Insertion-base Text Generation

Yizhe Zhang
@yizzhang at NLP, MSR AI

This Talk

- Sentence-level Insertion Generation
  - [ACL 2020] INSET: Sentence Infilling with INter-SEntential Transformer
  - https://arxiv.org/abs/1911.03892
  - Semantic-aware sentence insertion

- Word-level Insertion Generation
  - [under submission] POINTER: Constrained Text Generation via Insertion-based Generative Pre-training
  - https://arxiv.org/abs/2005.00558
  - Non-autoregressive generation from lexical constraints

- Orthogonal to each other

INSET: Sentence Infilling with INter-SEntential Transformer

Yichen Huang∗, Yizhe Zhang∗, Oussama Elachaqar, Yu Cheng

Sentence Infilling (w/ and w/o hints)

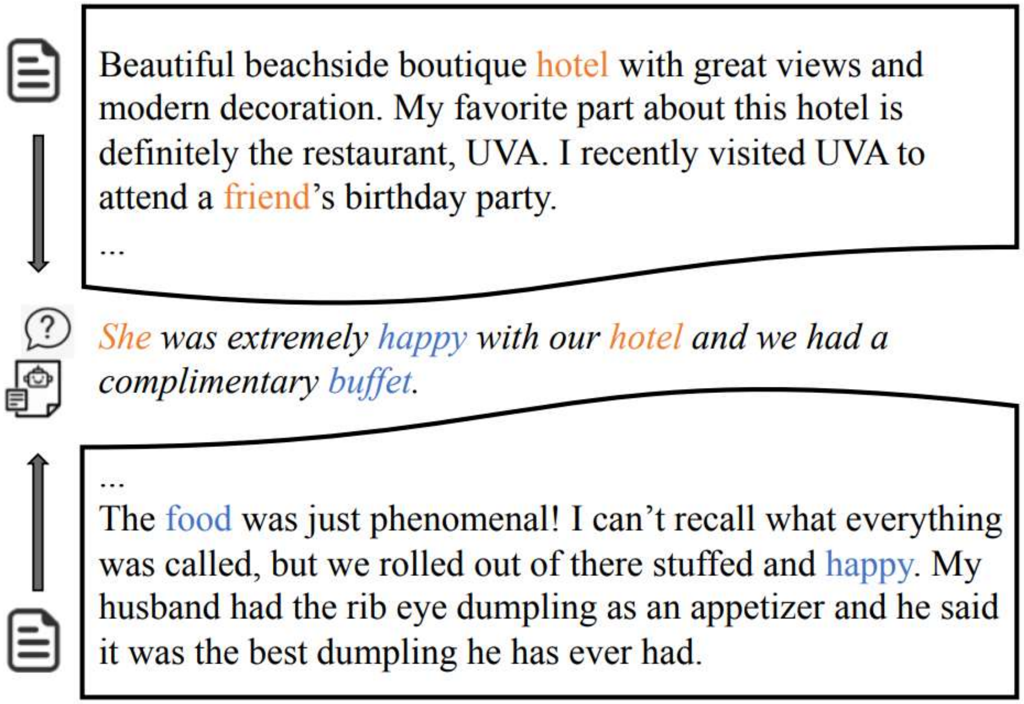


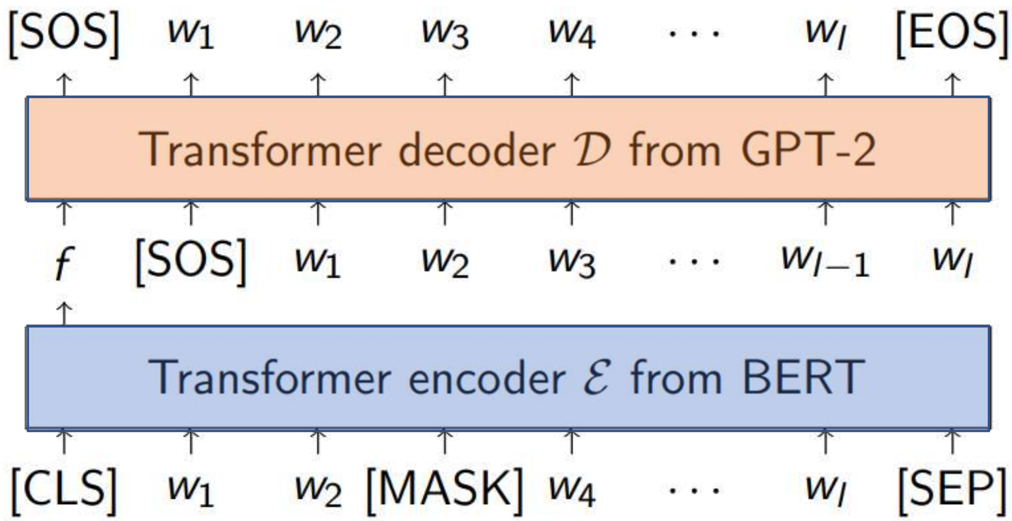
Figure: Sentence infilling: generating an intermediate sentence that provides a smooth semantic transition from the preceding to the following context. This example is generated by our model on the TripAdvisor dataset.

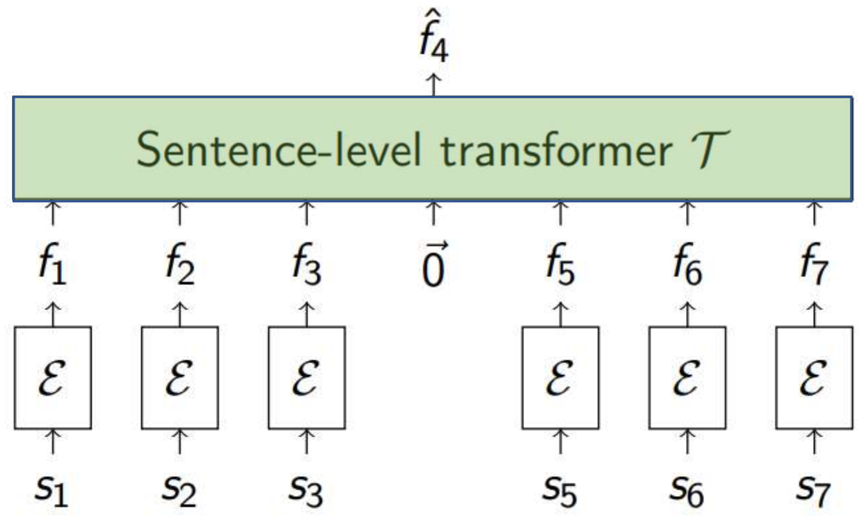
It is not necessary for the generated sentence to be close to the ground truth.

Possible scenarios

- **Document auto-completion:** suggesting missing bridging sentences in the surrounding context
- **Collaborative document writing:** unifying different writing styles from multiple authors
- **Note expansion**: extending a set of keywords to a full sentence, leveraging the surrounding context

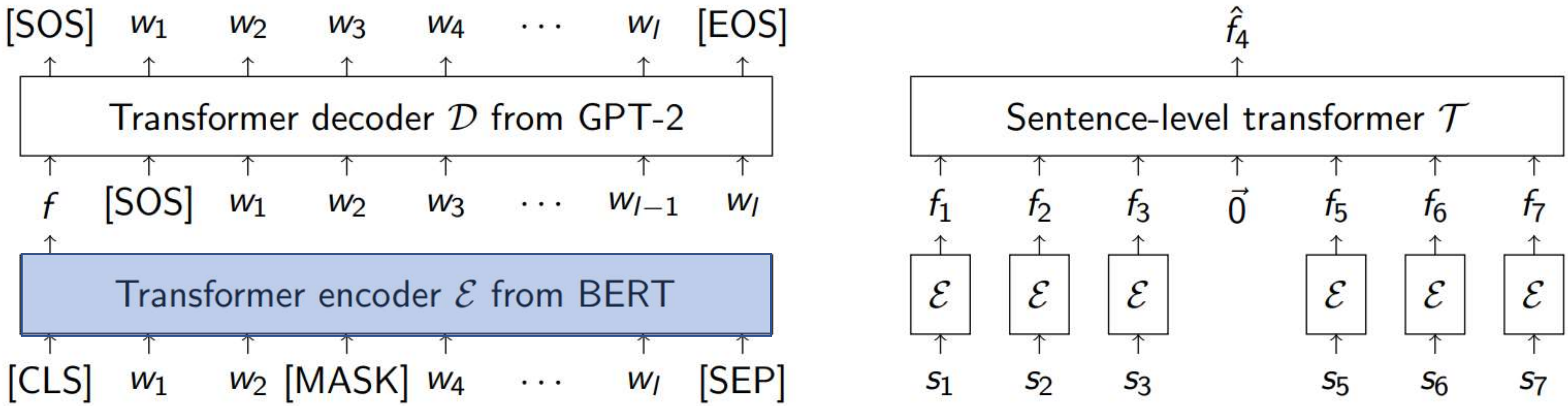
INSET: INter-SEntential Transformer





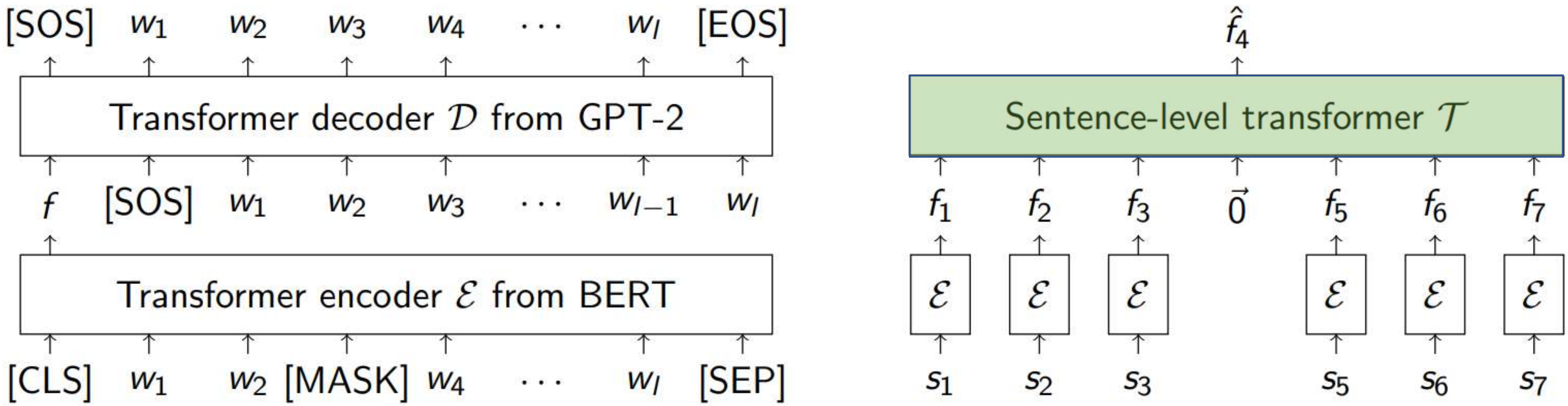
- ***Understanding*** (BERT-like encoder)
- ***planning*** (sentence-level Transformer)
- ***generation*** (GPT-like decoder)

INSET: INter-SEntential Transformer



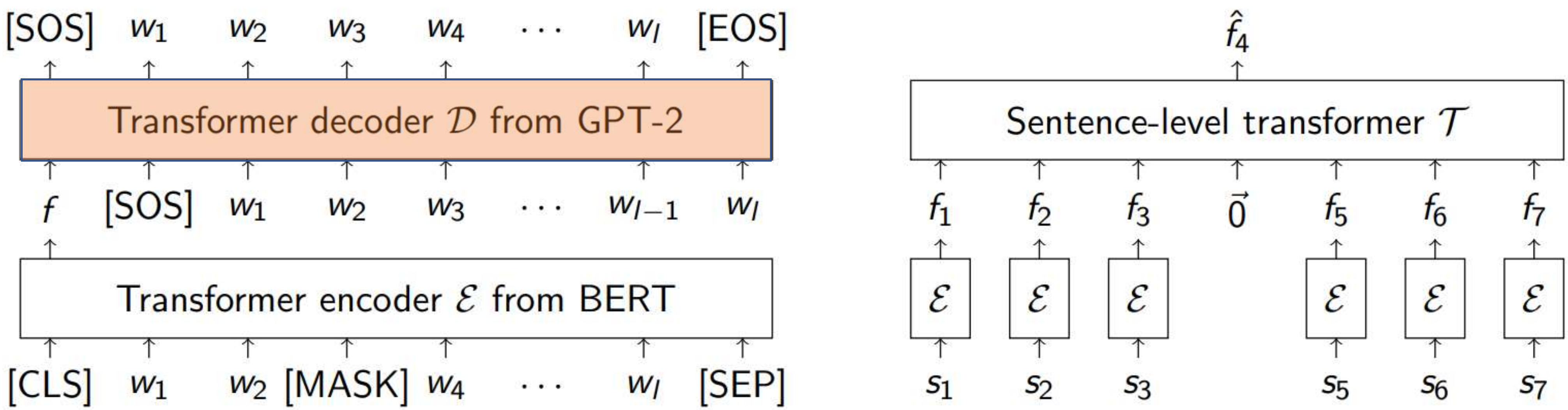
- ***Understanding*** (BERT-like encoder) : BERT-base size 110M
- A BERT-based encoder to map each sentence to the latent semantic space (768 dimension vector)

INSET: INter-SEntential Transformer



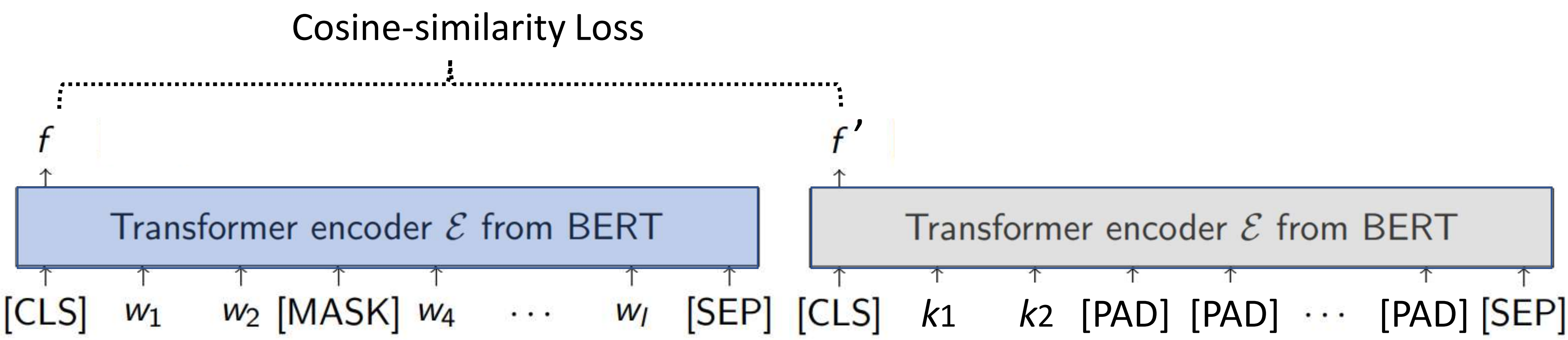
- ***planning*** (sentence-level Transformer) : BERT-base size, 110M
- A sentence-level semantic planner to infer the missing information that can bridge the semantics of preceding and following context.

INSET: INter-SEntential Transformer



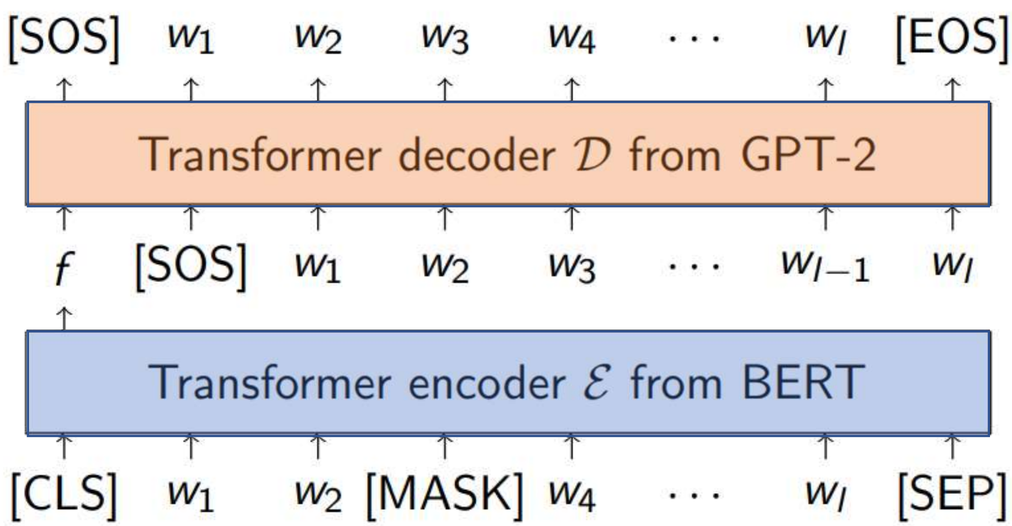
- ***generation*** (GPT-like decoder) : GPT-small size 117M
- A GPT-based generator (decoder) to map semantic features back to the text domain.

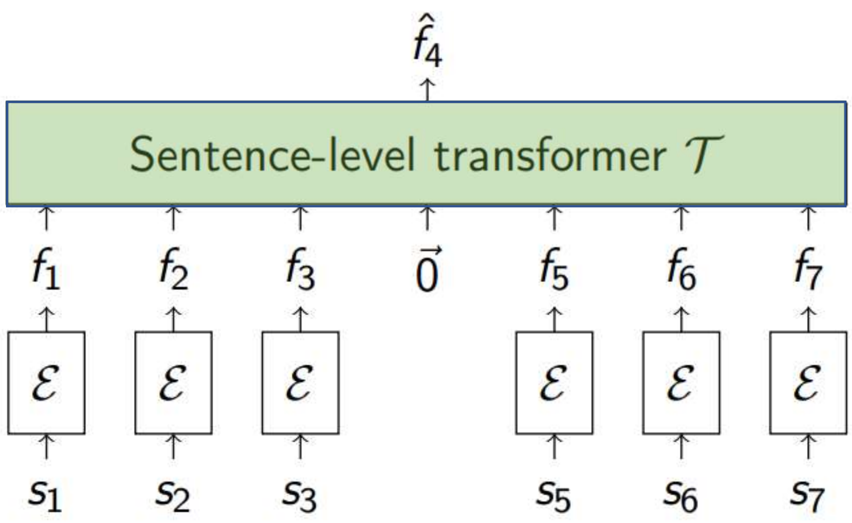
INSET: INter-SEntential Transformer (w/ keywords hint)



- ***Constraint feature encoder*** (BERT-like encoder) : BERT-base size 110M
- Distillation-like objective
- **Teacher:** fixed sentence encoder
- **Student:** constraint feature encoder with *no position embedding*.

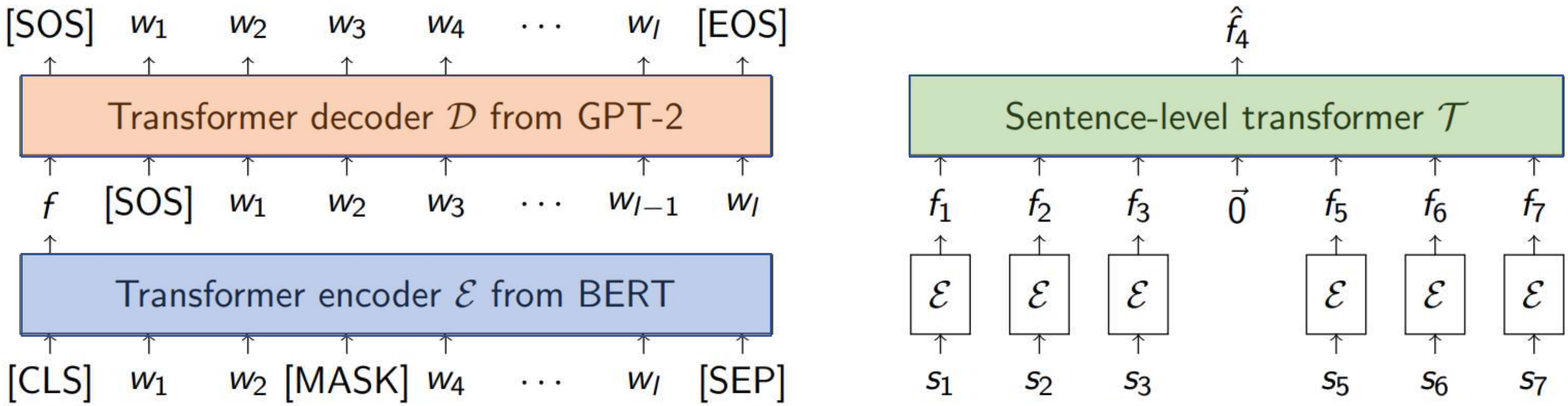
Model Training





- Train a denoising auto-encoder (DAE) for the encoder and decoder
- Train a sentence-level transformer for the planner
- Joint training is possible.

Advantages



- Good at capturing **long-term/semantic-level** inter-sentential correlation.
- Enable leveraging the **pre-trained models** (BERT, GPT-2)
- Can handle **long** text. Significant **reduction of computation** (time/memory)

Evaluation & Baseline

- **Evaluation:** 7 sentences, predict the 4th sentence. (w/ w/o keyword hints)

- **Dataset:**
  - TripAdvisor
    - One of the widely used datasets.
    - (Train/dev/test) = (1.1M / 62K / 533)
  - Recipe
    - Time-ordered procedure. Ideal for evaluating the inter-sentential planning/reasoning.
    - (Train/dev/test) = (1.1M / 56K / 500)

(Train/dev/test) = (1.1M/62K/533)

Metrics & Baseline

- **Evaluation:**
  - **Relevance:** Standard machine translation metrics, including BLEU, NIST, METEOR.
  - **Diversity**: Entropy (ENT-n) and Distinct score (DIST-n).
  - Human evaluation.

- **Baseline**
  - Text infilling (W. Zhu, Z. Hu, and E. Xing, Text Infilling, arXiv:1901.00158, 2019.)

Sentence representation learning

	example 1
A	The pool area was nice and sunbathing was great.
-	The pool area was nice and staff was great.
-	The pool area staff was nice and very helpful.
-	Front desk staff were very helpful and friendly.
B	Front desk staff were very nice and helpful.
	example 2
A	The service was attentive and we had the best food in town.
-	The service was attentive and we had a great room with plenty of food.
-	The room was spacious with good service and we had a queen bed.
-	The room was very spacious with queen beds.
B	The room was very spacious with 2 queen beds.

Table: Sentence interpolation. "A" and "B" are two sentences in the test set. The intermediate sentences are generated by interpolating between the latent-space representations of A and B.

Automatic evaluation

Dataset	Method	NIST N-2	NIST N-4	BLEU B-2	BLEU B-4	MET-EOR	Entropy E-4	Dist D-1	Dist D-2	Length
Trip	*Without keyword constraints:*									
	baseline[1]	0.54	0.54	4.29%	0.54%	5.85%	3.10	1.32%	2.23%	6.97
	INSET (full context)	**1.23**	**1.23**	**6.08%**	**0.96%**	**7.04%**	**8.13**	**16.30%**	**46.64%**	10.70
	INSET (less context)	1.02	1.02	4.74%	0.51%	5.83%	7.85	12.98%	41.39%	11.26
	With keyword constraints:									
	INSET (w/ context)	**3.09**	**3.15**	**20.14%**	**6.57%**	**16.48%**	**8.34**	**22.61%**	**63.60%**	11.23
	INSET (w/o context)	3.00	3.04	19.47%	6.07%	16.00%	8.16	20.51%	57.41%	11.12
	ground truth (human)	-	-	-	-	-	8.40	33.96%	79.84%	11.36
Recipe	baseline	0.67	0.68	3.91%	0.88%	5.23%	3.12	0.37%	0.47%	15.32
	INSET (ours)	**1.36**	**1.37**	**7.24%**	**1.33%**	**7.07%**	**7.99**	**20.12%**	**55.13%**	9.63
	ground truth (human)	-	-	-	-	-	8.22	29.21%	74.97%	10.55

Table: Automatic evaluation. "w/ context" indicates that the generation is based on both keywords and context. "w/o context" indicates that the generation is only based on keywords but not context. "Length" stand for the average generation length.

Human evaluation

system A	system B	criterion	prefer A (%)	same (%)	prefer B (%)
INSET (ours)	baseline	coherence fluency informativeness	**54.16** **43.38** **53.48**	13.76 26.98 18.79	32.07 29.64 27.72
INSET (ours)	ground truth	coherence fluency informativeness	27.87 21.78 27.49	15.69 31.38 21.92	**56.44** **46.84** **50.59**
INSET w/ keywords w/ context	ground truth	coherence fluency informativeness	18.50 17.82 20.54	23.45 29.78 26.13	**58.04** **52.39** **53.33**
INSET w/ keywords w/ context	INSET w/ keywords w/o context	coherence fluency informativeness	**37.71** 36.16 35.93	37.62 **37.87** **39.86**	24.68 25.97 24.21
INSET w/ keywords w/ context	INSET w/o keywords w/ context	coherence fluency informativeness	34.97 29.30 31.73	17.06 28.04 23.24	**47.97** **42.65** **45.03**

Table: Human evaluation. "w/(w/o) keywords" and "w/(w/o) context" indicate whether the generation is based on keywords and context, respectively. All numbers are percentages.

Generated examples

	example from TripAdvisor dataset	example from TripAdvisor dataset
preceding context	It was such a pleasure to see somthing new every night. It was not very crowded so we were able to get great seats at either the pool or the beach. The VIP sevice was great for dinner reservations and pillow service.	The walls are very thin. Since this is a family vacation type of hotel, people are up at the pool/bbq area/hallways during all hours of the night. Do not stay here if you need a quite night of sleep.
following context	Enjoyed the shrimp coctail and seafood salad delivered to us while enjoying the pool. All of us would not want to stay at another resort and are planning to go back again. Enjoy and Hola!Karen and FriendsMilford, CT	You have to take multiple elevators to go all the way to the 5th floor. My other complaint is that the hotel staff seemed a bit unprofessional. Not what I'm used to when I stay at Marriot properties.
ground truth	We did bring a lot of $1 for tipping and of course the service stepped up a notch more.	Also, the elevator situation is weird.
baseline	The staff was friendly and helpful.	The rooms are very clean and well kept.
INSET	The buffet dinner was amazing and we had the best food in the resort.	There is only one elevator block in the hotel.
+ keywords	$, service	elevator, situation
INSET (w/ keywords)	Service fee for the buffet dinner was $5.00 and we paid $5.00 extra for food service.	The elevator situation is extremely frustrating.

Table: Examples generated by our model and the baseline.

Summary

- We study the task of sentence infilling, which is analogous to the masked language modeling task for (pre-)training BERT, but ***at sentence-level***.

- INSET is designed to handle ***long-range inter-sentential*** correlation.

- INSET ***decouple*** three aspects of the task (understanding, planning, and generation).

POINTER: Constrained Text Generation via Insertion-based Generative Pre-training

Yizhe Zhang *, Guoyin Wang *, Chunyuan Li, Zhe Gan, Chris Brockett, Bill Dolan

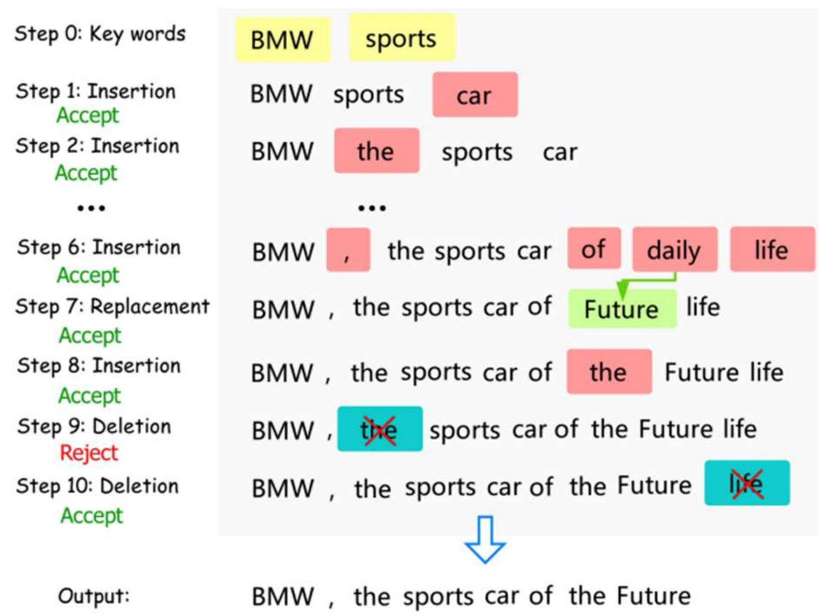
Hard-constrained Text Generation

- Generating sentence from keywords/key-phrases
- **Possible scenarios**: title generation, note expansion, story generation
- *Hard-constrained Text Generation*: **all** the predefined lexical constraints need to be present **in the given order**.

Previous works

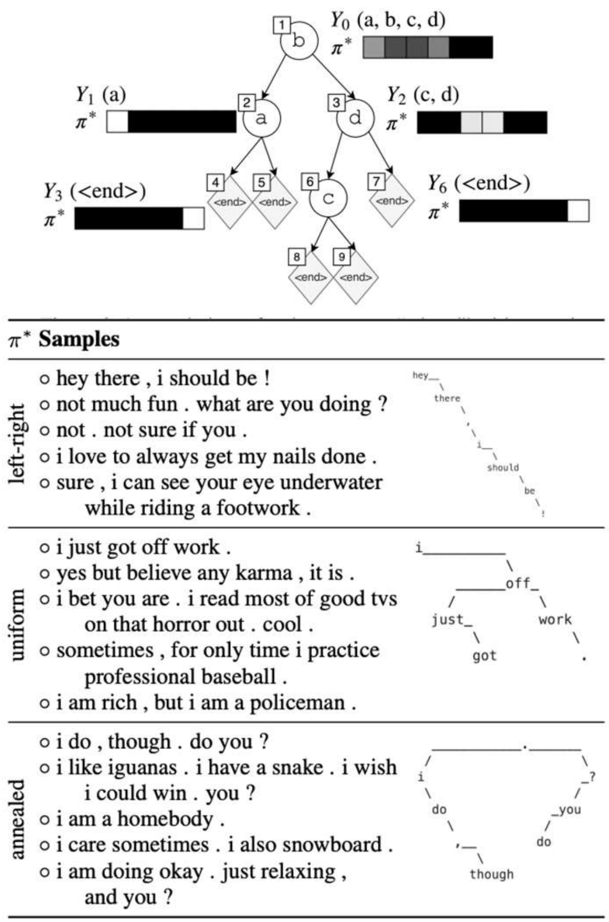
- **CGMH**: Constrained Sentence Generation by Metropolis-Hastings Sampling (AAAI 2019)
  - sampling-based approach
  - Words in a random position are either inserted, deleted or updated under a Metropolis-Hastings-like scheme.

- **Issue**:
  - Can easily get stuck into local optimal. (e.g. “Hong Kong”)
  - Slow inference.



Previous works

- **NMSTG**: Non-Monotonic Sequential Text Generation (ICML 2019)
  - Non-autoregressive generation approach
  - A tree-based text generation scheme: the model recursively generates words to its left and right, yielding a binary tree.
- **Issue**:
  - Sentence-to-Tree structure is a one-to-many mapping
  - Time complexity for inference is the same as autoregressive approach.



Microsoft Research AI

POINTER (PrOgressive INsertionbased TransformER)

Stage	Generated text sequence
0 (X^0)	sources sees structure perfectly
1 (X^1)	sources company sees change structure perfectly legal
2 (X^2)	sources suggested company sees reason change tax structure which perfectly legal .
3 (X^3)	my sources have suggested the company sees no reason to change its tax structure , which are perfectly legal .
4 (X^4)	my sources have suggested the company sees no reason to change its tax structure , which are perfectly legal .

At each stage, the algorithm inserts tokens **progressively:**

- From the original lexical constraints (X_0), first generates *high-level words* (e.g., informative nouns, verbs and adjectives)
- Then adding the *less informative words* (e.g. pronouns and prepositions)
- This process iterates until the generation is *converged* (no more edit).

POINTER (PrOgressive INsertionbased TransformER)

Stage	Generated text sequence
0 (X^0)	sources sees structure perfectly
1 (X^1)	sources company sees change structure perfectly legal
2 (X^2)	sources suggested company sees reason change tax structure which perfectly legal .
3 (X^3)	my sources have suggested the company sees no reason to change its tax structure , which are perfectly legal .
4 (X^4)	my sources have suggested the company sees no reason to change its tax structure , which are perfectly legal .

Our objectives:

- An intuitive top-down progressive generation.
- Allows better long-term planning/control.
- Can leverage pretrained BERT.
- Logarithm inference speed.

POINTER (PrOgressive INsertionbased TransformER)

Stage	Generated text sequence
0 (X^0)	sources sees structure perfectly
1 (X^1)	sources company sees change structure perfectly legal
2 (X^2)	sources suggested company sees reason change tax structure which perfectly legal .
3 (X^3)	my sources have suggested the company sees no reason to change its tax structure , which are perfectly legal .
4 (X^4)	my sources have suggested the company sees no reason to change its tax structure , which are perfectly legal .

Principles:

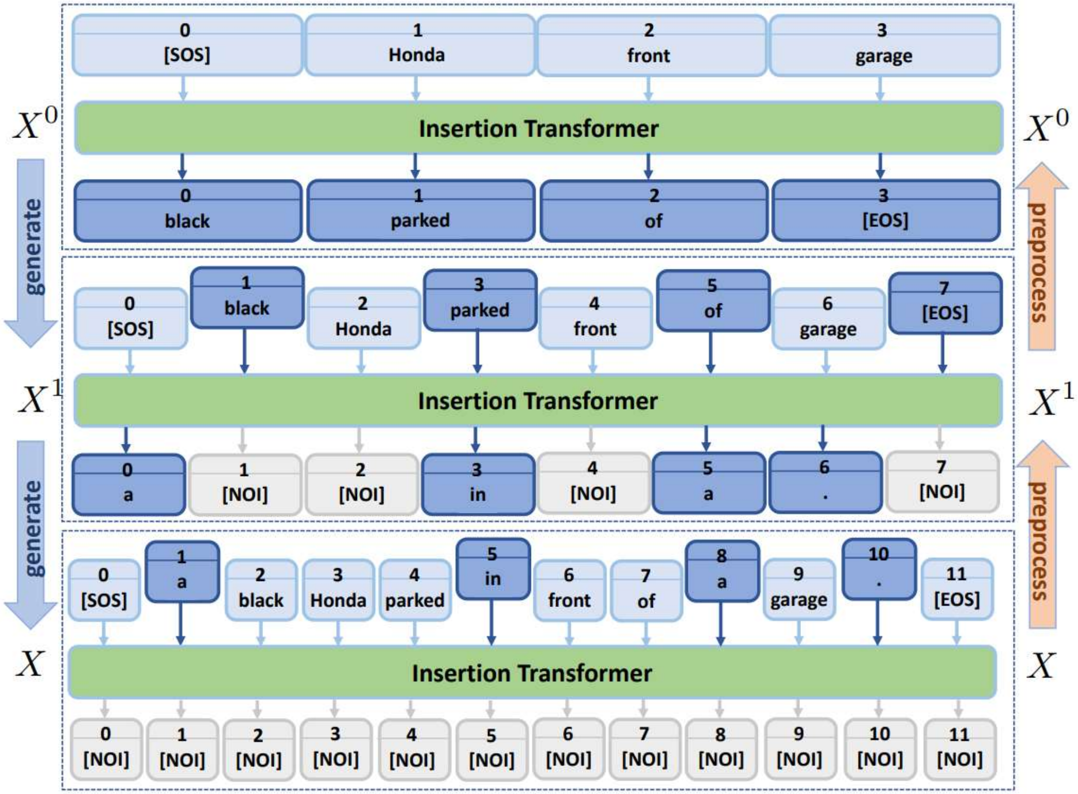
- *More important* tokens should be generated *earlier* => progressive
- Number of stage should be *small* => fast

Non-trivial to design a training objective like this!

Data preparation

Data preparation:

- Token Importance Scoring
- Data Instance Construction

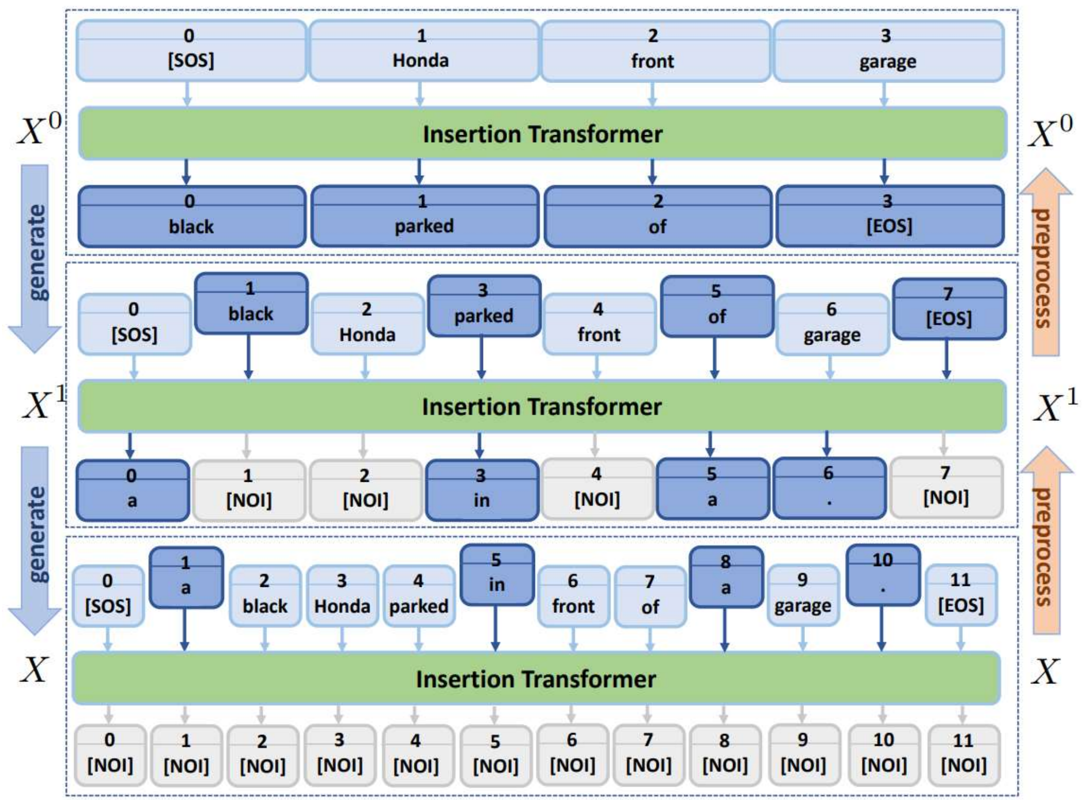


Data preparation

Data preparation:

- Token Importance Scoring
- Data Instance Construction

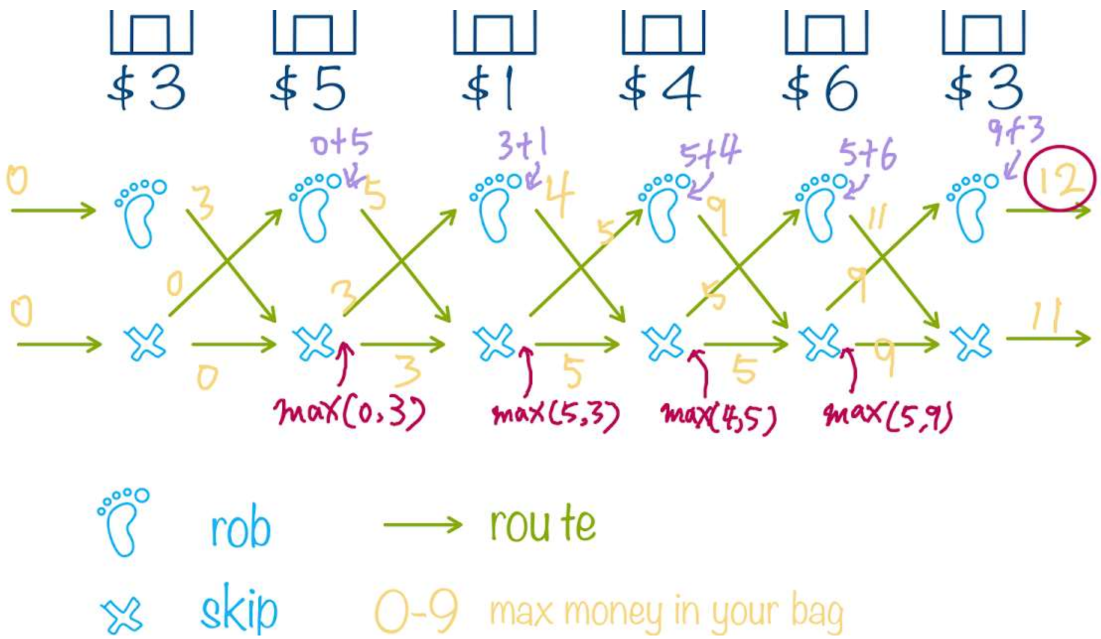
$$\alpha_t = \alpha_t^{\text{TF-IDF}} + \alpha_t^{\text{POS}} + \alpha_t^{\text{YAKE}}$$



Data preparation

Data preparation:

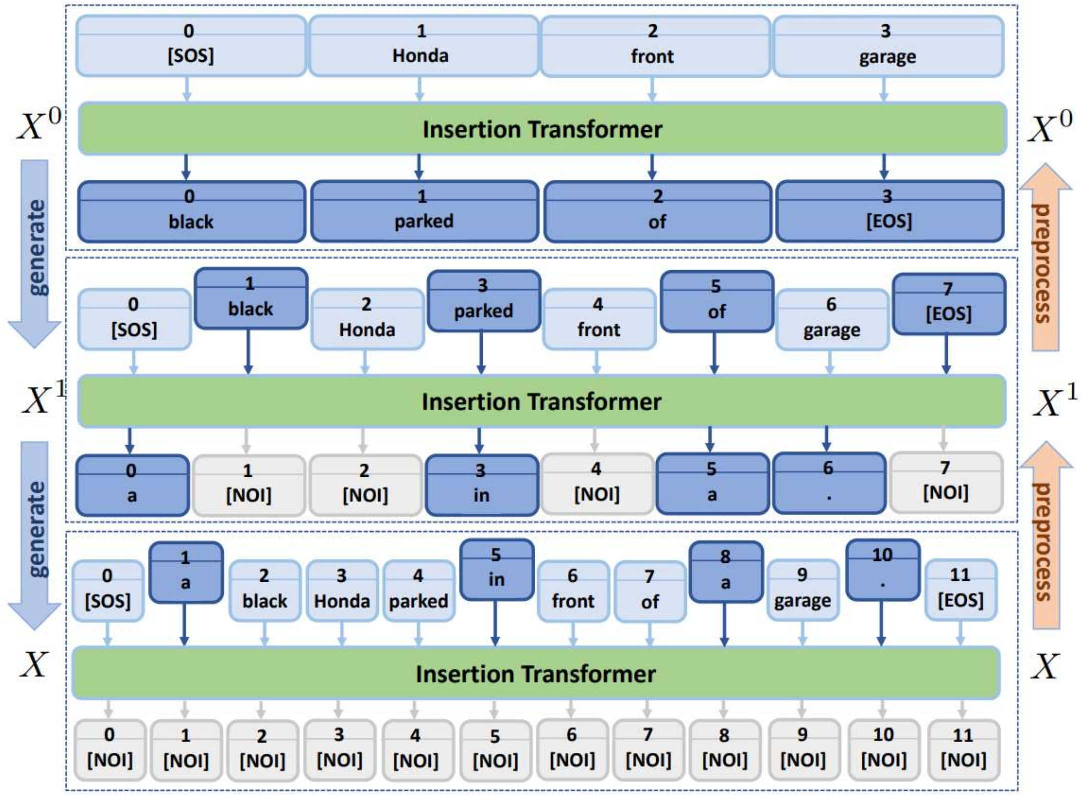
- Token Importance Scoring
- Data Instance Construction
    - Progressive => mask "important" words last
    - Fast => mask as many as possible
    - => House Robber Problem! (LEETCODE #198)



Model Training

Stage-wise Insertion Prediction:

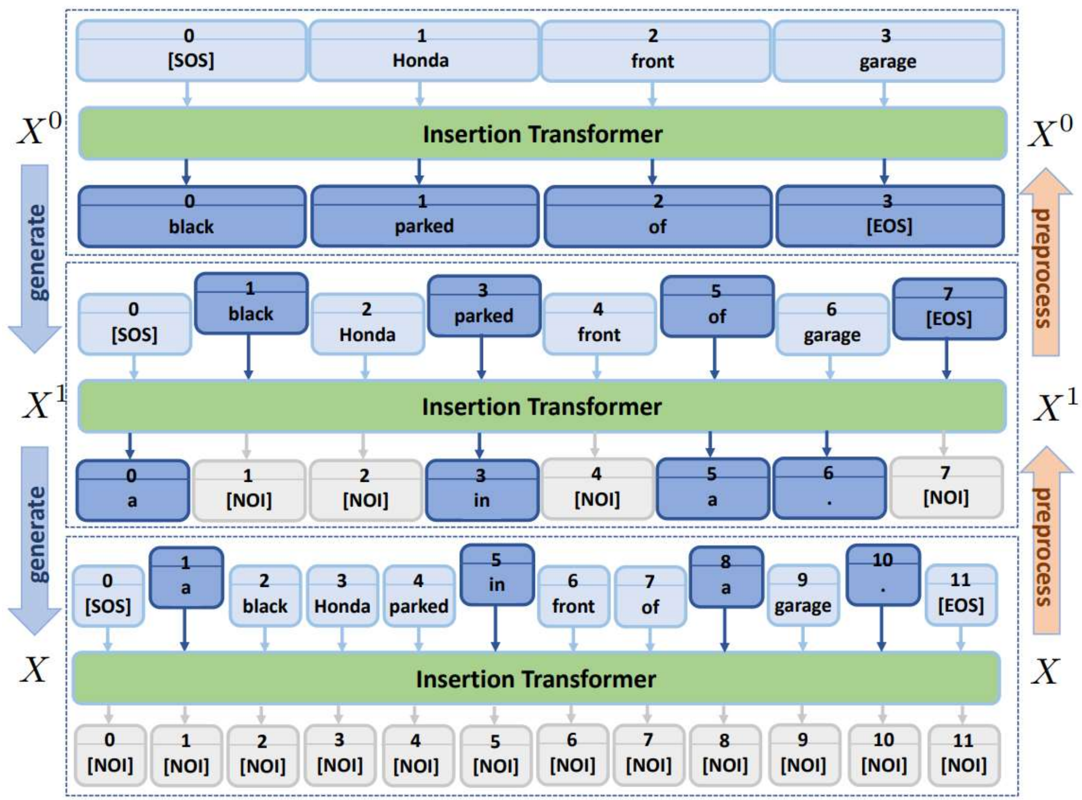
- BERT(MLM)-like objective
- Expanding the vocab with [NOI] for *non-insertion.*
- Large-scale Pre-training on Wiki.



Generation

- **Naïve Greedy Decoding:** conditional-independence at each stage

- **Inner-layer Beam Search (ILBS)**
    1) Generates top B token candidates by applying one evaluation step.
    2) Sweeps the generations to find the approximately optimal stage-wise decoding.



Evaluation

News dataset Method	NIST N-2	NIST N-4	BLEU B-2	BLEU B-4	METEOR	Entropy E-4	Dist D-1	Dist D-2	PPL.	Avg. Len.
CGMH	1.60	1.61	7.09%	1.61%	12.55%	9.32	**16.60%**	**70.55%**	189.1	14.29
NMSTG	2.70	2.70	10.67%	1.58%	13.56%	10.10	11.09%	65.96%	171.0	27.85
Greedy (base)	2.90	2.80	12.13%	1.63%	15.66%	**10.41**	5.89%	39.42%	97.1	47.40
Greedy (+Wiki,base)	3.04	3.06	13.01%	2.51%	**16.38%**	10.22	11.10%	57.78%	56.7	31.32
ILBS (+Wiki,base)	3.20	3.22	14.00%	2.99%	15.71%	9.86	13.17%	61.22%	66.4	22.59
Greedy (+Wiki, large)	**3.28**	**3.30**	**14.04%**	**3.04%**	15.90%	10.09	12.23%	60.86%	**54.7**	27.99
Human oracle	-	-	-	-	-	10.05	11.80%	62.44%	47.4	27.85

Yelp dataset Method	NIST N-2	NIST N-4	BLEU B-2	BLEU B-4	METEOR	Entropy E-4	Dist D-1	Dist D-2	PPL.	Avg. Len.
CGMH	0.50	0.51	4.53%	1.45%	11.87%	9.48	**12.18%**	**57.10%**	207.2	16.70
NMSTG	1.11	1.12	10.06%	1.92%	13.88%	10.09	8.39%	50.80%	326.4	27.92
Greedy (base)	2.15	2.15	11.48%	2.16%	**17.12%**	**11.00**	4.19%	31.42%	99.5	87.30
Greedy (+Wiki,base)	3.27	3.30	15.63%	3.32%	16.14%	10.64	7.51%	46.12%	71.9	48.22
ILBS (+Wiki,base)	3.34	3.38	16.68%	3.65%	15.57%	10.44	9.43%	50.66%	61.0	35.18
Greedy (+Wiki, large)	**3.49**	**3.53**	**16.78%**	**3.79%**	16.69%	10.56	6.94%	41.2%	**55.5**	48.05
Human oracle	-	-	-	-	-	10.70	10.67%	52.57%	55.4	50.36

Human evaluation

Semantics: *A and B, which is more semantically meaningful and consistent?*									
News dataset					Yelp dataset				
System A		Neutral	System B		System A		Neutral	System B	
POINTER(base)	**60.9%**	17.4%	21.8%	CGMH	POINTER(base)	**59.8%**	17.3%	23.0%	CGMH
POINTER(base)	**55.2%**	21.7%	23.1%	NMSTG	POINTER(base)	**57.5%**	23.0%	19.6%	NMSTG
POINTER(base)	21.7%	21.4%	**56.9%**	Human	POINTER(base)	26.8%	25.9%	**47.3%**	Human
Fluency: *A and B, which is more grammatical and fluent?*									
News dataset					Yelp dataset				
System A		Neutral	System B		System A		Neutral	System B	
POINTER(base)	**57.7%**	19.9%	22.4%	CGMH	POINTER(base)	**54.2%**	20.0%	25.8%	CGMH
POINTER(base)	**52.7%**	24.1%	23.2%	NMSTG	POINTER(base)	**59.0%**	22.8%	18.2%	NMSTG
POINTER(base)	16.6%	20.0%	**63.4%**	Human	POINTER(base)	24.0%	26.1%	**49.9%**	Human
Informativeness: *A and B, which is more informative?*									
News dataset					Yelp dataset				
System A		Neutral	System B		System A		Neutral	System B	
POINTER(base)	**70.4%**	12.8%	16.8 %	CGMH	POINTER(base)	**69.9%**	10.9%	19.3 %	CGMH
POINTER(base)	**57.7%**	18.7%	23.6%	NMSTG	POINTER(base)	**65.2%**	18.1%	16.7%	NMSTG
POINTER(base)	31.7%	19.0%	**49.4%**	Human	POINTER(base)	32.8%	19.0%	**48.2%**	Human

Generated examples and speed comparison

Keywords	**estate pay stay policy**
CGMH	an economic **estate** developer that could **pay** for it is that a **stay policy** .
NMSTG	as **estate** owners , they cannot **pay** for households for hundreds of middle - income property , buyers **stay** in retail **policy** .
POINTER (Greedy, base)	if you buy new buildings from real **estate** company, you may have to **pay** down a mortgage and **stay** with the **policy** for financial reasons .
POINTER (ILBS, base)	but no matter what foreign buyers do , real **estate** agents will have to **pay** a small fee to **stay** consistent with the **policy** .
POINTER (Greedy, Large)	but it would also be required for **estate** agents , who must **pay** a larger amount of cash but **stay** with the same **policy** for all other assets .

Table 3: Generated examples from the News dataset.

Keywords	**joint great food great drinks greater staff**
CGMH	very cool **joint** with **great food** , **great drinks** and even **greater staff** . ! .
NMSTG	awesome **joint** . **great** service. **great food** great **drinks**. good to **greater** and great **staff**!
POINTER (Greedy, base)	my favorite local **joint** around old town. **great** atmosphere, amazing **food**, delicious and delicious coffee, **great** wine selection and delicious cold **drinks**, oh and maybe even a **greater** patio space and energetic front desk **staff**.
POINTER (ILBS, base)	the best breakfast **joint** in charlotte . **great** service and amazing **food** . they have **great** selection of **drinks** that suits the **greater** aesthetic of the **staff** .
POINTER (Greedy, Large)	this is the new modern breakfast **joint** to be found around the area . **great** atmosphere , central location and excellent **food** . nice variety of selections . **great** selection of local craft beers , good **drinks** . quite cheap unless you ask for **greater** price . very friendly patio and fun **staff** . love it !

Table 4: Generated examples from the Yelp dataset.

Model	Training	Inference
CGMH	4382 toks/s	33h
NMSTG	357 toks/s	487s
POINTER	5096 toks/s	94s

Table 6: Speed comparison. “toks/s” represents tokens per second. Inference time is computed on 1000 test examples. POINTER uses (greedy, base)

Live demo

POINTER: Constrained Text Generation via Insertion Transformer

Paper Demo Github

Generating the sentence from the lexical constraints

Please input around 4-9 keyword constraints that are related with Yelp review.

Constraints

service perfect delicious chicken awesome good place

Generated sentence 1 :
amazing food and great service . my caesar salad was perfect and my steak salad was so delicious . i also recommend the chicken and the pepper steak , it is awesome . i highly recommend both . so good that i would recommend this place to everyone !

Generated sentence 2 :
amazing food and great service . the tomato soup was perfect and the greek salad was delicious . the greek salad and chicken salad are awesome . i highly recommend this for a good lunch . highly recommend this place !

Generated sentence 3 :
food and great customer service . i had short ribs cooked perfect and the greek salad and it was delicious , my bf also ordered chicken salad and it was awesome . i highly recommend anyone looking for good go this place !

Generated sentence 4 :
amazing food and great service . caesar salad was fresh and perfect . me and my husband had something delicious . the greek salad and chicken salad were awesome . i would recommend it for really good dinner i would recommend this place to everyone !

Generated sentence 5 :
amazing food and great service , the steak salad was perfect for me and my boyfriend and such delicious . we also ordered chicken thai salad , both were awesome and i highly recommend anyone looking for good pizza i would recommend this place to everyone !

Generate

Summary

- A simple yet powerful approach to **progressively** generating text.
- A pre-trained non-autoregressive model on wiki.
- Both automatic and human evaluation demonstrate the effectiveness of POINTER.