Personalized Transformer for Explainable Recommendation

Lei Li[1], Yongfeng Zhang[2], Li Chen[1]

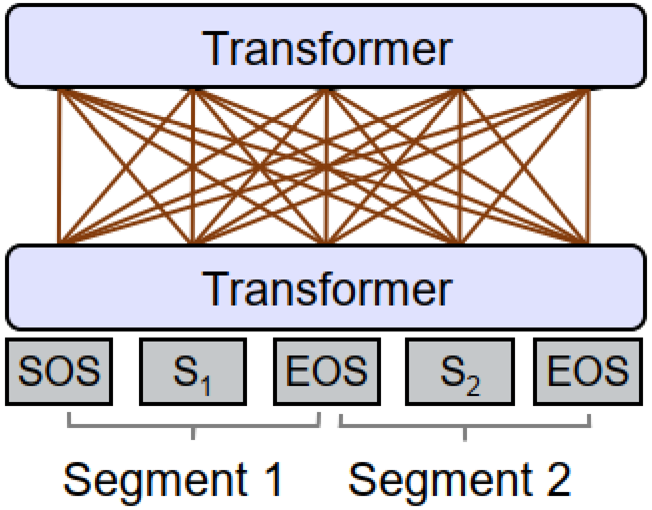
[1] Hong Kong Baptist University, [2] Rutgers University

csleili@comp.hkbu.edu.hk

August 4, 2021

Transformer (Vaswani et al., NIPS’17)

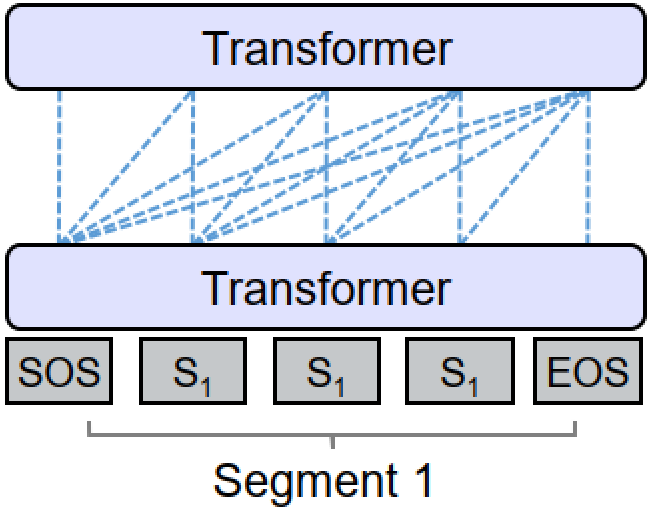
BERT (Devlin et al., NAACL’19)



Bidirectional model for classification

VS.

OpenAI GPT (Radford et al., 18)

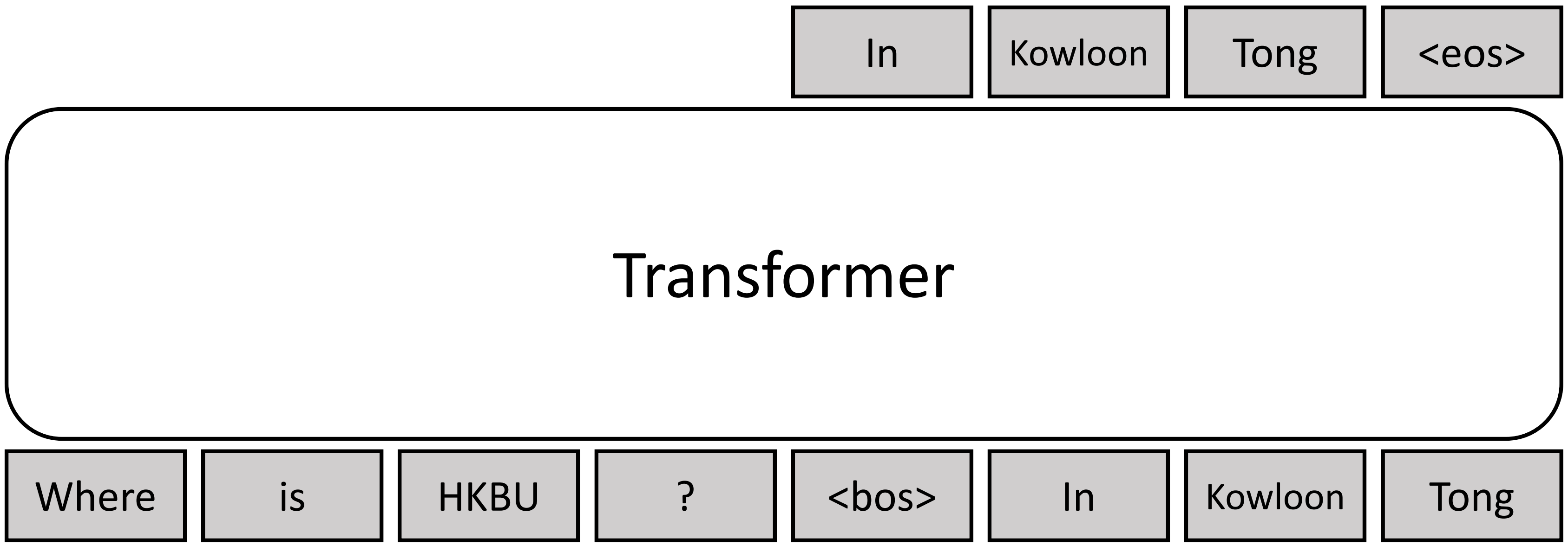


Unidirectional model for generation

Courtesy images from UNILM (Dong et al., NeurIPS’19)

Autoregressive Natural Language Generation

- Predict future tokens based on past tokens
  - Generate an output sequence, based on the given input sequence

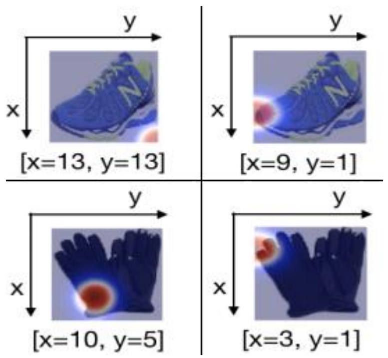


Explainable Recommendation

- Given a user-item pair, provide an **explanation** to justify why the item is recommended to the user
  - Pre-defined template (Zhang et al., SIGIR'14)
  - Image visualization (Chen et al., SIGIR'19)
  - Natural language sentence in this work
    - E.g., "*the style of the jacket is fashionable*"
  - ……

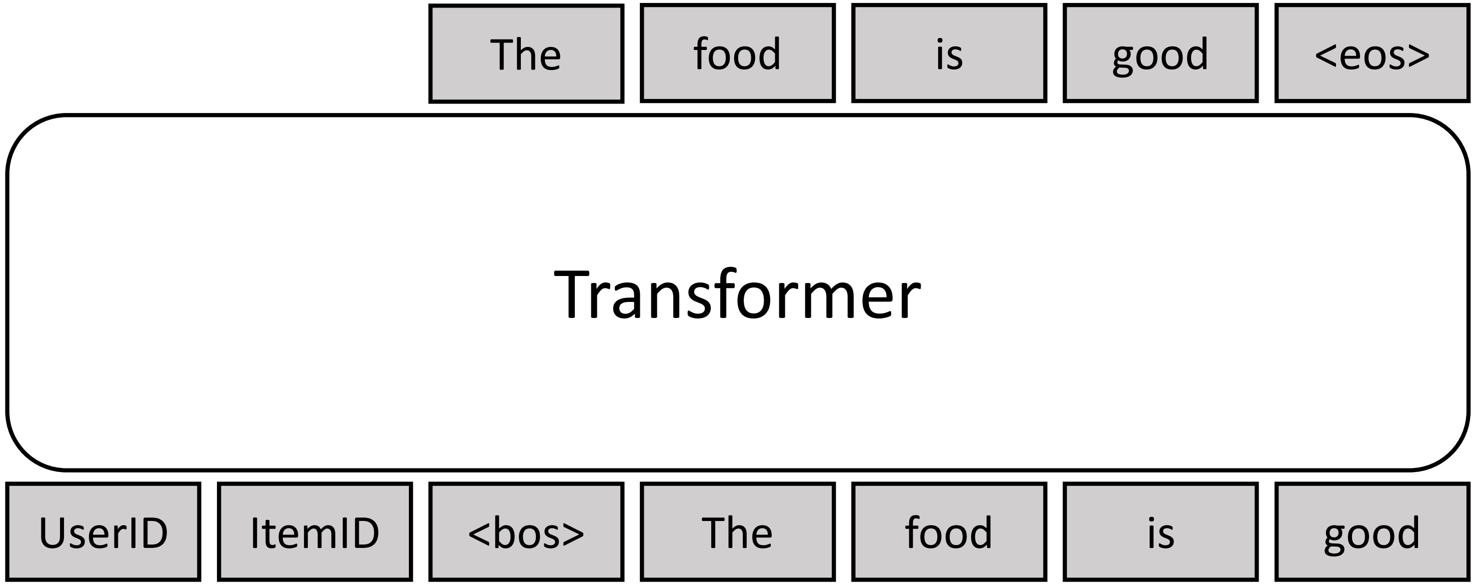
```
You might be interested in [feature],
on which this product performs well.
```

```
You might be interested in [feature],
on which this product performs poorly.
```



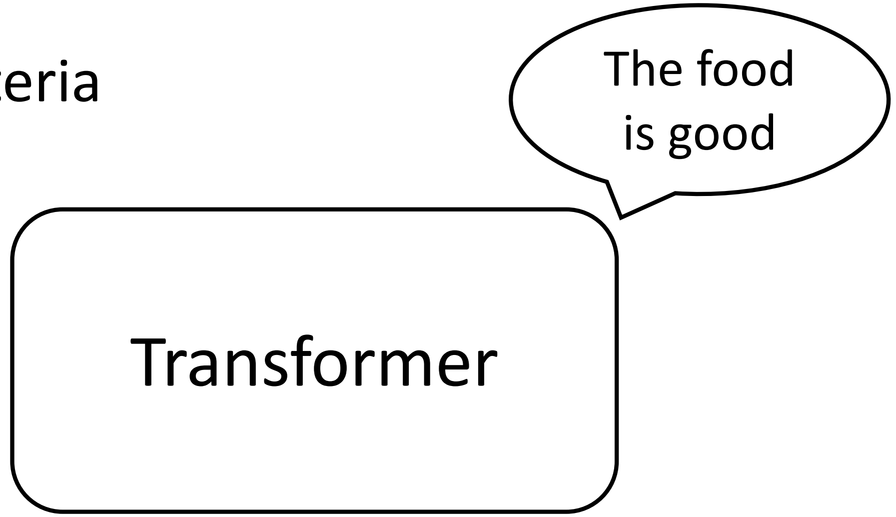
Transformer for Explanation Generation

- Consider the user-item pair as an input sequence
  - Regard IDs as tokens, similar to words



Problem Identification

- Identical generated explanations for almost every user-item pair
  - Adam   Main Canteen
  - Beth   Renfrew Cafeteria
  - Carol   Bistro Bon
  - David   Harmony Cafeteria
  - ……

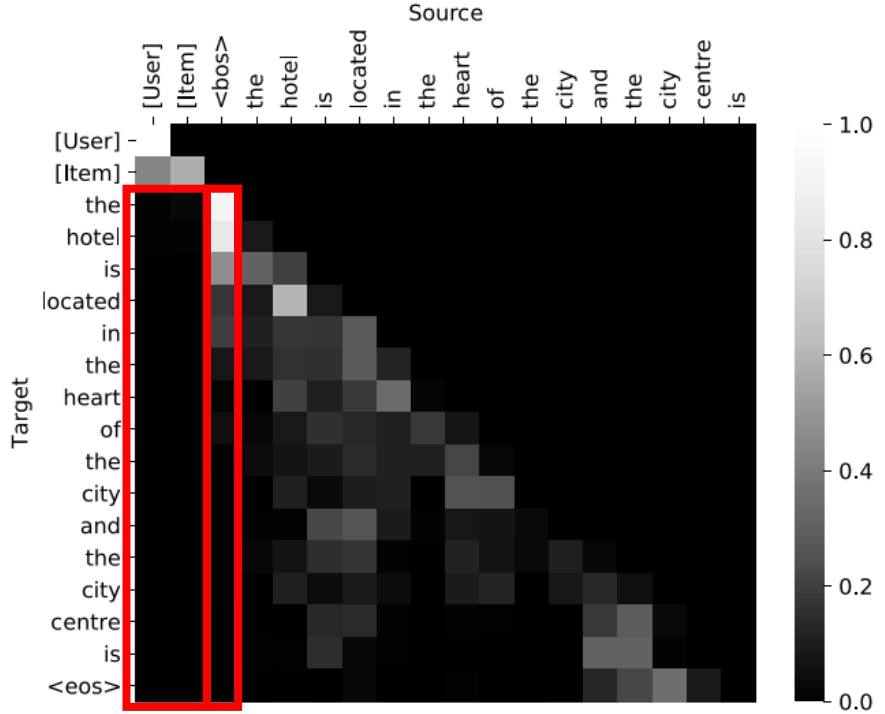


- **Less useful, if unable to explain the key specialty of each recommendation**
- **Could cause negative effects on users (**Tintarev and Mashoff, 15**)**

Attention Visualization

- The generation relies heavily on <bos>
  - The reason why all explanations are identical
- Attention weights on userID and itemID approach 0
  - Model insensitive to IDs

Why insensitive?



Problem Analysis

- Frequency mismatch between IDs and words
  - One user/item ID vs. hundreds of words in a review
  - An ID appears in only a few reviews
- IDs being regarded as uncommon words (OOV tokens)

Shelby S. Elite '2021
Saint Paul, MN
166 354 771

12/4/2019

6 photos

Ho Lee Fook was one of the best food spots I went to in HK. At first I was skeptical because sometimes the fusion or westernized type Asian restaurants are all for the look but don't taste great. But, Ho Lee Fook was beautiful inside and the food was amazing. We ordered the pan fried thick rolled noodles and the massive bone steak (forgot the actually name) but you won't miss it on the menu. The noodles were crispy and seasoned just right. The steak was so tender and delicious. It came with a jalapeño sauce on the plate which complimented it so well.

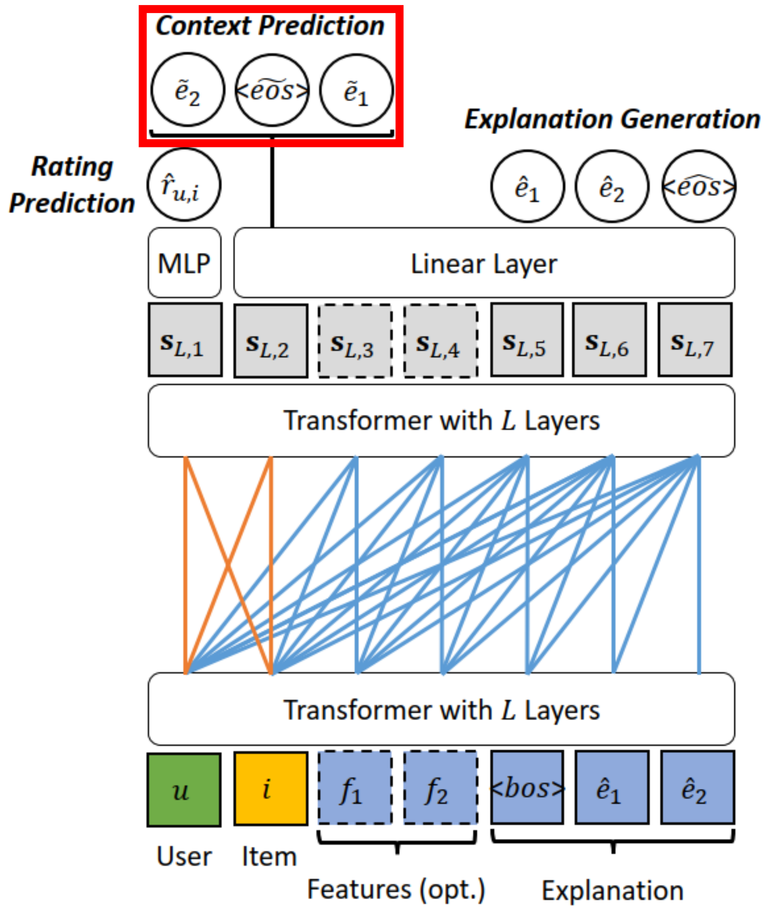
While being here I forgot I was in HK because everyone spoke English and the menu was also in English! The entrance is so cute with the lucky cats all on the walls.

If you are visiting HK or live there I definitely recommend giving this place a try! It is a little on the pricey side but for the atmosphere it is expected.

Restaurant review (yelp.com)

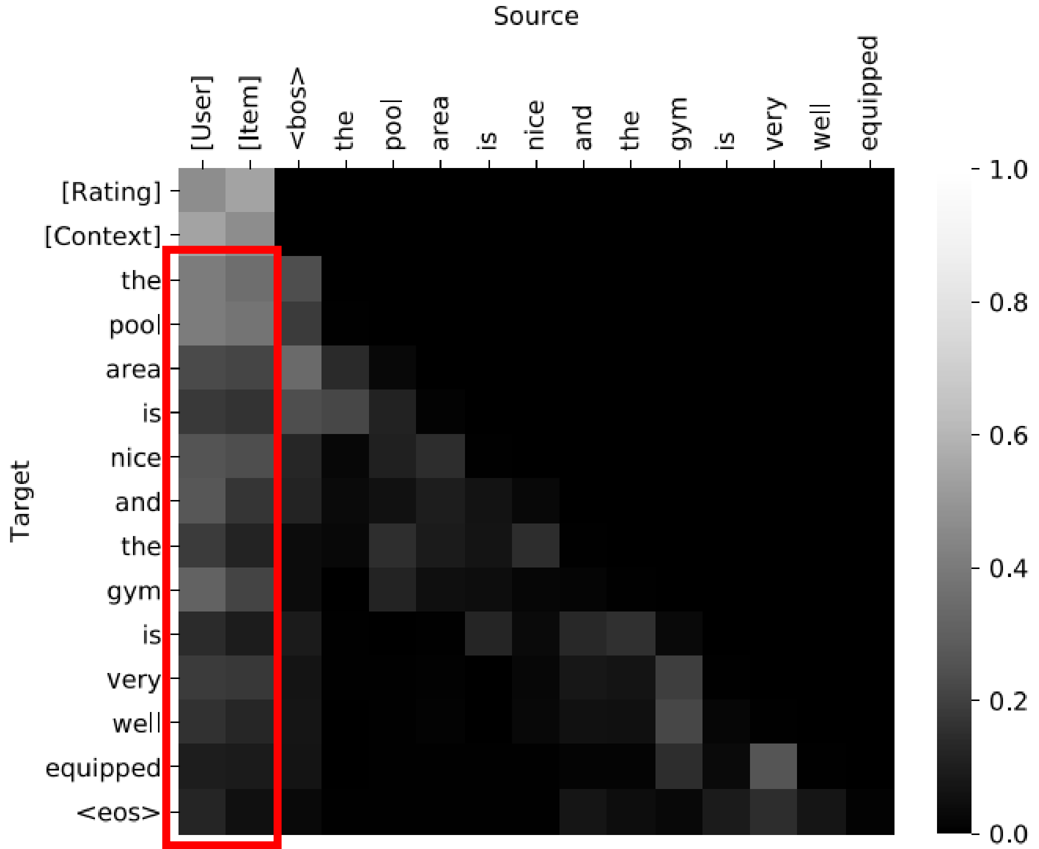
Solution: Context Prediction

- Bridge IDs and words, and give the former linguistic meanings



Attention Visualization Again

- Our model PETER can utilize IDs for generation
  - PETER: PErsonalized Transformer for Explainable Recommendation



Attention Masking

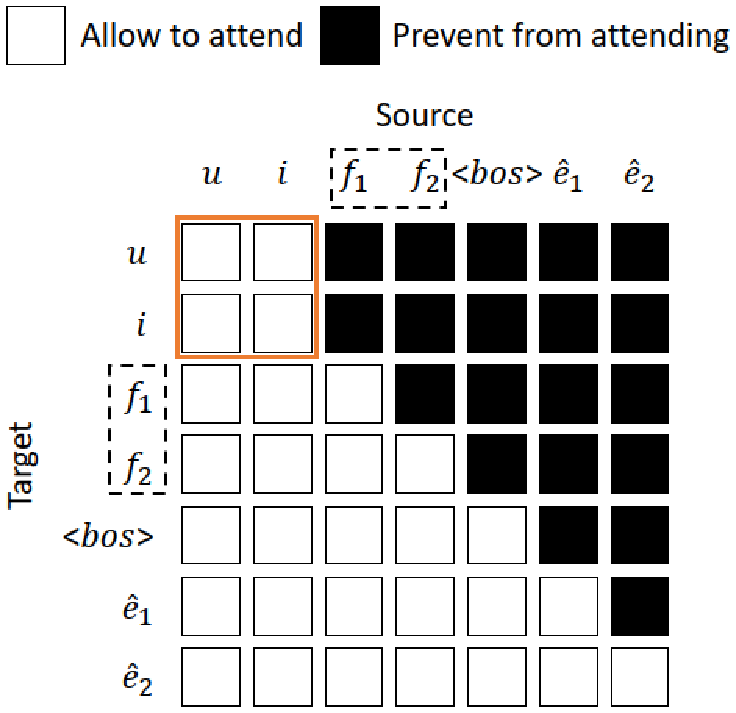
- Revise Left-to-Right attention masking matrix (call it PETER masking)
  - Allow the interaction between user and item for context prediction and recommendation

$$\mathbf{A}_{l,h} = \text{softmax}(\frac{\mathbf{Q}_{l,h}\mathbf{K}_{l,h}^{\top}}{\sqrt{d}} + \mathbf{M})\mathbf{V}_{l,h}$$

$$\mathbf{Q}_{l,h} = \mathbf{S}_{l-1}\mathbf{W}_{l,h}^{Q}, \mathbf{K}_{l,h} = \mathbf{S}_{l-1}\mathbf{W}_{l,h}^{K},$$

$$\mathbf{V}_{l,h} = \mathbf{S}_{l-1}\mathbf{W}_{l,h}^{V}$$

$$\mathbf{M} = \begin{cases} 0, & \text{Allow to attend} \\ -\infty, & \text{Prevent from attending} \end{cases}$$



Context Prediction & Explanation Generation

- Context prediction: predict explanation words in one step
  - With the item representation
- Explanation generation: generate them one by one
  - Linear layer

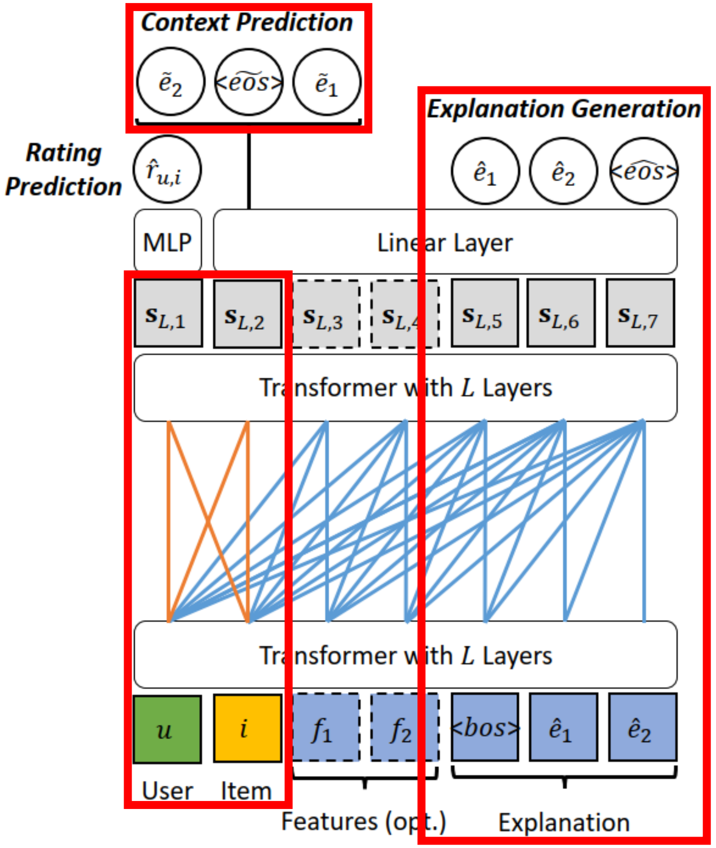
$$\mathbf{c}_t = \mathrm{softmax}(\mathbf{W}^v \mathbf{s}_{L,t} + \mathbf{b}^v)$$

  - NLL loss for context prediction

$$\mathcal{L}_c = \frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} \frac{1}{|E_{u,i}|} \sum_{t=1}^{|E_{u,i}|} -\log c_2^{e_t}$$

  - NLL loss for explanation generation

$$\mathcal{L}_e = \frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} \frac{1}{|E_{u,i}|} \sum_{t=1}^{|E_{u,i}|} -\log c_{2+|F_{u,i}|+t}^{e_t}$$



Recommendation & Targeted Explanation

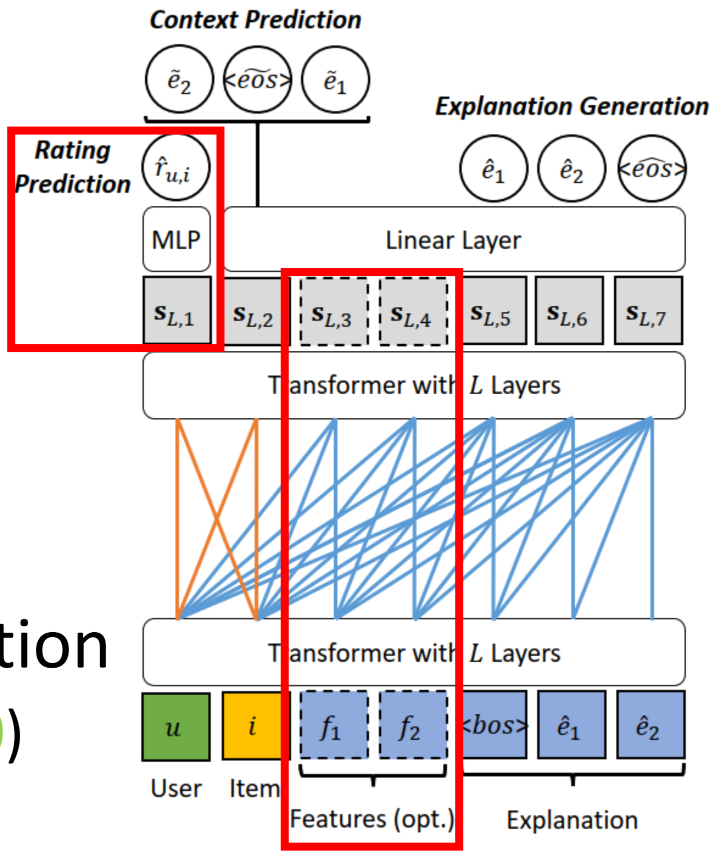
- Predict a rating score for the user-item pair
  - MLP with one hidden layer

$$\hat{r}_{u,i} = \mathbf{w}^r \sigma(\mathbf{W}^r \mathbf{s}_{L,1} + \mathbf{b}^r) + b^r$$

  - MSE loss for rating prediction

$$\mathcal{L}_r = \frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} (r_{u,i} - \hat{r}_{u,i})^2$$

- Incorporate features for targeted explanation generation
  - E.g., conversational recommendation (Chen et al., IJCAI'20)
  - Denoted as PETER+



Multi-task Learning

- Three tasks trained in an end-to-end manner
  - Explanation generation
  - Context prediction
  - Rating prediction

$$\mathcal{J} = \min_{\Theta}(\lambda_e \mathcal{L}_e + \lambda_c \mathcal{L}_c + \lambda_r \mathcal{L}_r)$$

Datasets (Li et al., CIKM'20)

- Yelp
  - Restaurant
- Amazon
  - Movies & TV
- TripAdvisor
  - Hotel
- The explanation is a review sentence containing features

	Yelp	Amazon	TripAdvisor
#users	27,147	7,506	9,765
#items	20,266	7,360	6,280
#records	1,293,247	441,783	320,023
#features	7,340	5,399	5,069
#records / user	47.64	58.86	32.77
#records / item	63.81	60.02	50.96
#words / exp	12.32	14.14	13.01

* **exp** denotes **explanation**.

Evaluation Metrics

- Recommendation
  - RMSE & MAE
- Explanation
  - Text quality: BLEU (Papineni et al., ACL’02) & ROUGE (Lin, ACL’04 Workshop)
    - Not equal to explainability (Chen et al., SIGIR’19 Workshop; Li et al., CIKM’20)
  - Explainability from the angle of item features (Li et al., CIKM’20)
    - Unique Sentence Ratio (USR)
    - Feature Matching Ratio (FMR)
    - Feature Coverage Ratio (FCR)
    - Feature Diversity (DIV)

Quantitative Analysis on Explanations

- Ours the best or comparable

	Explainability			Text Quality								
	FMR↑	FCR↑	DIV↓	USR↑	B1↑	B4↑	R1-P↑	R1-R↑	R1-F↑	R2-P↑	R2-R↑	R2-F↑
	TripAdvisor											
	Amazon											
	Yelp											
Transformer	0.06	0.06	2.46	0.01	7.39	0.42	**19.18**	10.29	12.56	1.71	0.92	1.09
NRT	0.07	0.11	2.37	0.12	**11.66**	0.65	17.69	12.11	13.55	1.76	1.22	1.33
Att2Seq	0.07	0.12	2.41	**0.13**	10.29	0.58	18.73	11.28	13.29	1.85	1.14	1.31
PETER	**0.08****	**0.19****	**1.54****	**0.13**	10.77	**0.73****	18.54	**12.20**	**13.77****	**2.02****	**1.38****	**1.49****
ACMLM	0.05	0.31	**0.95**	**0.95**	7.01	0.24	7.89	7.54	6.82	0.44	0.48	0.39
NETE	0.80	0.27	1.48	0.52	19.31	2.69	33.98	22.51	25.56	8.93	5.54	6.33
PETER+	**0.86****	**0.38****	1.08	0.34	**20.80****	**3.43****	**35.44****	**26.12****	**27.95****	**10.65****	**7.44****	**7.94****

IDs only

Metric problem

With features

Less useful, if unable to guarantee text quality

Qualitative Case Study on Explanations

- Context prediction task can indeed give IDs linguistic meanings
- Two tasks resemble one's drafting-polishing process
- Incorporated features further improve text quality

	Top-15 Context Words	Explanation
Ground-truth		the **rooms** are spacious and the bathroom has a large tub
PETER	<eos> the and a pool was with nice is very were to good in of	the pool area is nice and the gym is very well equipped <eos>
PETER+	<eos> the and a was pool with to nice good very were is of in	the rooms were clean and comfortable <eos>
Ground-truth		beautiful **lobby** and nice bar
PETER	<eos> the and a was were separate bathroom with shower large very had in is	the bathroom was large and the shower was great <eos>
PETER+	<eos> the and a was bathroom shower with large in separate were room very is	the lobby was very nice and the rooms were very comfortable <eos>

Efficiency Comparison

- Training minutes comparison with BERT-based model under the same settings
  - PETER+ is small (only 2 attention layers), so it takes much less training time
  - PETER+ is unpretrained, and thus requires more training epochs

	Time	Epochs	Time/Epoch
ACMLM	97.0	**3**	32.3
PETER+	**57.7**	25	**2.3**

Recommendation Performance

- Ours the best on the largest dataset with over 1 million records
- Ours comparable on small datasets
  - Not a problem to real applications, e.g., billion-scale users in Amazon

	Yelp		Amazon		TripAdvisor	
	R↓	M↓	R↓	M↓	R↓	M↓
PMF	1.09	0.88	1.03	0.81	0.87	0.70
SVD++	**1.01**	**0.78**	0.96	0.72	0.80	0.61
NRT	**1.01**	**0.78**	**0.95**	**0.70**	**0.79**	0.61
NETE	**1.01**	0.79	0.96	0.73	**0.79**	**0.60**
PETER	**1.01**	**0.78**	**0.95**	0.71	0.81	0.63

Ablation Study

- Prove the rationale of each component

Reduce to standard Transformer

	Explainability			Text Quality			Recommendation	
	FMR	FCR	DIV	USR	BLEU-1	BLEU-4	RMSE	MAE
Disable $\mathcal{L}_c$	0.06 ↓	0.03 ↓	5.75 ↓	$\boxed{0.01 ↓}$	15.37 ↓	0.86 ↓	0.80 ↑	0.61 ↑
Disable $\mathcal{L}_r$	0.07	0.14 ↑	2.90 ↑	0.10 ↑	16.16 ↑	1.15 ↑	3.23 ↓	3.10 ↓
Left-to-Right Masking	0.07	0.15 ↑	2.68 ↑	0.12 ↑	15.73 ↓	1.11	0.87 ↓	0.68 ↓
PETER	0.07	0.13	2.95	0.08	15.96	1.11	0.81	0.63

Recommendation and context prediction highly correlated

Block item information

Conclusion

- The 1st to enable Transformer with personalized natural language generation
  - Shed light on a broader scope of fields that also need personalization
    - E.g., personalized conversational systems
- Model small and unpretrained, but effective and efficient
  - Open up a new way of exploiting Transformer by designing good tasks instead of scaling up model size
- Design a task to connect IDs and words
  - Point out a way for Transformer to deal with heterogeneous inputs
    - E.g., **image generation** based on text in multimodal AI

References (1)

- Vaswani, Ashish, et al. "Attention is all you need." NIPS’17.
- Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." NAACL’19.
- Radford, Alec, et al. "Improving language understanding by generative pre-training." 2018.
- Dong, Li, et al. "Unified language model pre-training for natural language understanding and generation." NeurIPS’19.
- Chen, Xu, et al. "Personalized Fashion Recommendation with Visual Explanations based on Multimodal Attention Network: Towards Visually Explainable Recommendation." SIGIR’19.
- Zhang, Yongfeng, et al. "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis." SIGIR’14.
- Chen, Zhongxia, et al. "Towards Explainable Conversational Recommendation." IJCAI’20.
- Tintarev, Nava, and Judith Masthoff. "Explaining recommendations: Design and evaluation." Recommender systems handbook. 2015.

References (2)

- Li, Lei, et al. "Generate neural template explanations for recommendation." CIKM’20.
- Papineni, Kishore, et al. "BLEU: a method for automatic evaluation of machine translation." ACL’02.
- Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." ACL’04 Workshop.
- Chen, Hanxiong, et al. "Generate natural language explanations for recommendation." SIGIR’19 Workshop.

Q&A

Thank you!

csleili@comp.hkbu.edu.hk

lileipisces.github.io