Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ciencia de la Computación
IIC3633 - Sistemas Recomendadores

Personalized Transformer for Explainable Recommendation

Grupo 12: Daniel Alegría, Gabriel Catalán y Benjamín Faúndez

19/06/2025

Autores

Lei Li[1] **Yongfeng Zhang**[2] **Li Chen**[1]
[1]Hong Kong Baptist University, Hong Kong, China
[2]Rutgers University, New Brunswick, USA
[1]{csleili,lichen}@comp.hkbu.edu.hk
[2]yongfeng.zhang@rutgers.edu

Resumen

- Contexto
- Problema de recomendación
- Contribución
- Estado del arte y marco teórico
- Detalle solución
- Evaluación
- Referencias
- Preguntas

Contexto

- v1 Martes, 25 de Mayo 2021
- v2 Sábado, 5 de Junio 2021
- ¿Qué estaba pasando en la escena de la IA?

Contexto: NLG

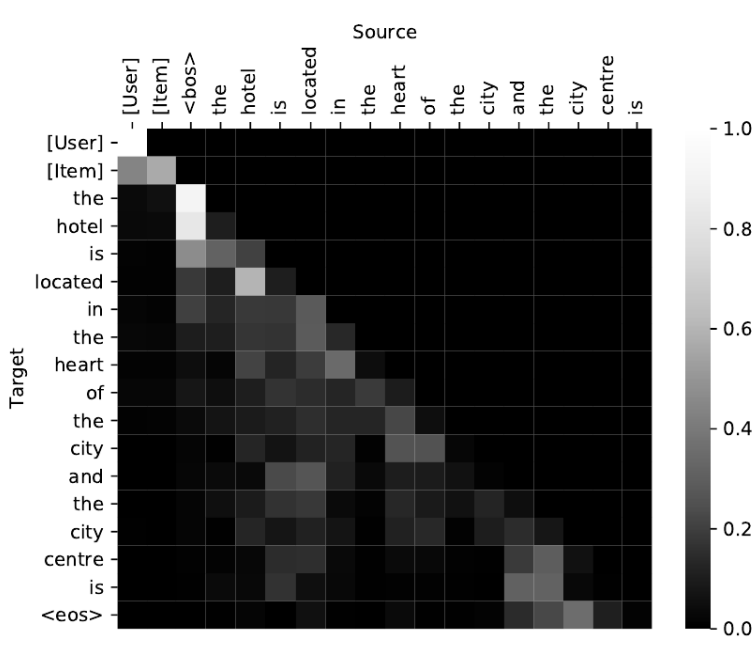
- Generación de lenguaje natural y sistemas recomendadores
- Usuarios no quieren solo una recomendación, sino que una explicación que justifique la recomendación

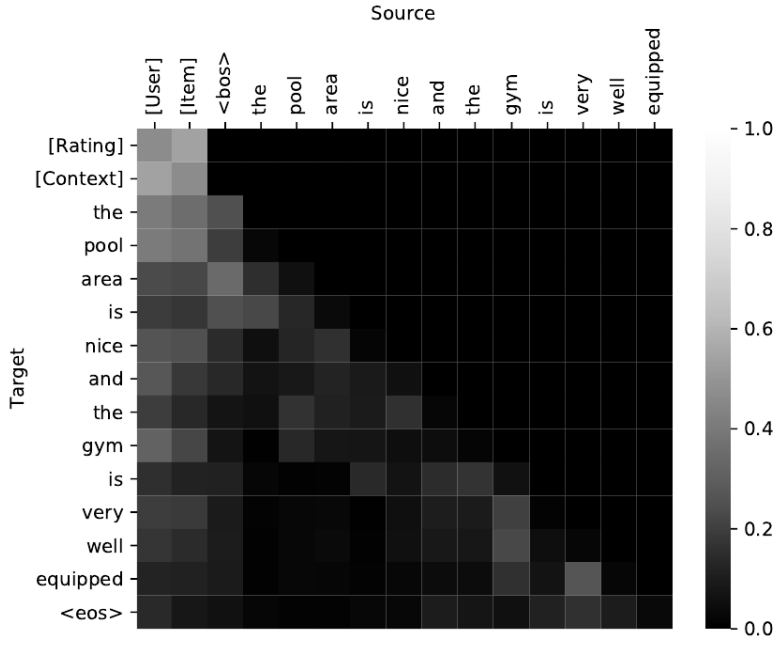
Contexto: Imaginen...

Problema de recomendación

- Oración $\hat{E}_{u,i}$ para un par (u, i)
- PETER también puede estimar $\hat{r}_{u,i}$
- Se pueden incorporar características de los artículos $F_{u,i}$

Limitaciones de los Transformers estándar





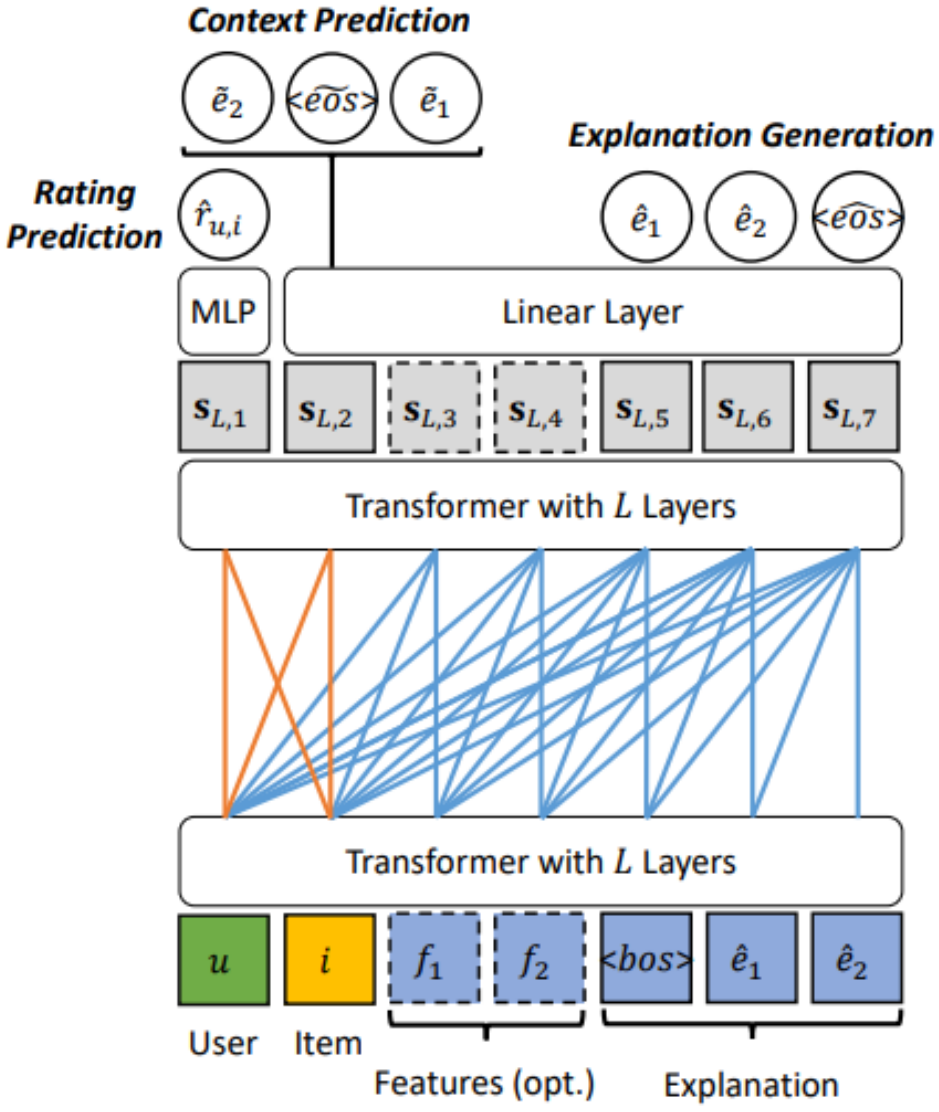
Contribución

- Según los autores: primer modelo basado en Transformer con generación de lenguaje natural personalizada
- Recomienda y explica
- Tarea: predicción de contexto
- Pequeño y eficiente

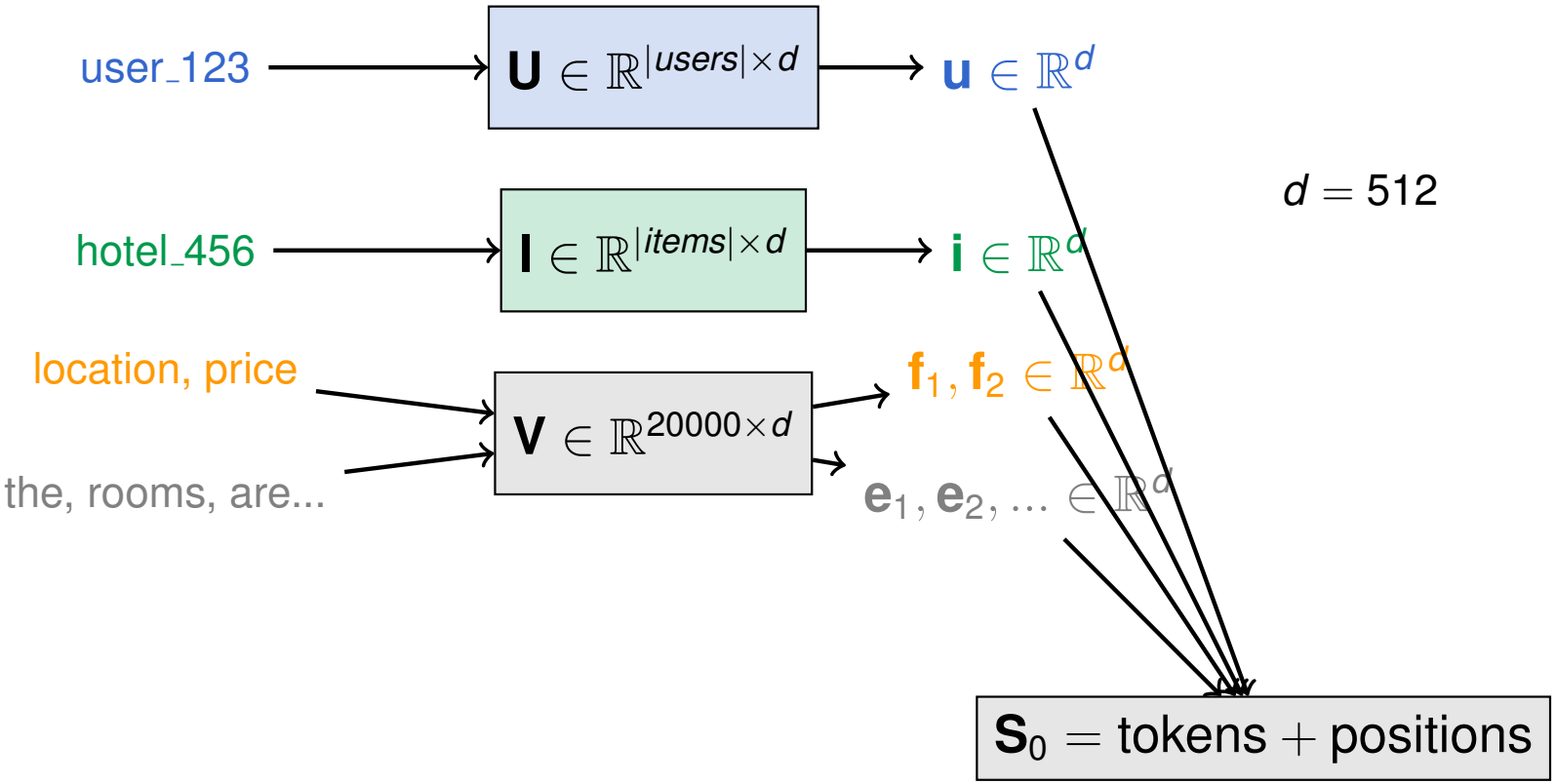
Estado del arte y marco teórico

- Recomendación explicable: LSTM; Hochreiter and Schmihuber, 1997. GRU; Cho et al., 2014
- Transformer: Vaswani et al., 2017. Devlin et al., 2019
- Generación personalizada: atributos personales; Zheng et al., 2020. Títulos de películas; Zhou et al., 2020. Características del artículo; Ni et al., 2019

Solución: Arquitectura PETER



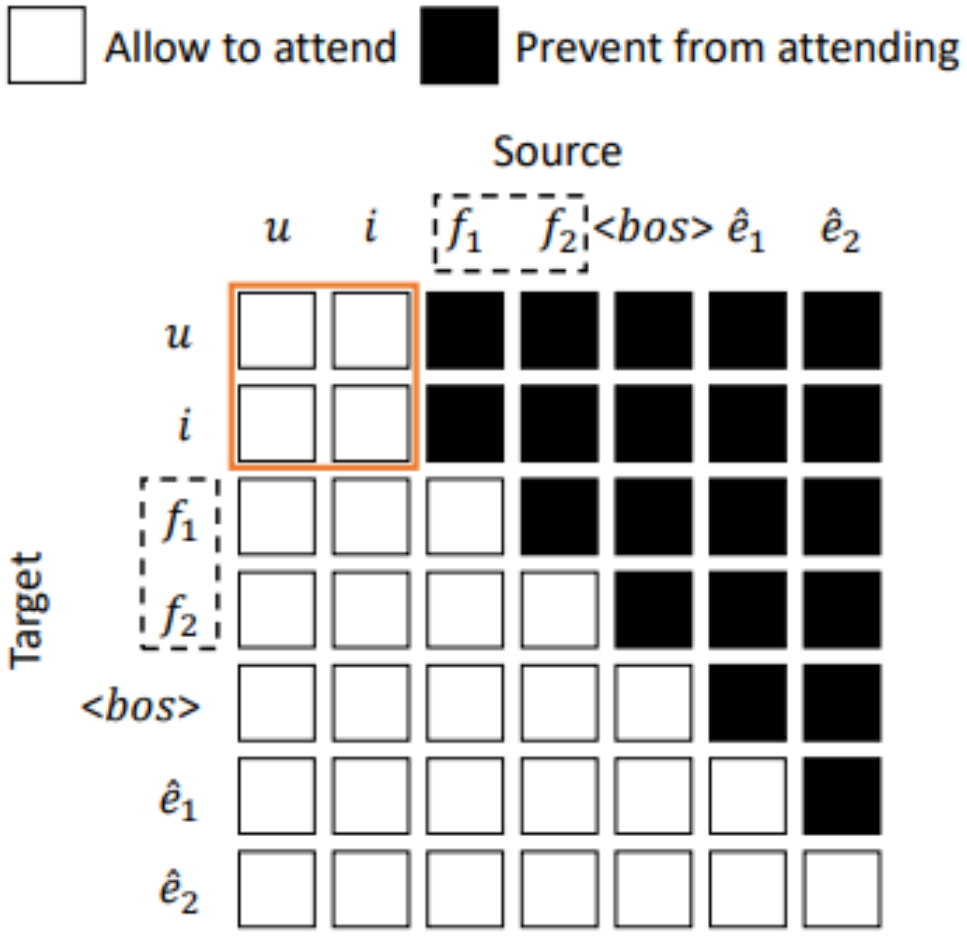
Representación de Entradas



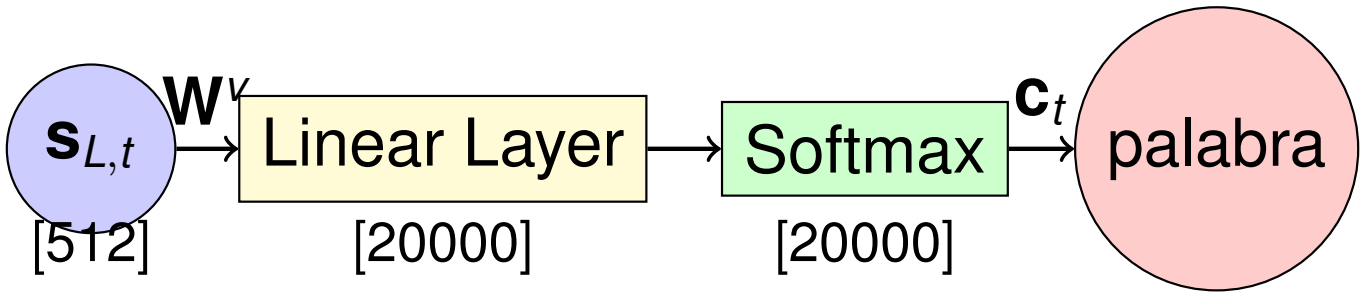
Ejemplo de secuencia:

$S = [u, i, f_1, f_2, \langle bos \rangle, e_1, e_2, ..., e_{15}]$

Máscara de atención de PETER



Generación de Explicación

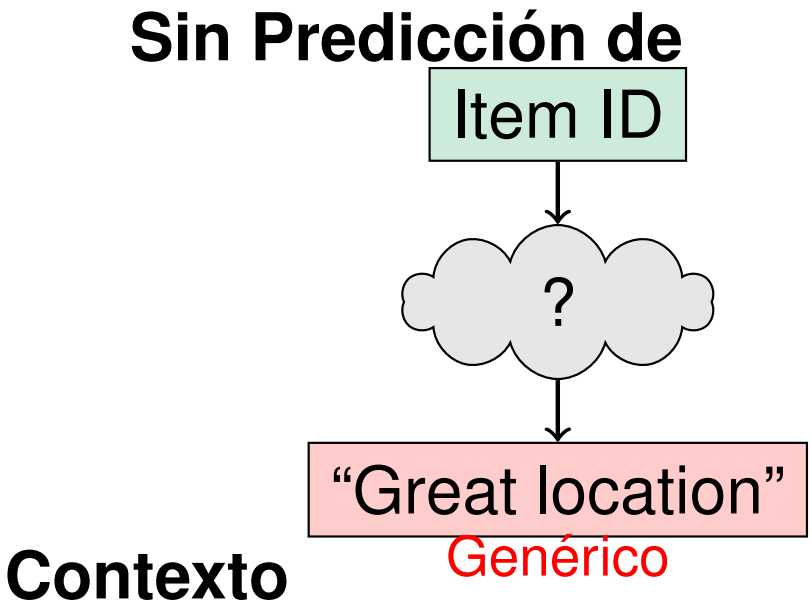


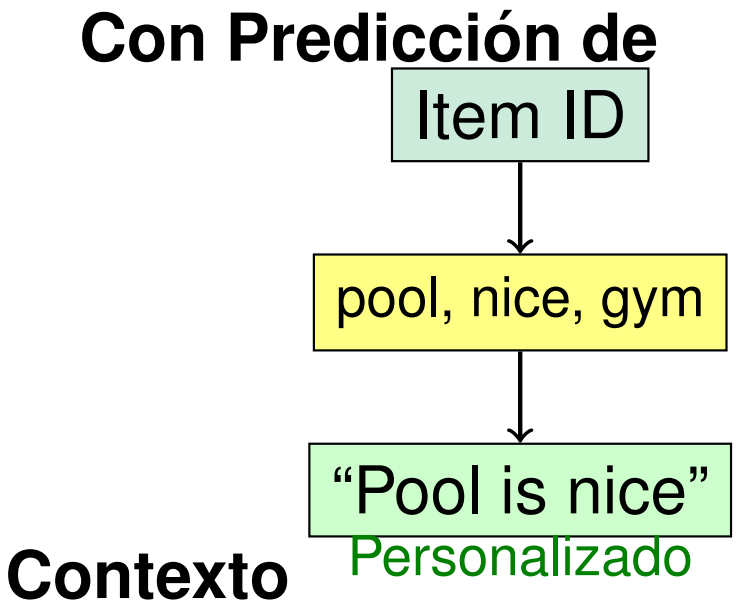
$$\mathcal{L}_e = -\frac{1}{|T|}\sum_{(u,i)\in T}\frac{1}{|E_{u,i}|}\sum_{t=1}^{|E_{u,i}|}\log c^{e_t}_{2+|F_{u,i}|+t}$$

Generación paso a paso:

$\langle bos \rangle \rightarrow$ "the" $\rightarrow$ "hotel" $\rightarrow$ "is" $\rightarrow$ ... $\rightarrow \langle eos \rangle$

Predicción de Contexto



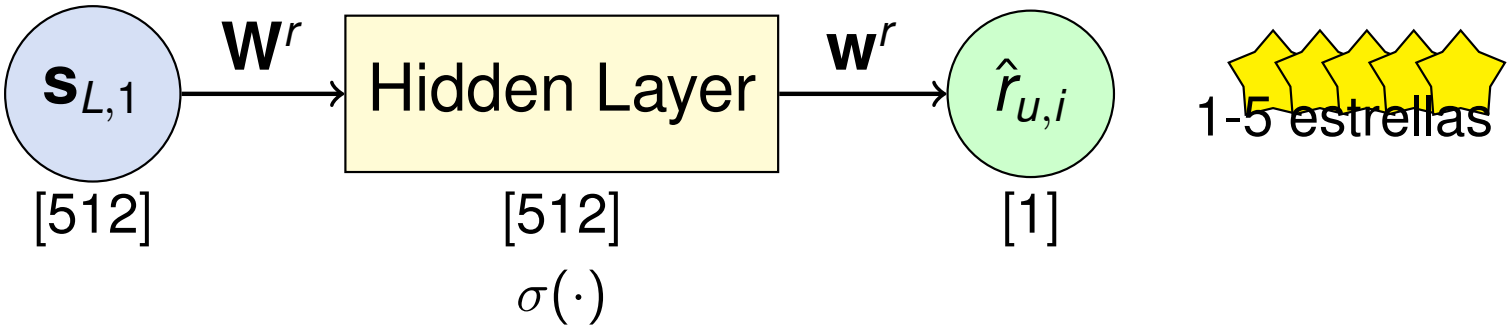


Mapeo ID $\rightarrow$ Palabras

Posición 2 (ítem) predice TODAS las palabras simultáneamente:

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

Predicción de Rating

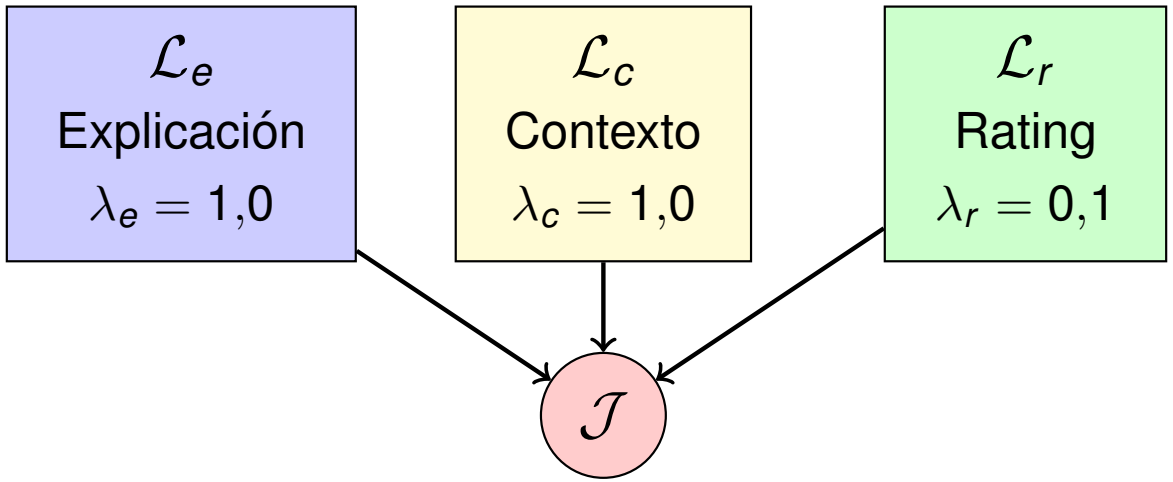


$$\hat{r}_{u,i} = \mathbf{w}^r \sigma(\mathbf{W}^r \mathbf{s}_{L,1} + \mathbf{b}^r) + \bar{b}^r$$

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

Aprendizaje Multitarea

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



Evaluación

- Dataset
- Métricas
- Modelos
- Resultados

Evaluación: Dataset

Se utilizaron tres Datasets

- Yelp
- Amazon
- TripAdvisor

	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

Evaluación: Métricas

- RMSE
- MAE
- BLEU-1
- BLEU-4
- ROUGE-1
- ROUGE-2
- USR
- FMR
- FCR
- DIV

Métricas: BLEU-n

Calcula qué tan parecida es la explicación del modelo con la explicación de testing utilizando n-gram

$$\text{BLEU Score} = BP \cdot \exp\left(\sum_{i=1}^{n} w_i \cdot \log(p_i)\right) \tag{1}$$

Donde:

- BP corresponde a una penalización si es que la explicación generada por el modelo es más corta que la baseline
- w_i corresponde al peso que se le da a cada i-gram
- p_i es la coincidencia de los i-gram del modelo con la i-gram baseline

Métricas: Rouge-n

Calcula distintas métricas en base a la superposición de los n-grams de la explicación generada y la explicación testing.

$$\text{Rouge score} = \{\textit{Precision}@\text{n-gram}, \textit{Recall}@\text{n-gram}, F1@\text{n-gram}\} \tag{2}$$

Donde:

- *Precision*@n-gram corresponde a la división entre el numero de n-grams que coinciden en la superposición y el numero de n-grams totales generado por el modelo

Métricas: Rouge-n

- *Recall*@n-gram corresponde a la división entre el numero de n-grams que coinciden en la superposición y el numero de n-grams totales del baseline.
- *F*1@n-gram es la métrica usual vista en clases usando las dos métricas anteriores.

Métricas: Unique Sentence Ratio (USR)

Calcula la división entre el número de explicaciones únicas y el número de explicaciones totales.

$$\text{USR} = \frac{|S|}{N} \tag{3}$$

Donde:

- S es el set de explicaciones únicas
- N es el numero de explicaciones totales.

Métricas: Feature Matching Ratio (FMR)

Mide cuándo una explicación generada tiene alguna característica de la explicación baseline

$$\text{FMR} = \frac{1}{N} \sum_{u,i} \delta(f_{u,i} \in \hat{S}_{u,i}) \tag{4}$$

Donde:

- N es el numero de explicaciones totales.
- $\hat{S}_{u,i}$ es la explicación generada para algún par usuario-item
- $f_{u,i}$ es la feature dada para ese usuario-item
- $\delta(x) = 1$ si x es TRUE y $\delta(x) = 0$ para cualquier otro caso.

Métricas: Feature Coverage Ratio (FCR)

Mide cuántas características diferentes hay dentro de todas las explicaciones.

$$\text{FCR} = \frac{N_g}{|\mathcal{F}|} \tag{5}$$

Donde:

- N_g es el numero de características diferentes dentro de las explicaciones generadas por el modelo
- $\mathcal{F}$ es el conjunto de todas las características.

Métricas: Feature Diversity (DIV)

Mide la intersección de las características entre cualquier par de explicaciones generadas.

$$\text{DIV} = \frac{2}{N \cdot (N-1)} \sum_{u,u',i,i'} \left| \hat{\mathcal{F}}_{u,i} \cap \hat{\mathcal{F}}_{u',i'} \right| \tag{6}$$

Donde:

- N es el numero de explicaciones totales.
- $\hat{\mathcal{F}}_{u,i}$ son el conjunto de las características presentes en una explicación generada por el modelo para el usuario u y para el item i
- $\hat{\mathcal{F}}_{u',i'}$ son el conjunto de las características presentes en una explicación generada por el modelo para el usuario u' y para el item i'

Evaluación: Modelos

Modelos de explicación

- Transformer
- NRT
- Att2Seq
- PETER
- *ACMLM
- *NETE
- *PETER+

Modelos de recomendación

- NRT
- PETER
- *NETE
- PMF
- SVD++

Todos los modelos con * aceptan características como input para el entrenamiento.

Evaluación: Resultados de explicabilidad

	Explainability			Text Quality								
	FMR↑	FCR↑	DIV↓	USR↑	B1↑	B4↑	R1-P↑	R1-R↑	R1-F↑	R2-P↑	R2-R↑	R2-F↑
						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****
						Amazon						
Transformer	0.10	0.01	3.26	0.00	9.71	0.59	19.68	11.94	14.11	2.10	1.39	1.55
NRT	**0.12**	0.07	2.93	0.17	**12.93**	0.96	**21.03**	13.57	**15.56**	2.71	1.84	2.05
Att2Seq	**0.12**	0.20	2.74	**0.33**	12.56	0.95	20.79	13.31	15.35	2.62	1.78	1.99
PETER	**0.12****	**0.21**	**1.75****	0.29	12.77	**1.17****	19.81	**13.80**	15.23	**2.80**	**2.08****	**2.20****
ACMLM	0.10	**0.31**	2.07	**0.96**	9.52	0.22	11.65	10.39	9.69	0.71	0.81	0.64
NETE	0.71	0.19	1.93	0.57	18.76	2.46	33.87	21.43	24.81	7.58	4.77	5.46
PETER+	**0.77****	**0.31****	**1.20****	0.46	**19.75****	**3.06****	**34.71****	**23.99****	**26.35****	**9.04****	**6.23****	**6.71****
						TripAdvisor						
Transformer	0.04	0.00	10.00	0.00	12.79	0.71	16.52	**16.38**	15.88	2.22	**2.63**	**2.34**
NRT	0.06	0.09	4.27	0.08	15.05	0.99	18.22	14.39	15.40	2.29	1.98	2.01
Att2Seq	0.06	**0.15**	4.32	**0.17**	15.27	1.03	18.97	14.72	15.92	**2.40**	2.03	2.09
PETER	**0.07****	0.13	**2.95****	0.08	**15.96****	**1.11***	**19.07**	16.09	**16.48****	2.33	2.17	2.09
ACMLM	0.07	**0.41**	**0.78**	**0.94**	3.45	0.02	4.86	3.82	3.72	0.18	0.20	0.16
NETE	0.78	0.27	2.22	0.57	22.39	3.66	35.68	24.86	27.71	10.20	6.98	7.66
PETER+	**0.89****	0.35	1.61	0.25	**24.32****	**4.55****	**37.48****	**29.21****	**30.49****	**11.92****	**8.98****	**9.24****

Evaluación: Análisis de tiempo

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

Evaluación: Análisis cualitativo

	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>

Evaluación: Resultados de recomendación

	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

Evaluación: Análisis de ablación

	Explainability			Text Quality			Recommendation	
	FMR	FCR	DIV	USR	BLEU-1	BLEU-4	RMSE	MAE
Disable $\mathcal{L}_c$	0.06 ↓	0.03 ↓	5.75 ↓	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

Referencias

- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory.
- Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

Referencias

- Yinhe Zheng, Rongsheng Zhang, Minlie Huang, and Xiaoxi Mao. 2020. A pre-training based personalized dialogue generation model with persona-sparse data.
- Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020. Improving conversational recommender systems via knowledge graph based semantic fusion.
- Piji Li, Zihao Wang, Zhaochun Ren, Lidong Bing, and Wai Lam. 2017. Neural rating regression with abstractive tips generation for recommendation.
- Li Dong, Shaohan Huang, Furu Wei, Mirella Lapata, Ming Zhou, and Ke Xu. 2017. Learning to generate product reviews from attributes.

Referencias

- Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects.
- Lei Li, Yongfeng Zhang, and Li Chen. 2020c. Generate neural template explanations for recommendation.
- Andriy Mnih and Russ R Salakhutdinov. 2007. Probabilistic matrix factorization.
- Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model.

Preguntas