강화학습 기반의 제로샷 텍스트 분류

장송밍 [1], 조인휘 [2]
[1,2] 한양대학교 컴퓨터소프트웨어학과

kagami@hanyang.ac.kr,iwjoe@hanyang.ac.kr

Zero-shot Text Classification based on Reinforced Learning

Zhang Songming[1], Inwhee Joe[2]
[1,2] Dept. of Computer Science, Hanyang University

요 약

전통적인 텍스트 분류 방법은 상당량의 라벨링된 데이터와 미리 정의된 클래스가 필요해서 그 적용성과 확장성이 제한된다. 그래서 이런 한계를 극복하기 위해 제로샷 러닝(Zero-shot Learning)이 등장했다. 텍스트 분류 분야에서 제로샷 텍스트 분류는 모델이 대상 클래스의 샘플을 미리 접하지 않고도 인스턴스를 분류할 수 있도록 하는 중요한 주제이다. 이 문제를 해결하기 위해 정책 네트워크를 활용한 심층 강화 학습(DRL) 기반 접근법을 제안한다. 이러한 방법을 통해 모델이 새로운 의미 공간에 효과적으로 적응하면서, 다른 모델들과 비교하여 제로샷 텍스트 분류의 정확도를 향상시킬 수 있었다. XLM-R 과 비교하면 최대 15.9%의 정확도 향상이 나타났다.

1. 서론

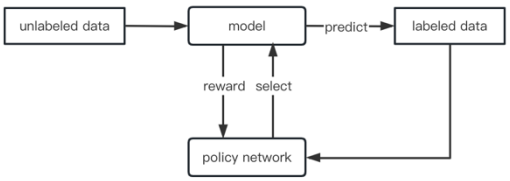
제로샷 텍스트 분류에서의 주요 도전 과제는 라벨이 없는 데이터를 보지 못한 클래스로 분류하는 것이다. 예를 들어, 모델이 훈련 중에 "행복" 클래스를 접했다면, 새로운 라벨이 없는 데이터를 "행복"으로 올바르게 분류할 수 있어야 한다. 그러나 "질투"와 같은 클래스를 이전에 보지 못했다면, 라벨이 없는 데이터를 "질투"로 정확하게 분류하는 것은 어려워진다. 대형 언어 모델은 이 문제를 해결하는데 실행 가능성을 제공한다. 대형 언어 모델은 방대한 양의 텍스트로 훈련되어 풍부한 의미 정보를 포착할 수 있다.

Yin 등 (2019)[1]은 텍스트 분류 작업을 텍스트 함의(text entailment) 문제로 형식화하는 혁신적인 접근 방식을 소개했다. 이 형식화는 자연어 추론(NLI)으로 훈련된 모델을 이용하여 이전에 보지 못한 다양한 하위 작업들에 대한 제로샷 텍스트 분류기로 활용할 수 있도록 한다. 예를 들어, 문장 "귀여운 개가 있다"가 "이 예시는 동물에 관한 것이다"를 함의하는지 여부를 판단함으로써 텍스트를 "동물"로 분류할 수 있다.

그러나, 언어 모델을 직접 사용하여 예측하는 것은 전통적인 지도 학습 방법에 비해 정확도가 상당히 낮다. 정확도를 향상시키기 위해 우리는 반지도 학습에서 널리 사용되는 의사 라벨(pseudo-label)[2] 기술을 활용하고 DRL 을 사용하여 모델을 훈련시켜, 정확도를 상당히 향상시켰다.

2. 모델

2.1 개요



(그림 1) 강화학습 기반의 제로샷 텍스트 분류 모델 개요.

우리는 텍스트 함의를 위해 BERT[3] 모델을 활용하였으며, 의사 라벨링 기술을 활용하여 라벨이 없는 데이터를 라벨이 있는 데이터로 변환하고 이를 다시 훈련 샘플로 통합하였다. 부정적인 예제의 수를 줄이기 위해 의사 라벨링된 데이터를 선택적으로 훈련 샘플로 포함하기 위해 DRL 을 적용하였다.

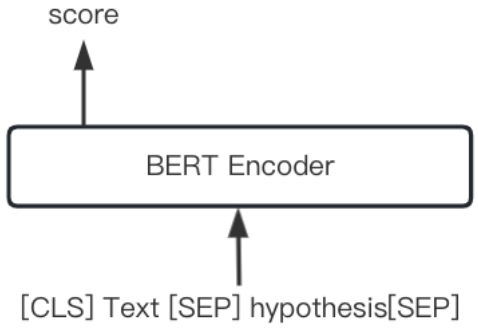
2.2 의사 라벨

의사 라벨링은 반지도 학습에서 자주 사용되는 방법으로, 라벨이 없는 데이터에 임시 라벨을 할당하는 것이다. 이 전략은 모델이 라벨이 없는 데이터에 대한 예측을 활용하여, 이를 훈련 과정에 통합할 수 있

는 의사 라벨로 취급한다.

2.3 BERT 를 텍스트 함의 모델로 사용하기

BERT 의 입력 시퀀스는 "[CLS] 텍스트 [SEP] 가설 [SEP]" 형식으로 포맷팅되며, [CLS]와 [SEP]는 BERT 의 특수 시작 및 구분자 토큰을 나타낸다.



(그림 2) BERT 의 입력.

$$S = sigmoid(W^T c \ + \ b) \tag{1}$$

$$Loss = \begin{cases} -log(S) & y' = y \\ -log(1-S) & y' \neq y \end{cases} \tag{2}$$

여기서 W 와 b 는 선형 레이어의 매개변수이며, c 는 [CLS]의 은닉 상태이다.

2.4 데이터 선택을 위한 강화 학습

잘못된 라벨이 부여된 데이터가 훈련 세트에 추가되면 모델의 정확도가 낮아질 수 있다. 따라서 우리는 DRL 을 활용하여 BERT 의 훈련 데이터로 고품질 데이터 하위 집합을 선택한다.

많은 DRL 알고리즘이 NLP 분야에 적용되었다. DQN, A3C, DDPG 등이 포함되며[4], 우리는 이러한 알고리즘들을 시도해 보았으며, 또한 정책 네트워크를 기반으로 한 새로운 알고리즘을 설계했다. 우리의 모듈에 대한 자세한 내용은 아래에 설명되어 있다.

2.4.1 상태

각각의 텍스트 x 에 대해, 예측 점수 y 를 얻는다. 가장 높은 점수를 가진 라벨이 의사 라벨로 선택된다. 시간 단계 t,에서 현재 상태 st 는 두 가지 구성 요소로 이루어진다: 예측 점수 p 와 들어오는 인스턴스의 표현 c (CLS 의 은닉 상태). 정책 네트워크는 p 와 c 를 입력으로 받고 선택할지 여부에 대한 확률을 출력한다.

2.4.2 액션

각 시간 단계마다, 에이전트는 현재 인스턴스를 선택할지 또는 거부할지를 결정한다. 시간 단계 t 에서, 만약 $a_t = 1$ 이면, 에이전트가 현재 인스턴스를 수용하고 훈련 세트에 추가한다는 것을 나타낸다. 그 반대로, $a_t = 0$ 이면 거부를 의미한다. 행동 값은 정책 네트워크 $P(a|s_t)$의 출력에서 샘플링된다.

2.4.3 보상

보상은 유효성 검사 세트의 텍스트 수반 모델에 의해 계산된다. 검증 세트는 두 부분으로 나뉜다: V_1 은 처음에 선택된 레이블이 지정된 가시 데이터로 구성되며, V_2 는 학습 과정에서 텍스트 연루 모델에 의해 추가되는 의사 레이블 데이터로 구성된다.

$$r = \frac{F^s - \mu^s}{\sigma^s} + \ \lambda \cdot \frac{F^u - \mu^u}{\sigma^u} \tag{3}$$

여기서 F^s 와 F^u 는 검증 세트 V_1 과 V_2 에서 평가한 macro-F1 점수를 나타낸다. σ 와 μ 는 각각 표준 편차와 평균을 나타낸다.

2.4.4 정책 네트워크와 최적화

정책 네트워크는 다층 퍼셉트론(MLP)이다. 정책 네트워크는 상태를 입력으로 받는다. 이 상태는 예측 신뢰도 s와 들어오는 인스턴스의 표현 c를 포함한다. 그리고 각 행동에 대한 확률을 생성한다.

$$z = ReLU(W_1^T c + W_2^T s \ + \ b_1) \tag{4}$$

$$P(a|s_t) = softmax(W_3^t z + b2) \tag{5}$$

여기서 W1, W2, W3, b1, b2 는 매개변수이며, P(a|st)는 정책 네트워크가 의사 라벨링된 데이터를 훈련 데이터로 선택하는 확률을 나타낸다.

목표는 최적의 데이터 선택 정책을 학습하여 기대 총 보상을 극대화하는 것이다. 이 목표는 다음과 같이 형식화될 수 있다:

$$J(\varphi) \ = \ E_{P\varphi(a|s)}[R(s,a)] \tag{6}$$

여기서 φ 는 정책 네트워크의 매개변수이다. 정책 그래디언트를 사용하여 φ 를 업데이트하기 위해, 우리는 상태-행동 가치 함수 R(s,a)를 활용하고 다음과 같은 최적화 과정을 적용한다:

$$\varphi \ \leftarrow \ \varphi \ + \ \eta \nabla_\varphi J \ (\varphi) \tag{7}$$

η 는 할인 학습률(discount learning rate)이다.

$$\nabla_\varphi J(\varphi) = \frac{r}{n} \sum_1^n \nabla_\varphi log(P(a|s)) \tag{8}$$

여기서 n 은 배치 크기(batch size)이고, r 은 보상(reward)이다. 정책 네트워크의 매개변수는 각 에포크

의 끝에서 업데이트된다.

3. 실험

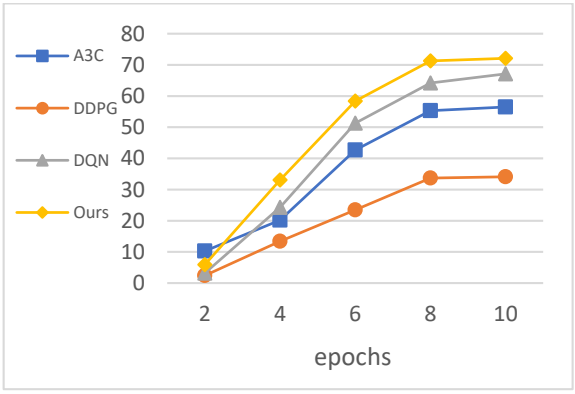
우리는 모델의 데이터 선택 구성 요소에 대해 다양한 강화 학습 알고리즘을 테스트했다. 또한, 우리는 다른 대형 언어 모델과 우리의 모델을 비교하여 제로샷 텍스트 분류 작업을 수행했다. 자세한 내용은 아래에 나와 있다.

3.1 데이터셋

우리는 총 세 개의 데이터셋을 실험하였다. 20 Newsgroups, Yahoo! News Dataset, 그리고 Sougou News Dataset 이다. 첫 두 개의 데이터셋은 영어로 이루어져 있고, Sougou News Dataset 은 중국어로 되어 있다.

3.2 데이터 선택을 위한 다양한 알고리즘의 성능

우리는 20 Newsgroups 데이터셋에서 다양한 알고리즘을 테스트하고 최종 분류 정확도를 사용하여 알고리즘 성능을 평가했다.



(그림 3) DRL 알고리즘의 성능.

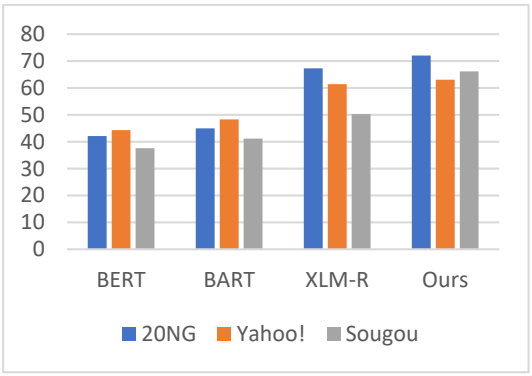
그림 3 에서 볼 수 있듯이, 우리의 접근 방식은 데이터 선택에서 가장 우수한 결과를 얻었다.

3.3 결과

또한 우리는 BART 와 XLM-R 을 포함한 다른 대형 언어 모델과 우리의 모델을 비교하였다.

BART[5]는 시퀀스-투-시퀀스 모델의 사전 훈련을 위한 노이즈 제거 오토인코더이다. 임의의 노이즈 함수로 텍스트를 손상시킨 다음, 원래 텍스트를 재구성하는 모델을 학습한다. 표준 Transformer 기반의 신경 기계 번역 아키텍처를 사용한다.

XLM-RoBERTa(XLM-R) [6]는 RoBERTa 의 다국어 버전이다. 이 모델은 100 개 언어의 CommonCrawl 데이터를 2.5TB 에 걸쳐 필터링하여 사전 훈련되었다.



(그림 4) 결과.

그림 4 에서 볼 수 있듯이, 우리의 모델은 다른 모델들과 비교하여 가장 우수한 결과를 달성했다. 특히, 사전 훈련 데이터가 제한된 중국어의 경우, 우리의 모델은 상당한 이점을 보였다.

4. 결론

본 논문에서는 의사 라벨링을 사용하고 DRL 을 데이터 선택에 적용하여 제로샷 텍스트 분류의 정확도를 향상시켰다. 그러나 비지도 학습과 비교하면 여전히 일부 라벨이 있는 데이터가 필요하다. 대부분의 데이터셋은 영어로 되어 있으며, 다른 언어의 고품질 데이터셋은 종종 부족하다. 향후 연구에서는 라벨이 있는 데이터에 대한 의존도를 줄이고, 라벨이 있는 데이터가 제한적인 상황에서 모델을 더 실용적으로 만드는 것을 목표로 한다.

참고문헌

[1] Yin, Wenpeng, Jamaal Hay, and Dan Roth. "Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach." arXiv preprint arXiv:1909.00161 (2019).

[2] Lee, Dong-Hyun. "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks." Workshop on challenges in representation learning, ICML. Vol. 3. No. 2. 2013.

[3] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[4] Uc-Cetina, Victor, et al. "Survey on reinforcement learning for language processing." Artificial Intelligence Review 56.2 (2023): 1543-1575.

[5] Lewis, Mike, et al. "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension." arXiv preprint arXiv:1910.13461 (2019).

[6] Conneau, Alexis, et al. "Unsupervised cross-lingual representation learning at scale." arXiv preprint arXiv:1911.02116 (2019).