task

Explainable Recommendation with Personalized Review Retrieval and Aspect Learning

method

Advisor : Jia-Ling, Koh

Speaker : Hsuan Lu

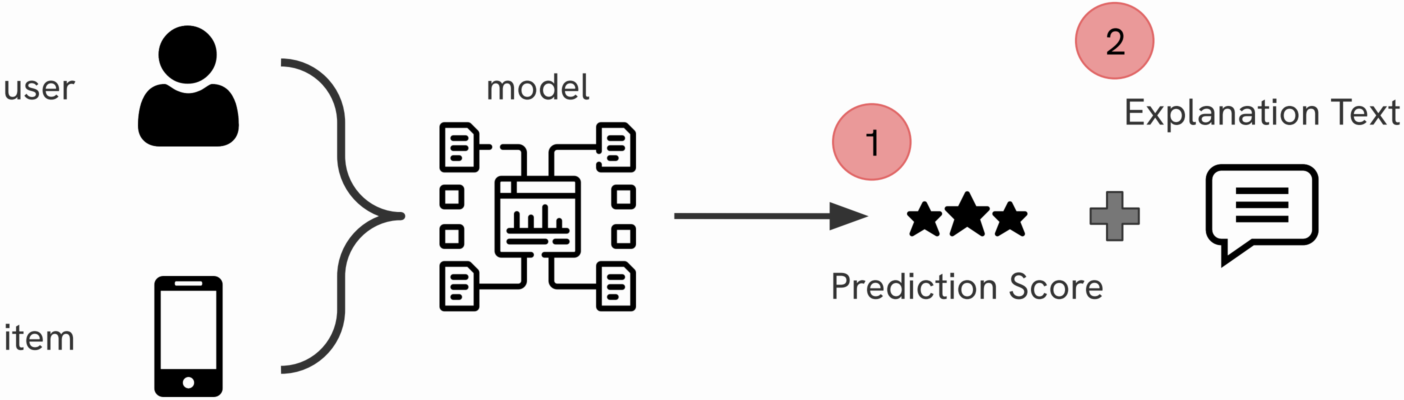
Source : ACL’2023

Date : 2024/02/02

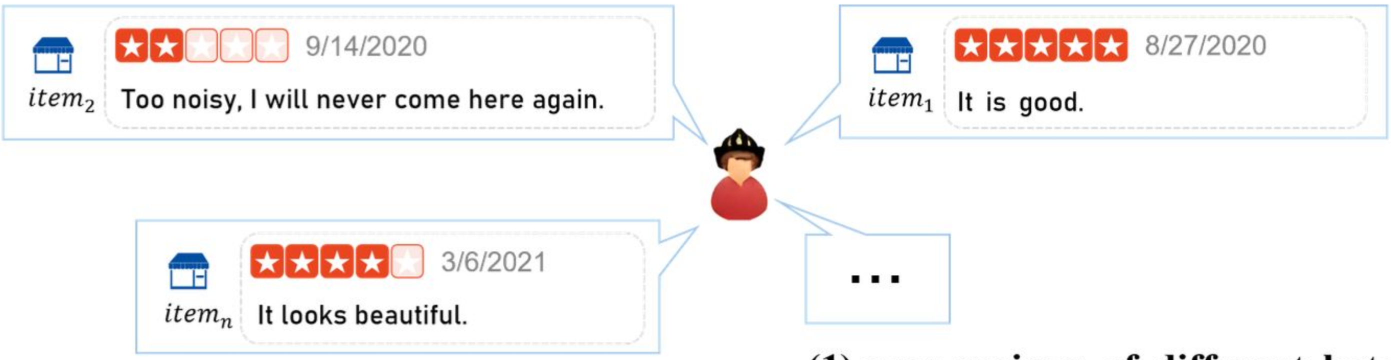
Outline

- Introduction
- Method
- Experiment
- Conclusion

Explainable Recommendation



Explainable Recommendation



(1) user reviews of different hotels

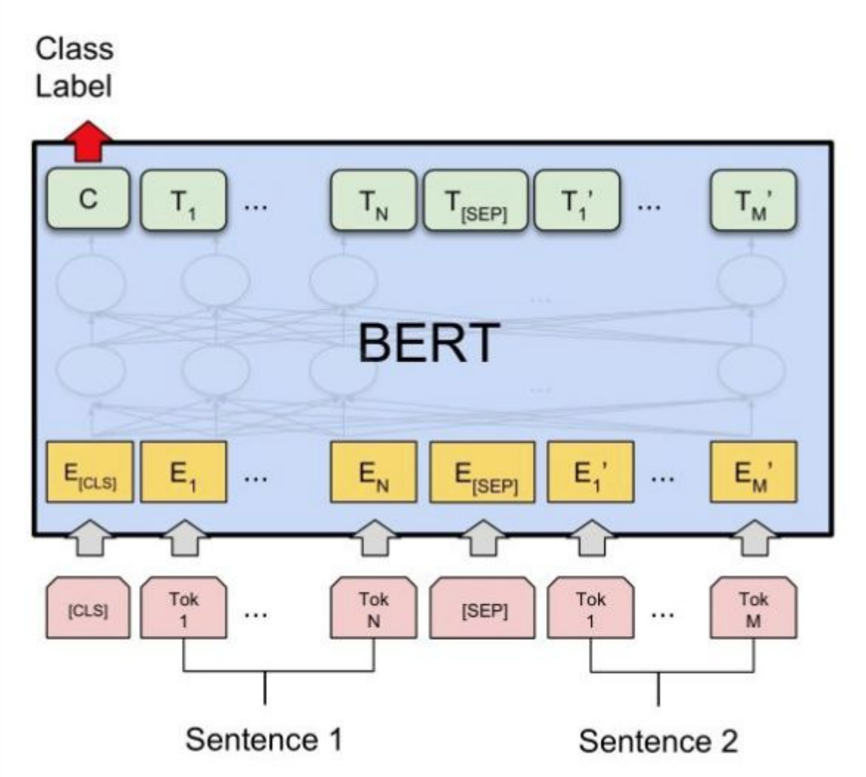
retrieval method The hotel is very comfortable to live in, and the environment is very pleasant. The only fly in the ointment is that the sound insulation is not very good.

Retrieval Corpus

Other: Wikipedia

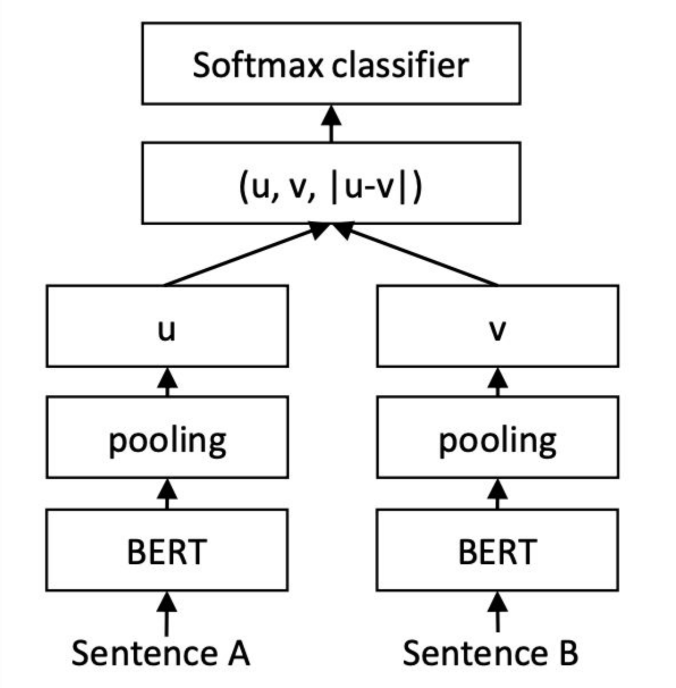
This paper: Reviews in the dataset

BERT

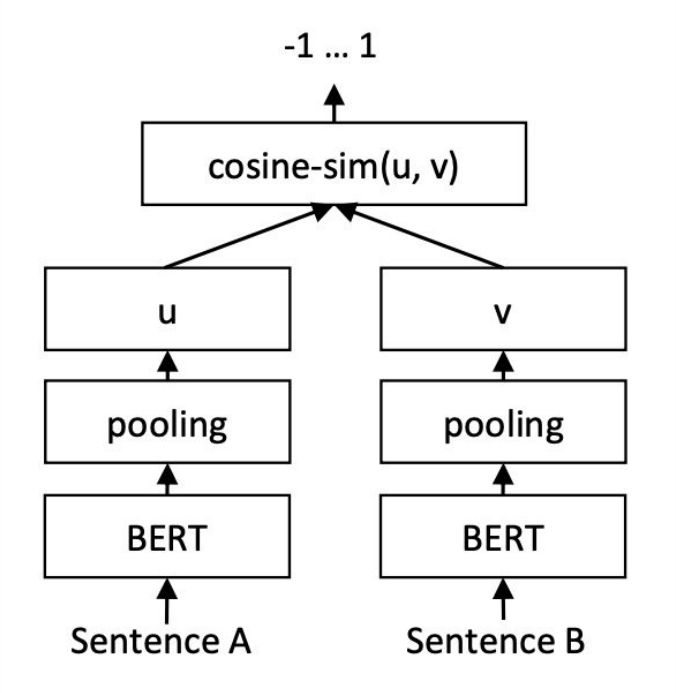


Class
Label
C
T_1
...
T_N
$T_{[SEP]}$
T_1'
...
T_M'
BERT
$E_{[CLS]}$
E_1
...
E_N
$E_{[SEP]}$
E_1'
...
E_M'
[CLS]
Tok 1
...
Tok N
[SEP]
Tok 1
...
Tok M
Sentence 1
Sentence 2

Sentence-BERT (SBERT)



training

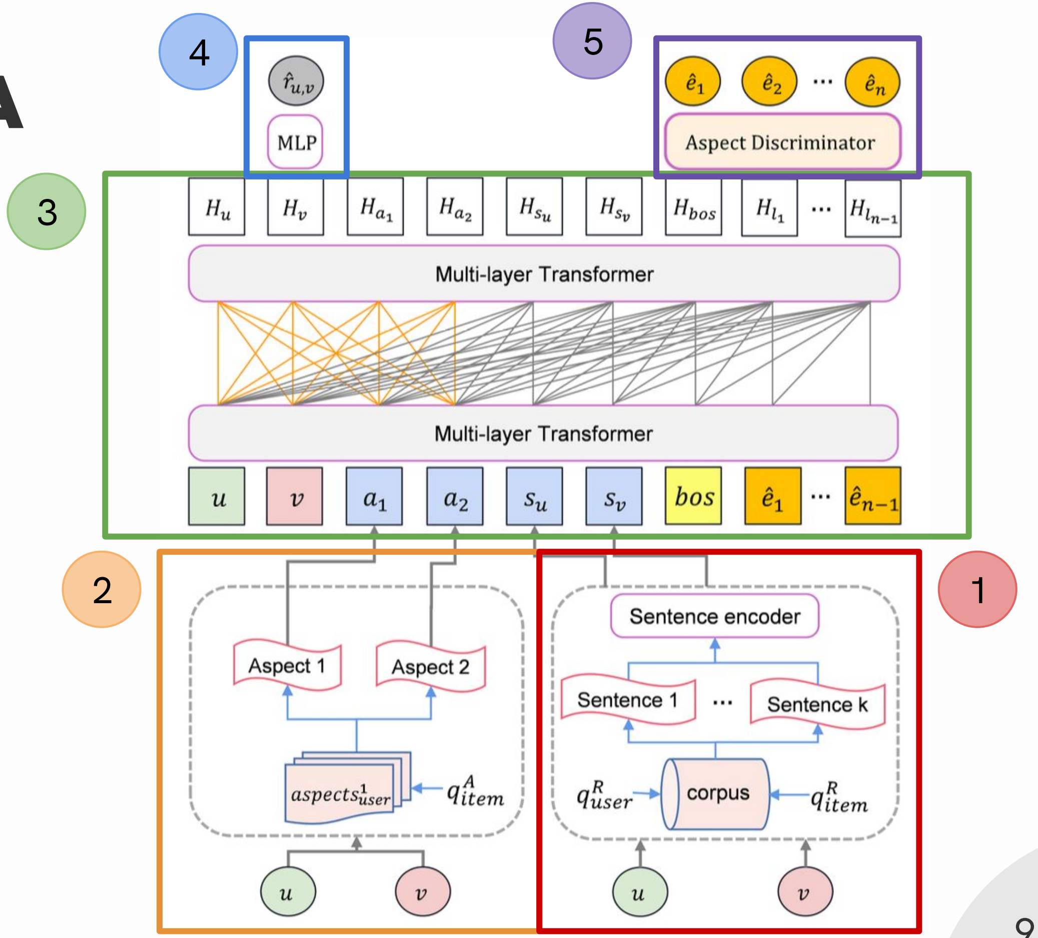


testing

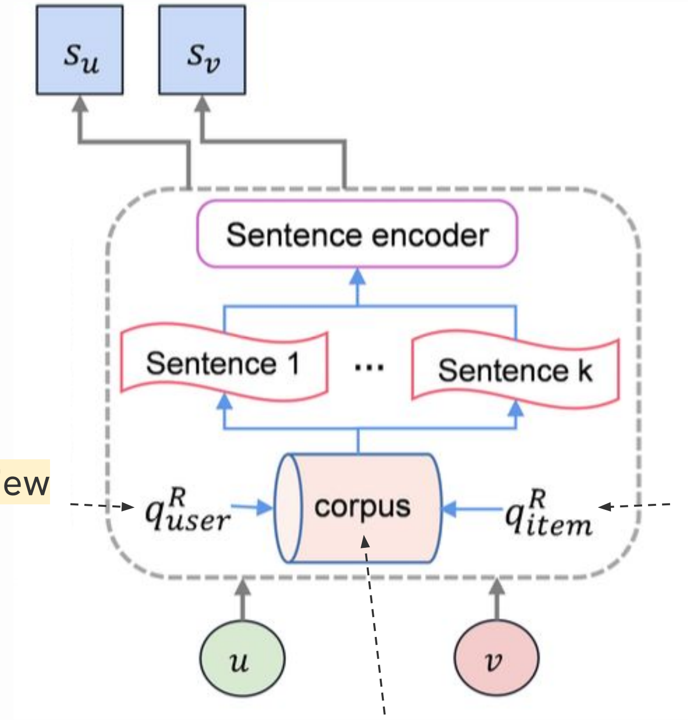
Outline

- Introduction
- Method
- Experiment
- Conclusion

Overview of ERRA



Retrieval Enhancement

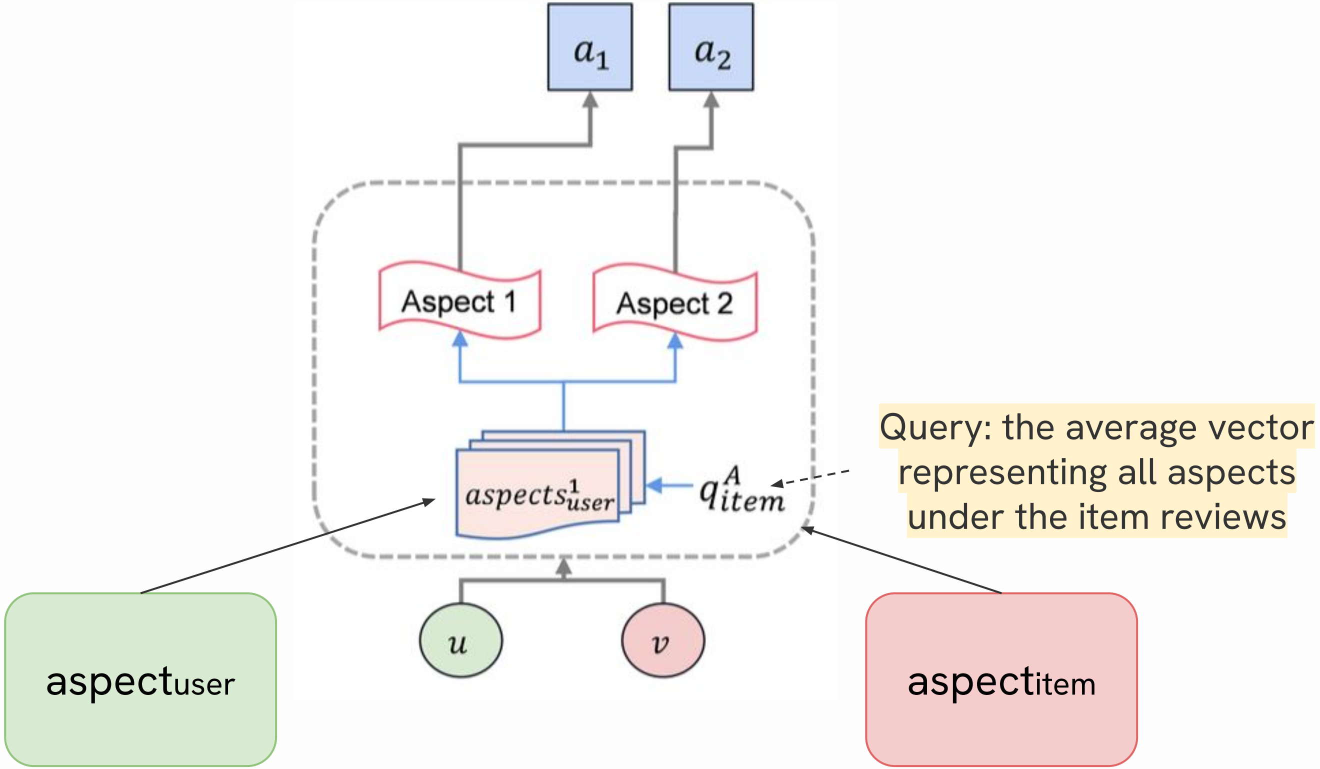


Query: the average of the review embedding of each user

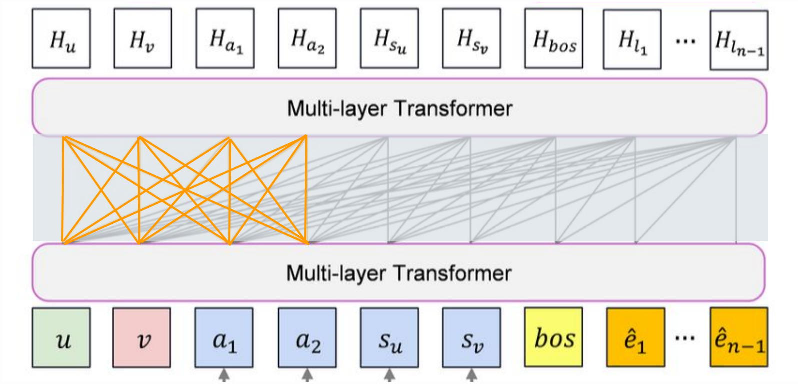
Query: the average of the review embedding of each item

from training set

Aspect Enhancement



Joint Enhancement Transformers



Multi-layer Transformer

$$\mathbf{A}_{i,h} = \text{softmax}\left(\frac{\mathbf{Q}_{i,h}\mathbf{K}_{i,h}^{\top}}{\sqrt{d}}\right)\mathbf{V}_{i,h} \qquad (1)$$

$$\mathbf{Q}_{i,h} = \mathbf{S}_{i-1}\mathbf{W}_{i,h}^{Q}, \mathbf{K}_{i,h} = \mathbf{S}_{i-1}\mathbf{W}_{i,h}^{K}, \qquad (2)$$

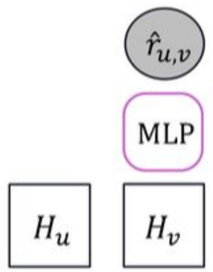
$$\mathbf{V}_{i,h} = \mathbf{S}_{i}\mathbf{W}_{i,h}^{V} \qquad (3)$$

$$\mathbf{S}_{i-1} \in \mathbb{R}^{|S|\times d} \qquad \mathbf{W}_{i,h}^{Q}, \mathbf{W}_{i,h}^{K}, \mathbf{W}_{i,h}^{V} \in \mathbb{R}^{d\times\frac{d}{H}}$$

i 前一層的 output

|S| : the length of the input sequence
d: the dimension of embeddings

Rating Prediction

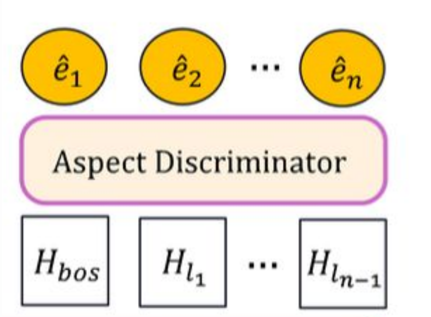


Multi-layer Perceptron (MLP)

$$\hat{r}_{u,v} = \mathrm{ReLU}\left([H_v, u_{id}, v_{id}]\mathbf{W}_{l,1}\right)\mathbf{W}_{l,2} \qquad (5)$$

$$\mathbf{W}_1 \in \mathbb{R}^{3d\times d}, \mathbf{W}_2 \in \mathbb{R}^{d\times 1}$$

Explanation Generation



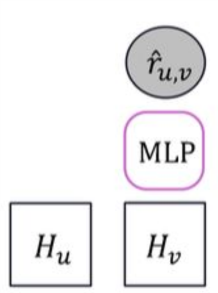
Auto-regressive Methodology

$$\mathbf{e}_t = \mathrm{softmax}\left(\mathbf{W}^v \mathbf{H}_{L,t} + \mathbf{b}^v\right) \tag{7}$$

the probability distribution over the vocabulary V

$\mathbf{W}^v \in \mathbb{R}^{|\mathcal{V}| \times d}$ **and** $\mathbf{b}^v \in \mathbb{R}^{|\mathcal{V}|}$

Loss Function of Rating Prediction



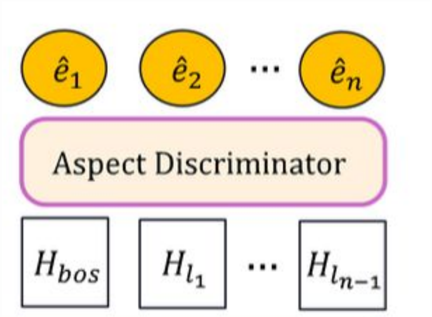
RMSE

Ground Truth Rating

$$\mathcal{L}_r = \frac{1}{|\mathcal{T}|} \sum_{(u,v)\in\mathcal{T}} (r_{u,v} - \hat{r}_{u,v})^2 \qquad (6)$$

training set

Loss Function of Explanation Generation

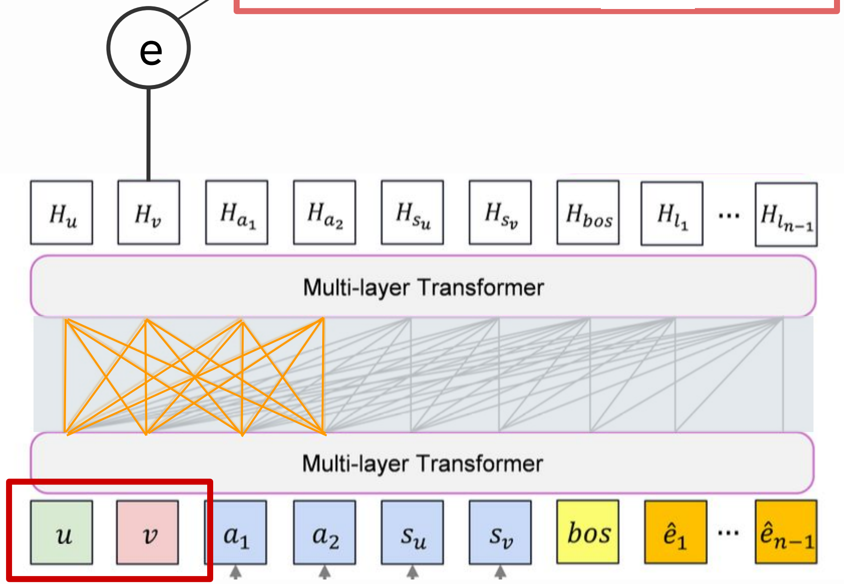


$$\mathcal{L}_g = \frac{1}{|\mathcal{T}|} \sum_{(u,v)\in\mathcal{T}} \frac{1}{|t_{len}|} \sum_{t=1}^{|t_{len}|} -\log e_{6+t}^{g_t} \quad (9)$$

training set

Loss Function of Context Prediction

$$e = \mathrm{softmax}\left(\mathbf{W}^v H_v^{g_{ti}} + \mathbf{b}^v\right)$$

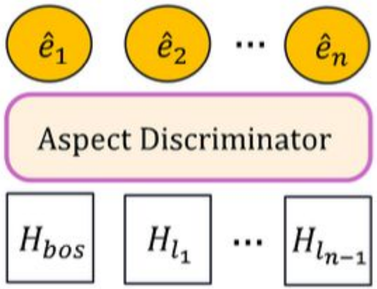


$$\mathcal{L}_c = \sum_{(u,v)\in\mathcal{T}} \frac{1}{|t_{len}|} \sum_{t=1}^{|t_{len}|} -\log H_v^{g_{ti}} \qquad (4)$$

training set

g_{ti} : only use the hidden vector of the position Hv to generate the i-th word

Loss Function of Aspect Discriminator



$$\mathcal{L}_a = \frac{1}{|\mathcal{T}|} \sum_{(u,v)\in\mathcal{T}} \frac{1}{|t_{len}|} \sum_{t=1}^{|t_{len}|} \left(-\tau_a \log e_{t,a}\right) \qquad (8)$$

training set

$$\boldsymbol{\tau} \in \mathbb{R}^{|\mathcal{V}|}$$

the aspects that interest this user

$$\boldsymbol{\tau} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Multi-Task Learning

the loss function of context prediction

the loss function of aspect discriminator

$$\mathcal{L} = pl\mathcal{L}_r + \lambda_c\mathcal{L}_c + gl\mathcal{L}_g + al\mathcal{L}_a + \lambda_l||\Theta||_2^2 \quad (10)$$

RMSE

the loss function of text generation

all the neural parameters

Outline

- Introduction
- Method
- Experiment
- Conclusion

Dataset

- Amazon
  - cell phones
- Yelp
  - restaurants
- TripAdvisor
  - hotels

Datasets	Yelp	Amazon	TripAdvisor
Number of users	27,147	157,212	9,765
Number of items	20,266	48,186	6,280
Number of reviews	1,293,247	1,128,437	320,023
Records per user	47.64	7.18	32.77
Records per item	63.81	23.41	50.96

Evaluation

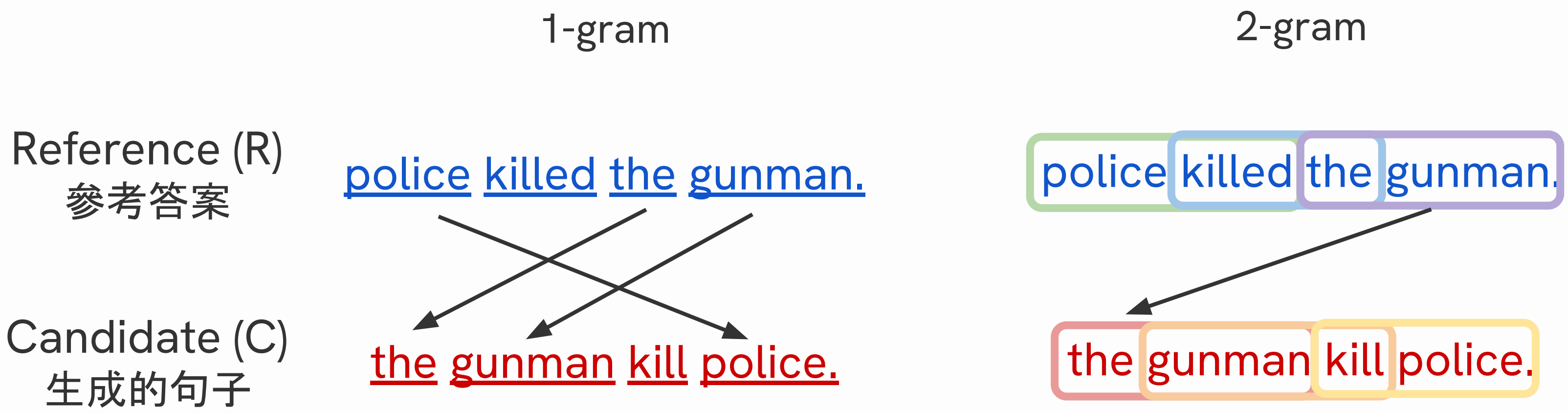
- Prediction
  - Root Mean Square Error (RMSE)
  - Mean Absolute Error (MAE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$$

- Explainability
  - BLEU
  - ROUGE
  - BERTscore

N-gram


1-gram
2-gram
Reference (R)
參考答案
police killed the gunman.
police killed the gunman.
Candidate (C)
生成的句子
the gunman kill police.
the gunman kill police.

BLEU-N—Precision

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log \boxed{p_n}\right).$$

Candidates中n-grams出現的次數

$$Count_{clip} = \min(Count, Max\_Ref\_Count)$$

Reference中n-gram出現最多的次數

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-}gram \in C} Count_{clip}(n\text{-}gram)}{\boxed{\sum_{C' \in \{Candidates\}} \sum_{n\text{-}gram' \in C'} Count(n\text{-}gram')}}.$$

Candidates中n-gram'的個數

1-gram

Reference (R)
參考答案

police killed the gunman.

Candidate (C)
生成的句子

the the the the the.

~~P1 = 5/ 1~~

P1 = ⅕

BLEU-N — Precision

$$\text{BLEU} = \boxed{\text{BP}} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right).$$

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}.$$

c: len(Candidate)
r: len(Reference)

1-gram

Reference (R)
參考答案

Going to play basketball in the afternoon.

Candidate (C)
生成的句子

Going to play basketball this afternoon.

BP = e ^ (1 - 7/6)

ROUGE

- ROUGE-2
  - R2-P
  - R2-R
  - R2-F
- ROUGE-L
  - RL-P
  - RL-R
  - RL-F

P : Precision
R : Recall
F : F1

ROUGE-2

N = 2

gramn 在C、R中共同出現的次數

ROUGE-N

$$= \frac{\sum_{S \in \{ReferemceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad (1)$$

Reference (R)
參考答案

police killed the gunman.

Candidate (C)
生成的句子

the gunman kill police.

ROUGE-2 = ⅓

ROUGE-2

Reference (R)
參考答案

police killed the gunman.

Candidate (C)
生成的句子

the gunman kill police.

R2-P = Count_match / Candidate_count = $\frac{1}{3}$

R2-R = Count_match / Reference_count = $\frac{1}{3}$

R2-F = $2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ = $(2 * \frac{1}{3} * \frac{1}{3}) / (\frac{1}{3} + \frac{1}{3})$

ROUGE-L

LCS: Longest Common Subsequence

m: len(reference)

n: len(candidate)

$$R_{lcs} = \frac{LCS(X,Y)}{m} \quad (2)$$

$$P_{lcs} = \frac{LCS(X,Y)}{n} \quad (3)$$

$$F_{lcs} = \frac{(1+\beta^2)R_{lcs}P_{lcs}}{R_{lcs} + \beta^2 P_{lcs}} \quad (4)$$

Reference (R)
參考答案

police killed <u>the</u> <u>gunman</u>.

Candidate (C)
生成的句子

<u>the</u> <u>gunman</u> kill police.

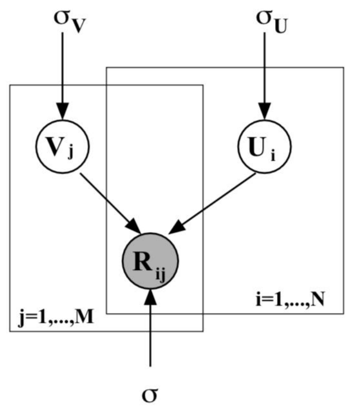
RL-R = ½

RL-P = ½

RL-F = (2 * ½ * ½) / (½ + ½)

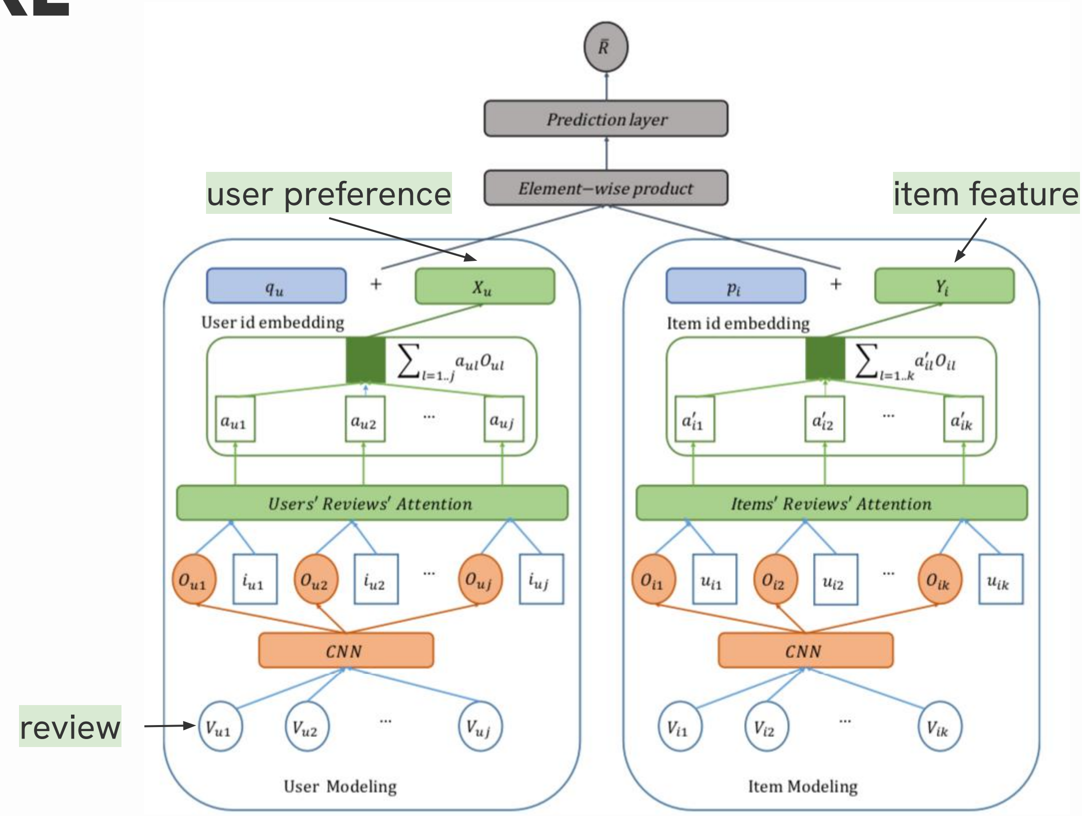
Baseline

- Factorization Method
  - PMF (Probabilistic Matrix Factorization)
  - SVD++
- Deep Learning Model
  - NARRE
  - NRT
  - CAML
  - ReXPlug
  - PETER



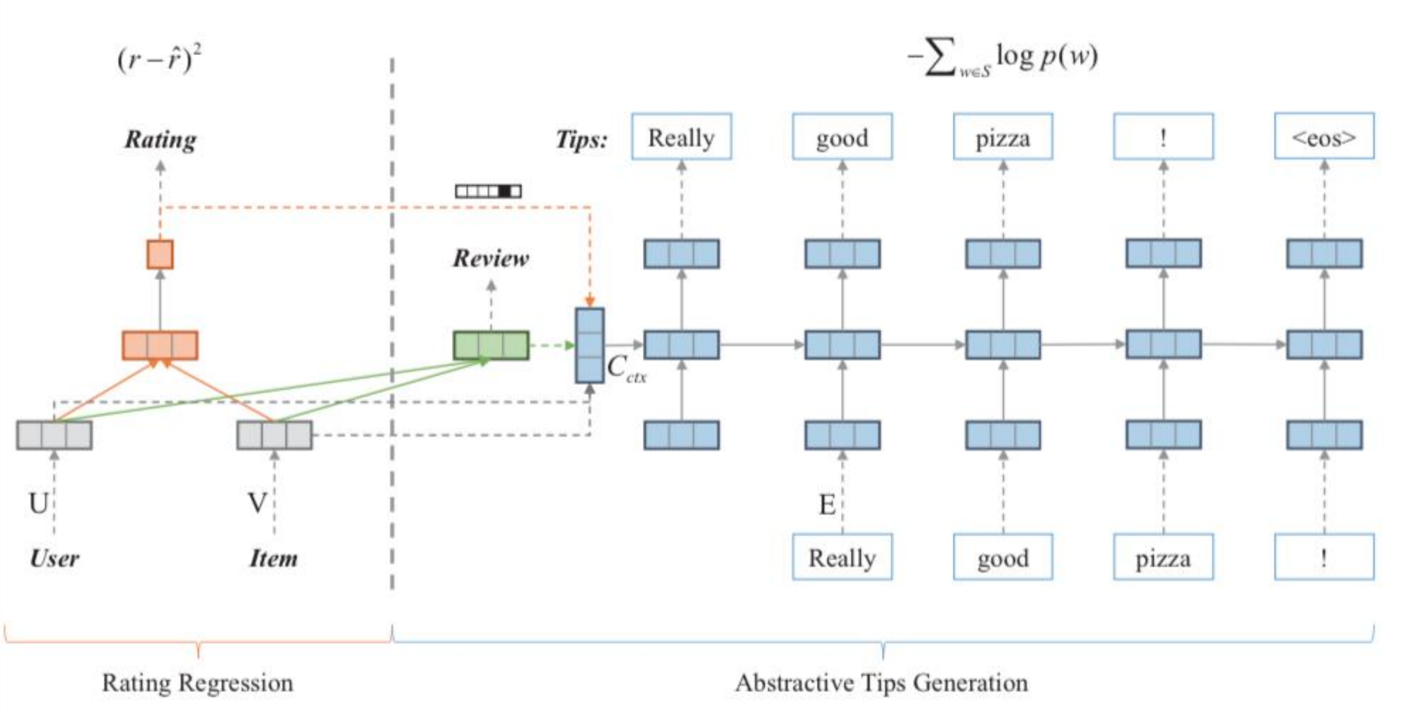
PMF

NARRE



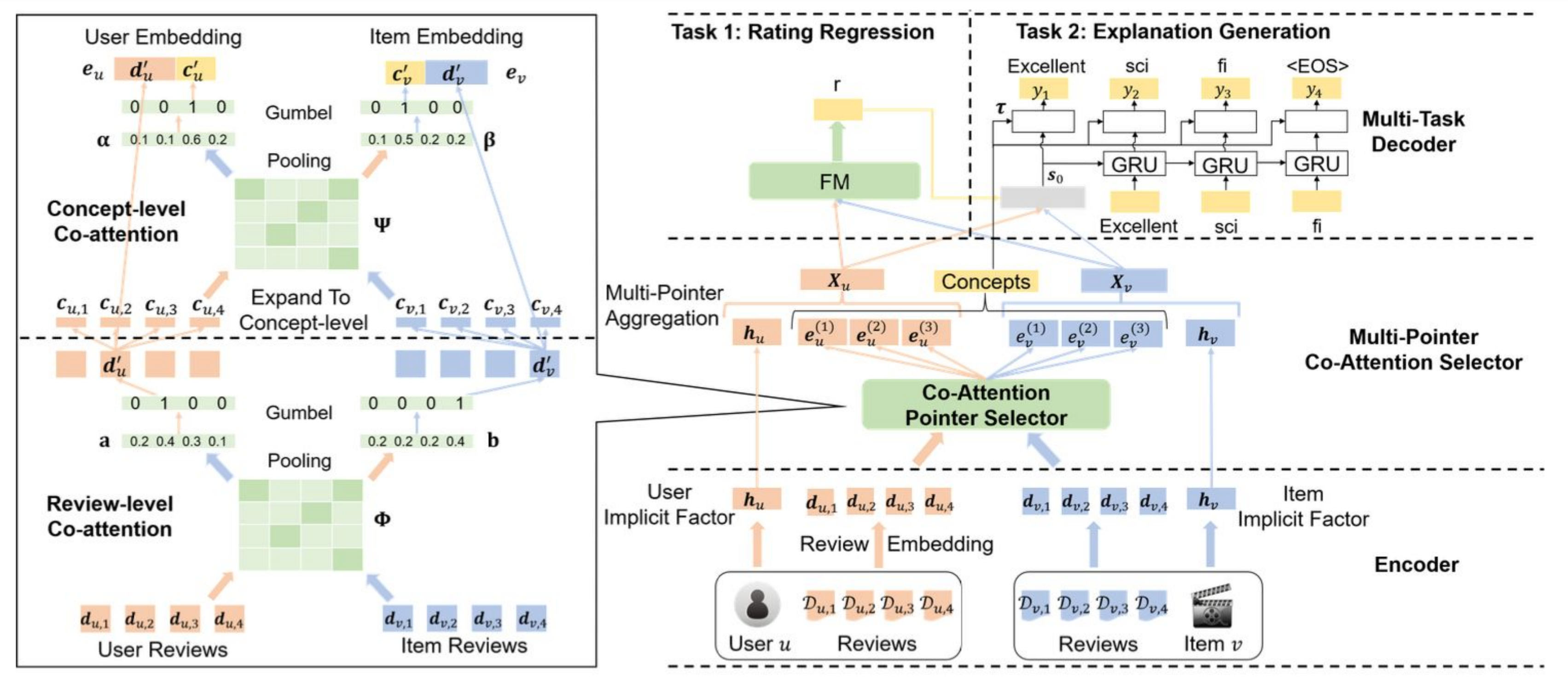
$\bar{R}$
Prediction layer
Element–wise product
user preference
item feature
q_u
+
X_u
User id embedding
$\sum_{l=1..j} a_{ul} O_{ul}$
a_{u1}
a_{u2}
a_{uj}
Users' Reviews' Attention
O_{u1}
i_{u1}
O_{u2}
i_{u2}
O_{uj}
i_{uj}
CNN
review
V_{u1}
V_{u2}
V_{uj}
User Modeling
p_i
+
Y_i
Item id embedding
$\sum_{l=1..k} a'_{il} O_{il}$
a'_{i1}
a'_{i2}
a'_{ik}
Items' Reviews' Attention
O_{i1}
u_{i1}
O_{i2}
u_{i2}
O_{ik}
u_{ik}
CNN
V_{i1}
V_{i2}
V_{ik}
Item Modeling

NRT



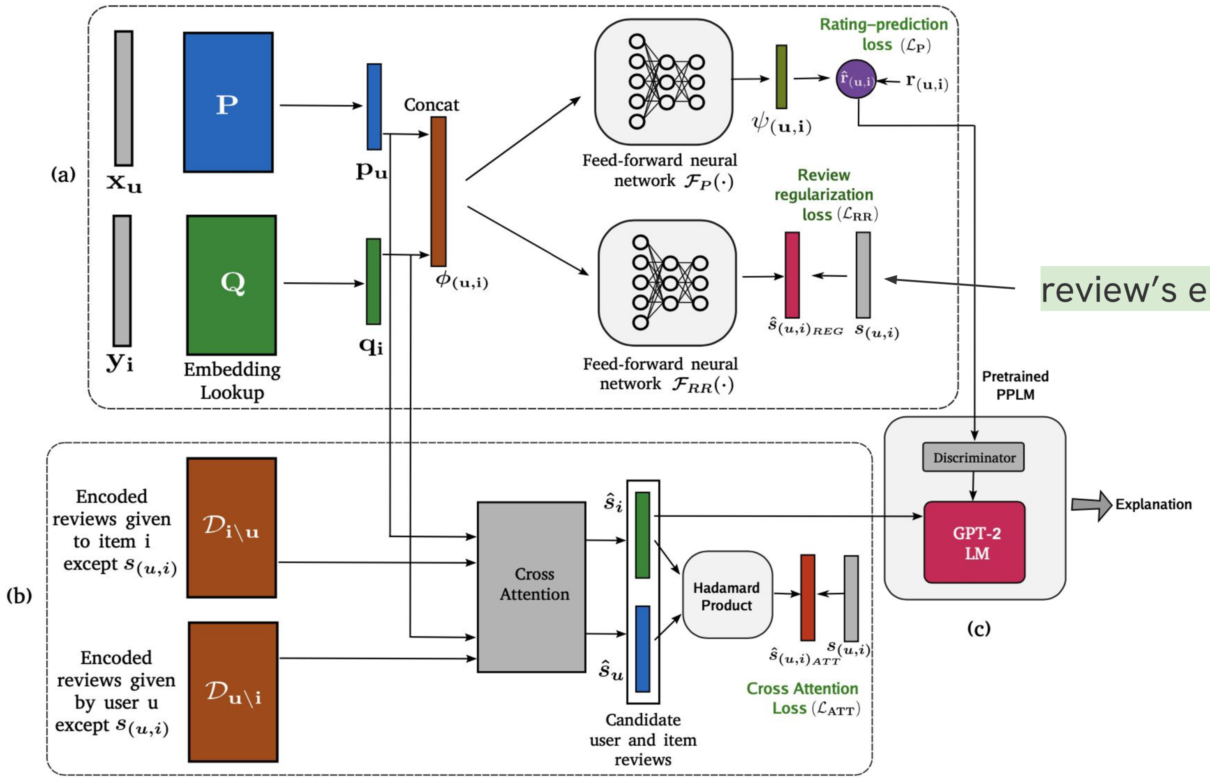
$(r-\hat{r})^2$
$-\sum_{w \in S} \log p(w)$
Rating
Tips:
Really
good
pizza
!
<eos>
Review
C_{ctx}
U
V
E
User
Item
Really
good
pizza
!
Rating Regression
Abstractive Tips Generation

CAML



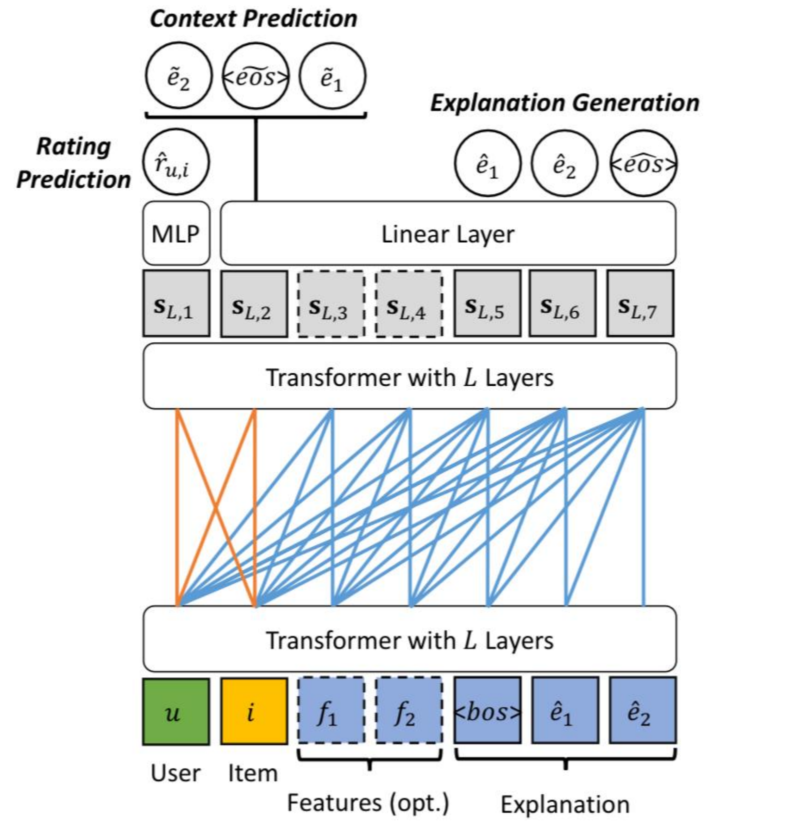
User Embedding
Item Embedding
Gumbel
Pooling
Concept-level Co-attention
Expand To Concept-level
Review-level Co-attention
User Reviews
Item Reviews
Task 1: Rating Regression
Task 2: Explanation Generation
FM
Excellent
sci
fi
<EOS>
Multi-Task Decoder
GRU
Concepts
Multi-Pointer Aggregation
Multi-Pointer Co-Attention Selector
Co-Attention Pointer Selector
User Implicit Factor
Review Embedding
Item Implicit Factor
Encoder
Reviews

ReXPlug



review's embedding

PETER



Context Prediction
$\tilde{e}_2$
<$\widetilde{eos}$>
$\tilde{e}_1$
Explanation Generation
Rating Prediction
$\hat{r}_{u,i}$
$\hat{e}_1$
$\hat{e}_2$
<$\widehat{eos}$>
MLP
Linear Layer
$\mathbf{s}_{L,1}$ $\mathbf{s}_{L,2}$ $\mathbf{s}_{L,3}$ $\mathbf{s}_{L,4}$ $\mathbf{s}_{L,5}$ $\mathbf{s}_{L,6}$ $\mathbf{s}_{L,7}$
Transformer with L Layers
Transformer with L Layers
u i f_1 f_2 <bos> $\hat{e}_1$ $\hat{e}_2$
User
Item
Features (opt.)
Explanation

Experiment - Rating Prediction

- R : RMSE
- M : MAE

	Yelp		Amazon		TripAdvisor	
	R ↓	M ↓	R ↓	M ↓	R ↓	M ↓
PMF	1.097	0.883	1.235	0.913	0.870	0.704
SVD++	1.022	0.793	1.196	0.871	0.811	0.623
NARRE	1.028	0.791	1.176	0.865	0.796	0.612
NRT	1.016	0.796	1.188	0.853	0.797	0.611
CAML	1.026	0.798	1.191	0.878	0.818	0.622
PETER	1.017	0.793	1.181	0.863	0.814	0.635
ERRA	**1.008**	**0.781**	**1.158**	**0.832**	**0.787**	**0.603**

Experiment - Explanation

GRU + co-attention

GRU

GPT

Transformer

v.s. PETER

Datasets	Metrics	Baselines				Ours	
		NRT	CAML	ReXPlug	PETER	ERRA	Improvement
Amazon	BLEU1	13.37	11.19	10.8	13.78	**14.38**	4.17%
	BLEU4	1.44	1.12	1.29	1.68	**1.88**	10.6%
Yelp	BLEU1	10.5	9.91	8.59	10.29	**10.71**	3.92%
	BLEU4	0.67	0.56	0.57	0.69	**0.73**	5.43%
TripAdvisor	BLEU1	15.78	14.43	12.64	15.33	**16.13**	5.9%
	BLEU4	0.85	0.86	0.71	0.89	**1.06**	15.8%

Expe

Datasets	Metrics	Baselines				Ours	
		NRT	CAML	ReXPlug	PETER	ERRA	Improvement
Amazon	BLEU1	13.37	11.19	10.8	13.78	**14.38**	4.17%
	BLEU4	1.44	1.12	1.29	1.68	**1.88**	10.6%
	R2-P	2.06	1.48	2.17	2.21	**2.71**	14.8%
	R2-R	2.08	1.23	1.12	2.02	**2.93**	17.6%
	R2-F	1.97	1.24	1.22	1.97	**2.57**	21.2%
	RL-P	12.52	9.32	9.20	12.62	**16.13**	19.7%
	RL-R	12.20	10.11	10.58	12.06	**14.41**	16.3%
	RL-F	10.77	8.11	8.73	11.07	**13.87**	18.1%
	BERT-S	75.4	74.9	75.3	76.2	**79.8**	4.5%
Yelp	BLEU1	10.5	9.91	8.59	10.29	**10.71**	3.92%
	BLEU4	0.67	0.56	0.57	0.69	**0.73**	5.43%
	R2-P	1.95	1.78	1.49	1.91	**2.03**	5.91%
	R2-R	1.29	1.05	1.07	1.31	**1.36**	3.6%
	R2-F	1.35	1.25	1.11	1.43	**1.48**	2.36%
	RL-P	15.88	14.25	13.32	16.07	**16.60**	3.19%
	RL-R	10.72	14.26	9.56	10.14	**11.23**	9.7%
	RL-F	9.53	9.16	8.70	10.26	**10.82**	5.1%
	BERT-S	83.6	83.2	82.2	83.3	**85.2**	2.2%
TripAdvisor	BLEU1	15.78	14.43	12.64	15.33	**16.13**	5.9%
	BLEU4	0.85	0.86	0.71	0.89	**1.06**	15.8%
	R2-P	1.98	1.49	1.61	1.92	**2.09**	8.1%
	R2-R	1.92	1.91	1.49	2.01	**2.15**	9.7%
	R2-F	1.9	1.92	1.61	1.94	**2.05**	5.3%
	RL-P	14.85	13.36	11.38	13.54	**15.40**	8.6%
	RL-R	14.03	12.38	10.22	14.75	**15.02**	1.81%
	RL-F	12.25	12.39	9.97	12.61	**13.17**	4.50%
	BERT-S	82.7	84.8	83.2	86.4	**88.1**	1.96%

Experiment - Case Study

Case 1 - Truth	The environment of this hotel is comfortable and the transportation is very convenient and the sound insulation effect is great. Aspects:(environment, comfortable) (hotel, insulation)
NRT	The environment of this hotel is best!
PETER	The hotel service is pretty good! looks very nice!
ERRA	The room **environment** is pretty **comfortable**! The **traffic** here is very **convenient**.

Case 2 - Truth	The screen of this phone is too small and his battery drains fast so I can't stand it. Aspects:(screen, too small) (battery, fast)
NRT	The phone is bad.
PETER	The phone is bad, It works poorly and I don't like it.
ERRA	I really **hate** this phone, the **battery doesn't last long**, the **screen** is **faulty**.

Case 3 - Truth	Delicious! The customer service is pretty good and the open all the way to 3 am in the morning. The prime foods are excellent! Aspects:(service, good) (foods, excellent)
NRT	The service is pretty good.
PETER	he tastes delicious! The service is pretty good!
ERRA	**excellent**! The **service** here is **pretty good**. The food here is **very delicious**! There are many unique foods in it and open **till dawn**.

Experiment - Human Evaluation

- Randomly selected 1000 samples and invited 10 annotators to assign scores.

	Measures	NRT	CAML	ReXPlug	ERRA
Range: [1, 5]	Fluency	2.73	2.92	3.11	**3.45**
Range: [-1, 1]	Kappa	(0.67)	(0.63)	(0.74)	(**0.79**)

評分者的一致性

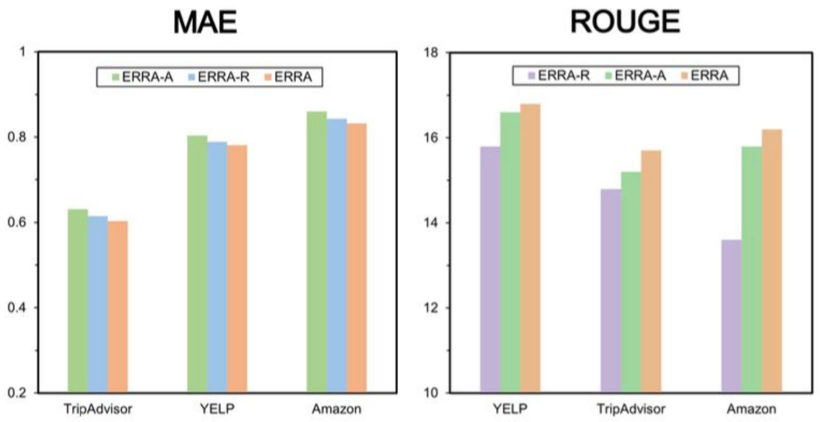
Kappa係數	一致性程度
< 0.4	差
0.4 ~ 0.6	一般
0.6 ~ 0.8	好
> 0.8	極佳

Ablation Analysis

- ERRA-A : removing the aspect enhancement module
- ERRA-R : removing the retrieval enhancement module

Prediction Loss ↓

Explanation Similarly ↑



(a) (b)

Outline

- Introduction
- Method
- Experiment
- Conclusion

Conclusion

- To address the issue of incorrect embedding induced by data sparsity, this paper incorporate personalized aspect information and rich review knowledge corpus into our model.