

Generating tweets by a VAE model

Student: Elena Voskoboy

SA: D.Sc. A.Savostyanov, M.Sc. I.Bondarenko,

A Hybrid Convolutional Variational Autoencoder for Text Generation

19.03.2020

- 1 **Convolutional Encoder** block encodes a sequence of binary input vectors to produce a sequence of binary output vectors. This block can process multiple symbols at a time.
- 2 **KL** (Kullback–Leibler divergence term, also called relative entropy) is a measure of how one probability distribution is different from a second, reference probability distribution.

VA modifies the conventional autoencoder framework in two key ways:

- 1 Firstly, a deterministic internal representation z of an input x is replaced with a posterior distribution $q(z|x)$.
- 2 Secondly, the posterior $q(z|x)$ is regularized with its KL divergence from a prior distribution $p(z)$.

The KL regularizer:

$$J_{vae} = KL(q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) - \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[\log p(\mathbf{x}|\mathbf{z})]$$

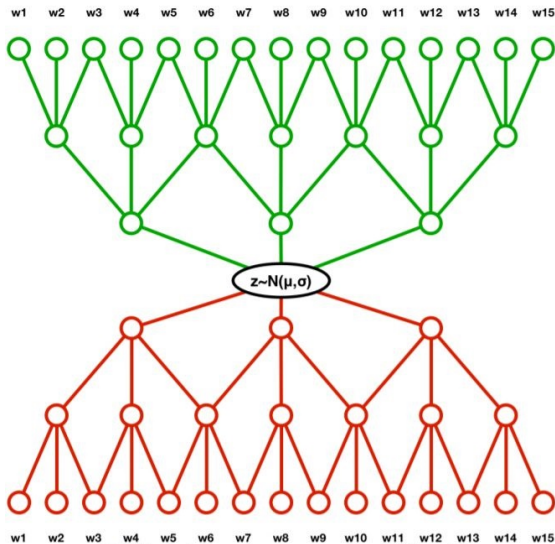
There are two primary motivations for choosing deconvolutional layers:

- ① Such layers have extremely efficient GPU implementations due to their fully parallel structure.
- ② Feed-forward architectures are easier to optimize, as the number of back-propagation steps is constant and potentially much smaller than in RNNs.

The model is composed of two relatively independent modules:

- 1 The first component is a standard VAE where the encoder and decoder modules are parametrized by convolutional and deconvolutional layers respectively.
- 2 The second component is a recurrent language model consuming activations from the deconvolutional decoder concatenated with the previous output characters. We consider a conventional LSTM network as recurrent function.

Model architecture



(a) Fully feed-forward component of our VAE model

- **Training data:** The standard Penn Treebank dataset. For training, fixed-size data samples are selected from random positions in the standard training and validation sets.
- **Test data:** A random sample tweets collected using the Twitter API to train the model and test it on a held out dataset of 10k samples.

- 5 convolutional layers with the ReLU non-linearity, kernel size 3 and stride 2 in the encoder.
- The number of feature maps is [128, 256, 512, 512, 512] for each layer respectively.
- The top layer is an LSTM with 1000 units. The baseline LSTM VAE model contained two distinct LSTMs both with 1000 cells.
- The model has comparable number of parameters: 10.5M for the LSTM VAE model.

@userid @userid @userid @userid @userid
@userid thanks for the follow
@userid @userid @userid @userid @userid
@userid @userid @userid @userid @userid
@userid thanks for the follow

@userid All the best!!
@userid you should come to my house tomorrow
I wanna go to the gym and I want to go to the beach
@userid and it's a great place
@userid I hope you're feeling better

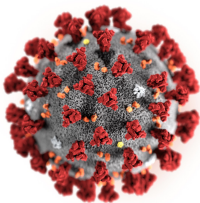
Examples of randomly generated tweets by a VAE model with a close to zero (top section) and larger than zero (bottom) KL term values.

@userid @userid @userid @userid @userid ...
I want to see you so much @userid #FollowMeCam ...
@userid @userid @userid @userid @userid ...
Why do I start the day today?

@userid thanks for the follow back
no matter what I'm doing with my friends they are so cute
@userid and I have to do that for a couple of days and then I can start with them
I wanna go to the UK tomorrow!! #feelinggood #selfie #instago
@userid @userid I'll come to the same time and it was a good day too xx

Random sample tweets generated by LSTM VAE
(top) and our Hybrid model (bottom).

Thank you for attention!



Take care of yourself!