

Mid Presentation

CSE 712 Symbolic Machine Learning - II

Group 3
Members:
Auninda Alam-21166050
Marjan Tahreen-21166049
Shohag Rana-21366015



INTRODUCTION

Detecting Attackable Sentences in Arguments

Yohan Jo, Seojin Bang, Emaad Manzoor, Eduard Hovy, Chris Reed

CONTRIBUTIONS

Introduced the problem of detecting attackable sentences in arguments

Analyzed driving reasons for attacks in arguments and the effects of sentence characteristics

The performance of machine learning models for detecting attackable sentences

LITERATURE REVIEW

Aristotle (2007) suggested three aspects of argument persuasiveness.

Wachsmuth et al. (2017b) summarized various aspects of argument quality studied in argumentation theory and NLP

Some research took empirical approaches and collected argument evaluation criteria from human evaluators (Habernal and Gurevych, 2016a; Wachsmuth et al., 2017a)

Some studies aimed to model the salience of individual sentences in attacked arguments (Jo et al., 2018; Ji et al., 2018)

Dataset		Train	Val	Test
Attacked	#posts #sentences #attacked	$\begin{array}{c} 25,839 \\ 420,545 \\ 119,254 \end{array}$	8,763 $133,090$ $40,163$	$\begin{array}{c} 8,558 \\ 134,375 \\ 40,354 \end{array}$
#posts Successful #sentences #successful		3,785 $66,628$ $8,746$	1,235 $20,240$ $2,718$	$^{1,064}_{17,129}_{2,288}$

Table 1: Data statistics. "Attacked" contains posts with at least one attacked sentence. "Successful" contains posts with at least one successfully attacked sentence.

F1 Personal opinion (28%)
F2 Invalid hypothetical (26%)
F3 Invalid generalization (13%)
F4 No evidence (11%)
F5 Absolute statement (7%)
F6 Concession (5%)
F7 Restrictive qualifier (5%)
F8 Other (5%)

(b) Motivating factors for attacks.

Dataset

Source

The Dataset was formed using the online discussions from the "Change My View (CMV)" subreddit.

Labelling

Each sentence in a post was labelled into three categories, i.e. successfully attacked, un-successfully attacked and unattacked.

Feature Extraction

- content
- external knowledge
- proposition types
- tone

Model

	Attacked			Successful		
	P@1	A@3	AUC	P@1	A@3	AUC
Random	35.9	66.0	50.1	18.9	45.0	50.1
Length	42.9	73.7	54.5	22.3	52.1	55.7
LR	47.1	76.2	61.7	24.2	54.5	59.3
(×) Content	45.2	74.4	58.1	24.0	52.6	57.0
(x) Knowledge	47.0	76.0	61.7	24.1	54.3	59.0
(×) Prop Type	46.7	75.9	61.5	24.4	53.6	59.0
(×) Tone	47.0	76.0	61.9	25.2	56.2	59.4
BERT	49.6	77.8	64.4	28.3	57.2	62.0
Humans [†]	51.7	80.1	-	27.8	54.2	_

Problem Formulation

- P@1
- A@3
- AUC

ML Models

- Logistic Regression
- BERT

Baseline Models

- Random
- Length

Results

Based on the Computational Model run 10 times

I'm typing this post mostly from anxiety Prediction (0.12) Personal (-0.20) considering recent events, but hopefully this pos-Topic37 (-0.21) will spark optimistic discussion that I don't see often in the news or online or such. With the appointment of KialoFreq (0.98) John Bolton as the National Security Adviser and Topic5 (0.39) KialoAttr (0.05) John Pompeo as the Secretary of State, two men known KialoExtr (-0.07) for hawkish and pro-war behavior in their previous tatements and actions, the US has appeared to take KialoFreq (0.75) more aggressive stance in foreign policy, seen with the Topic5 (0.39) expulsion of sixty Russian diplomats following minor Example (0.11) controversy in the United Kingdom, Also, despite KialoAttr (0.07) planned negotiations with Kim Jong-Un concerning the KialoExtr (-0.07) future of North Korea, the US, and NK's nuclear arsenal. President Trump has filled out his cabinet/diplomacy team Topic5 (0.39) with people who are in favor of things such as a KialoFreq (0.22) KialoAttr (0.13) regime change or attacking North Korea, further stirring Hypothetical (-0.06) things up for a potential falling out. If talks between KialoExtr (-0.11) the two nations break down, the US does not have much more of a reason to withhold from attacking KialoFreg (0.45) North Korea, which is a plan that seems to be Topic5 (0.39) KialoAttr (0.26) favorable among higher officials. Considering that KialoExtr (-0.05) this is also sort of a proxy scuffle between us and China/ Use of "We" (-0.18) Russia, attacking or otherwise provoking North Korea or Topic5 (0.39) Russia could lead to situations ranging from a QuestOther (0.39) worldwide economic downturn to nuclear holocaust. Is conflict the current trajectory of international Why/How (0.91) relations? How would we otherwise not engage Use of "We" (-0.18) in some sort of scuffle? Topic37 (-0.21)

Figure 2: Prediction visualization. Background color indicates predicted attackability (blue: high, red: low). Successfully attacked sentences are underlined. Features with high/low weights are indicated with blue/red.

LR and BERT Outperform

Both the LR and BERT models significantly outperform the baselines, while the BERT model performs best.

	Attacked			Successful		
	P@1	A@3	AUC	P@1	A@3	AUC
Random	35.9	66.0	50.1	18.9	45.0	50.1
Length	42.9	73.7	54.5	22.3	52.1	55.7
LR	47.1	76.2	61.7	24.2	54.5	59.3
(×) Content	45.2	74.4	58.1	24.0	52.6	57.0
(x) Knowledge	47.0	76.0	61.7	24.1	54.3	59.0
(×) Prop Type	46.7	75.9	61.5	24.4	53.6	59.0
(×) Tone	47.0	76.0	61.9	25.2	56.2	59.4
BERT	49.6	77.8	64.4	28.3	57.2	62.0
Humans [†]	51.7	80.1	_	27.8	54.2	_

Conclusion



