

Transitive Consistency Constrained Learning for Entity-to-Entity Stance Detection

Haoyang Wen, Eduard Hovy, Alexander Hauptmann

† Language Technologies Institute, Carnegie Mellon University

‡ School of Computing and Information Systems, The University of Melbourne
{hwen3, hovy, alex}@cs.cmu.edu

Venue: ACL 2024

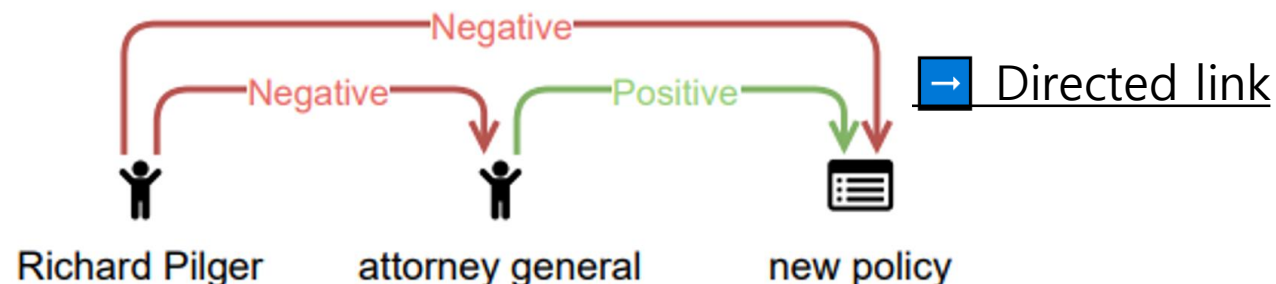
발제자 : 이다현(hyundai@soongsil.ac.kr)

HUMANE Lab

2024-08-28

Entity-to-Entity Stance Detection?

- 엔티티 쌍 간의 입장을 탐지하는 작업
- 기존의 입장 탐지: Target, Polarity가 중요
- 엔티티-투-엔티티 입장 탐지: Target, Source, Polarity가 존재



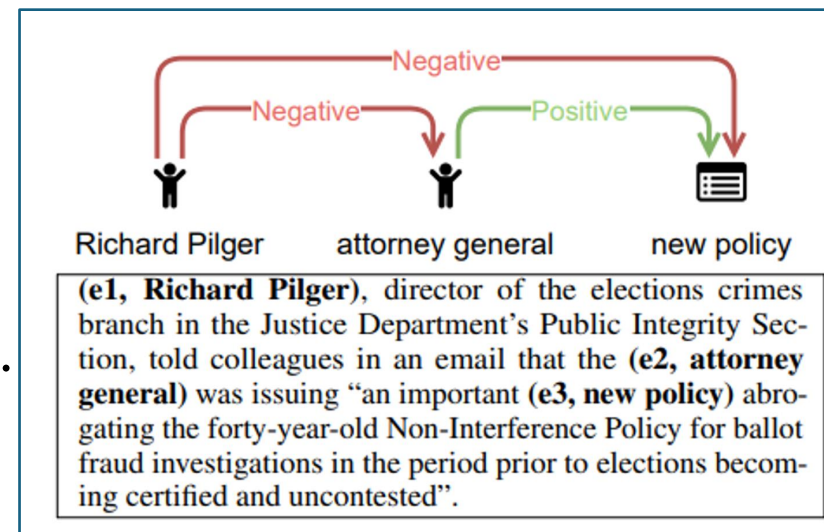
(e1, **Richard Pilger**), director of the elections crimes branch in the Justice Department's Public Integrity Section, told colleagues in an email that the (e2, **attorney general**) was issuing "an important (e3, **new policy**) abrogating the forty-year-old Non-Interference Policy for ballot fraud investigations in the period prior to elections becoming certified and uncontested".

Previous work and hypothesis

- 과거 연구는 하나의 엔티티 쌍만을 고려함
(Park et al., 2021; Zhang et al., 2022)
- 그러나, 엔티티 쌍 간의 입장은 서로 연관될 수 있음

- 가설

- "전이적 연관성"이 정치 뉴스에서는 흔하며,
더 나은 모델을 훈련시키기 위해 활용될 수 있다.




Transitive Consistency

Entity-to-Entity Stance Detection Frameworks

- 엔티티 쌍 간의 입장을 탐지하는 작업
- directed link를 통하여 stance의 source와 target까지 고려
- 이 연구에서는 이를 위한 두 개의 프레임워크를 제안
 1. Classification-based Framework
 2. Generation-based Framework

Classification-based Framework

- Input: sentence  entity-pair representation
- Output: Classification result
- 이 패러다임은 여러 relation extraction task에서 효과성을 입증
- (Eberts and Ulges, 2020; Wang and Lu, 2020; Wang et al.,2020; Wen and Ji, 2021)

Classification-based Framework

- 모델은 두 문장을 나타내는 토큰 시퀀스를 입력으로 받음
 - 문장에서 엔티티의 위치도 제공. 엔티티는 e_1, e_2 로 나타내며
각 엔티티의 위치는 p_1, p_2 로 표현
- 사전학습된 언어모델을 사용해 문장의 전체 입력에 대한 문맥화된 표현을 얻음 $\mathbf{H} = \text{PLM}(x)$
- 두 엔티티의 관계를 표현하기 위해 모델은 엔티티 위치 p_1, p_2 에서 문맥화된 표현을 연결 $\mathbf{c} = [\mathbf{h}_{p_1}; \mathbf{h}_{p_2}]$

Classification-based Framework

- 두 개의 피드 포워드 신경망 layer,
tanh 활성화 함수,
softmax 레이어를 거쳐 분류

$$p(s \mid e_1, e_2) = \text{softmax}(\mathbf{a}),$$
$$\mathbf{a} = \text{FFN}_2(\tanh(\text{FFN}_1(\mathbf{c})))$$

- 이 단계에서 두 엔티티간 입장을 예측

Classification-based Framework

- 방향과 극성을 모두 포함하는 라벨 사용
ex) entity 1이 entity 2에 긍정
- 엔티티 입력 순서를 고려하여 입장의 source와 target을 결정

Classification-based Framework

- 교차 엔트로피 손실 함수를 사용해 훈련

$$\mathcal{L}_s = - \sum \mathbb{I}_{s(e_1, e_2) = s_i} \log p(s = s_i \mid e_1, e_2)$$

Generation-based Framework

- 언어 모델이 입력된 문장에서 엔티티 간 입장을 자동 회귀적으로 생성하는 방식
- 모델: decoder-only

$$p(\mathbf{o} \mid \mathbf{x}, e_1, e_2) = \prod_{i=1}^{|\mathbf{o}|} p(o_i \mid \mathbf{o}_{<i}; \mathbf{T}(\mathbf{x}, e_1, e_2))$$

- Input
 - *Analyze the entity-entity stance in the following text: \n 문장 \n Entity 1: e1 \n Entity 2: e2 \n Stance:*
- Output
 - Entity 1 to Entity 2 / positive, negative, and neutral

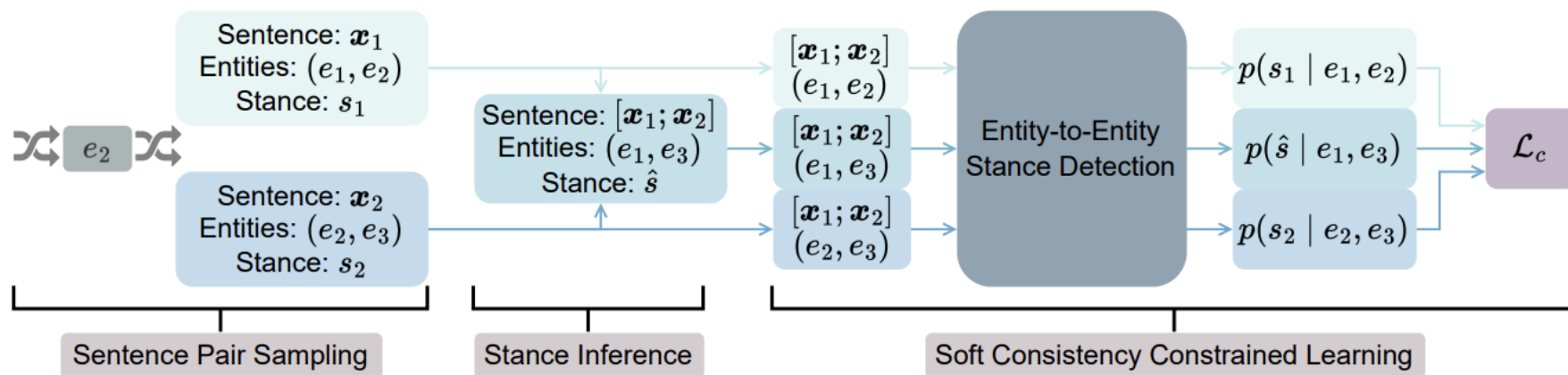
Generation-based Framework

- 출력 시퀀스의 log-likelihood를 최소화하는 방향으로 훈련

$$\begin{aligned}\mathcal{L}_s &= -\log p(\mathbf{o} \mid \mathbf{x}, e_1, e_2) \\ &= -\sum_{i=1}^{|\mathbf{o}|} \log p(o_i \mid \mathbf{o}_{<i}; \mathbf{T}(\mathbf{x}, e_1, e_2))\end{aligned}$$

Transitive Consistency Constrained Learning

- 가설: “전이적 연관성”이 정치 뉴스에서는 흔하며, 더 나은 모델을 훈련시키기 위해 활용될 수 있다.
- 목표: 기존에 존재하는 입장들 간의 전이성을 활용하여 예측된 입장을 최적화 함으로써 상관관계를 포착
- 방법: 두 문장에 나타난 엔티티들의 입장을 활용, cross-sentence stance prediction을 진행

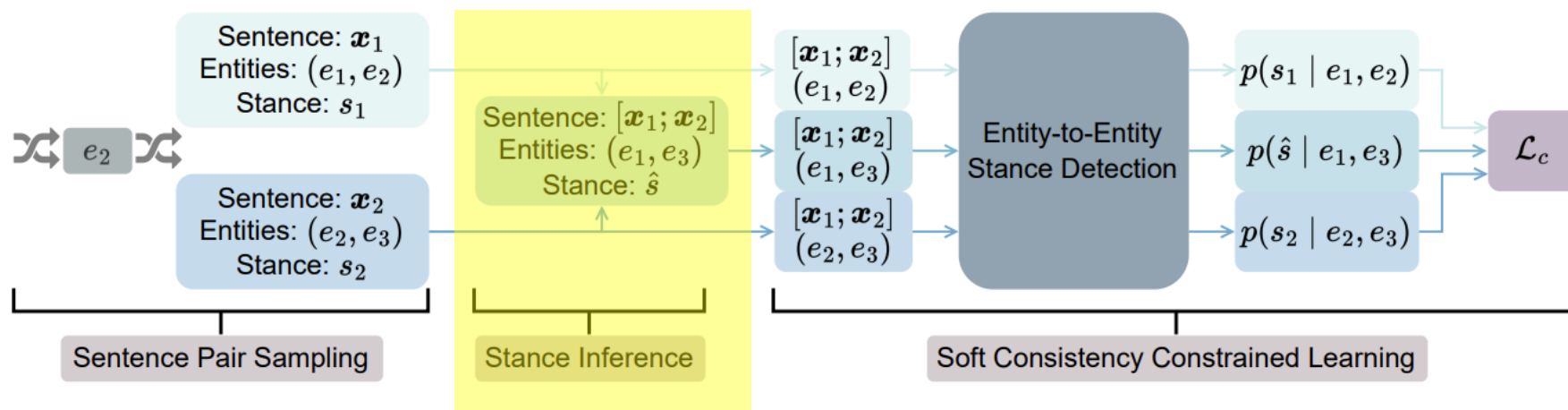


Transitive Stance inference

- 두개의 스텝

1. E1과 e3 사이 existing directed link가 존재하는가?

2. 입장의 극성 결정 $\hat{s}(e_1, e_3) = s(e_1, e_2) \oplus s(e_2, e_3).$



Transitive Stance inference

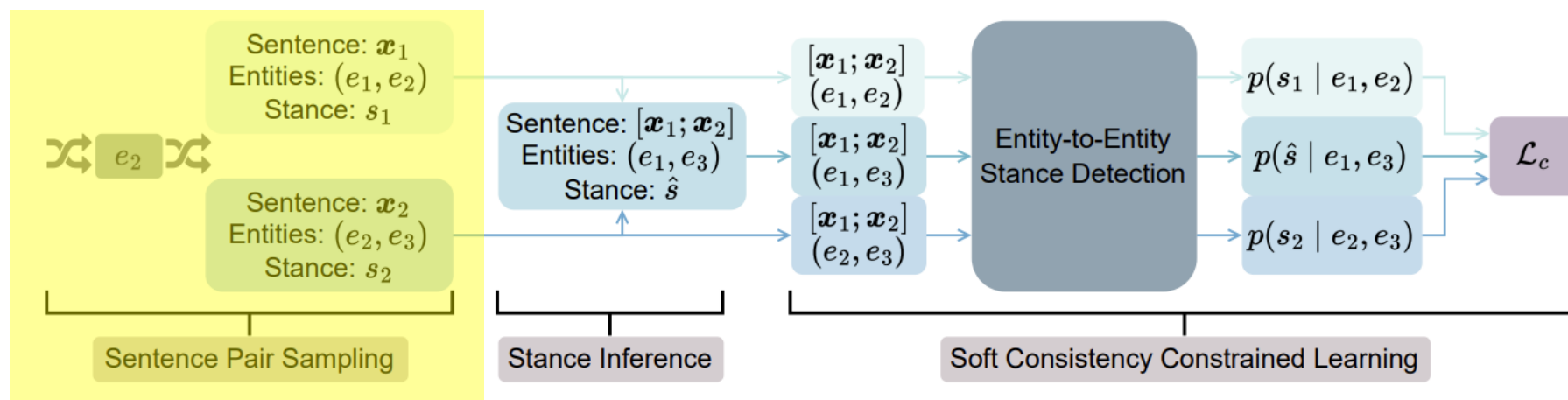
- 두개의 스텝

1. E1과 e3 사이 existing directed link가 존재하는가?
2. 입장의 극성 결정

		$e_2 \longrightarrow e_3$		$e_2 \longleftarrow e_3$	
		Positive	Negative	Positive	Negative
$e_1 \longrightarrow e_2$	Positive	$e_1 \xrightarrow{\text{Positive}} e_3$	$e_1 \xrightarrow{\text{Negative}} e_3$	-	-
	Negative	$e_1 \xrightarrow{\text{Negative}} e_3$	$e_1 \xrightarrow{\text{Positive}} e_3$	-	-
$e_1 \longleftarrow e_2$	Positive	-	-	$e_1 \xleftarrow{\text{Positive}} e_3$	$e_1 \xleftarrow{\text{Negative}} e_3$
	Negative	-	-	$e_1 \xleftarrow{\text{Negative}} e_3$	$e_1 \xleftarrow{\text{Positive}} e_3$

Two-Step sentence Pair sampling

- 기존 데이터는 문장 단위 라벨링에 집중함
하나의 문장에서 한 쌍의 엔티티를 뽑아서 라벨링
- 이 연구를 위해서는 입장 쌍과
하나의 엔티티를 공유하는 두 개의 엔티티 쌍이 필요



Two-Step sentence Pair sampling

1. Uniform sampling of shared entity

- 문장에 등장하는 엔티티들 중 하나를 무작위로 선택. 모든 엔티티가 동일한 확률로 선택되도록 하기 위해 균등 샘플링

2. Sentence pair sampling

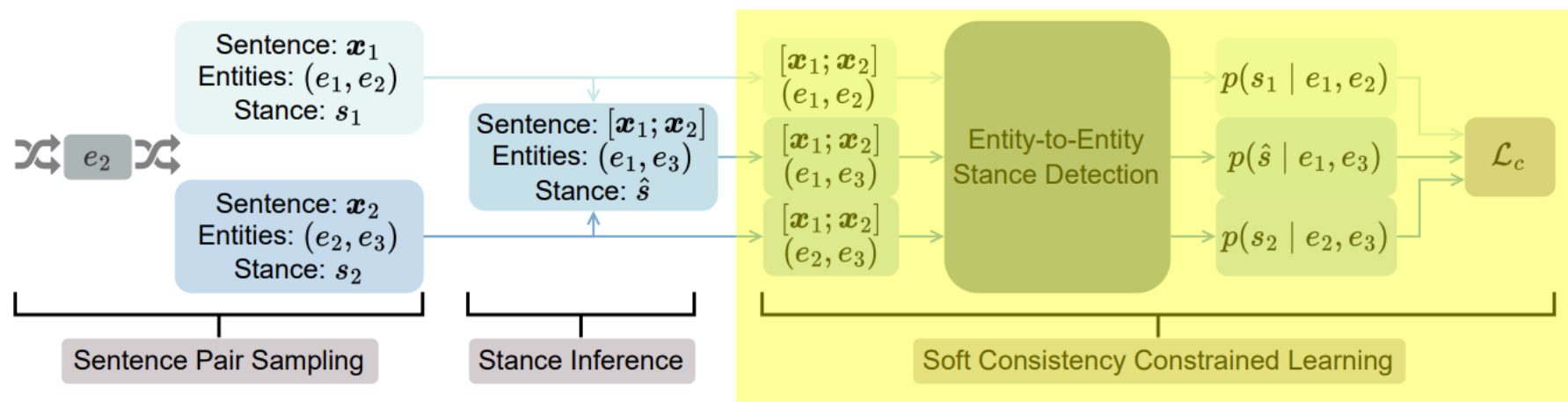
- 첫 번째 단계에서 선택된 엔티티가 포함된 모든 문장을 검색
- 그런 다음, 이러한 문장들 중 두 개의 문장을 선택하여 문장 쌍을 만듦

3. Validity check

- 두 문장에서 같은 엔티티 쌍이 있는 경우 샘플에서 제외
- 샘플링된 문장 쌍에서 전이적 관계를 학습할 수 없는 경우 샘플 무시

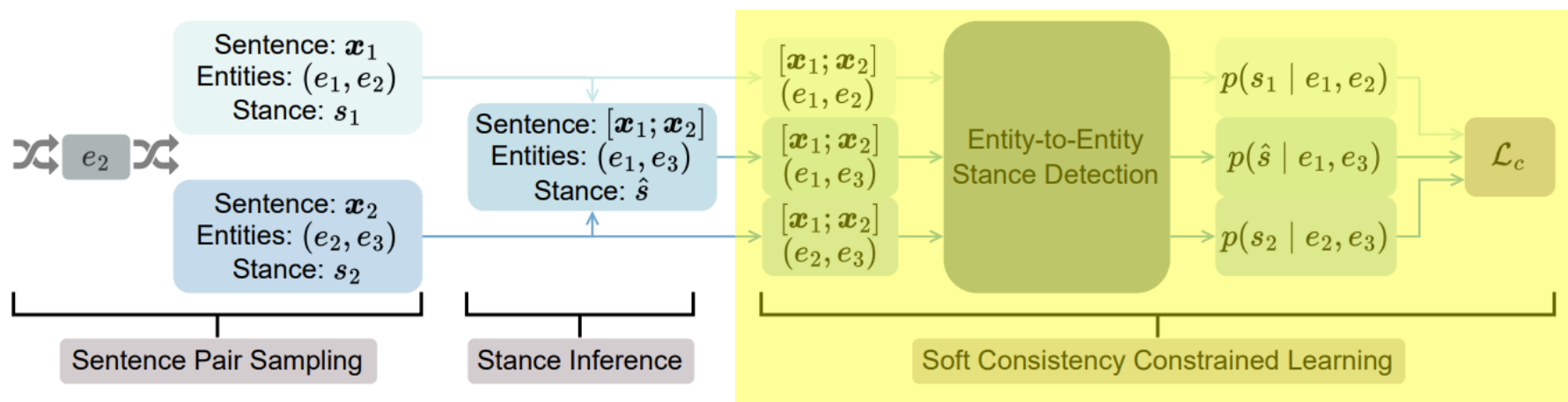
Soft consistency Constrained Learning

- Constrained Learning
 - 학습 과정에서 제약 조건을 추가하여
모델이 더 일관되고 특정 목표에 맞는 예측을 할 수 있도록 하는 방법



Soft consistency Constrained Learning

1. 주어진 문장 쌍과 엔티티 쌍이 있을 때, 엔티티들 간의 관계를 정규화하여 공유된 엔티티를 중심으로 입장 방향을 정리
2. 문장 쌍에 대해 세가지 확률 분포를 계산
 - 첫번째 문장 내 엔티티 간의 입장 확률, 두번째 문장 내 엔티티 간 입장 확률, 문장 간 엔티티 간의 입장 확률
 - 첫번째, 두번째는 ground truth stance, 세번째는 inferred stance



Soft consistency Constrained Learning

- 모델의 목표는 첫번째, 두번째 확률 분포에서 얻은 입장 정보의 곱이 세 번째 확률 분포와 일치하도록 하는 것
 - 이를 위해 L1 거리를 기반으로 한 추가적인 손실함수를 정의하여 모델의 예측을 유도

$$\begin{aligned}\mathcal{L}_c = & |\log p(s = s(e_{1,1}, e_{1,2}) | e_{1,1}, e_{1,2}) \\ & + \log p(s = s(e_{2,1}, e_{2,2}) | e_{2,1}, e_{2,2}) \\ & - \log p(s = \hat{s}(e_{1,1}, e_{2,2}) | e_{1,1}, e_{2,2})|\end{aligned}$$

- 최종적인 학습 목적 함수는 기존의 학습 목적 함수에 일관성 손실을 더하여 계산
 - 이 과정에서 람다 값을 사용해 일관성 제약 정도를 조절
 - 람다가 크면 엄격한 제약, 작으면 덜 제약

$$\mathcal{L} = \mathcal{L}_s + \lambda \mathcal{L}_c$$

Soft consistency Constrained Learning

- 생성 기반 방법으로 확장하는 방법

$$\mathcal{L}_c = |\log p(s = s(e_{1,1}, e_{1,2}) | e_{1,1}, e_{1,2}) + \log p(s = s(e_{2,1}, e_{2,2}) | e_{2,1}, e_{2,2}) - \log p(s = \hat{s}(e_{1,1}, e_{2,2}) | e_{1,1}, e_{2,2})|$$

- 모델이 생성한 텍스트에서 예측된 stance에 대한 log 확률을 계산해야 함.
- Ex) Entity 1 to Entity 2 positive
- Stance 라벨을 구성하는 Polarity, Entity 주요 요소를 선택하여 log 확률 계산
 $\log p(\text{stance}) = \log p(\text{polarity word}) + \log p(\text{entity 1}) + \log p(\text{entity 2})$

$$\mathcal{L} = \mathcal{L}_s + \lambda \mathcal{L}_c$$

Experiments

- 이 연구에서는 Transitive Consistency Constrained Learning을 제안
 - 엔티티 쌍과 스탠스 간의 Consistency를 파악하고,
Transitive Consistency를 강제하기 위해 추가적인 목표를 도입
- Key idea
 - A가 B에 긍정. B가 C에 긍정. A도 C에 긍정적일 “가능성”이 높다고 가정,
이러한 전이적 관계를 학습에 반영 → 보다 일관성 있는 Stance Detection 가능
- 이를 위해 두가지의 방법을 제안
 1. 엔티티 쌍의 representation으로부터 relation classification을 하는 것
 2. language model에 instruction tuning을 하는 것
→ stance를 autoregressively하게 generate함

Experiments – Dataset

- DSE와 SEESAW에서 실험을 진행
- 둘 다 정치적 뉴스에서 stance를 분석함
 - DSE (Park et al., 2021)
 - 모델이 입장 방향과 극성을 모두 예측해야 하며, 엔티티의 언급 위치도 포함됨
 - SEESAW (Zhang et al., 2022).
 - 원래는 엔티티 쌍과 그 입장을 함께 생성하는 실험을 위해 설계된 데이터
 - 엔티티 언급 수준의 정보와 중립 레이블은 제공하지 않음

Experiments

- Base models
 - Classification-based model: RoBERTa-base
 - Generation-based model: BLOOMZ-560m

Methods	Development Set		Test Set	
	Micro F ₁	Macro F ₁	Micro F ₁	Macro F ₁
LNZ (Combined)	69.40	65.16	70.55	53.58
LNZ (Context)	63.31	45.18	63.71	46.65
LNZ (EntityPrior)	59.14	44.27	58.53	40.63
DSE2QA (Complete)	78.92	67.51	77.26	66.17
DSE2QA (Pseudo)	80.72	68.27	79.73	67.66
POLITICS	85.45	71.94	84.19	71.12
Generation	83.92	70.14	83.25	70.12
+ Consistency Training	84.86	71.75	83.51	70.25
Classification	85.82	74.07	83.82	70.59
+ Consistency Training	86.67	74.41	85.19	72.50
BLOOMZ-176b + 25 samples	20.60	18.04	21.07	18.56

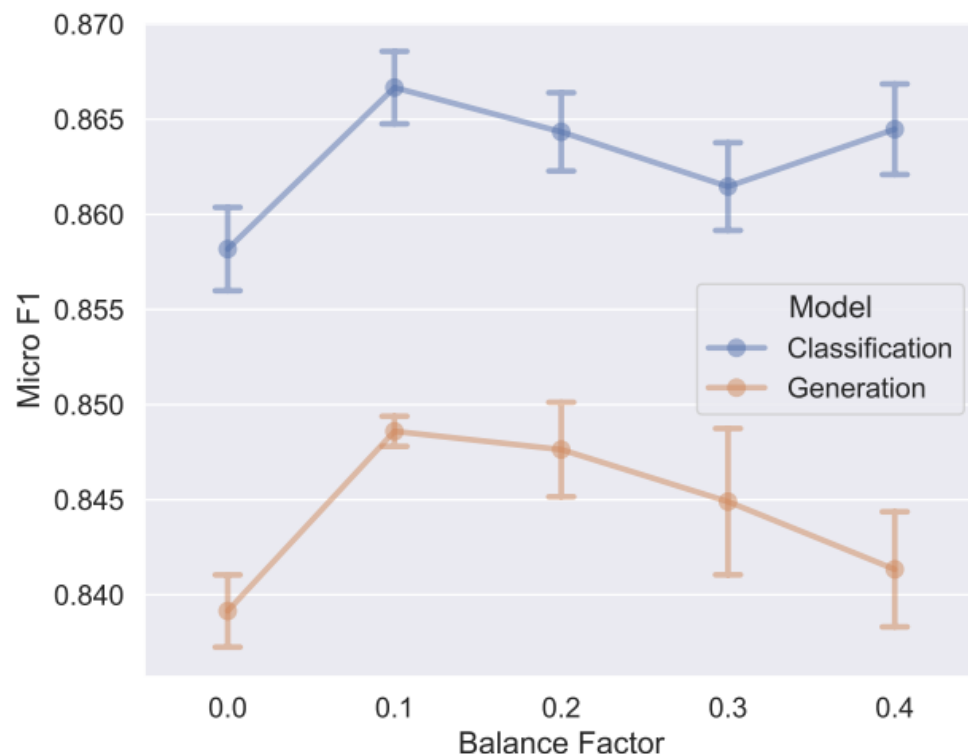
↑ Results on DSE

Methods	Micro F ₁
DSE2QA	83.35
POLITICS	84.02
Generation	80.35
+ Consistency Training	81.05
Classification	83.72
+ Consistency Training	84.11
BLOOMZ-176b + 10 samples	77.29

↑ Results on SEESAW

Quantitative Analysis

- Effects on Balance factor



- Effects on Consistency Learning

Training	Micro F_1	Macro F_1
Data Augmentation	84.55	71.91
Consistency Learning	85.19	72.50

- Effects on Sampling Methods

Sampling	Micro F_1	Macro F_1
Random Sampling	84.44	71.99
Two-Step Sampling	85.19	72.50

Challenges of Frozen Large Language Models

- BLOOM-176b를 Few-shot in-context learning
 - DSE 데이터셋에서 독립적인 레이블과 방향성을 예측하는 데 어려움을 겪음

Test Data Type	Micro F_1	Macro F_1
Full Label	21.07	18.56
- w/o Direction	26.74	26.19
- w/o Neutral	66.96	35.80
- w/o Both	77.62	71.73

Conclusion

- 분류 기반 방법과 생성 기반 방법 모두 엔티티 간 입장 감지에서 경쟁력 있는 성능을 보임
- 일관성 제약 학습은 분류 기반 방법과 생성 기반 방법 모두의 성능을 향상 시킴
- 제약 학습은 제약의 적용 정도를 조절하는 균형 요소에 민감함
- 대형 언어 모델은 few-shot in-context learning으로 복잡한 구조적 예측을 신뢰할 수 있는 수준으로 수행하지 못할 수 있음

나의 생각

- 알게된 점
 - 일관성 제약 학습이라는 개념을 알게 됨.
 - 텍스트에서 입장을 탐지하는 것이 아닌 엔티티 간 입장 탐지 작업에 대해 새로 알게 됨
- 아쉬운 점
 - 연구에서 사용하는 엔티티 간 입장 연관성이 항상 옳지는 않음
 - 현실 세계는 훨씬 복잡하므로 규칙 기반 방법으로는 일반화에도 어려움이 있을 것