

Detecting Contextomized Quotes in News Headlines by Contrastive Learning

Seonyeong Song, Hyeonho Song, Kunwoo Park, Jiyoung Han, Meeyoung Cha

1. Introduction

- 화자의 말을 그대로 복제하여 인용하는 것은 뉴스의 중립성, 사실성, 객관성을 강조함으로써 뉴스의 신뢰도를 높임
- 신뢰할 수 있는 기사에 비해 가짜 뉴스에서 직접 인용구를 더 많이 사용함. (Dalecki et al., 2009; Govaert et al., 2020)
- 우리 논문에서 연구할 부분이 Contextomy 를 탐지하는 것
 - Contextomy란?
 - 의미를 거두절미식으로 오역을 유도하는 인용구
 - 화자의 의도를 바꾸는 방식으로 문맥을 벗어난 단어를 인용하는 것

News headline quote	Body-text quotes	Label
"이대론 그리스처럼 파탄" (A debt crisis, like Greece, is on the horizon)	"건강할 때 재정을 지키지 못하면 그리스처럼 될 수도 있다" (If we do not maintain our fiscal health, we may end up like Greece) "강력한 '지출 구조조정'을 통해 허투루 쓰이는 예산을 아껴 필요한 곳에 투입해야 한다" (Wasted budgets should be reallocated to areas in need through the reconstruction of public expenditure)	Contextomized
"불필요한 모임 일절 자제" (Avoid unnecessary gatherings altogether)	"저도 백신을 맞고 해서 여름에 어디 여행이라도 한번 갈 계획을 했었는데..." (Since being vaccinated, I had planned to travel somewhere in the summer, but ...) "행사가 일단 다 취소됐고요..." (Services have been halted...) "어떤 행위는 금지하고 어떤 행위는 허용한다는 개념이 아니라 불필요한 모임과 약속, 외출을 일제 자제하고..." (It is not a matter of prohibiting or permitting specific activities, but of avoiding unnecessary gatherings, appointments, and going out altogether...)	Modified

1. Introduction

- 목표
 - 원래의 인용문에서 의미적 변화가 있는 단어들을 발췌하여 인용한 Contextomized quote를 탐지
 - headline 인용문이 본문 인용문과 의미적으로 일치하는지 여부를 분류
- 이를 해결하기 위해 Contrastive Learning 기법 사용

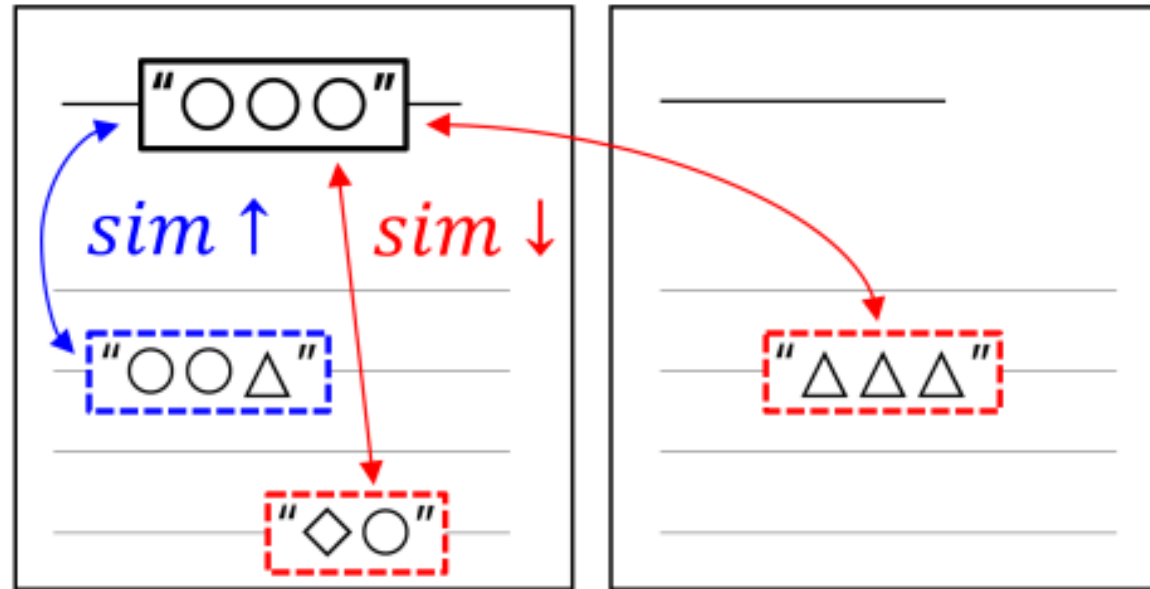
1. Introduction

- Contrastive learning(대조학습)이란?
 - 유사도가 높은 데이터 샘플쌍을 positive pair, 유사도가 낮은 데이터 샘플쌍을 negative pair
 - Positive pair끼리는 similarity를 높이도록 학습, negative pair끼리는 similarity를 낮추도록 학습
 - 즉, positive pair끼리는 가깝게, negative pair끼리는 멀게 학습시키는 기법
 - Vision쪽에서 최근 높은 성능을 보여주었고 최근 NLP에서 SimCSE가 좋은 성능을 보여줌
- SimCSE (Simple Contrastive Learning of Sentence Embeddings)
 - 이 논문은 positive sample을 생성하는 방식을 새롭게 제안함
 - Unsupervised : 하나의 sentence에 dropout을 적용한 문장을 positive sample로 사용

$$-\log \frac{e^{\text{sim}(\mathbf{h}_i, \tilde{\mathbf{h}}_i) / \tau}}{\sum_{j=1}^N e^{\text{sim}(\mathbf{h}_i, \tilde{\mathbf{h}}_j) / \tau}},$$

2. Data Methodology

- 이 연구에서 제안한 방법 QuoteCSE
 - Contribution: Positive pair와 negative pair를 정의하는 방법



2. Data Methodology

- 데이터 수집
 - 네이버에 8개 언론사(조선일보, 중앙일보, 동아일보, 한겨레, 연합뉴스, 경향신문, 머니투데이, 매일경제)에서 발행된 뉴스 약 40만 건을 크롤링
 - 정규표현식을 사용하여 직접 인용구 추출
 - 데이터 기간: 2005년 ~ 2019년
- 데이터 labeling
 - 언론학 전공 학부생 2명 고용하여 Headline의 직접 인용문이 Contextomy인지 아닌지를 분류하도록 훈련
 - Mass communication의 직원이 가이드라인 작성
 - 2주에 걸쳐 8번 반복 훈련 후 2000개 기사에 대해 Krippendorff's Alpha=0.93 달성
- 최종 labeled data
 - 본문에 인용문이 없거나, Headline 인용문과 완전히 동일한 경우 제외
 - 1,609개의 데이터 생성 (Contextomized quote: 823/ Modified quote: 786)

2. Data Methodology

- QuoteCSE
 - 기사 헤드라인에 존재하는 인용구 T
 - 본문에 존재하는 인용구 B
- Positive pair: T_1 과 $B_{1,i}$ (B_1 중 T_1 과 가장 가까운 인용구)
- Negative pair: T_1 과 $B_{1,j}$ (B_1 중 T_1 과 가장 가까운 인용구를 제외한 나머지 인용구 중 랜덤하게 하나 선택) -> hard negative
- 구성 방법
 - Positive pair와 negative pair를 결정하기 위해 SentenceBERT를 사용하였음
 - Positive pair 간의 유사도가 0.75 이상인 86,275개의 data만 사용
 - 훈련, 검증, 테스트를 8:1:1로 나눔

$$-\log \frac{e^{\text{sim}(\mathbf{h}_i, \mathbf{h}_i^+)/\tau}}{\sum_{j=1}^N \{e^{\text{sim}(\mathbf{h}_i, \mathbf{h}_j^+)/\tau} + e^{\text{sim}(\mathbf{h}_i, \mathbf{h}_j^-)/\tau}\}},$$

3. Evaluation

- 3-1) Evaluation on labeled data
 - 총 1609개의 data (contextomized: 823, modified: 786)
 - 64-dimensional hidden layer를 가진 classifier를 구축하여 평가
 - classifier의 input: $u, v, |u - v|, u * v$
 - u : 기사 헤드라인의 인용구의 embedding
 - v : 본문 텍스트 중 기사 헤드라인의 인용구와 가장 가까운 인용구의 embedding
 - QuoteCSE가 가장 높은 성능을 달성

	F1	AUC
BERT	0.665 ± 0.006	0.661 ± 0.006
SBERT	0.652 ± 0.008	0.621 ± 0.014
SimCSE-Quote	0.692 ± 0.006	0.689 ± 0.006
SimCSE-NLI	0.638 ± 0.007	0.639 ± 0.007
BERT fine-tune	0.730 ± 0.006	0.730 ± 0.006
QuoteCSE	0.763 ± 0.007	0.760 ± 0.008

3. Evaluation

- 3-2) Embedding quality
- Alignment
 - Positive embedding들이 얼마나 가까운지?
 - 작을수록 좋음
- Uniformity
 - embedding들이 얼마나 균등하게 분포하는지?
 - 작을수록 좋음
- QuoteCSE가 모든 면에서 최상의 결과를 보여줌

	Alignment (title-title)	Alignment (title-body)	Uniformity
BERT	0.638	0.738	-0.711
SBERT	0.227	0.329	-1.356
SimCSE-Quote	0.503	0.380	-2.176
SimCSE-NLI	0.319	0.260	-3.257
QuoteCSE	0.150	0.194	-3.562

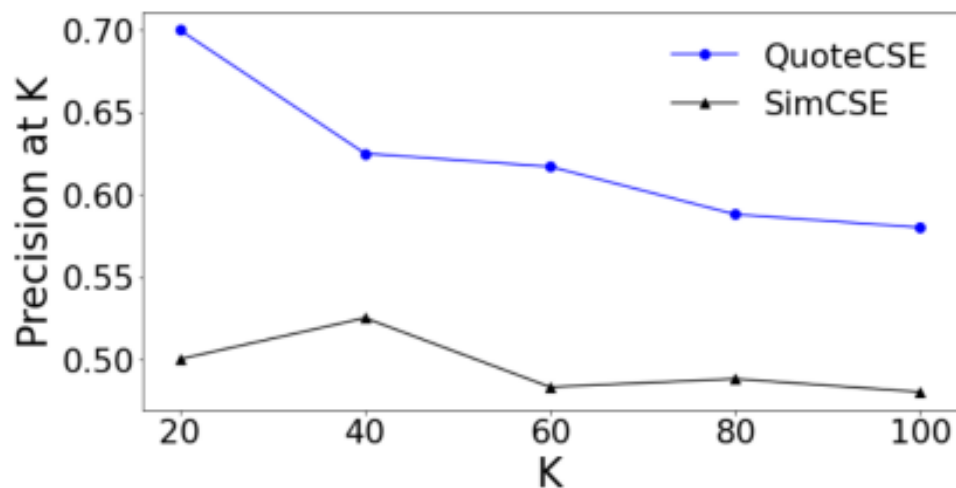
3. Evaluation

- 3-3) Ablation experiment
 - Negative sample과 positive sample의 영향력에 대해 실험
 - Negative sample이 없을 때와 positive sample이 없을 때 모두 감소
 - 이는 negative sample과 positive sample이 모두 중요함을 의미

Positive	Hard Negative	F1	AUC
QuoteCSE	QuoteCSE	0.763 \pm 0.007	0.760 \pm 0.008
SimCSE	QuoteCSE	0.690 \pm 0.005	0.682 \pm 0.005
QuoteCSE	—	0.674 \pm 0.007	0.674 \pm 0.006

3. Evaluation

- 3-4) Evaluation in the wild
 - 2021년 7,8월에 발행된 10,055개의 뉴스 기사로 평가
 - SimCSE, QuoteCSE를 사용하여 Contextomized 예측 점수 상위 100개의 뉴스 기사를 각각 뽑아 수동으로 평가
 - Contextomized라고 예상된 상위 K개 사례 중 얼마나 많은 경우가 정확하게 탐지되었는지를 보여줌
 - QuoteCSE는 상위 20개에 대해 0.7의 높은 precision을 보여줌
 - SimCSE는 신뢰도가 높은 경우에도 0.55 이하의 precision을 보여줌
- QuoteCSE가 실제에서 Contextomized 인용문을 잘 찾을 수 있다는 가능성을 보여줌



4. Limiation

- 실험한 서버의 환경상 사전학습 시 batch size 16이 최대였음
- Batch size가 클수록 모델이 더 많은 negative sample를 학습할 수 있기 때문에 batch size가 클 수록 성능 향상 가능성이 높음
- 대안책으로 MoCo를 실험
 - MoCo: Momentum Contrast for Unsupervised Visual Representation Learning
 - MoCo는 memory bank 방식을 사용하여 더 많은 양의 negative sample을 사용할 수 있음

→ MoCo 방법으로 실험해봤을 때 QuoteCSE가 SimCSE보다 잘 나왔지만, batch size를 사용하는 기존 방법에 비해 성능 개선은 없었음

	F1	AUC
BERT	0.665 ± 0.006	0.661 ± 0.006
SBERT	0.652 ± 0.008	0.621 ± 0.014
SimCSE-Quote	0.692 ± 0.006	0.689 ± 0.006
SimCSE-NLI	0.638 ± 0.007	0.639 ± 0.007
BERT fine-tune	0.730 ± 0.006	0.730 ± 0.006
QuoteCSE	0.763 ± 0.007	0.760 ± 0.008

	F1	AUC
MoCo: SimCSE	0.639 ± 0.009	0.652 ± 0.007
MoCo: QuoteCSE	0.748 ± 0.005	0.747 ± 0.005

5. Conclusion

- 화자의 의도를 왜곡하고 문맥을 벗어난 인용구는 저널리즘 윤리에 위배되고 여론을 오도할 수 있음
- 이런 인용구를 Contextomized 인용구라 정의하였고, Contrastive learning을 사용하여 headline 에서 Contextomized 인용구를 탐지하는 모델을 만들어 QuoteCSE라 정의했음
- 이 논문은 Contrastive learning에서 positive 뿐만 아니라 hard negative 정의의 중요성을 보였음
- 결론적으로 QuoteCSE는 SimCSE의 성능보다 더 높은 성능을 보였음

Thank You

감사합니다.