"I'M NOT MAD"

Commonsense Implications of Negation and Contradiction

Liwei Jiang — Antoine Bosselut — Chandra Bhagavatula — Yejin Choi

PAUL G. ALLEN SCHOOL W · AI2 Allen Institute for AI · Stanford University

AI Systems Struggle with Negations

Alex **doesn't** wear a mask in public

COMET "Alex is seen as **secretive**" ✗

Alex pays **without** eating

COMET "As a result, Alex feels **full**" ✗

Alex **opposes** racism

COMET "Alex intends to **be a racist**" ✗

- Natural language expresses **negation** in **complex** and **subtle** ways using diverse **syntactic**, **semantic** and **pragmatic** formulations (Xiang et al., 2016)
- **Negated** language expressions are **much less likely to appear** in text than affirmative statements (Reitan et al., 2015)

Knowledge Model of Negations

Eval Set	Train Set	↓ PPL	↑ BLEU-2	↑ P@10
ATOMIC	ATOMIC	9.30	**14.18**	**55.18**
	ATOMIC + ANION	**9.28**	14.05	*53.61
ANION-L	ATOMIC	10.87	10.86	35.84
	ATOMIC + ANION	**9.08**	**11.96**	****45.42**
ANION-S	ATOMIC	11.69	12.07	36.89
	ATOMIC + ANION	**9.80**	**13.22**	****46.88**
ANION-C	ATOMIC	12.02	14.32	46.70
	ATOMIC + ANION	**11.20**	**14.64**	****50.65**

Training on ANION helps inferences of negated events

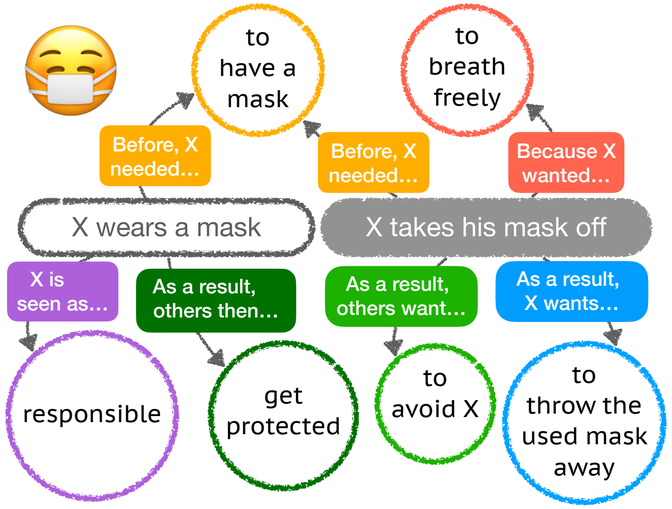
ANION

Array of commonseNse Inferences for Oppositions and Negations

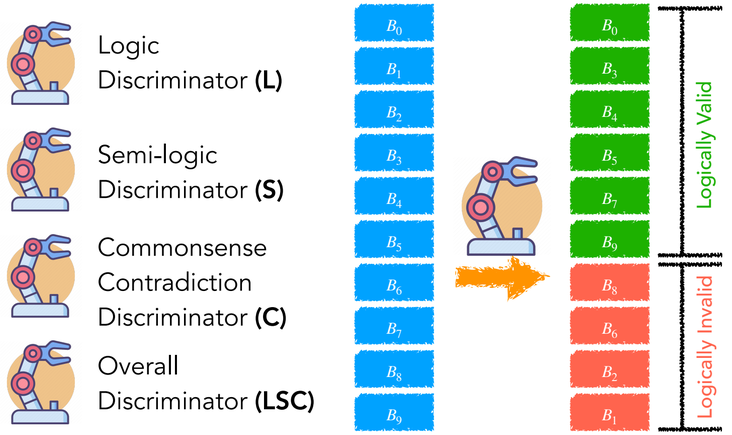
~**22K** new events, derived from ATOMIC, covering

- logical negation **(ANION-L)**
- semi-logical negation **(ANION-S)**
- commonsense contradiction **(ANION-C)**

~**627K** commonsense inferences around the new events, covering 9 relations



Discriminating Inconsistent Inferences



Eval Set		L	S	C	LSC
ANION-L	all	**55.69**	55.93	56.94	58.30
	valid	55.65	**56.18**	**57.26**	**59.07**
ANION-S	all	39.46	37.85	36.43	39.45
	valid	****46.3**	****41.9**	**37.54**	****45.5**
ANION-C	all	37.13	39.29	37.72	38.55
	valid	**37.48**	****44.5**	**39.03**	****44.9**

- Discriminators can identify **logic pitfalls** in inferences
- **Learning-based** and **discriminator-based** approaches are **complementary**

@liweijianglw

lwjiang@cs.washington.edu

https://liweijiang.me

Data