SIGIR 2019
21-25 July

The 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval
PARIS, FRANCE

Context Attentive Document Ranking and Query Suggestion

Wasi Uddin Ahmad
University of California, Los Angeles

Kai-Wei Chang
University of California, Los Angeles

Hongning Wang
University of Virginia

https://github.com/wasiahmad/context_attentive_ir

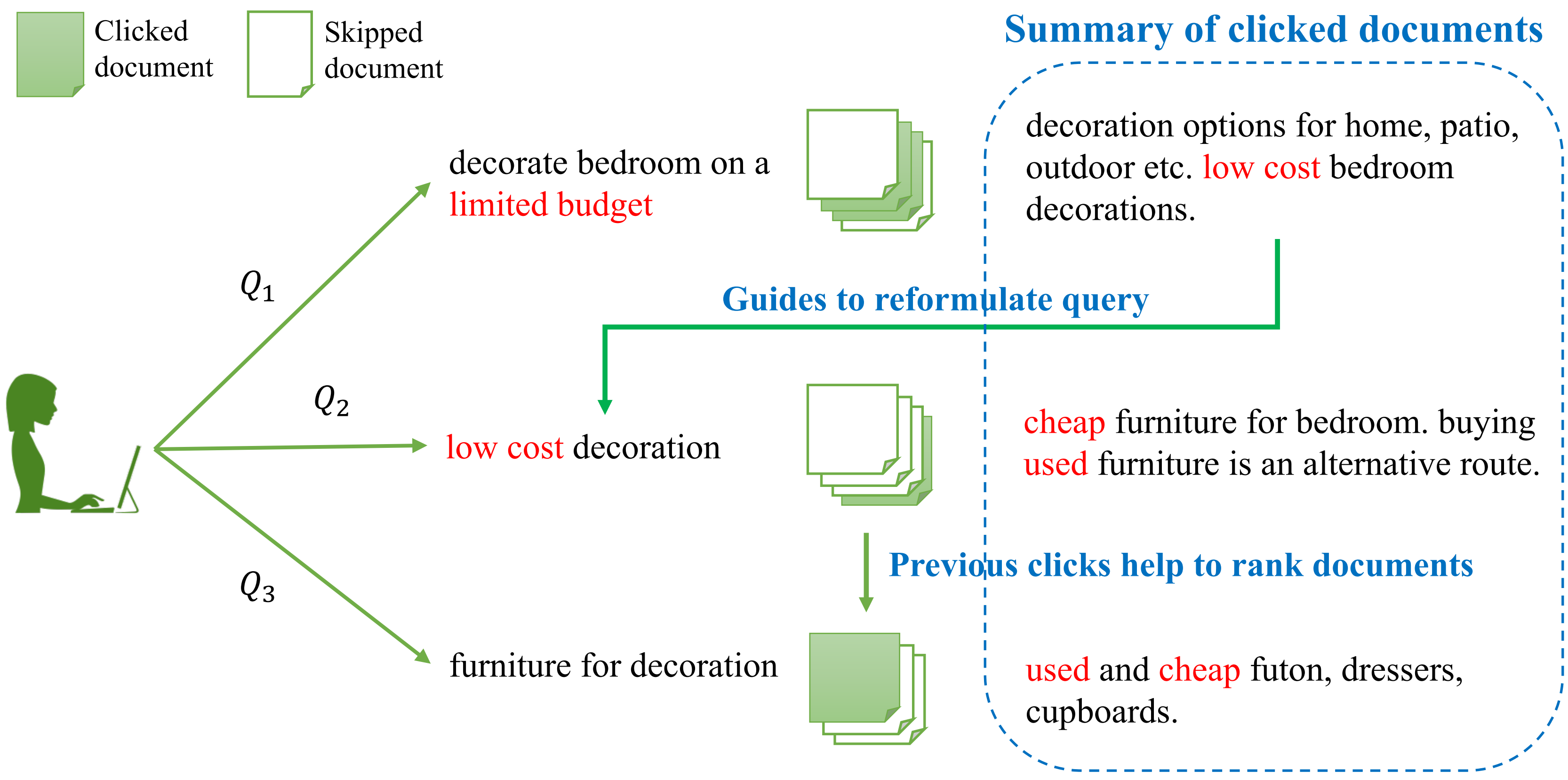
Codes will be released soon!

Search Logs Provide Rich Context to Understand Users' Search Tasks

5/29/2012	
5/29/2012 14:06:04	coney island cincinnati
5/29/2012 14:11:49	sas
5/29/2012 14:12:01	sas shoes
5/30/2012	
5/30/2012 12:12:04	exit #72 and 275 lodging
5/30/2012 12:25:19	6pm.com
5/30/2012 12:49:21	coupon for 6pm
5/31/2012	
5/31/2012 19:40:38	motel 6 locations
5/31/2012 19:45:04	hotels near coney island

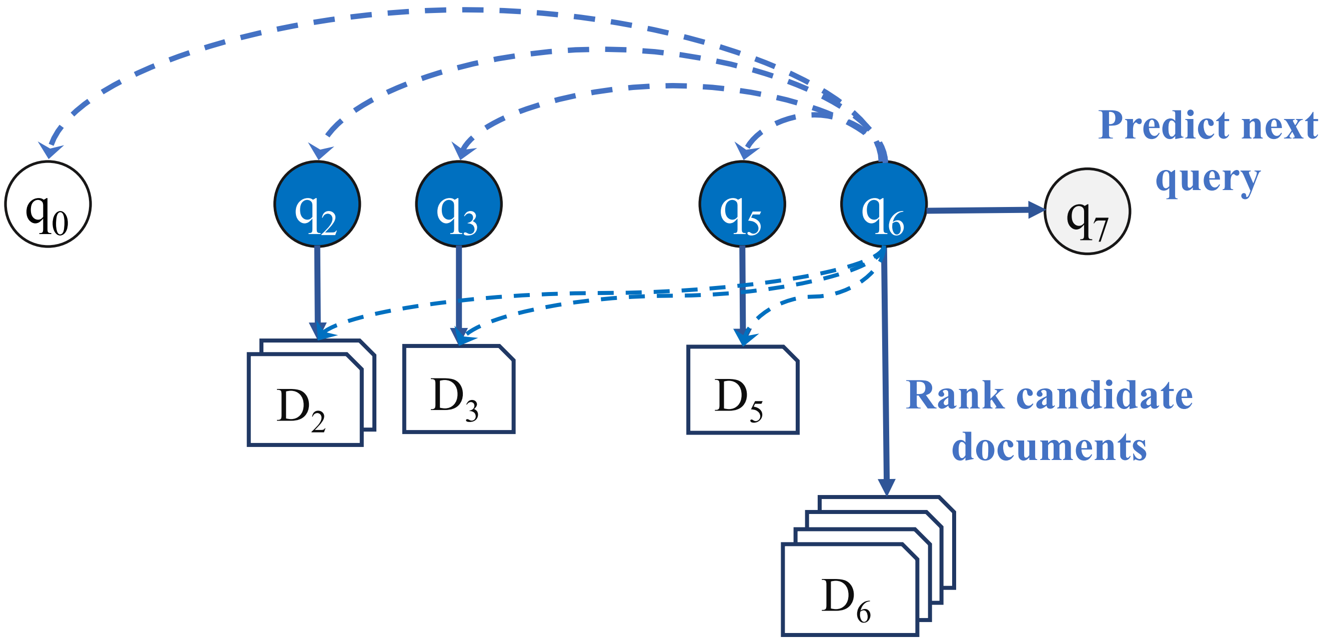
***Task**: an atomic information need that may result in one or more queries.*

Clicks Further Enrich Context to Understand Users' Search Tasks



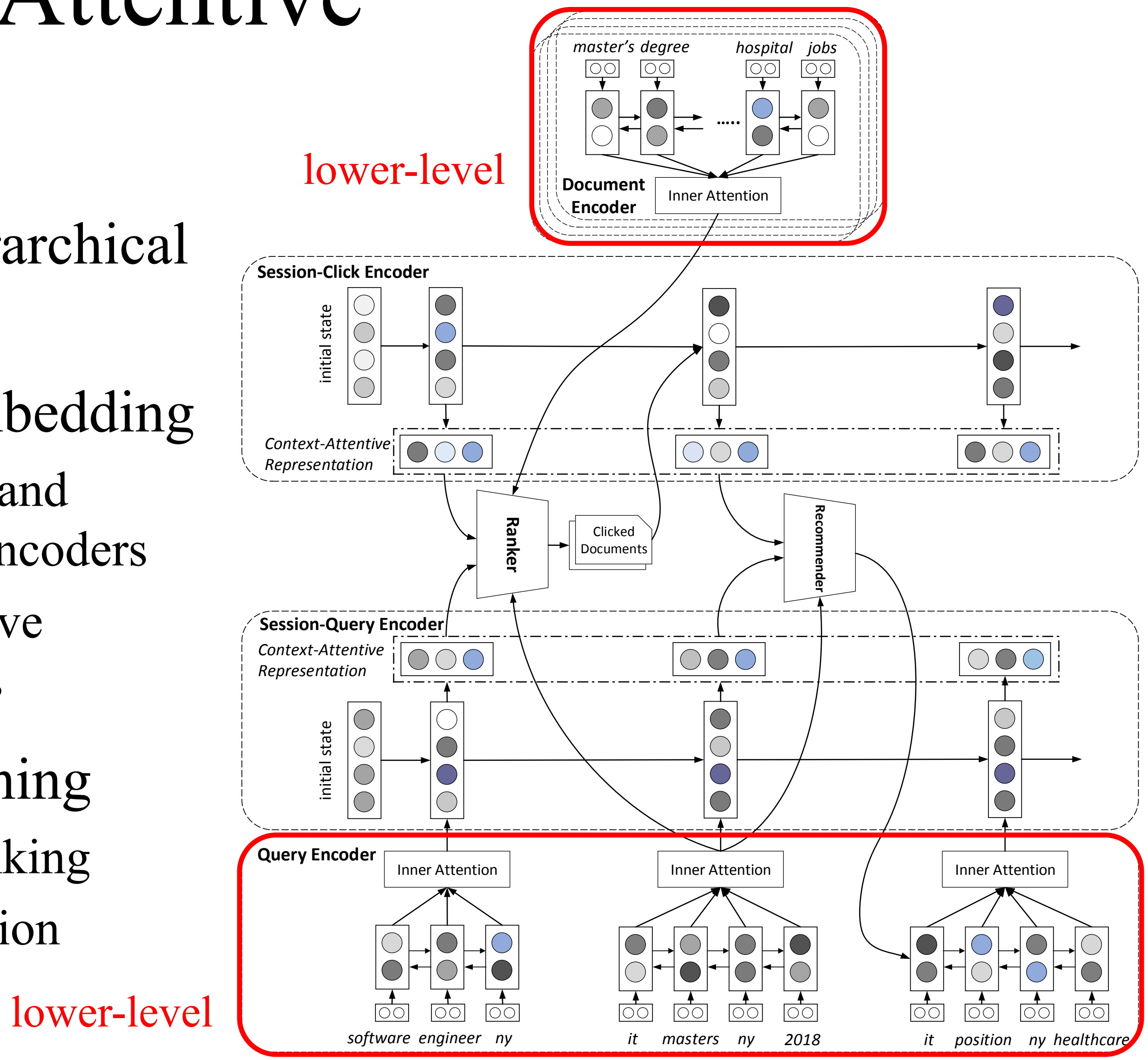
Our Proposal

- Attend on previous queries/clicks to perform retrieval tasks
- Learning to utilize context in multiple retrieval activities



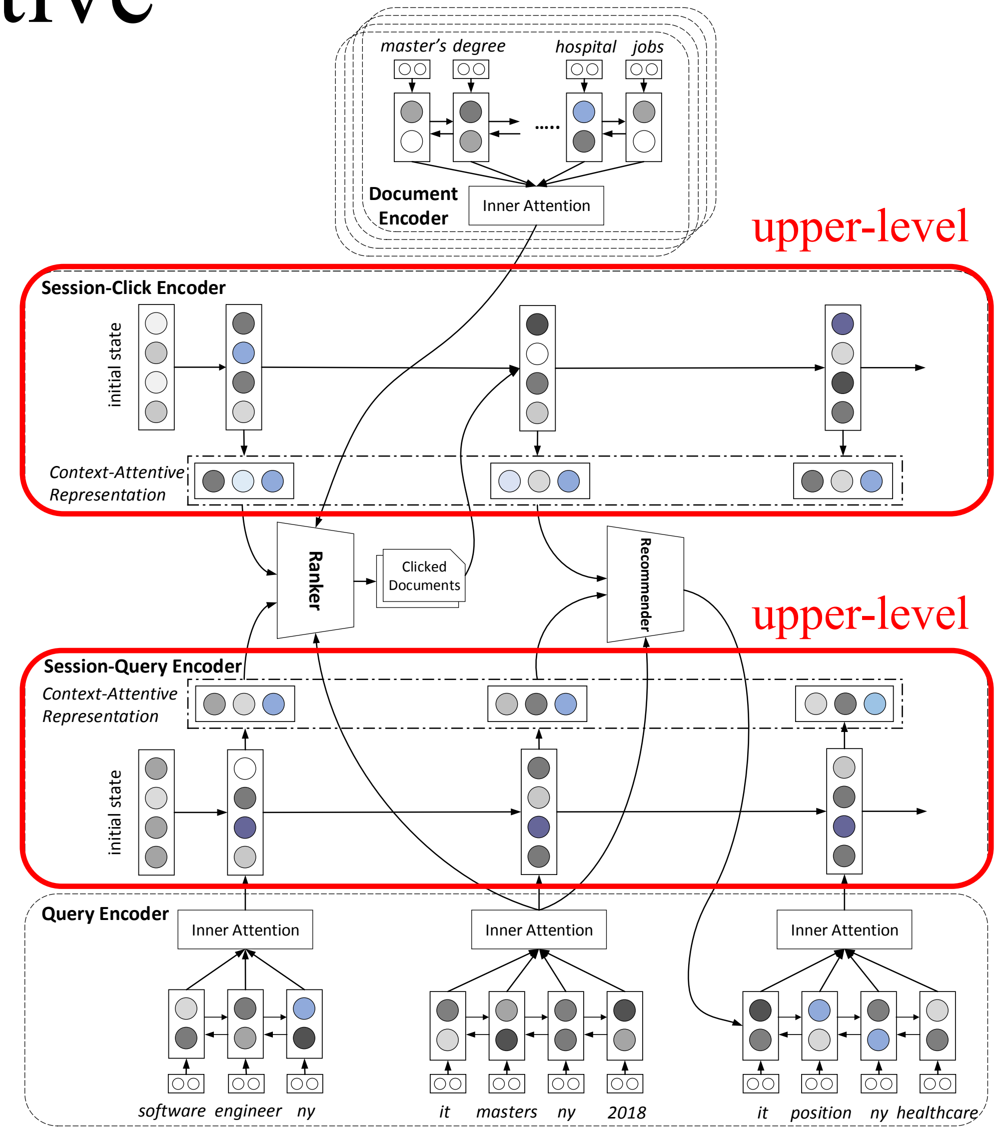
A Context Attentive Solution

- A two-level hierarchical structure
- Task context embedding
    - Session-query and session-click encoders
    - Context attentive representations
- Multi-task Learning
    - Document Ranking
    - Query Suggestion



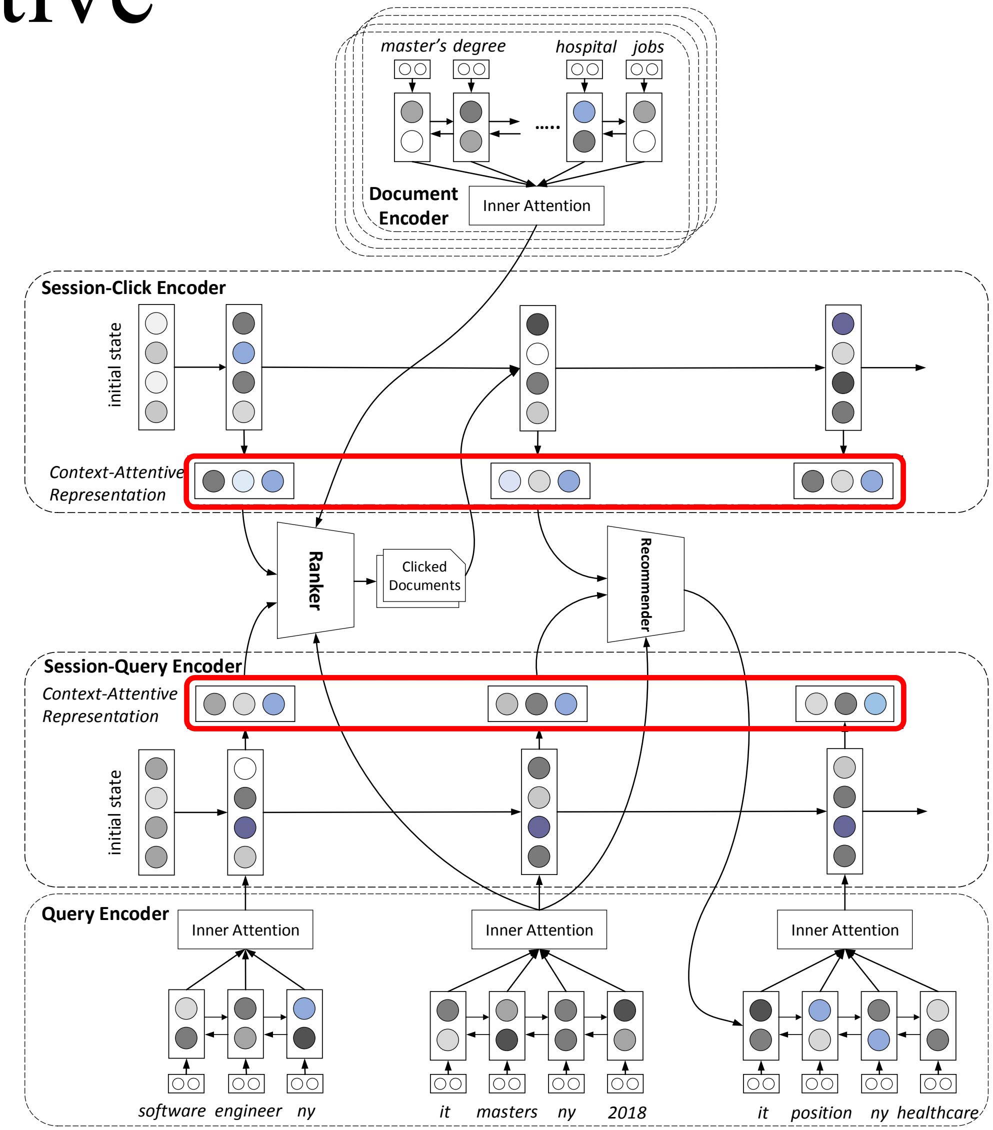
A Context Attentive Solution

- A two-level hierarchical structure
- Task context embedding
  - Session-query and session-click encoders
  - Context attentive representations
- Multi-task Learning
  - Document Ranking
  - Query Suggestion



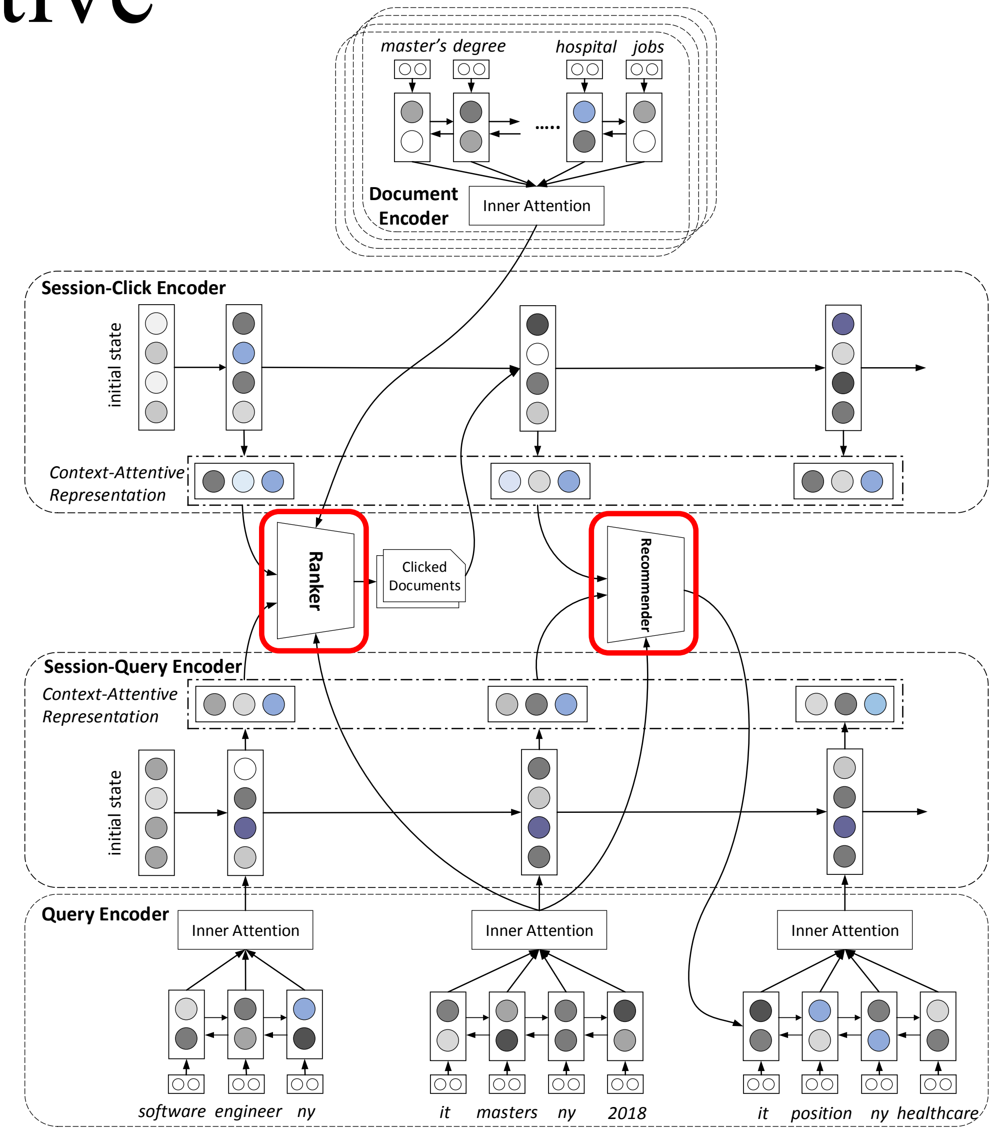
A Context Attentive Solution

- A two-level hierarchical structure
- Task context embedding
  - Session-query and session-click encoders
  - Context attentive representations
- Multi-task Learning
  - Document Ranking
  - Query Suggestion



A Context Attentive Solution

- A two-level hierarchical structure
- Task context embedding
  - Session-query and session-click encoders
  - Context attentive representations
- Multi-task Learning
  - Document Ranking
  - Query Suggestion



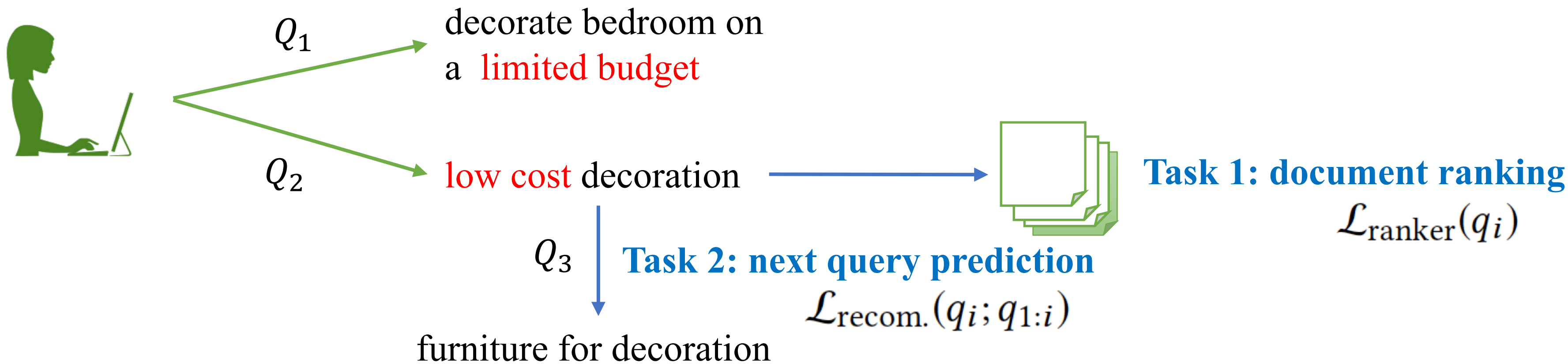
Multi-task Learning Objective

- Optimized via Regularized Multi-task Learning

Regularization

Negative log-likelihood

$$\boxed{\mathcal{L}_{R1}} + \boxed{\mathcal{L}_{R2}} + \frac{1}{N}\sum_{k=1}^{N}\sum_{q_i \in \mathcal{T}_k}\left(\boxed{\mathcal{L}_{\text{ranker}}(q_i)} + \boxed{\mathcal{L}_{\text{recom.}}(q_i; q_{1:i})}\right)$$



Experiments

Data Source

- AOL search log – 13 weeks search log of ~650k users
  - Background set – 5 weeks
  - Training set – 6 weeks
  - Validation and Test set – 2 weeks
- Aggregation of candidate documents for ranking
  - Candidates are sampled from the top 1000 documents retrieved by BM25 from a pool of 1M documents
- Task Segmentation
  - Averaging the query term vectors → Query embedding
  - Consecutive queries* with cosine similarity > 0.5

* within 30 mins interval

Data Statistics

- Only the document titles are utilized in the experiments

Dataset Split	Train	Validation	Test
# Task	219,748	34,090	29,369
# Query	566,967	88,021	76,159
Average Task Length	2.58	2.58	2.59
Average Query Length	2.86	2.85	2.90
Average Document Length	7.27	7.29	7.08
Average # Click per Query	1.08	1.08	1.11

Evaluation Metrics and Baselines

- Document Ranking
  - Metrics – MAP, MRR, NDCG@k (where k=1,3,10)
  - Baselines – BM25, QL, FixInt, DSSM, DUET, MNSRF etc.

- Query Suggestion
  - Metrics – MRR, F1, BLEU-k (where k=1,2,3,4)
  - Baselines – HRED-qs, Seq2seq+Attn, MNSRF etc.

MRR – assesses discrimination ability, rank a list of candidate queries that might follow a given query

Evaluation Metrics and Baselines

- Document Ranking
  - Metrics – MAP, MRR, NDCG@k (where k=1,3,10)
  - Baselines – BM25, QL, FixInt, DSSM, DUET, MNSRF etc.

- Query Suggestion
  - Metrics – MRR, F1, BLEU-k (where k=1,2,3,4)
  - Baselines – HRED-qs, Seq2seq+Attn, MNSRF etc.

F1, BLEU-k – assesses generation ability, measures overlapping between the generated query term sequence and ground-truth sequence.

Evaluation on Document Ranking

- *Vocabulary gap limits the performance*
- *Do not utilize any search context information*

Model	MAP	MRR	NDCG		
			@1	@3	@10
Traditional IR-models					
BM25	0.230	0.206	0.206	0.269	0.319
QL	0.195	0.166	0.166	0.213	0.276
Single-task Learning					
CDSSM	0.313	0.341	0.205	0.252	0.373
DUET	0.479	0.490	0.332	0.462	0.546
Match Tensor	0.481	0.501	0.345	0.472	0.555
Multi-task Learning					
M-NSRF	0.491	0.502	0.348	0.474	0.557
M-Match Tensor	0.505	0.518	0.368	0.491	0.567
CARS	**0.531**	**0.542**	**0.391**	**0.517**	**0.596**

Evaluation on Document Ranking

– *Do not utilize any search context information*

Model	MAP	MRR	NDCG		
			@1	@3	@10
Traditional IR-models					
BM25	0.230	0.206	0.206	0.269	0.319
QL	0.195	0.166	0.166	0.213	0.276
Single-task Learning					
CDSSM	0.313	0.341	0.205	0.252	0.373
DUET	0.479	0.490	0.332	0.462	0.546
Match Tensor	0.481	0.501	0.345	0.472	0.555
Multi-task Learning					
M-NSRF	0.491	0.502	0.348	0.474	0.557
M-Match Tensor	0.505	0.518	0.368	0.491	0.567
CARS	**0.531**	**0.542**	**0.391**	**0.517**	**0.596**

Evaluation on Document Ranking

+ *Jointly learns ranking and suggestion*
– *Utilizes query history but no click history*

Model	MAP	MRR	NDCG		
			@1	@3	@10
Traditional IR-models					
BM25	0.230	0.206	0.206	0.269	0.319
QL	0.195	0.166	0.166	0.213	0.276
Single-task Learning					
CDSSM	0.313	0.341	0.205	0.252	0.373
DUET	0.479	0.490	0.332	0.462	0.546
Match Tensor	0.481	0.501	0.345	0.472	0.555
Multi-task Learning					
M-NSRF	0.491	0.502	0.348	0.474	0.557
M-Match Tensor	0.505	0.518	0.368	0.491	0.567
CARS	**0.531**	**0.542**	**0.391**	**0.517**	**0.596**

Evaluation on Document Ranking

+ *Jointly learns ranking and suggestion*
+ *Utilizes both query and click history*

Model	MAP	MRR	NDCG		
			@1	@3	@10
Traditional IR-models					
BM25	0.230	0.206	0.206	0.269	0.319
QL	0.195	0.166	0.166	0.213	0.276
Single-task Learning					
CDSSM	0.313	0.341	0.205	0.252	0.373
DUET	0.479	0.490	0.332	0.462	0.546
Match Tensor	0.481	0.501	0.345	0.472	0.555
Multi-task Learning					
M-NSRF	0.491	0.502	0.348	0.474	0.557
M-Match Tensor	0.505	0.518	0.368	0.491	0.567
CARS	**0.531**	**0.542**	**0.391**	**0.517**	**0.596**

Evaluation on Query Suggestion

Model	MRR	F1	BLEU			
			1	2	3	4
Single-task Learning						
Seq2seq	0.422	0.077	8.5	0.0	0.0	0.0
Seq2seq + Attn.	0.596	0.555	52.5	30.7	18.8	11.4
HRED-qs	0.576	0.522	48.8	26.3	15.3	9.2
Multi-task Learning						
M-Match Tensor	0.551	0.458	41.5	20.6	11.5	7.0
M-NSRF	0.582	0.522	49.7	26.7	16.0	9.9
CARS	**0.614**	**0.589**	**55.6**	**36.2**	**25.6**	**19.1**

Do not utilize any context

Evaluation on Query Suggestion

Model	MRR	F1	BLEU			
			1	2	3	4
Single-task Learning						
Seq2seq	0.422	0.077	8.5	0.0	0.0	0.0
Seq2seq + Attn.	0.596	0.555	52.5	30.7	18.8	11.4
HRED-qs	0.576	0.522	48.8	26.3	15.3	9.2
Multi-task Learning						
M-Match Tensor	0.551	0.458	41.5	20.6	11.5	7.0
M-NSRF	0.582	0.522	49.7	26.7	16.0	9.9
CARS	**0.614**	**0.589**	**55.6**	**36.2**	**25.6**	**19.1**

Do not utilize click history

Evaluation on Query Suggestion

Model	MRR	F1	BLEU			
			1	2	3	4
Single-task Learning						
Seq2seq	0.422	0.077	8.5	0.0	0.0	0.0
Seq2seq + Attn.	0.596	0.555	52.5	30.7	18.8	11.4
HRED-qs	0.576	0.522	48.8	26.3	15.3	9.2
Multi-task Learning						
M-Match Tensor	0.551	0.458	41.5	20.6	11.5	7.0
M-NSRF	0.582	0.522	49.7	26.7	16.0	9.9
CARS	**0.614**	**0.589**	**55.6**	**36.2**	**25.6**	**19.1**

+ *Utilizes both in-task query and click history*

Ablation Study

- Modeling search context is beneficial
- Joint learning of related retrieval tasks results in improvements

CARS Variant	NDCG			BLEU	
	@1	@3	@10	1	2
CARS	0.391	0.517	0.596	55.6	36.2
CARS w/o Attn.	0.387*	0.515*	0.594*	48.6*	26.1*
Ablation on search context					
w/o Session Query	0.379	0.505	0.586	33.7*	14.2*
w/o Session Click	0.356	0.485	0.568	48.2*	25.6*
Ablation on joint learning					
w/o Recommender	0.379	0.505	0.585	-	-
w/o Ranker	-	-	-	55.9	36.9

∗ indicates that the attention in the query recommender

Ablation Study

- Modeling search context is beneficial
- Joint learning of related retrieval tasks results in improvements

Performance drops

CARS Variant	NDCG @1	NDCG @3	NDCG @10	BLEU 1	BLEU 2
CARS	0.391	0.517	0.596	55.6	36.2
CARS w/o Attn.	0.387*	0.515*	0.594*	48.6*	26.1*
Ablation on search context					
w/o Session Query	0.379	0.505	0.586	33.7*	14.2*
w/o Session Click	0.356	0.485	0.568	48.2*	25.6*
Ablation on joint learning					
w/o Recommender	0.379	0.505	0.585	-	-
w/o Ranker	-	-	-	55.9	36.9

∗ indicates that the attention in the query recommender

Ablation Study

- Modeling search context is beneficial
- Joint learning of related retrieval tasks results in improvements

CARS Variant	NDCG			BLEU	
	@1	@3	@10	1	2
CARS	0.391	0.517	0.596	55.6	36.2
CARS w/o Attn.	0.387*	0.515*	0.594*	48.6*	26.1*
Ablation on search context					
w/o Session Query	0.379	0.505	0.586	33.7*	14.2*
w/o Session Click	0.356	0.485	0.568	48.2*	25.6*
Ablation on joint learning					
w/o Recommender	0.379	0.505	0.585	-	-
w/o Ranker	-	-	-	55.9	36.9

Performance drops

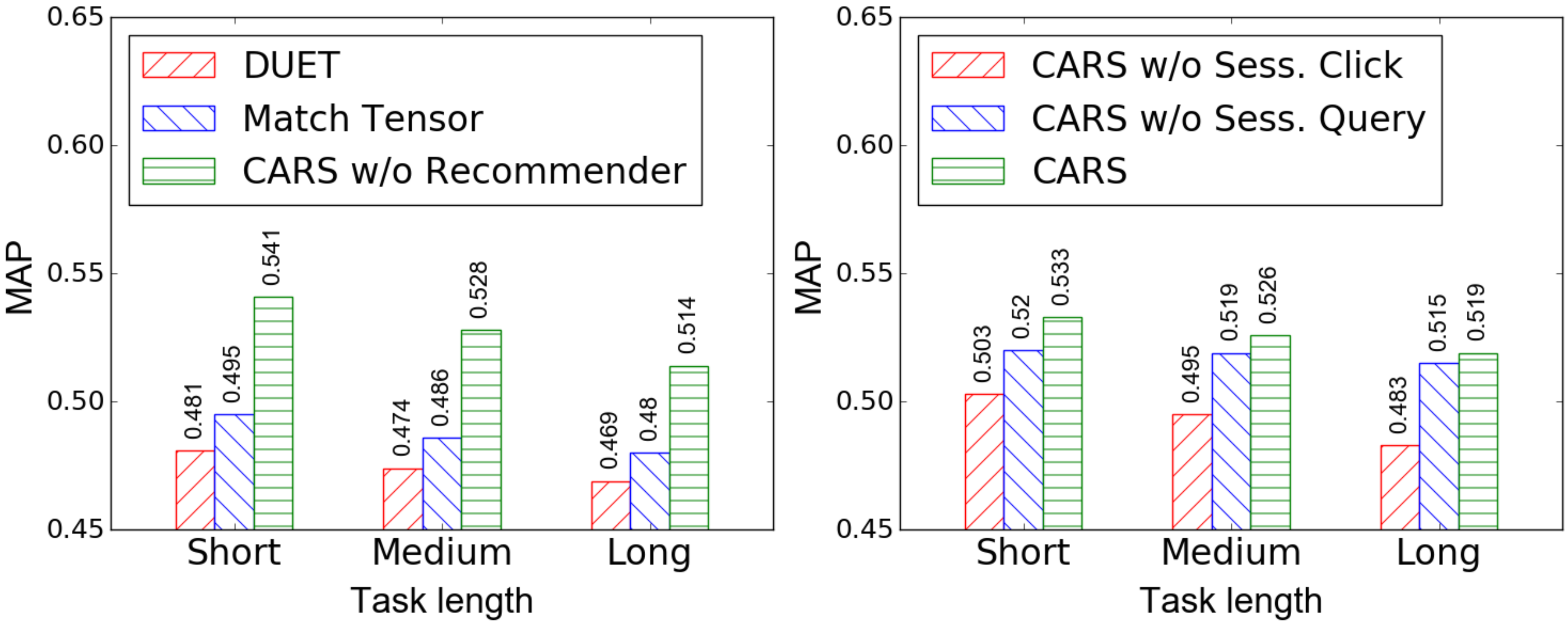
Performance increases

∗ indicates that the attention in the query recommender

Short tasks (with 2 queries)
Medium tasks (with 3–4 queries)
Long tasks (with 5+ queries)

Effect of Context Modeling

Hypothesis – longer tasks are intrinsically more difficult.

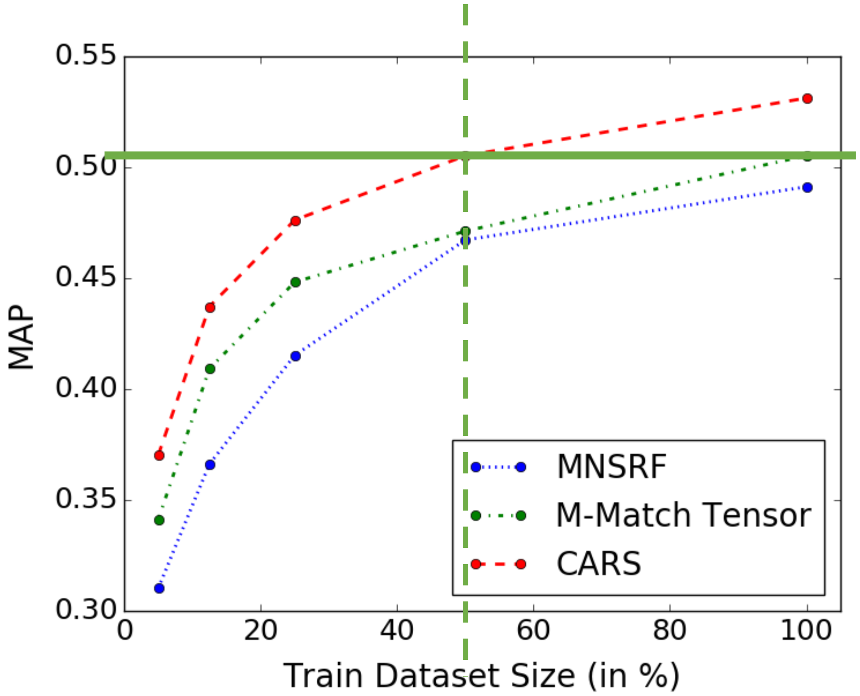


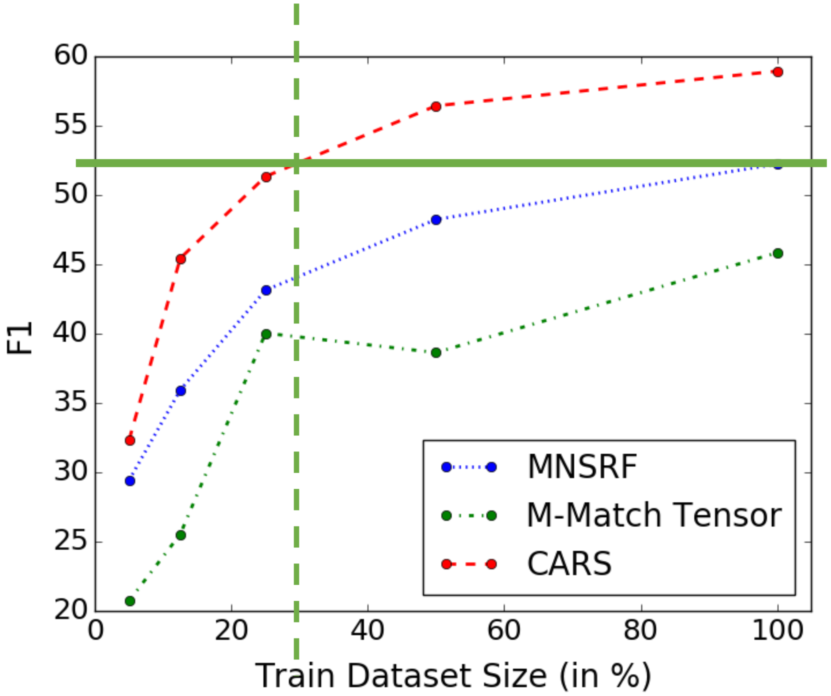
Effect of Context Modeling

- Context information helps more on short/medium tasks
- Longer tasks are intrinsically more difficult.

Sample Complexity

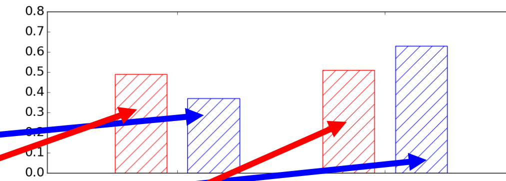
In terms of #parameters, CARS > MNSRF > M-Match Tensor



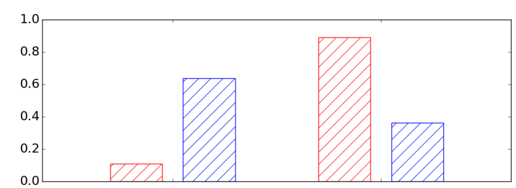


Case Analysis

How in-task previous queries and clicks impact predicting the next query and click for it?



Q1: **abortion** at **home**	Q2: **consequences** of **abortion** in future pregnancy	Q3: **abortion** laws in los **angeles**
D1. **private clinic** caring doctors and staff free consultations call us now D2. **not** every **pregnancy** is planned and sometimes a pregnancy is unwanted ✓D3. **natural home** remedies for abortion in early pregnancy ✓D4. **carrying** out an **abortion** at home is even more complex D5. have you come across unwanted **pregnancy** do you want natural **miscarriage**	D1. the effect of **pregnancy termination** on future reproduction ncbi ✓D2. **how** an **abortion** affects your chance of getting pregnant again D3. does **having** an abortion affect your future **fertility** D4. **abortion** does it affect **subsequent** pregnancies mayo clinic ✓D5. **medical abortion** won't affect future pregnancies abc news	



Conclusion & Future Works

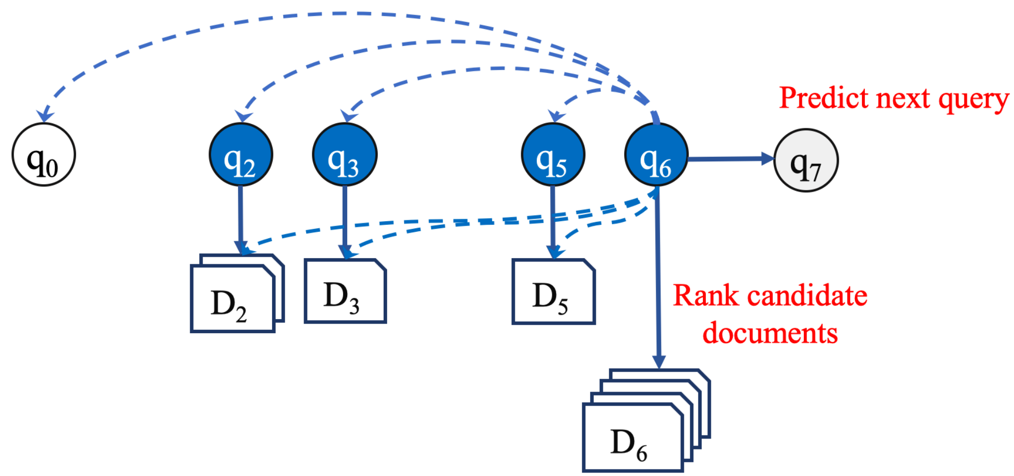
- A task-based approach of learning search context
  - Exploiting users' on-task search query and click sequence
  - Jointly optimized on two companion retrieval tasks

- Future works
  - Modeling across-task relatedness, e.g., users' long-term search interest
  - Apply to any scenario where a user sequentially interacts with a system

https://github.com/wasiahmad/context_attentive_ir

Codes will be released soon!

***In-task Context*: a richer way to understand users' search intent**



Thank You!

Q&A