Content & Context

in Argumentative Relation Classification

ArgMining 2019

Juri Opitz & Anette Frank
Heidelberg University

Argumentative Relation Classification

Marijuana should be legalized.

Legalizing marijuana can increase use by teens, with harmful results.

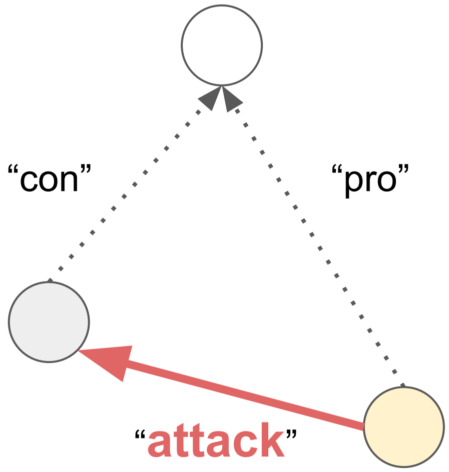
Admittedly,

However,

On the other hand,

...

Legalization allows the government to set age-restrictions on buyers.



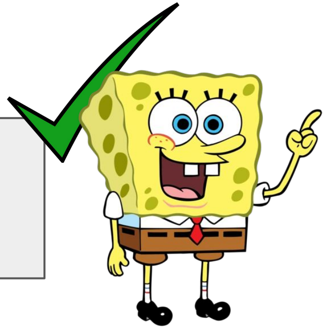
Intermediate insight

- in some cases, inspection of shallow discourse clues can help predict argumentative relations with high accuracy

single
-doc

(....)
AU-1 Moreover, AU-2
(....)

AU-2 supports AU-1

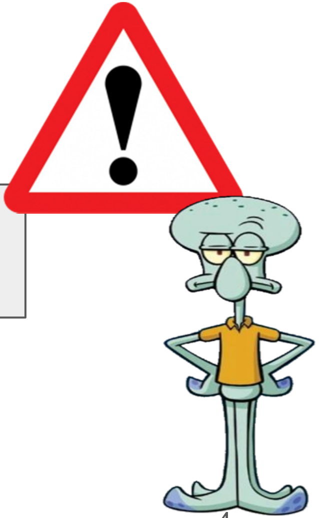


multi-
doc

(....)
AU-1
(....)

(....)
Moreover, AU-2
(....)

AU-2 supports AU-1



Research questions

- we want to investigate to what extent systems rely on shallow discourse clues
- where do we stand in **content-based** argumentative relation classification?
    - necessary for large scale **cross-document argumentative relation mining**
        - argumentative units for many debates can be mined from millions of documents scattered across the www
        - to assess relations between them we cannot rely on discourse clues but need systems which learn the **content/meaning of argumentative units**

Methodology

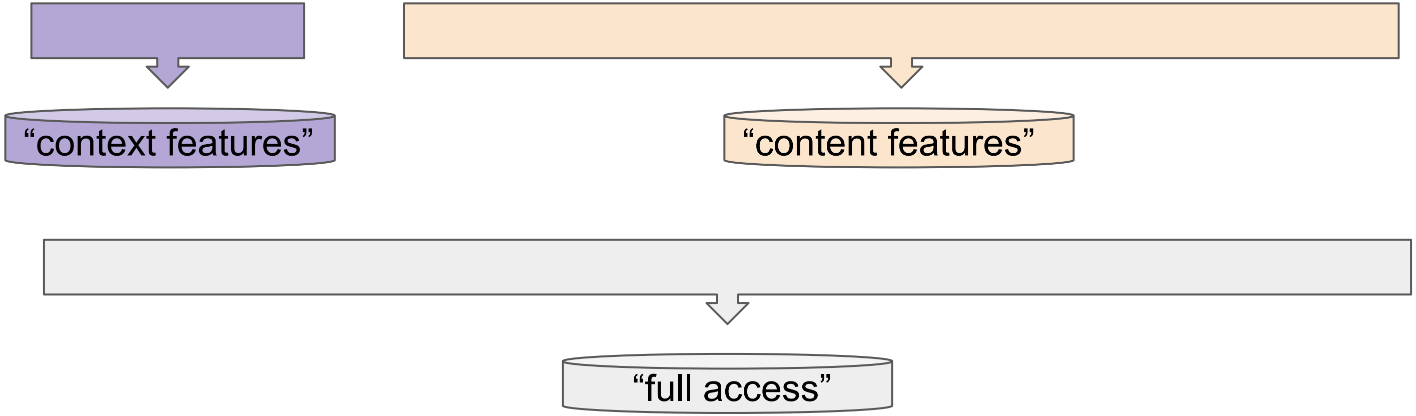
1. we replicate a competitive argumentative relation classifier:

   SVM (Stab and Gurevych, 2017) with

   i. discourse features
   ii. sentiment features
   iii. bag-of-word features
   iv. bag-of-production-rule features
   v. GloVe features
   vi. structural features
2. we extract these features from different spans
   a. features extracted from the **argumentative unit span** (“content”)
   b. features extracted from the **unit’s embedding context** (“context”)
   c. features extracted from both (“full-access”)

Data

- 402 Student essays (Stab and Gurevych, 2017)
- annotated with argumentative units and more than 3,000 relations
- class distribution: ca. 10% ‘attack’, ca. 90% ‘support’
- annotated unit spans correspond to argumentative clauses
    - “*On the one hand*, [AU: *Legalization can increase use by teens, with harmful effects*]”



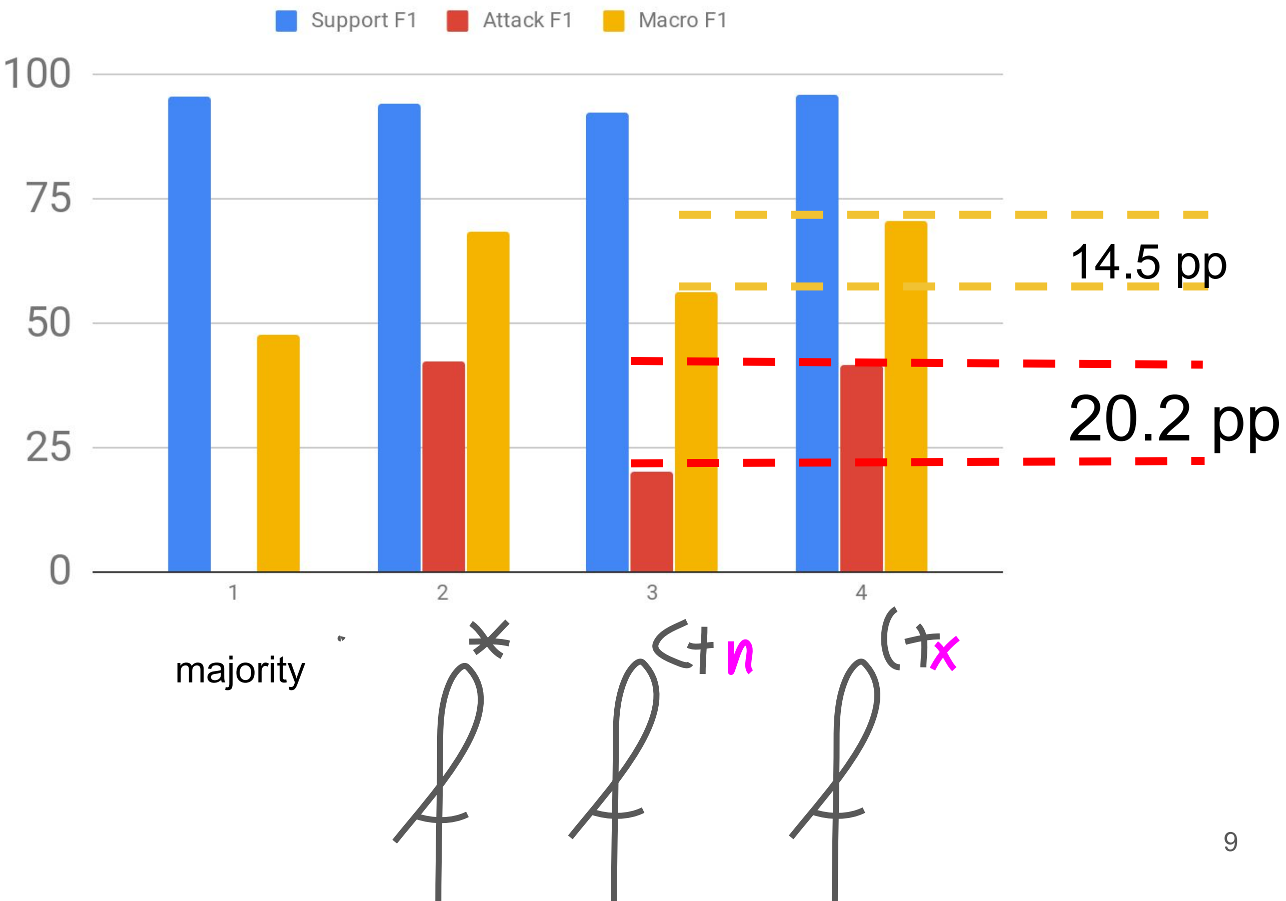
“context features”
“content features”
“full access”

$$f^{*}\left(\boxed{\text{AU-1}}, \boxed{\text{CTX}}, \boxed{\text{AU-2}}\right)$$

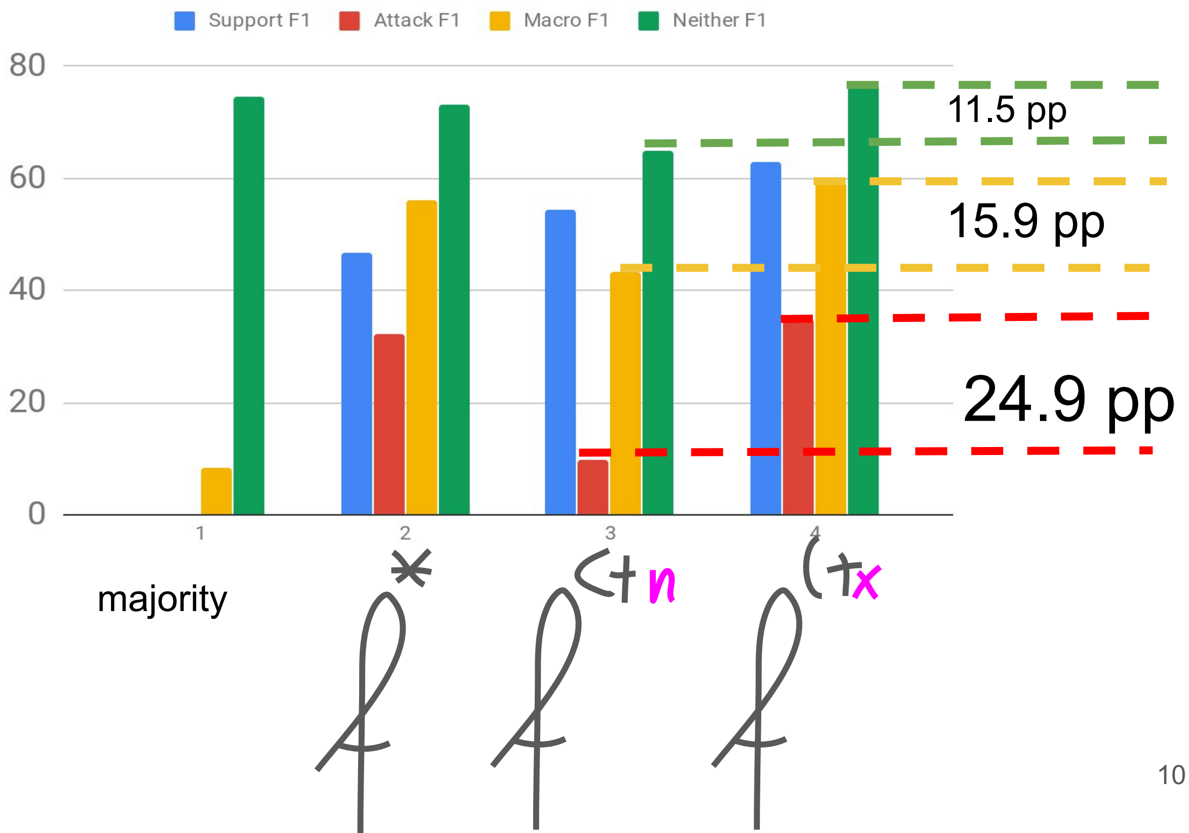
$$f^{ctx}\left(\boxed{\text{CTX}}\right)$$

$$f^{ctn}\left(\boxed{\text{AU-1}}, \boxed{\text{AU-2}}\right)$$

F1 Results: Attack vs. Support



F1 Results: Attack vs. Support vs. Neither



Is everything lost for $\ell^{<t\,n}$?

- No!
  - still outperforms majority baseline by a good margin
    - +9.5 pp. macro F1 in support vs. attack
    - +10.5 pp. macro F1 in support attack vs. neither

Cross document potential of f^{*}, f^{ctn} and f^{ctx}

To investigate, how the three systems port to a cross document scenario, we conduct two **simulation studies**:

- **random context**: we shuffle the contexts of testing instances to simulate porting to open world where AUs may appear in arbitrary contexts
- **no context:** we mask the contexts of testing instances for all three systems

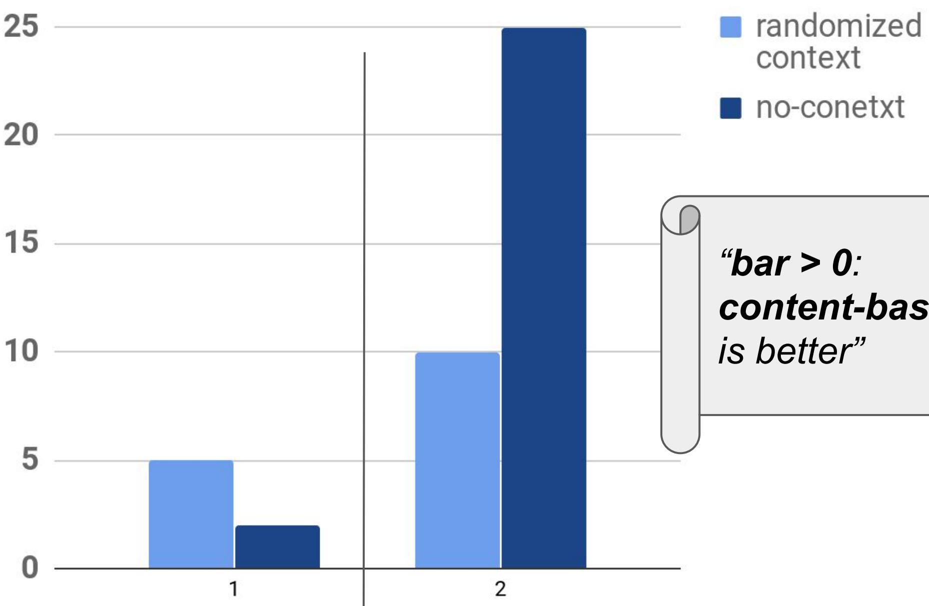
No-context

“<MASK>: *Legalization can increase use by teens, with harmful effects*]”

Randomized context

“Therefore, : *Legalization can increase use by teens, with harmful effects*]”

Macro F1 Results



$\ell^{(t+n)} - \ell^{*}$

$\ell^{(t+n)} - \ell^{(t+x)}$

*“**bar > 0**:*
content-based
is better”

Results

- we see the reverse picture:
    - **models which access context** (full-feature model and context-only model) **fall behind the content-based** system

Problem

Context can be exploited by systems in single-document contexts but can lead to confusion when discourse markers are **missing** or **cannot be trusted** (cross-document)

Insight

Context-focused systems are not safe for porting to cross-document scenarios

Recommendation

Develop content-based systems for cross-document scenarios

Take-Aways

We have shown that

- shallow discourse clues are very strong indicators for argumentative relations
- a very naive system that only sees context can strongly outperform a system which sees the content and also outperforms a system which sees everything

Insight 1

Good scores may not reflect capacity to model argumentative content

Insight 2

Argumentative relation classification needs better modeling of content

Conclusions

Need work towards **content-based** argumentative relation classification

- to address large scale argumentative relation mining **across document boundaries**
- Student essay data can serve as a first benchmark
    - task: predict relations based on the content of argumentative units, **mask context**
- Our results may serve as a **baseline**

Thank you for your attention!