

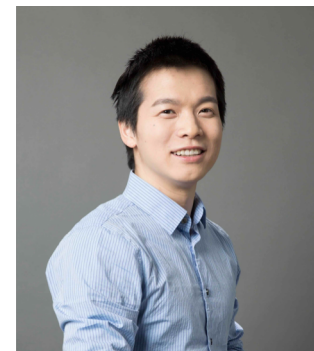
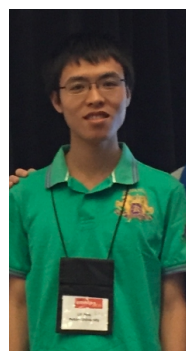
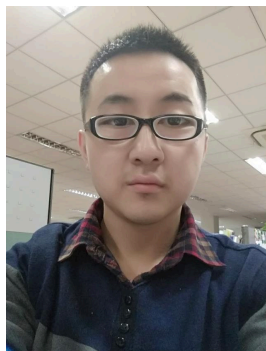


AAAI 2019

CGMH: Constrained Sentence Generation by Metropolis-Hastings Sampling

Ning Miao^{1,2}, Hao Zhou¹, Lili Mou³, Rui Yan², **Lei Li**¹

¹ByteDance, ²Peking University, ³AdeptMind Research



Outline

- Motivation and Constrained Sentence Generation
- Limitation of Existing Methods
- Overview of CGMH
- Experiments
 - Keyword to sentence
 - Paraphrasing
 - Sentence Error Correction
- Conclusion

Motivation: Design Advertisement Slogans



Keywords from Advertiser

Rin clothes bright



Advertisement Slogan

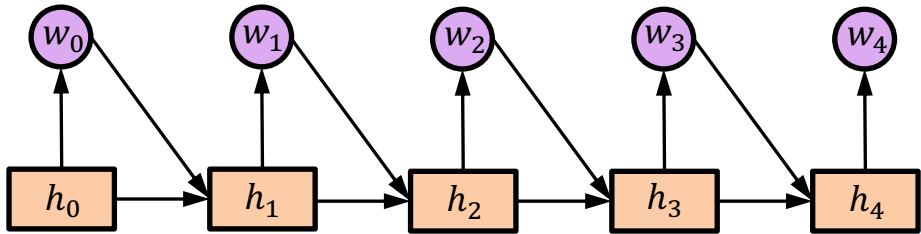


Problem Definition

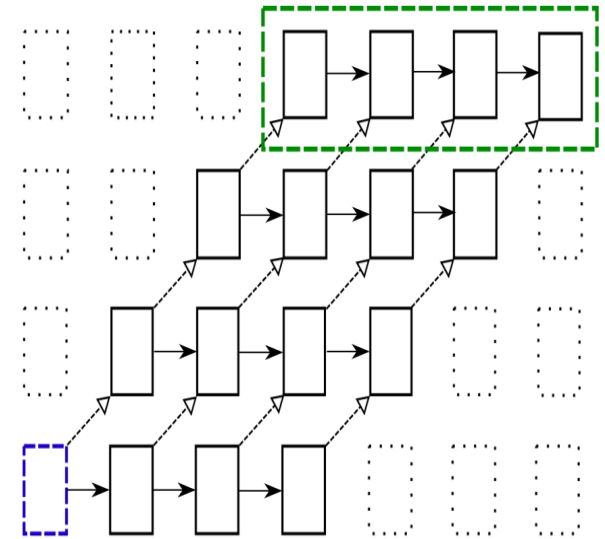
- Generating sentence satisfying constraints:
- Hard constrains: Keyword must occur in sentences (e.g. keyword2sentence task)
 - E.g. Juice -> Brand natural juice, specially made for you
- Soft constrains: Semantically similar to a given sentence (e.g. paraphrase task)
 - E.g. The movie is great -> It is one of my favorite movies

Existing Approach: Grid Beam Search or RNN with Separate decoding

- LSTM with Grid beam search

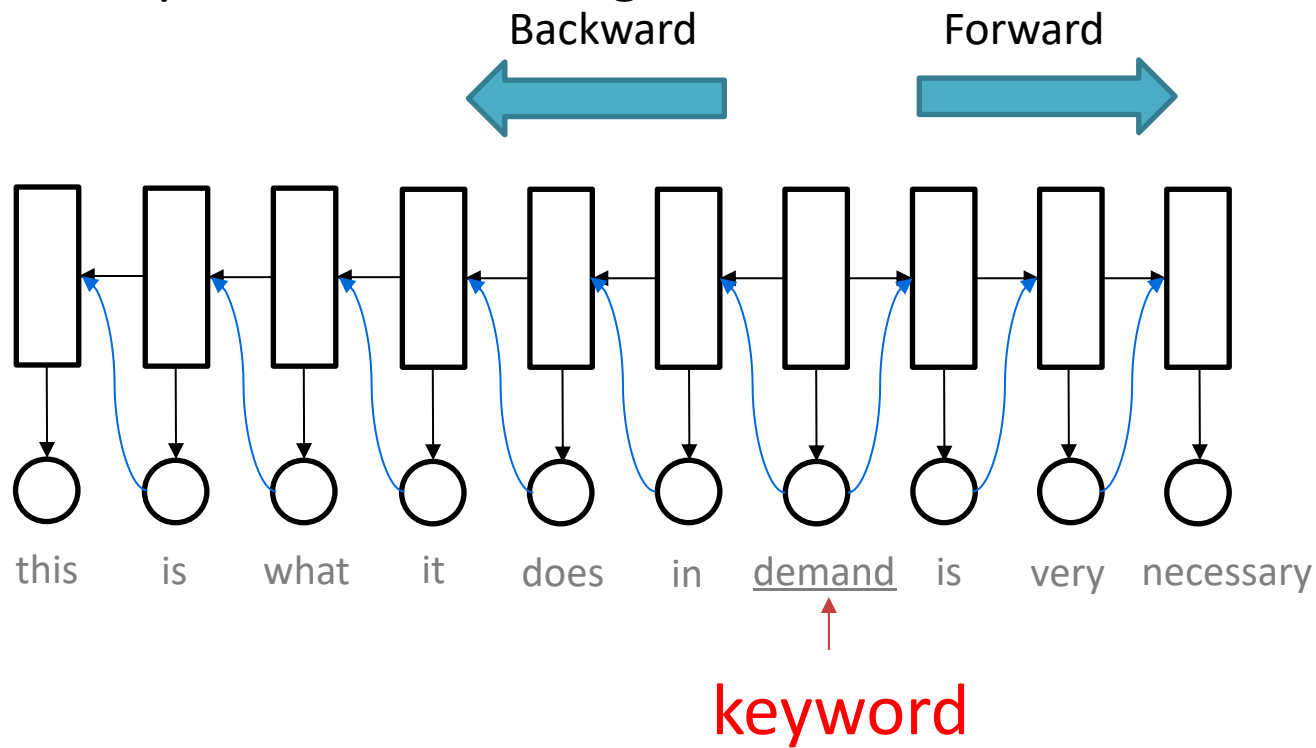


Constraint
3
Constraint
2
Constraint
1
No
constraint
s



LSTM with Separate Backward forward decoding: limited to one keyword

- LSTM w/ sep-B/F, which generates **independent backward and forward** sequences from the given word.



LSTM with Separate Backward forward decoding: limited to one

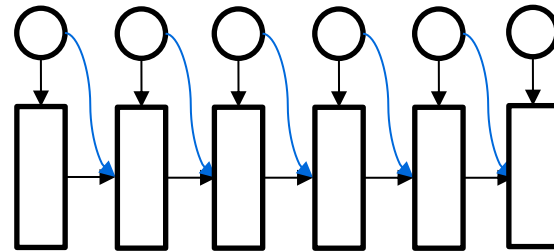
keyword

- LSTM w/ asyn-B/F, which **first generates the first half of a sentence and then generates another half** conditioned on the first half.

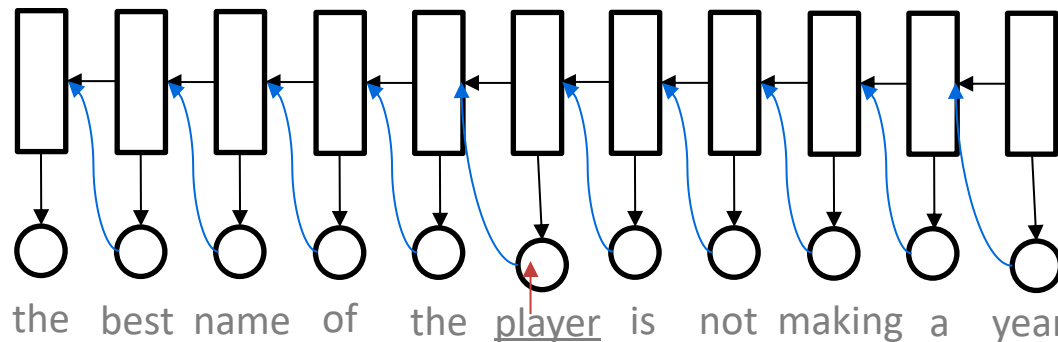
Step1: Forward



player is not making a year



Step2: Backward



keyword

Formulation in Proposed CGMH

- To generation samples (sentences) from the target distribution

$$\pi(x) = \prod_t P(x_t | x_{0:t-1}) \cdot \prod_i P_C^i(x)$$

language model probability

Indicator(0-1) function for constraints

Challenge

- To generation samples (sentences) from the target distribution

$$\pi(x) = \prod_t P(x_t | x_{0:t-1}) \cdot \prod_i P_C^i(x)$$

language model probability

Indicator(0-1) function for constraints

- $\pi(x)$ is high-dimensional, and no direct sampling method.

Main Idea of CGMH

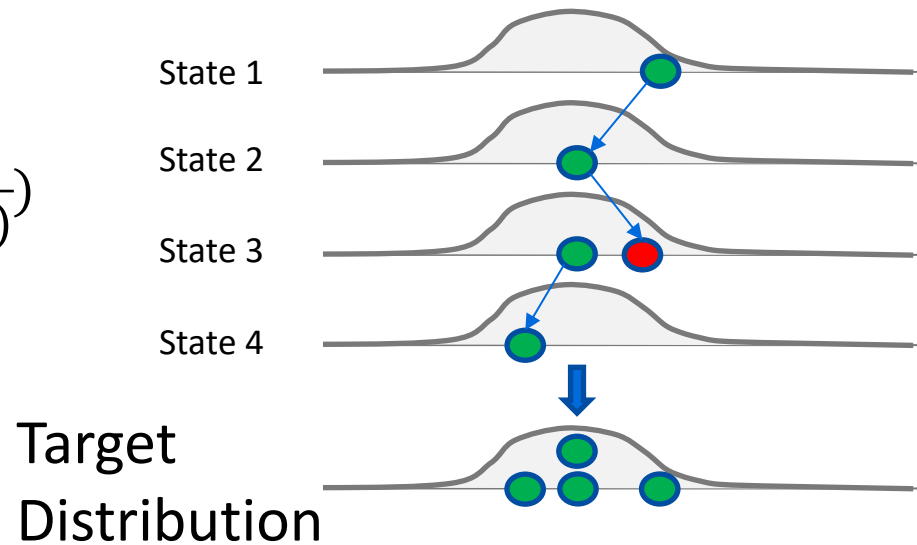
- Instead of sampling from $\pi(x)$ directly, generate samples iteratively:
 - Starting with initial keywords
 - next sentence based on modification of previous
 - action proposals to modify the sentences
- Metropolis-Hastings Algorithm

Background: Metropolis-Hastings sampling

- Metropolis-Hastings(MH) perform sampling by first **proposes** a transition, and then **accepts or rejects** the transition.

$$A(x'|x_{t-1}) = \min\left(1, \frac{\pi(x') \cdot g(x_{t-1}|x')}{\pi(x_{t-1}) \cdot g(x'|x_{t-1})}\right)$$

g is proposal distribution



CGMH

- CGMH performs constrained generation by:
1. Pretrain Language Model prob;
 2. Start from a initial sentence x_0 ;
 3. Propose a new sentence x_t from x_{t-1} , and
accept/reject the action. Action proposal include:
 - I. Replacement: change a word to another one
 - II. Insertion: add a word
 - III. Deletion: remove a word

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
------	--------	---------	-----------

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports
1	Insert	Accept	BMW sports car

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports
1	Insert	Accept	BMW sports car
2	Insert	Accept	BMW the sports car

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports
1	Insert	Accept	BMW sports car
2	Insert	Accept	BMW the sports car
...

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports
1	Insert	Accept	BMW sports car
2	Insert	Accept	BMW the sports car
...
6	Insert	Accept	BMW , the sports car of daily life

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports
1	Insert	Accept	BMW sports car
2	Insert	Accept	BMW the sports car
...
6	Insert	Accept	BMW , the sports car of daily life
7	Replace	Accept	BMW , the sports car of future life

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports
1	Insert	Accept	BMW sports car
2	Insert	Accept	BMW the sports car
...
6	Insert	Accept	BMW , the sports car of daily life
7	Replace	Accept	BMW , the sports car of future life
8	Insert	Accept	BMW , the sports car of the future life

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports
1	Insert	Accept	BMW sports car
2	Insert	Accept	BMW the sports car
...
6	Insert	Accept	BMW , the sports car of daily life
7	Replace	Accept	BMW , the sports car of future life
8	Insert	Accept	BMW , the sports car of the future life
9	Delete	Reject	BMW , the sports car of the future life

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports
1	Insert	Accept	BMW sports car
2	Insert	Accept	BMW the sports car
...
6	Insert	Accept	BMW , the sports car of daily life
7	Replace	Accept	BMW , the sports car of future life
8	Insert	Accept	BMW , the sports car of the future life
9	Delete	Reject	BMW , the sports car of the future life
10	Delete	Accept	BMW , the sports car of the future life

Illustration of CGMH

- CGMH performs Metropolis-Hastings **sampling directly in sentence space.**

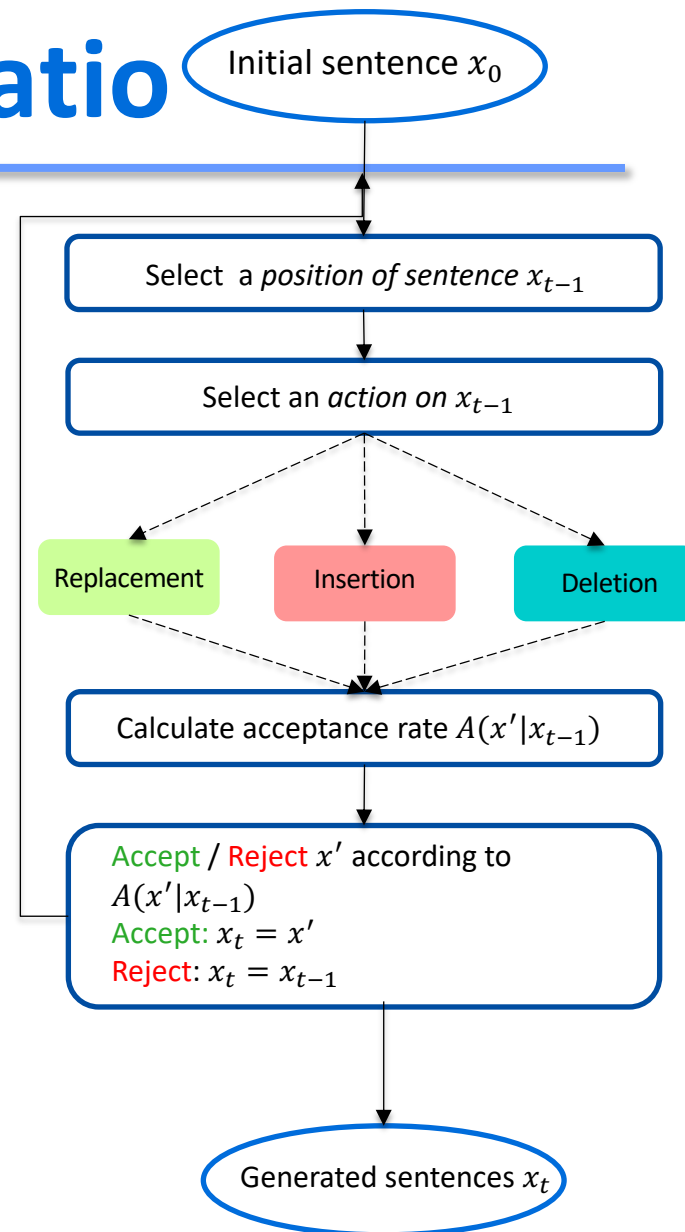
Step	Action	Acc/Rej	Sentences
0	[Input]		BMW sports
1	Insert	Accept	BMW sports car
2	Insert	Accept	BMW the sports car
...
6	Insert	Accept	BMW , the sports car of daily life
7	Replace	Accept	BMW , the sports car of future life
8	Insert	Accept	BMW , the sports car of the future life
9	Delete	Reject	BMW , the sports car of the future life
10	Delete	Accept	BMW , the sports car of the future life
11	[Output]		BMW , the sports car of the future

CGMH: Acceptance Ratio

- Calculate the acceptance rate:

$$A(x'|x_{t-1}) = \min\left(1, \frac{\pi(x') \cdot g(x_{t-1}|x')}{\pi(x_{t-1}) \cdot g(x'|x_{t-1})}\right)$$

- Accept x' with probability $A(x'|x_{t-1})$

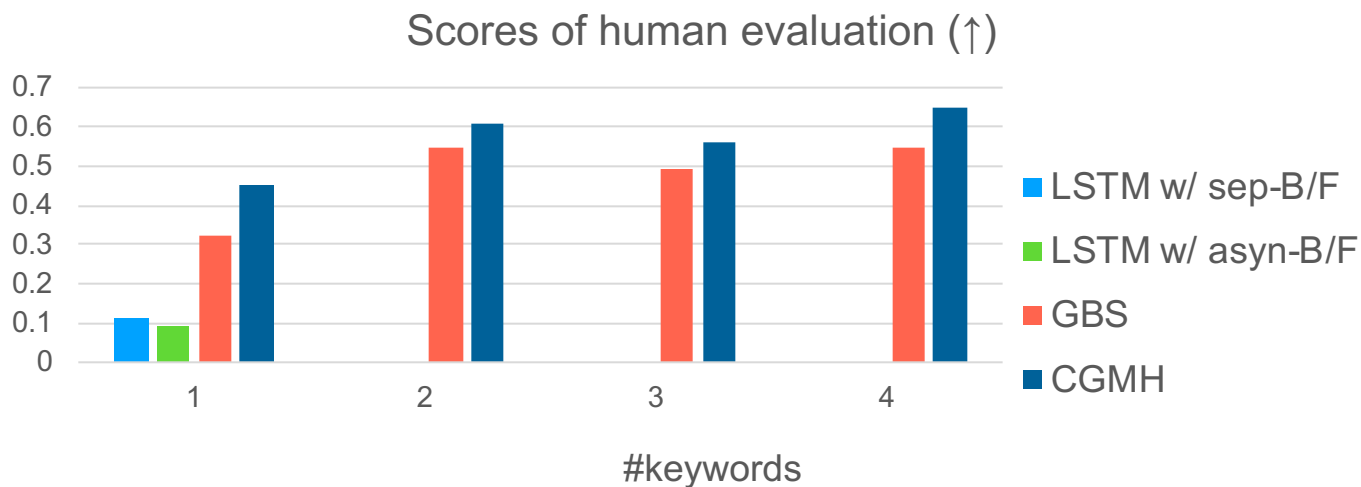
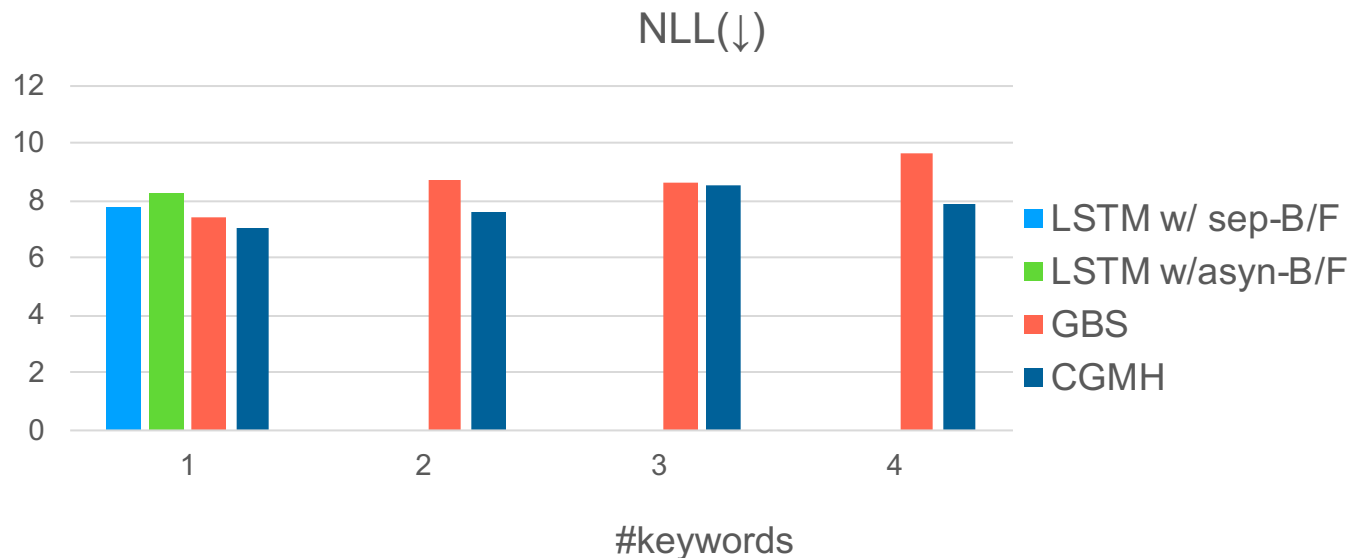


Experiment

- Keywords to Sentence Generation (*hard*)
- Unsupervised Paraphrase Generation (*soft*)
- Sentence Correction (*soft*)

Experiment – Keywords to Sentence Generation

➤ CGMH outperforms previous work in both NLL and human evaluations.



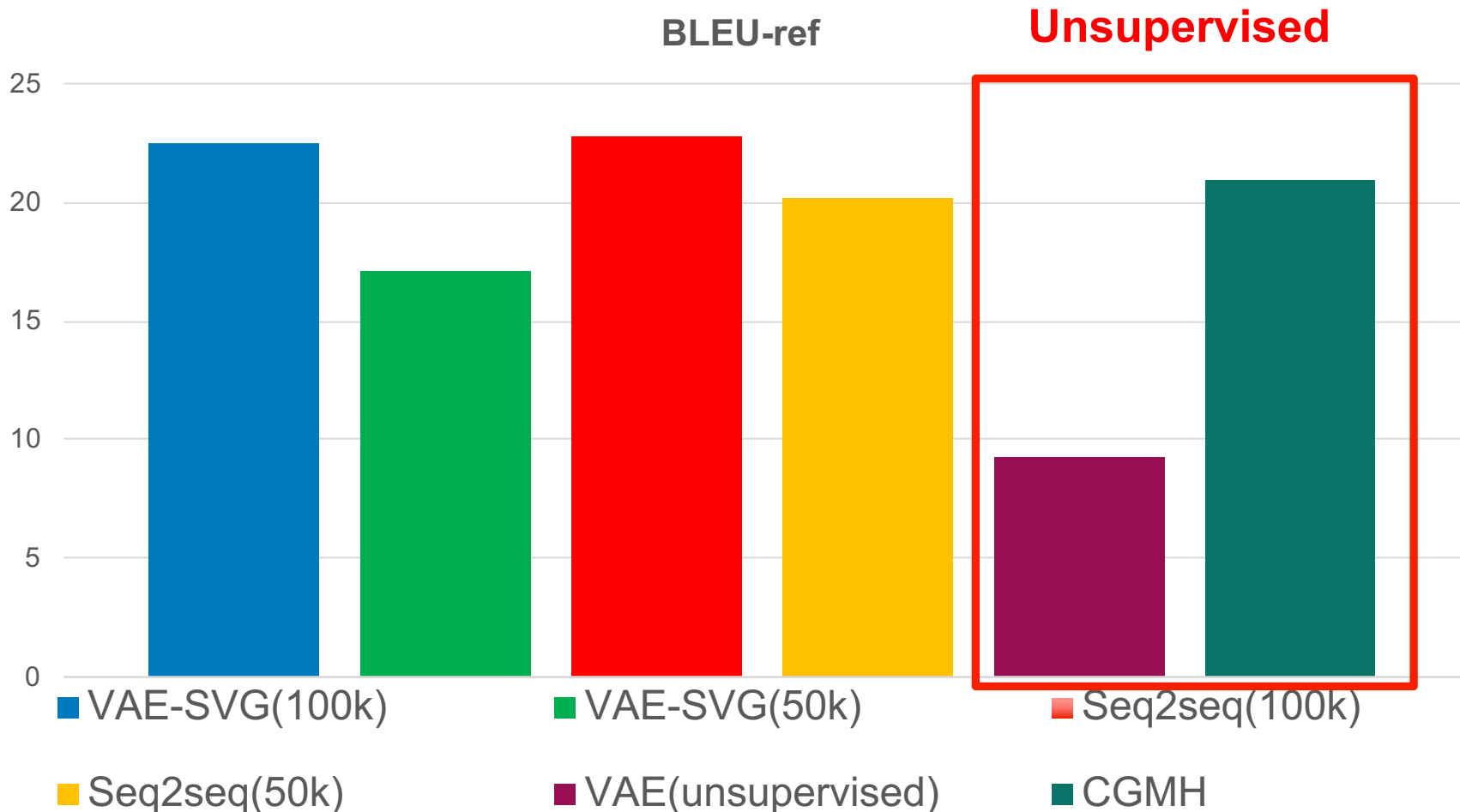
Keyword-to-Sentence: Showcase

➤ Examples of CGMH and GBS.

Keyword(s)	Generated Sentences	GBS
friends	My good friends were in danger .	But friends and family have been arrested .
project	The first project of the scheme .	The project , which is expected to be completed next year
have, trip	But many people have never made the trip .	But the trip has be completed .
lottery, scholarships	But the lottery has provided scholarships.	The lottery is a scholarship .
decision, build, home	The decision is to build a new home.	The decision builds a house for home .
attempt, copy, painting, denounced	The first attempt to copy the painting was denounced.	But attempt to copy painting will be denounced.

Experiment – Paraphrase Generation

- CGMH is the first unsupervised model to achieve comparable results with supervised models.



Experiment – Unsupervised Paraphrase Generation

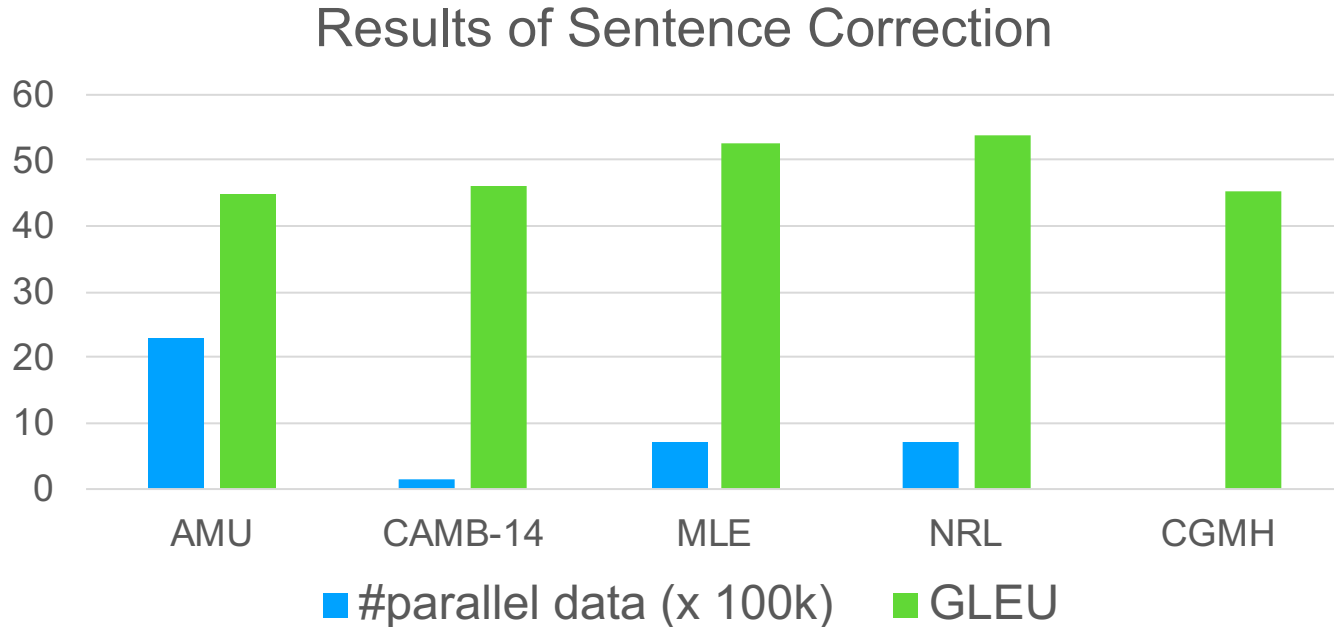
- CGMH is the first unsupervised model to achieve comparable results with supervised models.

Examples

- 1, what 's the best plan to lose weight ->
what 's the best way to slim down quickly
2. how should i control my emotion ->
how do i control my anger
3. why do my dogs love to eat tuna fish ->
why do my dogs like to eat raw tuna and raw fish

Experiment – Unsupervised Error Correction

- CGMH outperforms some of the supervised models trained on large parallel corpus.
- Dataset: JFLEG, 1501 sentences



Experiment – Unsupervised Error Correction

- CGMH outperforms some of the supervised models trained on large parallel corpus.

Erroneous sen1	Even if we are failed , we have to try to get a new things .
Reference sen1	Even if we all failed , we have to try to get new things .
Output sen1	Even if we are failing , we have to try to get some new things ..
Erroneous sen2	In the world oil price very high right now .
Reference sen2	In today 's world , oil prices are very high right now .
Output sen2	In the world , oil prices are very high right now .

Conclusion

- CGMH is a Monte Carlo sentence generation algorithm capable of dealing with various constrained sentence generation.
- Method beyond DL such as Bayesian method still works
- Unsupervised method can achieve comparable performance to supervised methods in paraphrasing

Thanks!

- Contact: lileilab@bytedance.com
- Code is available at:
<https://github.com/NingMiao/CGMH>

Reference

- [1] Hokamp, C., and Liu, Q. 2017. Lexically constrained decoding for sequence generation using grid beam search. In *ACL*.
- [2] Anderson, P.; Fernando, B.; Johnson, M.; and Gould, S. 2017. Guided open vocabulary image captioning with constrained beam search. In EMNLP.
- [3] Gupta, A.; Agarwal, A.; Singh, P.; and Rai, P. 2017. A deep generative framework for paraphrase generation. *arXiv preprint arXiv:1709.05074*.
- [4] Mou, L.; Yan, R.; Li, G.; Zhang, L.; and Jin, Z. 2015. Backward and forward language modeling for constrained sentence generation. *arXiv preprint arXiv:1512.06612*.
- [5] Bowman, S. R.; Vilnis, L.; Vinyals, O.; Dai, A.; Jozefowicz, R.; and Bengio, S. 2016. Generating sentences from a continuous space. In CoNLL.
- [6] Li, Z.; Jiang, X.; Shang, L.; and Li, H. 2017. Paraphrase generation with deep reinforcement learning. *arXiv preprint arXiv:1711.00279*.
- [7] Junczys-Dowmunt, M., and Grundkiewicz, R. 2016. Phrasebased machine translation is state-of-the-art for automatic grammatical error correction. *arXiv preprint arXiv:1605.06353*.
- [8] Felice, M.; Yuan, Z.; Andersen, Ø. E.; Yannakoudakis, H.; and Kochmar, E. 2014. Grammatical error correction using hybrid systems and type filtering. In *CoNLL*.
- [9] Napoles, C.; Sakaguchi, K.; Post, M.; and Tetreault, J. 2015. Ground truth for grammatical error correction metrics. In *ACL*.

Appendix – Properties of CGMH

➤ Property of CGMH:

1. *Detailed balance condition* is satisfied;
2. Sampling process is *irreducible* and *nonperiodic*.
3. So *ergodicity* of sampling from π is satisfied.

Appendix – $P_C(x)$ for different tasks

➤ For different tasks, we use different $P_C(x)$:

- Keywords2Sentence: $P_C(x) = 1_{\{x \text{ contains the keywords}\}}$
- Paraphrase: $P_C(x) = 1 / P_C^{KW}(x) / P_C^{KW}(x) P_C^{SIM}(x)$
- Correction: $P_C(x) = 1 / P_C^{WMA}(x)$

Appendix – Details of experiments

➤ Keywords to Sentence Generation (*hard*)

- Aim: To generate fluent sentences containing the given set of words.
- Dataset: A subset of One–Billion–Word Corpus (5M)
- Initial sentence: Keywords.

➤ Unsupervised Paraphrase Generation (*soft*)

- Aim: To generate sentences with similar meaning of the given one.
- Dataset: Quora(140k pairs of paraphrase sentences)
- Initial sentence: Original sentence.

➤ Sentence Correction (*soft*)

- Aim: To correct the errors in the given sentence.
- Dataset: A subset of One–Billion–Word Corpus (5M, base language model) and JFLEG(1501 sentences, for test only)
- Initial sentence: Erroneous sentence.