CS 6501 Natural Language Processing

Text Generation

Yangfeng Ji

November 12, 2018

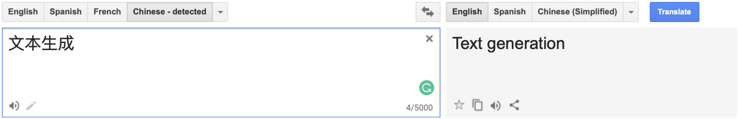
Department of Computer Science
University of Virginia

Overview

Introduction

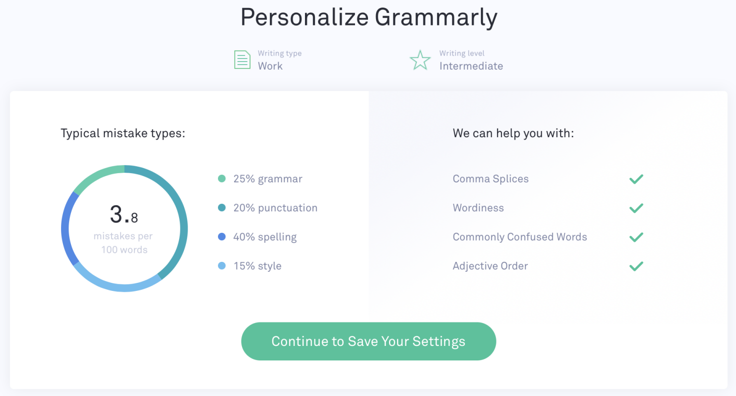
Text-to-text Generation

Machine translation

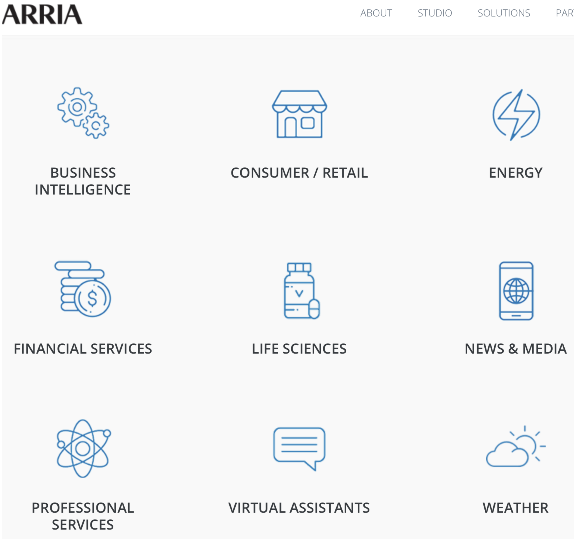


Text-to-text Generation (II)

Spelling, grammar and text correction



Data to Text Generation



Categories

Classical topics

- Text to text
  - machine translation
  - spelling, grammar and text correction
  - paraphrase generation
  - question generation
- Data to text
  - soccer reports
  - weather and financial reports
  - summaries of patient information in clinical contexts

Categories

New topics

- Image to text
  - image captioning
  - visual QA

Categories

New topics

- Image to text
  - image captioning
  - visual QA
- *Nothing* to text
  - decoding a RNN language model

Classic Text Generation

Example: Weather Report

The month was cooler and drier than average, with the average number of rain days. The total rain for the year so far is well below average. There was rain on every day for eight days from the 11th to the 18th, with mist and fog patches on the 16th and 17th. Rainfall amounts were mostly small, with light winds.

[Reiter and Dale, 2000]

Example: Weather Report

The month was cooler and drier than average, with the average number of rain days. The total rain for the year so far is well below average. There was rain on every day for eight days from the 11th to the 18th, with mist and fog patches on the 16th and 17th. Rainfall amounts were mostly small, with light winds.

The month was our driest and warmest August in our 24 year record, and our first rainless month. The 26th was the warmest August day in our record with 30.1 degrees, and our first hot August day (30). The month forms part of our longest dry spell: 47 days from 18th July to 2nd September. Rainfall so far is the same as at the end of July but now is very deficient. So far this has been our third driest year.

[Reiter and Dale, 2000]

NLG Tasks

NLG System Pipeline [Reiter and Dale, 2000]

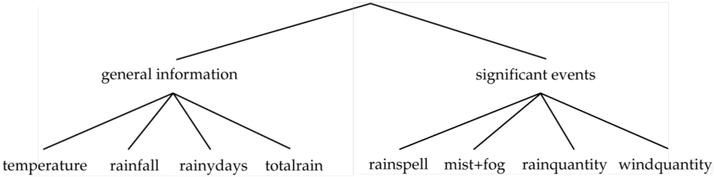
1. Content determination
2. Document structuring
3. Lexicalisation
4. Referring expression generation (REG)
5. Aggregation
6. Linguistic realisation
7. Structure realisation

NLG Tasks (II)

Module	Content task	Structure task
1. Document planning	Content determination	Document structuring
2. Microplanning	Lexicalisation REG	Aggregation
3. Realisation	Linguistic realisation	Structure realisation

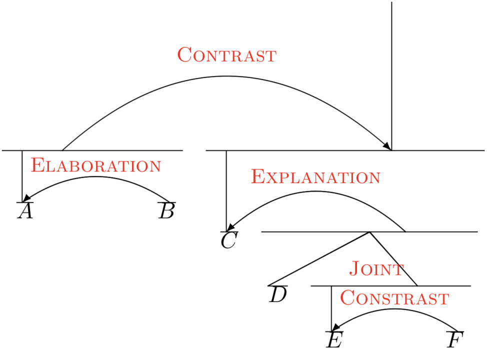
Document Structuring

It is concerned with the problem of imposing ordering and structure over the information to be conveyed.



[Reiter and Dale, 2000]

Document Structring (II)



[Although the food was amazing]^{A} [and I was in love with the spicy pork burrito,]^{B} [the service was really awful.]^{C} [We watched our waiter serve himself many drinks.]^{D} [He kept running into the bathroom]^{E} [instead of grabbing our bill.]^{F}

[Ji and Smith, 2017]

Lexicalisation

It is the problem of choosing the content words — nouns, verbs, adjectives and adverbs — that are required in order to express the selected content.

a. There was rain on every day for eight days from the 11th to the 18th.

b. There was rain on every day for eight days from the 11th.

c. There was rain on every day from the 11th to the 18th.

d. There was rain on the 11th, 12th, 13th, 14th, 15th, 16th, 17th, and 18th.

[Reiter and Dale, 2000]

Evaluation: BLEU

Let p_n be the n-gram precision score

$$\textsc{Bleu} = \mathrm{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \tag{1}$$

with

$$\mathrm{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{1-r/c} & \text{if } c \leq r \end{cases} \tag{2}$$

where r is the length of reference sentence and c is the translated sentence, with $N = 4$ and $w_1 = 1/N$.

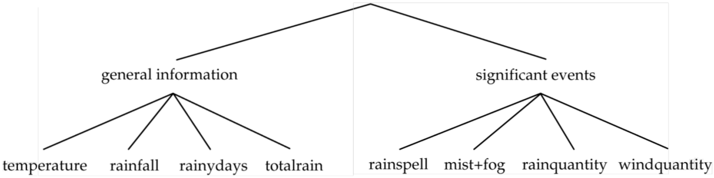
Evaluation

Some comments from E. Reiter

- Be cautious with metric-based evaluation
- Use human-based evaluation as well
- *"Overall, Overall, the evidence supports using BLEU for diagnostic evaluation of MT systems … , but does not support using BLEU outside of MT …"* [Reiter,]

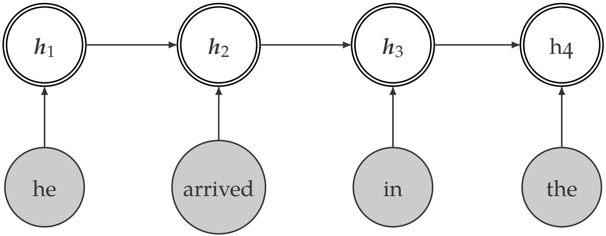
Limitation

Domain knowledge, such as

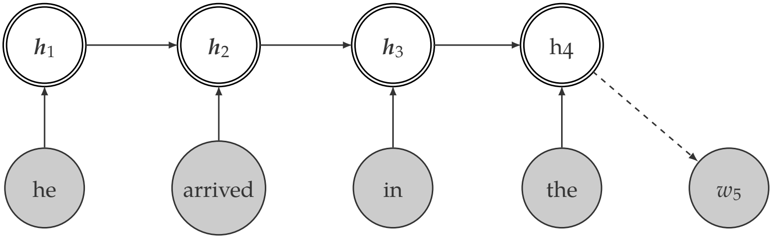


Neural Text Generation

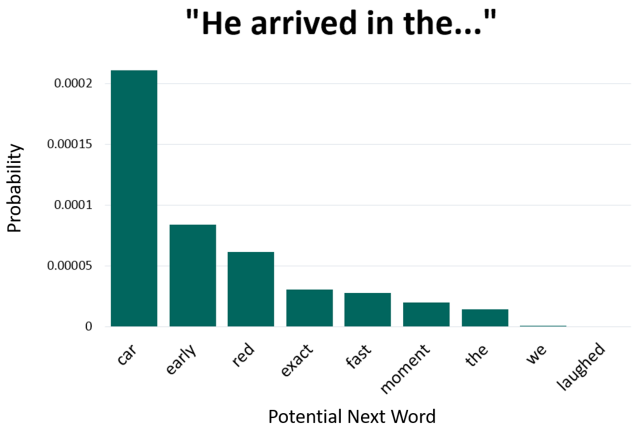
RNN Language Models



RNN Language Models



RNN Language Models (II)

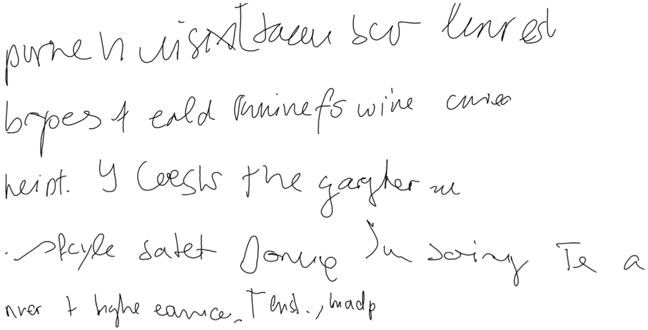


Generated Wikipedia data [Graves, 2013]

```
      <text xml:space="preserve">#REDIRECT [[Quartic war]]
</text>
    </revision>
  </page>
  <page>
    <title>ICSM</title>
    <id>14939</id>
    <revision>
      <id>42109942</id>
      <timestamp>2006-02-28T17:22:02Z</timestamp>
      <contributor>
        <username>Dtelclan</username>
        <id>26</id>
      </contributor>
      <minor />
      <comment>/* Possible catheterman */</comment>
      <text xml:space="preserve">[[Image:Isaac.org/ice.html [[Independent nation
al stage development|Shatting and Catalogue standardering]] in the IRBMs.

Up-2000 they called the SC 4220 system: he was swalloped early in Calvino, or si
nce each trial mentioned based on [[Balbov's new single-jarget|bit-oriann guess]
```

Generated handwritten texts [Graves, 2013]



Problems in Neural Text Generation

Three typical research problems

- How to increase diversity?
- How to incorporate context?
- How to deal with the training-decoding mismatch (exposure Bias)?

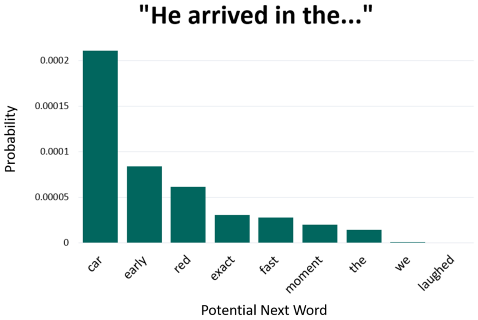
Problems in Neural Text Generation

Three typical research problems

- How to increase diversity?
- How to incorporate context?
- How to deal with the training-decoding mismatch (exposure Bias)?

Does classic NLG have the same problems?

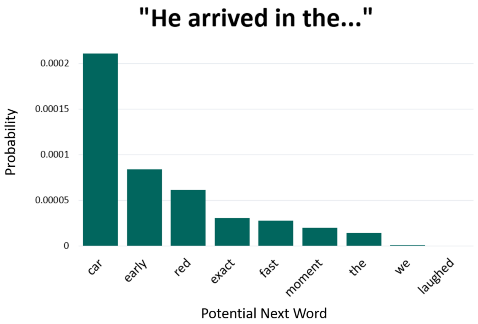
Diversity



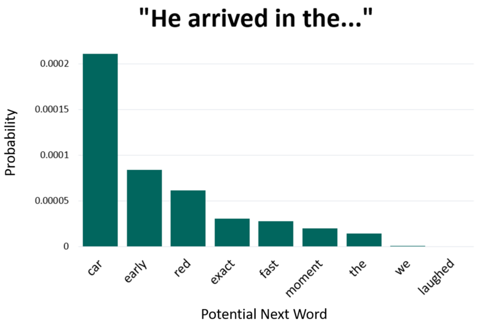
Solutions

- Random sampling + external evaluation
- Beam search
- Variational auto-encoder

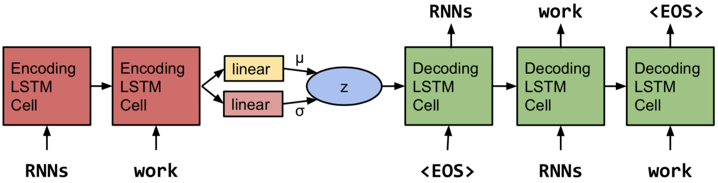
Random Sampling



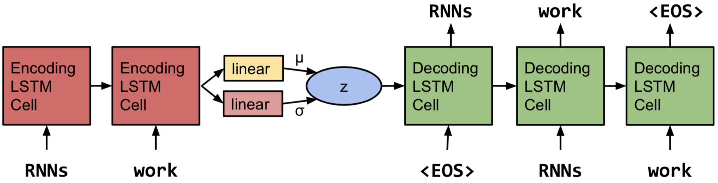
Beam Search


"He arrived in the..."
Probability
0.0002
0.00015
0.0001
0.00005
0
car
early
red
exact
fast
moment
the
we
laughed
Potential Next Word

Variational Auto-encoder



Variational Auto-encoder



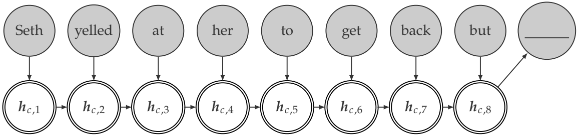
but now , as they parked out front and owen stepped out of the car , he could see _ _ _ _ _ _		
True: *that the transition was complete .*	**RNNLM:** *it , " i said .*	**VAE:** *through the driver 's door .*
you kill him and his _ _		
True: *men .*	**RNNLM:** *. "*	**VAE:** *brother .*
not surprising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

Problems

- ✓ How to increase diversity?
- How to incorporate context?
- How to deal with the training-decoding mismatch (exposure Bias)?

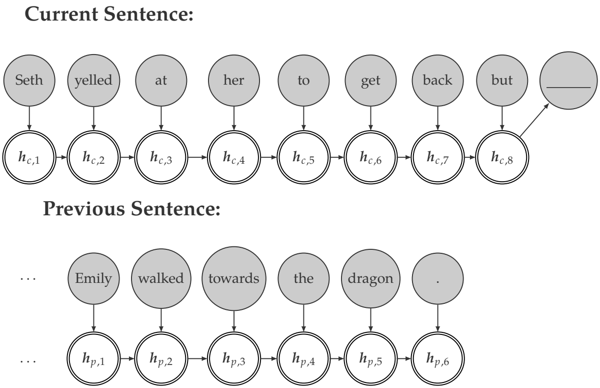
Contextual Information

Current Sentence:



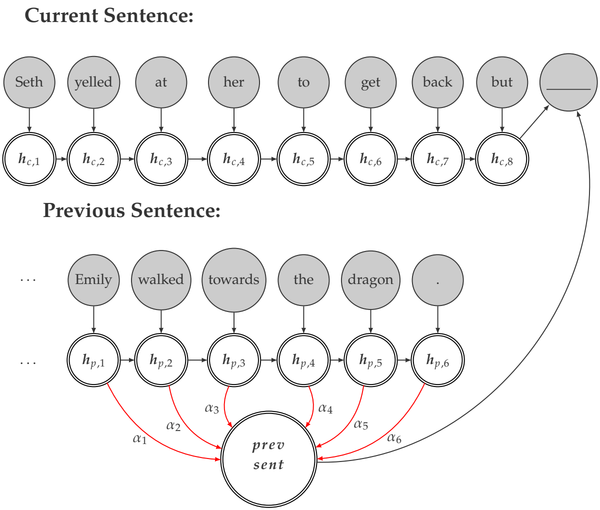
[Bahdanau et al., 2015]

Contextual Information



[Bahdanau et al., 2015]

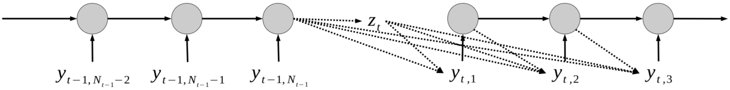
Contextual Information



[Bahdanau et al., 2015]

Contextual Information (II)

Semantic relations between sentences



$$p(y_{t,n+1} \mid z_t,\ \boldsymbol{y}_{t,<n},\ \boldsymbol{y}_{t-1}) = g\Big(\underbrace{\mathbf{W}_o^{(z_t)} \boldsymbol{h}_{t,n}}_{\substack{\text{relation-specific} \\ \text{intra-sentential context}}} + \underbrace{\mathbf{W}_c^{(z_t)} \boldsymbol{c}_{t-1}}_{\substack{\text{relation-specific} \\ \text{inter-sentential context}}} + \underbrace{\boldsymbol{b}_o^{(z_t)}}_{\substack{\text{relation-specific} \\ \text{bias}}} \Big)$$

`I like coffee. But ______`

[Ji et al., 2016]

Contextual Information (III)

Context	All of a sudden, [*Emily*]$_1$ walked towards [*the dragon*]$_2$.
Current Sentence	[*Seth*]$_3$ yelled at [*her*]$_1$ to get back but ________________

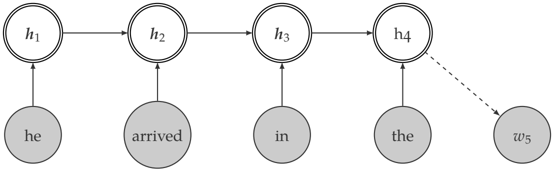
[Clark et al., 2018]

Problems

- ✓ How to increase diversity?
- ✓ How to incorporate context?
- ► How to deal with the training-decoding mismatch (exposure Bias)?

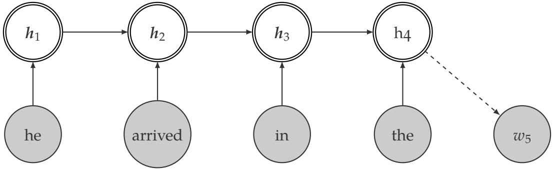
Exposure Bias

The model is never exposed to its own errors during training, and so the inferred histories at test-time do not resemble the gold training histories.



Exposure Bias

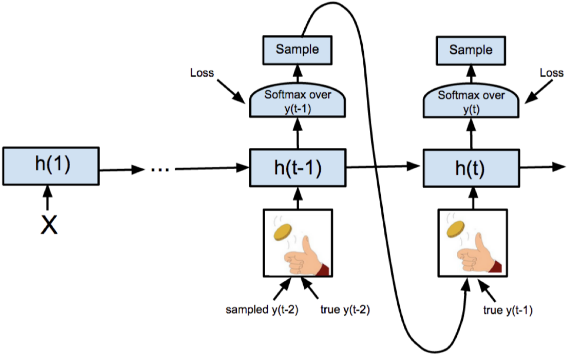
The model is never exposed to its own errors during training, and so the inferred histories at test-time do not resemble the gold training histories.



Solutions

- Scheduled Sampling
- Beam search Optimization [Wiseman and Rush, 2016]

Scheduled Sampling

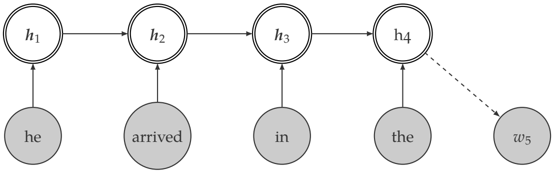


[Bengio et al., 2015]

Comparison

Comparison

(Basic) Neural text generation



Classical text generation

Module	Content task	Structure task
Document planning	Content determination	Document structuring
Microplanning	Lexicalisation REG	Aggregation
Realsation	Linguistic realisation	Structure realisation

Comparison (II)

- Classic NLG
  - You know what you want to say.
  - Focus on the best way of saying it in order to achieve a communication goal.

(Goldberg, 2018)

Comparison (II)

- Classic NLG
  - You know what you want to say.
  - Focus on the best way of saying it in order to achieve a communication goal.
- Neural NLG
  - Generate me some text given this input.
  - Yay it looks readable!!

(Goldberg, 2018)

Summary

Reference

Bahdanau, D., Cho, K., and Bengio, Y. (2015).
Neural machine translation by jointly learning to align and translate.
In *ICLR*.

Bengio, S., Vinyals, O., Jaitly, N., and Shazeer, N. (2015).
Scheduled sampling for sequence prediction with recurrent neural networks.
In *Advances in Neural Information Processing Systems*, pages 1171–1179.

Clark, E., Ji, Y., and Smith, N. A. (2018).
Neural text generation in stories using entity representations as context.
In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 2250–2260.

Graves, A. (2013).
Generating sequences with recurrent neural networks.
arXiv preprint arXiv:1308.0850.

Ji, Y., Haffari, G., and Eisenstein, J. (2016).
A latent variable recurrent neural network for discourse relation language models.
In *Proceedings of NAACL-HLT*, pages 332–342.

Ji, Y. and Smith, N. A. (2017).
Neural discourse structure for text categorization.
In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 996–1005.

Reiter, E.
A structured review of the validity of bleu.
Computational Linguistics, (Just Accepted):1–12.

Reiter, E. and Dale, R. (2000).