中国人民大學 RENMIN UNIVERSITY OF CHINA 高瓴人工智能学院 Gaoling School of Artificial Intelligence

Enhancing User Behavior Sequence Modeling by Generative Tasks for Session Search

Haonan Chen, Zhicheng Dou*, Yutao Zhu, Zhao Cao, Xiaohua Cheng, Ji-Rong Wen

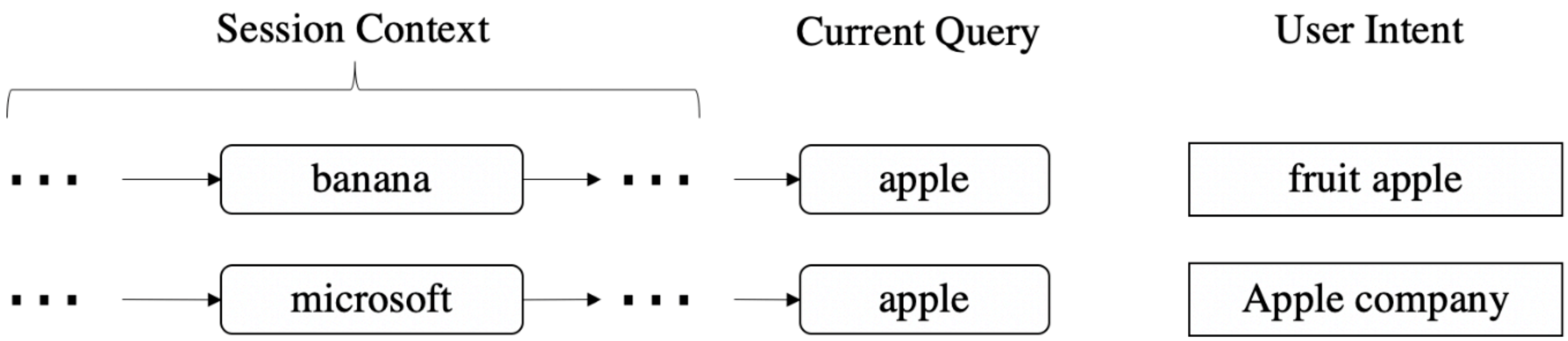
CIKM 2022

Haonan Chen hnchen@ruc.edu.cn

Gaoling School of Artificial Intelligence, Renmin University of China

*: Corresponding author.

Background

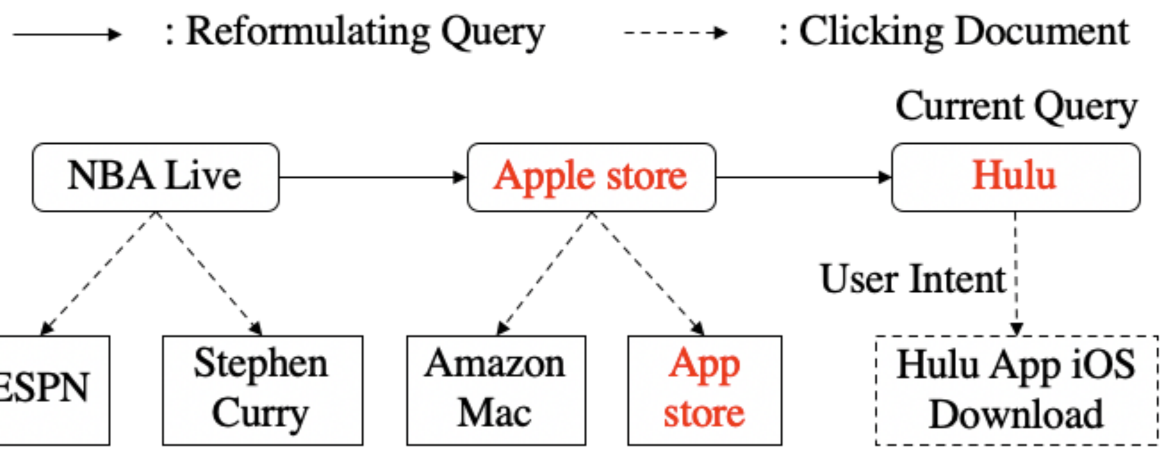


- To fulfill complex information needs, users will issue a sequence of queries, examine and interact with some of the results.
- Utilizing contextual information of the current search session is valuable for capturing a user's actual search intent.

Motivation

- Existing neural approaches in session search mainly used encoders (e.g., RNN, BERT) to encode user behaviors into a latent representation.

- However, the current session sequence may contain some useless information that could cause the encoders to misinterpret the real user intent.



Motivation

- A straightforward way to tackle this issue is to apply the auto-encoder technique, which enhances the encoder by making the decoder recover the whole session sequence.

- However, applying auto-encoder to session search is non-trivial in three aspects:
  - Generating the whole session sequence is challenging.
  - There is noise in the session sequence.
  - The user behaviors that represent the real search intent may be implicit in the current session sequence.

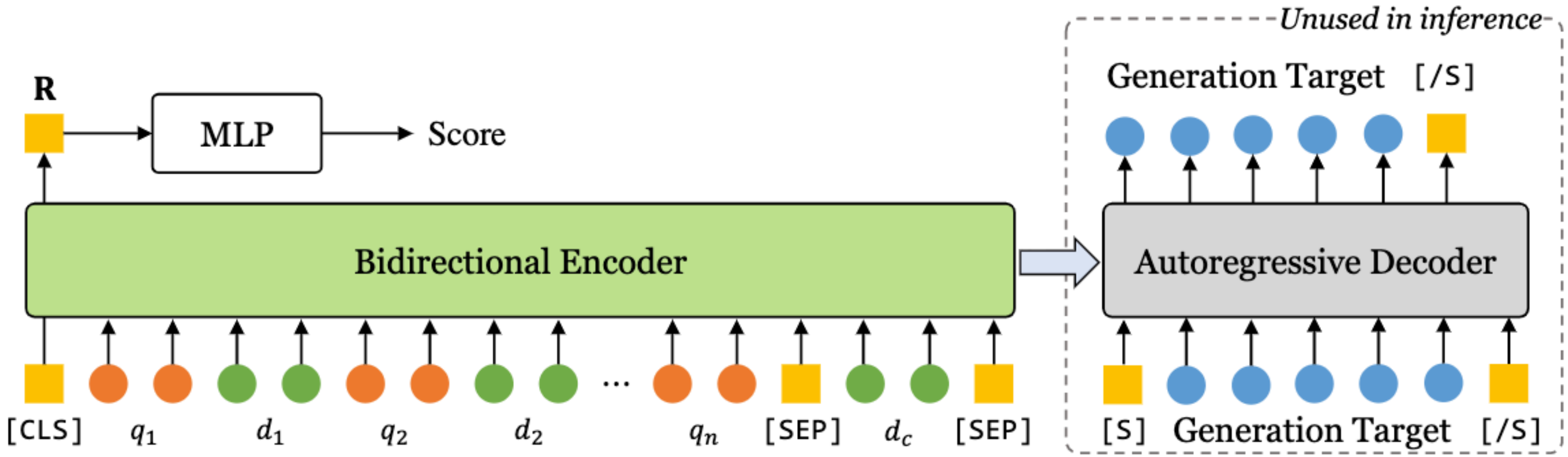
Idea

- An encoder-decoder structure + generative tasks (designed for SS):
  - Task 1: Predicting future queries.
  - Task 2: Predicting future clicked documents.
  - Task 3: Predicting a supplemental query.

- These tasks address the three challenges of auto-encoder as follows:
  - Task 1&2 attempt to generate future queries and documents separately.
  - We explore many potential generation targets, and propose these three tasks that can actually help the encoder.
  - All these tasks try to predict sequences that are not in the current session.

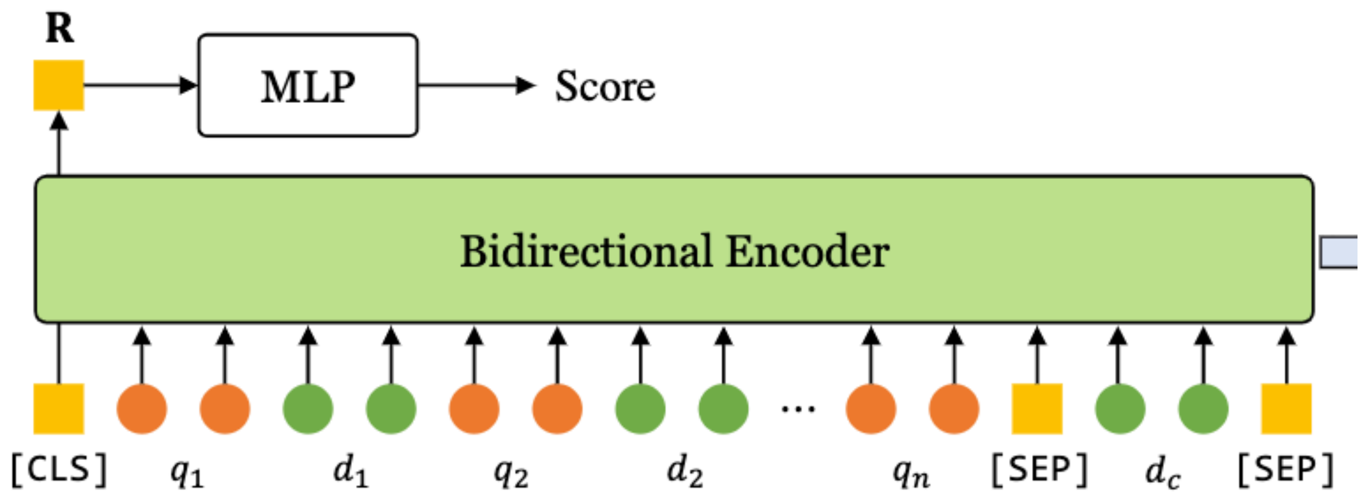
- ASE – Auto-Session-Encoder

Overview



- A standard encoder-decoder learning framework.
- Jointly learn ranking and generative tasks.
- Decoder not used in inference.

Modeling Session Sequence



- Constructing behavior sequence to input sequence.
  - $I = \{[\text{CLS}]q_1[\text{EOS}]d_1[\text{EOS}] \ldots q_n[\text{EOS}][\text{SEP}]d_c[\text{EOS}][\text{SEP}]\}$

- Ranking score: Applying an MLP to the output of "[CLS]" token.

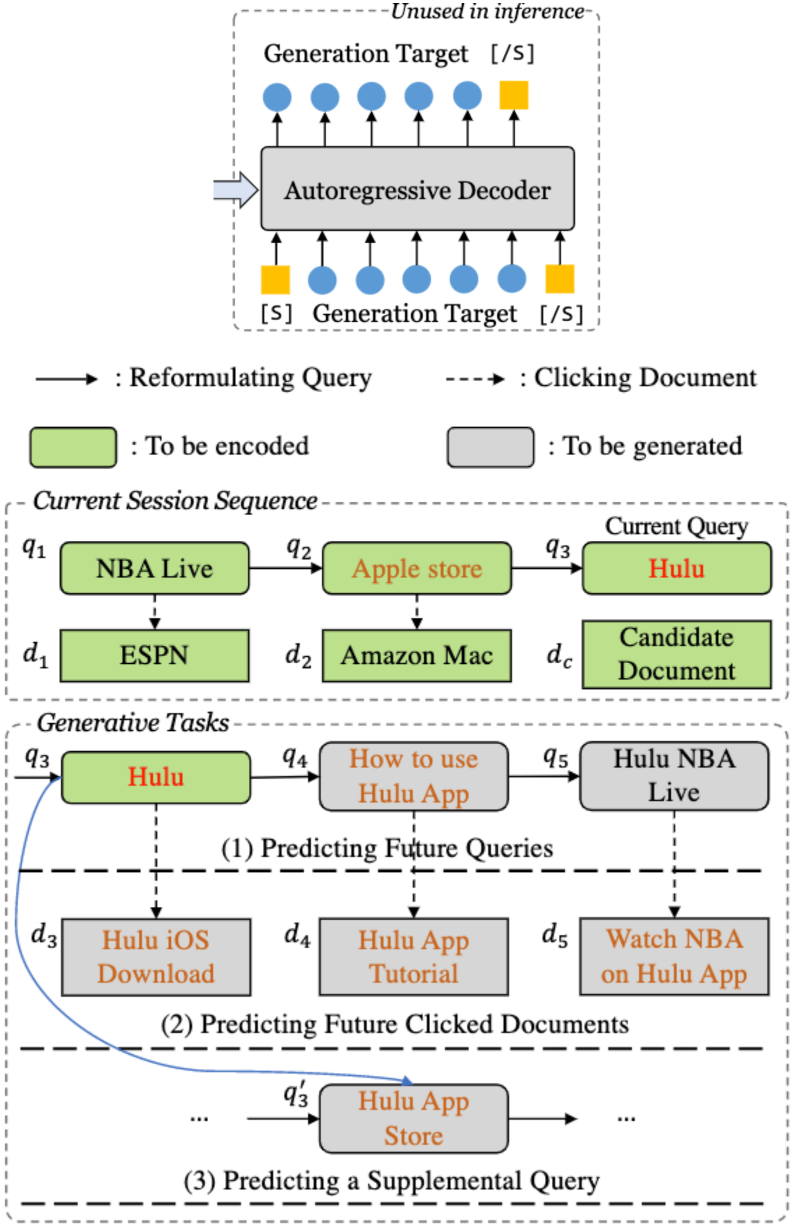
Generative Tasks

中国人民大學 RENMIN UNIVERSITY OF CHINA 高瓴人工智能学院 Gaoling School of Artificial Intelligence

- Predicting Future Queries
  - More accurate expression of search intent.
  - $GT_1 = q_{n+1}[\text{SEP}]q_{n+2}[\text{SEP}] \ldots q_{n+w}[\text{SEP}]$
  - w: prediction window size.

- Predicting Future Clicked Documents
  - More specific than keyword-based queries.
  - $GT_2 = d_n[\text{SEP}]d_{n+1}[\text{SEP}] \ldots d_{n+w-1}[\text{SEP}]$
  - Starting with current clicked document.

- Predicting a Supplemental Query
  - $\text{sup}(q_n', q_n) = \text{spe}(q_n', q_n) + \text{sim}(q_n', q_n)$
  - $GT_3 = q_n'[\text{SEP}]$



Optimization

- Pair-wise Ranking Loss:
  - $L_R(q_n) = \sum_{(d_c^+, d_c^-)\epsilon D_c} \max\left(0, \alpha - Score(d_c^+) + Score(d_c^-)\right)$
- Log-likelihood Generation Loss:
  - $L_g(GT) = -\sum_{j=1}^{|GT|} \log\left(Pr\left(w_j | w_{1:j-1}, S, d_c\right)\right)$
- Multi-task Learning with Uncertainty:
  - $L = \frac{L_R}{2\tau_r^2} + \log(\tau_r^2 + 1) + \sum_{g\epsilon G}\left(\frac{L_g}{2\tau_g^2} + \log\left(\tau_g^2 + 1\right)\right)$
  - τ: Uncertainty parameter, balancing losses of tasks.

Experimental Setup

- Dataset: AOL, Tiangong-ST.
- Baselines:
  - Ad-hoc ranking models: BM25, ARC-I, ARC-II, KNRM, Duet.
  - Context-aware ranking models: CARS, HBA-Transformers, HQCN, RICR, BERT, BART, COCA.
- Codes: https://github.com/haon-chen/ASE-Official

Experimental Setup

- Backbone: BART
  - BERT-like Encoder + GPT-like Decoder (only architectures).
  - Good for generation task, on par for discriminative task.
  - Fair comparisons with BERT-based baseline models.
    - Comparable parameters as BERT (6 layers encoder + 6 layers decoder v.s. 12 layers encoder).
    - Pre-trained on same data and same training settings as BERT.
- Can be applied to other Transformers-based encoder-decoder PLMs.

Results

中国人民大學 RENMIN UNIVERSITY OF CHINA 高瓴人工智能学院 Gaoling School of Artificial Intelligence

Table 2: Overall results on AOL and Tiangong-ST. "†" and "‡" denote the result is significantly worse than our ASE in t-test with p-value < 0.01 and p-value < 0.05 respectively. The best performance is in bold.

	AOL					
Model	NDCG@1	@3	@5	@10	MAP	MRR
Ad-hoc Ranking Models						
BM25	0.1195†	0.1862†	0.2136†	0.2481†	0.2200†	0.2271†
ARC-I	0.1988†	0.3108†	0.3489†	0.3953†	0.3361†	0.3475†
ARC-II	0.2428†	0.3564†	0.4026†	0.4486†	0.3834†	0.3951†
KNRM	0.2397†	0.3868†	0.4322†	0.4761†	0.4038†	0.4133†
Duet	0.2492†	0.3822†	0.4246†	0.4675†	0.4008†	0.4111†
Context-aware Ranking Models						
CARS	0.2816†	0.4117†	0.4542†	0.4971†	0.4297†	0.4408†
HBA	0.3773†	0.5241†	0.5624†	0.5951†	0.5281†	0.5384†
RICR	0.3894†	0.5267†	0.5648†	0.5971†	0.5338†	0.5450†
HQCN	0.3990†	0.5441†	0.5783†	0.6070†	0.5448†	0.5549†
BART	0.3908†	0.5414†	0.5797†	0.6108†	0.5450†	0.5551†
BERT	0.3990†	0.5440†	0.5818†	0.6123†	0.5471†	0.5572†
COCA	0.4024†	0.5478†	0.5849†	0.6160†	0.5500†	0.5601†
ASE	**0.4144**	**0.5682**	**0.6007**	**0.6283**	**0.5650**	**0.5752**

	Tiangong-ST					
Model	NDCG@1	@3	@5	@10	MAP	MRR
Ad-hoc Ranking Models						
BM25	0.6029†	0.6646†	0.7072†	0.8541†	0.7837†	0.8225†
ARC-I	0.7088†	0.7087†	0.7317†	0.8691†	0.8580‡	0.9159†
ARC-II	0.7131†	0.7237†	0.7379†	0.8732†	0.8611‡	0.9227†
KNRM	0.7198†	0.7421†	0.7660†	0.8857‡	0.8683	0.9130†
Duet	0.7577‡	0.7354†	0.7548†	0.8829‡	0.8663	0.9273‡
Context-aware Ranking Models						
CARS	0.7385†	0.7386†	0.7512†	0.8837‡	0.8556‡	0.9268‡
HBA	0.7612‡	0.7518†	0.7639†	0.8896‡	0.8615	0.9316‡
RICR	0.7670‡	0.7636‡	0.7740‡	0.8934‡	0.8147†	0.8937†
HQCN	0.7739‡	0.7682	0.7783	0.8976	0.8659	0.9328‡
BART	0.7380†	0.7464†	0.7574†	0.8853‡	0.8585‡	0.9294‡
BERT	0.7488†	0.7541‡	0.7651†	0.8890‡	0.8653	0.9316‡
COCA	0.7769	0.7576‡	0.7703‡	0.8932‡	0.8623	0.9382
ASE	**0.7884**	**0.7727**	**0.7839**	**0.8996**	**0.8701**	**0.9482**

Ablation Studies

中国人民大學 RENMIN UNIVERSITY OF CHINA 高瓴人工智能学院 Gaoling School of Artificial Intelligence

Predicting Future Queries (PFQ).
Predicting future Clicked Documents (PCD).
Predicting a Supplemental Query (PSQ).

Metric	w/o. PFQ		w/o. PCD		w/o. PSQ		ASE
NDCG@1	0.4100	-1.06%	0.4036	-2.61%	0.4102	-1.01%	**0.4144**
NDCG@3	0.5580	-1.80%	0.5570	-1.97%	0.5636	-0.81%	**0.5682**
NDCG@5	0.5933	-1.23%	0.5895	-1.86%	0.5957	-0.83%	**0.6007**
NDCG@10	0.6205	-1.24%	0.6180	-1.64%	0.6246	-0.59%	**0.6283**
MAP	0.5579	-1.26%	0.5546	-1.84%	0.5608	-0.74%	**0.5650**
MRR	0.5691	-1.06%	0.5650	-1.77%	0.5707	-0.78%	**0.5752**

Performances of Various Generative Targets

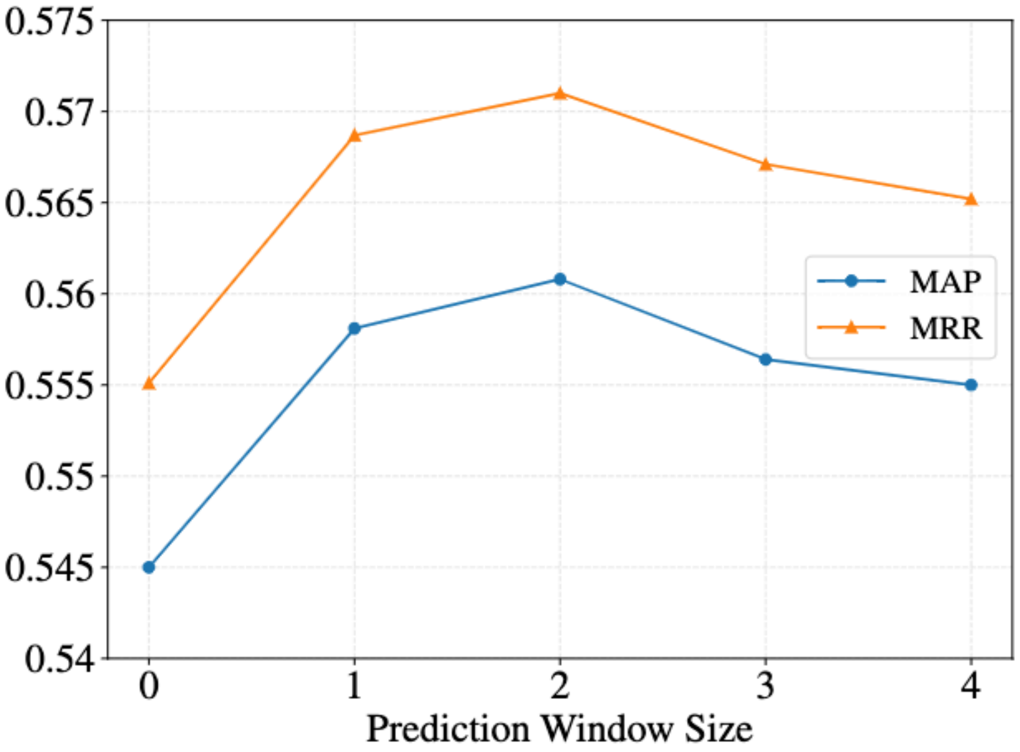
Current Session Sequence:

$S = \{q_1, d_1, q_2, d_2, \dots, q_n\}$

GT	NDCG@1	NDCG@10	MAP
- (BART)	0.3882	0.6124	0.5450
q_{n-1}	0.3849 -0.85%	0.6103 -0.34%	0.5427 -0.42%
q_n	0.3928 +1.84%	0.6077 -0.77%	0.5442 -0.15%
q_{n+1}	0.4004 +3.14%	0.6150 +0.42%	0.5516 +1.21%
d_{n-1}	0.3922 +1.03%	0.6104 -0.33%	0.5464 +0.26%
d_n	0.4022 +3.61%	0.6212 +1.44%	0.5548 +1.80%
d_{n+1}	0.4044 +4.17%	0.6206 +1.34%	0.5565 +2.11%
q'_n	0.3990 +2.78%	0.6151 +0.44%	0.5509 +1.08%

Effect of Prediction Window Size

Performances of the variant of ASE (BART with GT_1 and GT_2) with different values of w on AOL:



中国人民大学 RENMIN UNIVERSITY OF CHINA 高瓴人工智能学院 Gaoling School of Artificial Intelligence

Applicability

Application of the proposed generative tasks (*GT*s) to Other Transformer-based Encoder-Decoder Models:

Model	MAP	MRR	NDCG@3	NDCG@10
T5-small	0.5142†	0.5257†	0.5102†	0.5803†
T5-small + *GTs*	0.5246	0.5363	0.5232	0.5911
Improv.	+2.02%	+2.02%	+2.55%	+1.86%
BlenderBot-small	0.5465†	0.5570†	0.5470†	0.6108†
BlenderBot-small + *GTs*	0.5580	0.5685	0.5601	0.6220
Improv.	+2.10%	+2.06%	+2.39%	+1.83%

Thank You for Listening!
Q & A