

Neural Attention을 반영한 문장 생성 모델

이세희⁰, 이지형^{*}

⁰성균관대학교 전자전기컴퓨터공학과

e-mail: crimson12@skku.edu⁰, john@skku.edu^{*},

Sentence generation model with neural attention

Seihee Lee⁰, Jee-Hyung Lee^{*}

⁰Department of Electrical and Computer Engineering, Sungkyunkwan University

● 요약 ●

자연어 처리 분야에서 대화문 생성, 질의응답 등과 같은 문장생성과 관련된 연구가 꾸준히 진행되고 있다. 본 논문에서는 기존 순환신경망 모델에 Neural Attention을 추가하여 주제 정보를 어느 정도 포함시킬지 결정한 뒤 다음 문장을 생성할 때 사용하는 모델을 제안한다. 이는 기존 문장과 다음 문장의 확률 정보를 사용할 뿐만 아니라 주제 정보를 추가하여 문맥적인 의미를 넣을 수 있기 때문에, 더욱 연관성 있는 문장을 생성할 수 있게 도와준다. 이 모델은 적절한 다음 문장을 생성할 뿐만 아니라 추가적으로 어떤 단어가 다음 문장을 생성함에 있어 주제문장에 더 민감하게 반응하는지 확인할 수 있다.

키워드: 주목 (Attention), 문장 생성 (Sentence generation), 장단기 기억 네트워크 (Long Short Term Memory networks)

I. Introduction

오늘날 질의응답, 문서 요약, 설명문 생성, 대화문 등에서 문장 생성 모델에 관련된 연구가 활발히 진행되고 있다. 특히 딥러닝 기법이 발달함에 따라 자연어 처리 연구를 위하여 텍스트에 내재된 의미 및 구문을 벡터 공간에 표현하는 언어 모델링 연구가 활발하게 진행되고 있다. 대부분의 언어 모델링은 텍스트에 나타나는 단어들의 등장 패턴을 학습한다. 이와 같이 단어들의 등장 패턴과 이를 모델링하는 연구[2]는 문장의 패턴정보를 잘 학습하나, 문장 간의 연결성이나 전체 텍스트의 문맥적 의미를 파악하지는 못하는 단점이 있다. 이를 보완하기 위하여 본 연구는 주제 정보를 다음 문장 예측 시에 사용하여 모델링에 포함시킬 수 있도록 하는 모델을 제안한다. 실험을 통해 본 모델이 주제 정보를 포함하지 않았을 때 보다 더 나은 성능을 보였으며, 이는 제안모델이 보다 연관성 있는 문장을 생성함을 의미한다.

II. Related work

현재 문장 생성을 위한 모델링을 위해서 Attention, Sequence-to-Sequence 등 관련연구가 진행되고 있다. Bahdanau는 Attention을 처음 제안하였으며, 번역 문장을 생성함과 동시에 연관된 부분을 파악하였다.[5] Kyunghyun는 기존 Sequence to Sequence 모델을 확장하여 RNN Encoder-Decoder 모델을 제안하였다.[4] 이와 같은 문장 생성 모델에는 문맥 정보가 적게 들어있기 때문에, [3]과 유사한 기법을 사용하여 모델에 주제와 연관된 단어별 주제 정보를 포함시켜 문맥 정보를 파악할 수 있게 함으로써 더 나은 문장 생성 모델을 만들고자 한다.

III. Proposed Scheme

1. 제안 방법

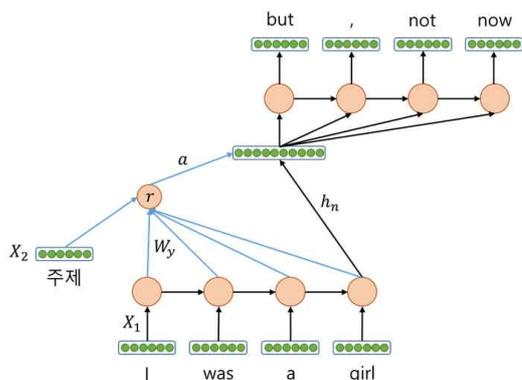


Fig. 2. Structure of Proposed Method

$$Y = [h_1, \dots, h_N] \quad (1)$$

$$r = \text{softmax}(Y \cdot W_y) \quad (2)$$

$$a = r \cdot X_2 \quad (3)$$

$$c = [a, h_n] \quad (4)$$

다음 단어 및 문장을 생성하기 위해 주제와 같은 외부 정보를 각 주어진 문장에 따라 적절히 대입시키기 위하여 위와 같은 모델을 제안한다. 이는 입력 문장 X_1 , 주제 벡터 X_2 로 구성되어 있다. 첫 번째 LSTM의 입력 값은 X_1 으로 각 출력 h_i 를 만든다. 이렇게 만들어진 h_i 는 서로 결합되어 주제벡터를 어느 정도 통과시킬지 결정하는 r 이 된다. 이렇게 만들어진 r 은 주제 벡터 X_2 의 통과정도를 결정하여 a 를 생성한다. 이렇게 만들어진 a 는 LSTM의 마지막 출력값 h_n 과 결합되어 다음 문장을 생성하는 벡터 c 가 된다. 이렇게 만들어진 c 는 다음 LSTM의 입력 값이 되어 다음 문장을 생성하게 된다.

2. 실험

실험 데이터는 Document Understanding Conference(DUC)1)에서 제공하는 DUC 2002와 DUC 2004를 사용하였다.

실험 결과는 다음 Table 1과 같다. 6,000줄의 문장을 사용하였고, 주제를 넣은 기법과 주제를 넣지 않은 모델을 사용하여 학습을 하였다. 총 50회의 epoch를 수행하였고, 실험 결과 제안기법이 좋은 성능을 보임을 알 수 있다.

Model	accuracy (%)
Without Topic	47.03
Proposed Method	78.04

Table 1. The Result of Accuracy

IV. Conclusions

본 논문에서는 기존 순환신경망 모델에 Neural Attention을 추가하여 주제 정보를 어느 정도 포함시킬지 결정한 뒤 다음 문장을 생성할 때 추가로 사용하는 모델을 제안하였다. 이는 Attention을 통해 주제와 같은 외부 정보를 포함시킨 LSTM 기반 모델이 적절한 문장 모델링을 하는 것을 보였다. 추후 다양한 데이터를 이용한 실험을 통해 여러 분야에 적용하여 검증하고자 한다.

Acknowledgment

본 논문은 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원(B0101-16-0559, 디지털 소상공인 지원을 위한 지역 비즈니스 전략 분석 및 맞춤형 영상홍보 창작 SW 플랫폼 개발)과 2014년도 미래창조과학부의 재원으로 한국연구재단-차세대정보 컴퓨팅기술개발사업의 지원을 받아 수행된 연구임(No. NRF-2014M3C4A7030503)

References

- [1] Sepp Hochreite, Jurgen Schmidhuber, "LONG SHORT-TERM MEMORY," Neural Computation, Vol. 9, No. 8, pp. 1735-1780, 1997
- [2] Rocktäschel, Tim, et al, "Reasoning about entailment with neural attention," arXiv preprint arXiv:1509.06664, 2015.
- [3] Mikolov, T. et al, "Distributed Representations of Sentences and Documents," ICML, Vol. 14, pp. 1188-1196, 2014
- [4] Cho, Kyunghyun, et al, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014
- [5] Bahdanau, Dzmitry, Kyunghyun Cho, Yoshua Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014

1) <http://www-nlpir.nist.gov/projects/duc/intro.html>