

Dual Bi-Directional Attention Flow를 이용한 한국어 기계이해 시스템

이현구[○], 김학수, 최정규*, 김이른*

강원대학교 컴퓨터정보통신공학과, LG 전자 SW센터 인공지능연구소*

nlpghlee@kangwon.ac.kr, nlpdrkim@kangwon.ac.kr, stanley.choi@lge.com, yireun.kim@lge.com

Korean Machine Comprehension using Dual Bi-Directional Attention Flow

Hyeon-gu Lee[○], Harksoo Kim, Jungkyu Choi*, Yi-reun Kim*

Kangwon National University Computer and Communication Engineering

AI Lab., SW Center, LG Electronics*

요 약

기계이해 시스템은 주어진 문서를 이해하고 질의에 해당하는 정답을 출력하는 방법으로 심층 신경망을 활용한 주의집중 방법이 발달하면서 활발히 연구되기 시작했다. 본 논문에서는 어휘 정보를 통해 문서와 질의를 이해하는 어휘 이해 모델과 품사 등장 정보, 의존 구문 정보를 통해 문법적 이해를 하는 구문 이해 모델을 함께 사용하여 기계이해 질의응답을 하는 Dual Bi-Directional Attention Flow 모델을 제안한다. 한국어로 구성된 18,863개 데이터에서 제안 모델은 어휘 이해 모델만 사용하는 Bi-Directional Attention Flow 모델보다 높은 성능(Exact Match: 0.3529, F1-score: 0.6718)을 보였다.

주제어: 기계이해, 질의응답 시스템, 주의집중

1. 서론

질의응답 시스템(Question Answering System)은 자연어로 이루어진 질의를 시스템이 이해하고 적절한 답변을 출력해주는 시스템이다. 최근 심층 신경망 기술이 발전함에 따라 이러한 질의응답 시스템의 방법 중 기계이해(Machine Comprehension) 방식의 질의응답이 활발하게 연구되고 있다. 기계이해 시스템은 정보가 포함된 문서와 그 문서에 포함된 정보를 질문하는 질의 간의 관계를 이해시키고 이를 통해 문서에 나타나는 정보의 위치를 찾아주는 질의응답 방법이다. 다양한 정보가 빠르게 증가하여 제한된 시간 안에 많은 정보 습득이 중요해진 만큼 정보 습득을 도와주는 기계이해 시스템이 활발히 연구되고 있다. 특히 영어권에서는 SQuAD(Stanford Question Answering Dataset)[1]를 사용하는 기계이해 시스템 경진대회를 통해 다양한 모델이 연구되고 있다. 본 논문에서는 한국어 문서와 질의간의 관계를 이해시키고 정답을 찾아내기 위해 어휘와 구문 정보를 함께 활용하는 한국어 기계이해 시스템을 제안한다.

2. 관련 연구

최근 SQuAD 데이터를 사용하는 기계이해 시스템이 활발하게 연구되고 있다. 문서 어휘에서 질의 어휘의 주의집중(Attention)과 질의 어휘에서 문서 어휘의 주의집중을 계산하여 정답의 위치를 반환하는 Bi-Directional Attention Flow[2], 문서에 질의의 주의집중을 계산한

후 self-matching을 통해 문서를 다시 확인하는 R-Net[3], 문서와 질의의 벡터를 내적하는 단순한 방법으로 관계를 계산하고 언어모델(Language Model)을 통해 정답을 찾아내는 Interactive AoA Reader[4] 등이 연구되고 있다. 또한 구문적 정보를 반영하기 위해 구조 정보를 임베딩하여 입력으로 반영한 연구[5], 개체명과 품사를 반영한 연구 등 언어의 특성을 반영하기 위한 연구[6]도 진행되었다. 본 논문에서는 Bi-Directional Attention Flow 모델을 한국어에 맞게 변형하고 품사 등장 정보 및 의존 구문 정보를 추가하여 성능을 향상시키는 Dual Bi-Directional Attention Flow를 제안한다.

3. Dual Bi-Directional Attention Flow

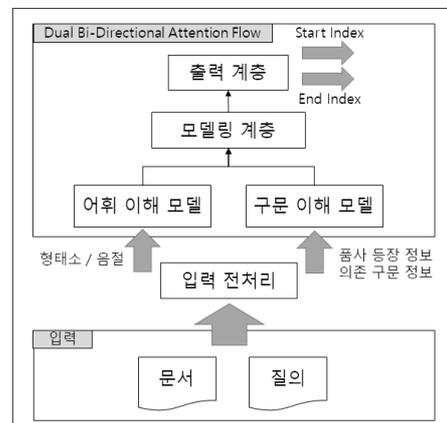


그림 1. 제안 모델의 구조도

[그림 1]은 제안 모델의 구조도를 보여준다. 제안 모델은 문서와 질의 사이의 관계를 이해하기 위해 어휘 정보를 통해 단어와 단어 간의 관계를 찾아내는 어휘 이해 모델과 품사 등장 정보 및 의존 구문 정보를 통해 문법적 관계를 찾아내는 구문 이해 모델로 구성된다. 어휘 이해 모델은 형태소와 음절을 통해 어절 임베딩을 생성하고 문서와 질의 간의 주의집중을 계산한다. 구문 이해 모델은 문장에 나타나는 어절의 품사 등장 정보와 의존 구문분석을 통해 얻어진 의존 구문 표지의 최단 경로 정보를 통해 임베딩을 생성하고 문서와 질의 간의 주의집중을 계산한다. 두 가지 모델을 통해 얻어진 어휘 주의집중 벡터와 구문 주의집중 벡터를 Long Short-Term Memory(LSTM) 순환신경망(Recurrent Neural Network)[7]으로 구성된 모델링 계층에 전달하여 어절 간의 정보를 모델링하고 출력 계층에서 정답에 해당하는 문서의 어절 위치를 반환한다.

3.1. 한국어 어휘 이해 모델

본 논문에서는 문서와 질의 사이의 관계를 이해시키기 위해 먼저 형태소와 음절을 통해 어절 임베딩을 생성한다. 다음으로 문서와 질의에 나타나는 어절 간의 주의집중을 계산하여 문서에 나타나는 어절과 질의에 나타나는 어절 간의 연관성을 찾아낸다. [그림 2]는 한국어 어휘 이해 모델이다.

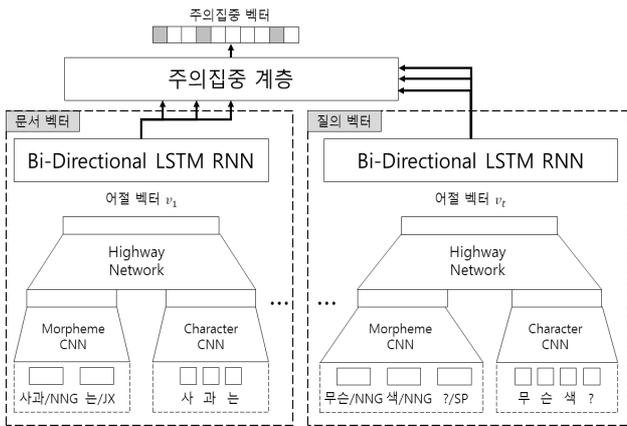


그림 2. 어휘 이해 모델

[그림 2]에서 형태소와 음절을 합성곱 신경망(Convolutional Neural Network)[8]을 통해 어절단위로 각각 임베딩 한 후 생성된 임베딩을 Highway Network[9]에 전달하여 형태소와 음절이 반영된 어절 임베딩을 생성한다. Highway Network는 LSTM과 같은 게이트 이론이 반영된 모델로 기울기 소실 문제(Vanishing gradient problem)[10]를 해결해 벡터를 확실하게 전달할 수 있게 하는 모델이다. 각 어절별 벡터가 생성된 후 양방향 LSTM 순환 신경망을 통해 어절에 문장 단위 정보를 반영한다. 마지막으로 생성된 문서의 어절 임베딩과 질의의 어절 임베딩을 주의집중 계층에 입력하여 주의집중 관계를 계산한다. 주의집중 계층은 식 (1)과 같이 계산된다.

$$\begin{aligned}
 V_{ij} &= \alpha(S_i, Q_j) \\
 \alpha(S_i, Q_j) &= W^T[S_i; Q_j; S_i \circ Q_j] \\
 a_i &= \text{softmax}(V_i) \\
 \tilde{Q}_i &= \sum_{j=0}^n a_{ij} Q_j \\
 b &= \text{softmax}(\max(V_i)) \\
 \tilde{S}_i &= \sum_{i=0}^m b_i S_i \\
 F_i &= \beta(S_i, \tilde{Q}_i, \tilde{S}_i) \\
 \beta(S, \tilde{Q}, \tilde{S}) &= [S; \tilde{Q}; S \circ \tilde{Q}; S \circ \tilde{S}]
 \end{aligned}
 \tag{1}$$

식 (1)에서 S_i 는 문서의 i 번째 어절의 벡터, Q_j 는 질의의 j 번째 어절의 벡터, $S_i \circ Q_j$ 는 S_i 와 Q_j 의 요소별 곱셈(elementwise multiplication)이다. \tilde{Q} 는 질의가 문서에 작용되는 주의집중 벡터, \tilde{S} 는 문서가 질의에 작용되는 주의집중 벡터이다. 즉, 식 (1)에서는 문서와 질의간의 주의집중 가중치를 구하고 문서의 벡터와 결합하여 질의가 문서의 어떤 어절에 중요하게 작용하는지를 찾아낸다.

3.2. 한국어 구문 이해 모델

3.1절에서 언급한 어휘간의 이해 외에도 문법적 관계를 이해하기 위해 어절의 품사 등장 정보 및 의존 구문 정보를 사용하는 구문 이해 모델을 제안한다. [그림 3]은 한국어 구문 이해 모델이다.

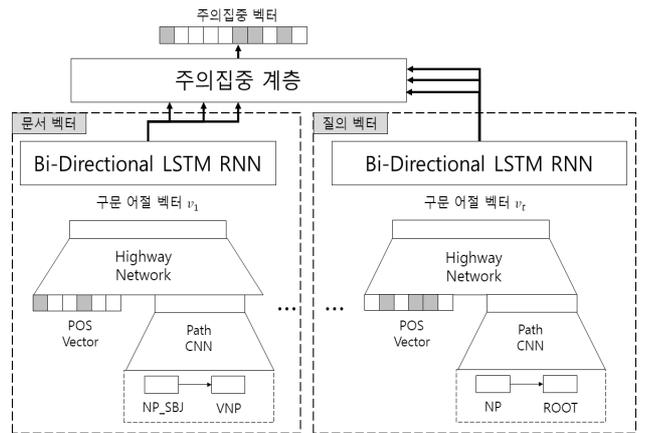


그림 3. 구문 이해 모델

[그림 3]의 품사 등장 정보는 현재 어절에 나타나는 품사를 이진 벡터(어절에서 나타나면 1 아니면 0; 세종 품사 45차원)로 표현하고 의존 구문 정보는 현재 어절에서 문장의 루트 어절까지의 의존 구문 표지 최단 경로 정보로 3.1절의 어휘 벡터 생성과 같이 합성곱 신경망을 적용하여 벡터를 생성한다. 생성된 의존 구문 정보 벡터와 품사 등장 이진 벡터를 Highway Network에 입력하여 구문 어절 벡터를 생성하고 양방향 LSTM 순환 신경망에 입력하여 문장 정보가 반영된 구문 어절 벡터를 생성한다. 마지막으로 3.1절과 마찬가지로 식 (1)을 적용하여

구문 정보가 반영된 문서와 질의의 구문 어절 벡터 간의 주의집중 관계를 계산한다. 계산된 주의집중은 문서에서 정답을 찾기 위해 질의의 품사와 의존 구문 표지가 문서의 어떤 품사, 의존 구문 표지와 연관되는지를 찾아낸다.

3.3. 한국어 기계이해 질의응답 시스템

본 논문에서 제안하는 모델은 한국어 어휘 이해 모델과 구문 이해 모델에서 계산된 주의집중 벡터를 통해 정답에 해당하는 문서의 어절 위치를 반환한다. 정답을 찾기 위한 계층은 어절별로 출력된 주의집중 벡터를 양방향 LSTM 순환신경망을 이용하여 문장단위로 정보를 찾아내는 모델링 계층과 찾아진 정보를 통해 정답의 시작 위치와 끝 위치를 반환하는 출력 계층으로 이루어진다. 모델링 계층은 어휘 이해 모델의 주의집중 벡터와 구문 이해 모델의 주의집중 벡터를 연결한 후 2층으로 구성된 양방향 LSTM 순환신경망에 입력하는 구조로 이루어져있다. 다음으로 모델링 레이어의 출력 결과를 정답의 어절 위치를 찾아내는 출력 계층의 입력으로 사용한다. 출력 계층은 식 (2)와 같이 계산한다.

$$\begin{aligned} start &= \text{softmax}(W_{start}^T [F_{lexicon}; F_{syntax}; M]) \\ end &= \text{softmax}(W_{end}^T [F_{lexicon}; F_{syntax}; M^2]) \end{aligned} \quad (2)$$

식 (2)에서 $F_{lexicon}$ 은 어휘 이해 모델의 주의집중 벡터, F_{syntax} 은 구문 이해 모델의 주의집중 벡터, M 은 모델링 계층의 결과 벡터, M^2 은 $start$ 값을 또 다른 양방향 LSTM 순환신경망에 입력하여 출력된 벡터이다.

3.4. 손실 함수

본 논문에서는 학습을 위해 식 (3)과 같은 손실 함수(loss function)를 사용한다.

$$L(\theta) = -\frac{1}{N} \sum_{i=0}^N \log(start_{y_i^{start}}) + \log(end_{y_i^{end}}) \quad (3)$$

식 (3)에서 $start_{y_i^{start}}$ 와 $end_{y_i^{end}}$ 는 출력 계층의 $start$ 와 end 의 확률 분포에서 실제 정답의 위치에 해당하는 실수값을 나타낸다.

4. 실험 및 평가

4.1. 실험 준비

본 논문에서는 한국어 기계이해 질의응답 시스템을 실험하기 위해 자체 제작한 질의응답 데이터 18,863개를 사용한다. 질의응답 데이터는 여러 문장으로 구성된 문서, 질의, 질의에 해당하는 답변 및 어절 위치로 구성되어 있으며 학습 데이터 18,456개, 평가 데이터 357개, 개발 확인 데이터 50개를 무작위로 나누어 사용한다. 실

험에서 사용한 형태소 분석기와 구문 분석기는 Kacteil 언어 분석기[11-12]를 사용했으며 성능은 형태소 정확도 95.21%, UAS 87.21% LAS 85.28%이다. [그림 4]는 실험에서 사용하는 데이터의 예시이다.

Context	데니스 매캘리스테어 리치(1941년 9월 9일 ~ 2011년 10월 12일)는 미국의 저명한 전산학자이자 현대 컴퓨터의 선구자이다. C와 유닉스로 알려져있다. 1983년에 켄 톰프슨과 “범용 운영체제 이론개발, 특히 유닉스 운영체제의 구현에 대한 공로”로 튜링상을 수상했다.
Question	1983년 누구와 함께 튜링상을 수상했나?
Answer	켄 톰프슨과
Index	start : 1 end : 3

그림 4. 실험 데이터의 예

4.2. 실험 평가

본 논문에서 모델의 성능을 확인하기 위해 기계이해 시스템에서 많이 사용되는 완전 일치율(Exact Match)과 형태소 단위의 F1-score를 성능 지표로 사용한다. 완전 일치율은 모델이 예측한 시작 위치에서부터 끝 위치까지의 형태소가 정답과 모두 일치하는 경우를 의미하고 형태소 단위 F1-score는 모델이 예측한 위치 사이에 존재하는 형태소가 실제 정답과 얼마나 일치하는지를 F1-score로 측정하는 성능이다. 실험에 사용되는 모델은 어휘 이해 모델만을 사용하는 Bi-Directional Attention Flow(BiDAF), 어휘 이해 모델과 구문 이해 모델을 함께 사용하는 제안 모델 Dual Bi-Directional Attention Flow(Dual BiDAF)가 있다. 표 1은 모델의 성능을 보여준다.

표 1. 모델별 성능 비교 (%)

모델	Exact Match	F1-Score
BiDAF	33.33	64.58
Dual BiDAF	35.29	67.18

표 1에서 보는 것과 같이 어휘 이해 모델에 구문 이해 모델을 같이 사용하는 Dual Bi-Directional Attention Flow가 어휘 이해 모델만 사용하는 Bi-Directional Attention Flow보다 완전 일치율 1.96%p, F1-Score 2.6%p 높게 나왔다. 즉, 어휘만을 사용했을 때보다 품사 등장 정보, 의존 구문 정보의 주의집중 벡터가 정답의 경계를 찾아내는데 도움을 주는 것을 알 수 있다.

영어권에서 연구되고 있는 SQuAD(Stanford Question Answering Dataset) 데이터를 사용하는 기계이해 시스템의 경우 완전 일치율 0.7, F1-score 0.8정도로 본 논문에서 제안한 한국어 모델과 큰 성능차이를 보이고 있다. 이는 SQuAD 학습 데이터가 87,599개로 매우 많은 양을 학습하는 반면 한국어 데이터는 18,456개 밖에 학습을

하지 않기 때문이다. 따라서 제안 모델이 데이터가 증가 될수록 성능이 올라 갈 수 있는 가능성을 보기 위해 학습 데이터의 개수에 따른 성능을 측정한다. [그림 5]는 학습 데이터의 개수 변화에 따른 성능의 차이를 나타낸다.

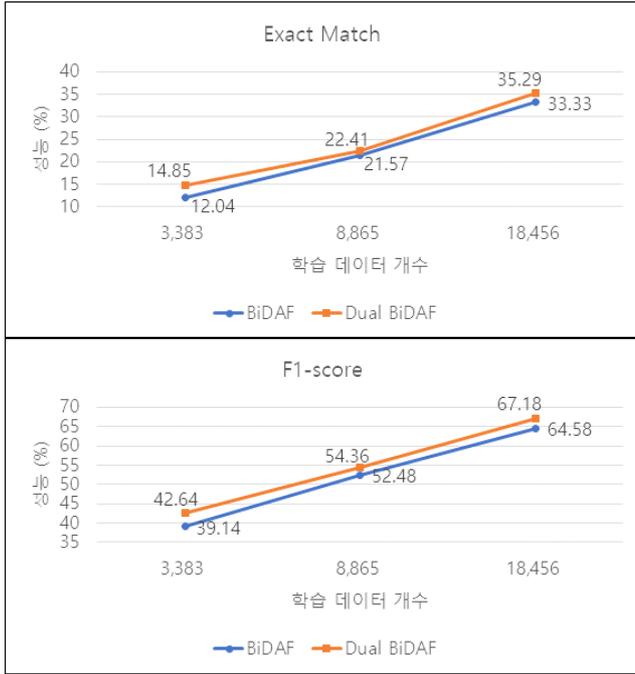


그림 5. 학습 데이터 개수에 따른 성능 변화

[그림 5]에서 학습 데이터의 개수가 증가됨에 따라 완전 일치율과 F1-score 성능이 모두 향상되는 것을 확인할 수 있다. 이는 심층 신경망 모델의 복잡함으로 인해 많은 양의 파라미터를 학습해야 하지만 데이터가 적을 경우 다양한 어휘 및 구문간의 관계를 파악하는 파라미터를 충분히 학습하지 못했기 때문이다. 즉, 많은 양의 학습 데이터를 확보할 경우 보다 높은 성능을 달성할 수 있을 것으로 예상된다.

5. 결론 및 향후 연구

본 논문에서는 어휘 이해만을 사용하는 기존의 Bi-Directional Attention Flow 모델에 구문 이해를 추가한 한국어 Dual Bi-Directional Attention Flow 모델을 제안하였다. 실험 결과 어휘 이해 모델만을 사용하는 Bi-Directional Attention Flow 모델에 품사 등장 정보, 의존 구문 정보를 주의집중 계층을 통해 반영해주는 구문 이해 모델을 추가하여 완전 일치율, F1-score가 각각 1.96%p, 2.6%p 향상되는 결과를 얻었다. 향후 연구로 완전 일치율의 성능을 향상시키기 위해 어절 간의 연결 오류를 형태소의 기능어를 통해 해결해 볼 예정이다.

감사의 글

본 연구는 LG전자 산학연구용역 과제의 지원을 받아

수행되었음.

참고문헌

- [1] P. Rajpurkar, J. Zhang, K. Lopyrev and P. Liang, SQuAD: 100,000+ Questions for Machine Comprehension of Text, *arXiv preprint arXiv:1606.05250*, 2016.
- [2] M. Seo, A. Kembhavi, A. Farhadi and H. Hajishirzi, Bidirectional attention flow for machine comprehension, *arXiv preprint arXiv:1611.01603*, 2016.
- [3] W. Wang, N. Yang, F. Wei, B. Chang and M. Zhou, Gated Self-Matching Networks for Reading Comprehension and Question Answering, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1. 2017.
- [4] Y. Cui, Z. Chen, S. Wei, S. Wang, T. Liu and G. Hu, Attention-over-Attention Neural Networks for Reading Comprehension, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1. 2017.
- [5] R. Liu, J. Hu, W. Wei, Z. Yang and E. Nyberg, Structural Embedding of Syntactic Trees for Machine Comprehension, *arXiv preprint arXiv:1703.00572*, 2017.
- [6] B. Pan, H. Li, Z. Zhao, B. Cao, D. Cai and X. He, MEMEN: Multi-layer Embedding with Memory Networks for Machine Comprehension, *arXiv preprint arXiv:1707.09098*, 2017.
- [7] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural computation* 9.8, pp. 1735-1780, 1997.
- [8] Y. Kim, Convolutional neural networks for sentence classification, *arXiv preprint arXiv:1408.5882*, 2014.
- [9] R. K. Srivastava, K. Greff and J. Schmidhuber, Highway networks, *arXiv preprint arXiv:1505.00387*, 2015.
- [10] S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6.02, pp. 107-116, 1998.
- [11] 최맹식, 김학수, “기계학습에 기반한 한국어 미등록 형태소 인식 및 품사 태깅”, *정보처리학회논문지 제 18-B권 1*, pp. 45-50, 2011.
- [12] 최맹식, 정석원, 김학수, “CRFs를 이용한 의존구조 분석 및 의존 관계명 부착”, *정보과학회논문지 : 소프트웨어 및 응용 41(4)*, pp. 302-308, 2014.