Joint prediction in MST style discourse parsing for argumentation mining

Andreas Peldszus  Manfred Stede

Applied Computational Linguistics, University of Potsdam

EMNLP 20.09.2015
Outline

1. Argumentation Mining
2. Dataset & Scheme
3. Models
4. Results for the attachment task
5. Results for all tasks
Outline

1. Argumentation Mining
2. Dataset & Scheme
3. Models
4. Results for the attachment task
5. Results for all tasks
What is argumentation mining?

Health insurance companies should naturally cover alternative medical treatments. Not all practices and approaches that are lumped together under this term may have been proven in clinical trials, yet it's precisely their positive effect when accompanying conventional 'western' medical therapies that's been demonstrated as beneficial. Besides many general practitioners offer such counselling and treatments in parallel anyway - and who would want to question their broad expertise?
What is argumentation mining?

[e1] Health insurance companies should naturally cover alternative medical treatments.

[e2] Not all practices and approaches that are lumped together under this term may have been proven in clinical trials,

[e3] yet it’s precisely their positive effect when accompanying conventional ‘western’ medical therapies that’s been demonstrated as beneficial.

[e4] Besides many general practitioners offer such counselling and treatments in parallel anyway -

[e5] and who would want to question their broad expertise?
What is argumentation mining?

[1] Health insurance companies should naturally cover alternative medical treatments.

[2] Not all practices and approaches that are lumped together under this term may have been proven in clinical trials.

[3] Yet it's precisely their positive effect when accompanying conventional 'western' medical therapies that's been demonstrated as beneficial.

[4] Besides many general practitioners offer such counselling and treatments in parallel anyway -

[5] and who would want to question their broad expertise?

Tasks:

- EDU segmentation
- ADU segmentation resp. argumentative relevance
- ADU type classification
- Relation identification
- Relation type classification
Outline

1. Argumentation Mining
2. Dataset & Scheme
3. Models
4. Results for the attachment task
5. Results for all tasks
Dataset: argumentative microtexts

Texts:
- 112 texts: collected in a controlled text generation experiment
- with professional parallel translation to English
- annotated with argumentation structures
- see [Peldszus and Stede, to appear]
- freely available, CC-by-nc-sa license
- https://github.com/peldszus/arg-microtexts

Properties:
+ about 5 segments long
+ each segment is arg. relevant
+ explicit main claim
+ at least one possible objection considered
Freeman’s theory, revised & slightly generalized: [Freeman, 1991, 2011] [Peldszus and Stede, 2013]

- node types = argumentative role
  - proponent (presents and defends claims)
  - opponent (critically questions)

- link types = argumentative function
  - support own claims (normally, by example)
  - attack other’s claims (rebut, undercut)
Health insurance companies should naturally cover alternative medical treatments.

Not all practices and approaches that are lumped together under this term may have been proven in clinical trials, yet it’s precisely their positive effect when accompanying conventional ‘western’ medical therapies that’s been demonstrated as beneficial.

Besides many general practitioners offer such counselling and treatments in parallel anyway - and who would want to question their broad expertise?

From complex structures...
Health insurance companies should naturally cover alternative medical treatments.

Not all practices and approaches that are lumped together under this term may have been proven in clinical trials, yet it’s precisely their positive effect when accompanying conventional ‘western’ medical therapies that’s been demonstrated as beneficial.

Besides many general practitioners offer such counselling and treatments in parallel anyway - and who would want to question their broad expertise?

From complex structures... ... to simple support-attack graphs.
Tasks tackled in this paper:

- **attachment (at)**
  464 pairs yes, 2000 pairs no

- **central claim (cc)**
  112 yes, 451 no

- **role (ro)**
  451 proponent, 125 opponent

- **function (fu)**
  290 support, 174 attacks
Tasks tackled in this paper:

- Health insurance companies should naturally cover alternative medical treatments.
- Not all practices and approaches that are lumped together under this term may have been proven in clinical trials, yet it's precisely their positive effect when accompanying conventional 'western' medical therapies that's been demonstrated as beneficial.
- Besides many general practitioners offer such counselling and treatments in parallel anyway - and who would want to question their broad expertise?

- **attachment (at)**
  464 pairs yes, 2000 pairs no

- **central claim (cc)**
  112 yes, 451 no

- **role (ro)**
  451 proponent, 125 opponent

- **function (fu)**
  290 support, 174 attacks
Tasks tackled in this paper:

- **attachment (at)**
  464 pairs yes, 2000 pairs no
- **central claim (cc)**
  112 yes, 451 no
- **role (ro)**
  451 proponent, 125 opponent
- **function (fu)**
  290 support, 174 attacks
Outline

1. Argumentation Mining

2. Dataset & Scheme

3. Models

4. Results for the attachment task

5. Results for all tasks
Models: Overview

- Baseline 1
- Baseline 2
- Simple classifier 1
- Simple classifier 2
- Evidence graph 1
- Evidence graph 2
- MSTparser 1
- MSTparser 2
- MSTparser 3
- MSTparser 4
- MSTparser 5
Models: Baseline 1

BL-first: attach to first

- tendency to state the central claim first in English-speaking debating tradition
- covers convergent argumentation, but not serial argumentation
- in 50 of the 112 texts (44.6%) the first segment is the central claim
- 176 of the 464 relations (37.9%) attach to the first segment
Models: Baseline 1

BL-first: attach to first

- tendency to state the central claim first in English-speaking debating tradition
- covers convergent argumentation, but not serial argumentation
- in 50 of the 112 texts (44.6%) the first segment is the central claim
- 176 of the 464 relations (37.9%) attach to the first segment
Models: Baseline 2

BL-preced.: attach to preceeding

- strong baseline, usually hard to beat
  [Muller et al., 2012]
- covers serial argumentation, but not convergent argumentation
- 210 of all 464 relations (45.3%) attach to the preceding segment
Models: Baseline 2

BL-preced.: attach to preceding

- strong baseline, usually hard to beat
  [Muller et al., 2012]
- covers serial argumentation, but not convergent argumentation
- 210 of all 464 relations (45.3%) attach to the preceding segment
Models: simple classifier

Simple:
- pair-wise classification
- log-loss linear model, SGD training

Procedure:
- predict edge score
- apply classification threshold
Models: simple classifier

**Simple:**
- pair-wise classification
- log-loss linear model, SGD training

**Procedure:**
- predict edge score
- apply classification threshold
Models: simple classifier

**Simple:**
- pair-wise classification
- log-loss linear model, SGD training

**Procedure:**
- predict edge score
- apply classification threshold
Models: simple classifier

Simple:
- pair-wise classification
- log-loss linear model, SGD training

Procedure:
- predict edge score
- apply classification threshold
Models: simple classifier

Simple:
- pair-wise classification
- log-loss linear model, SGD training

Procedure:
- predict edge score
- apply classification threshold
Models: features

features from source and target segment:
- lemma unigrams (*) and bigrams
- first three lemma
- POS tags (*)
- main verb mood & tempus
- dependency parse triples
- discourse connectives & relations
  [Stede, 2002] (*)
- position in text

(*) extracted also from adjacent segment

features of source target pair:
- linerarity (forward, backward)
- absolut distance
- relative distance
Models: features

features from source and target segment:
- lemma unigrams (*) and bigrams
- first three lemma
- POS tags (*)
- main verb mood & tempus
- dependency parse triples
- discourse connectives & relations
  [Stede, 2002] (*)
- position in text

(*) extracted also from adjacent segment

features of source target pair:
- linerarity (forward, backward)
- absolut distance
- relative distance
Models: features

features from source and target segment:
- lemma unigrams (*) and bigrams
- first three lemma
- POS tags (*)
- main verb mood & tempus
- dependency parse triples
- discourse connectives & relations
  [Stede, 2002] (*)
- position in text

(*) extracted also from adjacent segment

features of source target pair:
- linerarity (forward, backward)
- absolut distance
- relative distance
Models: simple classifier

Simple:
- pair-wise classification
- log-loss linear model, SGD training

Procedure:
- predict edge score
- apply classification threshold

*Might not be a tree!?*
Models: simple classifier

Simple:
- pair-wise classification
- log-loss linear model, SGD training

Procedure:
- predict edge score
- apply classification threshold

Not a tree for 85% of the texts! Too many edges.
Models: simple classifier + MST-decoding

Procedure:

- predict edge score
- apply minimum spanning tree algorithm
Models: simple classifier + MST-decoding

Procedure:
- predict edge score
- apply minimum spanning tree algorithm
Models: simple classifier + MST-decoding

Procedure:
- predict edge score
- apply minimum spanning tree algorithm
Models: simple classifier + MST-decoding

Procedure:

- predict edge score
- apply minimum spanning tree algorithm

[Chu and Liu, 1965, Edmonds, 1967]
[McDonald et al., 2005b]
[Baldridge et al., 2007, Muller et al., 2012]
Models: simple classifier + MST-decoding

Procedure:
- predict edge score
- apply minimum spanning tree algorithm

[Chu and Liu, 1965, Edmonds, 1967]
[McDonald et al., 2005b]
[Baldridge et al., 2007, Muller et al., 2012]
Attachment classification: results for German
Attachment classification: results for German

<table>
<thead>
<tr>
<th></th>
<th>BL-first</th>
<th>BL-preced.</th>
<th>simple</th>
<th>simple+MST</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 macro</td>
<td>.618</td>
<td>.662</td>
<td>.679</td>
<td><strong>.688</strong></td>
</tr>
<tr>
<td>attach F1</td>
<td>.380</td>
<td>.452</td>
<td><strong>.504</strong></td>
<td>.494</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.236</td>
<td>.325</td>
<td>.365</td>
<td><strong>.377</strong></td>
</tr>
</tbody>
</table>

Average scores of 10 repetitions of 5-fold CV.
Attachment classification: results for English

<table>
<thead>
<tr>
<th></th>
<th>BL-first</th>
<th>BL-preced.</th>
<th>simple</th>
<th>simple+MST</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 macro</td>
<td>.618</td>
<td>.662</td>
<td>.663</td>
<td><strong>.674</strong></td>
</tr>
<tr>
<td>attach F1</td>
<td>.380</td>
<td>.452</td>
<td><strong>.478</strong></td>
<td>.470</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.236</td>
<td>.325</td>
<td>.333</td>
<td><strong>.347</strong></td>
</tr>
</tbody>
</table>

Average scores of 10 repetitions of 5-fold CV.
Why restrict to only one feature of the graph?

- Health insurance companies should naturally cover alternative medical treatments.
  - Not all practices and approaches that are lumped together under this term may have been proven in clinical trials.
  - Yet it's precisely their positive effect when accompanying conventional 'western' medical therapies that's been demonstrated as beneficial.
  - Besides many general practitioners offer such counselling and treatments in parallel anyway.
  - And who would want to question their broad expertise?

Argumentation aspects annotated in the graph:
- attachment (at)
- central claim (cc)
- role (ro)
- function (fu)
Models: Joint prediction in one evidence graph

Procedure:

- predict attachment probability
- predict role, function, cc probability
- combine predictions in one score
- apply MST algorithm
- identify control claim
- derive final role class for each segment
Models: Joint prediction in one evidence graph

Procedure:
- predict attachment probability
- predict role, function, cc probability
- combine predictions in one score
- apply MST algorithm
- identify central claim
- derive final role class for each segment
Models: Joint prediction in one evidence graph

Procedure:
- predict attachment probability
- predict role, function, cc probability
- combine predictions in one score
- apply MST algorithm
- identify central claim
- derive final role class for each segment
Models: Joint prediction in one evidence graph

Procedure:
- predict attachment probability
- predict role, function, cc probability
- combine predictions in one score
- apply MST algorithm
- identify central claim
- derive final role class for each segment
Models: Joint prediction in one evidence graph

Procedure:

- predict attachment probability
- predict role, function, cc probability
- combine predictions in one score
- apply MST algorithm
- identify central claim
- derive final role class for each segment
Models: Joint prediction in one evidence graph

Procedure:
- predict attachment probability
- predict role, function, cc probability
- combine predictions in one score
- apply MST algorithm
- identify central claim
- derive final role class for each segment
Models: Joint prediction in one evidence graph

Procedure:
- predict attachment probability
- predict role, function, cc probability
- combine predictions in one score
- apply MST algorithm
- identify central claim
- derive final role class for each segment
Models: Joint prediction in one evidence graph

\[
\begin{align*}
\text{cc:} & \quad \text{yes}=3, \text{no}=7 \\
\text{fu:} & \quad \text{sup}=4, \text{att}=6 \\
\text{ro:} & \quad \text{pro}=2, \text{opp}=8 \\
\end{align*}
\]

Joint prediction for argumentation mining
Models: Joint prediction in one evidence graph

How to move segment-wise predictions into the edge weights?
Models: Joint prediction in one evidence graph

\[ w_{i,j} = \frac{\phi_1 \text{RO} + \phi_2 \text{FU} + \phi_3 \text{CC} + \phi_4 \text{AT}}{\sum \phi_n} \]
Models: Joint prediction in one evidence graph

AT = at(1,2) = .6

Probability of attachment from source to target
Models: Joint prediction in one evidence graph

Probability of proper function per edge type

\[\text{FU} = \begin{cases} \text{fu}(1)(\text{sup}) = .7 \\ \text{fu}(1)(\text{att}) = .3 \end{cases}\]
Models: Joint prediction in one evidence graph

Probability of not being the central claim

CC = cc(1)(no) = .8
cc(1)(no) = .8
cc(1)(yes) = .2
Models: Joint prediction in one evidence graph

RO = \frac{\text{pro}(1) \times \text{pro}(2) + \text{opp}(1) \times \text{opp}(2)}{.9 \times .1 + .1 \times .9} = .18

RO = \frac{\text{pro}(1) \times \text{opp}(2) + \text{opp}(1) \times \text{pro}(2)}{.9 \times .9 + .1 \times .1} = .82

Probability of preserved/switched role for sup/att edges
Models: Joint prediction in one evidence graph

\[ w_{i,j} = \frac{\phi_1 \text{RO} + \phi_2 \text{FU} + \phi_3 \text{CC} + \phi_4 \text{AT}}{\sum \phi_n} \]

AT = .6       AT = .6
FU = .7       FU = .3
CC = .8       CC = .8
RO = .18      RO = .82
Models: Joint prediction in one evidence graph

**EG-equal:**
- all base-classifiers weighted equally

**EG-best:**
- optimize base-classifier weighting with a simple evolutionary search
Models: Joint prediction in one evidence graph

EG-equal:
- all base-classifiers weighted equally

EG-best:
- optimize base-classifier weighting with a simple evolutionary search
Attachment classification: results for German
### Attachment classification: results for German

<table>
<thead>
<tr>
<th></th>
<th>BL-first</th>
<th>BL-preced.</th>
<th>simple</th>
<th>simple+MST</th>
<th>EG equal</th>
<th>EG best</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 macro</td>
<td>.618</td>
<td>.662</td>
<td>.679</td>
<td>.688</td>
<td>.712</td>
<td>.710</td>
</tr>
<tr>
<td>attach F1</td>
<td>.380</td>
<td>.452</td>
<td>.504</td>
<td>.494</td>
<td>.533</td>
<td>.530</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.236</td>
<td>.325</td>
<td>.365</td>
<td>.377</td>
<td>.424</td>
<td>.421</td>
</tr>
</tbody>
</table>
## Attachment classification: results for English

<table>
<thead>
<tr>
<th></th>
<th>BL-first</th>
<th>BL-preced.</th>
<th>simple</th>
<th>simple+MST</th>
<th>EG equal</th>
<th>EG best</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F1 macro</strong></td>
<td>.618</td>
<td>.662</td>
<td>.663</td>
<td>.674</td>
<td>.692</td>
<td><strong>.693</strong></td>
</tr>
<tr>
<td><strong>attach F1</strong></td>
<td>.380</td>
<td>.452</td>
<td>.478</td>
<td>.470</td>
<td>.501</td>
<td><strong>.502</strong></td>
</tr>
<tr>
<td><strong>κ</strong></td>
<td>.236</td>
<td>.325</td>
<td>.333</td>
<td>.347</td>
<td>.384</td>
<td><strong>.386</strong></td>
</tr>
</tbody>
</table>
Models: Comparison with MSTParser pipelines

Background:
- 1-best MIRA structured learning for non-projective dep. parsing
  [McDonald et al., 2005a]
  [Baldridge et al., 2007]
- same feature sets

Procedure:
- predict edge score
- apply MST decoding
- apply internal or external relation labeller
- derive final role class for each segment
Models: Comparison with MSTParser pipelines

Background:

- 1-best MIRA structured learning for non-projective dep. parsing
  [McDonald et al., 2005a]
  [Baldridge et al., 2007]

- same feature sets

Procedure:

- predict edge score
- apply MST decoding
- apply internal or external relation labeler
- derive final role class for each segment
Models: Comparison with MSTParser pipelines

Background:
- 1-best MIRA structured learning for non-projective dep. parsing
  [McDonald et al., 2005a]
  [Baldridge et al., 2007]
- same feature sets

Procedure:
- predict edge score
- apply MST decoding
- apply internal or external relation labeller
- derive final role class for each segment
Models: Comparison with MSTParser pipelines

Background:

- 1-best MIRA structured learning for non-projective dep. parsing
  [McDonald et al., 2005a]
  [Baldridge et al., 2007]
- same feature sets

Procedure:

- predict edge score
- apply MST decoding
  - apply internal or external relation labeller
- derive final role class for each segment
Models: Comparison with MSTParser pipelines

Background:

- 1-best MIRA structured learning for non-projective dep. parsing
  [McDonald et al., 2005a]
  [Baldridge et al., 2007]
- same feature sets

Procedure:

- predict edge score
- apply MST decoding
- apply internal or external relation labeller
- derive final role class for each segment
Models: Comparison with MSTParser pipelines

Background:
- 1-best MIRA structured learning for non-projective dep. parsing
  [McDonald et al., 2005a]
  [Baldridge et al., 2007]
- same feature sets

Procedure:
- predict edge score
- apply MST decoding
- apply internal or external relation labeller
- derive final role class for each segment
Models: Comparison with MSTParser pipelines

**MP:**
- normal features
- internal relation labeler

**MP+p:**
- normal features plus base classifier predictions
- internal relation labeler

**MP+p+r:**
- normal features plus base classifier predictions
- external relation labeler
Models: Comparison with MSTParser pipelines

**MP:**
- normal features
- internal relation labeler

**MP+p:**
- normal features **plus** base classifier predictions
- internal relation labeler

**MP+p+r:**
- normal features **plus** base classifier predictions
- **external** relation labeler
Models: Comparison with MSTParser pipelines

**MP:**
- normal features
- internal relation labeler

**MP+p:**
- normal features **plus** base classifier predictions
- internal relation labeler

**MP+p+r:**
- normal features **plus** base classifier predictions
- **external** relation labeler
## Attachment classification: results for German

<table>
<thead>
<tr>
<th>BL-first</th>
<th>BL-preced.</th>
<th>simple</th>
<th>simple+MST</th>
<th>EG</th>
<th>equal</th>
<th>EG best</th>
<th>MP</th>
<th>MP+p</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 macro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.662</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.679</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.688</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.712</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.710</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.724</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.728</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>attach F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.504</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.494</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.533</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.530</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.553</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.559</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Attachment classification: results for German

<table>
<thead>
<tr>
<th></th>
<th>BL-first</th>
<th>BL-preced.</th>
<th>simple</th>
<th>simple+MST</th>
<th>EG equal</th>
<th>EG best</th>
<th>MP</th>
<th>MP+p</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 macro</td>
<td>.618</td>
<td>.662</td>
<td>.679</td>
<td>.688</td>
<td>.712</td>
<td>.710</td>
<td>.724</td>
<td>.728</td>
</tr>
<tr>
<td>attach F1</td>
<td>.380</td>
<td>.452</td>
<td>.504</td>
<td>.494</td>
<td>.533</td>
<td>.530</td>
<td>.553</td>
<td>.559</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.236</td>
<td>.325</td>
<td>.365</td>
<td>.377</td>
<td>.424</td>
<td>.421</td>
<td>.449</td>
<td>.456</td>
</tr>
</tbody>
</table>
## Attachment classification: results for English

<table>
<thead>
<tr>
<th></th>
<th>BL-first</th>
<th>BL-preced.</th>
<th>simple</th>
<th>simple+MST</th>
<th>EG equal</th>
<th>EG best</th>
<th>MP</th>
<th>MP+p</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 macro</td>
<td>.618</td>
<td>.662</td>
<td>.663</td>
<td>.674</td>
<td>.692</td>
<td>.693</td>
<td>.707</td>
<td><strong>.720</strong></td>
</tr>
<tr>
<td>attach F1</td>
<td>.380</td>
<td>.452</td>
<td>.478</td>
<td>.470</td>
<td>.501</td>
<td>.502</td>
<td>.524</td>
<td><strong>.546</strong></td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.236</td>
<td>.325</td>
<td>.333</td>
<td>.347</td>
<td>.384</td>
<td>.386</td>
<td>.414</td>
<td><strong>.440</strong></td>
</tr>
</tbody>
</table>
Tasks tackled in this paper:

- **attachment (at)**
  
  464 pairs yes, 2000 pairs no

- **central claim (cc)**
  
  112 yes, 451 no

- **role (ro)**
  
  451 proponent, 125 opponent

- **function (fu)**
  
  290 support, 174 attacks
## Results for all levels: German

<table>
<thead>
<tr>
<th></th>
<th>simple</th>
<th>EG equal</th>
<th>EG best</th>
<th>MP</th>
<th>MP+p</th>
<th>MP+p+r</th>
</tr>
</thead>
<tbody>
<tr>
<td>maF1</td>
<td>.679</td>
<td>.712</td>
<td>.710</td>
<td>.724</td>
<td><strong>.728</strong></td>
<td><strong>.728</strong></td>
</tr>
<tr>
<td>κ</td>
<td>.365</td>
<td>.424</td>
<td>.421</td>
<td>.449</td>
<td><strong>.456</strong></td>
<td><strong>.456</strong></td>
</tr>
</tbody>
</table>
### Results for all levels: German

<table>
<thead>
<tr>
<th></th>
<th>simple</th>
<th>EG equal</th>
<th>EG best</th>
<th>MP</th>
<th>MP+p</th>
<th>MP+p+r</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>at</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maF1</td>
<td>.679</td>
<td>.712</td>
<td>.710</td>
<td>.724</td>
<td><strong>.728</strong></td>
<td><strong>.728</strong></td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.365</td>
<td>.424</td>
<td>.421</td>
<td>.449</td>
<td><strong>.456</strong></td>
<td><strong>.456</strong></td>
</tr>
<tr>
<td><strong>cc</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maF1</td>
<td>.849</td>
<td>.879</td>
<td><strong>.890</strong></td>
<td>.825</td>
<td>.855</td>
<td>.855</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.698</td>
<td>.759</td>
<td><strong>.780</strong></td>
<td>.650</td>
<td>.710</td>
<td>.710</td>
</tr>
</tbody>
</table>
## Results for all levels: German

<table>
<thead>
<tr>
<th></th>
<th>simple</th>
<th>EG equal</th>
<th>EG best</th>
<th>MP</th>
<th>MP+p</th>
<th>MP+p+r</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>at</strong></td>
<td>maF1</td>
<td>.679</td>
<td>.712</td>
<td>.710</td>
<td>.724</td>
<td>.728</td>
</tr>
<tr>
<td></td>
<td>$\kappa$</td>
<td>.365</td>
<td>.424</td>
<td>.421</td>
<td>.449</td>
<td>.456</td>
</tr>
<tr>
<td><strong>cc</strong></td>
<td>maF1</td>
<td>.849</td>
<td>.879</td>
<td><strong>.890</strong></td>
<td>.825</td>
<td>.855</td>
</tr>
<tr>
<td></td>
<td>$\kappa$</td>
<td>.698</td>
<td>.759</td>
<td><strong>.780</strong></td>
<td>.650</td>
<td>.710</td>
</tr>
<tr>
<td><strong>ro</strong></td>
<td>maF1</td>
<td><strong>.755</strong></td>
<td>.737</td>
<td>.734</td>
<td>.464</td>
<td>.477</td>
</tr>
<tr>
<td></td>
<td>$\kappa$</td>
<td><strong>.511</strong></td>
<td>.477</td>
<td>.472</td>
<td>.014</td>
<td>.022</td>
</tr>
</tbody>
</table>
Results for all levels: German

<table>
<thead>
<tr>
<th></th>
<th>simple</th>
<th>EG equal</th>
<th>EG best</th>
<th>MP</th>
<th>MP+p</th>
<th>MP+p+r</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>at</strong></td>
<td>maF1</td>
<td>.679</td>
<td>.712</td>
<td>.710</td>
<td>.724</td>
<td>.728</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>.365</td>
<td>.424</td>
<td>.421</td>
<td>.449</td>
<td>.456</td>
</tr>
<tr>
<td><strong>cc</strong></td>
<td>maF1</td>
<td>.849</td>
<td>.879</td>
<td><strong>.890</strong></td>
<td>.825</td>
<td>.855</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>.698</td>
<td>.759</td>
<td><strong>.780</strong></td>
<td>.650</td>
<td>.710</td>
</tr>
<tr>
<td><strong>ro</strong></td>
<td>maF1</td>
<td>.755</td>
<td>.737</td>
<td>.734</td>
<td>.464</td>
<td>.477</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>.511</td>
<td>.477</td>
<td>.472</td>
<td>.014</td>
<td>.022</td>
</tr>
<tr>
<td><strong>fu</strong></td>
<td>maF1</td>
<td>.703</td>
<td>.735</td>
<td><strong>.736</strong></td>
<td>.499</td>
<td>.527</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>.528</td>
<td>.573</td>
<td>.570</td>
<td>.293</td>
<td>.326</td>
</tr>
</tbody>
</table>

Peldszus, Stede (Uni Potsdam)
### Results for all levels: German

<table>
<thead>
<tr>
<th></th>
<th>simple</th>
<th>EG equal</th>
<th>EG best</th>
<th>MP</th>
<th>MP+p</th>
<th>MP+p+r</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>at</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maF1</td>
<td>.679</td>
<td>.712</td>
<td>.710</td>
<td>.724</td>
<td><strong>.728</strong></td>
<td><strong>.728</strong></td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.365</td>
<td>.424</td>
<td>.421</td>
<td>.449</td>
<td><strong>.456</strong></td>
<td><strong>.456</strong></td>
</tr>
<tr>
<td><strong>cc</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maF1</td>
<td>.849</td>
<td>.879</td>
<td><strong>.890</strong></td>
<td>.825</td>
<td>.855</td>
<td>.855</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.698</td>
<td>.759</td>
<td><strong>.780</strong></td>
<td>.650</td>
<td>.710</td>
<td>.710</td>
</tr>
<tr>
<td><strong>ro</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maF1</td>
<td><strong>.755</strong></td>
<td>.737</td>
<td>.734</td>
<td>.464</td>
<td>.477</td>
<td>.669</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.511</td>
<td>.477</td>
<td>.472</td>
<td>.014</td>
<td>.022</td>
<td>.340</td>
</tr>
<tr>
<td><strong>fu</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maF1</td>
<td>.703</td>
<td>.735</td>
<td><strong>.736</strong></td>
<td>.499</td>
<td>.527</td>
<td>.723</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.528</td>
<td><strong>.573</strong></td>
<td>.570</td>
<td>.293</td>
<td>.326</td>
<td>.557</td>
</tr>
<tr>
<td><strong>avg</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maF1</td>
<td>.747</td>
<td>.766</td>
<td><strong>.768</strong></td>
<td>.628</td>
<td>.647</td>
<td>.744</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>.526</td>
<td>.558</td>
<td><strong>.561</strong></td>
<td>.352</td>
<td>.379</td>
<td>.516</td>
</tr>
</tbody>
</table>
## Results for all levels: English

<table>
<thead>
<tr>
<th></th>
<th>simple</th>
<th>EG equal</th>
<th>EG best</th>
<th>MP</th>
<th>MP+p</th>
<th>MP+p+r</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>at</strong></td>
<td>maF1</td>
<td>.663</td>
<td>.692</td>
<td>.693</td>
<td>.707</td>
<td><strong>.720</strong></td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>.333</td>
<td>.384</td>
<td>.386</td>
<td>.414</td>
<td><strong>.440</strong></td>
</tr>
<tr>
<td></td>
<td>cc</td>
<td>maF1</td>
<td>.817</td>
<td>.860</td>
<td>.869</td>
<td>.780</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ</td>
<td>.634</td>
<td>.720</td>
<td>.737</td>
<td>.559</td>
</tr>
<tr>
<td><strong>ro</strong></td>
<td>maF1</td>
<td>.750</td>
<td>.721</td>
<td>.720</td>
<td>.482</td>
<td>.475</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>.502</td>
<td>.445</td>
<td>.442</td>
<td>.024</td>
<td>.015</td>
</tr>
<tr>
<td><strong>fu</strong></td>
<td>maF1</td>
<td>.671</td>
<td>.707</td>
<td>.710</td>
<td>.489</td>
<td>.514</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>.475</td>
<td>.529</td>
<td>.530</td>
<td>.254</td>
<td>.296</td>
</tr>
<tr>
<td><strong>avg</strong></td>
<td>maF1</td>
<td>.725</td>
<td>.745</td>
<td>.748</td>
<td>.615</td>
<td>.635</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>.486</td>
<td>.520</td>
<td>.524</td>
<td>.313</td>
<td>.353</td>
</tr>
</tbody>
</table>
Impact of joint prediction
Impact of joint prediction

Simulations of better base classifiers for English (dashed levels artificially improved):

- **x**: number of predictions overwritten with ground truth
- **y**: average $\kappa$ score in 10 iterations of 5fold CV

Peldszus, Stede (Uni Potsdam)

Joint prediction for argumentation mining

EMNLP 2015 35 / 37
Contributions:

- First data-driven model optimizing argumentation structure **globally**.
- First model for argumentation mining **jointly** tackling segment type classification, relation identification and relation type classification.

That's it!

Checkout the corpus: [https://github.com/peldszus/arg-microtexts](https://github.com/peldszus/arg-microtexts)

Checkout some evaluations scripts, parameters and (soon) predictions: [https://github.com/peldszus/emnlp2015](https://github.com/peldszus/emnlp2015)
Contributions:

- First data-driven model optimizing argumentation structure **globally**.
- First model for argumentation mining **jointly** tackling segment type classification, relation identification and relation type classification.

That’s it!

- Checkout the corpus:  
  https://github.com/peldszus/arg-microtexts
- Checkout some evaluations scripts, parameters and (soon) predictions:  
  https://github.com/peldszus/emnlp2015


