacing the deterministic function \( f(x) \) with a learned posterior recognition model, \( q(z|x) \). This model parametrizes an approximate posterior distribution over \( z \) (they employed a diagonal Gaussian) with a neural network conditioned on \( x \). The VAE learns sentences not as single points, but as soft ellipsoidal regions in latent space, forcing the sentences to latent the space rather than memorizing the training data as isolated sentences. If the VAE were trained with a standard autoencoder's reconstruction objective, it would learn to encode its inputs deterministically by making the variances in \( q(z|x) \) vanishingly small. Instead, the VAE uses an objective which encourages the model to keep its posterior distributions close to a prior \( p(z) \), generally a standard Gaussian \( (\mathfrak{N} = 0; \sigma^2 = 1) \). Additionally, this objective is a valid lower bound on the true log likelihood of the data, making the VAE a generative model. This objective takes the following form:

\[
L(\theta; x) = -\text{KL}(q(\mathfrak{N} = 0; \sigma^2 = 1) || p(z)) + \mathbb{E}_{q(z|x)}[\log p_\theta(x|z)] \\
\leq \log p(x). \tag{1}
\]

This forces the model to be able to decode plausible sentences from every point in the latent space that has a reasonable probability under the prior.

It was used diagonal Gaussians for the prior and posterior distributions \( p(z) \) and \( q(z|x) \), using the Gaussian reparameterization trick. The model was trained with stochastic gradient descent, and at each gradient step they estimated the reconstruction cost using a single sample from \( q(z|x) \), but compute the KL divergence term of the cost function in closed form.

5. Combining model between RNNLM and GAE

A variational autoencoder model was proposed Ref. 2. It was found that on a standard language modeling evaluation where a global variable was not explicitly needed, their model yielded similar performance to
existing RNNLMs. The model was also evaluated using a larger corpus on the task of imputing missing words. For this task, it was introduced a novel evaluation strategy using an adversarial classifier, sidestepping the issue of intractable likelihood computations by drawing inspiration from work on non-parametric two-sample tests and adversarial training. In this setting, the model's global latent variable allowed to do well where simpler models fail. In this model several qualitative techniques was introduced for analyzing the ability of their model to learn high level features of sentences. They could produce diverse, coherent sentences through purely deterministic decoding and that they could interpolate smoothly between sentences.

Fig 2. A VAE models from Ref. 2

The model of Ref. 2 allows it to explicitly model holistic properties of sentences such as:

1. style,
2. topic
3. high-level syntactic features

Samples from the prior over these sentence representations remarkably produce diverse and well-formed sentences through simple deterministic decoding. By examining paths through this latent space, we are able to generate coherent novel sentences that interpolate between known sentences. In order to solve the difficult learning problem, the method implemented in this model to demonstrate its effectiveness in imputing missing words, explore many interesting properties of the model's latent sentence space, and present negative results on the use of the model in language modeling.

A class of deep, directed generative models with Gaussian latent variables at each layer10. In order to allow for efficient and tractable inference, they introduced an approximate representation of the posterior over the latent variables using a recognition model that acts as a stochastic encoder of the data. For the generative model, they derive the objective function for optimization using variational principles; for the recognition model, they specify its structure and regularization by exploiting recent advances in deep learning. Using this construction, the model could learn the data by the gradient backpropagation that allowed for optimization of the parameters of both the generative and recognition models jointly.

Fig. 4 shows the corresponding computational graph of Fig. 3. Black arrows indicate the forward pass of sampling from the recognition and generative models. Solid lines indicate propagation of deterministic activations, dotted lines indicate propagation of samples. Red arrows indicate the backward pass for gradient computation. Solid lines indicate paths where deterministic backpropagation is used, dashed arrows indicate stochastic backpropagation.

Fig. 3. A plate notation of a variational RNN model

Fig. 4. A computational graph corresponding to Fig. 3

6. Discussion

So far, we tried to compare several methods proposed recently. Studies cited here might suggest a trends toward to generate sentences or topics with computational accountabilities with both RNNLMs and VAEs.

Additionally, the authors also aim at the application to narrative generation7. Conversation in narratives, such as folktales and novels, is used for a scene’s description where several characters appear in many cases. There are
novels that conversation is the central part in narrative progression, such as many novels by F. Dostoyevsky, E. Hemingway, and Y. Kawabata. Main effects of conversation of narratives are, for example, “realistic effect” (Hemingway and Kawabata) and “dramatic effect” (Dostoyevsky). Although the current version of an “Integrated Narrative Generation System (INGS)” that have been developing has not conversational part and the main narratological element is “event”, conversation, in addition to description and explanation, is a necessary and important element for completing a narrative’s structure and representation. Furthermore, although the INGS is currently an approach to narrative generation based on semiotic processing and knowledge, studying paternal processing approaches by neural networks stated in this paper is an important direction for aiming at the fluency of text generation or representation and the acquisition of diverse narrative knowledge including episodes and scripts.

7. Conclusion

We introduced computational explanations for sentence generations and showed relations between such studies based on neural networks and narrative studies. Despite many studies should be required, it seems to worth trying to go forward along these ways that we employed here. The RNNLMs and VAEs might converge in some ways in the near future, then the results that these method might bring us further steps.

References