

듀얼 포인터 네트워크 디코더를 이용한

정답 후보군 탐지 시스템

장영진⁰, 김학수, 김진태*, 왕지현*, 이충희*

강원대학교 컴퓨터정보통신공학과, (주)엔씨소프트

dan_yon@kangwon.ac.kr⁰, nlpdrkim@kangwon.ac.kr, kjt1505@ncsoft.com*, Korjhwang@ncsoft.com*,

forever.7374@ncsoft.com*

Answers Candidate Detection System using Dual Pointer Network Decoder

Youngjin Jang⁰, Harksoo Kim, Jintae Kim*, Jihyun Wang*, Chunghee Lee*

Kangwon National University computer and Communication Engineering

NCSoft Corp*

요 약

정답 후보군 탐지 모델은 최근 활발히 진행되고 있는 질의-응답 데이터 수집 연구의 선행이 되는 연구로 특정 질문에 대한 정답을 주어진 단락에서 추출하는 작업을 말한다. 제안 모델은 포인터 네트워크 디코더를 통하여 기존의 순차 레이블링 모델에서 처리할 수 없었던 정답이 겹치는 문제에 대해서 해결할 수 있게 되었다. 그리고 독립된 두 개의 포인터 네트워크 디코더를 사용함으로써, 단일 포인터 네트워크로 처리할 수 없었던 정답의 탐지가 가능하게 되었다.

주제어: 정답 후보군 탐지, 질문 생성, 기계독해, 듀얼 포인터 네트워크

1. 서론

최근 질의-응답 시스템의 연구가 활발히 진행되고 있다. 기계독해 시스템은 질의-응답 시스템의 대표적인 작업 중 하나로, 주어진 문서에서 질문에 대한 정답을 찾는 시스템이다. 이러한 기계독해 시스템은 문서의 이해, 질문 작성 등의 작업을 통해 데이터를 수집하기 때문에, 데이터를 구축하기 위해서는 높은 비용을 요구한다. 최근 이 비용 문제를 완화하기 위해 심층 학습(Deep Learning)을 이용한 질문 자동 생성 모델[1-2]이 연구되고 있다. 질문 생성 모델은 단락과 정답 정보를 기반으로 정답에 적합한 질문을 생성한다. 단락에서 정답 정보를 추출하는 작업 또한 수작업으로 진행하기에는 높은 비용이 요구되기 때문에 질문 생성 모델을 위해 정답 후보를 자동으로 추출하는 연구가 필요하다. 따라서 본 논문에서는 질문 생성 모델의 선행 연구로 듀얼 포인터 네트워크 디코더(Dual Pointer Network)[3]를 사용한 정답 후보군 탐지 시스템을 제안한다.

2. 관련 연구

정답 후보군 탐지에 대한 기존 연구는 심층 신경망을 이용한 방법[4]과 규칙 기반 시스템[2]이 있다. 규칙 기반 정답 탐지 시스템은 학습 데이터가 필요 없다는 장점을 가지고 있지만 규칙을 직접 구축해야하기 때문에 높은 비용을 발생시키며, 구절 형태의 정답(서술형 정답)을 탐지하기 어렵다는 단점이 있다. 이에 반해 심층 신경망 기반 정답 탐지 시스템은 학습 데이터를 요구한다

는 단점이 있지만 규칙을 따로 구축할 필요가 없고 학습 데이터에 따라 구절 형태의 정답 탐지가 가능하다는 장점이 있다. 본 논문에서는 단답형 정답뿐만 아니라 서술형으로 이루어진 정답 후보를 탐지하기 위하여 심층 학습 기반의 정답 후보군 탐지 모델을 제안한다. [5]는 멘션 탐지 작업에 포인터 네트워크[6]을 사용하여 순차 레이블링(Sequence Labeling) 모델에서는 다룰 수 없었던 정답 중첩 문제를 해결했다. [7]은 포인터 네트워크와 멀티헤드 주의 집중 방법(Multi-head Attention)을 의존 구문 분석에 적용하여 높은 성능을 보였다. 위의 연구를 통해 포인터 네트워크는 입력 열에 대응되는 위치를 출력하는데 최적화 되어있는 것을 알 수 있다. [3]은 포인터 네트워크를 관계추출 작업에 사용함으로써 입력 문장에 존재하는 모든 관계를 추출하고, 단일 포인터 네트워크로 처리할 수 없었던 하나의 주체가 두 개 이상의 객체를 갖는 경우에 대해서 두 개의 포인터 네트워크를 통해 문제를 해결하는 모델을 제안했다. 본 논문에서는 정답 중첩 문제를 해결하기 위해 [5]와 같이 포인터 네트워크를 정답 후보군 탐지 모델에 적용하고, 시작 위치가 같은 다수의 정답 후보를 탐지하기 위하여 두 개의 포인터 네트워크를 통해 문단 내에 존재하는 정답 후보를 추출한다.

그림 1은 포인터 네트워크가 가리키는 방향과 포인터 네트워크 디코더 수에 따른 정답 후보군 추출 예시를 보여준다. (a)는 정답의 시작 위치에서 정답의 끝 위치를 가리키는 단일 포인터 네트워크 디코더를 사용한 예시이고, (b)는 정답의 끝 위치에서 정답의 시작 위치를 가리

키는 단일 포인터 네트워크 디코더를 사용한 예시를 보여준다. a와 b를 통해 순차 레이블링 모델에서는 탐지할 수 없었던 안긴 정답((a):[“4대 군주이며 언어학자”, “언어학자”], (b):[“4대 군주”, “4대 군주이자 언어학자”])을 탐지하는 것을 확인할 수 있다. (c)는 (a)와 (b)에서 사용된 포인터 네트워크를 함께 사용했을 때의 정답 추출 예시를 보여준다. (c)를 통해서 기존의 단일 포인터 네트워크(a, b)로는 탐지할 수 없었던 [“4대 군주”], [“언어학자”]라는 정답 추가로 탐지하는 것을 확인할 수 있다.

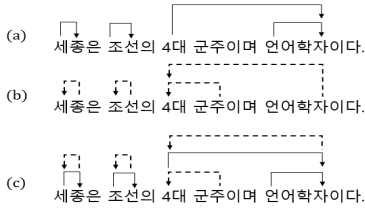


그림 1. 포인팅 방향과 포인터 네트워크 디코더 수에 따른 정답 후보 예시

3. 정답 후보군 탐지 모델

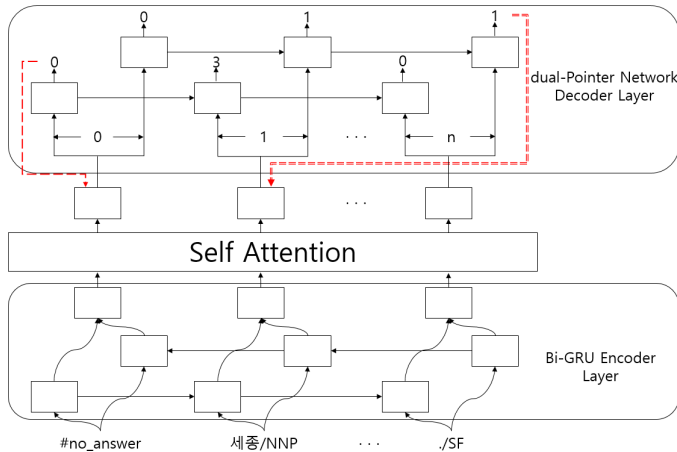


그림 2. 제안 모델 구조

그림 2는 본 논문에서 제안하는 모델의 구조를 보여준다. 제안 모델은 순환 신경망 인코더, 자가 주의 집중 (Self Attention)[8] 계층, 그리고 포인터 네트워크 디코더로 이루어져있다. 인코더는 양방향 순환 신경망[9]으로 구현되어 있으며, 자가 주의 집중 계층을 거친 출력 값은 포인터 네트워크에 입력된다. 포인터 네트워크로 구현된 두 개의 디코더는 각각 정답의 시작과 끝 위치에서 정답의 끝 위치와 시작 위치를 가리킴으로써, 정답 후보를 탐지한다.

3.1. 순환 신경망 인코더와 자가 주의 집중 계층

제안 모델의 순환 신경망 인코더는 GRU(Gated Recurrent Unit)[10]을 사용하며 형태소 벡터와 자질을 연결한 벡터를 입력 받는다. 인코더에 입력되는 자질은 품사 정보가 사용된다. 순환 신경망의 인코더를 통해 인코딩된 출력 값은 자가 주의 집중 계층을 통해 인코딩된 정보를 강조한다. 자가 주의 집중 계층에서 사용되는

수식은 (1)과 같다.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

d_k 는 은닉 벡터의 크기를 나타내며, Q, K, V 는 모두 인코더 출력 값을 의미한다.

3.2. 포인터 네트워크 디코더

제안 모델의 디코더는 독립된 두 개의 포인터 네트워크로 구성된다. 첫 번째 포인터 네트워크는 정답의 시작 위치에서 정답의 끝 위치를 가리키며, 또 다른 포인터 네트워크는 정답의 끝 위치에서 정답의 시작 위치를 가리킨다. 각 디코더는 자가 주의 집중이 반영된 출력 벡터 값과 각 스텝의 위치에 대응되는 벡터를 입력받는다. 현재 스텝의 위치가 정답의 시작이나 끝이 아닐 경우 0번째 위치인 “#no_answer”를 가리키게 된다.

4. 실험

4.1. 실험 준비

본 논문에서 실험한 데이터는 위키피디아, 나무위키 5,050 단락을 기반으로 수집된 질의-응답 데이터를 사용했다. 학습 데이터는 5,000개의 단락을 사용했으며, 실험 데이터는 학습 데이터에 사용되지 않은 단락 50개에서 사람이 직접 정답이 될 수 있는 후보군을 표시해 놓은 데이터를 사용했다. 실험에 사용한 형태소 임베딩은 20GB의 뉴스 기사로 사전 학습한 GloVe[11]를 사용했으며, 품사 정보와 스텝 위치의 벡터는 무작위로 초기화하여 사용했다. 제안 모델의 파라미터는 표 1과 같다.

표 1. 모델 파라미터

배치 크기	64
학습 횟수	50
은닉 크기	200
단어 임베딩 크기	100
단락 최대 길이	150
품사 정보 자질 크기	50
위치 정보 자질 크기	50
학습률	0.0001
드랍아웃	0.2

4.2. 실험 결과 및 분석

성능 평가 지표는 단어 완전 일치(Exact Match:EM)와 단어 부분 일치 측면에서 정밀도와 재현율 그리고 F1 점수를 사용했다. 실험 결과는 표 2와 표 3과 같다.

표 2. 단어 완전 일치 성능

	단어 완전 일치		
	정밀도	재현율	F1 점수
[4]	0.6600	0.2661	0.3793
제안 모델	0.6451	0.3225	0.4301

표 3. 단어 부분 일치 성능

	단어 부분 일치		
	정밀도	재현율	F1 점수
[4]	0.8235	0.4213	0.5575
제안 모델	0.8048	0.4805	0.6017

[4]는 단일 포인터 네트워크 디코더를 사용한 정답 후보군 탐지 모델이다. 표 2와 표 3에 따르면 단일 포인터 네트워크 디코더를 사용하는 것 보다 독립적인 두 개의 포인터 네트워크를 사용하는 것이 F1 점수 기준으로 성능이 더 높은 것을 확인 할 수 있다. 실험 결과에 따르면 제안 모델의 정밀도가 [3]과 비교하여 더 낮은 것을 확인할 수 있는데, 이런 결과의 원인은 제안 모델이 더 많은 정답 후보를 추출했기 때문인 것으로 판단된다. 실제로 [4]의 결과를 확인해보면 76개의 정답 후보를 추출하고 이 중 44개가 정답과 완전 일치했으며, 제안 모델은 98개의 정답 후보를 추출했고 54개가 정답과 일치했다. 아래의 표 4은 제안 모델과 [4]가 실제로 추출한 데이터이다.

표 4. 실제 정답 후보군 탐지 결과

단락	1983년 나이지리아, 말레이시아, 이란 진출.
실제 정답	“1983년”, “나이지리아”, “말레이시아”, “이란”, “나이지리아, 말레이시아, 이란”
[3]	“1983년”, “말레이시아, 이란”, “나이지리아, 말레이시아”
제안 모델	“1983년”, “말레이시아”, “나이지리아”, “이란”, “말레이시아, 이란”, “나이지리아, 말레이시아”

표 4에 따르면 제안 모델과 [4](포인터 네트워크 디코더를 이용한 정답 후보군 탐지 모델)는 순차 레이블링 모델에 추출할 수 없는 정답(안긴 정답)을 추출하는 것을 확인할 수 있다. 그리고 [4]와 제안 모델의 결과를 통해서 단일 포인터 네트워크로는 표현할 수 없는 정답을 두 개의 포인터 네트워크를 사용하여 추출했음을 알 수 있다.

5. 결론

본 논문은 두 개의 포인터 네트워크를 이용하여 입력 단락에서 정답 후보군을 탐색하는 모델을 제안했다. 표 2와 표 3을 통해 단일 포인터 네트워크를 사용하는 것 보다 두 개의 포인터 네트워크를 사용하는 것이 더 효과적인 것을 증명했다. 향후 연구로는 직접 수집한 서술형 질의응답 데이터와 규칙 기반 시스템을 통해 수집한 단답형 데이터를 기존 학습 데이터에 추가하여 정밀도를 보완할 수 있는 실험을 진행할 예정이다. 그리고 제안 모델과 질문 생성 모델과 통합하여 기계독해 시스템 데

이터를 구축하고 데이터를 평가하는 실험을 진행할 예정이다.

감사의 글

본 연구는 엔씨소프트 산학연구용역 과제의 지원을 받아 수행되었음.

참고문헌

- [1] Y. Kim, H. Lee, J. Shin, K. Jung, “Improving Neural Question Generation Using Answer Separation”, *The Third AAAI Conference on Artificial Intelligence(AAAI-19)*, pp.6602-6609, 2019.
- [2] V. Kumar, S. Muneeswaran, G. Ramakrishnan, Y. Li, “ParaQG: A System for Generating Questions and Answers from Paragraphs”, *arXiv preprint arXiv:1909.01642v1*, 2019.
- [3] 박성식, 김학수, “포인터 네트워크를 이용한 문장 내 다중 개체 간의 관계 추출”, *2019년 한국컴퓨터종합학술대회 논문집*, pp.506-508, 2019
- [4] 김진태, 최기현, 김학수, 왕지현, “포인터 네트워크와 자가 주의집중 방법을 이용한 정답 후보군 탐지 시스템”, *2019년 한국컴퓨터종합학술대회 논문집*, pp.572-574, 2018.
- [5] 박천음, 이창기, “포인터 네트워크를 이용한 멘션 탐지”, *정보과학회 논문지*, 44(8), pp774-781, 2017.
- [6] O. Vinyals, F. Meire, & J. Navdeep, “Pointer Networks”, *In Advances in Neural Information Processing Systems*, pp. 2692-2700, 2015.
- [7] 박성식, 오신혁, 김홍진, 김학수, “멀티헤드 어텐션과 포인터 네트워크를 이용한 한국어 의존 구문 분석”, *제 30회 한글 및 한국어 정보처리 학술대회 논문집*, pp.682-684, 2018.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, “Attention is All you Need”, *In Advances in neural information processing systems*, pp. 5998-6008, 2017
- [9] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks”, *IEEE Transactions on Signal Processing*, pp. 2673-2681, 1997
- [10] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling”, *arXiv preprint arXiv:1412.3555*, 2014
- [11] J. Pennington, R. Socher and C. D. Manning. "GloVe: Global Vectors for WordRepresentation", *2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.1532-1543, 2014