TREC CAsT 2021 - CFDA & CLIP Lab

The Multi-stage Pipeline for Conversational Search – Paraphrase Query Expansion and Multi-view Point-wise Ranking

Jia-Huei Ju†, Chih-Ting Yeh†, Cheng-Wei Lin†, Chia-Ying Tsao†, Jun-En Ding†
Ming-Feng Tsai‡ and Chuan-Ju Wang†

Presenter: (JH) Jia-Huei Ju

† Research Center for Information Technology Innovation, Academia Sinica,
‡ Department of Computer Science, National Chengchi University

Table of Contents

Introduction

Year 3 in CAsT

Turns	Topic #106: Utterances (u), canonical passage responses (r)
u_1	I just had a breast biopsy for cancer. What are the most common types?
r_1	More research is needed. Types Breast cancer can be: Ductal carcinoma: ...
u_2	Once it breaks out, how likely is it to spread?
r_2	Even though this condition doesn't spread, it's important to keep ...
u_3	How deadly is it?
r_3	**In 1999, a student opened fire at ..., In 2000, LCI was locked down...**(irrlevant)
u_4	**What? No, I want to know about the deadliness of lobular carcinoma in situ.**
r_4	It's sometimes difficult to separate the two conditions and in this case it will be des

1. More natural and explicit feedback.
2. Canonical responses.
3. Document corpus, but passage return.

Methodology

Multi-stage Pipeline for CAsT

	Reformulation			**Retrieving**	**Re-ranking**
	H^D_{-1}	H^Q_{-1}		H_0	H_1
	Document Expansion	Query Rewrite	Query Expansion	Sparse Retrieve	Point-wise Ranking
Module	Doc2query-T5	NTR-T5	PQE-T5	BM25 (Anserini)	MonoT5/MonoT5M

We leverage T5 pretrained model checkpoint on our methods.

- Doc2query-T5: Document expansion by query prediction.
- NTR-T5: Neural Transfer Reformulation.
- **PQE-T5: Paraphrase Query Expansion.**
- BM25(Anserini[1]): Classic sparse retrieval algorithm.
- **MonoT5/MonoT5M: T5 point-wise re-ranker.**

[1] https://github.com/castorini/anserini

How to apply these modules in CAsT?

Document Expansion (DE) Nogueira et al. [4, 5]

Expand each document d by its predicted queries.

$$d' = d \oplus (\hat{q}_1 \oplus \hat{q}_2, \ldots \hat{q}_{10}) \quad \text{,where } \hat{q} = \mathcal{F}_{DE}(d)$$

Neural Transfer Reformulation (NTR) [1, 3]

Reformulate raw utterance u into standalone $\bar{q}$ by the context $(u_{<i}, r_{<i})$

$$\bar{q}_i = \mathcal{F}_{QR}\Big(\Omega(u_{1:i-1}, r_{-3:-1})|||u_i\Big)$$

Paraphrase Query Expansion (PQE)

Expand each query q by its paraphrased question.

$$q' = q \oplus \mathcal{F}_{QE}(q)$$

Seq2seq Document Ranking model (MonoT5 [6]/ MonoT5M [2])

Calculate the relevance scores s_i of each query-passage pair (q_i, p).

$$s_i = \mathcal{F}_{Rank}(q, p_i)$$

Paraphrase Query Expansion

After the effective query reformulation by NTR,
we would like to explore the QE approach for further improvement.

- Consistent to the meaning of the original query. (emphasize)
- Add relevant new terms. (various)

$\Longrightarrow$ **Paraphrase Query Expansion (PQE)**

The expanded q' consist of reformulated query and its self-paraphrase.

$$q' = \bar{q} \oplus \mathcal{F}_{QE}(\bar{q})$$

User utterance u	What are some treatment options?
Reformulated query $\bar{q}$	What are some treatment options for light drinking during pregnancy?
Paraphrase query $\mathcal{F}_{QE}(\bar{q})$	What are some advices to stop drinking alcohol during pregnancy ?

Paraphrase Query Expansion

Paraphrase generation is literally like:

- Traditional seq2seq task for T5.
  - $\text{Question}_2 = \mathcal{F}_{QE}(\text{Question}_1)$
- Fine-tuned on *Quora Question Pair* dataset.
  - Duplicated pair only.

Used	Question_1 (source)	Question_2 (target)	is_duplicate
O	How should I prepare for CA final law?	How one should know that he/she completely prepare for CA final exam?	1
X	What happens if you drink a soda expired by six months?	What would happen if I ate only chocolate for 6 months?	0

monoT5M: fine-tuning with multiple views

We try the point-wise ranking model, monoT5M, which is a variant reranker of monoT5.

Same purpose, (of point-wise re-ranker)

$$s_i = \mathcal{F}_{Rank}(q, p_i)$$

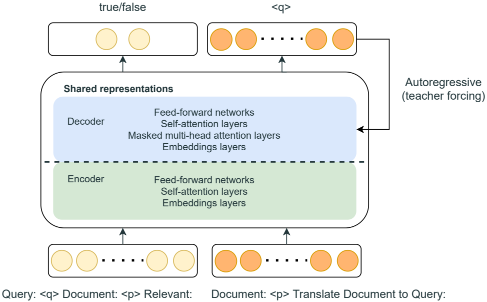
but different fine-tuning process:

Re-ranker	Source	Target
monoT5 [6]	Query <q> Document: <p> Relevant:	true/false
monoT5M [2]	Query <q> Document: <p> Relevant:	true/false
	Document: <p> Translate Document to Query:	<q>

By additionally fine-tuning on the passage-to-query task, we can make the ranking model more generalized.

monoT5M

monoT5M [2]: representation of query-passage pair with multiple views.
And it can be implemented by random sampling.



So, How's the effectiveness of PQE and monoT5M?

Experiments & Result

PQE: Retrieval Performance

Condition	CAsT 2020		CAsT 2021	
	MAP	Recall	d-MAP	d-Recall
Manual Rewrite (Baseline)				
BM25	0.1866	0.7304	0.2363	0.7487
+PQE	0.1870	0.7320	0.2456	0.7659
+PQE (POS-filter)	**0.1984**	**0.7334**	**0.2541**	**0.7723**
+PQE+DE (MRUN)	-	-	**0.2812**	**0.7839**
Automatic Rewrite (Baseline)				
BM25	0.1099	0.5209	0.1741	0.6242
+PQE	0.1099	0.5224	0.1760	0.6315
+PQE (POS-filter)	**0.1279**	**0.5524**	**0.1811**	**0.6322**
+ Our Rewrite (ARUN)	**0.1335**	**0.5947**	**0.2012**	**0.6895**

So far in the first-stage, the retrieval performance of PQE:

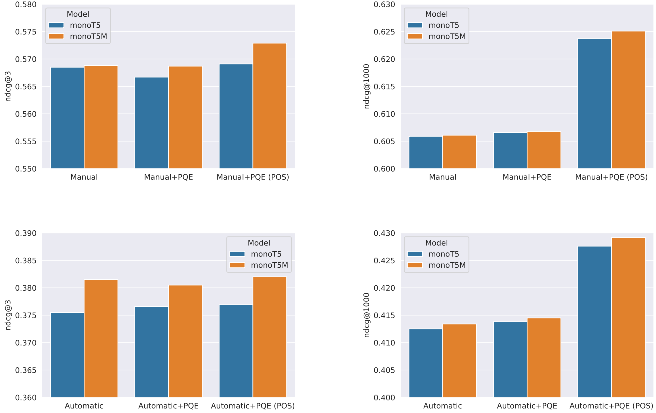
- Small improvement, but POS filter works.
- Work well with Document Expansion.

PQE: Full-ranking Performance

CAsT 2020	Retrieval					Re-ranking			
	nDCG@3	nDCG@500	nDCG	MAP	Recall	nDCG@3	nDCG@500	nDCG	MAP
Manual Rewrite (Baseline)									
BM25	0.2398	0.3985	0.4232	0.1866	0.7304	**0.5685**	0.6039	0.6059	**0.3958**
BM25 w/ PQE	**0.2413**	**0.3998**	**0.4244**	**0.1870**	**0.7320**	0.5667	**0.6041**	**0.6066**	0.3955
Automatic Rewrite (Baseline)									
BM25	0.1451	0.2599	0.2838	0.1099	0.5209	0.3755	0.4103	0.4125	0.2482
BM25 w/ PQE	**0.1473**	**0.2602**	**0.2844**	**0.1099**	0.5224	**0.3766**	**0.4116**	**0.4138**	**0.2487**

- Hard to tell the noise or signal in the first-stage.
  (Inconsistent performances b/w Retrieval and Re-ranking)

monoT5M: Re-ranking effectiveness



- monoT5M is More effective on **nDCG@3**.
- Especially on **Automatic Rewrite**.

Conclusion

Conclusion

Document/Query Expansion

- Higher recall of the document candidates.
- Inconsistency between Retrieval and Re-ranking.

Passage Re-ranking:

- monoT5M can perform better in the shallower depth.
- The improvement spaces are still large.

References i

[1] A. Elgohary, D. Peskov, and J. Boyd-Graber. Can you unpack that? learning to rewrite questions-in-context. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5918–5924, Hong Kong, China, Nov. 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1605. URL https://aclanthology.org/D19-1605.

[2] J.-H. Ju, J.-H. Yang, and C.-J. Wang. Text-to-text multi-view learning for passage re-ranking. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 1803–1807, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380379. doi: 10.1145/3404835.3463048. URL https://doi.org/10.1145/3404835.3463048.

[3] S.-C. Lin, J.-H. Yang, R. Nogueira, M.-F. Tsai, C.-J. Wang, and J. Lin. Conversational question reformulation via sequence-to-sequence architectures and pretrained language models, 2020.

[4] R. Nogueira, J. Lin, and A. Epistemic. From doc2query to docTTTTTquery. *Online preprint*, 2019.

[5] R. Nogueira, W. Yang, J. Lin, and K. Cho. Document expansion by query prediction. *arXiv preprint arXiv:1904.08375*, 2019.

[6] R. Nogueira, Z. Jiang, R. Pradeep, and J. Lin. Document ranking with a pretrained sequence-to-sequence model. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 708–718, Online, Nov. 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.63. URL https://aclanthology.org/2020.findings-emnlp.63.

Thank You!

Are there any questions you'd like to ask?

Jia-Huei Ju	jhjoo@citi.sinica.edu.tw
Ming-Feng Tsai	mftsai@nccu.edu.tw
Chuan-Ju Wang	cjwang@citi.sinica.edu.tw