R3 : Refined Retriever-Reader Pipeline for MultiDoc2Dial

S. Bansal* S. Agarwal* S. Tripathi* S. Gururaja* A. Veerubhotla* R. Dutt T. Mitamura E. Nyberg

Language Technologies Institute - Carnegie Mellon University

Introduction & Summary

- MultiDoc2Dial is a conversational question answering dataset that grounds dialogues in multiple documents, coming from the same domain. The task requires a model to determine a relevant document to answer an user's utterance given the dialogue history and generate a response to it.
- Consists of 30k query turns; grounded in 4k passages across 4 public domains for MDD-SEEN task. For MDD-UNSEEN task, dialogues are from a domain not seen during training.
- Baseline : Retrieval-Augmented Generator (RAG) model which uses a fine-tuned dense passage retrieval (DPR) model to find relevant passages and a pre-trained BART model to generate the response by marginalizing over the passage scores.
- Our proposed approach employs sparse representations for passage retrieval, a passage re-ranker, fusion-in-decoder architecture for generation, and a curriculum learning training paradigm.
- Our approach shows a **12-point improvement in BLEU** score compared to the baseline RAG model and our submission is **ranked 1st in the MDD-UNSEEN**.

Results

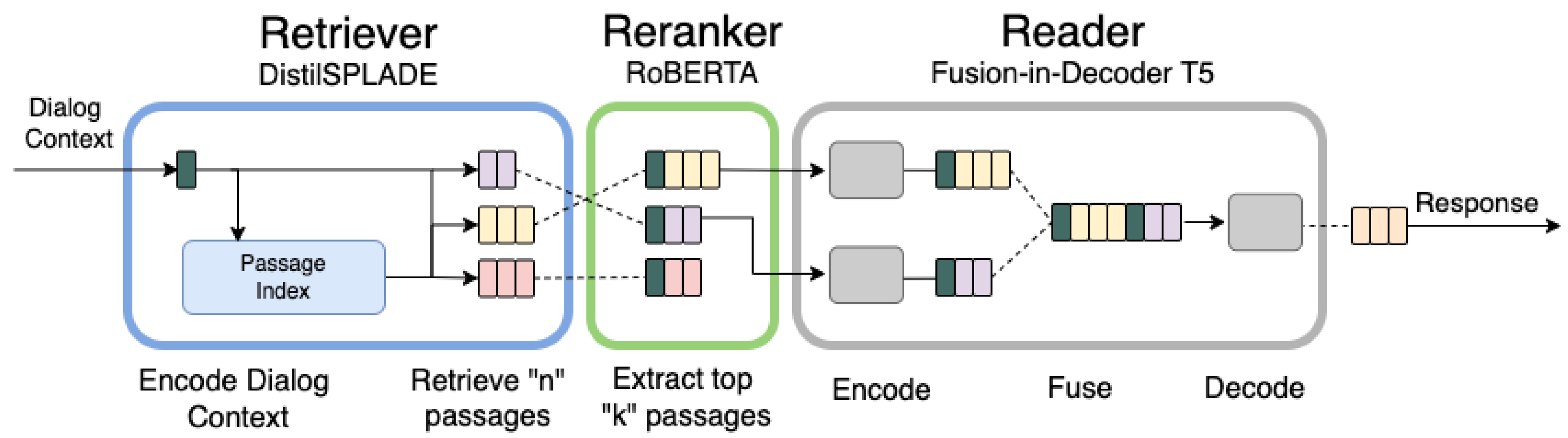
Model	R@10	R@100
DPR-FT (Baseline)	73.2	92.8
SPLADE-max-FT	75.1	93.9
DistilSPLADE-FT	77.0	94.8
DistilSPLADE-FT+DPR-FT(Neg)	**78.6**	**94.9**
DistilSPLADE-FT+DPR-FT(Neg) + Reranker	**85.7**	**94.9**

Table 1: Performance of the retriever for different model configurations at R@10 and R@100.

Model	Reader	EM	F1	BLEU	RougeL
Baseline	BART	3.6	33.8	19.2	31.4
DistilSPLADE + RAG	BART	4.8	38.5	23.7	36.2
DistilSPLADE + FiD	T5	5.1	42.3	29.7	40.2
DistilSPLADE + FiD + RR	T5	5.5	43.1	30.1	41.1
DistilSPLADE + FiD + RR + CL (**M1**)	T5	5.3	**43.3**	**31.1**	**41.4**
DistilSPLADE + FiD + Top10	T5	5.3	42.4	30.5	40.6
DistilSPLADE + FiD + Top10 + RR	T5	5.5	42.5	30.4	40.7
DistilSPLADE + FiD + Top10 + RR + CL (**M2**)	T5	**5.6**	43.0	30.5	41.0
M1 (on Shared Task MDD-SEEN test)	T5	-	46.2	31.8	44.2
M2 (on Shared Task MDD-UNSEEN test)	T5	-	33.0	25.0	32.0

Table 2: Model performance on the validation split for EM, F1, BLEU and RougeL.

Methodology



We employ the standard retriever-reader architecture used in open-domain question answering.

Retriever (DistilSPLADE) : This model takes the user's current turn and dialogue context as the query to select top-n relevant passages. We use DistilSPLADE as our retriever, which augments the query and passages, subsequently projecting them to a sparse vector in the vocabulary space, which gave an improvement of 5% over the baseline DPR for R@10.

Reranker (RR) : The passages retrieved by the bi-encoder are then passed through a RoBERTA based cross-encoder. The model is trained to capture the relevance between the query and the passage through cross-attention. Applying the reranker on top of the retrieved passages results in a increase of 7% over the bi-encoder for R@10.

Reader (FiD) : An abstractive T5-based fusion in decoder (FiD) model is used to generate agent responses. The FiD model encodes all the top-k reranked passages one-by-one and concatenates them to form the input to the decoder. The decoder then learns to collect evidence from multiple passages to generate the response.

Curriculum Learning (CL) : We also experiment with training our model using a curriculum learning approach, by bucketing the data into "easy", "medium" and "hard" samples, based on the reader model's performance. Then, we train in four phases, with increasing levels of difficulty.

Top 10 (Top10) : To add extra supervision to the reader, we passed the ground truth passage at the top while training the reader.

Conclusion & Future Directions

- Sparse DistilSPLADE retriever and Fusion-in-decoder (FID) as the reader.
- Cross-attention based reranker to further boost recall scores.
- Refine the training process through curriculum learning to handle the diverse complexity.

Future directions : Better dialogue modelling, response ranking and including previous grounded knowledge in modeling.