FiD를 적용한 end-to-end 한국어 오픈 도메인 질의 응답

End-to-End Korean Open-domain Question Answering Using FiD

강대욱, 나승훈, 김태형, 류휘정, 장두성

dwkng@jbnu.ac.kr, nash@jbnu.ac.kr, taehyeong2019.kim@kt.com,
hwijung.ryu@kt.com, dschang@kt.com

전북대학교, KT

Introduction

최근 신경망을 사용한 오픈 도메인 질의 응답 연구가 증가하고 있음

두 개의 BERT 인코더를 이용해 질의와 문서의 유사도를 구하는 연구는 BM25와 TF-IDF를 뛰어넘는 성능을 보여줌

검색 모델을 통해 얻은 문서를 통해 정답을 유추할 때 문서에서 정답의 위치를 예측하는 것 보다 문서와 질의를 통해 정답을 생성하는 방식이 최근 더 뛰어난 성능을 얻고 있음

본 논문에서는 FiD와 RAG의 특성을 합쳐 검색모델과 생성 모델을 end-to-end로 학습하는 연구를 진행

Related Works

RAG는 검색 모델을 통해 얻은 문서 각각을 이용해 문장들을 생성한 후 문장들의 확률을 취합하여 최종 문장을 생성

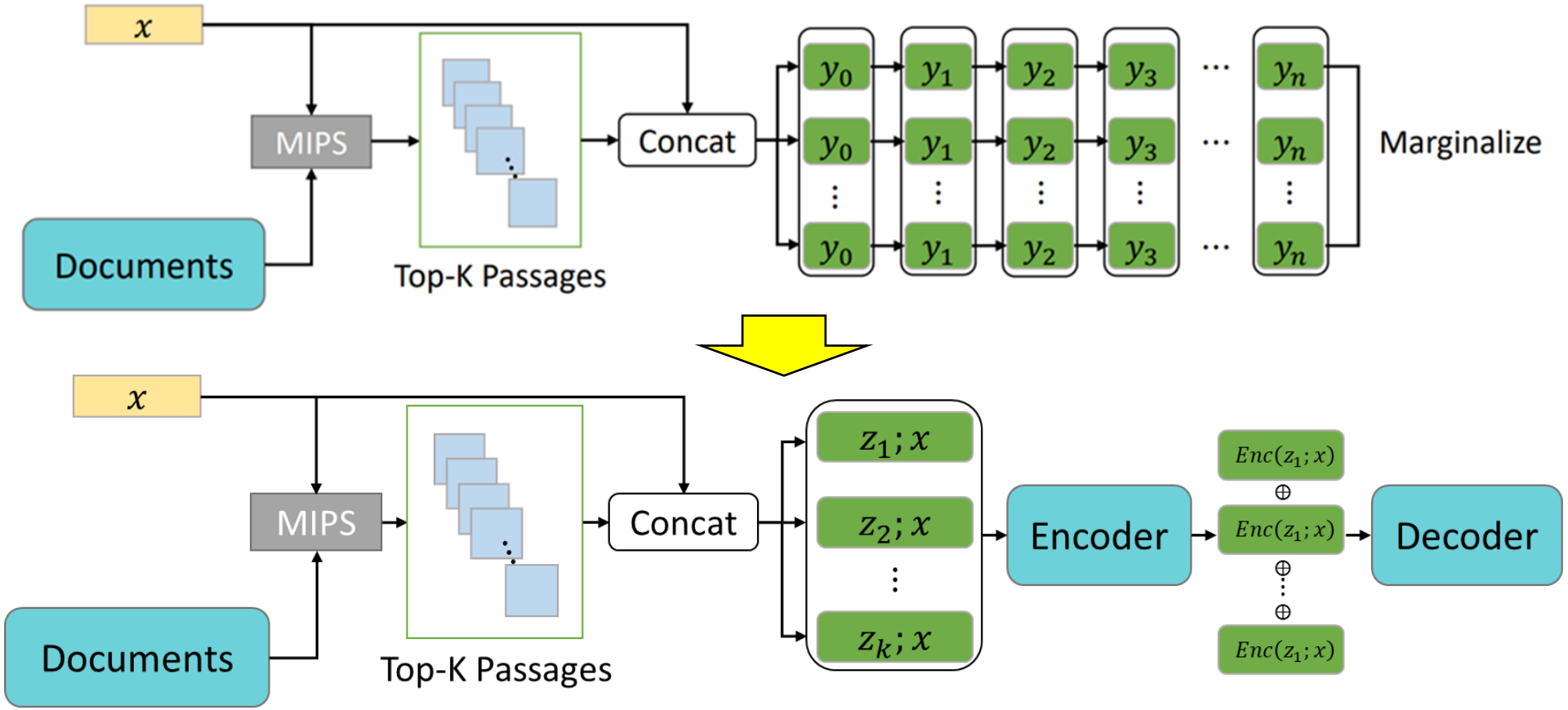
Fusion-in-Decoder는 생성 모델의 인코더로 문서 각각을 인코딩한 뒤 모두 이어붙여 한 번에 디코딩하여 문장을 생성

그러나 Fusion-in-Decoder는 현재 높은 성능을 보이고 있지만 검색 모델을 학습할 수 없음

Methods

Fusion-in-Decoder 방식으로 생성 모델에 저장된 지식을 활용하며 RAG의 end-to-end 학습 특성을 유지

RAG의 Marginalize 부분을 제거하고 문서 유사도를 디코더 내부 교차 어텐션시 합산해 end-to-end 특성 유지



Methods

질의 x에 대한 Top-K개 문서 z를 MIPS 연산 이후 각 문서에 질의 x를 이어붙여 인코더 입력 생성

생성한 인코더 입력을 인코더에 입력한 뒤 인코더의 출력을 모두 이어붙임

$$\mathbf{d}(z) = \mathrm{BERT}_d(z), \quad \mathbf{q}(x) = \mathrm{BERT}_q(x)$$

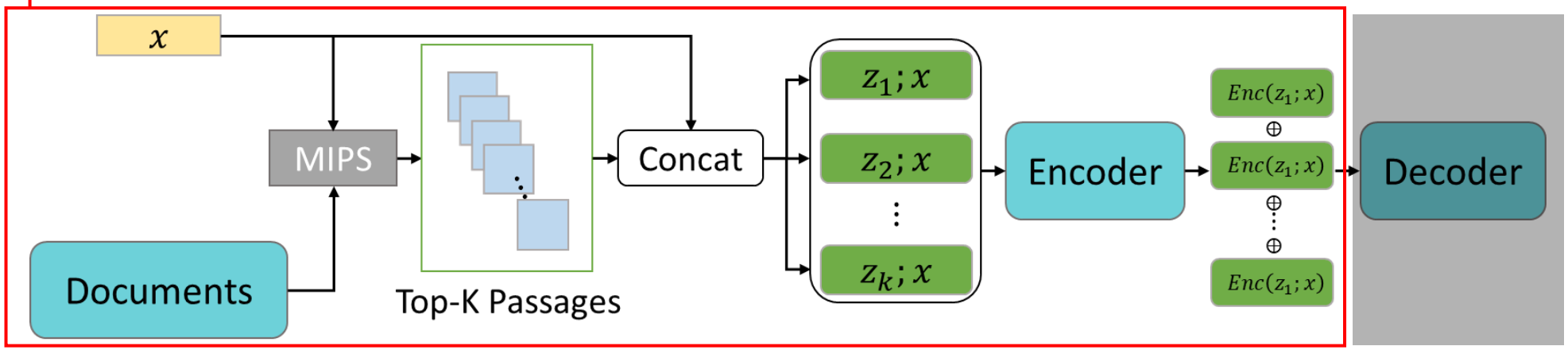
$$p_\eta(z|x) = d(z)^\intercal q(x)$$

$$p_\eta(Z|x) = \mathrm{top{-}k}\left(p_\eta(z|x)\right), \qquad Z = \{z_1, z_2, \ldots, z_k\}$$

$$z_i; x \;\; for \; z_i \in Z$$

$$E_o^{(i)} = Enc(z_i; x)$$

$$E_c = E_o^{(1)} \oplus \cdots \oplus E_o^{(k)}$$

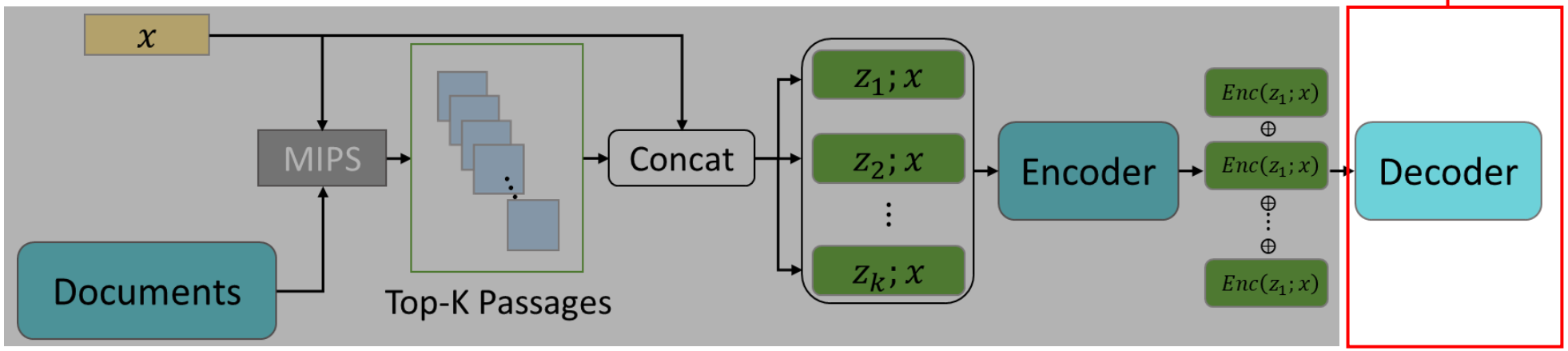


Methods

이어붙인 인코더 출력을 디코더에 입력해 교차 어텐션 수행

$$p_\theta(y_i|x, z, y_{1:i-1}) \approx Dec(y_{1:i-1}, E_c, p_\eta(Z|x))$$

$$p_{RAG-FiD}(y|x) \approx \prod_i^N p_\theta(y_i|x, z, y_{1:i-1})$$



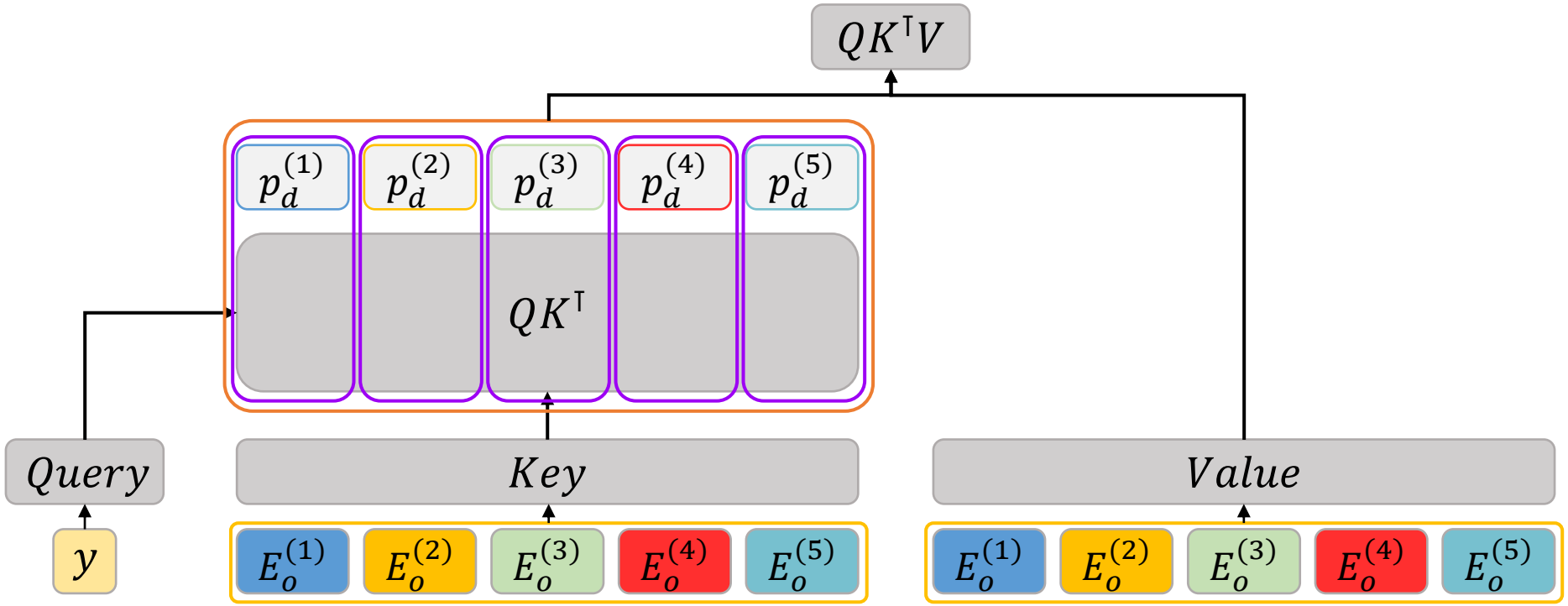
Methods

매 디코더 층에서 교차 어텐션시 문서의 유사도 p_d를 곱하여 문서의 유사도를 모델에 반영

$$p_d(Z|x) = 1 + Softmax\left(\frac{p_\eta(Z|x)}{T}\right) - mean(Softmax\left(\frac{p_\eta(Z|x)}{T}\right))$$

$$p_d^{(i)} = p_d(z_i|x), z_i \in Z$$

$$p_d = [p_d^{(1)}; \dots; p_d^{(i)}; \dots; p_d^{(k)}]$$



Experiments

REALM의 인코더를 Retriever로 사용

한국어 T5를 Generator로 사용

한국어 위키피디아 20년 5월 1일자 덤프를 외부 지식으로 사용

KTQA 데이터에서 20 epoch 동안 미세조정 후 성능 측정

문서 유사도를 곱할 때 Temperature를 주어 반영 정도를 조절

블록 수	블록 당 평균 문장 수	블록 당 평균 단이 수
87,233	4.94	67.81

표 1. 한국어 위키피디아 데이터 구성

	$\|\mathcal{D}\|$	Avg.$\|\mathcal{A}\|$
Train	15,900	1.36
Dev	900	1.35
Test	1,800	1.35

표 2. KTQA 데이터 구성

Results

실험 결과 교차 어텐션에서 유사도를 곱해주는 방식이 성능을 해치는 것을 확인할 수 있음

Model	Temperature	All		Has Answer	
		EM	F1	EM	F1
RAG-Token	X	**53.14**	**66.53**	**67.41**	**86.08**
RAG-Sequence	X	50.25	62.93	64.30	81.59
RAG-FiD	1	17.97	42.12	5.67	50.56
	2	15.08	42.28	11.37	50.95
	5	29.88	51.11	24.09	60.02
	10	23.7	51.52	18.73	64.47

KTQA에서의 EM,F1 성능 평가