Generating Sentences from a Continuous Space

1 Main Idea

The paper presents an alternative strategy to language modeling using RNNs. The authors attempt to use this to impute missing words in a sentence, to interpolate between the latent representations of 2 sentences, and also generate sentences by sampling from the space of the latent representation's prior probability. [1]

2 Background

- The authors differentiate between types of sentence embeddings. **Sequence Autoencoders** have RNNs as encoders and decoders are are just used to regenerate the original text. **Skip-Thought models** are similar but the target sentence is different from the original sentence. **Paragraph Vector models** simply predict the words that are present in a given sentence.
- Variational autoencoders impose a prior distribution on the latent representation, but a standard autoencoder does not.
- The latent representation is isuualy parametrized by a diagonal Gaussian distribution.

3 Method

- The prior distribution of the latent representation acts as a regularizer.
- Unclear as to what a 'global latent representation' is. Intuitively, each sentence would have its own representation.
- The authors suggest KL-term annealing, which involves having a cost function like

$$L(\theta; x) = \alpha(-KL(q_\theta(z|x)||p(z))) + E_{q_\theta(z|x)}[\log p_\theta(x|z)]$$

where the value of α is raised from 0 to 1 during the course of the training. This can be thought of as a steady progression from a standard autoencoder to a VAE.

- Word-level dropout used to force the decoder to rely primarily on the latent space for the sentence generation.

4 Observations

- The paper doesn't talk about the presence of dead-zones in the latent space. This should be more of a problem due to the discrete nature of word representation.

References

[1] Samuel R Bowman, Luke Vilnis, Oriol Vinyals, Andrew M Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. *CoNLL 2016*, page 10, 2016.