GerCCT: An Annotated Corpus for Mining Arguments in German Tweets on Climate Change

Robin Schaefer & Manfred Stede
Applied Computational Linguistics
University of Potsdam, Germany
{robin.schaefer | stede}@uni-potsdam.de

Introduction: Motivation, Related Work & Contributions

Motivation: Twitter is used for debates on controversial issues like climate change. This renders it an interesting medium for argument mining (AM).

Related work on AM on Twitter used simplistic argument schemes with a focus on single argument components (claim or evidence):

- general argument detection [1]
- evidence type detection [2]
- claim detection [3]

Our previous work showed that a high degree of subjectivity regarding the claim/evidence definitions complicates the annotation task [4].

Hypothesis: Annotating argument properties reduces the degree of subjectivity due to more precise definitions.

Our Contributions: We present GerCCT [i] (n=1,200), the German Climate Change Tweet Corpus, a new tweet resource annotated for argumentation.

1. Three argument annotation layers:
   - properties
   - components (claim/evidence)
   - general argument (= claim and/or evidence)
2. Annotation of sarcasm and toxic language to facilitate filtering of non-argumentative tweets
3. Training of first classification models on the annotated corpus

Note: *Components* and *general argument* are abstracted from property annotations. Also, this poster does not show our sarcasm/toxic language results due to limited space. The reader is referred to the paper for details.

The Corpus: IAA, Examples & Class Distribution

Annotation Procedure:

- Annotators (A1,A2) labelled classes for 300 tweets for IAA calculation. Then, annotators proceeded individually until the full corpus size (n=1,200) was reached.
- All annotations are conducted on the tweet level.
- Each tweet can be annotated with more than one class.

Annotation Class	Krippendorff's α
Unverifiable Claim	0.63
Verifiable Claim	0.64
Reason	0.41
External Evidence	0.83
Internal Evidence	0.40
Argument	0.71
Claim	0.69
Evidence	0.64

Note: We use the following abbreviations: Unverifiable Claim=UC; Verifiable Claim=VC; External Evidence=EE; Internal Evidence=IE.

	Tweet Examples	Annotations
1)	Such random prices render the public transport unappealing and expensive. [...] [link]	UC EE
2)	You cannot negotiate with nature. This is why you cannot prepare a climate protection package like a trade agreement. It's about science and its laws are non-negotiable. [...]	reason UC VC
3)	The biggest issue for the climate are people. There are calculations based on the assumption of 50 tonnes of CO2 per person. The planet is suffering from overpopulation To not get kids is the best you can do for the environment. [...]	UC VC reason
4)	It already starts with the definitions. [...] What is climate change denial? I personally don't know anyone who doubts the human influence	UC IE

Absolute Occurrences and Proportions of Argument Components and Properties

Set	Argument		Claim		Evidence		UC		VC		Reason		EE		IE	
A1	219	0.73	205	0.68	77	0.26	186	0.62	65	0.22	29	0.10	48	0.16	3	0.01
A2	211	0.70	203	0.68	67	0.22	177	0.59	78	0.26	23	0.08	43	0.14	2	0.01
Full Corpus	844	0.70	784	0.65	295	0.25	703	0.59	244	0.20	132	0.11	165	0.14	11	0.01

Classification Experiments

Feature Types and Classification Algorithms:

- Feature types: Unigrams, BERT Embeddings
- Algorithms: XGBoost, Logistic Regression (LR), Softmax, Naive Bayes, SVM

For the Unigram approach, we lowercased the tweets, removed punctuation and replaced links with a placeholder ([link]). Also, we experimented with fine-tuning the BERT architecture.

Classification Results: F1 Macro Scores

Approach	Argument	Claim	Evidence	UC	VC	Reason	EE
Majority	0.41	0.40	0.43	0.37	0.44	0.47	0.46
Unigrams + XGBoost	0.66	0.63	0.72	0.63	0.56	0.54	0.81
BERT + LR	0.69	0.68	0.74	0.66	0.62	0.55	0.84
BERT (ft) + Softmax	**0.70**	**0.73**	**0.77**	**0.70**	**0.69**	**0.60**	**0.86**

Discussion

- IAA for most classes is promising and higher than in our previous study [4].
- Fine-tuned BERT model approach performs best for all argument classes.
- Major weaknesses:
  1. Reason shows low IAA due to a high degree of subjectivity and lack of explicit linguistic markers.
  2. IE shows low IAA due to rare use in the corpus and linguistic markers comparable to claims.

Resources & References

Resources:
[i] Annotated Corpus: `https://doi.org/10.5281/zenodo.6479492`

References:
[1] Bosc, Cabrio & Villata (2016). *DART: a Dataset of Arguments and their Relations on Twitter.*
[2] Addawood & Bashir (2016). *"What is your Evidence?" a Study of Controversial Topics on Social Media.*
[3] Wührl & Klinger (2021). *Claim Detection in Biomedical Twitter Posts.*
[4] Schaefer & Stede (2020). *Annotation and Detection of Arguments in Tweets.*