Learning goals

- Concepts
  - Various properties of argumentation to be assessed
  - Theoretical notions of argumentation quality
  - The subjective nature of argumentation properties

- Methods
  - Route kernels and more for stance and myside bias
  - Feature-based and neural methods for schemes and fallacies
  - Classification, regression, and graph analyses for quality

- Associated research fields
  - Argumentation theory and rhetoric
  - Computational linguistics

- Within this course
  - How to understand aspects of the (previously mined) arguments and their structure.
Outline

I. Introduction to computational argumentation
II. Basics of natural language processing
III. Basics of argumentation
IV. Argument acquisition
V. Argument mining
VI. Argument assessment
VII. Argument generation
VIII. Applications of computational argumentation
IX. Conclusion

VI. Argument assessment

a) Introduction
b) Stance and bias
c) Schemes and fallacies
d) Quality in theory
e) Absolute and relative quality assessment
f) Objective and subjective quality assessment
g) Conclusion
What is argument assessment?

- **Argument(ation) assessment**
  - Coverage term for analysis tasks that detect, classify, rate or otherwise judge specific properties of argumentative units, arguments, or argumentative texts.

  "If you wanna hear my view, I think that the EU should allow sea patrols in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people."

- **Why argument assessment?**
  - Argumentative structure alone is not sufficient for many applications.
  - Often, some understanding is needed of how an arguments relates to an issue, how it works, and how good or important it is.
What properties of argumentation to assess?

- **What is meant by argumentation properties?**
  - Properties that reflect an understanding of aspects of argumentation.
  - These properties can be formalized as labels, scores, additional text fragments, or similar.

- **Selected properties to assess**
  - **Subjectiveness.** Stance, myside bias, emotions, ...
  - **Reasoning.** Argumentation schemes, fallacies, warrants, enthymemes, ...
  - **Argumentation quality.** Logical, rhetorical, and dialectical strength, ...
  - **Content-related.** Issues, aspects, frames, creation date, ...
  - **Style-related.** Genre, authorship, discourse modes, rhetorical moves, ...
  - **Structure-related.** Argumentative depth, claim centrality and diviseness, ...

- **Notice**
  - Where mining ends and assessment starts, is not perfectly unambiguous.

For example, classifying evidence types might be seen as assessment.
Subjectiveness: Stance and myside bias

**Stance** (recap)

- The overall position held by a person towards some target, such as an object, statement, or issue.
  - Near-synonyms: Viewpoint, view, standpoint, stand, position.
- To have/take a stance on a target means to be pro or con towards it.

Con towards death penalty.  
The death penalty must be abolished.

Pro towards the left claim.  
It doesn’t deter people from violence.

**Myside bias**

- The focus on information that confirms one’s stance, giving disproportionally less attention to information that contradicts the stance.
  - Near synonym: Confirmation bias
- An argumentative text with myside bias gives only reasons supporting its stance, no counter-considerations.
Reasoning: Argumentation schemes and fallacies (recap)

- **Argumentation scheme**
  - The form of inference from an argument’s premises to its conclusion.
  - Around 60 deductive, inductive, and especially abductive schemes exist.

- **Example schemes**
  - Argument from example
  - Argument from consequence
  - Argument from position to know

- **Fallacy**
  - An argument with some (often hidden) flaw in its reasoning, i.e., it has a failed or deceptive scheme.

- **Example types of fallacies**
  - Ad-hominem. Attacking the opponent instead of attacking her arguments.
  - Appeal to ignorance. Taking lack of evidence as proof for the opposite.

---

**Conclusion**  
*A is true.*

**Major premise**  
*Source E is in a position to know about things in a subject domain S with proposition A.*

**Minor premise**  
*E asserts that A is true (in domain S).*
# Overview of argumentation schemes

## A common set of argumentation schemes (Walton et al., 2008)

- Argument from witness testimony
- Argument from popular opinion
- Argument from popular practice
- Argument from example
- Argument from composition
- Argument from division
- Argument from oppositions
- Argument from alternatives
- Argument from verbal classification
- Argument from definition to verbal classification
- Argument from vagueness of a verbal classification
- Argument from arbitrariness of a verbal classification
- Argument from interaction of act and person
- Argument from values
- Argument from the group and its members
- Practical reasoning argument
- Argument from waste
- Argument from sunk costs
- Argument from correlation to cause
- Argument from sign
- Argument from evidence to a hypothesis
- Argument from consequences
- Argument from threat
- Argument from fear appeal
- Argument from danger appeal
- Argument from need for help
- Argument from distress
- Argument from commitment
- Ethotic argument
- Generic ad hominem argument
- Pragmatic inconsistency argument
- Argument from inconsistent commitment
- Circumstantial ad hominem argument
- Argument from bias
- Bias ad hominem argument
- Argument from gradualism
- Slippery slope argument
Argumentation quality

- **Argumentation quality**
  - Natural language argumentation is rarely logically *correct* or *complete*.
  - Quality reflects how *good* a unit, argument, or argumentation is.

"If you wanna hear my view, I think that the EU should allow sea patrols in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people."

- **Observations**
  - **Goal orientation.** What is important, depends on the goal of argumentation.
  - **Granularity.** Quality may be addressed at different levels of text granularity.
  - **Dimensions.** Several dimensions of quality may be considered.
Argumentation quality: Theory and in practice

- **Quality in theory**
  - The normative view of quality in terms of cogency, reasonableness, or similar.
  - Suggests to use *absolute* quality ratings.

- **Quality in practice**
  - Quality is decided by the effectiveness on (some type of) people.
  - *Relative* comparisons are often more suitable.

  ”Is a strong argument an effective argument which gains the adherence of the audience, or is it a valid argument, which ought to gain it?“

  (Perelman and Olbrechts-Tyteca, 1969)

- **Unresolved questions**
  - Should quality be aligned with how we *should* or how with we *do* argue?
  - Is this actually so different? (more on this below)
I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

Next section: Stance and bias

a) Introduction
b) Stance and bias
c) Schemes and fallacies
d) Quality in theory
e) Absolute and relative quality assessment
f) Objective and subjective quality assessment
g) Conclusion
What is stance and myside bias classification?

- **Stance classification**
  - The classification of the stance of a (span of) text towards a given target.
  - **Input.** An argumentative unit or text, and a target in terms of an issue or claim.
  - **Output.** Whether the respective text is *pro* or *con*.
    Sometimes, also classes such as *neutral* or *not relevant* are considered.

  **Target:** Sea patrols

  ”If you wanna hear my view, I think that the EU should allow sea patrols in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people.”

- **Myside bias classification**
  - The classification of an argumentative text as to whether it misses opposing viewpoints (myside bias) or not.
  - **Input.** An argumentative text.
  - **Output.** Whether the text has *myside bias* or *no myside bias*.

Not a standard task in computational argumentation, but important in argumentative writing support.
How good are humans in stance classification?

- What is the stance of the claims on the right to the issues on the left?

<table>
<thead>
<tr>
<th>Argument Assessment, Henning Wachsmuth</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;We should ban boxing.&quot;</td>
</tr>
<tr>
<td>&quot;It is sometimes right for the govern-</td>
</tr>
<tr>
<td>ment to restrict freedom of speech.&quot;</td>
</tr>
<tr>
<td>&quot;We should embrace multiculturalism.&quot;</td>
</tr>
<tr>
<td>&quot;Advertising is harmful.&quot;</td>
</tr>
<tr>
<td>&quot;Boxing remains the 8th most deadly sport.&quot;</td>
</tr>
<tr>
<td>&quot;Human rights can be limited or even pushed aside during times of national emergency.&quot;</td>
</tr>
<tr>
<td>&quot;Unity is seen as an essential feature of the nation and the nation-state.&quot;</td>
</tr>
<tr>
<td>&quot;Marketing creates consumerism and waste.&quot;</td>
</tr>
</tbody>
</table>

What makes the task challenging?

- Stance can be expressed without mentioning the issue.
- The contrastiveness of discussed concepts needs to be accounted for.
- Positive stance can be expressed with negative sentiment and vice versa.

But stance and sentiment polarity often correlate.
Overview of stance classification

- **How to model stance classification computationally?**
  - Standard text classification trained on texts for specific issues.
  - Relation-like classification with the issue as one input.

- **Common features** (Somasundaran and Wiebe, 2010, Hasan and Ng, 2013)
  - Bag-of-words. Distribution of words or word n-grams
  - Core vocab. Terms from subjectivity lexicons
  - POS. Distribution of part-of-speech tags
  - Discourse. Connectives and relations between units
  - Sentiment. Aspect-based or topic-directed polarity
    ... among many others

- **Specific stance classification approaches**
  - Exploit author knowledge in dialog (Ranade et al., 2013)
  - Exploit opposing views in dialog (Hasan and Ng, 2013)
  - Stance as sentiment and contrast of text and issue targets (Bar-Haim et al., 2017a)
  - Route kernels for stance based on overall structure (Wachsmuth et al., 2017f)


Stance as sentiment and contrast (Bar-Haim et al., 2017a)

- **Task**
  - Given any issue and any claim relevant to the issue, classify the stance of the claim on the issue.
  
  The issue is also supposed to have a claim-like phrasing.

- **Data**
  - 55 issues from iDebate, and 2394 claims from Wikipedia.
  - The target of each claim and its sentiment polarity (positive or negative) were manually annotated for training.

- **Approach in a nutshell**
  1. Identify the target of the issue and the claim.
  2. Classify the sentiment polarity towards each targets.
  3. Determine whether the targets are contrastive or not.
  4. Derive stance from sentiment and contrast.

  Actually, Bar-Haim et al. (2017a) start with the issue target and sentiment polarity given already.

<table>
<thead>
<tr>
<th>Issue</th>
<th>Advertising is harmful.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim</td>
<td>Marketing creates consumerism and waste.</td>
</tr>
</tbody>
</table>

\[
\text{claim sentiment} \times \text{contrastiveness} \times \text{issue sentiment} \approx \text{stance}
\]
Stance as sentiment and contrast: Approach

- **Identify targets** $t_c$ and $t_i$ of claim and issue
  - **Candidate targets.** Any noun phrase
  - **Features.** Position in parse tree, relation to sentiment phrase, Wikipedia title or not, ...
  - **Supervised classifier.** Logistic regression

- **Score polarities** $p(t_c)$ and $p(t_i)$ in $[-1,1]$
  - **Lexicon-based.** Find sentiment terms and polarity shifters from lexicons
  - **Scoring.** Based on distance to targets

- **Score contrastiveness** $c(t_c, t_i)$ in $[-1,1]$
  - **Features.** Polarity shifters, relatedness measures, Wikipedia headers, ...
  - **Supervised classifier.** Random forest.

- **Score stance** $s = p(t_c) \cdot c(t_c, t_i) \cdot p(t_i)$

$s$ can be thresholded to decide when to actually classify stance.

---

**Issue:** ”Advertising is harmful. “

**Claim:** ”Marketing creates consumerism and waste. “

**Issue:** ”Advertising is harmful. “ $-1$

**Claim:** ”Marketing creates consumerism and waste. “ $-1$

**Advertising** $\leftrightarrow$ **Marketing** $1$

$s = -1 \cdot 1 \cdot -1 = 1$
Stance as sentiment and contrast: Results

- **Evaluation** more in (Bar-Haim et al., 2017a)
  - **Data.** 25 issues (1039 claims) for training, 30 issues (1355 claims) for testing
  - **Baseline.** Claim classification with SVM using unigram and sentiment features
  - **Measure.** Accuracy at coverage depending on $s$ threshold (here 0.2–1.0)

<table>
<thead>
<tr>
<th>Approach</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.717</td>
<td>0.709</td>
<td>0.691</td>
<td><strong>0.668</strong></td>
<td>0.632</td>
</tr>
<tr>
<td>Sentiment only</td>
<td>0.770</td>
<td>0.749</td>
<td>0.734</td>
<td>0.632</td>
<td><strong>0.632</strong></td>
</tr>
<tr>
<td>Sentiment + contrast</td>
<td><strong>0.847</strong></td>
<td><strong>0.793</strong></td>
<td><strong>0.740</strong></td>
<td>0.632</td>
<td><strong>0.632</strong></td>
</tr>
</tbody>
</table>

- **Observations**
  - Reliable for confident cases, but does not beat baseline if all are classified.
  - The hardest cases are those where stance is expressed without sentiment.

- **Extended approach** (Bar-Haim et al., 2017b)
  - Automatic lexicon expansion and use of sentiment in surrounding context.

| Bar-Haim et al. (2017b) | 0.935 | 0.856 | 0.776 | 0.734 | 0.691 |
Overview of myside bias classification

How to model myside bias classification computationally?

• Conceptually, a standard text classification task.
• Argumentative structure naturally may be predictive for myside bias.

Approaches to myside bias classification

• Supervised classification of essays using various features (Stab and Gurevych, 2016)
• Route kernels for essay myside bias using overall structure (Wachsmuth et al., 2017f)
Supervised classification of myside bias  
(Stab and Gurevych, 2016)

- **Task**
  - Given a persuasive student essay, classify it as having myside bias or not.

- **Approach**
  - Polynomial SVM on up to six feature types:
    1. **Unigrams.** Word 1-grams
    2. **Dependency.** Triples from dependency tree
    3. **Production.** Rules from constituency tree
    4. **Opposition.** Presence of opposing words
    5. **Sentiment.** Lexicon-based overall sentiment
    6. **Relations.** Types of discourse relations

- **Data**
  - 402 essays, 251 w/ bias, 151 w/o bias

- **Results**
  - About three out of four cases correct.

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o Unigrams</td>
<td>0.733</td>
</tr>
<tr>
<td>w/o Dependency</td>
<td>0.765</td>
</tr>
<tr>
<td>w/o Production</td>
<td>0.760</td>
</tr>
<tr>
<td>w/o Opposition</td>
<td>0.736</td>
</tr>
<tr>
<td>w/o Sentiment</td>
<td>0.756</td>
</tr>
<tr>
<td>w/o Relations</td>
<td>0.757</td>
</tr>
<tr>
<td><strong>All features</strong></td>
<td><strong>0.755</strong></td>
</tr>
<tr>
<td><strong>Best set (1+3+4)</strong></td>
<td><strong>0.770</strong></td>
</tr>
<tr>
<td>Majority baseline</td>
<td>0.624</td>
</tr>
</tbody>
</table>
Continued: Stance and bias

I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

a) Introduction

b) Stance and bias

c) Schemes and fallacies

d) Quality in theory

e) Absolute and relative quality assessment

f) Objective and subjective quality assessment

g) Conclusion
The death penalty is a legal means that as such is not practicable in Germany.

For one thing, inviolable human dignity is anchored in our constitution, and further no one may have the right to adjudicate upon the death of another human being.

Even if many people think that a murderer has already decided on the life or death of another person, this is precisely the crime that we should not repay with the same.

(Peldszus and Stede, 2016)
Route kernels for stance and bias  
(Wachsmuth et al., 2017f)

§ Task
• Given a monological argumentative text, classify stance and myside bias without knowing the issue discussed.

§ Hypothesis
• The overall structure of argumentative texts is decisive for stance and myside bias.

§ Research questions
1. How to jointly model sequential and hierarchical overall argumentation?
2. What model has most impact on the two tasks?

§ Approach in a nutshell
• Start from argumentative structure of a text.
• Model overall structure with route kernels, a variation of tree kernels.
• Classify stance and myside bias based on overall structure.
Route kernels for stance and bias: Tasks and data

- **Myside bias on AAE-v2**
  (Stab and Gurevych, 2016)
  - 402 persuasive student essays
  - 15.1 units per text, proprietary argument model
  - 251 myside bias, 151 no myside bias

- **Stance on Arg-Microtexts**
  (Peldszus and Stede, 2016)
  - 112 short argumentative texts
  - 5.1 units per text, model of Freeman (2011)
  - 46 pro stance, 42 con stance, 24 unlabeled

- **For comparison: Genre on Web Discourse**
  (Habernal and Gurevych, 2015)
  - 340 argumentative web texts
  - 3.4 units per text, modified model of Toulmin (1958)
  - 216 comments, 46 blog posts, 73 forum posts, 5 articles
Route kernels for stance and bias: A unified model

- **Map specific models to unified model**
  - Order nodes according to position.
  - Encode stance towards parent as node label.
  - Model relations between node *pairs* only.
  - The root implicitly defines main claim.

- **Pros and cons**
  - Sequential structure captured
  - Same analyses on all corpora
  - Comparisons across corpora
  - Simpler argument mining (hypothesized)
  - Partly less expressive

- **In this talk, only unified model**
  - For experiments with specific models, see paper. (Wachsmuth et al., 2017f)
Route kernels for stance and bias: Visualizing structures

- **Myside bias on AAE-v2**
  - no myside bias → High impact of modeling hierarchical structure?

- **Stance on Arg-Microtexts**
  - con stance → Medium impact of modeling both types of structure?

- **Genre on Web Discourse**
  - forum posts → Low impact of modeling sequential structure if any?
Background: Route kernels  (see also lecture part V)

- **Kernel methods in machine learning** (recap)
  - Kernels represent instances in a task-specific implicit feature space.
  - Kernel functions compute similarities used by classifiers, such as SVMs.
  - **Tree kernels** capture hierarchical structures.

- **Route kernels**
  - **Route kernels** capture both sequential and hierarchical structure.  (Aiolli et al., 2009)
  - A tree kernel with edge labels, indicating node positions relative to their siblings.

- **Adapted route kernel for argumentation**
  - Modeling of all paths starting from the root of a tree.
  - A polynomial kernel "combines“ paths to capture full overall structure.
  - Positions are relative to parent node.
Route kernels for stance and bias: Approach

- **Overall structure as a positional tree**
  - A tree $T = (V, E)$ where nodes in $V$ represent argumentative units and edges in $E$ a relation between two units.
  - **Node labels.** Each node is labeled as *pro* or *con*.
  - **Edge labels.** Node position in a text relative to parent node.

- **Kernel function for overall structure**
  - Let two trees $T = (V, E)$ and $T' = (V', E')$ be given.
  - The similarity of the trees is defined as:

$$K_{\xi \pi}(T, T') = \left( \sum_{v \in V} \sum_{v' \in V'} \delta(\xi(v), \xi(v')) \cdot \delta(\pi(v), \pi(v')) \right)^d$$

1 for identical paths, 0 otherwise
Node label path from root to $v$
Edge label path from root to $v$
Degree of polynomial (2 best in experiments)

Sum all pairs of paths of the two trees
Normalization over maximum possible score
Route kernels for stance and bias: Evaluation

Overall argumentation approaches

- **frequencies**
  - linear kernel

- **sequences**
  - subsequence kernel

- **hierarchies**
  - tree path kernel

- **routes**
  - adapted route kernel

Baseline approaches

- **majority**
  - always majority class

- **pos**
  - linear kernel

- **tokens**
  - linear kernel

Experiments on ground-truth argument corpora

- SVM for each kernel evaluated in repeated 10-fold cross-validation.
- Hyperparameters of SVM tuned on training set with balanced class weights.
Route kernels for stance and bias: Results

- **Myside bias accuracy on AAE-v2**

  - 77.0 (Stab and Gurevych, 2016)
  - 62.4 \( \pm \) 63.3 \( \pm \) 70.5
  - 83.4 frequencies
  - 87.9 sequences
  - 97.1 hierarchies
  - 95.8 routes
  - 97.1 (best blue + best red)

- **Stance accuracy on Arg-Microtexts**

  - 52.3 \( \pm \) 58.8 \( \pm \) 65.2
  - 49.7 frequencies
  - 52.2 sequences
  - 59.8 hierarchies
  - 66.7 routes
  - 69.8 (best blue + best red)

- **Genre accuracy on Web Discourse**

  - 64.5 \( \pm \) 74.0 \( \pm \) 75.6
  - 62.6 frequencies
  - 64.5 sequences
  - 58.1 hierarchies
  - 53.4 routes
  - 75.7 (best blue + best red)
Stance and bias: Discussion

- **Effective stance and myside bias classification**
  - Approaches to stance achieve an accuracy < 0.8 in most settings.
  - Stance is subjective, so a notably higher accuracy may not be feasible.
  - Too few approaches to myside bias exist to make a conclusive statement.

- **Impact of argumentative structure**
  - At least for entire argumentative texts, modeling overall structure is important.
  - Theoretically, modeling hierarchical structure “solves” myside bias.
  - Practically, the impact depends on the effectiveness of argument mining.

- **Stance classification, an independent task**
  - Stance classification is also studied apart from computational argumentation.
  - Not in all literature on the topic, arguments are considered explicitly.
  - Still, the notion of stance implies a controversial and, thus, argumentative context.
Next section: Schemes and fallacies

I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

a) Introduction
b) Stance and bias
c) Schemes and fallacies
d) Quality in theory
e) Absolute and relative quality assessment
f) Objective and subjective quality assessment
g) Conclusion
What is scheme and fallacy detection?

- **Scheme classification**
  - The assignment of an argumentation scheme to an argument from a given scheme set.
  - **Input.** An argument, usually with annotated structure.
  - **Output.** The argument with assigned scheme.

- **Fallacy detection**
  - The identification of arguments being a fallacy of a type from a set of types.
  - **Input.** An argument, possibly with annotated structure.
  - **Output.** Whether or not the argument is a fallacy of a certain type.
Example: Correct or fallacious argumentation scheme?

- How good are humans in analyzing schemes?
  - Is the following example a correct instance of argument from position to know?
  - Check the critical questions below.

<table>
<thead>
<tr>
<th>Conclusion</th>
<th>Cigarettes are not addictive.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major premise</td>
<td>James W. Johnston (the CEO of RJ Reynolds Tobacco Company) is an expert on tobacco.</td>
</tr>
<tr>
<td>Minor premise</td>
<td>Johnston testified before Congress that tobacco is not an addictive substance.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conclusion</th>
<th>A is true.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major premise</td>
<td>Source E is in a position to know about things in a subject domain S with proposition A.</td>
</tr>
<tr>
<td>Minor premise</td>
<td>E asserts that A is true (in domain S).</td>
</tr>
</tbody>
</table>

- Critical questions
  - Is Johnston in a position to know about cigarette addictiveness? yes
  - Did Johnston assert that it’s true that cigarettes are addictive? (yes)
  - Is Johnston a reliable source? no!
Overview of scheme and fallacy detection

- **Schemes and fallacies in argumentation**
  - Core ways of describing how the reasoning in an argument works or how it is flawed.

- **How to model scheme classification?**
  - Conceptually, a multi-class classification task.
  - The few approaches so far realized it as a one-vs.-all or one-vs.-one task.

- **How to model fallacy detection?**
  - Conceptually, the same.
  - The few works dealing with fallacies so far consider only specific types of fallacies.

- **Selected approaches to schemes and fallacies**
  - **Scheme classification with tailored features** (Feng and Hirst, 2011; Lawrence and Reed, 2016)
  - **Ad-hominem argument detection on the web** (Habernal et al., 2018)

... along with several human annotation studies of schemes and fallacies
**Classifying schemes with tailored features**  
(Feng and Hirst, 2011)

**Task**
- Given the premises and conclusion of an argument, assign one scheme from a set of given schemes.

**Research question**
- How visible is the scheme of an argument in its text and its structure?

**Data**
- The Araucaria corpus with 658 mixed argumentative texts, annotated for Walton’s argumentation schemes. (Walton et al., 2008)
- Only the five most frequent schemes considered. (see next slide)

**Approach in a nutshell**
- Compute features tailored to argumentation schemes.
- Classify schemes with standard supervised learning.
### Argument from verbal classification

**Minor pr.**  
*a* has property *F*.

**Major pr.**  
For all *x*, if *x* has property *F*, then *x* can be classified as having property *G*.

**Conclusion**  
*a* has property *G*.

### Argument from cause to effect

**Minor pr.**  
In this case, *A* occurs.

**Major pr.**  
Generally, if *A* occurs then *B* will occur.

**Conclusion**  
*B* will occur.

### Practical reasoning

**Minor pr.**  
I have a goal *G*.

**Major pr.**  
Carrying out this action *A* is a means to realize *G*.

**Conclusion**  
I ought to carry out *A*.

### Argument from consequences

**Major pr.**  
If *A* is done, good (bad) consequences will occur.

**Conclusion**  
*A* should (not) be done.
Classifying schemes with tailored features: Examples

”Censorship is [...] the hallmark of an authoritarian regime. For example, one of Nazi Germany’s first acts was to burn all the books [...] which offended their sensibilities, beliefs, and values.“

”[You shouldn’t build a] road into the heart of the Amazon. [This] will likely result in commercialization and destruction of the valuable Amazon habitat.“

”If we want to stop the counterfeit products, we have to make new products more unique.“

”[The] Iraq war [is] illegal. There is no law [...] that sanctions attacks on guys because you have good reason to believe they are bad, and could threaten you.

“The crisis [of a party] is likely to have an effect on other opposition parties. The public's disappointment with the [party] will lead to an erosion of confidence in the opposition.“

(credit to Jonas Bülling for these examples)
Classifying schemes with tailored features: Approach

- **Approach**
  - C4.5 decision tree for supervised classification.
  - Feature engineering for all argumentation schemes.

- **Features tailored to all schemes**
  - **Location.** Relative positions and distances of premises and conclusion
  - **Statistics.** Premise/conclusion length ratio, number of premises
  - **Structure.** Linked or convergent (given in ground-truth!)

- **Features tailored to specific schemes**
  - **Cue phrases,** e.g., ”for example“, ”result“, ”want“
  - **Indicating patterns,** e.g., causal WordNet relations
  - **Sentiment.** Positive and negative words
  - **Word similarity** between central words in premise and conclusion
Classifying schemes with tailored features: Results

- **10-fold cross-validation**
  - **One-against-others.** 50% target scheme, 50% others (for all schemes)
  - **One-against-all.** 50% scheme A, 50% scheme B (for all scheme pairs)

- **Results (accuracy)**

<table>
<thead>
<tr>
<th>Features</th>
<th>Acc.</th>
<th>Example</th>
<th>Practical reas.</th>
<th>Cause to effect</th>
<th>Consequ.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal classific.</td>
<td>0.632</td>
<td>0.860</td>
<td>0.983</td>
<td>0.856</td>
<td>0.642</td>
</tr>
<tr>
<td>From consequ.</td>
<td>0.629</td>
<td>0.869</td>
<td>0.979</td>
<td>0.867</td>
<td></td>
</tr>
<tr>
<td>Cause to effect</td>
<td>0.704</td>
<td>0.806</td>
<td>0.942</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical reas.</td>
<td>0.908</td>
<td>0.931</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From example</td>
<td>0.906</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Observations**
  - **High effectiveness for some schemes, but two schemes were confused often.**
    Both less training data and less clear linguistic indicators may be reasons.
  - **Ultimately, focusing on five schemes limits the applicability of the approach.**
Continued: Schemes and fallacies

I. Introduction to computational argumentation
II. Basics of natural language processing
III. Basics of argumentation
IV. Argument acquisition
V. Argument mining
VI. Argument assessment
VII. Argument generation
VIII. Applications of computational argumentation
IX. Conclusion

a) Introduction
b) Stance and bias
c) Schemes and fallacies
d) Quality in theory
e) Absolute and relative quality assessment
f) Objective and subjective quality assessment
g) Conclusion
Ad-hominem arguments on the web (Habernal et al., 2018)

That's an ad hominem fallacy Calvin!!

"YOU'RE FACE IS AN AD HOMINEM!!"
Ad-hominem arguments on the web: Task and data

What is an ad-hominem argument?
- An argument that attacks the author of an argument, not the argument itself.
- According to a study, 20% of all news comments are uncivil. (Coe et al., 2014)

Research questions
- How well can ad-hominem be identified automatically?
- What triggers ad-hominem in discussions?

Data
- 2M posts from Reddit ChangeMyView
- 3866 posts (0.2%) contain ad-hominem arguments
  Ad-hominem is deleted by moderators, but was made available to Habernal et al. (2018).

Reddit ChangeMyView (CMV)
- An opinion poster (OP) states a view.
- Others argue for the opposite.
- OP gives Δ to convincing posts.

Deltas from OP
CMV: Trump has done nothing of substance since being elected to office.
This is kind of a counter to the other post made recently about Trump being a great president.
He pointed out things like the economy, which was growing before he was in office, it is still growing.
Ad-hominem arguments on the web: Examples

"Reading comprehension is your friend"  "Ever have discussions with narcissistic idiots on the internet? They are so tiring"  "You still refuse to acknowledge that you used a strawman argument against me"

"Thank you so much for all your pretentious explanations"  "To say that people intrinsically understand portion size is idiotic."  "You started with a fallacy and then deflected."

"boy"  "Did you even read this?"  "Read what I posted before acting like a pompous ass"

"Again, how old are you?"  "You're making the claims, it's your job to prove it. Don't you know how debating works?"  "Please don't waste peoples time pretending to know what you're talking about"

"You have no capability to understand why"  "Wow. Someone sounds like a bit of an anti-semite"  "Do you even know what you're saying?"

"How can you explain that? You can't because it will hurt your feelings to face reality"  "You're just a dishonest troll"  "Your second paragraph is fairly idiotic"

"Willful ignorance is not something I can combat"  "You're trash at debating."  "Possible lie any harder?"

"You're too dishonest to actually quote the verse because you know it's bullshit"  "If you can't grasp the concept, I can't help you"  "Can you also use Google?"

"You're using troll tactics"  "sir"  "Your just an asshole"
Ad-hominem arguments on the web: Identification

- **Distribution of ad-hominem on CMV**
  - 75% threads with $\leq 2$ ad-hominem (but some with $>50$)
  - 49% threads stop after ad-hominem
  - 66% ad-hominem "out of the blue" (but one after 57 posts)
  - 23% ad-hominem against OP
  - 12% ad-hominem from OP

- **Types of ad-hominem on CVM**
  - Ad-hominem were annotated in 400 arguments by 7 crowdworkers.
  - 15 types were identified manually in their annotations.

- **Identification of ad-hominem**
  - **Manual.** 100 balanced arguments (50 ad-hominem) were classified by 6 workers.
  - **Computational.** 7242 balanced arguments were classified by two neural classifiers (Bi-LSTM, CNN).

Accuracy

- Manual: 0.88
- Bi-LSTM: 0.78
- CNN: 0.81

Types of ad-hominem:
- Vulgar insult: 31%
- Illiteracy insult: 13%
- Condescension: 7%
- Ridiculing and sarcasm: 7%
- "Idiot" insults: 7%
- Accusation of stupidity: 4%
- Denial of no arguing skills: 4%
Background: Attention in neural networks (see also lecture part V)

- **Attention**
  - A mechanism of RNNs (especially LSTMs) that quantifies interdependencies between different parts of input and output.
  - The key idea is to retain all hidden states of an input while creating the output.
  - This allows learning to focus on input parts relevant to the output.

- **Self-attention**
  - Quantification of interdependencies within the input only.
    - In NLP, usually this means between the words of a sentence.
  - An RNN with self-attention can provide weight values that represent the relevance it gives to different parts of an input.
Ad-hominem arguments on the web: Triggers

- **Prediction of ad-hominem**
  - Self-attentive LSTM trained on 2852 argument 3-tuples.
  - **Accuracy.** 0.72
  - Manual attention analysis:

  (OOV_comment_begin) If only you would n’t rely on [fallacious](http://OOV) arguments( http://OOV) to make your point. So no, I do n’t realize how stupid and naive I am. All I ’ve realized is that you are n’t actually prepared to have an actual discussion.

  (OOV_comment_begin) What god do you believe in? And it ’s not a fallacy when it ’s very comparable to the most popular gods.

- **Terms with much attention**
  - Mostly topic-independent rhetorical devices
  - A few loaded keywords, such as ”rape” or ”racist”
  - Partly meta about argumentation

---

**vulgar intensifiers**  "... the fuck..."  
**direct imperatives**  "You should..."

**missing evidence**  "unsupported claims!"

**bad argumentation**  "You’re grasping at straws"

...
Effective scheme and fallacy classification

- Some schemes are reflected in words, others require deeper understanding.
- Many schemes have never been approached so far.
- Finding ad-hominem seems doable, but this may not hold for other fallacies.

Limited computational research

- While extensively studied in theory, computational research on schemes and fallacies is rare so far.
  - For schemes, one reason lies in the complexity of getting ground-truth data.
  - The high number of less frequent schemes is a particular problem in this regard.
- For fallacies, their detection is often just hard, even for humans.

Why studying schemes and fallacies?

- Knowing the scheme means to understand how an argument reasons.
- Schemes clarify what has been left implicit, allowing to find enthymemes.
- Fallacies define a way of judging quality: a good argument should not be fallacious — although there are exceptions. (Hamblin, 1970)
Next section: Quality in theory

I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

a) Introduction
b) Stance and bias
c) Schemes and fallacies
d) Quality in theory
e) Absolute and relative quality assessment
f) Objective and subjective quality assessment
g) Conclusion
## Survey of existing research

Based on Wachsmuth et al. (2017b)

<table>
<thead>
<tr>
<th>Argumentation theory</th>
<th>Assessment approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toulmin (1958)</td>
<td>Walton et al. (2008)</td>
</tr>
<tr>
<td>Persing and Ng (2013)</td>
<td>Persing and Ng (2015)</td>
</tr>
<tr>
<td>Feng et al. (2014)</td>
<td>Hoeken (2001)</td>
</tr>
<tr>
<td>Persing and Ng (2014)</td>
<td>Park et al. (2015)</td>
</tr>
<tr>
<td>Persing et al. (2010)</td>
<td>Tan et al. (2016)</td>
</tr>
<tr>
<td>Habernal and Gurevych (2016)</td>
<td></td>
</tr>
</tbody>
</table>
Three main quality aspects (recap)

**Logic**

An argument is *cogent* if its premises are *relevant* to its conclusion, individually *acceptable*, and together *sufficient* to draw the conclusion. “

Blair (2012)

**Dialectic**

A dialectical discussion derives its *reasonableness* from a dual criterion: *problem validity* and *intersubjective validity*. “

van Eemeren (2015)

**Rhetoric**

In making a speech, one must study three points: the *means of producing persuasion*, the *style* or language to be used, and the proper *arrangement* of the various parts. “

Aristotle (2007)
Unification of views

Focus on theory
- validity
- soundness
- level of support
- amount of evidence
- well-formedness
- global coherence
- thesis clarity
- prompt adherence
- evaluable
- appropriateness of style

Focus on accepted
- premise acceptability
- local/probative relevance
- premise sufficiency
- amount of rebuttal
- strength
- clarity of style
- credibility
- evaluability
- effect veness

Prefer general
- intersubjective acceptability
- cogency
- amount of rebuttal
- rational acceptability
- reason-ability
- dialectical sufficiency
- dialectical acceptability
- arrangement
- organization
- emotional appeal
- persuasiveness

Unify names
- argument acceptability
- global/dialectical relevance
- argument relevance
- prominence
- satisfactoriness
- convincingness
- arrangement
- organization
- emotional appeal
- persuasiveness

Argumentation quality
A taxonomy of argumentation quality

- **Local acceptability**
  - Level of support: Braunstain et al. (2016)
  - Amount of evidence: Rahimi et al. (2014)
  - Sufficiency: Stab and Gurevych (2017)

- **Local relevance**
  - Thesis clarity: Persing and Ng (2013)
  - Prompt adherence: Persing and Ng (2014)
  - Global coherence: Feng et al. (2014)

- **Cogency**
  - Evaluability: Park et al. (2015)

- **Reason-ability**
  - Global acceptability: Cabrio and Villata (2012)
  - Argument relevance: Wachsmuth et al. (2017a)

- **Global relevance**
  - Organization: Persing et al. (2010)
  - Rahimi et al. (2015)

- **Global acceptability**

- **Global sufficiency**

- **Effectiveness**
  - Credibility: Persing and Ng (2015)
  - Persuasiveness: Tan et al. (2016)
  - Winning side: Wang et al. (2016)
  - Convincingness: Habernal and Gurevych (2016)

- **Arrangement**
  - Prompt adherence: Persing and Ng (2014)

- **Clarity**
  - Level of support: Braunstain et al. (2016)
  - Evaluability: Park et al. (2015)

- **Appropriateness**

- **Emotional appeal**

- **Argument strength**

- **Local sufficiency**

- **Global sufficiency**

- **Local relevance**

- **Global relevance**

Argument Assessment, Henning Wachsmuth
Quality dimensions in the taxonomy

- **A cogent argument.** Has acceptable, relevant, and sufficient premises.
  - Local acceptability. The premises are worthy being believed as true.
  - Local relevance. The premises are relevant to the conclusion.
  - Local sufficiency. The premises are sufficient to draw the conclusion.

- **Effective argumentation.** Persuades the target audience.
  - Credibility. Makes the authors worthy of credence.
  - Emotional appeal. Makes the audience open to be persuaded.
  - Clarity. Is linguistically clear and as simple as possible.
  - Appropriateness. Linguistically matches the audience and issue.
  - Arrangement. Presents content in the right order.

- **Reasonable argumentation.** Is acceptable, relevant, and sufficient.
  - Global acceptability. Worthy to be considered in the way stated.
  - Global relevance. Contributes to resolution of issue.
  - Global sufficiency. Adequately rebuts potential counterarguments.

Notice: cogency also adds to effectiveness, and cogency and effectiveness also add to reasonableness.
The Dagstuhl-15512 ArgQuality corpus

- **Corpus based on the taxonomy**
  - 320 debate portal arguments (Habernal and Gurevych, 2016)
  - 10 per issue/stance pair
  - 3 annotators per argument
  - Score from [1,3] for all 15 dimensions

- **Agreement**
  - Krippendorff’s $\alpha$ limited
  - Majority agreement very high

- **Correlations**
  - Overall quality correlates most with reasonableness (.86), cogency (.84), and effectiveness (.81).
  - Several intuitive correlations exist.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Mean</th>
<th>$\alpha$</th>
<th>Maj.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cogency</strong></td>
<td>1.6</td>
<td>.44</td>
<td>92%</td>
</tr>
<tr>
<td>local acceptability</td>
<td>1.9</td>
<td>.46</td>
<td>91%</td>
</tr>
<tr>
<td>local relevance</td>
<td>2.3</td>
<td>.47</td>
<td>92%</td>
</tr>
<tr>
<td>local sufficiency</td>
<td>1.5</td>
<td>.44</td>
<td>93%</td>
</tr>
<tr>
<td><strong>effectiveness</strong></td>
<td>1.4</td>
<td>.45</td>
<td>94%</td>
</tr>
<tr>
<td>credibility</td>
<td>1.7</td>
<td>.37</td>
<td>96%</td>
</tr>
<tr>
<td>emotional appeal</td>
<td>1.9</td>
<td>.26</td>
<td>94%</td>
</tr>
<tr>
<td>clarity</td>
<td>2.1</td>
<td>.35</td>
<td>90%</td>
</tr>
<tr>
<td>appropriateness</td>
<td>2.1</td>
<td>.36</td>
<td>88%</td>
</tr>
<tr>
<td>arrangement</td>
<td>1.8</td>
<td>.39</td>
<td>93%</td>
</tr>
<tr>
<td><strong>reasonableness</strong></td>
<td>1.6</td>
<td>.50</td>
<td>96%</td>
</tr>
<tr>
<td>global acceptability</td>
<td>1.9</td>
<td>.44</td>
<td>95%</td>
</tr>
<tr>
<td>global relevance</td>
<td>2.0</td>
<td>.42</td>
<td>90%</td>
</tr>
<tr>
<td>global sufficiency</td>
<td>1.2</td>
<td>.27</td>
<td>98%</td>
</tr>
<tr>
<td><strong>overall quality</strong></td>
<td>1.6</td>
<td>.51</td>
<td>94%</td>
</tr>
</tbody>
</table>
Next section: Absolute and relative quality assessment

I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

a) Introduction
b) Stance and bias
c) Schemes and fallacies
d) Quality in theory
e) Absolute and relative quality assessment
f) Objective and subjective quality assessment
g) Conclusion
What is argumentation quality assessment?

- **Argumentation quality assessment**
  - Identification of indisputable flaws or requirements of argumentation.
  - Judgment about a specific quality dimension.
  - Determination whether argumentation successfully achieves its goal.

  ![linguistically clear? effective in persuading?](image)

- **Observations**
  - **Choice of comparison.** Dimensions can be assessed *absolutely* or *relatively*.
  - **Subjectivity.** Perceived quality depends on the view of the reader/audience. (and maybe also on the author/speaker)

- **How to approach quality assessment?**
  - **Input.** Argumentative text, metadata (e.g., author), external knowledge, ...  
  - **Techniques.** Supervised classification/regression, graph-based analyses, ...

  Multiple example approaches discussed below.
Importance of quality assessment

- Why assessing argumentation quality?
  - Mining arguments and understanding the reasoning is not enough in practice.
  - For successful argumentation, we need to choose the "best" arguments.
  - Critical for any application of computational argumentation.

  "In some sense, the question about the quality of an argument is the ‘ultimate’ one for argumentation mining."

  (Stede and Schneider, 2018)

- Example applications
  - Argument search. What argument to rank highest?
  - Writing support. How good is an argumentative text, what flaws does it have?
  - Automatic decision making. Which arguments outweigh which others?
Absolute vs. relative assessment

How to assess a quality dimension computationally?

- **Absolute rating.** Assignment of a score from a predefined scale. Typical scales: Integers (possibly with half-points): 1–3, 1–4, 1–5, 1–10, -2–2, ... Real valued: [0,1], [-1,1]
- **Relative comparison.** Given two instances, which of them is better.

"If you wanna hear my view, I think that the EU should allow sea patrols in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people."

Observations

- Both allow for ranking assessed instances.
- Absolute ratings entail relative comparisons and they imply a maximum and minimum.

Absolute vs. relative assessment

- A relative assessment is often much easier.
- Still, absolute ratings are widely spread and often work well.

"It‘s the main job of the EU to save people‘s lives, no matter whether they belong here. “
Problem
• Can we predict whether an argument(ation) is good (cogent, effective, ...)?
• Can we rate how good it is?

Main idea
• See quality assessment as a standard classification or regression task.
• Learn what linguistic feature or metadata speaks for quality?

Existing approaches
• Persuasiveness. Prediction based on interaction of participants (Tan et al., 2016)
• Organization. Assessment based on quality-oriented features (Persing et al., 2010)
  Analog approaches for thesis clarity, prompt adherence, and argument strength (Persing and Ng, 2013–2015).
• Amount of evidence. Count of evidence supporting conclusion (Rahimi et al., 2014)
... among other approaches
Absolute quality rating: Dimensions covered here

- **local acceptability**
- **local relevance**
- **cogency**
- **effectiveness**
- **appropriateness**
- **credibility**
- **clarity**
- **organization**
- **arrangement**
- **argument strength**
- **persuasiveness**
- **winning side**
- **emotional appeal**
- **level of support**
- **amount of evidence**
- **sufficiency**
- **thesis clarity**
- **prompt adherence**
- **global coherence**
- **evaluability**
- **argument acceptability**
- **argument prominence**
- **argument relevance**
- **global sufficiency**
- **global relevance**
- **global acceptability**
- **argumentation quality**

- **Braunstain et al. (2016)**
- **Rahimi et al. (2014)**
- **Stab and Gurevych (2017)**
- **Persing and Ng (2013)**
- **Persing and Ng (2014)**
- **Feng et al. (2014)**
- **Park et al. (2015)**
- **Rahimi et al. (2015)**
- **Persing et al. (2010)**
- **Boltužić and Šnjajder (2015)**
- **Wachsmuth et al. (2017a)**
- **Cabrio and Villata (2012)**
- **Feng et al. (2014)**
- **Park et al. (2015)**
- **Rahimi et al. (2015)**
- **Braunstain et al. (2016)**
- **Rahimi et al. (2015)**
- **Stab and Gurevych (2017)**
- **Persing and Ng (2013)**
- **Persing and Ng (2015)**
- **Tan et al. (2016)**
- **Wang et al. (2016)**
- **Zhang et al. (2016)**
- **Habernal and Gurevych (2016)**
Rating quality based on interaction  (Tan et al., 2016)

- Task
  - In a discussion, what will persuade someone open to be persuaded?

- Approach
  - Analysis of correlations between linguistic, interaction, and meta-discussion features with persuasion.
  - Prediction based on features as to whether persuasion will happen.

- Data
  - 20k+ discussions from Reddit ChangeMyView.
  - Discussion. An opinion poster (OP) states a view, others argue against, OP gives Δ to convincing arguments.

- Selected results
  - Accuracy. 69% in balanced setting.
  - Insights. Some interactions and many participants help; appropriate style, not to similar to OP’s style most persuasive.
Rating quality based on mining  (Wachsmuth et al., 2016)

- **Task**
  - Given a persuasive essay, score argumentation-related quality dimensions.

- **Dimensions**  (Persing et al., 2010; Persing and Ng, 2013–2015)
  - **Organization.** How well is the argumentation arranged?
  - **Thesis clarity.** How easy to understand is the thesis?
  - **Prompt adherence.** How close does the essay stay to the issue?
  - **Argument strength.** How strong is the argument made for the thesis?

- **Research question**
  - Can we leverage argument mining to assess the argumentation quality of persuasive essays?

- **Data**
  - 800–1003 essays with scores in [1,4] annotated for each dimension
Background: Argumentative writing support

1. Essay (input)

2. Argumentative structure

3. Essay scoring

4. Argumentation quality

5. Suggestion (output)

The new state of the art in argument mining for argumentation quality assessment is the first study of argument mining of suggestions for improvements based on the output of mining for two quality dimensions.

Our approach is based on the output of mining for paragraph-level argumentation into essay argumentation.

The benefit of this structure has rarely been evaluated.

Novel feature types are used for argumentative discourse units (ADUs) with sentence-level argumentative discourse units (ADUs).

Mean squared errors (mean squared errors in green significantly improve the state of the art with a confidence of over 90%)

<table>
<thead>
<tr>
<th>ADU Type</th>
<th>Our Approach</th>
<th>State-of-the-art baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flows</td>
<td>0.228</td>
<td>0.481</td>
</tr>
<tr>
<td></td>
<td>0.257</td>
<td>0.259</td>
</tr>
</tbody>
</table>

For poets and literate people of yore it was a common idea to transcend reality or to go beyond it by using their imagination. Whereas now in the 21st century and in "The Day the Earth Stood Still", the idea of using imagination to create worlds that exist only in our minds becomes one with Nature, and cruise wherever he wants using his imagination. Whereas now in the 21st century and in "The Day the Earth Stood Still", the idea of using imagination to create worlds that exist only in our minds becomes one with Nature, and cruise wherever he wants using his imagination.

For example, if we indulge in entertaining the idea of the film "The Day the Earth Stood Still", we can easily imagine a world where humans can become one with Nature, and cruise wherever he wants using his imagination. Whereas now in the 21st century and in "The Day the Earth Stood Still", the idea of using imagination to create worlds that exist only in our minds becomes one with Nature, and cruise wherever he wants using his imagination.

Some people say that in our modern world, dominated by science and technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?

Our approach involves the use of standard features on argumentative discourse units (ADUs) for the learning of mining...

Experimental set-up

Our approach of Flows: Sentiment flow patterns...

For example, if we indulge in entertaining the idea of the film "The Day the Earth Stood Still", we can easily imagine a world where humans can become one with Nature, and cruise wherever he wants using his imagination. Whereas now in the 21st century and in "The Day the Earth Stood Still", the idea of using imagination to create worlds that exist only in our minds becomes one with Nature, and cruise wherever he wants using his imagination.

With the touch of a button could unleash Armageddon...
Rating quality based on mining: Mining and analysis

**Mining**

- **Task.** Classify sentence-level units as thesis, conclusion, premise, or none
- **Data.** AAE corpus (Stab and Gurevych, 2014a)
- **Approach.** SVM with different standard features

<table>
<thead>
<tr>
<th>Approach</th>
<th>Accuracy</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority baseline</td>
<td>52.5</td>
<td>36.1</td>
</tr>
<tr>
<td>Stab and Gurevych (2014b)</td>
<td>77.3</td>
<td>72.6</td>
</tr>
<tr>
<td><strong>Mining approach</strong></td>
<td>74.5</td>
<td><strong>74.5</strong></td>
</tr>
</tbody>
</table>

**Analysis**

- **Task.** Compute most common unit role flows
- **Data.** All paragraphs of all 6085 essays in ICLE corpus (Granger et al., 2009)

<table>
<thead>
<tr>
<th>Unit role flows</th>
<th>Average</th>
<th>First</th>
<th>Last</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conclusion, Premises</td>
<td>25.1%</td>
<td>–</td>
<td>13.1%</td>
</tr>
<tr>
<td>Conclusion, Premises, Conclusion</td>
<td>17.0%</td>
<td>–</td>
<td>27.2%</td>
</tr>
<tr>
<td>None, thesis</td>
<td>3.4%</td>
<td>25.9%</td>
<td>–</td>
</tr>
<tr>
<td>Premises, Conclusion</td>
<td>2.9%</td>
<td>–</td>
<td>2.7%</td>
</tr>
</tbody>
</table>
Prompt

"Some people say that in our modern world, dominated by science and technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?"

Essay

"If we take a look back in time we are in a position to see man dreaming, philosophizing and using his imagination of whatever comes his way. We see man transcending his ego I a way and thus becoming a God - like figure. And by putting down these sacred words, what is taking shape in my mind is the fact that using his imagination Man is no longer this organic and material substance like his contemporary counterpart who is putting his trump card on science, technology and industrialization but Man is a way transcends himself through his imagination.

For instance, if we take into account the Renaissance or Romantic periods of mankind and close our eyes we could see Shakespeare applying his imagination in the fancy world of his comedies: elf and nymphs circling the stage making it a dream that will lost forever in our minds. We could even hear their high-pitched weird chuckle piercing with a gentle touch our ears, but "open those eyes that must eclipse the day" and you'll see the high-tech wiping out every trace of the human elevated spirit that have dominated over the previous centuries. What we see now is "deux aux machina" or the fake "God from the machine“ who with the touch of a button could unleash Armageddon.

As a logical conclusion to my essay I would like to put only one thing: Wouldn't it be better if imagination makes the world go round". If I was to answer this question, the answer would be positive, but given the aquisitive or consumer society conditions we live in let's make a match between imagination and science. It would be somewhat more realistic."

Rating quality based on mining: Example essay

| Organization | 3.0 |
| Thesis clarity | 2.0 |
| Prompt adherence | 4.0 |
| Argument strength | 2.0 |
Rating quality based on mining: Approach and results

### Assessment

- **Approach.** SVM based on argument-specific and standard features

- **Evaluation.** Mean squared error for each quality dimension

<table>
<thead>
<tr>
<th>Approach</th>
<th>Organization</th>
<th>Clarity</th>
<th>Adherence</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average baseline</td>
<td>0.349</td>
<td>0.469</td>
<td>0.291</td>
<td>0.266</td>
</tr>
<tr>
<td>Persing et al. (2010–2015)</td>
<td>0.175</td>
<td><strong>0.369</strong></td>
<td><strong>0.197</strong></td>
<td>0.244</td>
</tr>
<tr>
<td><strong>Assessment approach</strong></td>
<td><strong>0.164</strong></td>
<td>0.425</td>
<td>0.216</td>
<td><strong>0.226</strong></td>
</tr>
<tr>
<td>— Unit role flows</td>
<td>0.234</td>
<td>0.461</td>
<td>0.247</td>
<td>0.242</td>
</tr>
<tr>
<td>— Unit role composition</td>
<td><strong>0.194</strong></td>
<td>0.457</td>
<td>0.239</td>
<td>0.239</td>
</tr>
<tr>
<td>— Function flows</td>
<td>0.220</td>
<td>0.478</td>
<td>0.255</td>
<td>0.251</td>
</tr>
<tr>
<td>— Content features</td>
<td>0.336</td>
<td><strong>0.425</strong></td>
<td>0.231</td>
<td><strong>0.236</strong></td>
</tr>
</tbody>
</table>
Continued: Absolute and relative quality assessment

I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

a) Introduction
b) Stance and bias
c) Schemes and fallacies
d) Quality in theory
e) Absolute and relative quality assessment
f) Objective and subjective quality assessment
g) Conclusion
Relative quality comparison: Overview

- **Problem**
  - Rating the quality of an argument in isolation may be hard or even doubtful.
  - Is there an easier or more realistic way to assess quality?

- **Main idea**
  - Often, we are only interested in the best available argument.
  - Then, it’s enough to compare the quality of an argument to others.
  - **Dilemma.** Unclear in the end whether the best argument is good.

- **Existing approaches**
  - **Winning side.** Prediction of the debate winner from debate flow. (Zhang et al., 2016)
  - **Winning side.** Prediction of the winner from content and style (Wang et al., 2016)
  - **Convincingness.** Argument quality comparison with SVM and Bi-LSTM. (Habernal and Gurevych, 2016)
  - **Level of support.** Ranking of arguments by support of claim. (Braunstain et al., 2016)
Relative quality comparison: Dimensions covered here

Argumentation quality

- local acceptability
- global acceptability
- argument acceptability

- local relevance
- global relevance
- argument prominence

- local sufficiency
- global sufficiency
- argument relevance

- cogency
- reason-ability
- argumentation quality

- clarity
- local cogency
- global cogency

- effectiveness
- persuasiveness
- winning side

- appropriateness
- evaluability

- credibility
- argument strength
- convincingness

- arrangement
- organization

- amount of evidence
- sufficiency
- thesis clarity

- prompt adherence
- global coherence

- Habernal and Gurevych (2016)
- Persing and Ng (2013)
- Feng et al. (2014)
- Braunstain et al. (2016)
- Stab and Gurevych (2017)
- Rahimi et al. (2014)
- Park et al. (2015)
- Boltužić and Šnajder (2015)
- Wachsmuth et al. (2017a)
- Persing et al. (2010)
- Rahimi et al. (2015)
- Tan et al. (2016)
- Wang et al. (2016)
- Zhang et al. (2016)
Comparing quality based on debate flow (Zhang et al., 2016)

- **Task**
  - Given a full Oxford-style debate, which opponent wins?

- **Approach**
  - Mining of supporting points each side.
  - Modeling of the "conversational flow": When does a side puts forward own points, when does it attack opponent points?
  - Logistic regression classifier with features capturing the flow.

- **Data**
  - 108 Intelligence² debates (117 turns on average).
  - Winning side and audience feedback given.

- **Results**
  - **Accuracy.** Approach (0.65) beats audience feedback (0.6).
  - **Insights.** Attacking the opponent’s points better than focus on own points.
Comparing quality with SVM and Bi-LSTM (Habernal and Gurevych, 2016)

- **Task**
  - Given two arguments with the same topic and stance, which one is more convincing?

- **Supervised learning approaches**
  - **SVM.** SVM with RBF kernel using various linguistic features.
  - **Bi-LSTM.** Bi-directional long short-term memory neural network.
    Notice: The focus of the paper was not the approaches but the data construction.

- **Crowdsourced data**
  - 16,927 pairs of 1052 debate portal arguments for 32 topic-stance pairs.
  - Each annotated 5 times for convincingness (most reliable annotation taken).
    Reliability can be estimated with MACE (Hovy et al., 2013). Annotators also had to give reasons.

- **Results in 32-fold cross-validation**
  - **Accuracy.** SVM (0.78) beats Bi-LSTM (0.76). Human performance 0.93.
  - **Insights.** Surface features like capitalization easy, ”inverted“ sentiment hard.
Absolute vs. relative assessment ~ Theory vs. practice

- **Data representing theory** (Wachsmuth et al., 2017b)
  - Absolute expert ratings
  - Normative guidelines
  - 15 predefined quality dimensions

- **Data representing practice** (Habernal and Gurevych, 2016)
  - Relative lay comparisons
  - No guidelines
  - 17+1 resulting reason labels

**Empirical comparison of theory and practice** (Wachsmuth et al., 2017d)

- 736 argument pairs are available with ratings and labels.
- Compute Kendall’s $\tau$ correlations of all dimensions and reasons.
How different is assessment in theory and in practice?

**Selected insights**
- Convincing correlates most with **overall quality** (0.64).
- Generally high "correlations" between 0.3 and 1.0.
- Perfect: Global acceptability + attacking/abusive (1.0).
- Mostly very intuitive, such as clarity + unclear (0.91).
- Top **overall quality** for well thought through (mean score 1.8 of 3).
- Lowest **overall quality** for off-topic (mean score 1.1 of 3).
- Few unintuitive results, e.g., "only" 0.52 for credibility + no credible evidence.
- Local sufficiency + global sufficiency hard to separate.

**Conclusions**
- Theory and practice match more than expected.
- Theory can guide quality assessment in practice.
- Practice indicates what to focus on to simplify theory.
I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

Next section: Objective & subjective quality assessment

a) Introduction
b) Stance and bias
c) Schemes and fallacies
d) Quality in theory
e) Absolute and relative quality assessment
f) Objective and subjective quality assessment
g) Conclusion
The role of participants in argumentation (recap)

- **Author (or speaker)**
  - Argumentation is connected to the person who argues.
  - The same argument is perceived differently depending on the author.

  "The EU should allow rescue boats. Many innocent refugees will die if there are no rescue boats."

- **Reader (or audience)**
  - Argumentation often targets a particular audience.
  - Different arguments and ways of arguing work for different readers.

  "According to a recent UN study, the number of rescue boats had no effect on the number of refugees who try."

- **Questions**
  - May the assessment ignore the author/speaker? And the reader/audience?

  The author/speaker is unknown in some application scenarios, but rarely the reader/audience is.
Objective and subjective quality assessment

- **Subjectiveness of quality assessment**
  - Many dimensions are inherently subjective.
  - Quality depends on the subjective weighting of different aspects of an issue.
  - Also, it depends on preconceived opinions.

- **Example: Which argument is more relevant?**
  - "The death penalty legitimizes an irreversible act of violence. As long as human justice remains fallible, the risk of executing the innocent can never be eliminated."

  - "The death penalty doesn’t deter people from committing serious violent crimes. The thing that deters is the likelihood of being caught and punished."

- **Two ways to approach this problem** (both detailed below)
  - Either, focus on properties that can be assessed "objectively".
  - Or, include a model of the reader/audience in the quality assessment.
Objective quality assessment: Overview

- **Problem**
  - How to assess quality without learning from subjective annotations?
  - What are objective argumentation quality indicators?

- **Main idea**
  - Assess quality based on the structure induced by the set of all arguments.
  - Works for both for absolute and relative assessment.
  - **Dilemma.** Evaluation on subjective annotations? A solution is to rely on majority assessments of many annotators.

- **Existing approaches**
  - **Acceptability.** Assessment based on the attack relations. (Cabrio and Villata, 2012)
  - **Relevance.** Assessment based on reuse of units. (Wachsmuth et al., 2017a)
  - **Prominence.** Assessment based on argument frequency. (Boltužić and Šnajder, 2015)
Objective quality assessment: Dimensions covered here

Argumentation quality

- local acceptability
- global acceptability
- cogency
- reason-ability
- global relevance
- local relevance
- effectiveness
- credibility
- appropriateness
- emotional appeal
- arrangement
- validity
- prompt adherence
- global coherence
- evaluability
- thesis clarity
- amount of evidence
- sufficiency
- level of support
- argument acceptability
- argument prominence
- argument relevance
- organization
- argument strength
- persuasiveness
- winning side
- convincingness

References:
- Persing and Ng (2013)
- Feng et al. (2014)
- Rahimi et al. (2014)
- Stab and Gurevych (2017)
- Braunstain et al. (2016)
- Habernal and Gurevych (2016)
- Park et al. (2015)
- Persing et al. (2010)
- Wachsmuth et al. (2017a)
- Cabrio and Villata (2012)
- Boltužić and Šnajder (2015)
- Wachsmuth et al. (2017a)
Objective assessment based on attacks \cite{cabrio2012}

**Background: Abstract argumentation framework** \cite{dung1995}

- A directed graph where nodes represent arguments and edges attack relations between arguments.
- Graph analysis reveals whether to accept an argument.
  - **Accepted.** If all arguments attacking it are rejected.
  - **Not accepted.** If an accepted argument attacks it.

Extensions with weightings and with support+attack exist.

**Approach**
- Given a set of arguments, use textual entailment algorithm to classify attacks.
- Assess acceptability of arguments following Dung’s framework.

**Evaluation**
- Tested on 100 argument pairs from idebate.org, 45 attacking each other.
  - **Attack classification.** Accuracy 0.67
  - **Acceptability assessment.** Accuracy 0.75
Objective assessment based on reuse  (Wachsmuth et al., 2017a)

- **Task**
  - Given a set of arguments, which one is most relevant to some issue?
  - **Problem.** Relevance is highly subjective.

- **Research question**
  - Can we develop an "objective" measure of relevance?

- **Key hypothesis**
  - The relevance of a conclusion depends on what other arguments across the web use it as a premise.
  - **Rationale.** Author cannot control who "cites" a conclusion in this way.

- **Approach**
  - Ignore content and reasoning of arguments (for now).
  - Derive relevance structurally from the reuse of conclusions at web scale.
Objective assessment based on reuse: Argument graph

The death penalty doesn’t deter people from committing serious violent crimes.

A survey of the UN on the relation between the death penalty and homicide rates gave no support to the deterrent hypothesis.

Page et al. (1999)

”PageRank, a method for rating web pages objectively and mechanically, effectively measuring human interest “

Conclusion

Premises

The death penalty should be abolished.

It does not deter people from committing serious violent crimes.

Even if it did, is it acceptable to pay for predicted future crimes of others?
Objective assessment based on reuse: Approach

- **Original PageRank score** of a web page $d$ (Page et al., 1999)

$$p(d) = (1 - \alpha) \cdot \frac{1}{|D|} + \alpha \cdot \sum_i \frac{p(d_i)}{|D_i|}$$

- **Adapted PageRank score** of an argument unit $c$ (Wachsmuth et al., 2017a)

$$\hat{p}(c) = (1 - \alpha) \cdot \frac{p(d)}{|A|} + \alpha \cdot \sum_i \frac{\hat{p}(c_i)}{P_i}$$

- **Argument relevance** is aggregation of premise scores
  - Minimum, average, maximum, or sum
Objective assessment based on reuse: Data

- **No use of argument mining here**
  - Evaluation of PageRank without noise.

- **A ground-truth argument graph**
  - 57 argument corpora from [www.aifdb.org](http://www.aifdb.org).
  - Merged all arguments except for duplicates.
  - 17,877 arguments, 31,080 different units.
  - PageRank computed based on assumption that units match if they span the same text.

- **Benchmark rankings**
  - Since no objective relevance assessments exist, use average assessments as a proxy.
  - 110 arguments for 32 general claims.
  - 2–6 arguments per claim.
  - Ranked by seven annotators (mean Kendall’s $\tau = 0.36$, highest $\tau = 0.59$).
Objective assessment based on reuse: Examples

"Strawberries are the best choice for your breakfast meal."

"Berries are superfoods because they’re so high in antioxidants without being high in calories, says Giovinazzo MS, RD, a nutritionist at Clay health club and spa, in New York City."

"One cup of strawberries, for instance, contains your full recommended daily intake of vitamin C, along with high quantities of folic acid and fiber."

"Technology has enhanced the daily life of humans."

"The use of technology has revolutionized business."

"The internet has enabled us to widen our knowledge."

"Technology has given us a means of social interaction that wasn't possible before."
Objective assessment based on reuse: Results

- **Evaluation of unsupervised ranking approaches**

<table>
<thead>
<tr>
<th>PageRank of premises</th>
<th>Frequency of premises</th>
<th>Similarity of units</th>
<th>Sentiment of premises</th>
<th>Number of premises</th>
<th>Random ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{p} )</td>
<td>( \sum )</td>
<td>( c \sim P )</td>
<td>( \smiley )</td>
<td>(</td>
<td>P</td>
</tr>
</tbody>
</table>

  each for minimum, average, maximum, and sum aggregation

- **Experiment on ground-truth graph**

  - Rank arguments with each approach.
  - Correlate with benchmark rankings.

- **Results**

  - PageRank best (with sum aggregation).
  - Notable correlation despite ignorance of content and reasoning.

  best results for each ranking approach

<table>
<thead>
<tr>
<th>#</th>
<th>Approach</th>
<th>Kendall’s ( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PageRank</td>
<td>0.28</td>
</tr>
<tr>
<td>2</td>
<td>Number</td>
<td>0.19</td>
</tr>
<tr>
<td>3</td>
<td>Sentiment</td>
<td>0.12</td>
</tr>
<tr>
<td>4</td>
<td>Frequency</td>
<td>0.10</td>
</tr>
<tr>
<td>5</td>
<td>Similarity</td>
<td>0.02</td>
</tr>
<tr>
<td>6</td>
<td>Random</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Continued: Objective & subjective quality assessment

I. Introduction to computational argumentation

II. Basics of natural language processing

III. Basics of argumentation

IV. Argument acquisition

V. Argument mining

VI. Argument assessment

VII. Argument generation

VIII. Applications of computational argumentation

IX. Conclusion

- a) Introduction
- b) Stance and bias
- c) Schemes and fallacies
- d) Quality in theory
- e) Absolute and relative quality assessment
- f) Objective and subjective quality assessment
- g) Conclusion
Inclusion of Subjectivity: Overview

Problem
• Ultimately, effective argumentation requires to consider the target audience.
• Humans would barely argue without doing so.

Main idea
• Model the target audience within quality assessment.
• This also includes to have audience-specific ground-truth annotations.

Missing approaches
• Audience model have rarely been included explicitly so far.
• Implicitly, some annotated corpora may actually represent specific audiences.
• Recent studies analyze the quality perception of different audiences.

Studies
• Different personalities. Effectiveness of emotional vs. rational arguments.
  (Lukin et al., 2017)
• Different ideologies. Effectiveness of news editorials. (El Baff et al., 2018)
Effectiveness based on personality  (Lukin et al., 2017)

- **Hypothesis**
  - People with different personalities are open to different types of arguments.

- **Study**
  - Impact of personality on the effectiveness of emotional and factual arguments.
  - **Personality.** Here, the "Big Five".

- **Data**
  - 5185 arguments from online dialogs.
  - **Quality.** Each annotated for whether it changed the belief (to pro, to con, neither).
  - **Personality.** Each annotator did Big Five test.

- **Selected insights**
  - Agreeable people easiest to predict ($F_1 \sim .48$), extroverted hardest ($F_1 \sim .44$).
  - Factual arguments best for agreeable people, emotional best for open people.
Effectiveness based on ideology (El Baff et al., 2018)

- **Effects of news editorials**
  - News editorials are said to shape public opinion, but they rarely *change* a reader’s prior stance.
  - Rather, they challenge or reinforce stance — or neither.

- **Dialectical notion of argumentation quality**
  - A good editorial reinforces one side and challenges the other.
  - Or it challenges both sides.
Effectiveness based on ideology: Data

- **Hypothesis**
  - Prior stance depends on political ideology (and personality).
  - Ideology needs to be known to assess the effectiveness of news editorials.

- **Study**
  - Impact of ideology (and personality) on the effectiveness of news editorials.
  - **Ideology.** Here, conservative vs. liberal.

- **Data**
  - 1000 editorials from NYTimes.
  - **Quality.** Each annotated for persuasive effect by 3 conservatives and 3 liberals.
  - **Ideology.** All 24 annotators (in total) did the Political Typology Quiz.
  - **Personality.** Also, Big Five test was taken.

![Graph showing the distribution of annotations by ideological groups]

- Conserving
  - Core Conservatives
  - Country First Conservatives
  - Market Skeptic Republicans
  - New Era Enterprisers

- Liberals
  - Devout and Diverse
  - Disaffected Democrats
  - Opportunity Democrats
  - Solid Liberals

Argument Assessment, Henning Wachsmuth
Effectiveness based on ideology: Results

- Majority effect distribution in the corpus
  - Conservatives vs Liberals
  - Change stance
  - Strongly challenging: 33, 72; Somewhat challenging: 35, 71; No effect: 275, 269; Somewhat reinforcing: 1282, 708; Strongly reinforcing: 1402, 798
  - Challenge & Challenge: 1% Challenge & Reinforce: 5% Reinforce & Reinforce: 44% Challenge & No_effect: 2% Reinforce & No_effect: 38% No effect & No_effect: 10%

- Effect depending on ideology and personality
  - Kendall's $\tau$ correlation with challenge/reinforce
  - Agreeability: -0.14, 0.04, 0.06, 0.16, 0.26
  - Conscientiousness: -0.14, 0.04, 0.06, 0.16, 0.26
  - Extraversion: -0.14, 0.04, 0.06, 0.16, 0.26
  - Neuroticism: -0.14, 0.04, 0.06, 0.16, 0.26
  - Openness: -0.14, 0.04, 0.06, 0.16, 0.26

Argument Assessment, Henning Wachsmuth
Effect assessment depending on ideology (El Baff et al., 2020)

- **Task**
  - Given a news editorial and a reader’s ideology, predict the persuasive effect.

- **Approach**
  - SVM using five style feature types:
    - **LIWC.** Psyche-related words
    - **NRC.** Emotion/Sentiment words
    - **MPQA-S.** Subjective words
    - **MPQA-A.** Argumentative words
    - **ADUs.** Distribution of ADU types
  + Lemma n-grams for comparison to content

- **Data**
  - As above, 80% training, 20% test.

- **Results**
  - Only for liberals, significant micro F₁ gains over random baseline achieved.

<table>
<thead>
<tr>
<th>Features</th>
<th>Conserv.</th>
<th>Liberals</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIWC</td>
<td>0.26</td>
<td>0.40</td>
</tr>
<tr>
<td>NRC</td>
<td>0.29</td>
<td>0.39</td>
</tr>
<tr>
<td>MPQA-S</td>
<td>0.28</td>
<td>0.38</td>
</tr>
<tr>
<td>MPQA-A</td>
<td>0.29</td>
<td>0.41</td>
</tr>
<tr>
<td>ADUs</td>
<td>0.31</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>Best style set</strong></td>
<td>0.37</td>
<td>*0.49</td>
</tr>
<tr>
<td>Lemma n-grams</td>
<td><strong>0.38</strong></td>
<td>*0.49</td>
</tr>
<tr>
<td><strong>Best overall</strong></td>
<td>0.36</td>
<td><strong>0.54</strong></td>
</tr>
<tr>
<td>Random baseline</td>
<td>0.34</td>
<td>0.26</td>
</tr>
</tbody>
</table>

For liberals, style seems at least as discriminative as content.
Next section: Conclusion

I. Introduction to computational argumentation
II. Basics of natural language processing
III. Basics of argumentation
IV. Argument acquisition
V. Argument mining
VI. Argument assessment

| a) Introduction          |
| b) Stance and bias      |
| c) Schemes and fallacies|
| d) Quality in theory    |
| e) Absolute and relative quality assessment |
| f) Objective and subjective quality assessment |
| g) Conclusion           |

VII. Argument generation
VIII. Applications of computational argumentation
IX. Conclusion
Conclusion

- **Argument assessment**
  - Classification of issue-related subjectiveness properties.
  - Interpretation of the reasoning of an argument.
  - Judgment of several quality dimensions of an argument.

- **Subjectiveness and reasoning**
  - Stance, bias, argumentation schemes, fallacies, and more.
  - Stance classification is a major and extensively studied task.
  - Reasoning-related methods are still limited.

- **Argumentation quality**
  - Several dimensions are considered in theory and practice.
  - Absolute rating and relative comparison approaches exist.
  - Subjectiveness may be included or somewhat circumvented.
References


References


References

References


References


References


References


