INSET: Sentence Infilling with INter-SEntential Transformer

Yichen Huang*, Yizhe Zhang*, Oussama Elachaqar, Yu Chen

ACL 2020

Microsoft Research AI

Availability

- INSET: Sentence Infilling with INter-SEntential Transformer[ACL 2020]
  - https://arxiv.org/abs/1911.03892
- GitHub repository:
  - https://github.com/dreasysnail/INSET
  - Stay tuned for our updates!
- Public demo
  - Under development, available soon.
- Contact us @
  - yichuang@mit.edu, yizzhang@microsoft.com

Sentence Infilling (w/ and w/o hints)

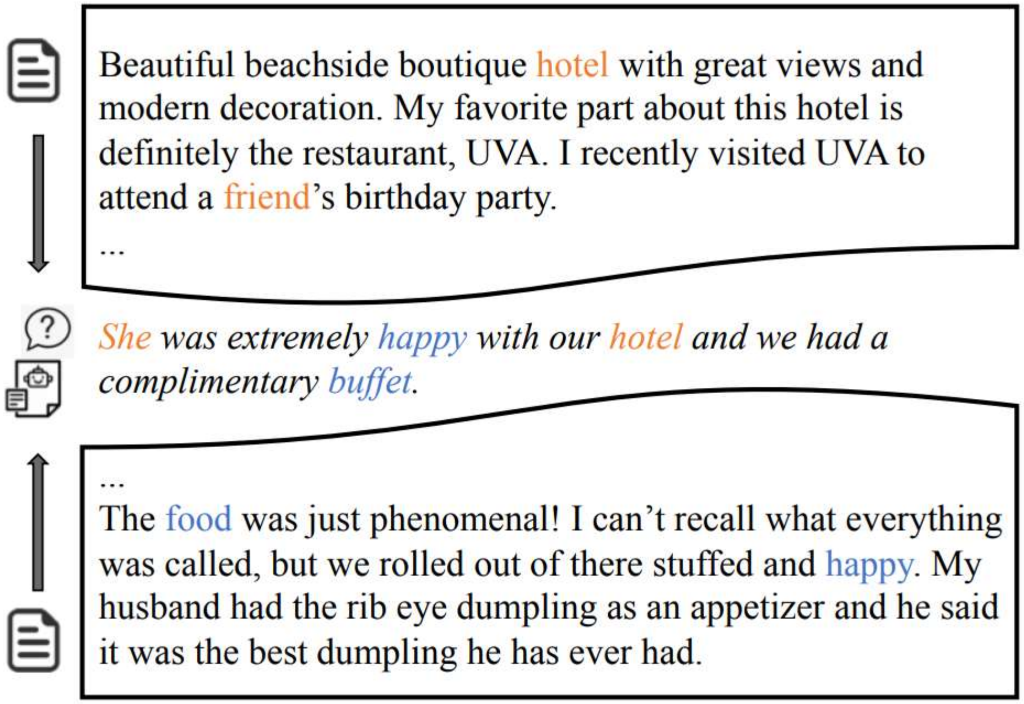


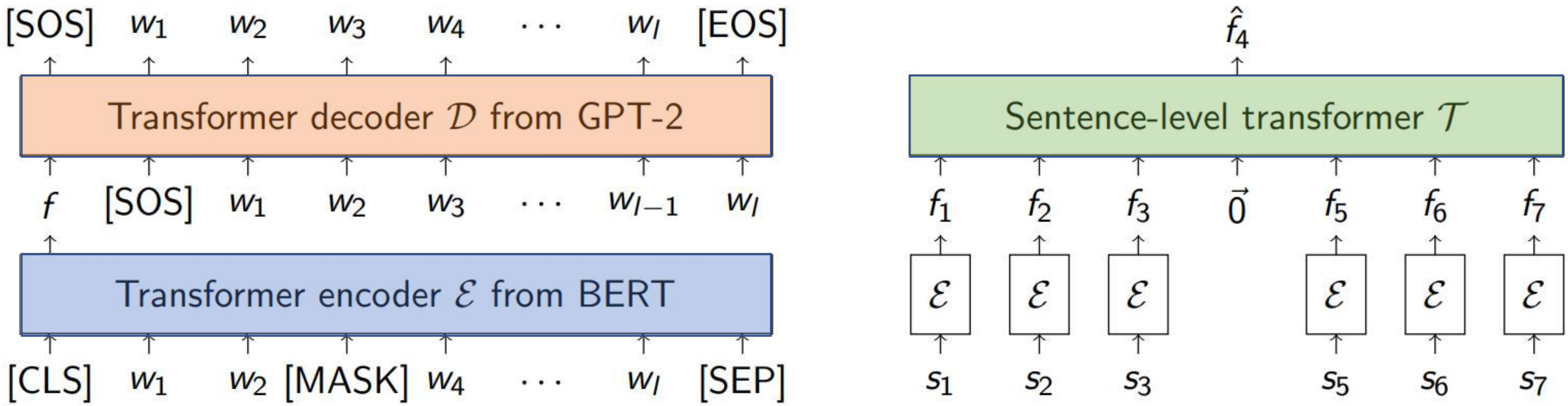
Figure: Sentence infilling: generating an intermediate sentence that provides a smooth semantic transition from the preceding to the following context. This example is generated by our model on the TripAdvisor dataset.

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

Possible scenarios

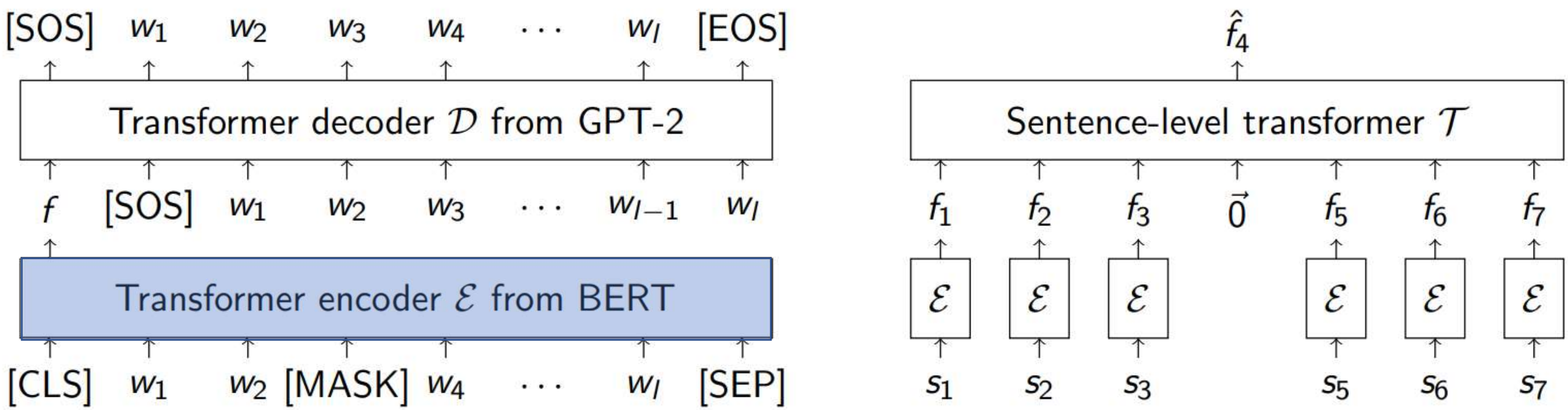
- **Document auto-completion:** suggesting missing bridging sentences in the surrounding context
- **Collaborative document writing:** unifying different writing styles from multiple authors
- **Note expansion**: extending a set of keywords to a full sentence, leveraging the surrounding context

INSET: INter-SEntential Transformer



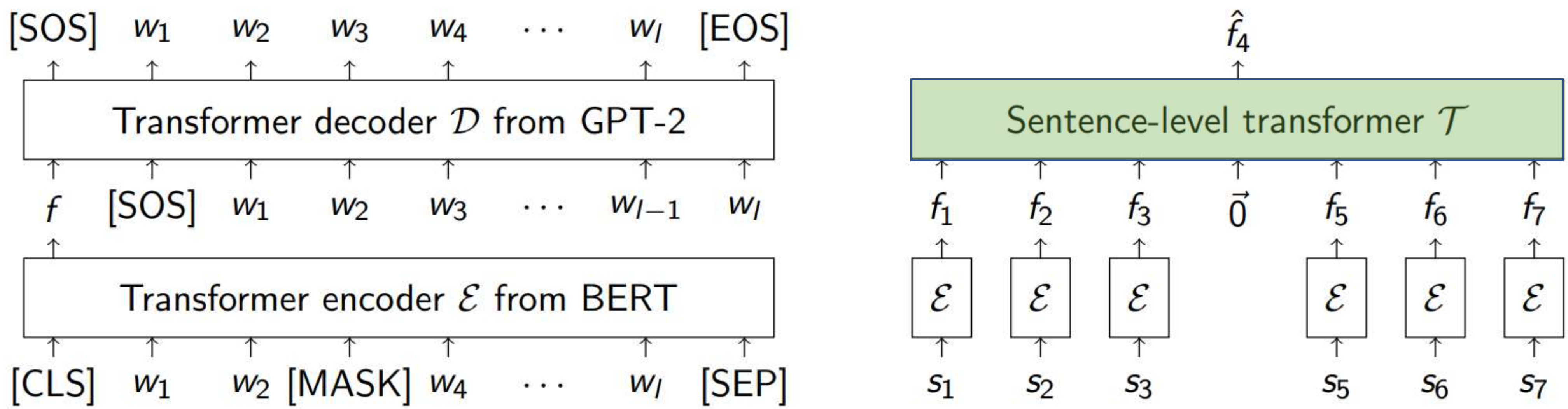
- ***Understanding*** (BERT-like encoder)
- ***planning*** (sentence-level Transformer)
- ***generation*** (GPT-like decoder)

INSET: INter-SEntential Transformer



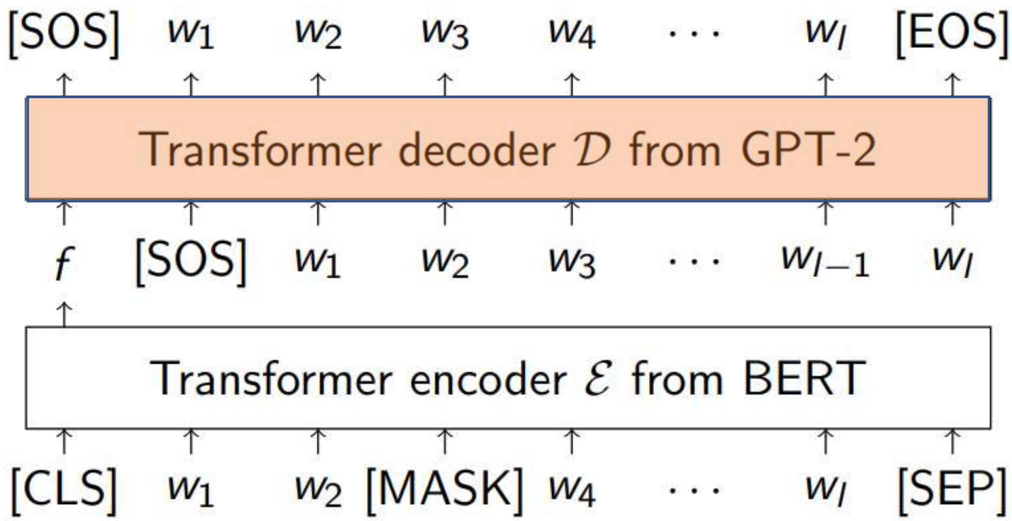
- ***Understanding*** (BERT-like encoder) : BERT-base size 110M
- a BERT-based encoder to map each sentence to the latent semantic space (768 dimension vector)

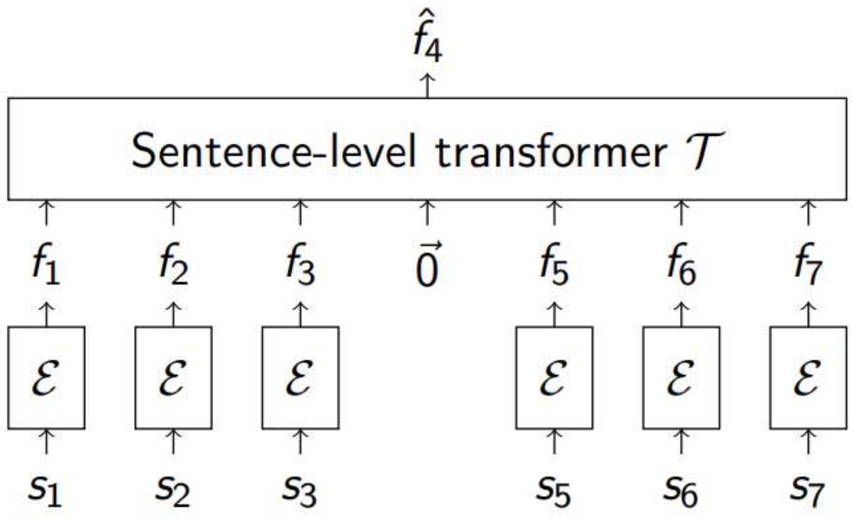
INSET: INter-SEntential Transformer



- ***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.

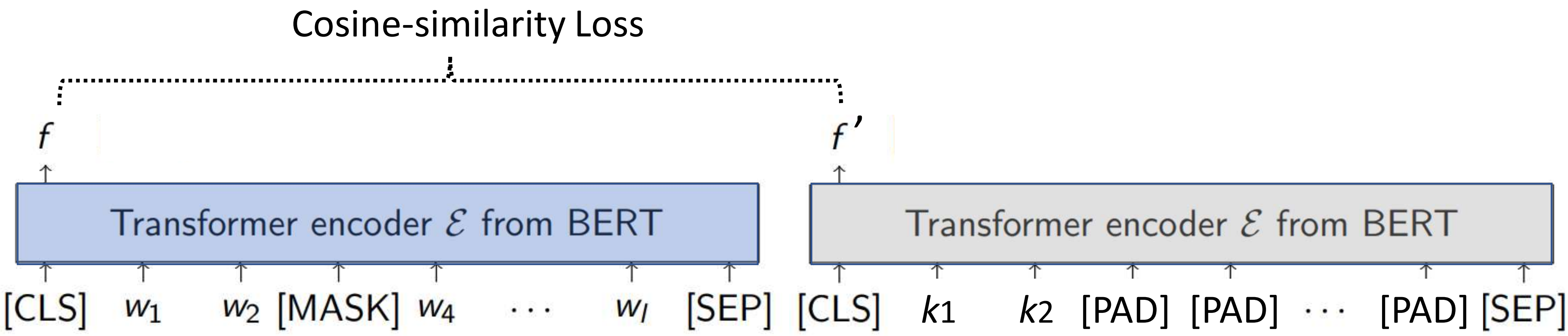
INSET: INter-SEntential Transformer





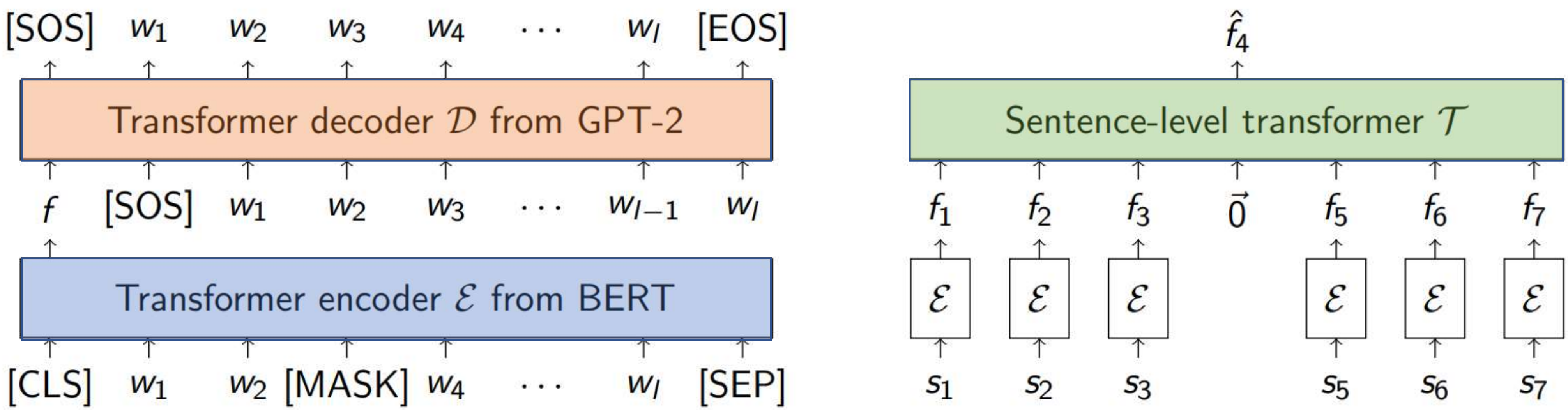
- ***generation*** (GPT-like decoder) : GPT-small size 117M
- a GPT-based generator (decoder) to map semantic features back to the text domain.

INSET: INter-SEntential Transformer



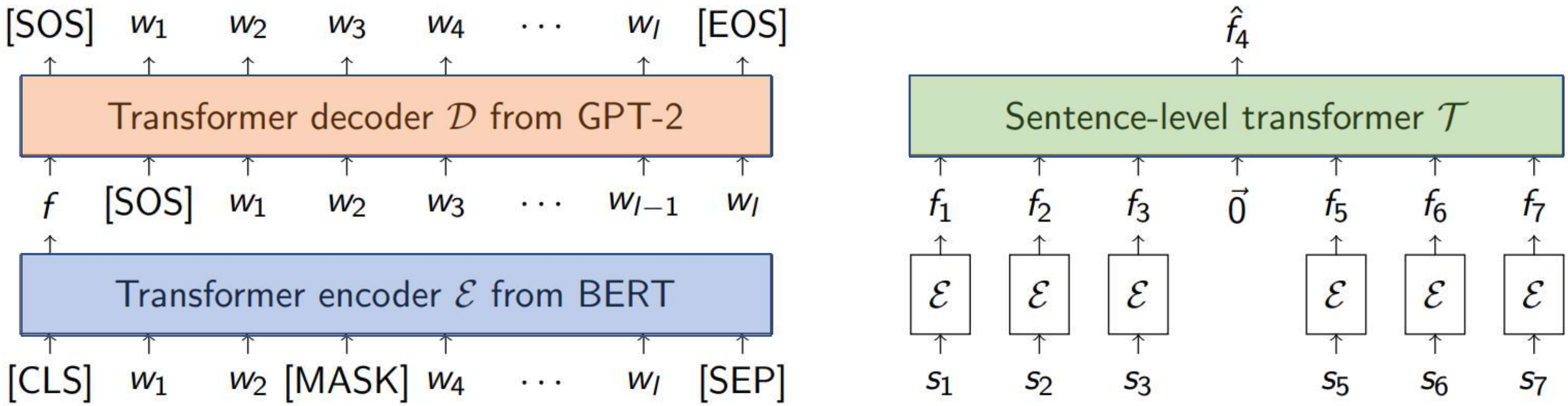
- ***Constraint feature encoder*** (BERT-like encoder) : BERT-base size 110M
- Distillation-like procedure
- **Teacher:** fixed sentence encoder
- **Student:** constraint feature encoder with *no position embedding*.

INSET: INter-SEntential Transformer



- 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)
    - (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)

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

- Sentence infilling requires the model to handle ***long-range inter-sentential*** correlation and to process high-level semantic information.

- We propose a framework called **INSET** to ***decouple*** three aspects of the task (understanding, planning, and generation).

- We demonstrate the effectiveness of our approach using automatic and human evaluation.