Generating Sentences from a Continuous Space

Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz and Samy Bengio

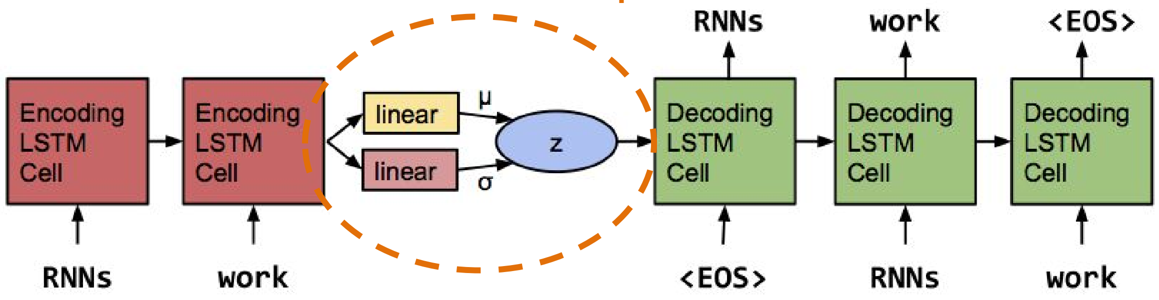
Presenter: Shuhei Iitsuka

Introduction

Variational Autoencoder for Natural Language Sentences.

- Recurrent Neural Network Language Model (RNNLM): the state-of-the-art generative model for natural language.
- Drawback: word-by-word generation ---> cannot capture the global characteristics of a **sentence**.
- Brings the idea of Variational Autoencoder (VAE) into natural language sentences.

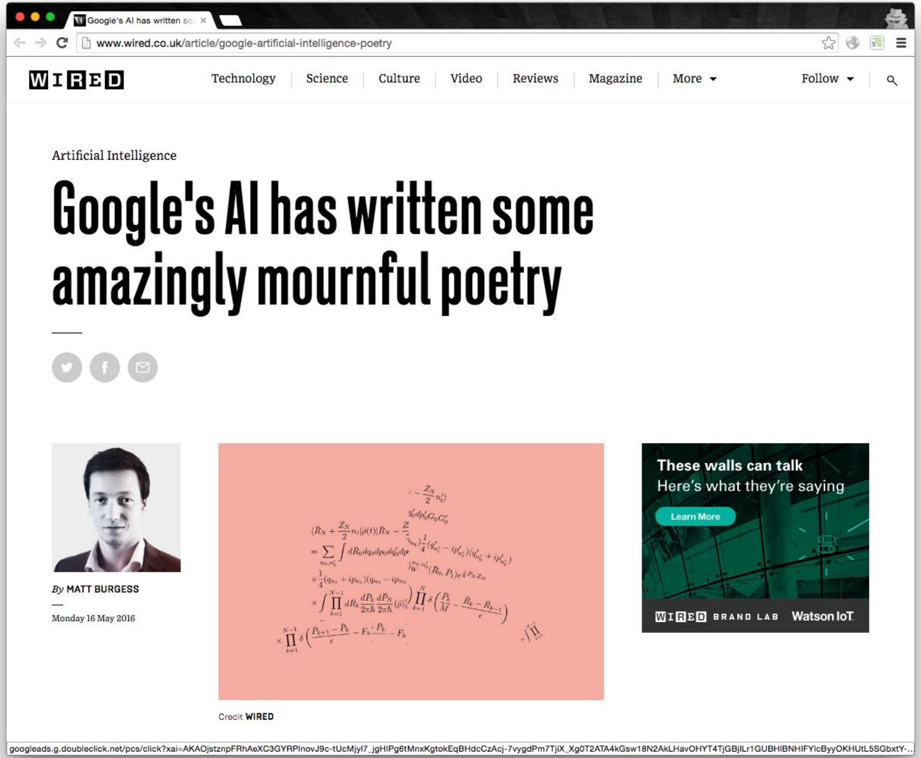
Not a representation vector,
but a continuous latent space.



Introduction: Contribution

- Proposed the variational autoencoder architecture for text sentence.
- Evaluated the performance on the language modeling task and the missing word imputing task comparing to RNNLM.
- Conducted qualitative / quantitative analysis on the proposed model.

Google Poetry



"there is no one else in the world.
there is no one else in sight.
they were the only ones who mattered.
they were the only ones left.
he had to be with me.
she had to be with him.
i had to do this.
i wanted to kill him.
i started to cry.
i turned to him. --"

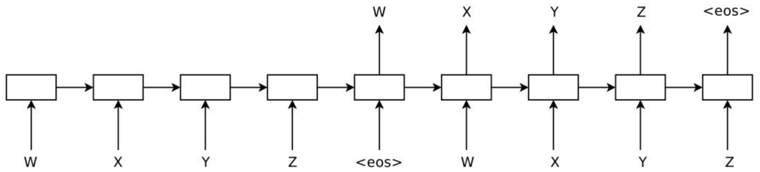
Agenda

- Introduction
- Background
  - Mapping a sentence to a latent space
  - Variational Autoencoder
- VAE for sentences
- Results
  - Language modeling
  - Imputing missing words
- Analysis
  - Dropout effect
  - Sampling from the posterior
  - Homotopies
- Conclusion

Background: Mapping a sentence to a latent space

Sequence Autoencoders

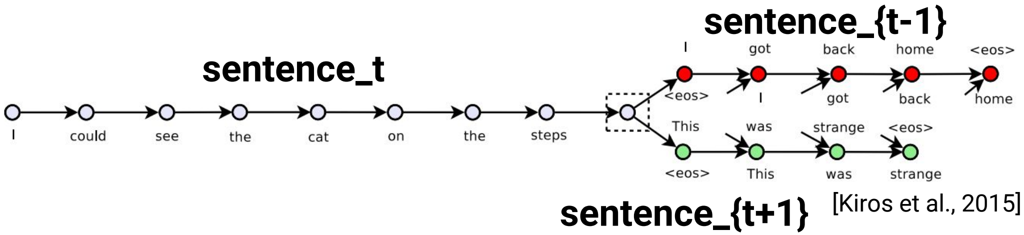
- Using seq2seq architecture as an autoencoder.
- Both encoder and decoder are RNNs.



[Dai and Le, 2015]

Skip Thought Vector

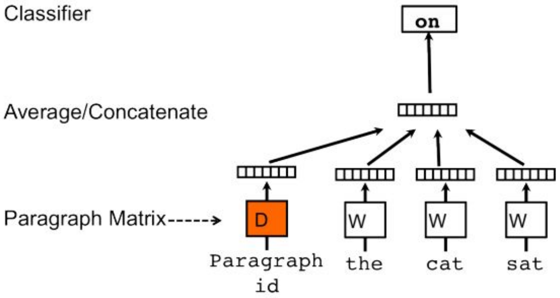
- One encoder and two decoder to predict previous sentence and next sentence.



[Kiros et al., 2015]

Paragraph Vector

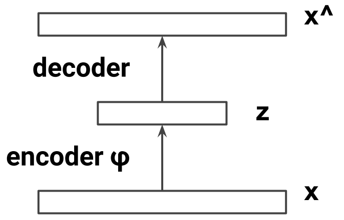
- non-RNN model.
- Paragraph matrix D → sentence, Vocabulary matrix W → word.
- Trains W to predict next words using D as contexts / memory.



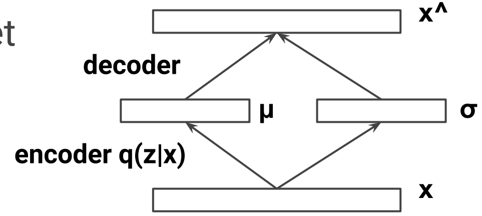
[Le and Mikolov, 2014]

Background: Variational autoencoder

- A regularized version of the standard autoencoder.
- Learns inputs not as single points, but as **regions** in the latent space.
  - Replace the deterministic encoder φ_{enc} with a probability distribution, $q(z|x)$.
- KL divergence is introduced to make $q(z|x)$ close to a prior $p(z)$.
  - If learned with just a reconstruction loss, variances in $q(z|x)$ get extremely small (=$q(z|x)$ becomes deterministic).



Standard autoencoder

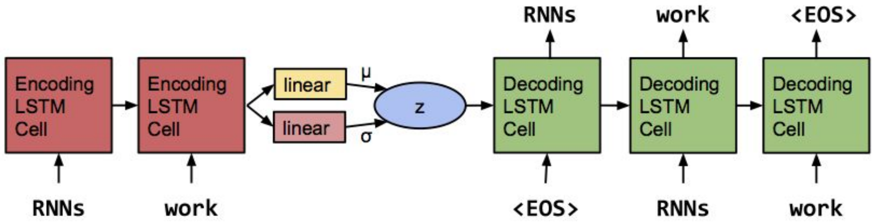


Variational autoencoder

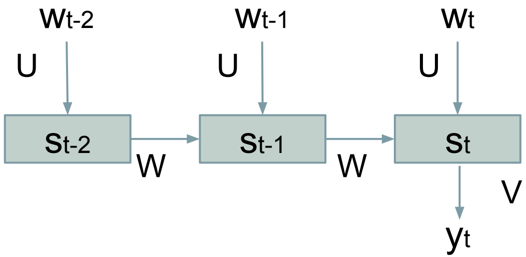
[Presenter Illustrated]

VAE for sentences

- Authors introduce the VAE for text which uses single-layer LSTM RNNs for both the encoder and decoder.
- The decoder can be regarded as a special version of RNNLM conditioned by the hidden code.
- Explored some variations → no significant improvements.
    - Concatenating z every time step.
    - Soft-plus activation
    - Feedforward between (encoder, latent variable) and (decoder, latent variable). (?)



Variational Sequence Autoencoder



RNNLM [Presenter Illustrated]

VAE for sentences: similar models

- Variational Recurrent Autoencoder (VRAE) [Fabious and van Amersfoort, 2014]
- Continuous latent variables with RNN-style modeling [Bayer and Osendorfer, 2015][Chung et al., 2015]

→ **Latent variables are separated per timestep.**
→ Not suitable to capture the general sentence characteristics.

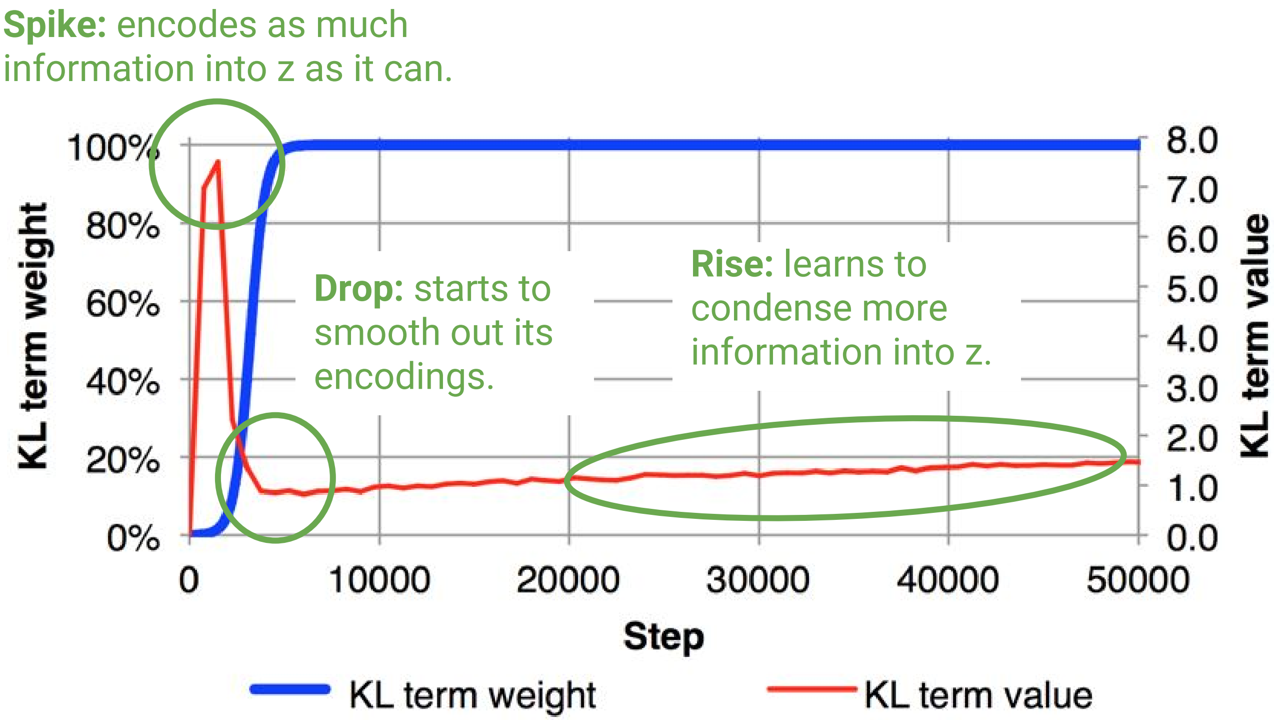
- VAE-based document-level language model [Miao et al., 2015]

→ **Input texts are models as bags of words.**
→ Not suitable to capture the sequential characteristics.

Optimization Challenges

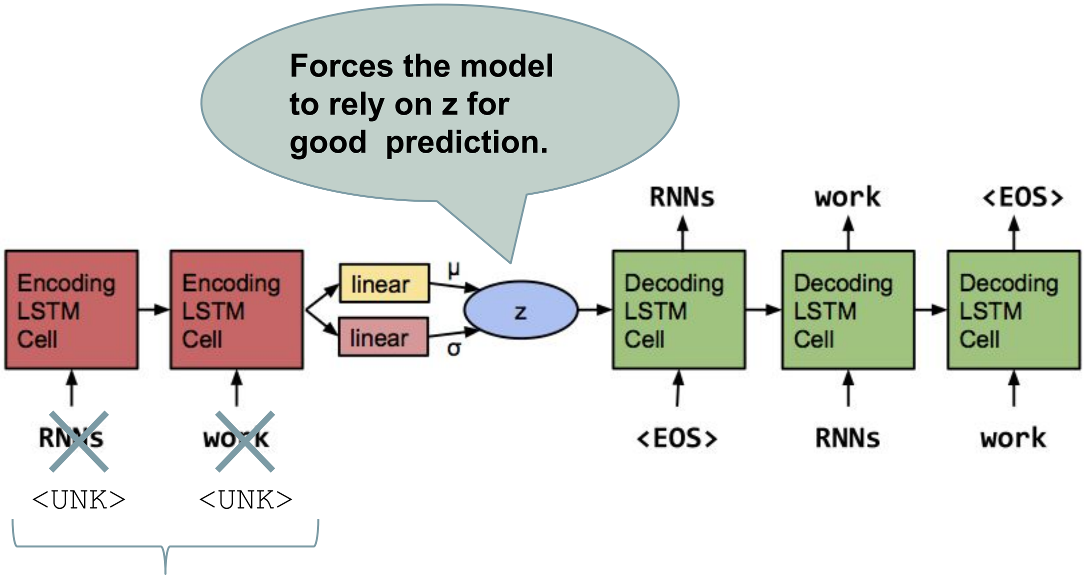
- KL divergence tends to become zero. (= $q(z|x) \sim p(x)$, which equals to RNNLM)
- Gives up capturing z and goes after explaining each sample with an optimized decoder.
- Authors introduce two techniques to migrate this issue.
    - KL cost annealing
    - Word dropout and historyless decoding

Optimization technique 1: KL cost annealing



Optimization technique 2: Word dropout and historyless decoding

Weaken the decoder by replacing a fraction of words with the generic unknown word token UNK.



Parameterized as a keep rate $k \in [0, 1]$.

Results: Language modeling

Objective: Is a continuous latent variable helpful for a standard task (language modeling)?

Methods:

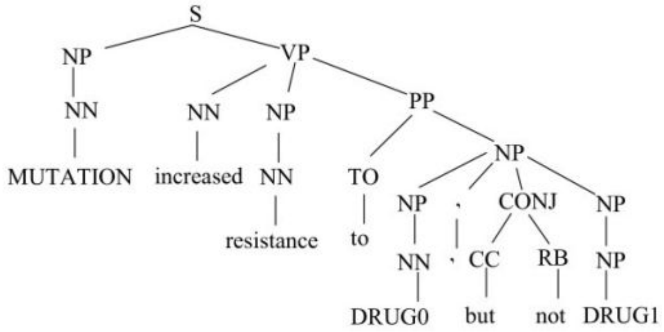
Baseline: RNNLM (non-variational)
Proposal: VAE (variational)

Dataset: Penn Treebank

Measurement: Likelihood

Baseline: True test likelihood
Proposal: Variational lower bound of likelihood

Disadvantage for VAE



Penn Treebank output example [Bui et al., 2010]

Results: Language modeling

Model	Standard				Inputless Decoder			
	Train NLL	Train PPL	Test NLL	Test PPL	Train NLL	Train PPL	Test NLL	Test PPL
RNNLM	100 –	95	**100** –	**116**	135 –	600	135 –	> 600
VAE	98 (2)	100	101 (2)	119	120 (15)	300	**125** (15)	**380**

NLL: negative log-likelihoods, PPL: perplexities. Lower is better for both. **() shows the KL loss.**

- VAE's performance is slightly worse than RNNLM.
- Without optimization technique (KL annealing and dropout), the performance becomes the same as RNNLM.

Note: 101 - 2 = 99 (<100) for reconstruction loss.

Inputless Decoder (== dropout keep rate = 0)

- KL loss becomes huge, but VAE shows better performance than RNNLM.

Results: Imputing missing words

Claim: VAE's global sentence features are suited to do the task of imputing missing words.

Computing the likelihood is intractable. (Need to calculate for every vocabulary V each step!)
→ Introduces **adversarial evaluation** for quantitative evaluation.

Adversarial evaluation: How well the output can deceive the classifier to train.

Example of imputing task and outputs from each model.

but now , as they parked out front and owen stepped out of the car , he could see _ _ _ _ _ _		
True: *that the transition was complete .*	**RNNLM:** *it , " i said .*	**VAE:** *through the driver 's door .*
you kill him and his _ _		
True: *men .*	**RNNLM:** *. "*	**VAE:** *brother .*
not surprising , the mothers dont exactly see eye to eye with me _ _ _ _		
True: *on this matter .*	**RNNLM:** *, i said .*	**VAE:** *, right now .*

Results: Imputing missing words

Dataset: Book corpus (mostly fiction)... 80M sentences after pruning.
Classifier to train: A bag-of-unigrams & LSTM logistic regression classifier.

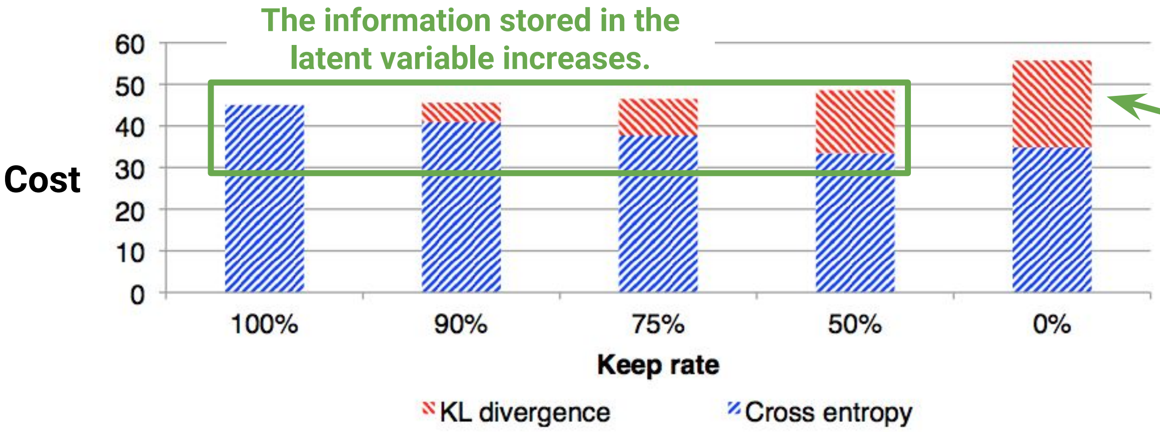
- VAE shows lower adversarial error (= much deceptive) than RNNLM.
- Negative log-likelihood is comparable.
    - RNNLM can make NLL lower by making “safe” answers.
      (e.g. complete the sentence with “, he said.”)

→ RNNLM yields less diverse samples than VAE, but produces natural sentences by favoring generic high-probability endings.

Model	Adv. Err. (%) Unigram	Adv. Err. (%) LSTM	NLL RNNLM
RNNLM (15 bm.)	28.32	38.92	46.01
VAE (3x5 bm.)	**22.39**	**35.59**	46.14

Adv. Error ∈ [0, 0.5]. 0.5 means all imputed samples are detected (not deceptive at all). 0 means all imputed samples are not detected. Lower is better.

Analysis: The impact of word dropout



With the extreme settings, the likelihood decreases.

Too typical, not much topics are captured.

Far less typical. Has a clear topic.

100% word keep	**75% word keep**
“ no , ” he said . *“ thank you , ” he said .*	*“ love you , too . ”* *she put her hand on his shoulder and followed him to the door .*
50% word keep	**0% word keep**
“ maybe two or two . ” *she laughed again , once again , once again , and thought about it for a moment in long silence .*	*i i hear some of of of* *i was noticed that she was holding the in in of the the in*

Ceases to be grammatically correct.

Analysis: Sampling from the posterior

- Analysis on reconstructed input texts.
    - Note: This is **variational**. It’s not re-outputting memorized samples.
- It seems capturing the length and topic of given sentences.
- As the sentence gets longer, the output becomes various.

INPUT	**we looked out at the setting sun .**	**i went to the kitchen .**	**how are you doing ?**
MEAN	*they were laughing at the same time .*	*i went to the kitchen .*	*what are you doing ?*
SAMP. 1	*ill see you in the early morning .*	*i went to my apartment .*	*“ are you sure ?*
SAMP. 2	*i looked up at the blue sky .*	*i looked around the room .*	*what are you doing ?*
SAMP. 3	*it was down on the dance floor .*	*i turned back to the table .*	*what are you doing ?*

Analysis: Homotopies

- Homotopy = linear interpolation. VAE can generate a sentence between two points in the latent space!
- Outputs are grammatical and have consistent topic.
- Similar syntax and topic but flipped sentiment can be problematic sometimes.

Sequence autoencoder outputs:

i went to the store to buy some groceries .
i store to buy some groceries .
i were to buy any groceries .
horses are to buy any groceries .
horses are to buy any animal .
horses the favorite any animal .
horses the favorite favorite animal .
horses are my favorite animal .

Grammatically incorrect

Variational autoencoder outputs:

" i want to talk to you . "
"i want to be with you . "
"i do n't want to be with you . "
i do n't want to be with you .
she did n't want to be with him .

Grammatically correct

he was silent for a long moment .
he was silent for a moment .
it was quiet for a moment .
it was dark and cold .
there was a pause .
it was my turn .

Grammatically correct

Analysis: Homotopies

there is no one else in the world .
there is no one else in sight .
they were the only ones who mattered .
they were the only ones left .
he had to be with me .
she had to be with him .
i had to do this .
i wanted to kill him .
i started to cry .
i turned to him .

this was the only way .
it was the only way .
it was her turn to blink .
it was hard to tell .
it was time to move on .
he had to do it again .
they all looked at each other .
they all turned to look back .
they both turned to face him .
they both turned and walked away .

im fine .
youre right .
" all right .
you 're right .
okay , fine .
" okay , fine .
yes , right here .
no , not right now .
" no , not right now .
" talk to me right now .
please talk to me right now .
i 'll talk to you right now .
" i 'll talk to you right now .
" you need to talk to me now .
" but you need to talk to me now .

convinced, understood

unconvinced, suspicious

[presenter interpretation]

Conclusion

This paper introduces the use of variational autoencoder for natural language sentences.

- Evaluated that novel techniques are useful for successful training
- Found that the model is effective for imputing missing words
- Showed the model can produce diverse sentences with smooth interoperation.

Future work

- Factorization of the latent space
- Incorporating conditional settings
- Learning sentence embeddings with semi-supervised fashion learning
- Going beyond adversarial evaluation to a fully adversarial training objective