Towards Neural Controllable Text Generation

Jingjing Li

Ph.D. Oral Defense

Supervisor: Prof. Irwin King and Prof. Michael R. Lyu

Committee:

Prof. Young Fung Yu

Prof. Lam Wai

Prof. Shou-De Lin

Outline

- ❑ Introduction
- ❑ Improving Semantic Fidelity in Text Generation
- ❑ Unsupervised Controllable Generation by Learning from Search
- ❑ Unsupervised Controllable Generation by Iterative Revision
- ❑ Conclusion

Outline

- ❑ Introduction
- ❑ Improving Semantic Fidelity in Text Generation
- ❑ Unsupervised Controllable Generation by Learning from Search
- ❑ Unsupervised Controllable Generation by Iterative Revision
- ❑ Conclusion

Controllable Text Generation

- Text Generation is to generate fluent and natural text.
    - Applications: summarization, dialogue generation, story generation...

Controllable Text Generation

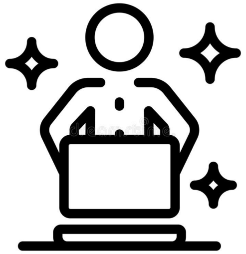
- Text Generation is to generate fluent and natural text.
  - Applications: summarization, dialogue generation, story generation...
- Controllable Text Generation (CTG) is to generate text whose attributes can be controlled.

Controllable Text Generation

- Text Generation is to generate fluent and natural text.
  - Applications: summarization, dialogue generation, story generation...
- Controllable Text Generation (CTG) is to generate text whose attributes can be controlled.
  - Example: *informal* vs. *formal* expression

Wow , I’m very dumb in my observation skills.....

I do not have good observation skills.



Controllable Text Generation

- Text Generation is to generate fluent and natural text.
  - Applications: summarization, dialogue generation, story generation...
- Controllable Text Generation (CTG) is to generate text whose attributes can be controlled.
  - Example: *negative* vs. *positive* comments

The burgers were over cooked to the point the meat was crunchy.

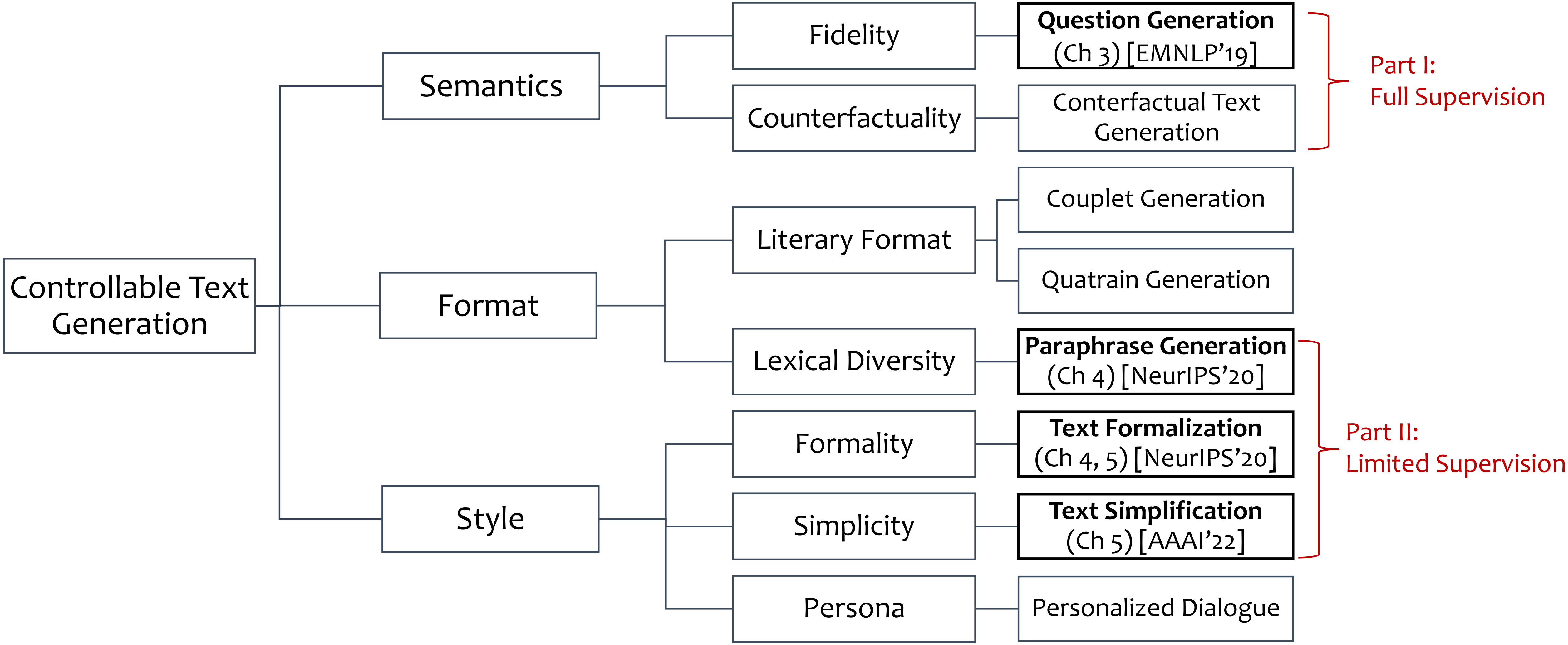
The burgers were perfectly cooked and I like the juicy meat!

Controllable Text Generation

- Text Generation is to generate fluent and natural text.
  - Applications: summarization, dialogue generation, story generation...
- Controllable Text Generation (CTG) is to generate text whose attributes can be controlled.
- Applications of CTG
  - Paraphrase generation (lexical diversity)
  - Text style transfer (style)
  - Text simplification (simplicity)
  - Grammatical error correction (syntax, grammar)
  - Text detoxification (toxic contents)
  - … …

Taxonomy

- Our Contributions



Part I: Supervised Text Generation

- Problem Setting
  - Training data: $\{(\mathbf{x}^{(m)}, \boldsymbol{y}^{(m)})\}_{m=1}^{M}$, learn a mapping function from $\boldsymbol{x}$ to $\boldsymbol{y}$.
- Solution
  - sequence-to-sequence framework
- Challenges
  - Semantic fidelity
  - Entity accuracy, hallucination
  - …

Outline

- ❑ Introduction
- ❑ Improving Semantic Fidelity in Text Generation
- ❑ Unsupervised Controllable Generation by Learning from Search
- ❑ Unsupervised Text Attribute Transfer by Iterative Revision
- ❑ Conclusion

Semantic Fidelity in QG: Introduction

- Question generation (QG) is to generate a *question* from a *reference sentence* and a specified *answer* within the reference sentence.

… "Oxygen is used in cellular respiration and released by photosynthesis, which uses the energy of sunlight to produce oxygen from water." …

- What life process produces oxygen in the presence of light?

- Photosynthesis uses which energy to form oxygen from water?

Semantic Fidelity in QG: Introduction

- Challenges
  - Identify question-related context words
  - The generated question should be relevant to the given answer.

Sentence: The daily mean temperature in January, the area's coldest month, is 32.6 °F (0.3 °C); however, temperatures usually drop to 10 °F (-12 °C) several times per winter and reach 50 °F (10 °C) several days each winter month.
Reference Question: What is New York City 's daily January mean temperature in degrees celsius ?

- Existing solutions
  - Zhou et al. (2017) uses BIO tagging scheme;
  - Sun et al. (2018) proposes proximity-based answer position encoding;

Semantic Fidelity in QG: Motivation

1. Proximity-based answer-aware approaches can't tackle with sentences with complex structure.
   - Example

Sentence: The daily mean temperature in January, the area's coldest month, is 32.6 °F (0.3 °C); however, temperatures usually drop to 10 °F (-12 °C) several times per winter and reach 50 °F (10 °C) several days each winter month.

Reference Question: What is New York City 's daily January mean temperature in degrees celsius ?
Baseline Prediction: What is the coldest temperature in Celsius ?

   - Experiment verification

Distance	B1	B2	B3	B4	MET	R-L
0∼5 (36.6% of #)	45.08	30.19	22.06	16.52	21.91	47.33
5∼10 (36.2% of #)	41.55	27.53	19.83	14.74	20.55	43.81
>10 (27.2% of #)	35.60	21.67	14.75	10.38	16.70	37.53

Table 1: Performance for the average relative distance between the answer fragment and other non-stop sentence words that also appear in the ground truth question

Semantic Fidelity in QG: Motivation

2. Answer-related structured relation can help keep the generated question to the point.

Sentence: The daily mean temperature in January, the area's coldest month, is 32.6 °F (0.3 °C); however, temperatures usually drop to 10 °F (-12 °C) several times per winter and reach 50 °F (10 °C) several days each winter month.

Structured Answer-relevant Relation: (The daily mean temperature in January; is; 32.6 °F (0.3 °C))

Reference Question: What is New York City 's daily January mean temperature in degrees celsius ?

- Verification:

	Sentence	Answer-relevant Relation
Avg. length	32.46	13.04
# overlapped words	2.87	1.86
Copy ratio	8.85%	14.26%

Semantic Fidelity in QG: Methodology

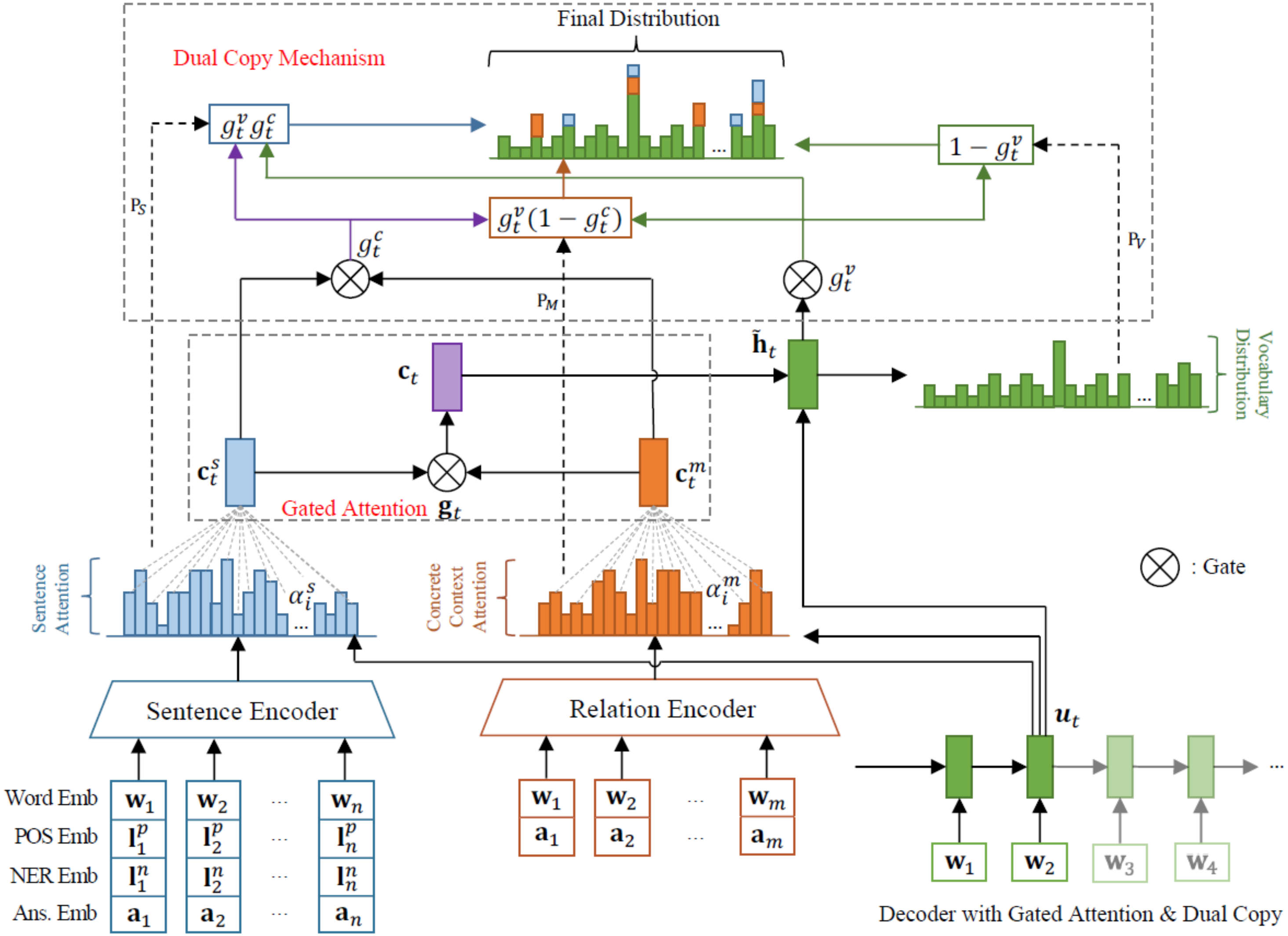
- Step 1: Answer-Relevant Relation Extraction
  - Relation:
    - Triple: (object1, object2, relation)

Sentence: The daily mean temperature in January, the area's coldest month, is 32.6 ˚F (0.3 ˚C) ; however, temperatures usually drop to 10 ˚F (-12 ˚C) several times per winter and reach 50 F (10 ˚C) several days each winter month.
Structured Answer-relevant Relation: • 0.95 (The daily mean temperature in January; is; 32.6 ˚F (0.3 ˚C)) • 0.94 (temperatures; drop; to 10 ˚F (12 ˚C); several times per winter; usually) • 0.90 (temperatures; reach; 50 ˚F)

- N-ary relation selection criterion
  1. Include answer phrase;
  2. Get high confidence score;
  3. Contain maximum non-stop words.

Semantic Fidelity in QG: Methodology

- Step 2: Proposed Framework



Semantic Fidelity in QG: Dataset & Evaluation

- Dataset
  - **S**tanford **Qu**estion **A**nswering **D**ataset (SQuAD)

	Du Split	Zhou Split
# pairs (Train)	74689	86635
# pairs (Dev)	10427	8965
# pairs (Test)	11609	8964
Sentence avg. tokens	32.56	32.72
Question avg. tokens	11.42	11.31

- Evaluation Metrics
  - BiLingual Evaluation Understudy (**BLEU**)
  - Metric for Evaluation of Translation with Explicit Ordering (**METEOR**)
  - Recall-Oriented Understudy for Gisting Evaluation (**ROUGE-L**)

Semantic Fidelity in QG: Experiments

- Main Result

	Du Split (Du et al., 2017)						Zhou Split (Zhou et al., 2017)					
	B1	B2	B3	B4	MET	R-L	B1	B2	B3	B4	MET	R-L
s2s (Du et al., 2017)	43.09	25.96	17.50	12.28	16.62	39.75	-	-	-	-	-	-
NQG++ (Zhou et al., 2017)	-	-	-	-	-	-	-	-	-	13.29	-	-
M2S+cp (Song et al., 2018)	-	-	-	13.98	18.77	42.72	-	-	-	13.91	-	-
s2s+MP+GSA (Zhao et al., 2018)	43.47	28.23	20.40	15.32	19.29	43.91	**44.51**	29.07	21.06	15.82	19.67	44.24
Hybrid model (Sun et al., 2018)	-	-	-	-	-	-	43.02	28.14	20.51	15.64	-	-
ASs2s (Kim et al., 2019)	-	-	-	16.20	19.92	43.96	-	-	-	16.17	-	-
Our model	**45.66**	**30.21**	**21.82**	**16.27**	**20.36**	**44.35**	44.40	**29.48**	**21.54**	**16.37**	**20.68**	**44.73**

- Our model achieves significant improvements over proximity-based answer-aware models (Zhou et al. & Sun et al.).
- Our model is a general one to jointly leverage structured & unstructured knowledge.

Semantic Fidelity in QG: Analysis

- Performance Improvement Analysis

	Hybrid			Our Model		
	BLEU	MET	R-L	BLEU	MET	R-L
0∼5 (36.6% of #)	28.46	21.10	47.33	29.69	22.45	48.15
5∼10 (36.2% of #)	25.91	20.55	43.81	27.08	21.03	44.21
>10 (27.2% of #)	20.60	16.70	37.53	22.05	17.41	38.40

  - Structured relation improves cases where contexts words are far from answer phrase.
  - The improvement increases when distance changes from ‘0~5’ to ‘>10’.

Semantic Fidelity in QG: Case Study

- Case 1: QG with Answer-Relevant Relation

answer

Sentence: The daily mean temperature in January, the area's coldest month, is 32.6 °F (0.3 °C); however, temperatures usually drop to 10 °F (-12 °C) several times per winter and reach 50 F (10 °C) several days each winter month. **Structured Answer-relevant Relation:** (The daily mean temperature in January; is; 32.6 °F (0.3 °C))
Gold Question: What is New York City 's daily January mean temperature in degrees celsius ? **Baseline:** What is the coldest temperature in Celsius ? **Ours:** In degrees Celsius , what is the average temperature in January ?

= wrong context words = correct context words

Semantic Fidelity in QG: Case Study

- Case 2: QG with Diverse Relations

answer

Sentence: In July 1960, NASA Deputy Administrator Hugh L. Dryden announced the Apollo program to industry representatives at a series of Space Task Group conferences.
Relation 1: (Hugh L. Dryden; [is] Deputy Administrator [of]; NASA) **Question 1:** Who was the NASA Deputy Administrator in 1960 ? **Relation 2:** (NASA Deputy Administrator Hugh L. Dryden; announced; the Apollo program to industry representatives at a series of Space Task Group conferences; In July 1960) **Question 2:** Who announced the Apollo program to industry representatives ?

= answer-related relation 1 = answer-related relation 2

Semantic Fidelity in QG: Conclusion

- We propose a novel framework to combine *unstructured sentences and structured answer-relevant relations* for question generation;
- Our proposed framework can be applied as an extension of other question generation model.
- Given multiple facts within one sentence, our model can generate *diverse questions* by verifying the input of relation encoder.

Part II: CTG with Limited Supervision

➢ Stage 1:
Supervised Neural
Controllable Text Generation

→ ?

- Our Reflections
  - Neural approaches rely heavily on training data *quantity* and *quality*.
  - Requiring much human annotation cost.
- How about low-resource setting?

Part II: CTG with Limited Supervision

➢ Stage 1: *Supervised* Neural Controllable Text Generation	→	➢ Stage 2: Neural Controllable Text Generation with *Limited Supervision*

- Problem Setting
  - Training data: $\{(\mathbf{x}^{(m)})\}_{m=1}^{M}$ and $\{(\boldsymbol{y}^{(n)})\}_{n=1}^{N}$, how to get a mapping function from $\boldsymbol{y}$ to $\boldsymbol{x}$?
- Challenges
  - no parallel corpus
- Applications
  - Low-resource NLG
  - Cold start for new projects/applications

Outline

- ❑ Introduction
- ❑ Improving Semantic Fidelity in Text Generation
- ❑ Unsupervised Controllable Generation by Learning from Search
- ❑ Unsupervised Text Attribute Transfer by Iterative Revision
- ❑ Conclusion

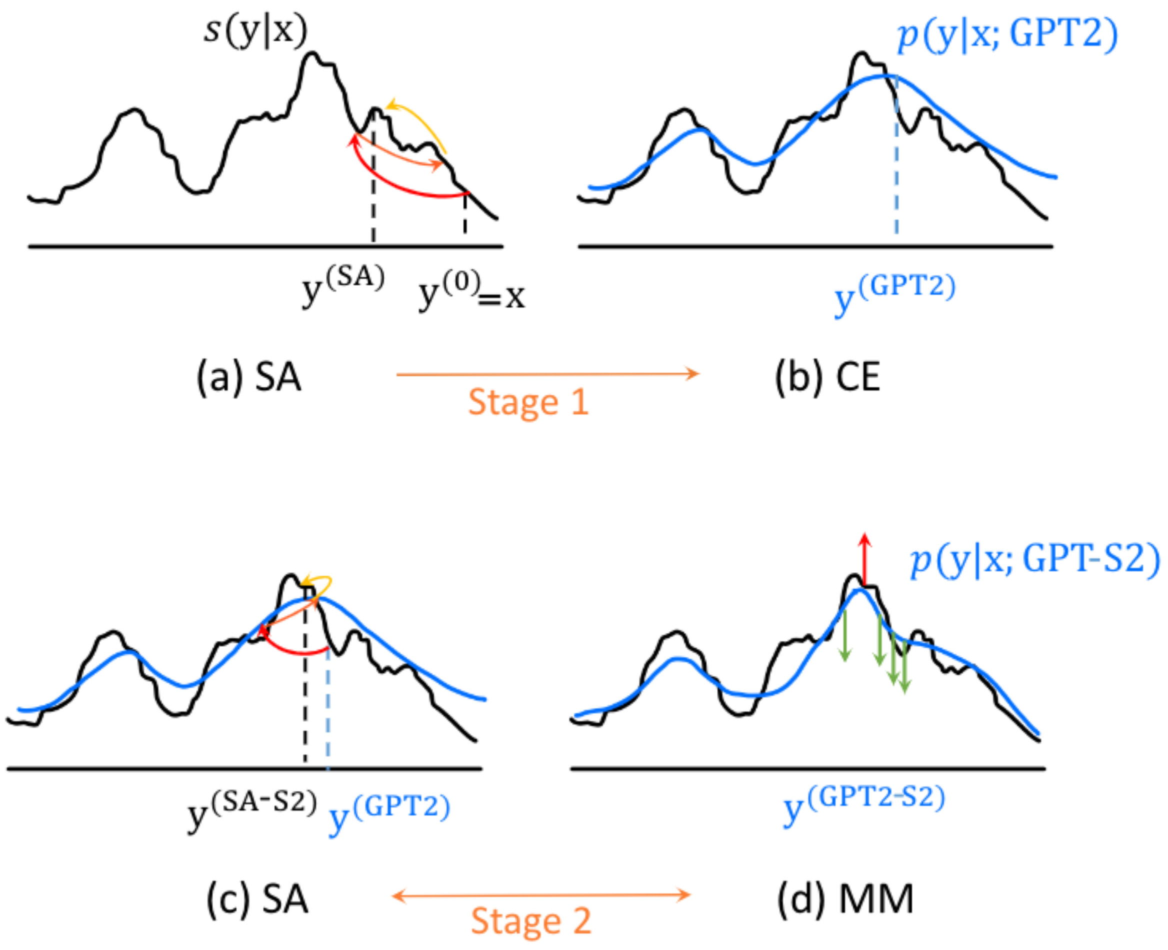
Unsupervised Controllable Generation: Background

- Existing Solution to Unsupervised CTG
  - Search-based approaches
  - RL-based methods

- Drawbacks
  - Slow in inference: ~100 iterations of propose-and-reject
  - Search could be noisy
    - Objective is defined heuristically
    - Local search in a discrete space

UCTG by Learn from Search: Overview

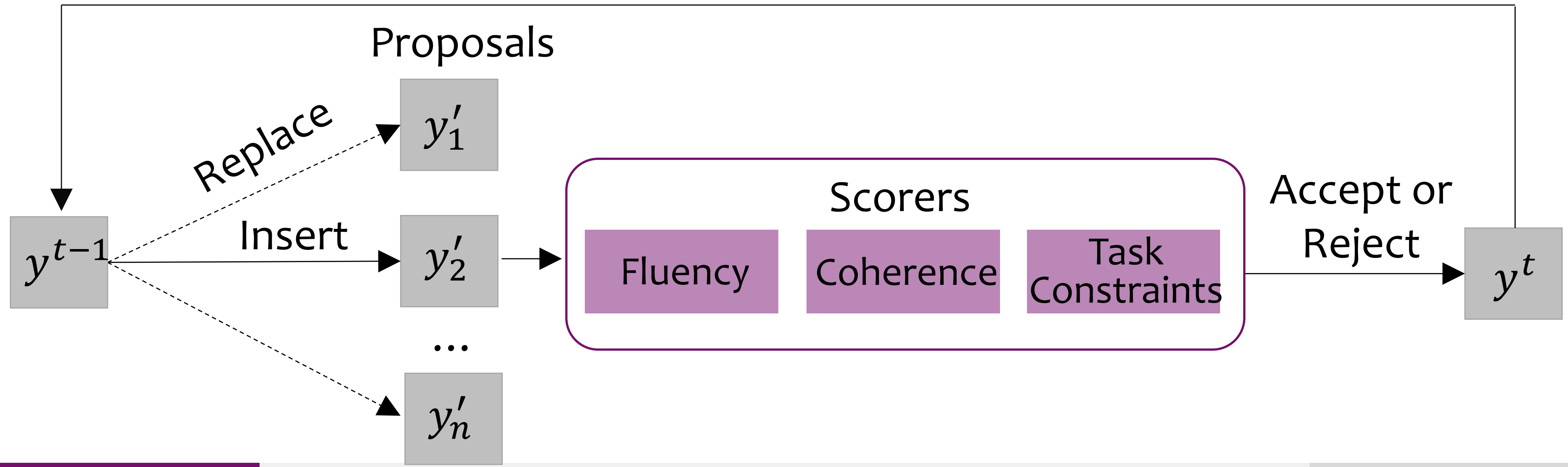
- Our Proposal
  - ➢ Search module
    - Search for target sentences, then learn from the search results.
    - Method: simulated annealing
  - ➢ Learning module
    - Two stages
      1. Word-level cross entropy learning
      2. Seq-level max-margin learning
    - Method: Seq2seq framework
      - Efficient in inference
      - Cross-entropy loss smooths out noise



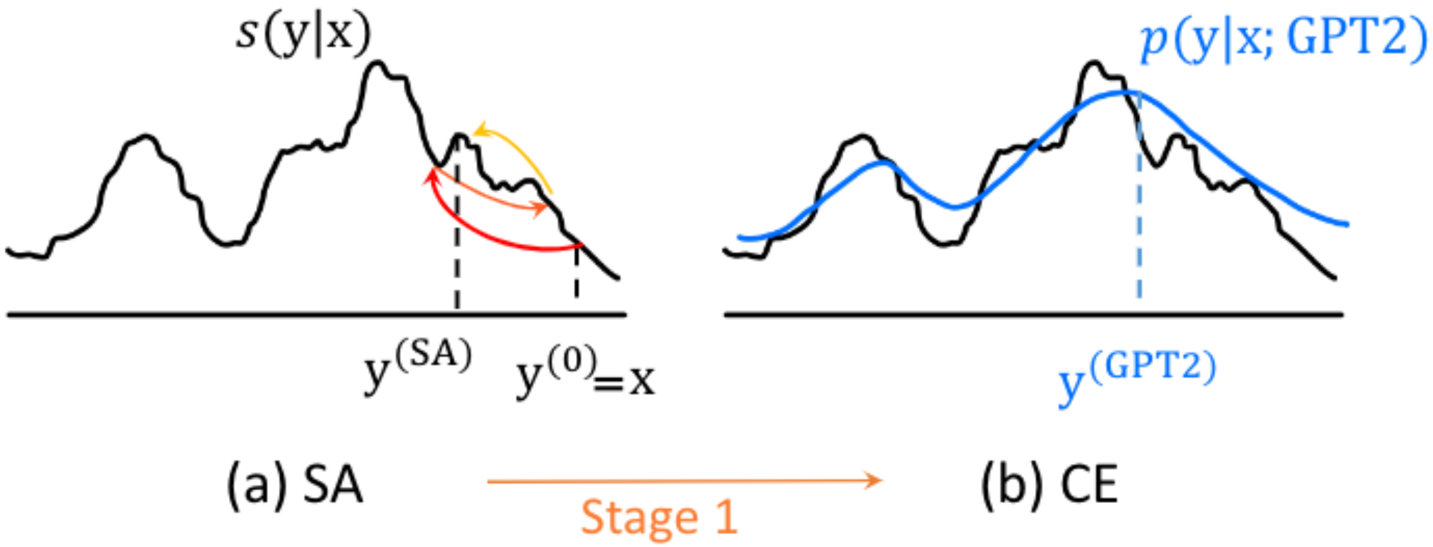
- Simulated Annealing (SA) Search
  - The system performs local search towards a heuristically defined objective (scorers): $s(\mathrm{y}|\mathrm{x}) = s_{\mathrm{lm}}(\mathrm{y}) \cdot s_{\mathrm{semantic}}(\mathrm{x}, \mathrm{y}) \cdot s_{\mathrm{task}}(\mathrm{y}, \cdot)$
  - At every step, the system proposes new sentences by local edits (replace/insert/delete) on the input, and decides to accept or reject according to the scores and current temperature:

$$p(\mathrm{accept}|\mathrm{y}', \mathrm{y}^{(t)}, \mathrm{x}, T) = \min\{1, \exp(\frac{s(\mathrm{y}'|\mathrm{x}) - s(\mathrm{y}^{(t)}|\mathrm{x})}{T})\}$$

  - Search process

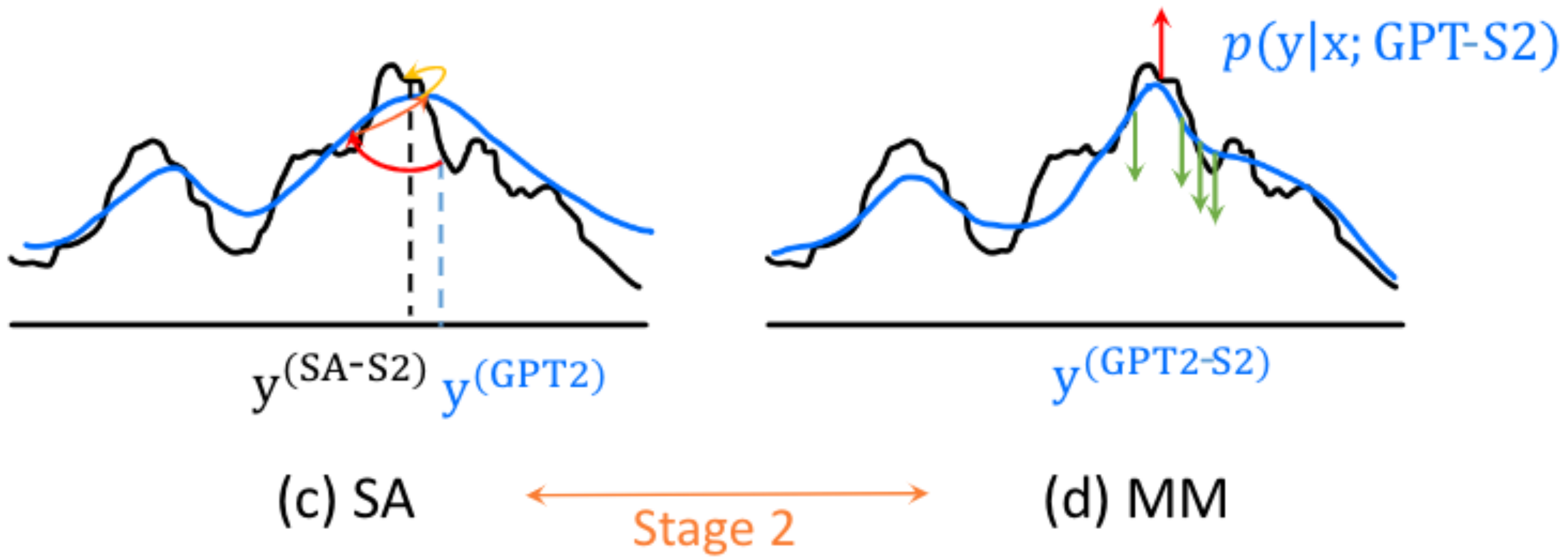


- Stage 1: Word-level Cross-Entropy (CE) Learning
  - Initialize an autoregressive text generator with search results
  - Training objective: $$J_{\mathrm{CE}} = -\sum_{i=1}^{N}\sum_{v\in\mathcal{V}} y_{i,v}^{(\mathrm{SA})} \log p_{i,v}^{(\mathrm{GPT2})}$$
  - CE loss is equivalent to minimize $\mathrm{KL}(\widehat{\boldsymbol{y}}_i^{(\mathrm{SA})} \| \boldsymbol{p}_i^{(\mathrm{GPT2})})$. Due to the asymmetry nature, a GPT2 model can smooth out the noise of stochastic search.



UCTG by Learn from Search: Learning Module

- Stage 2: Sequence-level Max-Margin (MM) Learning
  - Alternate between search and learn to bootstrap the performance
  - Max-margin learning: $$J_{\text{MM}} = \sum_{\text{y}^- \in \widetilde{Y}, \text{y}^- \neq \text{y}^+} \max\{0, E(\text{y}^+) - E(\text{y}^-) + \Delta\}$$
  - Compared with CE, MM corrects the prediction of highly-probable but low-scored samples.



UCTG by Learn from Search: Experiments

- UCTG Task 1: Text Formalization
  - Goal: transduce the formality style of input text
  - Dataset: Grammarly's Yahoo Answers Formality Corpus (GYAFC)
  - Evaluation metrics
    - Fluency: perplexity (PPL)
    - Semantic equivalence: BLEU
    - Formality score

Methods	PPL↓	BLEU	Formality	H-mean	G-mean
Supervised					
LSTM-attn	**23.42**	**69.36**	**87.39**	**77.34**	**77.85**
Unsupervised					
BackTrans	183.7	1.23	31.18	2.37	6.13
StyleEmb	114.6	8.14	12.31	9.80	10.01
MultiDec	187.2	13.29	8.18	10.13	10.42
CrossAlign	44.78	3.34	67.34	6.36	14.99
DelRetrGen	88.52	24.95	56.96	34.70	37.69
Template	197.5	43.45	37.09	40.02	40.14
UnsupMT	55.16	39.28	66.29	49.33	51.02
DualRL	66.96	54.18	58.26	56.15	56.18
TGLS (Ours)	**30.26**	**60.25**	**75.15**	**66.88**	**67.29**

UCTG by Learn from Search: Experiments

- UCTG Task 2: Paraphrase Generation
  - Goal: control the lexical diversity between input and output text
  - Dataset: Quora
  - Evaluation metrics
    - BLEU
    - iBLEU

Methods	iBLEU	BLEU
Supervised		
RL-NN	14.83	20.98
DAGGER	18.88	28.42
GPT2	19.19	26.92
Distant supervised		
Round-Trip MT (GPT2)	11.24	16.33
Round-Trip MT (Transformer)	14.36	20.85
Unsupervised		
VAE	8.16	13.96
CGMH	9.94	15.73
UPSA	12.02	18.18
SA w/ PLM (Ours)	14.52	21.08
TGLS (Ours)	**17.48**	**25.00**

UCTG by Learn from Search: Experiments

- Ablation Study
  - Search < Search + CE
  - 2nd stage: CE < MM
- Efficiency Analysis
  - Training: ~ 2 (SA + Seq2Seq)
  - Inference: 6-10x speedup than SA

Methods	iBLEU	BLEU	Inference Time (sec/sample)
SA	14.52	21.08	5.46
SA+CE	14.97	23.25	0.06
SA+CE+SA	15.41	21.48	2.62
SA+CE+SA+CE	15.70	21.70	0.37
SA+CE+SA+MM (full)	**17.48**	**25.00**	0.43

UCTG by Learn from Search: Conclusion

- We propose a novel search-and-learning framework for unsupervised text generation tasks.
- The proposed framework can be applied to different tasks, if the resemblance between source and target texts can be measured by a heuristically defined scoring function.
- We successfully Incorporate large-scale pretrained language models (GPT2, RoBERTa) into our framework.
- Our model outperforms unsupervised baseline methods on paraphrasing and text formalization.

Part II: CTG with Limited Supervision

➢ Stage 1:
Supervised Neural Controllable Text Generation

→

➢ Stage 2:
Neural Controllable Text Generation with *Limited Supervision*

- How to further improve the efficiency?
  - In-place edit!
  - phrasal replacement instead of word-level replacement!

Outline

UCTG by Iterative Revision: Background

- Text Revision
  - include family of natural language generation tasks, where the source and target sequences share moderate resemblance in *surface form* but differentiate in *attributes*.
- Problem Formulation
  - Given an input sequence X with attribute z, transfer it to another sequence X^* with the target attribute z^*.
- Challenges
  - Sequence-to-sequence transduction is not applicable with non-parallel data
  - Utilizing the transferrable power of pretrained models to text revision

UCTG by Iterative Revision: Overview

- We propose a an iterative in-place editing approach for text revision, named **OREO** (**O**n-the-Fly **RE**presentation **O**ptimization)
  - Training for OREO: Multi-task Fine-tuning
  - Inference of OREO: On-the-Fly Representation Optimization

Informal: Wow , I am very dumb in my observation skills ......

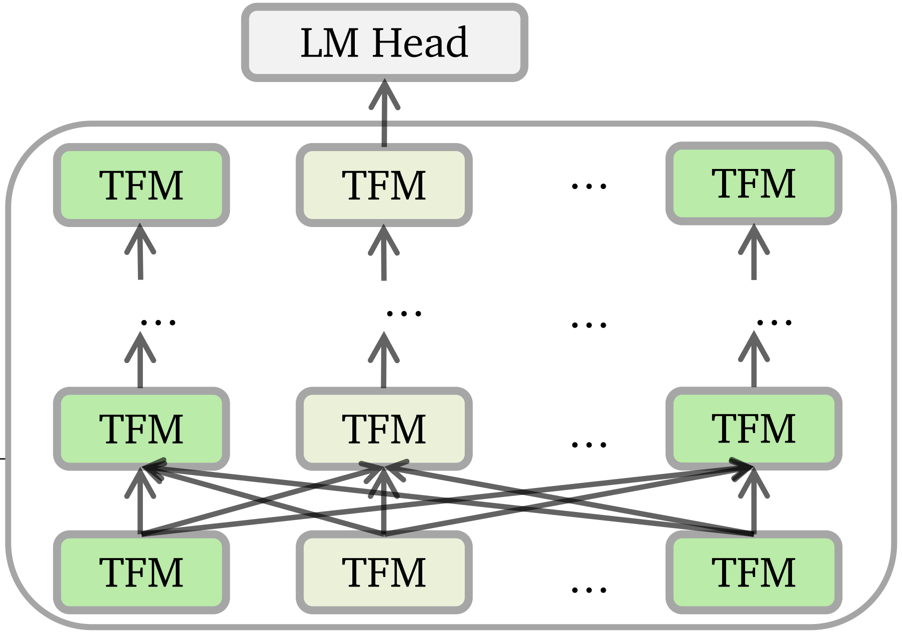
[Delete]: ~~Wow,~~ I am very dumb in my observation skills ......
[Replace]: I do not have good observation skills ......
[Delete]: I do not have good observation skills ~~......~~

Formal: I do not have good observation skills .

UCTG by Iterative Revision: Training Stage in OREO

- Training for OREO: Multi-task Finetuning
  1. Masked language modeling
     - variant-length span replacement
       - Append special token [LM-MASK] to the selected span to a fixed length
       - Set [PAD] as the target token and remove it from output text

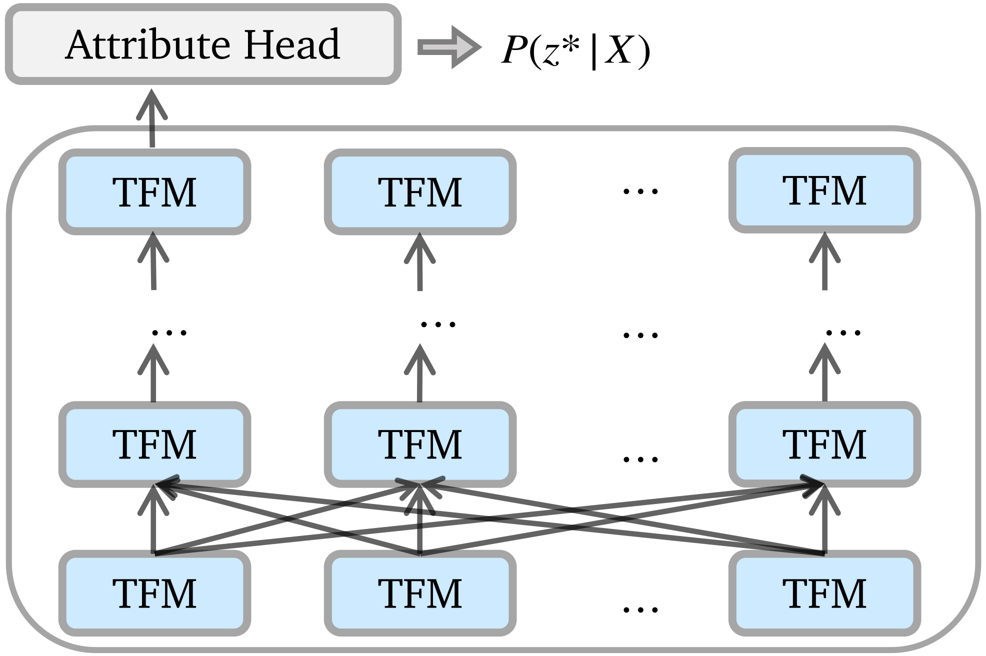
input:	*<u>Good luck</u> to you!*
padded masked input:	*[LM-MASK] [LM-MASK] [LM-MASK] to you!*
target:	*Good luck [PAD]*



UCTG by Iterative Revision: Training Stage in OREO

- Training for OREO: Multi-task Finetuning
  2. Attribute classification
     - Aggregate the representations of [CLS] token from all layers

$$P_\theta(Z|X) = \text{Softmax}(W_{\text{Att}}^T[H_0^0, H_0^1, ..., H_0^L])$$



UCTG by Iterative Revision: Inference Stage in OREO

- Inference of OREO: On-the-Fly Representation Optimization
  1. Span selection
  2. Text revision
     - Step 1: Representation Optimization
     - Setp 2: Span Replacement

Algorithm 1: Text revision with OREO

Input: An input sentence $X^{(0)}$;
Set target attribute z^*, threshold δ, maximum iteration number I;
A fine-tuned RoBERTa with parameters θ, including an attribute head W_{Att} and a LM head W_{LM}

Output: An output sentence X^*

Initialize: $i = 0$, $\zeta^{(0)} = P_\theta(z^*|X^{(0)})$

while $i < I$ and $\zeta^{(i)} < \delta$ **do**

▷ Span selection

Calculate $\zeta^{(i)} = P_\theta(z^*|X^{(i)})$ and $\mathcal{L}$ (4)

Calculate $a^{(i)}$ (6) and select $t, N = \underset{t,N}{\operatorname{argmax}}\, a^{(i)}_{t:t+N}$

▷ Representation optimization

Insert K `[LM-MASK]`s after $X^{(i)}_{t:t+N}$, then we have $X'^{(i)}$ as the input of RoBERTa at the next step

Calculate $H^{(i)}$, $P_{W_{\text{Att}}}(z^*|H^{(i)})$ and $\mathcal{L}'$ (4)

Update $H^{(i+1)}$ with $\nabla_{H^{(i)}}\mathcal{L}'$ (3)

▷ Span replacement

Replace the selected span $X'^{(i)}_{t:t+N}$ with `[LM-MASK]`s

$X^{(i+1)}_{\backslash t:t+N+K} = X'^{(i)}_{\backslash t:t+N+K}$

▷ The unselected part keep fixed

Infill a new span

$X^{(i+1)}_{t:t+N+K} = \underset{X_{t:t+N+K}}{\operatorname{argmax}}\, P_{W_{\text{LM}}}(X_{t:t+N+K}|H^{(i+1)}_{\backslash t:t+N+K})$

▷ Approximate by greedy decoding

Remove the `[PAD]` tokens in the new span, then we have $X^{(i+1)}$

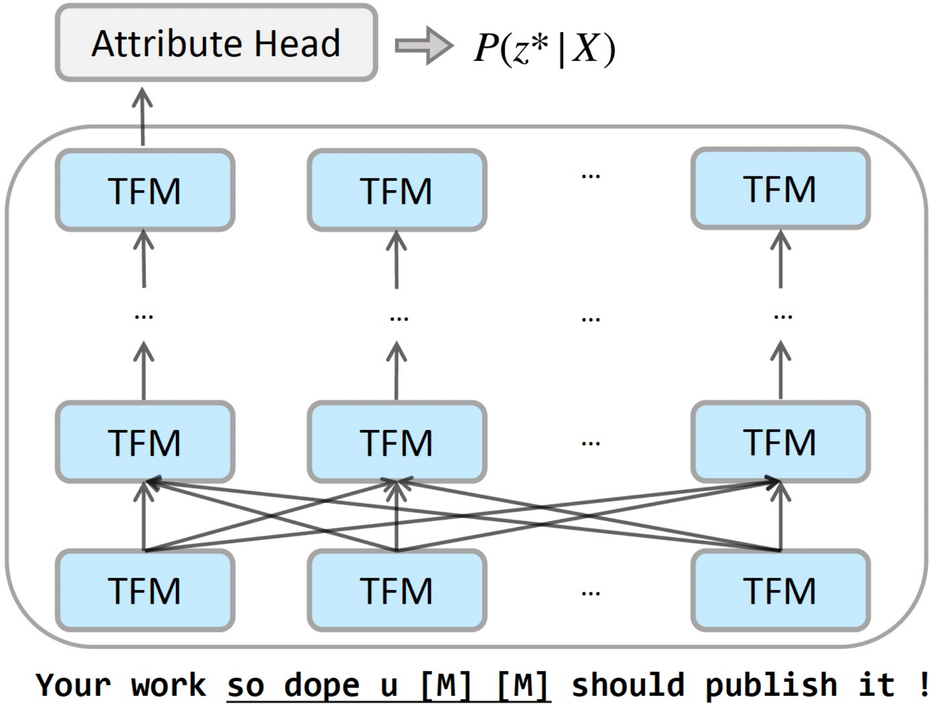
Return: $X^* = X^{(j)}$, where $j = \underset{j}{\operatorname{argmax}}\, \zeta^{(j)}$

1. Span Selection
   - Strategy: gradient-guided selection

$$\mathcal{L} = -\log P_{W_{\text{Att}}}(z^*|H^{(i)})$$

$$a_t^{(i)} = \|\nabla_{H_t^{0(i)}} \mathcal{L}\|$$

$$a_{t:t+N}^{(i)} = \frac{\sum_{n=1}^{N} a_{t+n}^{(i)}}{N+c}$$



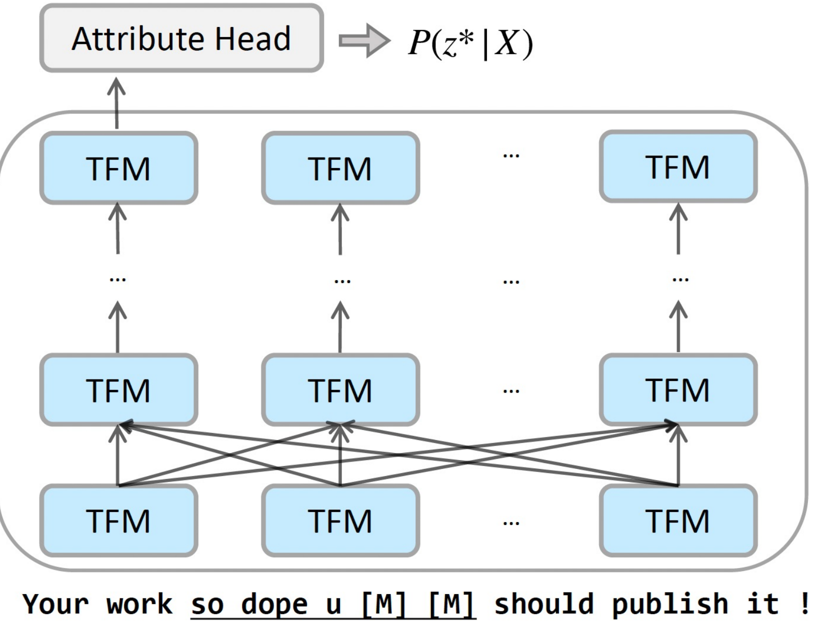
Step 1(a)

   - Advantages
     - Agnostic to revision algorithm
     - Allow use to insert [LM-MASK] tokens in advance
     - Enable human-in-the-loop generation

2. Text Revision

- Step 1: Representation optimization

$$\mathcal{L} = -\log P_{W_{\text{Att}}}(z^*|H^{(i)})$$



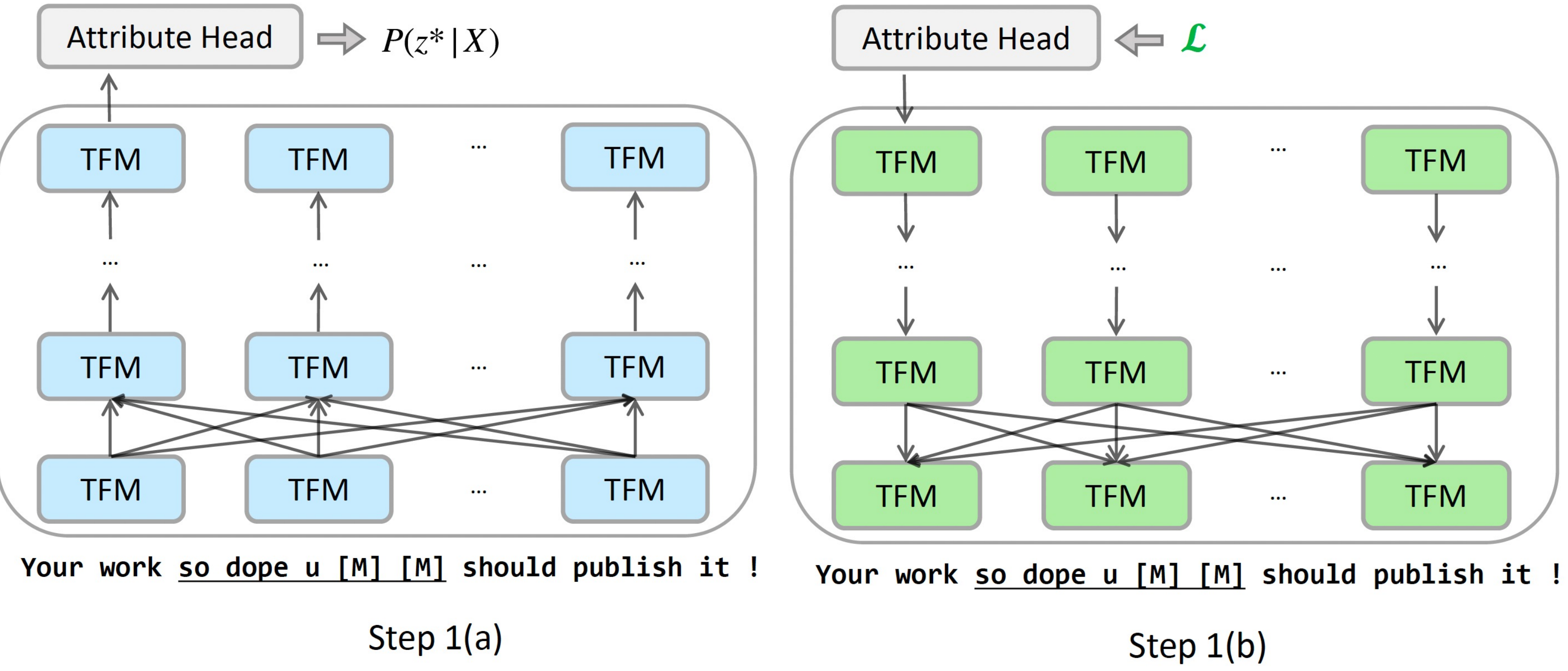
Step 1(a)

UCTG by Iterative Revision: Inference Stage in OREO

2. Text Revision

- Step 1: Representation optimization

$$\mathcal{L} = -\log P_{W_{\text{Att}}}(z^* | H^{(i)}) \quad \Rightarrow \quad H^{(i+1)} = H^{(i)} - \lambda \frac{\nabla_{H^{(i)}} \mathcal{L}}{\|\nabla_{H^{(i)}} \mathcal{L}\|}$$

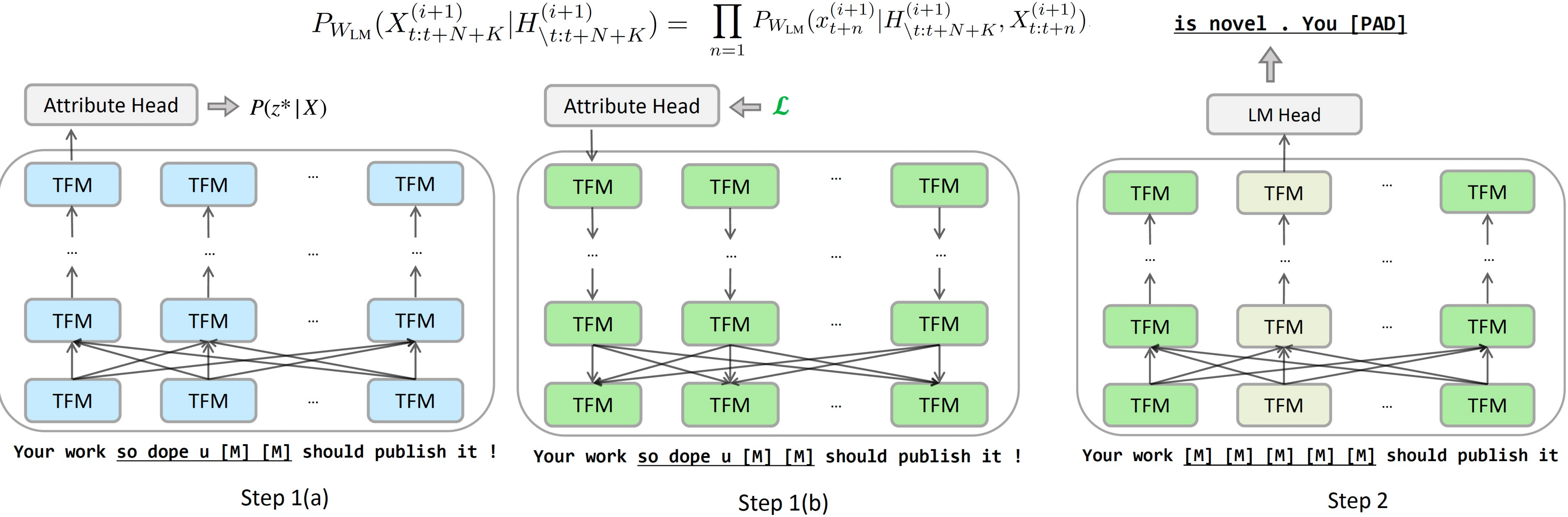


UCTG by Iterative Revision: Inference Stage in OREO

2. Text Revision

- Step 2: New span replacement

$$P_{W_{\mathrm{LM}}}(X^{(i+1)}_{t:t+N+K} | H^{(i+1)}_{\backslash t:t+N+K}) = \prod_{n=1}^{N+K} P_{W_{\mathrm{LM}}}(x^{(i+1)}_{t+n} | H^{(i+1)}_{\backslash t:t+N+K}, X^{(i+1)}_{t:t+n})$$



UCTG by Iterative Revision: Experiments

- UCTG Task 1: Text Simplification
  - Goal: revise the complex text into simpler language
  - Dataset: Newsela-Turk
  - Evaluation metrics: SARI, FKGL, SLen

Methods	SARI	Add	Keep	Delete	FKGL↓	SLen
Supervised						
Complex (input)	22.3	0.0	67.0	0.0	12.8	23.2
$\text{Transformer}_{\text{BERT}}$	36.0	3.3	54.9	49.8	**8.9**	16.1
EditNTS	37.4	1.6	61.0	49.6	9.5	16.9
Hybird-NG	38.2	2.8	57.0	54.8	10.7	21.6
ControlTextSimp	41.0	**3.4**	63.1	56.6	11.5	22.2
Unsupervised						
UNTS	39.9	1.5	60.5	57.7	11.2	22.0
OREO	**45.2**	2.3	**69.4**	**64.0**	11.4	**23.5**

UCTG by Iterative Revision: Experiments

- UCTG Task 2: Text Formalization
  - Goal: transduce the formality style of input text
  - Dataset: Grammarly's Yahoo Answers Formality Corpus (GYAFC)
  - Evaluation metrics
    - Fluency: perplexity (PPL)
    - Semantic equivalence: BLEU
    - Formality score

Methods†	BLEU	Formality	H-mean	G-mean
Human reference	100.0	95.20	97.49	97.52
CrossAlign	4.77	75.9	8.98	19.03
StyleEmbded	8.71	28.3	13.32	15.70
MultiDec	14.04	21.32	16.93	17.30
UnsupMT	37.36	76.88	50.28	53.59
MASKER	47.73	58.86	52.71	53.00
OREO (Ours)	**57.63**	**80.71**	**67.24**	**68.20**

UCTG by Iterative Revision: Experiments

- Human Evaluation
  - Formality
  - Coherency
  - Fluency

	Formality	Coherency	Fluency
MASKER	2.74	2.94	3.31
OREO	3.42	3.33	3.41
Human	**3.69**	**3.67**	**3.78**

- Ablation Study
  1) recomputing all hidden states when infilling span
  2) updating the hidden states with Gaussian noise
  3) without updating the hidden states
  4) randomly selecting span

	BLEU	Formality	H-mean	G-mean
Full	**57.63**	**80.71**	**67.24**	**68.20**
(1)	55.50	69.67	61.78	62.18
(2)	56.55	69.14	62.21	62.53
(3)	56.47	67.94	61.68	61.94
(4)	45.30	55.03	49.69	49.93

UCTG by Iterative Revision: Case Study

- Example cases

Complex Input	OREO
kraft announced monday that it will remove artificial food coloring, notably yellow no. 5 and yellow no. 6 dyes, from its iconic product by january 2016.	kraft announced monday that it will stop using some of the chemicals, such as yellow no. 5 and yellow no. 6 dyes, from its iconic product by january 2016.
still, recent trends suggest seattle is doing a better job of holding onto those kids, according to sightline institute, a think tank based in seattle.	still, recent studies suggest seattle is doing a better job of holding onto those kids, according to sightline institute, a group that studies people in seattle.
Informal Input	**OREO**
tell him, and it wouldn't seem psycho cuz u have kno each other for a long time	Tell him, and it will not even seem awkward you two have known each other for a long time

UCTG by Iterative Revision: Case Study

- Human-in-the-loop
  - Let user to decide the span to be edited and ask OREO to revise

informal:	The same guy you wanna be in a relationship with?
OREO:	The same guy you want to be in a relationship with?
1_{st} **Edit**:	The very same guy you want to be in a relationship with?
2_{nd} **Edit**:	Is this the same guy you want to be in a relationship with?
3_{rd} **Edit**:	Is this the same person that you want to be in a relationship with?

informal:	Then see if shes open for a dinner & a movie.
OREO:	Then see if she will accompany you for a dinner or perhaps a movie.
1_{st} **Edit**:	Inquire her if she will accompany you for a dinner or perhaps a movie.

UCTG by Iterative Revision: Conclusion

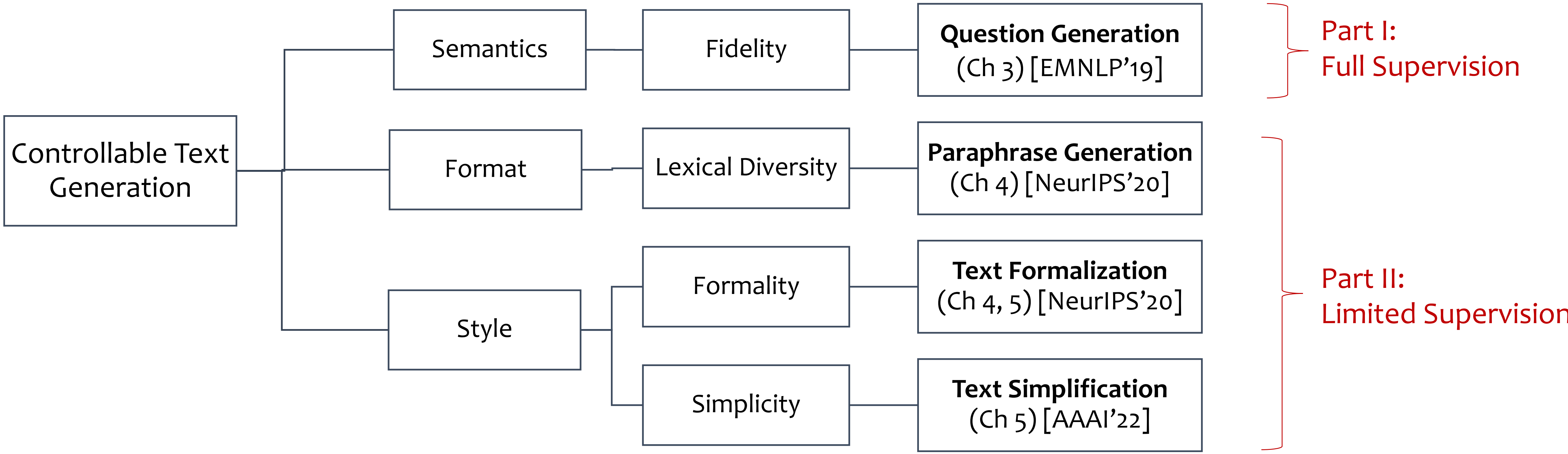
- We propose an efficient mask-and-infill method with on-the-fly optimized representation for text revision;
- Our approach has strong performance on text formalization dataset GYAFC-fr and text simplification dataset Newsela-Mturk;
- Our editing system can also produce meaningful revisions when interacting with human beings.

Outline

- ❑ Introduction
- ❑ Improving Semantic Fidelity in Text Generation
- ❑ Unsupervised Controllable Generation by Learning from Search
- ❑ Unsupervised Text Attribute Transfer by Iterative Revision
- ❑ Conclusion

Conclusion

- We investigate the problem of neural controllable text generation across different dimensions of control factors.
- We explore the setting of generation task from bipartite settings: supervised and unsupervised learning.
- All the studies consistently approach a general and efficient solution to NCTG.



Publications (as first author)

1. **Jingjing Li**, Yifan Gao, Lidong Bing, Irwin King, Michael R. Lyu, Improving Question Generation With to the Point Context, in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP 2019), pages 3216-3226, 2019.

2. **Jingjing Li**, Zichao Li, Lili Mou, Xin Jiang, Michael R. Lyu, and Irwin King. Unsupervised Text Generation by Learning from Search, in Advances in Neural Information Processing Systems (NeurIPS 2020), volume 33, pages 10820–10831, 2020.

3. **Jingjing Li**, Zichao Li, Tao Ge, Irwin King, and Michael R. Lyu. Text Revision by on-the-fly Representation Optimization, in Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI 2022), volume 36, pages 10956–10964, 2022.

Full Publication List

[1] Contrastive Learning for Concept Relatedness Estimation, AAAI 2023.

[2] Text Revision by On-the-Fly Representation Optimization, AAAI 2022 & ACL 2022.

[3] A Unified Strategy for Multilingual Grammatical Error Correction with Pre-trained Cross-Lingual Language Model, IJCAI 2022.

[4] Open-Retrieval Conversational Machine Reading, ArXiv 2021.

[5] Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading, EMNLP 2020.

[6] Unsupervised Text Generation by Learning from Search, NeurIPS 2020.

[7] Counting the Frequency of Time-Constrained Serial Episodes in a Streaming Sequence, Information Sciences 2019.

[8] Improving Question Generation with To the Point Context, EMNLP 2019.

Thank you!