IMHO Fine-Tuning Improves Claim Detection

Tuhin Chakrabarty, Christopher Hidey,
Kathleen Mckeown

Natural Language Processing
Columbia University

WHY IS CLAIM DETECTION IMPORTANT ?

Claims are the central component of any argument.
Detecting claims can be useful for:

- Automatic scoring of persuasive essays

- Fact-checking in news articles

- Generating arguments

WHAT IS A CLAIM ?

- Assertion that deserves our attention (Toulmin, 2003)

- A statement that is in dispute and that we are trying to support with reasons (Govier, 2010)

DETECTING CLAIMS CAN BE CHALLENGING

- Claims can be phrased very differently across different domains or datasets (e.g. student essay vs. web-discussion forum)

- Most datasets for claim detection are very small

OUR CONTRIBUTIONS

We address these challenges by:

- Introducing a new large dataset of distant-labeled claims
- Using this dataset to fine-tune a pre-trained model for the task of claim detection
- Showing improvements for claim detection on diverse argumentation datasets

EXISTING ARGUMENTATION DATASETS

TYPE	DATASET	#CLAIMS	NO OF SENTENCES	PERCENTAGE OF CLAIMS
MONOLOGUE	MICROTEXT (MT)	112	449	24.94
	PERSUASIVE ESSAY (PE)	2108	7116	29.62
DIALOGUE	WEB DISCOURSE (WD)	211	3899	5.41
	CHANGE MY VIEW (CMV)	1206	3451	34.0

EXISTING ARGUMENTATION DATASETS

TYPE	DATASET	#CLAIMS	NO OF SENTENCES	PERCENTAGE OF CLAIMS
MONOLOGUE	MICROTEXT (MT)	**112**	449	24.94
	PERSUASIVE ESSAY (PE)	2108	7116	29.62
DIALOGUE	WEB DISCOURSE (WD)	**211**	3899	**5.41**
	CHANGE MY VIEW (CMV)	1206	3451	34.0

DIFFERENCE IN PHRASING ACROSS DATASETS

MT : 50% of claim sentences contain the modal verb **'should'**

The death penalty **should** be abandoned everywhere

PE : Most claims contain discourse markers

In my opinion, although using machines have many benefits, we

cannot ignore its negative effects.

WD : Most claims contain explicit sentiment

I regard single sex education as **bad**.

PRIOR WORK

Daxenberger et al (2017) recognize these challenges using the same data sets

- In-domain experiments using CNN and Logistic Regression obtained state of art performance
- Cross-domain experiments with same models trained/ tested on different datasets decreased performance
- Analysis identified shared lexical properties useful for heterogeneous or unknown domains

OUR CONTRIBUTIONS

We address these challenges by:

- **Introducing a new large dataset of distant-labeled claims**
- Using this dataset to fine-tune a pre-trained model for the task of claim detection
- Showing improvements for claim detection on diverse argumentation datasets

DISTANT LABELED CORPUS FOR CLAIMS

- How do we find a corpus with language which is indicative of claims?

- It is common for users on Reddit platforms to express their opinions and claims using **IMO / IMHO (In my (Humble) Opinion)**?

DISTANT LABELED CORPUS FOR CLAIMS

Are boneless wings technically nuggets?

Unanswered

↑ PM_ME_UR_SCOOTER   1 point · 3 months ago
↓ IMHO, the main thing about "boneless wings" is that they're going to be sauced up like buffalo wings.

   💬 Reply   Give Award   Share   Report   Save

DISTANT LABELED CORPUS FOR CLAIMS

- How do we find a corpus with language which is indicative of claims? IMHO!
    - Automatic data collection of 5.5 million self-labeled opinions
    - Introduces minimal linguistic bias

- How can we incorporate this dataset into a claim detection model?
    - IMHO dataset contains no negative examples
    - Transfer knowledge using a language model (LM)

TRANSFER LEARNING FOR NLP

- Leveraging prior knowledge from one domain and task into a different domain and task.
  - Contextualized representations [ELMO (Peters et al, 2018), BERT (Devlin et al, 2018)]
  - Language model fine-tuning [ULMFIT (Howard and Ruder, 2018)]

OUR CONTRIBUTIONS

We address these challenges by:

- Introducing a new large dataset of distant-labeled claims
- **Using this dataset to fine-tune a pre-trained model for the task of claim detection**
- Showing improvements for claim detection on diverse argumentation datasets

TRANSFER LEARNING MODEL FOR CLAIM DETECTION

Universal Language Model Fine-Tuning (ULMFiT):

a) LM pre-training on large *general domain* corpus

b) *Task-specific* LM fine-tuning on corpus used for classification

c) *Task-specific* classifier fine-tuning on corpus used for classification

TRANSFER LEARNING MODEL FOR CLAIM DETECTION

Universal Language Model Fine-Tuning (ULMFiT):

a) LM pre-training on large *general domain* corpus

b) ~~*Task-specific* LM fine-tuning on **corpus used for classification**~~

b) *Task-specific* LM fine-tuning on **IMHO corpus**

c) *Task-specific* classifier fine-tuning on corpus used for classification

OUR CONTRIBUTIONS

We address these challenges by:

- Introducing a new large dataset of distant-labeled claims
- Using this dataset to fine-tune a pre-trained model for the task of claim detection
- **Showing improvements for claim detection on diverse argumentation datasets**

ULMFIT SCORES FOR CLAIM DETECTION

CLAIM-F1/MACRO-F1 SCORES

DATASET	CNN	TASK SPECIFIC LM FINE TUNING	IMHO LM FINE TUNING
MICROTEXT (MT)	67.3/78.6	67.0/78.1	**71.2/81.1**
PERSUASIVE ESSAY(PE)	61.1/73.6	64.0/74.1	**64.3/74.8**
WEB DISCOURSE (WD)	28.9/62.6	28.5/62.7	**33.2/65.2**
CHANGE MY VIEW(CMV)	53.5/63.8	57.0/66.4	**57.8/67.8**

QUALITATIVE ANALYSIS

Sentences from each dataset and nearest neighbor in IMHO

DATASET	SENTENCE
MT	That's why they **should be taken off the market**, unless they're unbreakable .
IMHO	**Should be taken off the market**
PE	However, flourishing tourism in a place **can be very rewarding** in terms of local economy
IMHO	It **can be very rewarding**.
CMV	Other than **anecdotal** evidence, **I have**n't **seen** anything **to support this claim**
IMHO	**I have** personally seen no evidence **to support this claim**, but that's just my **anecdotal** experience .

POTENTIAL MODEL BIAS ?

Example of Premises mis-classified as Claims

DATASET	SENTENCE
MICROTEXT (MT)	Video games namely FIFA in my case , can fascinate young people for hours **more** intensively and emotionally **than any** sport in the world
	It would be fair to make them into an Olympic event
PERSUASIVE ESSAYS (PE)	Last but not the least using public transportation is **much** safer **than** using private transportation
	There is no reason that governments should hesitate to invest in public transportation, a healthy, safe and economical way of transporting

OBSERVATIONS & POTENTIAL SOLUTIONS

Discriminatory phrases (e.g. **should**, **must be**) and comparative statements (e.g. **much safer than**, **more ... than any**) often indicative but lack of context can mislead

LM with context sentences or jointly modeling context (e.g. by predicting relations between claims and premises) may help

CONCLUSION

- Created new dataset by collecting over 5 million opinionated claims, validated on a variety of claim detection domains
- Showed that LM fine-tuning on IMHO rather than individual datasets, leads to statistically significant improvement over previous state of-the-art
- Provided empirical results and qualitative analysis showing features indicative of claims that transfer across data-sets

QUESTIONS

To obtain data please email us at

tc2896@columbia.edu

chidey@cs.columbia.edu