

BERT를 이용한 한국어 단어 의미 모호성 해소

윤준영^o, 신형진, 박정연, 이재성
충북대학교

{junyoung292, hjshin, parkjeongyeon, jasonlee}@cbnu.ac.kr

Korean Word Sense Disambiguation Using BERT

Jun Young Youn^o, Hyeong Jin Shin, Jeong Yeon Park, Jae Sung Lee
Chungbuk National University

요 약

단어의 의미 모호성을 해결하기 위한 연구는 오랫동안 지속되어 왔으며, 특히 최근에는 단어 벡터를 이용한 연구가 활발하게 이루어져왔다. 본 논문에서는 문맥 기반 단어 벡터인 BERT를 이용하여 한국어 단어 의미 모호성을 해소하기 위한 방법을 제안하고, 그 실험 결과를 기존의 한국어 단어 의미 모호성 연구 결과와 비교한다.

주제어: 단어 의미 모호성 해소, BERT, 단어 벡터

1. 서론

단어 의미 모호성 해소는 여러 가지 의미를 가진 단어가 문장내에 쓰일 경우, 그 문맥에 맞는 의미를 선택하는 작업이다. 이 작업은 사전(dictionary) 정보나 규칙을 이용하여 처리하는 방법, 말뭉치 정보를 이용하여 통계적으로 처리하는 방법 등 다양하게 연구되어 왔다 [1-4]. 특히 최근에 단어 벡터[5]를 이용하는 방법이 연구되고 있으며, 이는 어휘나 규칙을 이용하는 방법에 비해 부족한 어휘나 문맥 정보 문제를 어느 정도 해결하여 상대적으로 성능이 높았다[6-8].

단어 벡터를 이용한 의미 모호성 해소는 다른 방법과 유사하게 기본적으로 대상 단어의 문맥 정보를 이용하여 처리한다. 즉, 대상 단어의 좌우에 있는 문맥 단어들의 단어 벡터를 합하여 베이지안 규칙으로 대상 단어의 의미와 비교[7,8]하거나, 좌우 문맥 단어들을 SVM의 특성(feature)로 사용하여 의미를 분류하는 방법[6]을 사용한다.

단어 벡터는 최근 문맥을 고려한 단어 벡터로 ELMo[9], BERT[10] 등이 개발되어, 의미 모호성 해소를 포함한 여러 가지 자연어 처리의 성능 향상에 기여하였다. 특히 의미 부착 말뭉치를 지도학습 데이터로 사용하고, 도메인 의존적인 데이터의 부족 문제를 보완하기 위해 사전의 단어 정의(gloss) 정보를 사용하는 통합적 방법이 새로운 의미 모호성 해소 방법으로 제안되어 우수한 성능을 내고 있다[11, 12, 13]. 본 연구는 이런 최근의 연구 결과를 한국어에 적용해 보기 위한 초기 연구로, 한국어 단어 의미 모호성 해소 문제를 새로 개발한 한국어 BERT에 적용하여 보고, 그 결과를 기존의 한국어 연구 결과와 비교한다.

2. BERT 언어 모델

BERT 모델은 언어 모델의 하나로 주의(attention)을 기반으로 하는 트랜스포머를 사용한다[10]. 기학습된 언어 모델을 이용하면 이를 사용하는 자연어 처리 시스템의 성능을 향상시키는데, 정밀 조정(fine tuning)이나 특성(feature) 추가를 통해 이루어진다. BERT는 정밀 조정을 통해 다양한 응용에 활용되고 있다. 즉, 문장쌍 분류, 단일 문장 분류, 질의응답, 단일 문장 태깅 등의 작업에 활용되고 있다.

BERT는 근본적으로 언어에 독립적이며, 영어, 중국어, 한국어 등 다국어를 지원하고 있다. 그러나, 한국어와 같은 형태소 발달 언어는 띄어쓰기 단위로 단어를 처리할 경우 그 어휘 수가 많아, 보다 많은 학습데이터가 필요하다. 따라서, 비교적 적은 학습 말뭉치로 효과적인 BERT 벡터를 생성하기 위해서는 적절한 전처리(형태소 분석이나 스테밍 등)를 통해 입력 어휘를 어느 정도 정규화 할 필요가 있다.

전자통신연구소(ETRI)에서는 한국어 BERT(korBERT)를 개발하였으며, 이러한 한국어 특성을 고려하여, 어절 단위로 학습한 korBERT와 형태소 단위(품사 태그 포함)로 학습한 korBERT의 두 가지 버전을 제공하고 있다.

3. BERT를 이용한 한국어 단어 의미 모호성 해소

단어 의미 모호성 해소는 대상 단어가 가지는 다양한 의미중 특정 문맥에서의 의미를 선택하는 것이므로 본 논문에서는 이 문제를 단일 문장 태깅 문제로 정의하였다. 이를 위해 형태소 단위로 기학습한 korBERT를 사용하며, 기 학습된 BERT 모델을 단어 의미 모호성 해소 과제로 정밀 조정을 한다. 그림 1은 본 논문에서 제안한 단어 의미 모호성 해소 모델이며, 이 모델은 형태소 분석된 한국어 입력을 받아 그에 대응되는 의미태그(번호)를 출력한다. 의미 모호성이 없는 단일 의미를 가진 단

어는 0, 여러 의미를 가진 단어는 주어진 문맥에 맞는 단어 의미 번호(1이상의 양의 정수)를 출력한다.

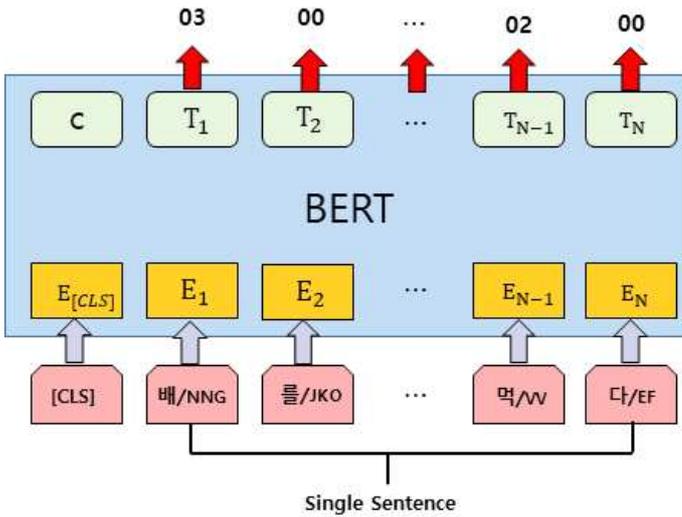


그림 1. korBERT기반 단어 의미 모호성 해소 모델

4. 실험 및 평가

실험은 세종 의미 부착 말뭉치[14]를 사용하였으며, 이 말뭉치의 통계정보는 표1과 같다.

표 1. 세종 의미 부착 말뭉치 통계

문장 수	823,650
어절 수	9,524,183
의미 부착 단어 수	3,892,113
동형이의어 어휘 수	17,078
동형이의어당 평균 의미 수	1.6

평가는 0 태그를 제외한 의미 번호의 일치 여부를 F1-score로 측정했으며, 5배수 교차검증으로 평가한 후 결과의 평균으로 계산하였다. 이런 평가는 모든 동형이의어 토큰에 대해 동일하게 평가했으므로 마이크로 평균에 해당된다. (동형이의어별로 평균을 낸 후, 모든 동형이의어의 평균을 다시 평균으로 계산한 매크로 평균과 구분됨). 이 결과를 기존의 단어 벡터를 이용한 모델 연구[8]의 결과와 비교한 것이 표 2이다.

표 2. 기존 연구 결과와 본 연구 결과의 비교(F1) (*는 문맥 단어들이 의미가 부착된 상태에서 평가된 것으로 실제 환경과는 차이가 있음)

단어 벡터 모델 [8]	단어(word) 공간	96.03
	복합(hybrid) 공간	96.18
	의미(sense) 공간*	96.53
BERT 모델 (본 연구)	korBERT	96.41

표2의 결과에서 보듯이 가장 성능이 높은 모델은 단어 벡터 모델 중 의미 공간 모델로 F1 96.53%를 보였다. 하지만, 이 모델은 좌우 문맥이 이미 정답 의미태그가 붙여진 것을 이용하여 대상 단어의 의미 모호성을 해소한 것이므로 실용적이지 않은 모델이다. 따라서, 좌우 문맥이 의미태그가 없는 일반 단어를 사용하는, 단어 공간 모델과 복합 공간 모델, korBERT모델을 비교하는 것이 실질적인 모델들에 대한 비교가 된다. 이 결과를 보면 koBERT 모델이 F1 96.41%로 다른 두 모델에 비해 가장 높은 것을 알 수 있다.

5. 결론

본 논문에서는 BERT 모델을 이용하여 한국어 단어 의미 모호성 해소 문제를 해결하기 위한 방법을 설명하였고, 실험 결과를 기존 연구와 비교하였다. 그 결과 제안한 방법이 실제적인 단어 의미 모호성 해소에서 기존의 한국어 의미 모호성 해소 방법보다 우수함을 보였다.

향후 연구로는 기존 연구에서 성능 향상을 보였던, 사전(dictionary) 정보를 활용하는 방법[13]을 한국어에도 적용해 보고, 기존의 영어 중심의 연구 결과와 비교해 볼 계획이다.

감사의 글

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2013-2-00131, 휴먼 지식증강 서비스를 위한 지능진화형 Wise QA 플랫폼 기술 개발)

참고문헌

- [1] M. Lesk, "Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone," In Proceedings of the 5th Annual International Conference on Systems Documentation, New York, USA, 1986, pp.24-26.
- [2] D. Yarowsky, "Word-sense disambiguation using statistical models of Roget's categories trained on large corpora," In Proceedings of the 14th Conference on Computational Linguistics-Volume 2, Association for Computational Linguistics, Stroudsburg, PA, USA, 1992, pp. 454-460.
- [3] W. Gale, K. Church, and D. Yarowsky, "A method for disambiguating word senses in a large corpus," Computers and the Humanities, vol. 26, no.5-6, pp. 415.439, 1992.
- [4] D. McCarthy, "Word sense disambiguation: An overview," Language and Linguistics compass, vol. 3, no. 2, pp. 537-558, 2009.
- [5] Word2Vec web page. 2017. <https://code.google.com/p/word2vec/>
- [6] I. Iacobacci, M. T. Pilehvar, and R. Navigli,

- "Embeddings for word sense disambiguation: An evaluation study," In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016.
- [7] M. Y. Kang, B. Kim and J. S. Lee, "Word sense disambiguation using Embedded Word Space," Journal of Computing Science and Engineering, vol. 11, no. 1, pp. 32-38, 2017.
- [8] M. Y. Kang, T. H. Min, J. S. Lee, "Sense Space for Word Sense Disambiguation," KIISE, The First International Workshop on Dialog Systems (BigComp 2018 workshop), 2018.01.
- [9] M. E. Peters et al., "Deep contextualized word representations." arXiv preprint arXiv:1802.05365 (2018).
- [10] J. Devlin et al., "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805 2018.
- [11] F. Luo, T. Liu, Z. He, Q. Xia, Z. Sui, and B. Chang, "Leveraging gloss knowledge in neural word sense disambiguation by hierarchical co-attention," In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 1402-1411.
- [12] F. Luo, T. Liu, Q. Xia, B. Chang, and Z. Sui. "Incorporating glosses into neural word sense disambiguation," arXiv preprint arXiv:1805.08028.
- [13] L. Huang, C. Sun, X. Qiu, X. Huang, "GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge," 2019. arXiv:1908.07245
- [14] 국립국어원, "21세기 세종프로젝트 최종성과물, 수정판", 2011.