

복사 메커니즘과 강화 학습을 적용한 BERT 기반의 문서

요약 모델

황현선^{0,†}, 이창기[†], 고우영^{††}, 윤한준^{††}

강원대학교[†], ETRI 부설 연구소^{††}

{hhs4322, leeck}@kangwon.ac.kr, {gwy876, hjyoon}@nsr.re.kr

BERT-based Document Summarization using Copying-Mechanism and Reinforcement Learning

Hyunsun Hwang^{0,†}, Changki Lee[†], Woo-Young Go^{††}, Han-Jun Yoon^{††}

Kangwon National University Dept. of Computer Science[†], The Affiliated Institute of ETRI^{††}

요약

문서 요약은 길이가 긴 원본 문서에서 의미를 유지한 채 짧은 문서나 문장을 얻어내는 작업을 의미한다. 딥러닝을 이용한 자연어처리 기술들이 연구됨에 따라 end-to-end 방식의 자연어 생성 모델인 sequence-to-sequence 모델을 문서 요약 생성에 적용하는 방법들이 연구되었다. 본 논문에서는 여러 자연어처리 분야에서 높은 성능을 보이고 있는 BERT 모델을 이용한 자연어 생성 모델에 복사 메커니즘과 강화 학습을 추가한 문서 요약 모델을 제안한다. 복사 메커니즘은 입력 문장의 단어들을 출력 문장에 복사하는 기술로 학습데이터에서 학습되기 힘든 고유 명사 등의 단어들에 대한 성능을 높이는 방법이다. 강화 학습은 정답 단어의 확률을 높이기 위해 학습하는 지도 학습 방법과는 달리 연속적인 단어 생성으로 얻어진 전체 문장의 보상 점수를 높이는 방향으로 학습하여 생성되는 단어 자체보다는 최종 생성된 문장이 더 중요한 자연어 생성 문제에 효과적일 수 있다. 실험결과 기존의 BERT 생성 모델 보다 복사 메커니즘과 강화 학습을 적용한 모델의 Rouge score가 더 높음을 확인 하였다.

주제어: 생성 요약, BERT, Copy mechanism, 강화 학습

1. 서론

문서 요약이란 길이가 긴 문서에서 의미를 유지한 채 짧은 문서나 문장으로 표현하여 사람이 이해하기 쉬운 형태로 만드는 것을 의미한다. 딥러닝이 자연어처리에 적용되기 전 요약 기술들은 어려운 요약 생성 보다는 문서에서 중요한 문장만을 판단하여 추출하는 추출요약 위주로 연구되었다[1]. 그러나 딥러닝의 sequence-to-sequence 모델을 통한 end-to-end 방식의 자연어 생성 기술들이 연구된 이후 해당 모델을 이용한 생성 요약 모델들이 연구되었다[2,3].

Sequence-to-sequence 모델은 문장을 입력 받아 인코더로 인코딩 한 뒤 해당 정보를 이용하여 디코더로 한 단어씩 생성하여 문장을 출력하는 하나의 신경망으로 이루어진 end-to-end 방식의 자연어 생성 모델이다[4,5]. Sequence-to-sequence 모델은 주로 RNN(Recurrent Neural Network)을 이용하고 성능 향상을 위해 주의 집중 방법(attention mechanism)을 이용한다[6]. 주의 집중 방법은 긴 입력 문장으로 인한 성능 저하를 해결하기 위한 기술로 디코더에서 입력 단어들의 중요도를 계산하는 기술이다. 이로 인해 sequence-to-sequence 모델은 긴 입력 문장에서 높은 성능을 낼 수 있었으며 이후 다양한 자연어처리 기술에 응용 되었다. 응용 모델 중의 하나인 Transformer 모델은 sequence-to-sequence 모델의 기본 구조인 RNN을 변경하여 multi-head attention 기술만으로

인코더와 디코더를 구성한 모델로 자연어 생성 문제에서 높은 성능을 보여주었으며 해당 모델의 인코더를 이용한 사전학습 모델인 BERT도 많은 자연어처리 태스크에서 연구되고 있다[7,8]. 또 다른 응용 모델인 복사 메커니즘(Copying mechanism)[9,10]은 자연어 생성 문제에서 발생하는 Out-of-vocabulary(OOV) 문제와 학습데이터에서 거의 학습되지 않는 고유명사 문제 등을 해결하기 위해 디코더의 출력 위치에서 복사를 위한 또 다른 attention network를 이용하여 입력 단어들의 스코어를 계산하여 입력 단어들을 해당 출력에 복사할 지를 결정한다.

지도 학습은 기계학습에서 자주 쓰이는 학습 방법 중 하나로 원하는 출력 결과를 얻기 위해 어떤 출력 결과가 나와야 하는지 정해진 정답 학습데이터를 학습하게 된다. 그러나 정답 학습데이터가 없거나 원하는 결과가 어떤 연속적인 행위의 결과인 경우 지도 학습 방법으로는 학습하기 어렵다는 문제가 있다. 강화 학습은 이러한 문제에 적용하는 방법으로 연속적인 행위를 선택한 뒤 그 최종 결과의 보상(reward) 점수를 높이는 방향으로 학습한다.

본 논문에서는 sequence-to-sequence 모델을 이용한 생성 요약 모델의 성능을 높이기 위해 다음의 기술들을 적용한다. Multi-head attention 기반의 transformer 모델을 이용하여 sequence-to-sequence 모델의 인코더와 디코더를 구성하고 인코더의 경우 최근 자연어처리에서 높은 성능을 보이는 사전학습 모델인 BERT를 적용한다.

디코더에서는 입력 문장의 고유명사 등이 출력 문장에서 다시 등장하는 요약 생성 문제 특성에 따라 적용할 수 있는 복사 메커니즘을 적용하며 이때 최신 복사 메커니즘인 Pointer-Generator 모델을 적용한다[10]. 추가적으로 sequence-to-sequence 모델의 지도학습의 문제점인 학습과 실제 적용될 때의 디코더 입력 단어가 다른 문제를 해결하기 위해 강화 학습을 적용한다.

2. 관련 연구

문서 요약은 크게 추출 요약과 생성 요약의 두가지 방법론으로 연구되었다. 추출 요약은 원본 문서의 문장들 중에서 중요하다고 판단되는 문장들을 추출하는 요약 방법이고, 생성 요약은 원본 문서의 내용을 이해하여 중요한 내용을 포함하는 새로운 요약문을 생성하는 요약 방법이다[3]. [2]에서는 딥러닝을 이용한 자연어 생성 모델인 sequence-to-sequence 모델을 이용하여 한국어 뉴스 기사 생성 요약 모델을 제안하였다. [3]에서는 문서 요약 생성의 성능을 높이기 위해 문서 요약 추출을 시행한 후 추출 결과를 요약 생성 모델의 입력으로 넣어 요약문을 생성하였다.

BERT는 자연어처리를 위한 사전학습 모델 중 하나로 자연어 생성 문제에서 좋은 성능을 보인 Transformer 모델[7]의 인코더를 이용하여 언어 모델로 산전 학습한 것이다. BERT는 연속적인 문장 쌍을 대용량 말뭉치로 사전 학습하여 학습된 모델을 다른 자연어처리 태스크에 적용하여 fine-tuning 한다는 특징이 있다. [11]과 [12]에서는 BERT를 이용한 문서 요약 모델을 제안하였다. [11]에서는 추출과 생성 모델의 인코더 부분을 BERT로 구성하였고 생성 모델의 디코더는 GRU와 copying mechanism을 적용하였다. [12]에서는 문서 요약을 위한 새로운 BERT 사전 학습 방법을 제안하며 [11]과 마찬가지로 인코더는 BERT로 구성하며 디코더는 Transformer의 디코더로 구성하였다.

복사 메커니즘(copying mechanism)은 자연어 생성 모델에서 입력 단어가 출력 단어에 복사 될 수 있게 하는 기술로 입력 문장과 출력 문장의 언어가 같거나 중요한 고유 명사들이 출력 문장에서도 등장하는 문제에 적합한 기술로 문서 요약 생성에 적용 될 수 있다. [2]에서는 sequence-to-sequence 모델에 copying mechanism을 적용한 문서 생성 요약 모델을 제안 하였다. 제안된 모델은 기존의 모델보다 고유 명사등의 문제에 강건한 성능을 보였다. [13]에서는 기계 번역 분야에서 좋은 성능을 보인 Transformer 모델에 복사 메커니즘을 적용하여 문서 요약 생성 모델을 제안하였다.

강화 학습은 어떠한 연속적인 행위의 선택을 결정하기 위해 정답을 학습하는 것이 아닌 최종 결과의 보상 점수를 높이는 쪽으로 학습하는 기계학습 방법이다. [14]에서는 sequence-to-sequence 모델의 특징 중 하나인 디코더에서 현재의 출력 단어를 결정할 때 이전 출력단어에 의존하는 모델의 문제점을 지적하였다. 현재 이러한 문제점은 beam search 알고리즘을 이용하여 해결을 시도하고 있으나 실제 모델의 학습에서는 디코더가 정답 입력 단

어를 학습한다는 문제가 있다. [14]에서는 디코더가 출력 단어를 결정하는 것을 행위 선택의 연속이라고 보고 자연어 생성 모델에 강화 학습을 적용 하였다. [15]에서는 마찬가지로 sequence-to-sequence 모델을 이용한 문서 요약 생성 모델에 강화 학습을 적용하는 방법을 제안 하였다.

본 논문은 BERT 인코더와 Transformer 디코더로 문서 요약 모델을 제안한 [12]의 연구에 기반한다. [12]의 모델에 추가적으로 copying mechanism을 적용하여 고유 명사 등에 강한 요약 생성 모델을 제안하며 추가적으로 sequence-to-sequence 모델의 문제점이었던 학습과 실제 적용될때의 디코더 입력 단어가 다른 문제에 대해 [15]의 강화 학습 방법을 적용하였다.

3. 복사 메커니즘과 강화 학습을 적용한 BERT 기반의 문서 요약 생성 모델

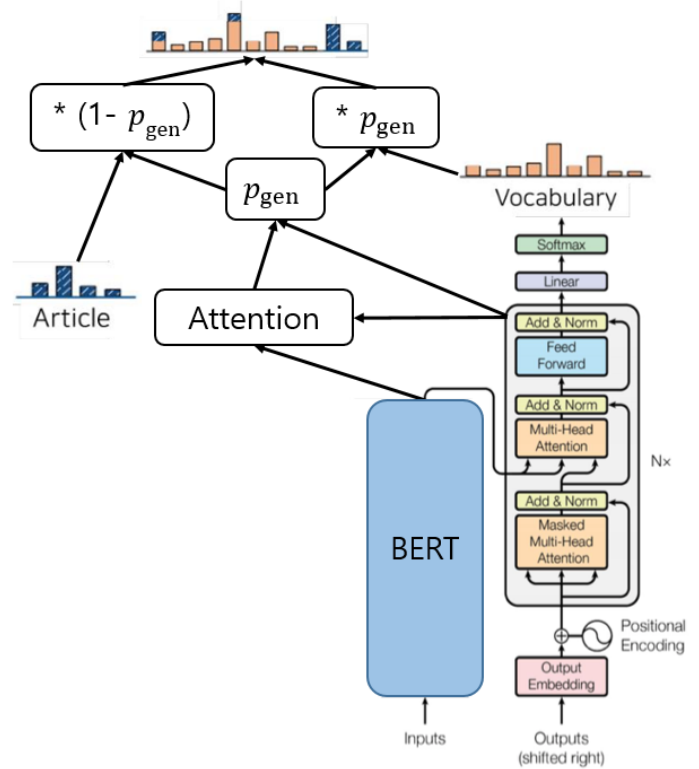


그림 1. 복사 메커니즘을 적용한 BERT 기반의 문서 요약 생성 모델 그림

3.1 BERT 기반 Sequence-to-sequence 모델

그림 1은 본 논문에서 제안하는 복사 메커니즘을 적용한 BERT 기반의 문서 요약 생성 모델의 전체 그림이다. 기존의 Transformer 기반의 sequence-to-sequence 모델 [7]에서 인코더 부분을 BERT로 대체하게 된다. 입력 문장은 BERT 모델을 통해 인코딩 되며 BERT의 마지막 레이어를 입력 문장의 인코딩 정보로서 활용한다. 해당 정보를 가지고 Transformer 디코더가 디코딩을 수행하게 되며 마지막 출력 단어 생성부분에서 복사 메커니즘이 적

용되어 최종 출력 단어를 결정하게 된다

3.2 Pointer-Generator 복사 메커니즘 모델

복사 메커니즘(copying mechanism)은 [10]의 방법을 적용하였다. Pointer-generator 모델의 수식은 다음과 같이 표현 될 수 있다.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

$$p_{\text{gen}} = \sigma(W_g[h_{\text{dec}}, \text{Attention}(W_c^Q h_{\text{dec}}, W_c^K h_{\text{enc}}, W_c^V h_{\text{enc}})])$$

$$P_{\text{vocab}}(w) = \text{softmax}(W_{\text{output}} h_{\text{dec}})$$

$$P(w) = P_{\text{gen}} P_{\text{vocab}}(w) + (1 - P_{\text{gen}}) \sum_{i:w_i=w} a_i^t$$

h_{dec} 는 디코더의 히든 스테이트(hidden state)이며 복사 메커니즘을 적용하지 않는 모델의 경우 여기에 가중치 W_{output} 을 곱하고 softmax 함수를 적용하여 최종 출력 단어의 확률 $P_{\text{vocab}}(w)$ 을 계산하게 된다. Pointer-generator 모델은 여기에 입력 단어들을 복사할 것인가를 결정할 확률 P_{gen} 을 soft switch로서 계산하게 된다. P_{gen} 은 0과 1사이의 값으로 해당 디코딩 시간 t 에서 복사 스코어 a^t 를 또 다른 attention network로 구하여 해당하는 출력 단어 확률에 P_{gen} 을 곱하고 더하여 최종 출력 단어 확률인 $P(w)$ 를 계산한다. 결과적으로 기존 출력 단어 확률에 attention network를 이용하여 복사 스코어를 계산한 뒤에 단순히 더하여 복사 메커니즘을 구현한 [9]와는 다르게 Pointer-generator 모델은 복사 스코어를 계산한 뒤에 복사할 것인가를 soft switch 방식을 통해 결정하도록 구현하여 좀더 정밀한 복사가 가능한 모델이다.

3.3 강화 학습

본 논문에서는 sequence-to-sequence 모델을 이용한 문서 요약 생성 모델에 강화 학습을 적용하였다. 강화 학습은 기본적으로 행위를 정할 agent(딥러닝 모델)과 해당 agent가 정한 action(행위)가 존재하고 이에 대한 environment(환경)이 정해지면 보상(reward) 점수를 계산한다. 이때 강화학습의 목표는 보상을 최대화 하도록 agent를 수정하게 된다. 이를 sequence-to-sequence 모델에 대응 시키면 agent는 sequence-to-sequence 모델이 되고, action은 sequence-to-sequence 모델의 각 시간별 출력 단어가 된다. Environment는 디코더가 생성해낸 문장의 보상 점수를 계산하는 함수가 되며 여기서 보상은 생성된 문장의 Rouge 스코어가 된다. 이에 따라 문서 요약 생성을 위한 강화학습은 다음과 같은 수식으로 표현 될 수 있다.

$$L_{ml} = - \sum_{t=1}^{n'} \log p(y_t^* | y_1^*, \dots, y_{t-1}^*, x)$$

$$L_{rl} = (r(\hat{y}) - r(y^s)) \sum_{t=1}^{n'} \log p(y_t^s | y_1^s, \dots, y_{t-1}^s, x)$$

$$L_{mixed} = \gamma L_{rl} + (1 - \gamma) L_{ml}$$

L_{ml} 은 기존의 지도학습 방식의 sequence-to-sequence 모델의 목적함수로 각 디코딩 시간의 정답 출력 단어(y^*)에 대한 loss를 계산한다. L_{rl} 은 강화학습을 위한 목적함수이다. 이때 $r()$ 은 보상을 계산하는 함수로 생성된 sample 출력 문장의 rouge 1,2,L 스코어의 평균에서 bigram과 unigram 단어의 반복 비율의 평균을 뺀 값을 보상 점수로 계산한다. 강화학습을 위해 생성된 sample 문장인 y^s 는 greedy search 알고리즘을 수정하여 출력 단어 확률에서 sampling을 사용하여 구하고, variance를 줄이기 위한 baseline인 \hat{y} 은 greedy search 알고리즘을 이용하여 구한다. 최종적으로 파라미터 γ 를 이용하여 기존의 지도학습 방법의 목적함수와 강화학습의 목적함수를 결합하여 목적함수 L_{mixed} 를 실제 학습에 사용한다.

4. 실험 결과

본 논문에서 제안한 문서 요약 생성 모델의 성능을 평가하기 위해 [3]에서 구축한 한국어 뉴스 기사(학습셋 109,558 문서, 개발셋 3,786 문서, 평가셋 3,705문서)를 사용하였다. BERT 모델은 [16]에서 사용된 ETRI의 KorBERT를 사용하였다. 모델들은 모두 Pytorch로 구현하였으며 요약 성능은 Rouge 스코어를 이용하여 측정 하였다. [3,11]에서의 연구와 동일하게 추출요약을 수행하여 추출요약의 결과를 입력 문장으로 사용하였다.

표 1. 각 요약 생성 모델들의 성능표

Model	γ	Beam	R-1	R-2	R-L
BERT 추출(MMR) + BERT encoder + GRU decoder + Copy[11]	-	5	32.85	15.78	29.44
BERT 추출(MMR) + BERT encoder +Transformer decoder	0	5	37.13	17.00	32.39
		10	37.08	17.19	32.51
BERT 추출(MMR) + BERT encoder +Transformer decoder + Copy	0	5	37.27	17.37	32.48
		10	37.10	17.50	32.53
BERT 추출(MMR) + BERT encoder +Transformer decoder + Copy + Reinforce	0.05	5	37.06	17.43	32.47
		10	37.05	17.66	32.69
	0.1	5	37.50	18.01	33.00
		10	37.29	17.97	33.03
	0.2	5	37.01	17.53	32.56
		10	36.93	17.75	32.73
	0.3	5	36.97	17.54	32.50
		10	37.09	17.79	32.82
	0.5	5	37.19	17.67	32.64
		10	37.12	17.73	32.78

참고문헌

표 1은 각각의 요약 생성 모델들의 성능표이다. [11]의 모델은 BERT 기반 인코더에 GRU decoder와 복사 메커니즘을 사용한 모델이다. R-1, R-2, R-L은 Rouge-1, Rouge-2, Rouge-L 스코어를 나타낸다. BERT 기반 인코더와 Transformer 디코더를 사용한 모델의 경우 [11]의 모델 보다 높은 성능을 보여 GRU 디코더 보다 Transformer 디코더가 더 좋은 성능을 보임을 알 수 있다. 해당 요약 생성 모델에 [10]의 Pointer-generator 방식의 복사 메커니즘을 적용한 모델의 경우 0.02~0.37의 Rouge 성능 향상을 보여 요약 생성 모델에서 복사 메커니즘이 유용함을 확인하였다. 강화학습의 경우 학습 비율(3.3절의 파라미터 γ)을 조절해가며 실험하였다. 강화학습을 적용하였을 때 전체적으로 비슷한 성능을 보였으나 강화학습 비율 0.1에서 Rouge-1 37.50, Rouge-2 18.01, Rouge-L 33.00로 적용하지 않은 모델보다 높은 Rouge 성능을 보였다.

표 2는 각 요약 모델별 생성 예제이다. Ground Truth는 사람이 만든 정답 요약문이며 그 아래는 모델들의 Beam size 별 요약문을 나타낸다. 복사 메커니즘이 적용되지 않은 모델의 경우 Beam size 1일 때에는 ‘딸기/NNG 제철/NNG 딸기/NNG 를/JKO’ 등의 잘못된 단어들을 생성하는 것을 볼 수 있으며 Beam size 5일 때에도 ‘베리밀 크프로프로그래/NNP’ 라고 잘못된 단어를 생성하는 문제점을 확인할 수 있다. 그러나 복사 메커니즘이 적용된 모델의 경우 Beam size 1일 때에는 ‘베리밀크프로마쥬/NNP’ 로 고유명사를 정확하게 생성해 내는 것을 볼 수 있으며 그 아래 복사 확률(copy prob)에서 해당 단어를 생성할 때 복사 확률이 높아지는 것을 볼 수 있고, 이는 복사 메커니즘이 고유명사 등의 문제를 해결 할 수 있음을 알 수 있다. Beam size 5의 경우 정답 요약문과 유사한 문장을 생성해 냈으며 강화학습이 적용된 모델도 비슷한 요약문을 만들어냈다.

5. 결론

본 논문에서는 BERT 모델과 Transformer 디코더를 결합한 생성 요약 모델의 성능을 높이기 위해 복사 메커니즘과 강화학습을 적용하였다. 실험 결과 복사 메커니즘이 생성 요약 성능을 향상 시킬수 있음을 확인하였으며 강화학습의 경우도 생성 요약 모델의 성능이 향상 될 수 있음을 확인하였다.

향후 연구로는, 본 논문에서 개발한 모델을 다른 자연어처리 태스크에 적용해 볼 예정이다.

감사의 글

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2013-2-00131, 휴먼 지식증강 서비스를 위한 지능진화형 Wise QA 플랫폼 기술 개발). 이 논문은 NSR의 지원을 받아 수행된 연구임.

- [1] Jae-Min Yoon, You-Jin Chung, Jong-Hyeok Lee, "Automatic Extractive Summarization of Newspaper Articles using Activation Degree of 5W1H," Journal of KIISE, SA, Vol. 31, No. 4, pp. 505-515, 2004.
- [2] 최경호, 이창기. "Copy Mechanism과 Input Feeding을 이용한 End-to-End 한국어 문서요약", 제28회 한글 및 한국어 정보처리 학술대회, 2016
- [3] 전재원, 황현선, 이창기. "추출 요약과 생성 요약을 결합한 2단계 문서 요약", KCC, 2019, 581-583
- [4] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems, 2014.
- [5] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078, 2014.
- [6] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473, 2014.
- [7] Vaswani, Ashish, et al. "Attention is all you need." Advances in Neural Information Processing Systems. 2017.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805, 2018.
- [9] Gu Jiatao, et al, "Incorporating copying mechanism in sequence-to-sequence learning." arXiv preprint arXiv:1603.06393 (2016).
- [10] See, Abigail, Peter J. Liu, and Christopher D. Manning. "Get to the point: Summarization with pointer-generator networks." arXiv preprint arXiv:1704.04368 (2017).
- [11] 전재원, 황현선, and 이창기. "BERT 를 이용한 2 단계 한국어 문서 요약." 한국정보과학회 학술발표논문집 (2019): 347-349.
- [12] Liu, Yang, and Mirella Lapata. "Text summarization with pretrained encoders." arXiv preprint arXiv:1908.08345 (2019).
- [13] 전동현, 강인호. "복사-메커니즘과 추론 단계의 페널티를 이용한 Copy-Transformer 기반 문서 생성 요약." 제31회 한글 및 한국어 정보처리학술대회 논문집 (2019).
- [14] Ranzato, Marc'Aurelio, et al. "Sequence level training with recurrent neural networks." arXiv preprint arXiv:1511.06732 (2015).
- [15] Paulus, Romain, Caiming Xiong, and Richard Socher. "A deep reinforced model for abstractive summarization." arXiv preprint arXiv:1705.04304

(2017).

[16] J. Bae, C. Lee, S. Lim, and H. Kim, "Korean Semantic Role Labeling with BERT," KCC 2019.

표 2. 각 요약 생성 모델들의 생성 예제

Ground Truth		'/SS 아시아/NNP 의/JKG 알프스/NNP '/SS 라고/JKQ 불리/VV 는/ETM 일본 /NNP 홋카이도/NNP 의/JKG 명물/NNG '/SS 르타오/NNP 치즈/NNG 케익/NNG '/SS 이/JKS 봄/NNG 딸기/NNG 를/JKO 만나/VV 았/EP 다/EF ./SF
BERT 추출(MMR) + BERT encoder + Transformer decoder	Beam size 1	향긋/XR 하/XSA ㄴ/ETM 딸기/NNG 와/JC 달콤/XR 하/XSA ㄴ/ETM 맛/NNG 이/JKS 일품/NNG 이/VCP ㄴ/ETM 르타오/NNP 가/JKS 싱그럽/VA ㄴ/ETM 봄 /NNG 에/JKB 잇/VV 어/EC 이번/NNG 에/JKB 는/JX 딸기/NNG 제철/NNG 딸기 /NNG 를/JKO 출시/NNG 하/XSV 았/EP 다/EF ./SF
	Beam size 5	르타오/NNP 코리아/NNP 가/JKS 봄/NNG 시즌/NNG 을/JKO 맞/VV 아/EC 새콤달콤/XR 하/XSA ㄴ/ETM 딸기/NNG 가/JKS 등뿔/MAG 올라가/VV ㄴ/ETM '/SS 베리밀크프로그래/NNP '/SS 를/JKO 출시/NNG 하/XSV 았/EP 다 /EF ./SF
BERT 추출(MMR) + BERT encoder + Transformer decoder + Copy	Beam size 1	봄/NNG 을/JKO 맞/VV 아/EC 고소/XR 하/XSA ㄴ/ETM 딸기/NNG 를/JKO 가 득/MAG 담/VV 은/ETM 신/XPN 제품/NNG '/SS 베리밀크프로마쥬/NNP '/SS 가/JKS 출시/NNG 되/XSV 어/EC 눈길/NNG 을/JKO 끝/VV ㄴ다/EF ./SF
	Beam size 1 (copy prob)	봄/NNG_:0.526 을/JKO_:0.000 맞/VV_:0.000 아/EC_:0.000 고:0.819 소:0.135 /X:0.004 R_:0.000 하/XSA_:0.000 ㄴ/ETM_:0.000 딸:0.982 기 /NNG_:0.999 를/JKO_:0.002 가득/MAG_:0.005 담/VV_:0.000 은 /ETM_:0.000 신/XPN_:0.150 제품/NNG_:0.000 '/SS_:0.129 베:0.999 리:0.865 밀:0.970 크:0.989 프로:1.000 마:0.997 쥬:1.000 /NNP_:0.992 '/SS_:0.256 가/JKS_:0.001 출시/NNG_:0.007 되/XSV_:0.000 어 /EC_:0.000 눈길/NNG_:0.001 을/JKO_:0.000 끝/VV_:0.000 ㄴ다 /EF_:0.000 ./SF_:1.000 [SEP]:1.000
	Beam size 5	딸기/NNG 덕후/NNG 들/XSN 의/JKG 취향/NNG 을/JKO 저격/NNG 하/XSV = /ETM 르타오/NNP 치즈/NNG 신/XPN 제품/NNG 이/JKS 나오/VV 았/EP 다 /EF ./SF
BERT 추출(MMR) + BERT encoder + Transformer decoder + Copy + Reinforce	Beam size 5	딸기/NNG 덕후/NNG 들/XSN 의/JKG 취향/NNG 을/JKO 저격/NNG 하/XSV = /ETM 르타오/NNP 케익/NNG 신/XPN 제품/NNG 이/JKS 출시/NNG 되/XSV 았 /EP 다/EF ./SF