

Rectifier Network 기반 학습된 심볼릭 제약을 반영한 뉴럴 자연언어처리

홍승연⁰¹, 나승훈², 신중훈³, 김영길⁴

¹²전북대학교, ³⁴한국전자통신연구원

hongsy034@hanmail.net, nash@jbnu.ac.kr, jhshin82@etri.re.kr, kimyk@etri.re.kr

Learning Symbolic Constraints Using Rectifier Networks for Neural Natural Language Processing

Seung-Yean Hong⁰¹, Seung-Hoon Na², Jong-Hoon Shin³, Young-Kil Kim⁴

¹²Jeonbuk National University, ³⁴ETRI

요약

자연언어처리 문제에서 딥러닝 모델이 좋은 성능을 보이고 있고 딥러닝 결과는 구조화된 결과를 내놓는 경우가 많다. 딥러닝 모델 결과가 구조적인 형태를 가지는 경우 후처리 통해 특정 구조에 맞는 제약을 가해주는 경우가 일반적이다. 본 논문에서는 이러한 제약을 규칙에 기반하지 않고 직접 학습을 통해 얻고자 하였다.

주제어: 제약, 규칙, ReLUs

1. 서론

최근 자연언어처리 문제 해결을 위해 많은 사람들은 딥러닝 모델을 이용하여 문제를 해결하고자 한다. 자연언어처리에서 트리, 그래프와 같이 특정한 구조를 가지는 결과를 내놓는 경우가 많다. 딥러닝 결과가 특정 구조를 만족하지 않는 경우 특정 구조를 만족하도록 규칙에 기반하여 제약을 가해준다. 본 논문에서는 규칙에 기반하여 구조에 제약을 가하지 않고 제약을 직접 학습하여 학습된 결과를 사용할 수 있도록 실험을 진행하였다.

본 논문에서는 [1]의 선행 연구를 이용하여 의미역 결정 문제에 제약 학습을 하였고 유의미한 성과를 얻었다.

2. 관련 연구

딥러닝 이전의 자연언어처리 문제들은 규칙에 기반하여 처리되어 생성된 결과물이 제약을 벗어나는 경우가 없었지만 딥러닝 모델을 사용하면 제약을 벗어나는 결과가 발생하여 후 처리를 통해 결과물을 처리해주어야 하는 문제가 발생하였다. 이러한 문제 해결을 위해 [2-3] 논문은 딥러닝 모델 내부에서 제약을 학습하여 따로 후처리가 하지 않도록 하는 모델을 제안하였다. GBI(gradient based inference) 모델[2]은 제약을 만족하는 경우에 그래디언트를 0으로 하여 업데이트가 되지 않도록 학습을 진행하는 모델이다. Distillation 모델[3]은 Distillation 방법을 통해 제약 정보를 전달하여 제약을 학습하는 모델이다. [4]의 논문은 [2-3]을 이용하여 다양한 한국어 자연언어처리 문제에 제약을 정의하고 적용한 논문이다. 학습 단계에서는 제약을 이용하여 학습을 진행하고 이때 제약은 규칙에 기반한 제약을 사용한다.

[1]는 본 연구에서 제약 학습을 위해 사용하는 모델로 제약 학습을 위해 Rectifier Network를 정의하고 제약을

직접 학습한 연구이다. 본 연구에서는 Rectifier Network를 직접 구현하고 학습을 진행하여 기존 [4]의 연구에서 규칙을 이용하여 제약을 가하는 부분을 Rectifier Network 모델의 결과를 이용하도록 수정하여 실험을 진행하였다.

BERT[5]는 대용량의 말뭉치를 이용하여 학습한 Transformer[6] 기반 언어 모델로 최근에 다양한 Task에 적용되어 state-of-the-art를 갱신하고 있다. BERT에 기반 하여 많은 논문이 나오고 있고 그 중 Roberta[7]는 기존의 BERT에서 문장 예측을 제거하고 Masking을 Dynamic하게 하는 확장된 BERT이다. 본 논문에서 사용한 RoBERTa 모델은 학습시키기 위해 약 15GB 위키피디아 코퍼스를 사용하였고 입력은 형태소-태그 단위를 입력으로 하고 단어장에 없는 경우는 BPE로 토큰화하여 처리한 모델이다. 본 논문에서 사용한 의미역 결정 모델은 선행 학습된 RoBERTa를 통해 fine-tuning한 결과이다.

의미역 결정은 질의응답, 기계번역, 정보 추출 등 다양한 자연언어처리 응용에서 성능 향상을 위한 자질로 사용되는 중요한 연구로 활발한 연구가 이루어지고 있다[8-10].

3. Rectifier Network을 이용한 심볼릭 제약 학습 및 뉴럴 자연언어처리

본 연구에서는 제약을 만족하는지를 판단하도록 학습을 진행하고자 한다. α 는 학습 가능한 파라미터, $\phi(x, y)$ 는 뉴럴 네트워크의 표상을 나타낸다.

$$y^* = \max_{y \in y_x} \alpha \cdot \phi(x, y) \quad (1)$$

제약이 K개가 존재할 때 K개의 제약을 모두 만족하는 최적의 결과 y^* 를 찾을 수 있도록 학습을 진행한다. 이를 위해 [1]에서는 Rectifier Networks를 제안하였고 본 논문에서는 의미역 결정 태스크에서 제약을 정의한

후 Rectifier Networks를 적용하였다.

3.1 Rectifier Networks

제약 c 가 정의되어 있고 입력 주어졌을 때 제약을 만족하는 결과를 얻기 위해 Boolean variable z 를 정의한다. z 는 수식 (2)와 같이 정의된다. $\psi(x, y)$ 은 입력을 나타낸다. 입력이 주어졌을 때 제약 개수 K 만큼 ReLUs hidden layer를 적용하여 제약을 모두 만족하는지를 결정한다. $c_k(x, y)$ 는 k 번째 제약이 만족할 경우 1이고 그렇지 않을 경우는 -1이 된다. 이를 통해 z 가 1이 되기 위해서는 모든 제약을 만족하여야 한다. (2)의 수식은 (3)과 같은 결과를 도출하기 위해 ReLUs를 적용하여 재정의한 수식이다.

$$z = \text{sgn}\left(1 - \sum_{k=1}^K R(w_k \psi(x, y) + b_k)\right) \quad (2)$$

$$z = c_1(x, y) \wedge \dots \wedge c_k(x, y) \quad (3)$$

수식 (4)와 같이 일차 부등식 형태로 정의되고 0보다 큰 경우는 모든 제약을 만족하는 것을 나타낸다. 이를 통해 적은 파라미터로 다양한 제약을 학습할 수 있다.

$$1 - \sum_{k=1}^K R(w_k \psi(x, y) + b_k) \geq 0 \quad (4)$$

3.2 의미역 결정 제약 학습

의미역 결정 태스크의 제약 학습을 위해 제약을 설정하였다. 서술어와 논항들의 관계가 기술된 격틀 사전을 사용하여 격틀 사전에 서술어와 논항 관계가 정의되어 있지 않으면 제약을 위반하는 것으로 정의하였다. 그림 1과 같이 격틀 사전에 없는 ARG2의 경우는 제약을 위반한 경우이다. 격틀 사전에는 부가 의미역 정보는 가지고 있지 않기 때문에 부가 의미역은 제외하였다.

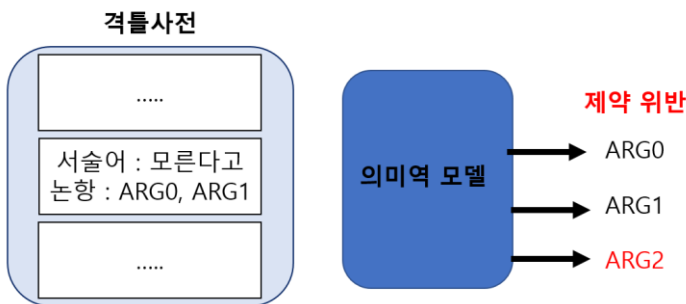


그림 1. 의미역 결정 제약

학습을 위해 Positive example과 Negative example을 추출하였다. Positive example은 아래 그림2와 같이 제약을 만족하는 술어와 논항 관계를 Positive로

추출하였고 만족하지 않는 관계를 Negative로 추출하여 제약 학습을 위한 데이터로 사용하였다. 학습은 추출한 Positive example과 Negative example을 통해 Positive는 z 가 0보다 큰 값을 가지도록 학습을 진행하였다.

술어	논항	
모른다고	ARG0	POS
모른다고	ARG1	POS
모른다고	ARG2	NEG
....

그림 2. Positive example/Negative example

수식 5는 제약이 학습된 경우 예시를 보여준다. 1차 부등식 형태로 나타낼 수 있고 서술어 “모른다고”, ARG0, ARG1, ARG2만 정의 되어 있을 때의 예시이다. x_1 은 “모른다고”, x_2 은 ARG0, x_3 은 ARG1, x_4 은 ARG2이다.

$$3.53x_1 - 1.90x_2 - 2.84x_3 - 4.12x_4 + 0.32 \geq 0 \quad (5)$$

$3.53 - 1.90 + 0.32$ 는 “모른다고”, ARG0가 나타났을 때 0보다 큰 값을 보여 제약을 만족한다. 하지만 $3.53 - 4.12 + 0.32$ 는 0보다 작은 값을 가져 제약을 만족하지 않는 것을 알 수 있다.

4. 실험

실험을 위해 Korean Propbank[11]의 Newswire 말뭉치만을 사용하여 학습 데이터를 추출하였다. 19602 문장은 학습 데이터, 1152문장은 개발 데이터, 2305문장은 평가 데이터로 사용하였다. 제약 학습을 위해 학습 데이터와 개발 데이터에서 Positive example 6024개, Negative example 8048개를 추출하여 학습을 진행하였다. 제약 학습은 의미역 결정 실험과 달리 긍정/부정만을 판단하기 때문에 성능 평가는 ACC로 측정하였다. 긍정은 제약을 만족하는 것을 나타내고 부정은 제약을 만족하지 않는 것을 나타낸다.

표 1. 제약 학습 결과

모델	ACC
Rectifier Network	98%

98%의 성능을 보여 대부분의 데이터를 올바르게 금/부정 분류함을 확인하였다.

아래 실험은 모델 내부에서 제약을 학습 하여 따로 후 처리를 하지 않고 제약을 학습하는 두 모델의 성능을 보여준다. 기존의 [4]의 연구에서는 학습 단계에서 제약을 가할 때 규칙을 이용하고 있다. 규칙을 이용한 모델은 Rule base라고 표시하였고 학습한 제약을 이용한 모델은 Learning constraint로 표시하였다. 서술어 인식

및 분류, 논항 인식 및 분류 중 논항 인식 및 분류의 결과로 평가 지표는 F1-score로 측정하였다.

표 2. 의미역 결정 실험 결과

모델	Dev F1	Test F1
Rule base		
RoBERTa GBI[4]	86.99%	85.75%
RoBERTa Distillation[4]	86.47%	85.47%
Learning constraint		
RoBERTa GBI	87.02%	85.69%
RoBERTa Distillation	86.66%	85.34%

Dev에서는 제약을 학습한 결과를 사용한 모델의 성능이 높게 나타났지만 Test에서는 규칙을 사용한 모델의 성능이 높음을 확인하였다.

감사의 글

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (R7119-16-1001, 지식증강형 실시간 동시통역 원천기술 개발)

5. 결론

본 연구에서는 제약 학습을 위해 Rectifier Network를 사용하여 실험을 진행하였고 규칙을 통해 얻는 성능과 비슷한 성능을 보여 제약 학습 모델이 잘 작동함을 확인하였다. 향후에 의미역 결정뿐 아니라 제약이 적용되는 다양한 자연어처리 문제에 적용할 예정이다.

참고문헌

- [1] Pan, Xingyuan, Maitrey Mehta, and Vivek Srikumar. "Learning Constraints for Structured Prediction Using Rectifier Networks." arXiv preprint arXiv:2006.01209, 2020.
- [2] Z. Hu, X. Ma, Z. Liu, E. Hovy, E. Xing, Harnessing Deep Neural Networks with Logic Rules, ACL, 2019
- [3] Jay Yoon Lee, Sanket Vaibhav Mehta, Michael Wick, Jean-Baptiste Tristan, Jaime Carbonell, Gradient-based Inference for Networks with Output Constraints, AAAI, 2019
- [4] 홍승연, 나승훈, 신종훈, 김영길, "심볼릭 제약을 반영한 뉴럴 자연어처리", KCC, 2020.
- [5] J. Devlin, M. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", arXiv:1810.04805, 2018.
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need", NIPS, 2017.
- [7] Y. Liu, M. Ott, N. Goyal, J.i Du, Mandar S. Joshi, D. Chen, O. Levy, M. Lewis, Luke S. Zettlemoyer, and V.

Stoyanov, "Roberta: A robustly optimized bert pretraining approach", ArXiv, abs/1907.11692, 2019.

[8] 박광현, 나승훈, 신종훈, 김영길, "BERT를 이용한 한국어 자연어처리: 개체명 인식, 감성분석, 의존 파싱, 의미역 결정", KCC, 2019.

[9] 홍승연, 나승훈, 신종훈, 김영길, "RoBERTa와 스택-포인터 네트워크를 이용한 한국어 의미역 결정", KSC, 2019.

[10] 민진우, 나승훈, 신종훈, 김영길, "Capsule Networks를 이용한 한국어 의미역 결정", KCC, 2020.

[11] Palmer, M. Ryu, S. Choi, j. Yoon, S. Jeon, "Korean propbank", LDC Catalog No.:LDC2006T03 ISBN, 1-58563, 2006.